UNIVERSITÀ DI SIENA

1240

DIPARTIMENTO DI BIOTECNOLOGIE, CHIMICA E FARMACIA

DOTTORATO DI RICERCA IN SCIENZE CHIMICHE E FARMACEUTICHE

CICLO XXXIII

COORDINATORE: PROF. MAURIZIO TADDEI

APPLICATION OF COMPUTATIONAL METHODS FOR THE DESIGN AND OPTIMIZATION OF SRC FAMILY KINASE INHIBITORS

SETTORE SCIENTIFICO-DISCIPLINARE: CHIM/08

TUTOR:

PROF. FABRIZIO MANETTI

CO-TUTOR:

DR. MATTIA MORI

DOTTORANDO:

RAFFAELE PASSANNANTI

ANNO ACCADEMICO: 2019-2020

This thesis is dedicated to Professor Maurizio Botta

ABSTRACT

Kinases constitute a family of enzymes involved in many cellular signalling pathways thanks to their ability to catalyse the transfer of a phosphate group from a donor, ATP or GTP, to an acceptor, a protein or a lipid. Src family kinases (SFKs), that represent the topic of this PhD work, belong to the cytosolic tyrosine kinases group and act by transferring the phosphate group to a tyrosine of the acceptor protein. SFKs play a key role in controlling cellular processes such as proliferation, apoptosis, angiogenesis, and invasion. Hyperactivation of SFKs was observed in many human cancers, which makes these enzymes very attractive targets for the treatment of these diseases.

In this PhD work, the application of molecular modelling techniques allowed the design and optimization of SFK inhibitors. In a first project, the ability of the pyrazolo[3,4-*d*]pyrimidine **ALM4-23** to stabilize c-Src in the three different conformational states, was evaluated by computational techniques. The resulting insights were used to design **ALM4-23** derivatives with potentially improved inhibitory activity against c-Src. In a second project, a structure-based virtual screening procedure was applied to a database of commercially available compounds to identify new c-Src inhibitors. The 3D structure of the kinase in an inactive conformation was used because potential advantages in terms of selectivity and drug residence time. In the last project, the pyrazolo[3,4-*d*]pyrimidine Fyn inhibitor, **Si308**, was studied to evaluate its ability to inhibit other SFK members, such as Lyn and Blk. Once confirmed the ability of **Si308** to act as a multitarget SFKs inhibitor, a library of its derivatives was synthesized and tested against B- and T-cell lymphoma cell lines, in which Fyn, Lyn and Blk play a pivotal role.

SUMMARY

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Kinases

Phosphorylation of proteins and lipids is at the centre of many cellular signalling pathways such as those responsible of cellular growth, differentiation, proliferation, and apoptosis. This reaction is promoted by kinases, a family of enzymes that catalyse the transfer of a phosphate group from a donor, generally adenosine triphosphate (ATP) or less frequently guanosine triphosphate (GTP), to an acceptor, a protein or a lipid. In the case of proteins as acceptors, the phosphate group is transferred to the hydroxyl group of a serine, or a threonine, or a tyrosine.[1] This allows the classification of these enzymes in protein tyrosine kinases and protein serine/threonine kinases.[2] Moreover, the acceptor can be represented by another molecule of the same protein that catalyses the reaction (autophosphorylation process) or by a different protein. In the case of lipids as acceptors, phosphorylation occurs at the polar heads of membrane lipids.

The deregulation of these enzymes makes them the protagonists of many diseases such as cancer, diabetes, cardiovascular and neurodegenerative disorders. In this regard, many kinase inhibitors are clinically used and many others are in clinical trials for the treatment of these diseases.[1]

The human genome encodes for 518 protein kinases[1,3] and about 20 lipid kinases (FIGURE 1). Sphingolipid kinases and phosphoinositide kinases belong to the class of lipid kinases, in which an important role is played by the phosphatidylinositol-3-kinases (PI3Ks) of class IA, able to phosphorylate in position 3 the inositol of phosphatidylinositol-4,5-bisphosphate (PIP2) and allow the obtaining of the triphosphate derivative PIP3, which in turns permits the binding of many proteins at the cell membrane and consequently the transduction of the signal of many receptor tyrosine kinases (RTKs). Several mutations in PI3Ks are associated with tumors, such as colon and breast cancer.

The 518 protein kinases are in turn classified into 478 eukaryotic protein kinases (ePKs), and in 40 atypical protein kinases (aPKs). This first classification is based on the absence of sequence similarity between the kinase domains of the members of these two groups. Nevertheless, aPKs have catalytic activity. Examples of aPKs are pyruvate dehydrogenase kinase, bromodomain kinases, bcr and the phosphatidylinositol-3-kinase-related kinases (PIKKs). In the latter family one of the most representative member is the mammalian target of

rapamycin (mTOR), a serine/threonine kinase expressed in pathways that regulate gene expression, proliferation, motility and cell growth.

The ePKs are further classified according to the sequence similarity of the kinase domains into eight groups:

- **Tyrosine kinase (TK) group** contains RTKs and cytosolic TKs. Among RTKs there are human epidermal growth factor receptor (HER/EGFR) family, insulin receptor (IR) and insulin-like growth factor 1 receptor (IGF1-R), platelet-derived growth factor receptors (PDGFRs) and fibroblast growth factor receptors (FGFRs), all involved in the progression of several tumors and therefore targets of many inhibitors that are presently used for their treatment. Src, Abl and JAK kinase families belong to the cytosolic TKs group.[1] The first one will be discussed in the following chapters. Regarding the second family, the main member is represented by Abl kinase which can participate to Bcr-Abl fusion protein, primary target of chronic myelogenous leukemia (CML), as well as pharmacological target of imatinib (Gleevec), the first clinically approved kinase inhibitor.[1,4] JAK family kinases are activated by the interaction with RTKs, and can regulate gene expression.

- **Tyrosine kinase like (TKL) group** consisting of both receptor and cytosolic kinases. The members of this group have similar sequences to those of the TK group. They are mainly serine/threonine kinases. In this group there are interleukin-1 receptor-associated kinase (IRAK), RAF kinases, LIM domain kinase (LIMK), and transforming growth factor beta (TGFβ) receptors.

- **STE group** consisting of three families: STE20 (MAP4K), STE11(MAP3K) and STE7 (MAP2K). Among the best-known members there are p21-activated kinases (Paks), whose overexpression was reported in various forms of cancer.

- **CK1 group** consists of casein kinase 1 (CK1) and homologues. These are serine/threonine kinases involved in the phosphorylation of substrates that in turn control gene expression and cytoskeleton function.

- **AGC group** whose name is derived from the families PKA, PKG and PKC. Akt family (PKB), which is part of this group, is particularly known for its role in the regulation of many cellular processes.

- **CAMK group** or Ca^{2+}/calmodulin-dependent kinases, are proteins that under basal conditions are self-inhibited and are activated following the binding of Ca^{2+}/calmodulin complex. Examples of proteins belonging to this group are the CHK1 and CHK2 kinases, which are involved in processes such as cell cycle blocking and DNA repair.

- **CMGC group** includes cyclin-dependent kinases (Cdks) that regulate cell cycle progression, mitogen-activated protein kinases (MAPKs) that regulate processes such as proliferation, differentiation and apoptosis, glycogen synthase kinases (GSKs) that participate in the inflammatory process, and Cdk-like kinases.

- **Receptor guanylyl cyclase (RGC) group** is the least populated group and involves enzymes that do not possess some of the fundamental residues for the phosphorylation process (pseudokinases). All enzymes in this group convert GTP to cyclic GMP.

- There are also 83 ePKs that have a poor sequence similarity with the eight groups just described. Among these we have CK2 which has the ability to use both GTP and ATP as a phosphate group donor and IKKs which are involved in the inflammation process.[1]
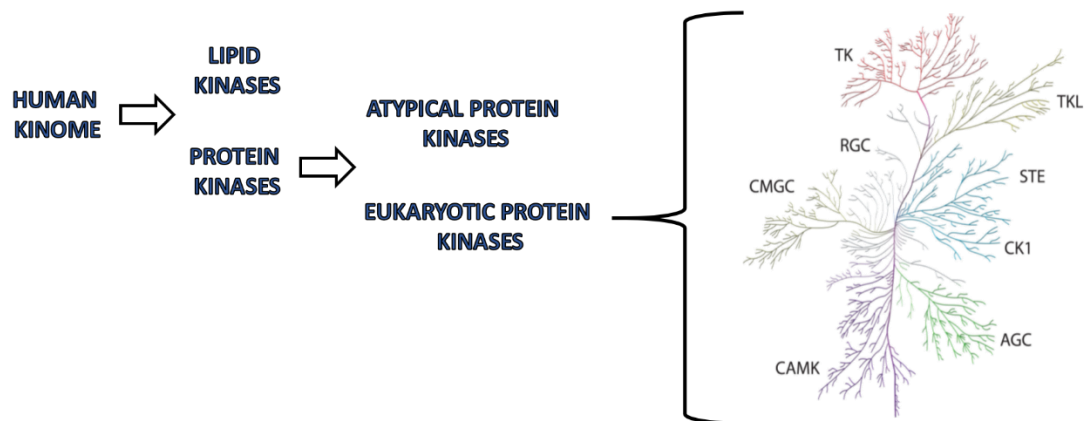
Figure 1. *Human kinome (adapted from* [1]*) .*

1.1.1 Src family kinases

1.1.1.1 Members, structure, and activation process

SFKs belong to the cytosolic TKs group.[5,6] Studies on this family began with the identification of v-Src, the transforming protein of oncogenic chicken retrovirus, Rous sarcoma virus (RSV). v-Src is the viral homologue of the cellular protein c-Src,[2,5,7] the best studied member of its family,[7] which in turn includes other ten proteins having high sequence similarity with it: Fyn, Yes, Yrk, Blk, Fgr, Hck, Lck, Lyn and Frk subfamily proteins Frk/Rak and Iyk/Bsk.[5] These enzymes are able to regulate cellular events such as proliferation, differentiation, survival, adhesion and migration due to their ability to interact with different cell receptors and intracellular targets.[5,8]

SFK members can be classified into three groups according to their district of expression:

- c-Src, Fyn and Yes are expressed ubiquitously.
- Blk, Fgr, Hck, Lck and Lyn are mainly expressed in hematopoietic cells. However, Lck and Lyn have been found also in neurons.
- Frk subfamily proteins are expressed principally in epithelial-derived cells.

These kinases are localized intracellularly in different places, e.g. c-Src was found in focal adhesions and endosomes, while other members like Fgr and Frk were found in the nucleus. In addition, for c-Src, Fyn and Lyn were found several isoforms.

From a structural point of view, SFKs are constituted by six functional regions (FIGURE 2).
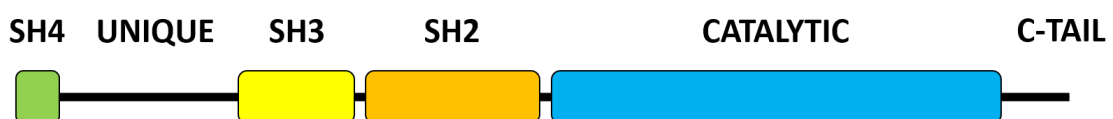
Figure 2. *SFKs domains (adapted from* [5]*).*

Going from the N-terminal to C-terminal end, there are respectively: SH4 (Src homolgy) domain, unique domain, SH3 domain, SH2 domain, catalytic or kinase domain, and C-terminal negative regulatory tail:[5]

- **SH4 domain** consists of approximately 15 amino acid residues in the various family members. It is characterized by a glycine residue in position 2 (Frk is an exception) that, after removal of the first amino acid, is myristoylated. The myristoylation allows the anchorage of these enzymes to the cell membranes. Moreover, in SH4 domain there are also cysteine residues (c-Src and Blk are an exception) that undergo to

palmitoylation,[5,6,9] an event which could reinforce the binding to the membranes, as well as, increases the affinity of the protein in question for particular intracellular targets. In this domain there is also a good percentage of basic residues, which could contribute to binding with the cell membrane, probably establishing electrostatic interactions with negatively charged phospholipids.

- **Unique domain** which has between 40 and 70 amino acids in the various family members.[6] As the name suggests, this domain is the least conserved in SFKs,[5,6] and together with the previous one, has not undergone to an extensive structural analysis over the years.[9] Its function is still unclear, but it is thought to be important for interaction with receptors or proteins that are specific for each family member.[5,9] Another hypothesis regarding the functionality of the unique domain is that it can act as a simple spacer keeping the rest of the protein away from the membrane.[6]

  In addition, serine and threonine residues which are subject to reaction of phosphorylation have been identified in the unique domain of c-Src and Lck. However, the specific function of these phosphorylated residues has not yet been clarified.[5,6]

- **SH3 domain** consists of approximately 50 amino acid residues in the various family members.[5] This domain is able to bind proline-rich sequences.[5,6] The folding of this domain is well known and is constituted by a β-barrel architecture containing five β-sheets and two loops (FIGURE 3). These loops are constituted by hydrophobic residues and represent the recognition site for proline-rich sequences. SH3 domains preferentially bind the PxxP motif, which in complex with SH3 domains, adopts a polyproline type II helical conformation.[9] This domain is fundamental for intra and intermolecular interactions that regulate the catalytic activity of SFKs.

- **SH2 domain** consists of approximately 100 amino acid residues[2] and it is also crucial for intra and intermolecular interactions that govern SFKs activity.[5] SH2 domain is able to bind amino acid sequences containing a phosphotyrosine.[5,6] Structural studies shown that this domain consists of a central portion made by β-sheets, with an adjacent α-helix on each side (FIGURE 3). These structural elements, together with connecting loops, form two pockets, one that coordinates the phosphotyrosine and another that binds the hydrophobic residues C terminal to it. The phosphotyrosine recognition pocket is highly conserved within the SH2 domains of SFKs and contains an arginine residue that forms electrostatic interactions with phosphorylated tyrosine. C-terminal recognition pocket is less conserved among family members, reflecting the differences between SH2 domains in the recognition of specific sequences with a phosphorylated tyrosine residue.

Binding studies of a phosphopeptide library have established that the SH2 domains of SFKs bind preferentially the sequences with the pY-E-E-I motif.[9]

- **Catalytic or kinase domain** represents most of the C-terminal region, as well as the domain responsible of catalytic activity.[5,6] SFKs have a kinase domain with a bilobate structure that is characteristic of all protein kinases. The N-terminal or small lobe is constituted by five β-sheets and a single α-helix named helix αC, which, as we will see later, is involved in the SFKs regulatory mechanism. The C-terminal or large lobe is mainly composed of α-helices and contains the regulatory activation loop (A loop),[9] where there is a conserved tyrosine residue which, once autophosphorylated, allows the assumption of the active conformation in the SFKs (FIGURE 3).[2,6,9]

- **C-terminal negative regulatory tail** consisting of about 20 amino acids in the various family members. Among these, there is a conserved tyrosine whose phosphorylation negatively regulates the catalytic activity of SFKs (FIGURE 3).[6] This reaction of phosphorylation is performed by Src-specific kinase Csk or its counterpart Chk.[9]
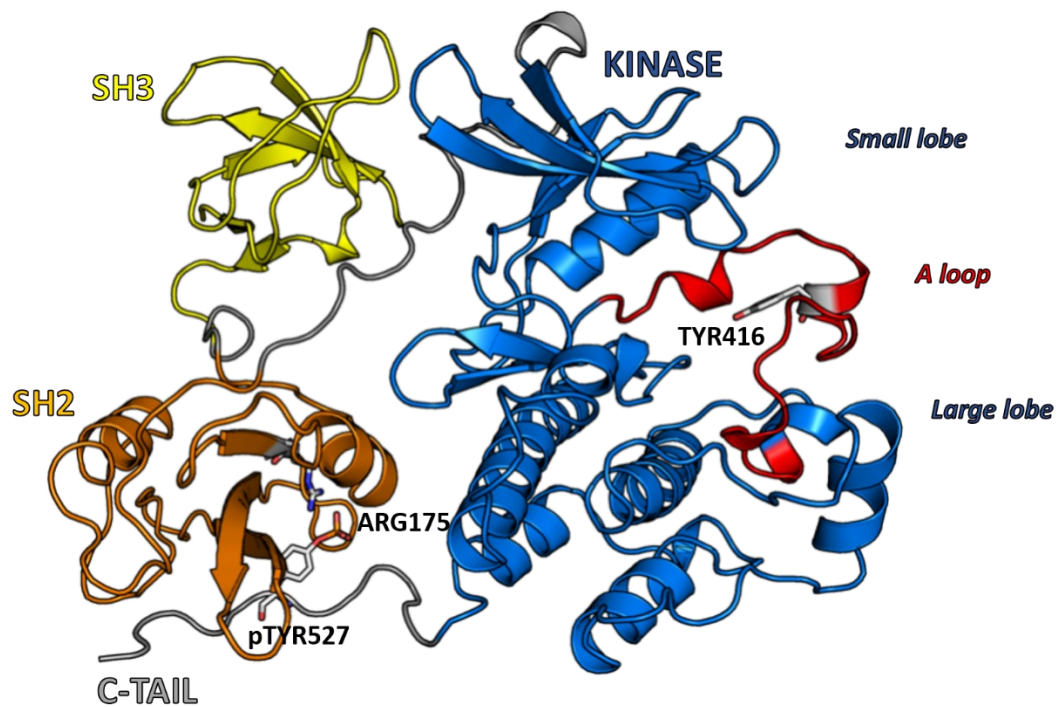
Figure 3. 3D structure of c-Src in autoinhibited conformation (PDB ID: 2SRC).

In vivo, SFKs are phosphorylated either to the tyrosine of the activation loop (in their activated state) or to the tyrosine of the C-terminal end (in their self-inhibited state).[9] The latter state is favoured in basal conditions.[2,5]

The SFKs activation process (using c-Src enumeration) is schematically reported in FIGURE 4. In the self-inhibited state of these kinases (left panel), the tyrosine of the C-terminal tail

(Tyr527) is phosphorylated and both SH3 and SH2 are placed behind the catalytic domain.[5] Specifically, as shown in FIGURE 3, the phosphorylated tail binds SH2 domain, while SH3 binds the loop that connects SH2 domain with the catalytic one. Therefore, the enzymatic inactivation of SFKs does not involve the coverage of the catalytic pocket by SH2 or SH3 domain, but it is the assembly of these two behind it that leads to the formation of an inhibitory folding to the A-loop, named A-loop helix. The formation of A-loop helix is able to:

- Produce the displacement of helix αC from the catalytic pocket. This event leads to the removal from the active site of a glutamic acid residue crucial for the catalytic process, Glu-310.
- Interfere with the binding of peptide substrates.
- Protect the Tyr-416 of the A-loop from phosphorylation.

The overall result is the inactivation of these enzymes.[10] It is important to say that in the self-inhibited conformation, intramolecular sequences bound to SH3 and SH2 domains do not represent 'high-affinity' binding partners. For example, in c-Src the linker segment bound to SH3 domain does not have the PxxP motif, and the C-terminal phosphorylated tail bound to SH2 domain has a glycine instead of an isoleucine in pY+3 position. The sub-optimal intramolecular interactions that hold together the self-inhibited conformation of SFKs, make these enzymes easily activable.[9] In fact, the displacement of SH2 and/or SH3 domains, either by dephosphorylation of the C-terminal tail through tyrosine phosphatase, or by competitive binding of optimal SH2/SH3 ligands [5,10] allows the kinase domain to stretch the A-loop, and expose the Tyr-416 for phosphorylation (central panel).[10] Finally, the phosphorylation of the Tyr-416 (right panel), which occurs as a result of an autophosphorylation process,[2,10] permits a conformational reorganization of the whole A-loop, which removes the steric barrier for binding of the protein substrate and allows the C-helix to return into the active site, reconstructing the catalytic machinery.[10]
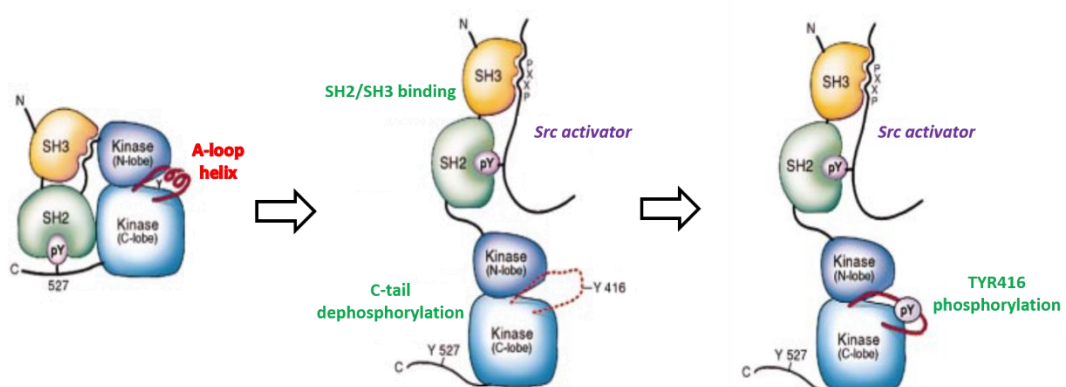
Figure 4. *Activation process of SFKs (adapted from [10]).*

1.1.1.2 c-Src kinase

c-Src was subjected to deep investigations in last decades due to its involvement in the development and progression of numerous types of cancer.[11] As previously mentioned, c-Src is expressed ubiquitously even if several studies indicate that the levels of this protein are 5-200 times higher in platelets, neurons and osteoclasts. In neurons, two specific isoforms containing 6 or 11 amino acid insertions at the SH3 domain have been identified. However, to date, there is no evidence of a functional difference between the different isoforms.

The human c-Src gene encodes for a protein of 536 amino acids, whereas the chicken gene encodes for a protein of 533 amino acids.[2] The two proteins have a sequence identity of 99.6% [11] and the three additional amino acids of the human protein are near the N-terminal end. Historically, the chicken protein numbering system is the most widely used in literature, even when studies concern the human protein.

c-Src is located intracellularly in different compartments, e.g. in fibroblasts, is bound to endosomes, perinuclear membranes, secretory vesicles and to the cytoplasmic side of the cell membrane where it can interact with a wide range of enzymes.

From a physiological point of view, c-Src plays a central role in processes such as cell differentiation, survival and proliferation. For example, the stimulation of the epidermal growth factor receptor (EGFR), which leads to cell proliferation, involves the activation of c-Src. This enzyme is also involved in the regulation of other events such as cell adhesion and cell motility.[2] The two main cellular structures that regulate these last two functions are focal adhesions and adherens junctions (FIGURE 5).
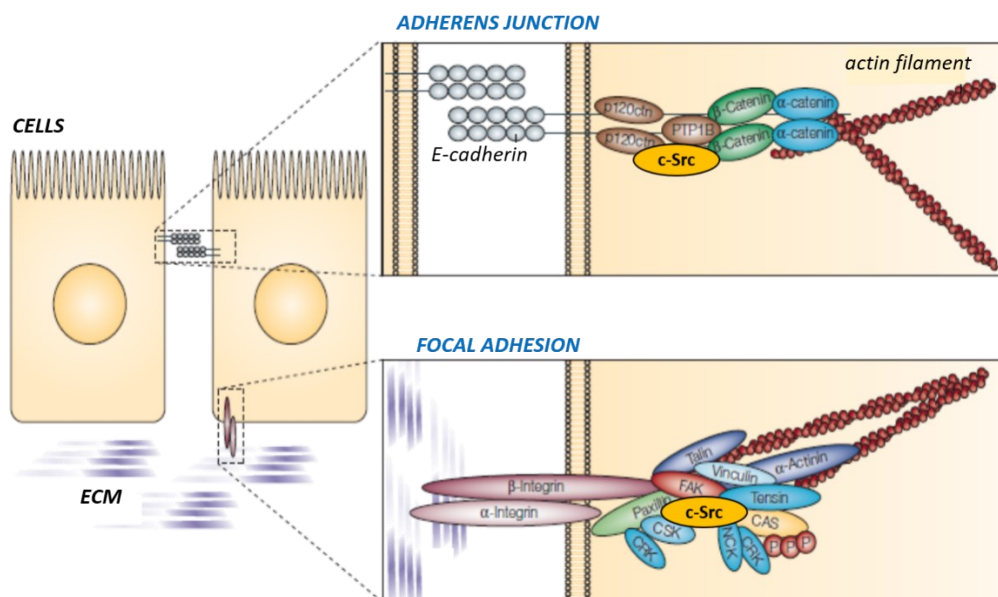
Figure 5. *Focal adhesions and adherens junctions (adapted from [12]).*

Focal adhesions allow the adhesion of the cell with the proteins of extracellular matrix (ECM), and participate in intracellular signalling pathways that regulate gene expression and proliferation. In focal adhesions, the heterodimers of α and β integrin bind with their extracellular domains the proteins of the ECM, while with cytoplasmic domains bind a complex consisting of more than 50 proteins, which has the function of associating the integrins with the actin filaments of the cytoskeleton. Several molecules including c-Src can associate to this complex, which when activated is able to stimulate the turnover of focal adhesions, promoting in this way cell motility. On the contrary, when motility is not required, Csk protein associates to this complex inactivating c-Src via phosphorylation of the C-terminal tail.

Cell-cell adhesion is ensured by other structures known as adherens junctions that involve the binding of E-cadherin molecules present on two adjacent cells. E-cadherin molecules are intracellularly connected with a complex constituted by p120 catenin, β-catenin and α-catenin, which allows the binding to actin filaments. c-Src can associate with this complex and when activated (generally by C-terminal dephosphorylation operated by the tyrosine phosphatase PTP1B) is able to stimulate the destruction of the adherent junctions, promoting in this way cell motility.[12]

Deregulation of the abovementioned cellular events is generally at the basis of the development and progression of neoplasms. In this context, hyperactivation of c-Src, caused by its over-expression or direct increase of its specific activity, has been identified in several tumors such as colon rectal cancer, breast cancer, ovarian cancer, lung cancer, prostate cancer, neuroblastoma, glioblastoma, as well as, hepatocellular, pancreatic, gastric, and oesophageal malignancies.[2,7,12–14] In tumors, hyperactivation of this kinase can be due to (FIGURE 6):

- Reduced expression of Csk, as evidenced in hepatocellular carcinoma.[7,12]
- Hyperexpression of phosphatase such as PTP1B, as evidenced in breast cancer.
- Hyperexpression of non-receptor tyrosine kinase such as FAK, which binding to SH2 and SH3 domains of c-Src, is responsible of the displacement of self-inhibiting intramolecular interactions.[12]
- Hyperexpression or hyperactivation of growth-factor receptors (EGFR, PDGFR, ErbB2, FGFR, colony-stimulating factor 1 or CSF-1R and hepatocyte growth factor or c-Met), as evidenced in tumors such as breast cancer, colon rectal cancer, prostate cancer, glioblastoma, neuroblastoma, etc., which interacting with SH2 domain of c-Src, displace self-inhibiting intramolecular interactions.[11,12]
- Deregulation of ubiquitin-proteasome pathway which lead to increased protein levels.

In addition, c-Src hyperactivation can be triggered by the development of point mutations, which have been identified only in some cancers such as colon and endometrial cancer.
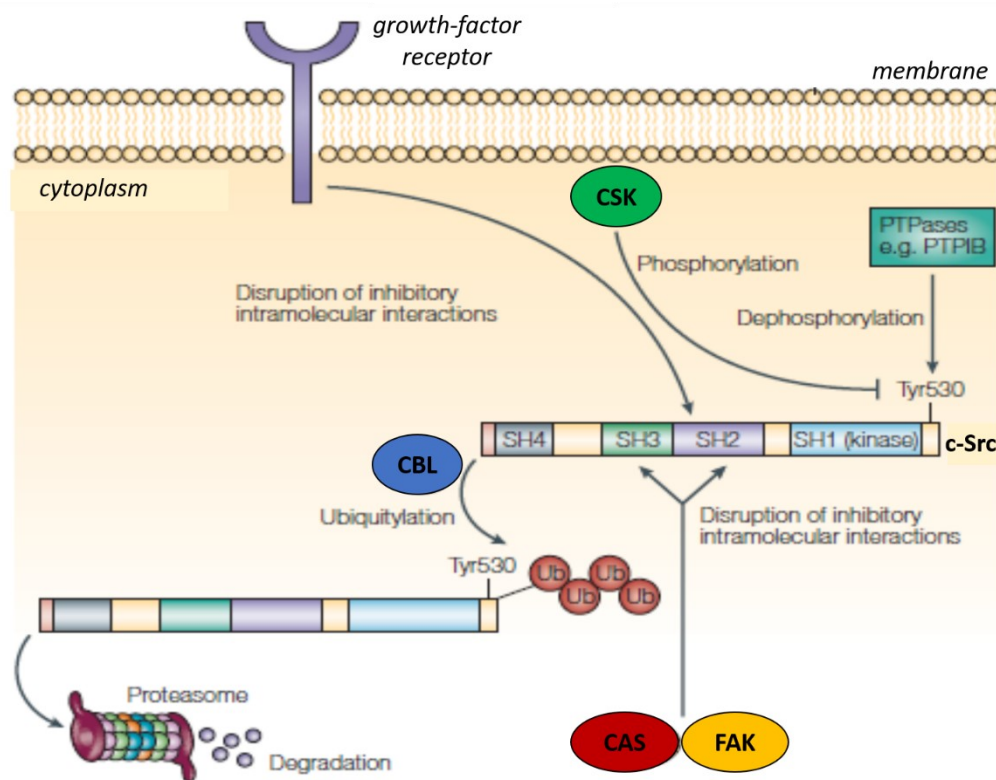
Figure 6. Cellular mechanisms which regulate c-Src activity (adapted from [12]).

c-Src determines the behaviour of tumor cells through interaction with various binding partners (FIGURE 7). Specifically, it promotes the increase of processes of motility and invasiveness (which in turn determine the development of metastatic phenotype) through:

- Phosphorylation of E-cadherins, which induces ubiquitylation and subsequent endocytosis of these molecules. The result is the loss of adherent junctions between adjacent cells.

- Phosphorylation of FAK, which then phosphorylates substrates such as paxillin, Cas and the p190-RHOGAP complex. These can subsequently induce changes to the actin filaments of the cytoskeleton, which result in destruction of the focal adhesions. c-Src leads to similar changes to cytoskeleton also independently from FAK, through its interaction with p120 catenin and cortactin.

- Phosphorylation of RRAS which inhibits the functionality of the integrins.

- Phosphorylation of FAK also stimulates the signalling pathway of c-JUN N-terminal kinase (JNK) which can induce the expression of MMP2/9 metalloproteinases, that are responsible of an increased degradation of ECM proteins.

Moreover, c-Src promotes the process of tumor angiogenesis, activating STAT3 which causes an increase in the expression of vascular endothelial growth factor (VEGF).[12] In addition, the activation of STAT3 induces the expression of the antiapoptotic protein Bcl-x$_L$, as well as the

expression of other genes that stimulate cell cycle progression, contributing in this way to oncogenesis. Furthermore, c-Src also promotes tumor proliferation by interacting with the Ras/MAPK pathway.[13]
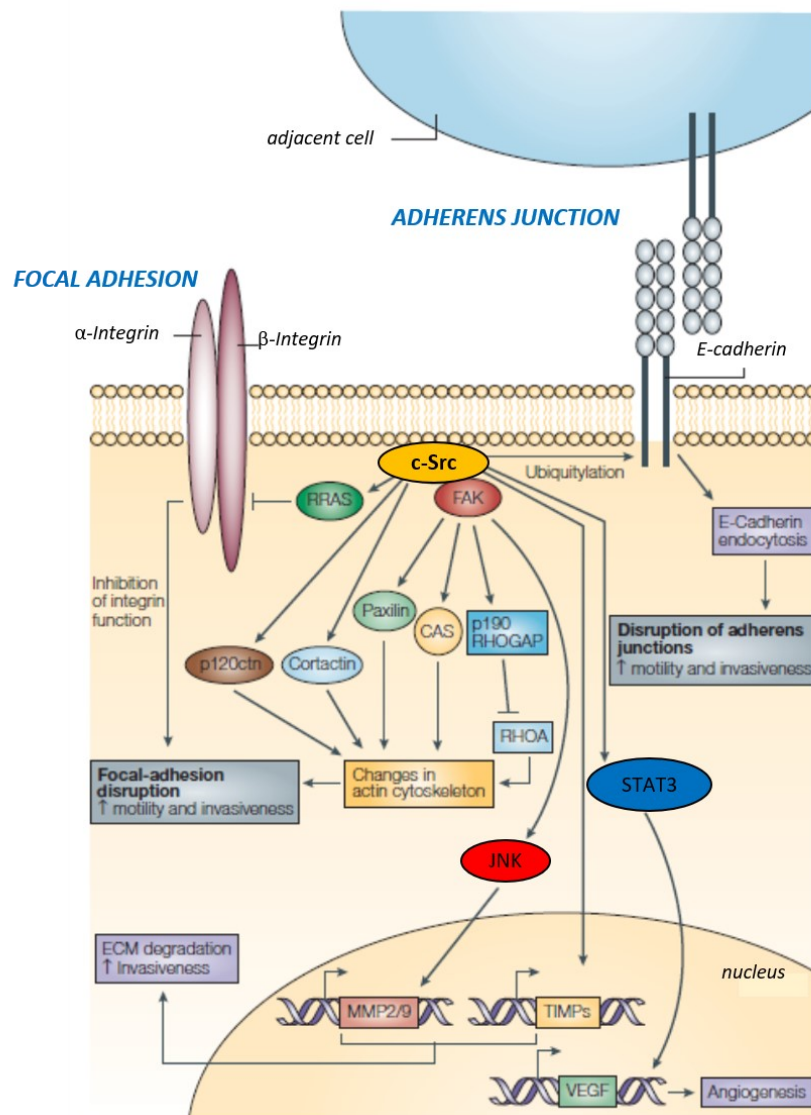
Figure 7. *Effects of c-SRC on tumour-cell behaviour (adapted from [12]).*

1.1.1.3 Src family kinases and Lymphomas

Lymphomas are neoplasms that involve the lymphoreticular system[15] and are classified based on the cell of origin in B-cell, T-cell, and natural killer (NK-) cell lymphomas. The velocity of growth of these neoplasms determines the distinction in indolent lymphomas, such as Follicular lymphoma and cutaneous T-cell lymphoma (CTCL), and aggressive lymphomas, such as diffuse large B-cell lymphoma (DLBCL) and peripheral T-cell lymphoma (PTCL).[16]

Histological analysis has enabled the most common classification in non-Hodgkin lymphomas (NHLs) and Hodgkin lymphomas (HLs),[17] the first ones representing 90% of all malignant lymphomas, while the second ones the remaining 10%. Recent studies showed a dramatic growth in the diagnosis of NHLs, surpassed only by lung cancer in women and malignant melanoma in both men and women.[18] The main risk factors for NHLs include hereditary immunodeficiency disorders, states of strong immunosuppression (HIV/AIDS, organ transplantation), infectious agents (Epstein-Barr Virus, human T-cell lymphotrophic virus-1, hepatitis C virus, Helicobacter pylori),[19] and autoimmune disorders (rheumatoid arthritis, Sjögren's syndrome, systemic lupus erythematosus myositis, Hashimoto's thyroiditis, celiac disease/dermatitis herpetiformis).[20]

In lymphomas derived from B- or T-cells the deregulation of the signalling pathways downstream to the B- or T-cell receptors is frequently observed.[21,22] In these deregulated pathways, different members of SFKs play a critical role, as demonstrated by several studies:

- Simultaneous inhibition with masitinib of the three SFKs Lyn, Fyn and Blk leads to the complete block of B-cell receptor (BCR) oncogenic signals in germinal center B-cell (GCB) and activated-B cell (ABC) DLBCL cells. The inhibition was effective not only on lymphoma cell lines but also in patient-derived xenograft models.[23] Moreover, SFKs inhibition also represents a strategy to overcome the mechanisms of resistance to signalling inhibitors currently in clinical use for the treatment of aggressive B-cell lymphomas, such as the Btk inhibitor ibrutinib.[23,24]

- Activating mutation of Fyn (Tyr531His) was found in PTCL and the treatment with dasatinib of cells expressing the mutant, reduces cell growth.[25] Moreover, in CTCL, another form of T-cell lymphoma, the involvement of Blk was reported,[26] an outcome that was further confirmed by the inhibition of CTCL proliferation *in vitro* and *in vivo* by dasatinib.[27,28]

1.1.1.4 Src family kinase inhibitors on market

Protein kinases represent one of the main pharmacological target of the last 20 years, with 48 targeted drugs approved by the Food and Drug Administration (FDA) (data updated to March 2019). Most of these compounds are administered orally with the only exception of netarsudil and temsirolimus which are administered as eye drops and intravenously respectively. Of these 48 drugs, 25 are inhibitors of receptor tyrosine kinases, 10 inhibitors of non-receptor tyrosine kinases, and 13 inhibitors of serine/threonine kinases.

All FDA approved kinase inhibitors are directed toward malignancies except 7 inhibitors (baricitinib, fostamatinib, ruxolitinib, nintedanib, sirolimus, tofacitinib, and netarsudil) which are used to treat non-cancer diseases.[29]

Most of these compounds are able to inhibit multiple members of the kinase family, including SFKs. Among them (FIGURE 8):

- Bosutinib is a Bcr-Abl, Src, Lyn, Hck, Kit, and PDGFR inhibitor and has been approved for treatment of CML and ALL (acute lymphoblastic leukemia).
- Dasatinib is an inhibitor of Bcr-Abl, Src, Lck, Fyn,Yes, PDGFR, and other kinases. It has been approved for treatment of CML.
- Ponatinib is an inhibitor of Bcr-Abl, SFKs, PDGFR, VEGFR and other kinases. It has been approved for treatment of CML and ALL.
- Vandetanib is an inhibitor of EGFR, SFKs, VEGFR, RET and other kinases. It has been approved for treatment of medullary thyroid carcinoma.[11]
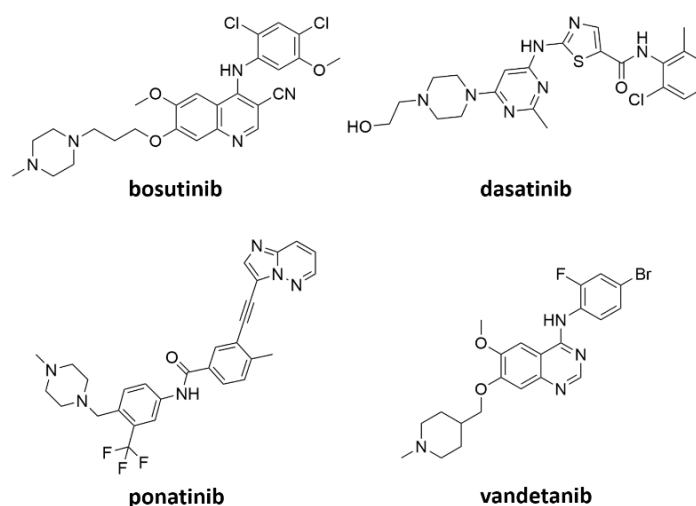
Figure 8. Chemical structures of SFKs inhibitors on market.

1.1.2 Conformational plasticity of protein kinases

Protein kinases are enzymes able to adopt at least two extreme functional conformations: one active and one inactive.[30] The most remarkable structural differences between these conformations occur at the kinase domain, which consists of a highly conserved bilobate folding, where the small lobe, rich in β sheets, is connected through the hinge loop to the large lobe, rich in α helixes. The active site, where ATP binds, is located between these two lobes (FIGURE 9). In this domain, the structurally important elements for catalysis are represented by:

- Glycine-rich P loop, the conserved β3 lysine, and the conserved glutamate in helix αC (FIGURE 9), which all help to coordinate the ATP phosphate groups.

- A loop which allows the binding of the peptide substrate. In addition, at its N-terminal end is located the conserved tripeptide motif Asp-Phe-Gly or DFG (FIGURE 9) able through DFG-Asp to coordinate Mg^{2+},[3] which in turn coordinates the β and γ phosphate groups of ATP.

- Conserved aspartate in the large lobe of non-receptor protein tyrosine kinases which acts as a catalytic base (FIGURE 9), deprotonating the tyrosine of peptide substrate and facilitating in this way its nucleophilic attack to the γ- phosphorus atom of ATP.[11]
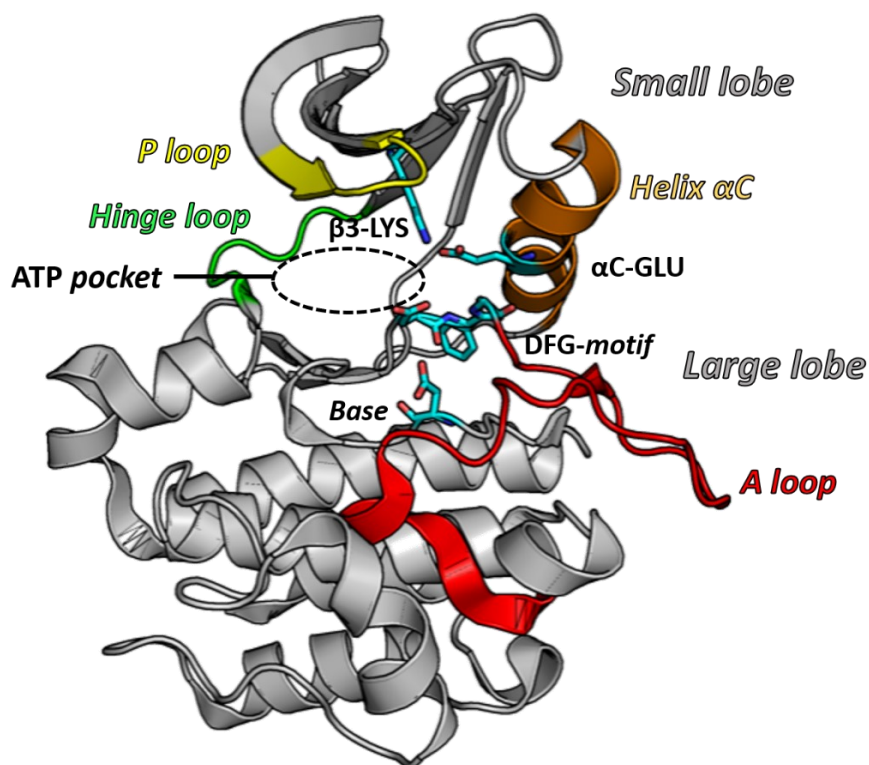
Figure 9. *Kinase domain (1Y57): structural elements necessary for catalysis.*

1.1.2.1 The active conformation

The active conformation of protein kinases is characterized by several structural peculiarities within the kinase domain: (i) orientation of DFG-Asp in the active site (DFG-Asp-in) where it can coordinate the ATP-Mg^{2+} complex (ii) salt bridge between β3 lysine and helix αC glutamate,[3,31] (iii) extended conformation of activation loop which facilitates the binding of the peptide substrate.[11,31]

In addition, recent studies led to the identification in the kinase domains of two conserved networks of hydrophobic amino acids called respectively regulatory (R) spine and catalytic (C) spine which are assembled in the active conformation of all protein kinases (FIGURE 10).[3] The C-spine consists of eight non-consecutive hydrophobic residues and is completed by the adenine ring of ATP while R-spine consists of four non-consecutive hydrophobic residues. Functionally, these spines play a pivotal role, in fact when these are assembled, all the catalytically important residues are properly positioned and ready for catalysis.[32]
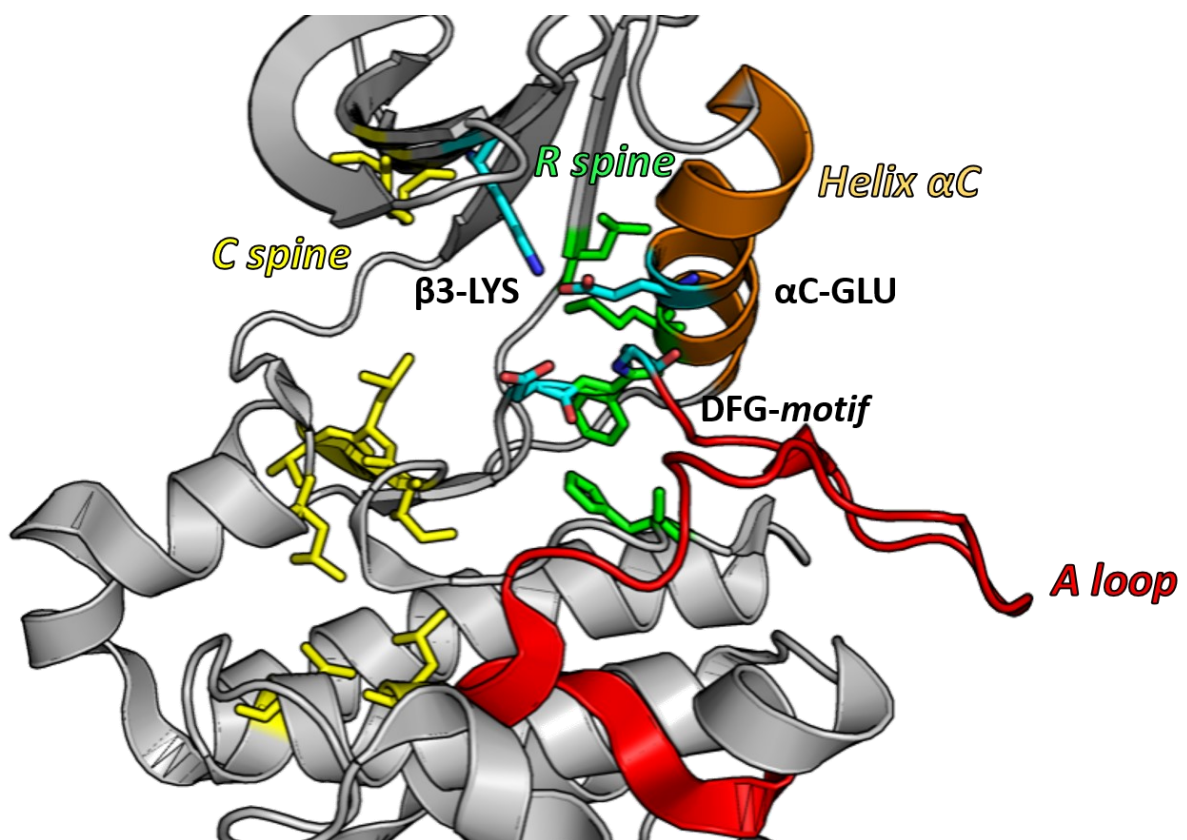
Figure 10. Kinase domain in active conformation (1Y57).

1.1.2.2 The inactive conformations

While all kinases resemble each other in their active conformation, these enzymes can become inactive assuming different conformations.[3] The best known are the Src/Cdk-like and Abl/c-Kit-like inactive conformations also named C-helix out (CHO) and DFG-out inactive conformations.

Regarding the CHO inactive conformation (corresponding to previous described self-inhibited state of SFKs), it is characterized by significant structural differences within the kinase domain in respect to the active conformation: the activation loop is folded into a conformation which does not allow the binding of the peptide substrate. The folding of the activation loop determines also an outward rotation of the helix αC (responsible of the loss of the salt bridge between the helix αC glutamate and β3 lysine)[3,31] and the disassembly of the R spine (FIGURE 11);[11,32,33] the overall result of such rearrangements is the enzymatic inactivation.
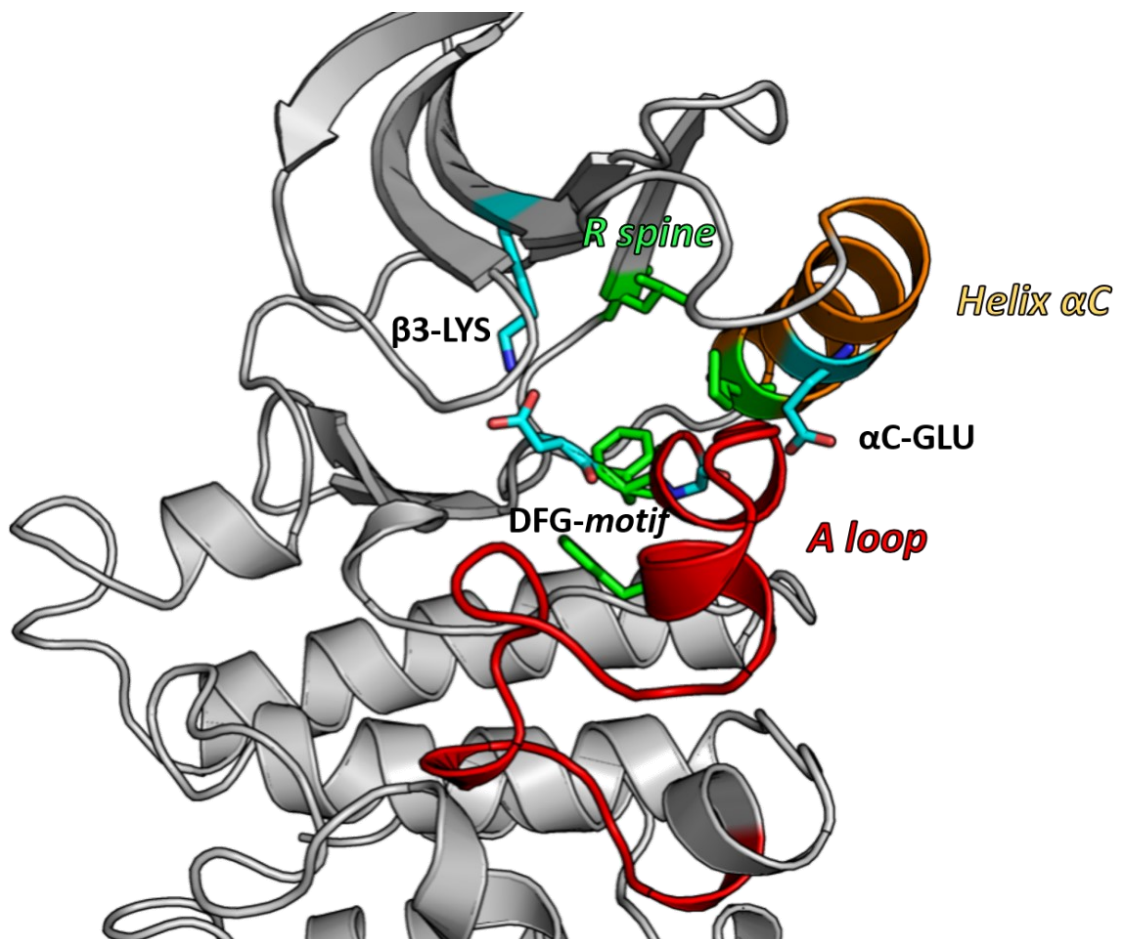
Figure 11. *Kinase domain in CHO inactive conformation (2SRC).*

Moreover, the outward rotation of the helix αC enlarges the ATP binding site in respect to the active conformation, creating an additional binding pocket under the loop connecting the helix αC and the β3 sheet of the small lobe (FIGURE 12A-12B).[3,31]
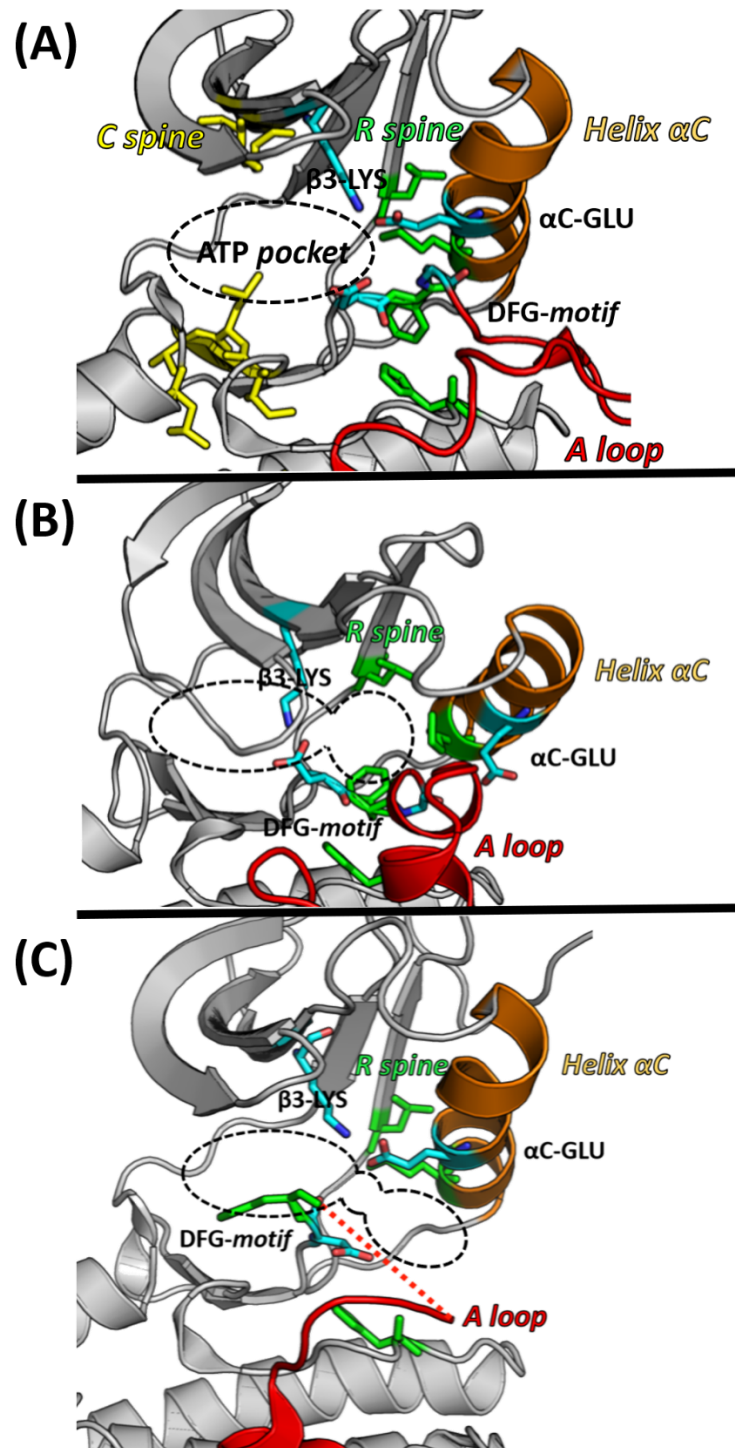
Figure 12. ATP pocket in the kinase domain of the active conformation (A), and of the CHO (B) and DFG-out (C) inactive conformations.

Regarding the DFG-out inactive conformation, compared with the active one, it is characterized within the kinase domain by the conservation of the salt bridge between β3 lysine and the helix αC glutamate,[3,31] and by a 180° rotation of the DFG motif. Again, the result is the enzymatic inactivation because DFG-Asp is not directed towards the active site where it is replaced by DFG-Phe. The disposition assumed by DFG motif produces the disassembly of the R spine and the subsequent formation of an additional hydrophobic pocket in ATP binding site,[3] named DFG pocket (FIGURE 12C). Moreover, also in this inactive conformation, the activation loop is able to assume a conformation that is not suitable for binding of peptide substrate (FIGURE 13).
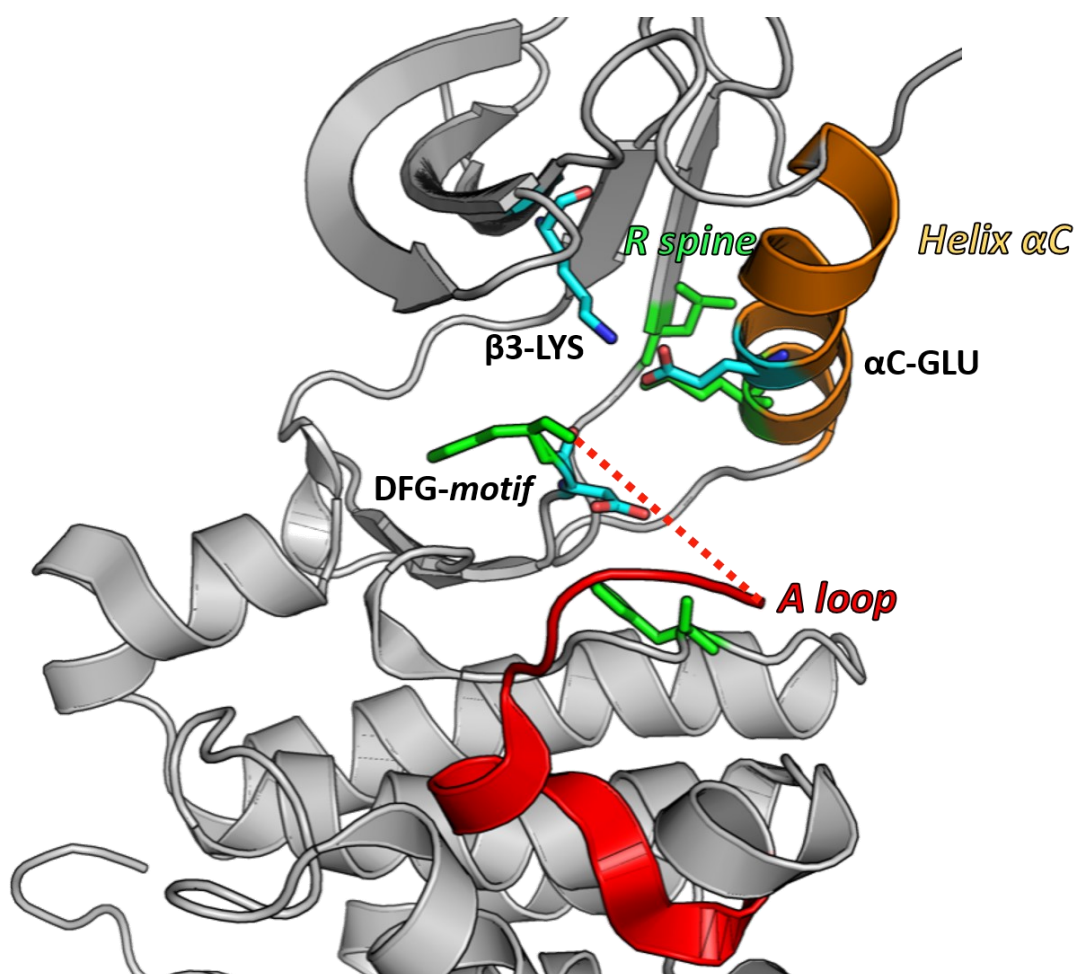
Figure 13. *Kinase domain in DFG-out inactive conformation (3G6G).*

Recent structural studies have identified, for some kinases, a third inactive conformation called CHO/DFG-out, having the features of both inactive conformations described above.[31]

1.1.2.3 Factors affecting conformational stability, type of kinase inhibitors and selectivity issue

The stability of a conformation can be influenced by several factors which can lead to a change in its population and consequently determine the overall kinase activity. Best known factors are binding of regulatory proteins and substrates, aminoacidic mutations and events such as activation loop phosphorylation. Inhibitors can also perturb the dynamic ensemble of kinase conformations, stabilizing a specific one. Based on the stabilized conformation of the target kinase, inhibitors are classified into:

- Type-I inhibitors, which compete with ATP for the active conformation. Examples are dasatinib, gefitinib and erlotinib (FIGURE 14).[3]
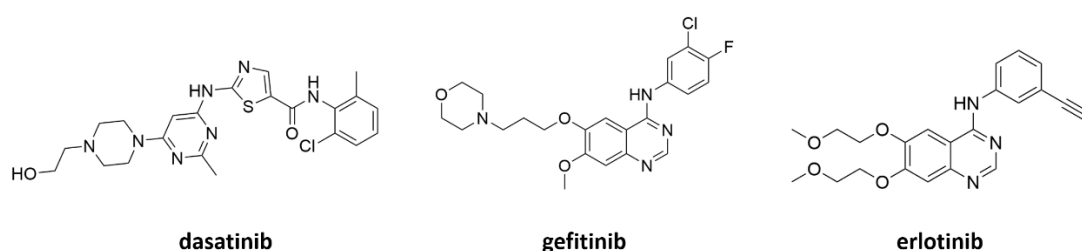
Figure 14. *Examples of Type-I inhibitors.*

- Type-I½ inhibitors, which stabilize the CHO inactive conformation binding the ATP binding pocket and the additional pocket characterizing this conformation. Lapatinib and vemurafenib are examples of Type- I½ inhibitors (FIGURE 15).[31]
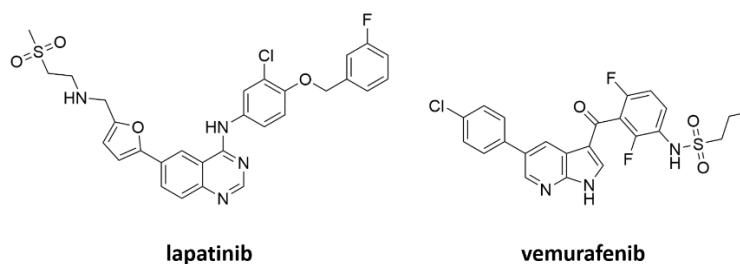
Figure 15. *Examples of Type-I½ inhibitors.*

- Type-II inhibitors, which stabilize the DFG-out inactive conformation binding the ATP binding pocket and the additional allosteric pocket which forms as result of DFG motif flip. Imatinib, sorafenib and nilotinib are examples of type-II inhibitors[3,31,34] (FIGURE 16).
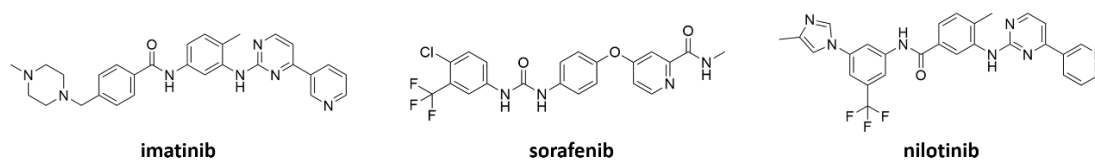
Figure 16. Examples of Type-II inhibitors.

- Type-III inhibitors, which stabilize the DFG-out inactive conformation but, differently from Type II inhibitors, only bind to the additional allosteric pocket of this conformation.[3,34] Type III inhibitors generally serve as starting points for the development of potent Type II inhibitors (FIGURE 17).[34]
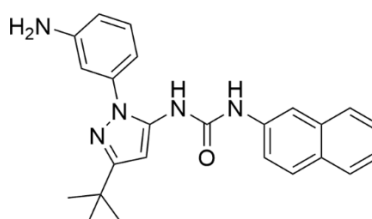
RL38

Figure 17. Example of Type-III inhibitor.

- Type-IV inhibitors, which stabilize an inactive conformation binding to allosteric sites that are distant from the active site. An example of a Type-IV inhibitor is GNF-2, which binding to Abl myristate pocket, is able to stabilize the kinase in DFG-out inactive conformation (FIGURE 18).
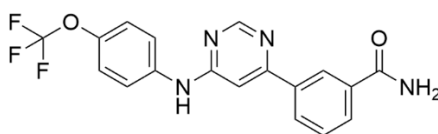
GNF-2

Figure 18. Example of Type-IV inhibitor.

Traditionally, the development of kinase inhibitors is focused on ATP-competitive small molecules (Type-I inhibitors) which, however, are affected by selectivity issues because of the high conservation of the ATP binding pocket in the kinome. Moving away from classical ATP-competitive inhibitors and targeting potentially less conserved sites, which for example become accessible when the kinase adopts an inactive conformation, can represent a solution to the selectivity problem as well as add other pharmacological advantages such as increased drug

target residence time.[35] In a recent study, for example, a series of selective c-Src macrocyclic peptide inhibitors were identified. These macrocycles act as Type-I½ inhibitors binding the ATP binding pocket and the additional pocket which is exposed in the CHO inactive conformation. Further studies have shown that the 60-80 times higher selectivity of these compounds for c-Src in respect to strictly related Hck kinase, can be partly attributed to this additional pocket. In effect, in c-Src, this pocket is lined with a leucine (Leu297), while in Hck this residue is replaced by a methionine (Met297) which reduces the size of the pocket not permitting the binding of macrocycles. Another example is represented by the exceptionally selective MEK1/2 kinase inhibitor PD318088 that stabilizes the CHO inactive conformation of the target kinase binding only the additional pocket. Similarly, targeting the allosteric pocket of the DFG-out conformation allows type-II inhibitors to form additional interactions (to those made with the ATP binding pocket) with potentially less conserved sites. Imatinib, prototype of Type-II inhibitor, is selective for its pharmacological target.[3]

2 AIM OF THE RESEARCH

Hyperactivation of SFKs was observed in many human cancers and given their crucial role in controlling cell proliferation, apoptosis, angiogenesis, and invasion, SFKs are considered attractive targets for the treatment of such diseases.[36,37] Based on these considerations, the aim of this PhD work was the design and optimization of SFK inhibitors. Specifically, it is divided in three projects:

- In the first project described in section three, the application of several computational methods allowed the evaluation of the mode of action of a pyrazolo[3,4-*d*]pyrimidine c-Src inhibitor and the design of its derivatives with potentially improved inhibitory activity against c-Src.

- In the second project described in section four, a structure-based virtual screening (SBVS) procedure was performed on a database of commercially available compounds with the aim to identify new Type-II c-Src inhibitors.

- In the third project described in section five, the application of docking and molecular dynamics (MD) simulations coupled with enzymatic assays allowed us to elucidate the ability of the pyrazolo[3,4-*d*]pyrimidine Fyn inhibitor **Si308** to act also as Lyn and Blk inhibitor. Once confirmed the multitarget SFKs inhibition, a small library of **Si308** derivatives was synthesized with the aim of identifying new compounds active on B- and T-lymphoma cell lines in which Fyn, Blk and Lyn play a critical role.

3 ALM4-23: EVALUATION OF ITS MODE OF ACTION AGAINST C-SRC AND DESIGN OF DERIVATIVES ENDOWED WITH POTENTIALLY IMPROVED INHIBITORY ACTIVITY

3.1 State of the art

Hyperactivation of c-Src plays a key role in the proliferation and invasiveness of many cancers.[38,39] Consequently, in the last 20 years, many small molecule c-Src inhibitors were developed starting from different chemical scaffolds such as purines, anilinoquinazolines, quinolinecarbonitriles, benzotriazines, and thiazole-carboxamides. In this context, the research group of Professor Maurizio Botta works from long time to the design of c-Src inhibitors endowed with pyrazolo[3,4-*d*]pyrimidine scaffold.[39] Many examples of pyrazolo[3,4-*d*]pyrimidine inhibitors of TKs were reported in literature. Among them PP1 and PP2 (FIGURE 19) are very strong and selective inhibitors of SFKs.

PP1: R = Me
PP2: R = Cl
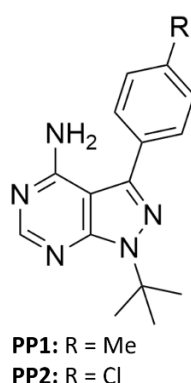
Figure 19. Chemical structure of PP1 and PP2.

The first group of pyrazolo[3,4-*d*]pyrimidine inhibitors of c-Src published by Botta's research group dates back to 2004. These compounds showed also an interesting antiproliferative activity towards epidermoid carcinoma cell lines.[38] Since then, a library of more than 400 members was synthesized,[40] with many of them able to act as c-Src inhibitors and exert anti-proliferative and pro-apoptotic effects towards several cancer cell lines (osteosarcoma,[41]

prostate cancer,[42,43] neuroblastoma,[39,44] glioblastoma,[37,40] rhabdomyosarcoma,[36] mesothelioma,[45] medulloblastoma,[46] and medullary thyroid carcinoma).[47]

In a recent study, Botta's research group published the X-ray crystal structure of the c-Src kinase domain in complex with **3** (4O2P), a pyrazolo[3,4-*d*]pyrimidine having an inhibitory activity against c-Src in the low micromolar range and a good ADME profile (TABLE 1).
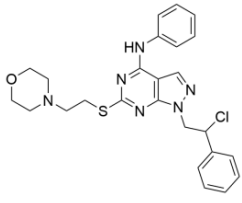
Compound 3	**c-Src K_i (µM)**	**Aqueous solubility (µg·mL^{-1})**	**Metabolic stability (%)**	**Gastrointestinal P_{app} (10^{-6} cm·s^{-1})**
	0.21	1.7	95	10

Table 1. Chemical structure, c-Src inhibitory activity and ADME properties of 3.

Interestingly, although the binding mode of **3** within the ATP binding pocket of the two protein chains found in the crystallographic cell unit is substantially the same, the c-Src kinase domain assumes different conformations: in chain A is in a CHO inactive conformation while in chain B is in an active conformation. Starting from this crystallographic complex, Monte Carlo free energy perturbation calculations were successfully applied to improve the activity of the hit compound.[39]

Recently, in the same research group, a new series of compounds resulting from a hybridization process between in house pyrazolo[3,4-*d*]pyrimidines and the Type I multikinase inhibitor dasatinib (IC$_{50}$ against c-Src of 0.4 nM)[48] were designed and synthesized. **ALM4-23** (FIGURE 20) represents one of the most interesting of the series for its potent inhibitory activity against c-Src (IC$_{50}$ = 24 ± 4 nM).

pyrazolo[3,4-*d*]pyrimidine   **dasatinib**   **ALM4-23**
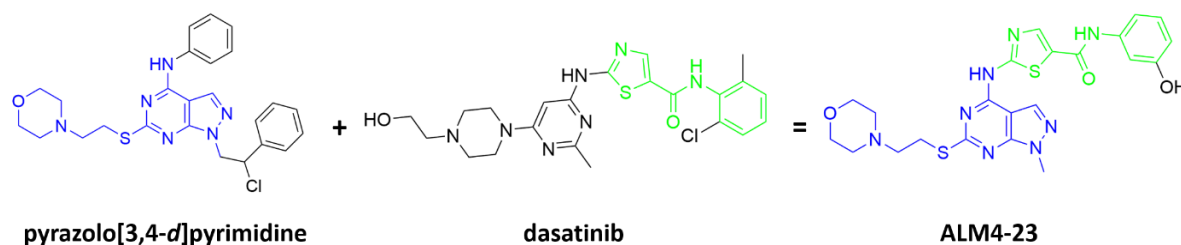
Figure 20. Hybridization process which led to ALM4-23.

As previously reported, kinases are enzymes able to adopt at least two extreme functional conformations. In the specific case of c-Src, at the best of our knowledge, the 3D structure of its kinase domain was reported in Protein Data Bank in an active conformation (e.g. 4O2P chain B)[39] and in both inactive CHO (e.g. 4O2P chain A)[39] and DFG-out conformations (e.g. 2OIQ)[49]. These structural data exemplify the ability of this kinase to assume three extreme functional conformational states: one active and two inactive.

3.2 Aim of the project

The first objective of the project was the evaluation of the mode of action of **ALM4-23**, or rather to elucidate which of the three possible conformational states of c-Src is stabilized by the inhibitor. Acquired information was then used to achieve the second objective, consisting in the rational design of **ALM4-23** derivatives endowed with potentially improved inhibitory activity towards c-Src.

3.3 Workflow

The evaluation of the mode of action of **ALM4-23** was carried out according to the following protocol:

- The compound was docked in the ATP binding site of kinase in active conformation, CHO inactive conformation, and DFG-out inactive conformation.
- The resulting complexes were then subjected to Gaussian Accelerated Molecular Dynamics (GaMD) simulations to further assess which of them was the most stable.

The design of **ALM4-23** derivatives was carried out according to the following protocol:

- The trajectory of the complex with higher stability was clustered to evaluate the most persistent interactions between ligand and the ATP binding pocket. Furthermore, the theoretical free energy of binding (ΔG_{bin}) of the complex was evaluated.
- The binding mode of compound obtained from the representative frame of most populated cluster was compared with the isosurfaces representative of the energetically most favourable sites of interaction between ATP binding site and probe atoms. This comparison simplified the design of **ALM4-23** derivatives.
- The proposed compounds were docked in the ATP binding site obtained from the representative frame of the most populated cluster. The resulting docking complexes were then subjected to GaMD simulations and evaluation of theoretical ΔG_{bin} to evaluate

the binding stability of the new derivatives within the ATP binding site and their potential increase in activity against c-Src.

3.4 Materials and methods

3.4.1 Protein preparation

The chain B of the X-ray structure 4O2P was prepared by using the Protein Preparation Wizard[50] workflow of the Maestro Suite Version 11.2. Crystallographic water molecules were removed, while hydrogen atoms, side chains, and missing loops were added using the software Prime of the Maestro Suite,[51,52] and bond orders were assigned. Moreover, Epik[53] was used to assign the proper tautomeric and ionization states of the co-crystallized ligand at pH 7.4. Finally, the network of hydrogen bonds was optimized at pH 7.4 and the complex was subjected to energy minimization with the OPLS3 force field, until the root mean square deviation (RMSD) of the heavy atoms was below 0.30 Å. Chain A of the X-ray structure 3EL8[54] was optimized with the same procedure.

Prime was used to build the composite model consisting of chains A and B of the X-ray structure 4DGG[55]. Before the creation of the model, the water molecules were removed from the structure, and the α carbon atoms of the two chains were aligned using Pymol software[56]. Subsequently, the FASTA sequence of chain A was used as query, while the coordinates of chain A (whole complex) and chain B (amino acid residues from 403 to 410) were in turn employed as templates in the building process of the composite model. The loop refinement task was then applied for the only loop not derived from templates (amino acid residues from 411 to 425), employing the serial loop sampling method and selecting "Ultra Extended" as level of accuracy (recommended for loops with more than 10 residues). Moreover, the side chain conformation of residues not derived from templates was optimized using the "Default" sampling algorithm. Finally, the complex was subjected to optimization of the of hydrogen bond network and to energy minimization as previously described for the selected chains of 4O2P e 3EL8.

3.4.2 Docking simulations

3.4.2.1 Self-docking

Docking simulations were performed by using Glide software[57] (SP procedure). The receptor grids were generated from the previously prepared crystallographic complexes, centred on the corresponding co-crystallized ligands, and treating the protein as rigid. The co-crystallized

ligands were energy minimized by using MacroModel[58] and finally docked within the corresponding receptor grid, with flexible ligand sampling.

RMSD calculation between the docked and the experimental poses was performed with smart_rms of Gold suite[59] considering only the ligand heavy atoms.

3.4.2.2 Ligand preparation and molecular docking of ALM4-23

The structure of the compound was sketched in Maestro and prepared in terms of tautomeric and ionization states by using the following tools implemented in LigPrep software:[60]

- Epik at pH 7.4
- Neutralizer
- Ionizer at pH 7.4

The compound was finally docked within the receptor grids described in the self-docking section and by using the same docking procedure.

3.4.2.3 Ligand preparation and molecular docking of ALM4-23 derivatives

The structures of compounds were sketched in Maestro and prepared in terms of tautomeric and ionization states by using the same procedure reported for **ALM4-23**. The receptor grid for docking calculation was generated from the representative frame of the most populated cluster related to the GaMD trajectory of the complex **ALM4-23**/c-Src active conformation. In particular, the frame was first stripped of waters and counterions and then the receptor grid was produced and used to dock the prepared compounds. The receptor grid generation and molecular docking procedures are the same described in the self-docking section.

3.4.3 GaMD simulations

GaMD is an enhanced sampling method where a harmonic boost potential is added to smoothen the system potential energy surface. In this way, the transitions between the low energy states are accelerated, allowing to obtain a complete conformational sampling of the system under study. Differently from other enhanced sampling methods, it does not require the definition of the reaction coordinates along which to apply the potential, which could represent an advantage especially when the operator does not have an expert knowledge of the system to study.[61]

3.4.3.1 Complexes with ALM4-23

For each complex, Maestro was employed to extract the ligand from the binding site, delete all the hydrogen atoms of the protein, and cap the N and C terminal residues.

Antechamber[62] (AM1-BCC method) was used to assign atomic partial charges and atom types to the ligand, while Parmchk2 was employed to produce the force field modification file having the force field parameters not provided by Generalized Amber Force Field (GAFF)[63], that was used to parameterize the ligand.

Leap module of Amber 16 was used for the parameterization of the protein by using ff14SB force field[64]. Hydrogen atoms were added to the complex before solvation in a TIP3P cuboid water box with a 20 Å distance between the walls and the protein. The system was neutralized by adding a proper number of counterions.

All the following described steps were performed by using Amber 16. Two steps of energy minimization were carried out to relax the solvated complex. In the first one, the complex was blocked with a positional restraint weight of 100 kcal/mol·Å^2, permitting in this way the minimization of the water molecules. 5000 steps of energy minimization were performed, the first 3000 with the steepest descent algorithm and the remaining 2000 with the conjugate gradient algorithm. In the second one, the overall structure was energy minimized through 7000 steps of steepest descent algorithm and 3000 with the conjugate gradient algorithm.

MD simulation with periodic boundary conditions at constant volume was carried out for 500 ps to heat the solvated complex to 300 K. A positional restraint weight of 2 kcal/mol·Å^2 was set on the protein to avoid large fluctuation during this heating step. Subsequently, 1.5 ns of MD with periodic boundaries conditions at constant pressure were run at 300 K with the aim to achieve density equilibration (the pressure and the temperature of system were kept constant by using the Berendsen barostat and the Langevin thermostat). These 1.5 ns were characterized by two steps (respectively of 500 ps and 1000 ps) in which positional restrains with a decreasing weight of 2 kcal/mol and 0.5 kcal/mol·Å^2 were set on the protein.

GaMD simulation was then performed at constant temperature (300 K) and pressure (1 atm) without restraints, adding dual boost (both on the dihedral and total potential energy) and setting the threshold energy E to V_{max}. GaMD simulation consist of:

- 10 ns of conventional MD simulation
- 50 ns of equilibration after the addition of dual boost potential
- 500 ns of production

During the simulation, the SHAKE algorithm was applied to restrain bonds containing hydrogen atoms, the electrostatic interactions were calculated through Particle Mesh Ewald (PME) method with a cut-off of 10 Å for nonbonded interactions, and 2 fs time step was used. RMSD calculation was performed by using the program cpptraj[65] implemented in Amber16.

3.4.3.2 Complexes of the ALM4-23 derivatives

The simulation protocol was the same described in section 'Complexes with **ALM4-23**', with the only exception for the production step of GaMD simulation that was set to 150 ns for each complex.

3.4.4 Cluster Analysis

The module cpptraj was used to perform cluster analysis. In each analysed trajectory the hierarchical agglomerative approach (average-linkage) was used as clustering algorithm, and the RMSD of the atoms of the amino acids around 6 Å from the ligand was used as distance metric. The choice of the cluster count was made:

- Evaluating the plot of the evolution of the RMSD, calculated on the atoms of the amino acid residues around 6 Å from the ligand, during the GaMD production time and with respect to last frame of minimization. The evaluation of this plot allowed to understand approximately the number of clusters present.
- Taking the cluster count able to divide the trajectory in accordance with the RMSD plot previously produced.

In TABLE 2 is reported the cluster count used to cluster the GaMD production phase and the complex of related trajectory.

COMPLEX	CLUSTER COUNT
ALM4-23/c-Src	2
Compound **1**/c-Src	2
Compound **2**/c-Src	2
Compound **3**/c-Src	3
Compound **4**/c-Src	2
Compound **5**/c-Src	2
Compound **6**/c-Src	2
Compound **7**/c-Src	3
Compound **8**/c-Src	2

Compound **9**/c-Src	2
Compound **10**/c-Src	2

Table 2. *Number of clusters for each complex.*

3.4.5 MM-GBSA analysis

In this computational method the combination of molecular mechanics calculations and continuum solvation models allows the evaluation of ΔG_{bin} and its decomposition in contributions deriving from different groups of atoms.[66,67]

For each analysed trajectory, the calculation of ΔG_{bin} of the complex was made by using the single trajectory approach[67] and employing 200 frames. In particular, these are the frames around the representative one of the most populated cluster, obtained by clustering the production phase of the GaMD trajectory, with the aim of evaluating the most persistent interactions between the compound and the ATP-binding site. The details of cluster analysis performed are reported in cluster analysis section. Moreover, the entropic term was not calculated because it is computationally expensive and subject to large approximation.[68]

For the complex between **ALM4-23** and c-Src in its active conformation, the MM-GBSA approach was also applied to carry out a pairwise decomposition of ΔG_{bin} by using for the calculation the same 200 frames employed in the evaluation of binding free energy.

3.4.6 Affinity grid maps

AutoDock employs affinity grid maps that show the interaction energy between specific atom types, generally of the ligand, and the whole binding site that is discretized through a grid. This approach allows to speed up the docking calculations because calculation of the interaction energies at every step of docking process is not necessary. Moreover, the visual inspection of these affinity grid maps can be exploited to obtain useful suggestions in the ligand optimization process,[69] which represents the reason of their application in this work.

The conformation of the protein within the representative frame of the most populated cluster, related to the GaMD trajectory of the complex **ALM4-23**/c-Src active conformation, was employed to calculate affinity grid maps. In detail, the .pdbqt files for protein and ligand (belonging to the above mentioned frame) were generated separately by using AutoDock Tools[70]. Subsequently, the grid box having 54 points in each dimension with a spacing between them of 0.375 Å and centred on the ligand, was generated. Finally, AutoGrid version 4.2.6 was used to calculate the interaction energy for the following atom types (named in the text as probe

atoms): A, OA, NA, HD, C, F, Br and Cl. The visualization of the isosurfaces representative of the energetically most favourable sites of interaction between the atom types and ATP binding site was made by using the AutoDock plugin for PyMol[69].

3.5 Results and discussion

3.5.1 Selection of the crystallographic structures

Three crystallographic structures of the c-Src kinase domain in active conformation, CHO inactive conformation, and DFG-out inactive conformation were selected from the Protein Data Bank to perform docking studies. Among the several 3D structures of the kinase present in the database, the attention was focused on those where the catalytic domain is co-crystallized with ligands having a pyrazolo[3,4-*d*]pyrimidine scaffold. In this way, ATP binding sites adapted to ligands structurally similar to the compound under study were passed to the subsequent molecular docking studies, potentially increasing their success rate.

In this selection process, given the high sequence identity (99.6%),[11] both the human and chicken kinase structures were taken into account. Particularly, the kinase domains of these proteins have only two differences: a methionine and an aspartate in chicken correspond to a threonine and a glutamate in human (FIGURE 21). Moreover, these mutated residues are sited in regions exposed to solvent and away from the ATP binding site, which makes reasonable the use of chicken structures in the following computational studies aimed to understand the behaviour of **ALM4-23** and derivatives towards the human protein.
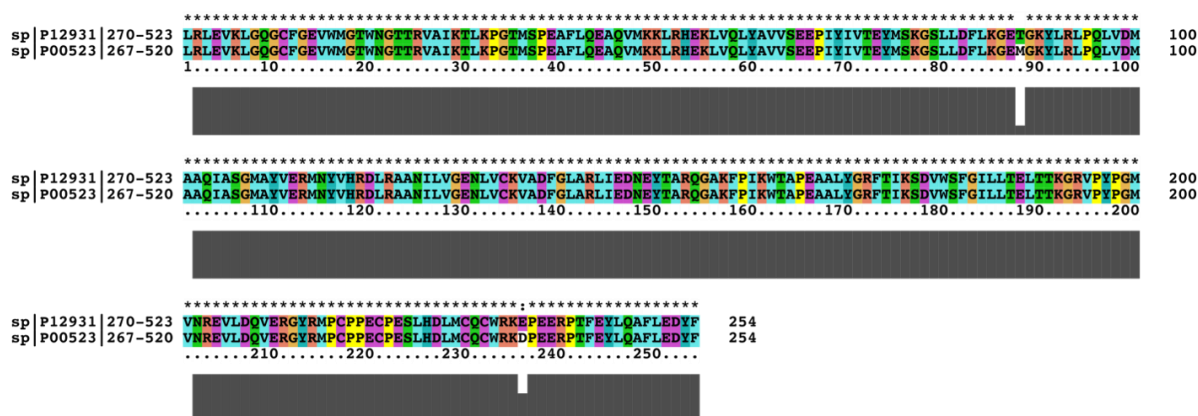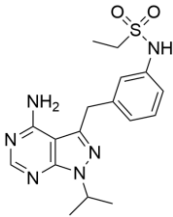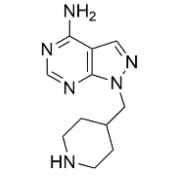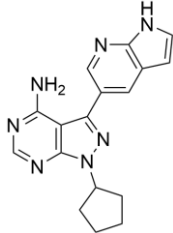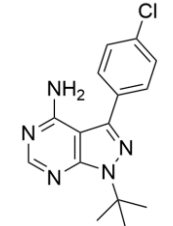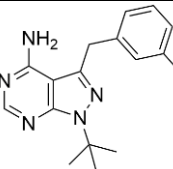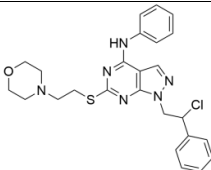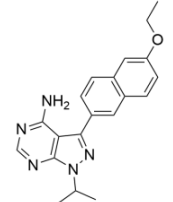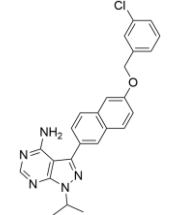
Figure 21. *Sequence alignment of the catalytic domain of human and chicken c-Src (UniProtKB codes: P12931 and P00523, respectively). Bars below the sequence alignment correspond to the degree of amino acid conservation among the sequences (full bar: residues identity; empty bar: completely different residues).*

Sixteen structures representative of the kinase domain of c-Src co-crystallized with pyrazolo[3,4-*d*]pyrimidines were identified within the Protein Data Bank (TABLE 3).

PDB	LIGAND	CHAINS	CONFORMATION	ORGANISM	RESOLUTION	MUTATION
3SVV		A-B	ACTIVE/ACTIVE	*Gallus gallus*	2.2 Å	T338C
3UQG		A-B	ACTIVE/ACTIVE	*Gallus gallus*	2.2 Å	
3EN4		A-B	ACTIVE/ACTIVE	*Gallus gallus*	2.55 Å	
3GEQ		A-B	ACTIVE/ACTIVE	*Gallus gallus*	2.2 Å	R388A
4LGG		A-B	ACTIVE/ACTIVE	*Gallus gallus*	2.41 Å	T338G
4O2P		A-B	CHO/ACTIVE	*Gallus gallus*	2.1 Å	
3UQF		A-B	CHO/CHO	*Gallus gallus*	2.27 Å	
4DGG		A-B	CHO/CHO	*Gallus gallus*	2.65 Å	

3EN5		A-B	-/CHO	*Gallus gallus*	2.66 Å	
3EN7		A-B	CHO/-	*Gallus gallus*	2.81 Å	
3QLF		A-B	-/CHO	*Gallus gallus*	2.75 Å	L317I
4K11		A	CHO	*Homo sapiens*	2.3 Å	T338G
3EL7		A	DFG-out	*Gallus gallus*	2.8 Å	
3EL8		A-B	DFG-out /ACTIVE	*Gallus gallus*	2.3 Å	
4LGH		A-B	-/-	*Gallus gallus*	2.8 Å	T338G
3EN6		A-B	-/-	*Gallus gallus*	2.39 Å	

***Table 3**. 3D structures of the c-Src kinase domain co-crystallized with pyrazolo[3,4-d]pyrimidines available within PDB. The conformational state of the kinase domain is not reported when, because of the low resolution of the ATP binding sites, was not possible to decipher it.*

Chain B of the X-ray crystal structure 4O2P was selected as representative of the c-Src kinase domain in active conformation because of its higher resolution among the available structures but also for the high structural similarity that characterizes the co-crystallized compound and the pyrazolo[3,4-*d*]pyrimidine moiety of **ALM4-23**.

Although it does not belong to the crystal structure with best resolution available, chain A of 4DGG was selected as representative of the c-Src kinase domain in CHO inactive conformation. The catalytic domain of this chain is co-crystallized with the pyrazolo[3,4-*d*]pyrimidine able to induce the creation of the largest additional binding pocket, due to its C3 moiety. In fact, this moiety represents the sterically most demanding substituent of the co-crystallized ligands with the protein in this conformation, directed towards helix αC. Structures with better resolution but co-crystallized with pyrazolo[3,4-*d*]pyrimidines endowed with less demanding groups, have a smaller ATP binding site, potentially influencing the final results of docking simulations. Also the catalytic domain of chain B has the CHO conformation and is characterized by an incomplete helix αC (well resolve in chain A), and by the complete activation loop folding (missed in chain A) typical of this inactive conformation. Therefore, chain B was used for the reconstruction of the missing folded areas of chain A to obtain a more complete CHO catalytic domain.

Chain A of the 3EL8 crystal structure was selected as representative of the c-Src kinase domain in DFG-out inactive conformation because of its highest resolution among the available structures. The selected complexes were finally optimized before performing docking studies.

3.5.2 Docking studies of ALM4-23

Docking studies were performed using Glide software to understand the possible binding mode of **ALM4-23** in the ATP binding site of the three conformational states of c-Src. The reliability of the computational docking protocol was verified by simulating the ligand-protein interactions of the selected crystallographic structures (self-docking). Specifically, ligands were extracted from the prepared complexes, energy minimized and finally docked in the corresponding ATP binding site. Results showed the ability of docking protocol to perfectly reproduce the experimental binding mode of 4DGG and 3EL8 ligands (RMSD between the docking pose and the experimental binding mode of 0.753 Å and 0.610 Å respectively). The predicted binding mode for 4O2P ligand is also highly comparable to the experimental pose, with a slight difference concerning the disposition of the solvent exposed morpholinoethylthio substituent (RMSD between the docking pose and the experimental binding mode of 2.804 Å) (FIGURE 22).
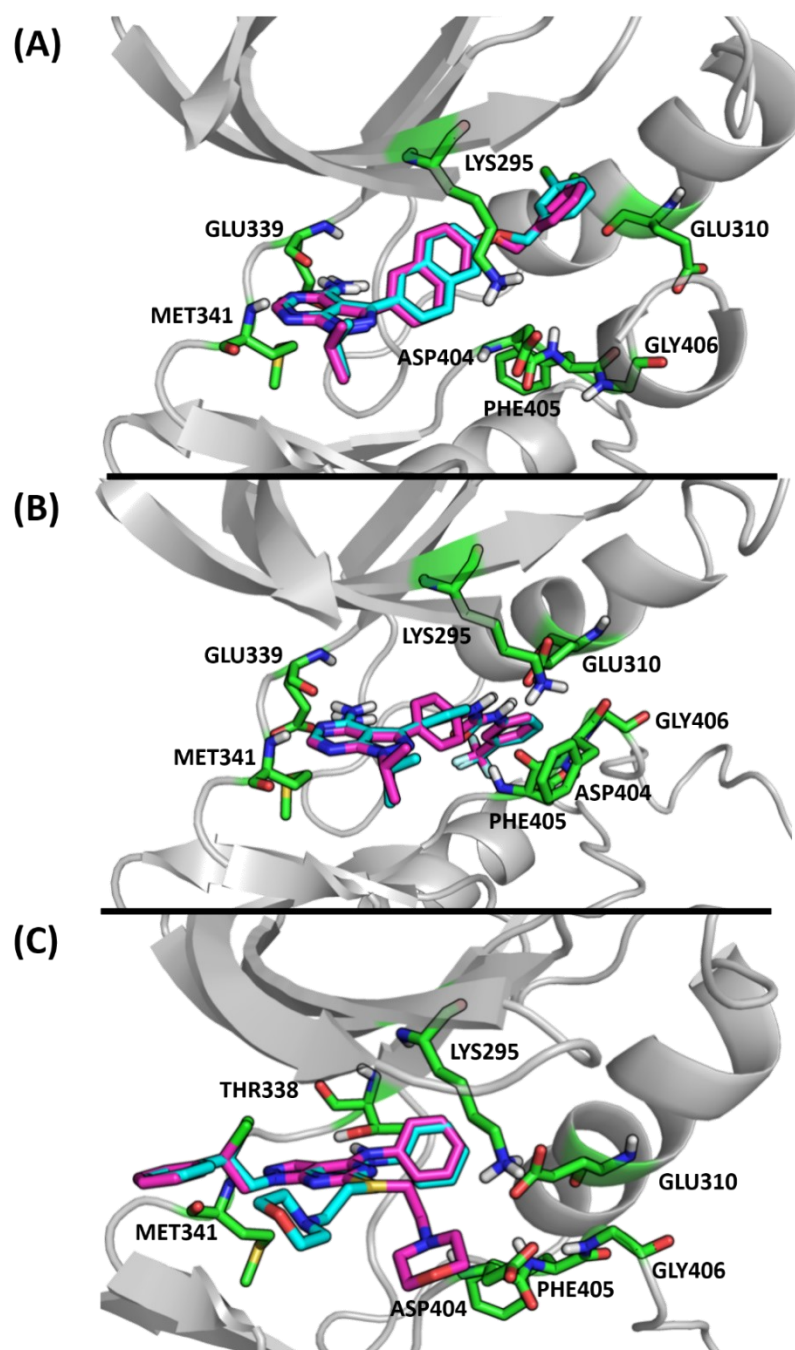
Figure 22. Predicted binding mode for the ligands of 4DGG (A), 3EL8 (B) and 4O2P (C) in the corresponding ATP binding site. The experimental and simulated binding modes are reported in cyan and magenta sticks, respectively. In green sticks are reported the interacting residues and those representative of the conformational state of protein.

Once the reliability of the computational docking protocol was verified, it was applied to dock **ALM4-23** in the ATP binding site of the three c-Src conformations. Before docking calculations, the compound was sketched and optimized in terms of tautomeric and ionization states at pH 7.4.

Docking results showed that **ALM4-23** interacts with the ATP pocket of active conformation through hydrogen bonds between the amide group and gatekeeper residue Thr338, the aromatic

37

amine and Met341 of hinge region, the phenol group and Glu310 of helix αC, the phenol group and Asp404 of DFG-motif, the protonated morpholine and Asp348. A further hydrogen bond is established with Met341 through its thiazole ring (FIGURE 23A).

The binding pose in the ATP pocket of DFG-out inactive conformation shows that the compound interacts through hydrogen bonds between the pyrazole ring and Met341, the aromatic amine and Thr338, the phenol group and Asp404 (FIGURE 23B). Conversely, **ALM4-23** interacts with the ATP pocket of CHO inactive conformation by establishing only one hydrogen bond (with Thr338) thanks to its phenol group (FIGURE 23C).
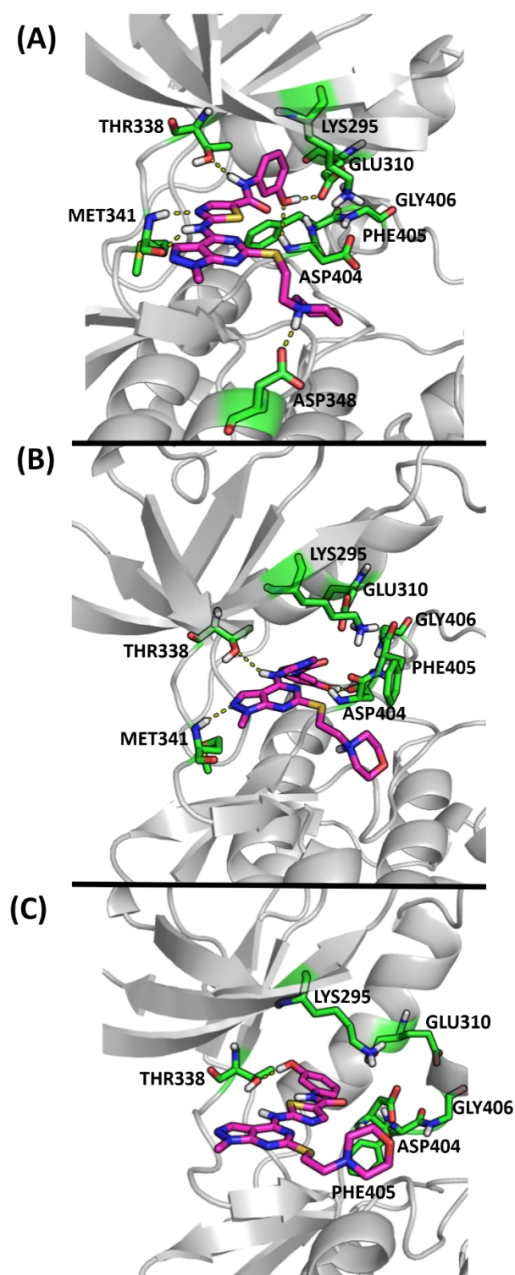
Figure 23. *Predicted binding mode of **ALM4-23** in the ATP binding site of c-Src active conformation (A), DFG-out inactive conformation (B) and CHO inactive conformation (C). Ligand is reported in magenta sticks. Green sticks show residues important for interactions with the ligand and residues representative of the conformational state of protein.*

The described binding poses are associated with Glide score values that highlight the potential ability of compound to better interact with the kinase active conformation (TABLE 4).

COMPLEX	GLIDE SCORE (kcal/mol)
ALM4-23/Active conformation	-11.6
ALM4-23/DFG-out conformation	-9.9
ALM4-23/CHO conformation	-6.4

*Table 4. Glide score related to the predicted binding modes of **ALM4-23** in the ATP binding pocket of the three c-Src conformational states.*

Interestingly, the binding mode of **ALM4-23** in the kinase active conformation is highly similar to the experimental pose of its parent compound dasatinib in the c-Src ATP pocket (FIGURE 24). This last observation reinforces the suggestion that **ALM4-23** could act by stabilization of the c-Src active conformation (as dasatinib, that is a Type I inhibitor, does), and further confirms the reliability of the docking results.
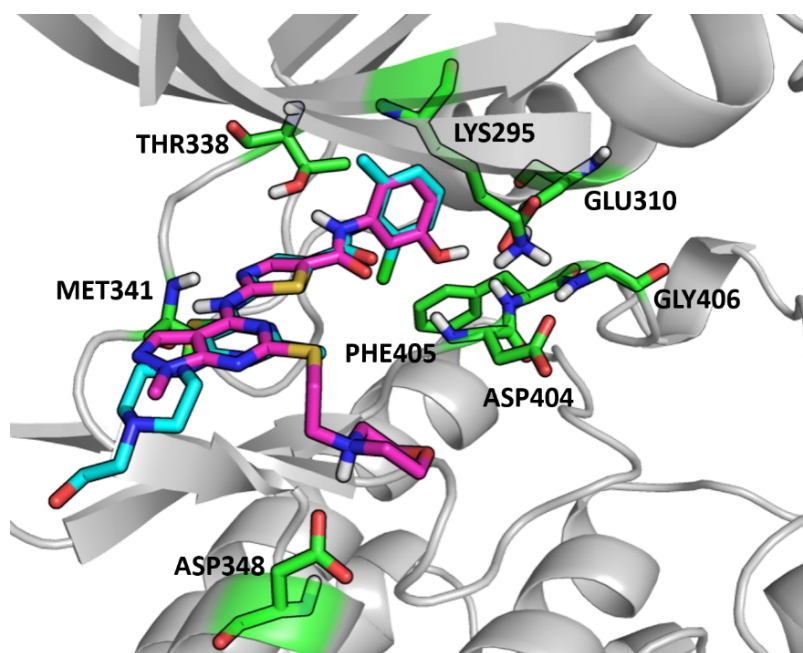
*Figure 24. Overlay of the **ALM4-23** predicted binding mode and the experimental binding mode of dasatinib (3G5D) in the ATP pocket of c-Src active conformation. Dasatinib and **ALM4-23** are reported in cyan and magenta sticks, respectively. In green sticks are reported the residues that interact with **ALM4-23** and those representative of the conformational state of protein.*

3.5.3 GaMD studies of ALM4-23 complexes, cluster analysis and evaluation of ΔG_{bin}

GaMD simulations were performed on the three docking complexes to evaluate more in depth their stability and to obtain further information about ligand-protein interactions. The reasons behind the use of this enhanced sampling technique are mainly two:

- Among its several applications, it was successfully used to evaluate the different stability of ligand-protein complexes in which the ligands were characterized by progressively lower binding affinity against their biological target.

- Being an enhanced sampling method, it allows for a better exploration of the conformational/energy landscape of the ligand-protein complexes, compared to conventional MD (cMD).[71] This should improve the degree of accuracy of the subsequent thermodynamic calculations.

Before simulations, docking complexes were solvated in a box of explicit water molecules containing counterions to neutralize the charge. Subsequently, the systems were relaxed through energy minimization, heated to 300 K and density equilibrated to produce a 560 ns GaMD trajectory for each complex (comprised of 10 ns cMD, 50 ns equilibration after adding the boost potential and 500 ns production).

Visual inspection of GaMD trajectories confirms the suggestions arising from docking simulations, showing a great stability of the complex between the compound and the active conformation and a very low stability of the complexes with inactive conformations, where the ligand loses its binding modes. In this regard, FIGURE 25 shows the evolution of the RMSD calculated on the atoms of the ligand during the GaMD production time with respect to the docking pose in each complex. In the complex with kinase active conformation, ligand RMSD is stable around values of 2 Å for the entire simulation time, which is representative of the great binding stability of **ALM4-23** within the ATP binding site. On the contrary, in the complexes with inactive conformations, the binding stability within the ATP binding sites is very low, as evidenced by the RMSD values that increase significantly up to 12 and 10 Å.
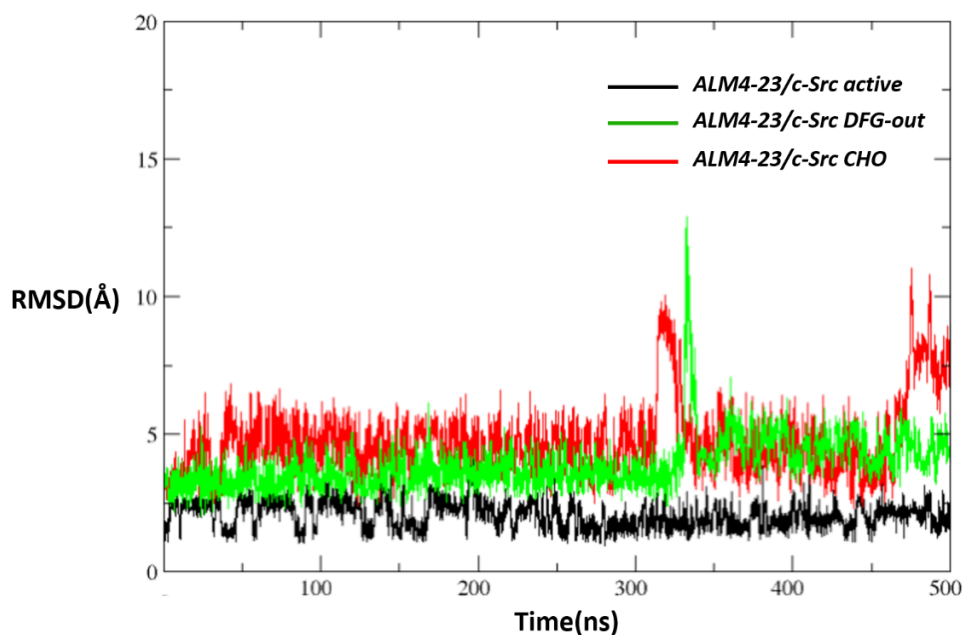
Figure 25. *Ligand RMSD in the three complexes.*

Overall, the results obtained from the computational studies show the trend of **ALM4-23** to stabilize the c-Src active conformation as a typical Type I inhibitor.

Once the mode of action of **ALM4-23** was clarified, the production phase of the GaMD trajectory related to the complex with the active conformation was submitted to cluster analysis to evaluate the most persistent interactions made by the compound within the ATP binding site. In this clustering process, the RMSD of the atoms of the amino acid residues around 6 Å from the ligand was used to evaluate the similarity between the frames.

In the representative frame of the most populated cluster (100 % of the total population), the compound retains overall the same interactions highlighted by molecular docking calculations (with Thr338, Met341, Glu310, Asp404) losing completely only the one with Asp348, probably due to the solvent exposure of the morpholine group (FIGURE 26).
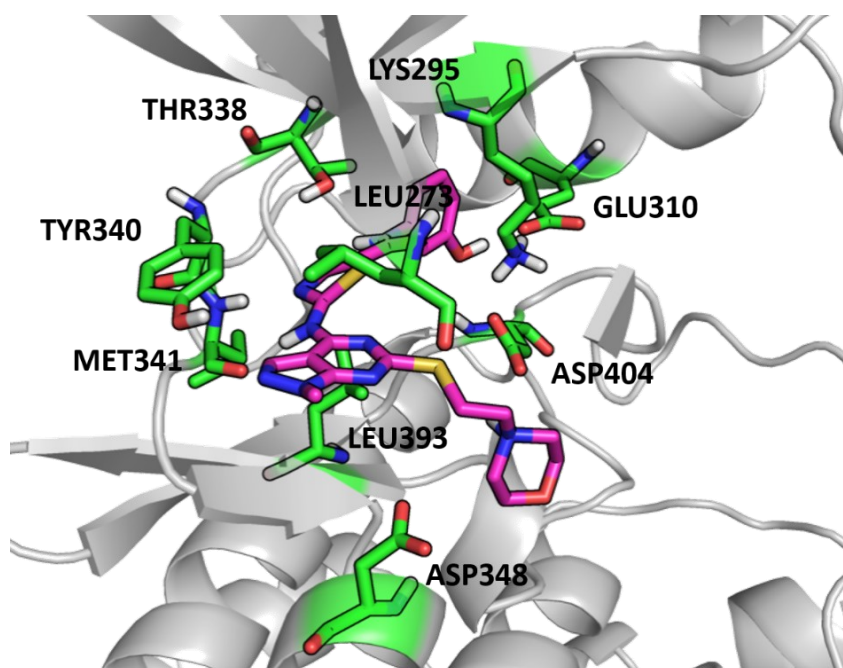
Figure 26. *Binding mode of **ALM4-23** in the representative frame of the most populated cluster. Ligand is reported in magenta sticks while in green sticks are reported the interacting residues.*

To obtain more detailed information on the complex, an energetic estimation of the ligand-protein interactions was also carried out through the pairwise decomposition of ΔG_{bin} implemented in the MM-GBSA approach. For the calculation, 200 frames around the representative one of the most populated cluster, were used. The energetically most relevant interactions (lower than -2.5 kcal/mol) made by **ALM4-23** in the kinase ATP binding site are the hydrogen bonds with Met341, Glu310, Thr338, the salt bridge with Asp348 (although the high standard deviation associated to its energy value underlines the intermittence of this interaction), the cation-π interaction with Lys295, and the hydrophobic interactions with Leu273, Tyr340 and Leu393 (TABLE 5).

INTERACTING AMINO ACID RESIDUE	ΔG_{bin} (kcal/mol)
Leu273	-4.929 ± 0.693
Lys295	-4.171 ± 0.990
Glu310	-6.896 ± 0.948
Thr338	-2.972 ± 0.810
Tyr340	-3.815 ± 0.742
Met341	-7.099 ± 0.647
Asp348	-9.864 ± 3.414
Leu393	-3.392 ± 0.412

Table 5*. Hot spots in ligand-protein interactions.*

Finally, the MM-GBSA approach was applied to calculate the ΔG_{bin} of the complex under study, corresponding to -60.5092 ± 4.6782 kcal/mol. For the calculation the same 200 frames employed before were used.

3.5.4 Rational design of ALM4-23 derivatives

The representative frame of most populated cluster identified from previous cluster analysis exemplifies also the most probably conformational state of the ligand-protein complex under study. Therefore, this snapshot was used for the rational design of **ALM4-23** derivatives endowed with potentially improved inhibitory activity towards c-Src. For this purpose, the conformation of the protein within the frame was first employed to calculate the affinity grid maps, describing the interaction energy between specific probe atoms and the protein ATP binding site (TABLE 6).

PROBE ATOMS	MINIMUM ENERGY (kcal/mol)	ISOSURFACE COLOR
A (aromatic carbon)	-0.89	Lime
OA (oxygen that accepts h bond)	-1.65	Red
NA (nitrogen that accepts h bond)	-1.37	Blue
HD (hydrogen that donates h bond)	-0.72	Cyan
C (aliphatic carbon)	-0.99	Orange
F (fluorine)	-0.61	Yellow
Br (bromine)	-1.55	Violet purple
Cl (chlorine)	-1.30	Green

Table 6. *Probe atoms for which the affinity maps were calculated, along with their minimum interaction energy with ATP binding site and the colour of the associated isosurface.*

Subsequently, the binding mode of **ALM4-23** within the same frame was compared with the isosurfaces representative of the energetically most favourable sites of interaction between the ATP binding site and the probe atoms. This comparison simplified the rational design of **ALM4-23** derivatives.

The first class of proposed derivatives is characterized by the substitution of hit compound moieties (atoms or groups) with groups that, based on the visualized isosurfaces, should lead to an improvement of the interactions with the ATP binding site (TABLE 7).
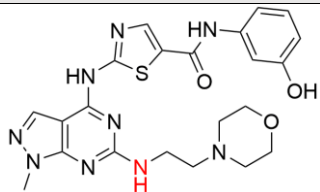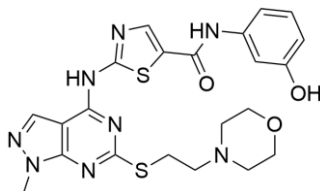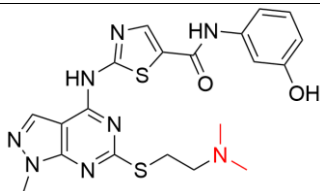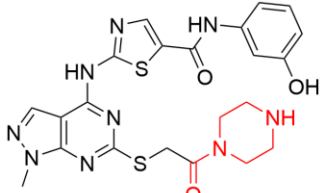
HIT COMPOUND	DESIGNED DERIVATIVES
 ALM4-23	 Compound **1** Compound **2** Compound **3**

Table 7. *Chemical structures of **ALM4-23** and first class of proposed derivatives. Structural changes with respect to **ALM4-23** are reported in red.*

Compound **1** is characterized by the substitution of the thiol sulfur atom with a secondary amine. This point mutation was suggested by the isosurfaces at contour level respectively of -1.2 kcal/mol, -1.2 kcal/mol and -0.6 kcal/mol for the interaction of OA, NA and HD probe atoms with the ATP binding site, that highlight how this sulfur atom falls in an area of the ATP binding site where the presence hydrogen bond donor groups is preferred for the interaction with the protein (FIGURE 27).
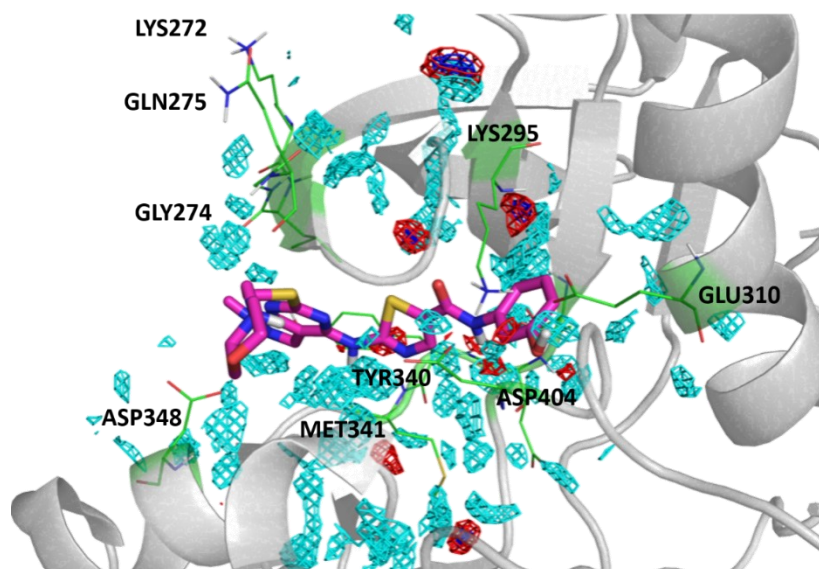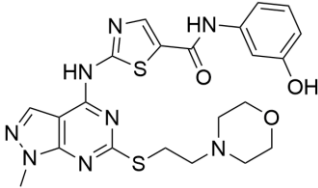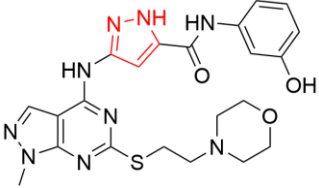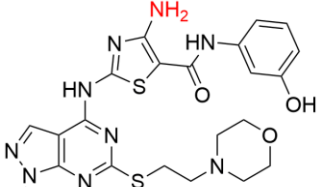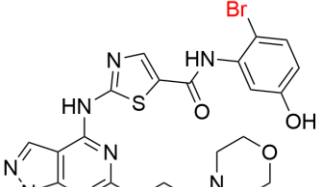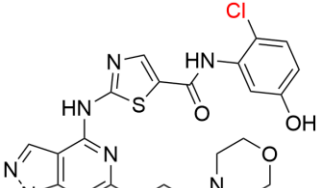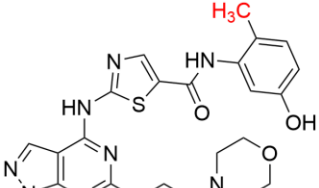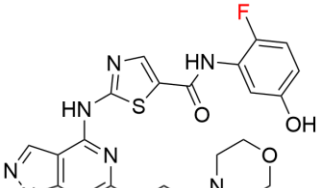
Figure 27. *Cyan, red, and blue meshes represent the isosurfaces for the interaction of HD, OA, and NA probe atoms with the ATP binding site at contour level reported in the text. Green lines show the residues of ATP binding site, while the binding mode of* **ALM4-23** *is represented by magenta sticks.*

Compound **2** resulted from the molecular simplification of morpholine ring, whose its oxygen atom falls in the previously described area of the ATP binding site where hydrogen bond donor groups are preferred for the interaction with the protein (FIGURE 27).

Compound **3** is characterized by the substitution of the morpholinomethyl moiety with a piperazine amide, to verify the effect on the inhibitor activity of the shifting of the hydrogen bond donor group towards an area more exposed to the solvent and in which these groups are preferred for the interaction with protein.

The second class of proposed derivatives is characterized by the addition of atoms or groups that can potentially reinforce the binding with ATP binding site by establishing additional interactions within it (TABLE 8).

HIT COMPOUND	DESIGNED DERIVATIVES
ALM4-23	Compound **4**
	Compound **5**
	Compound **6**
	Compound **7**
	Compound **8**
	Compound **9**
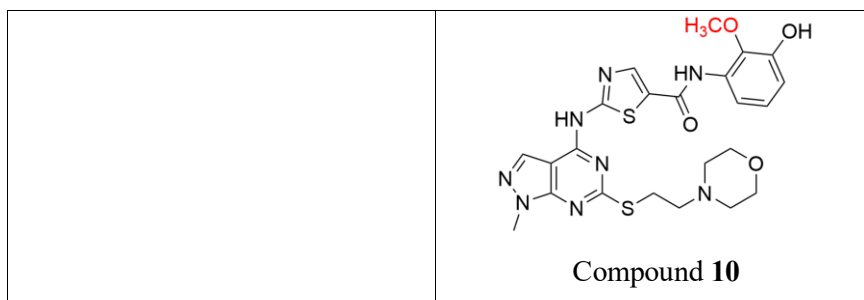
Compound **10**

*Table 8. Chemical structures of **ALM4-23** and second class of proposed derivatives. Structural changes with respect to **ALM4-23** are reported in red.*

Compound **4** is characterized by replacement of the thiazole with a pyrazole ring. This substitution was suggested by the isosurfaces at contour level of -1.2 kcal/mol, -1.2 kcal/mol, -0.6 kcal/mol, and -0.4 kcal/mol for the interaction of NA, OA, HD and A probe atoms with the ATP binding site, that highlight how a heterocyclic ring endowed with adjacent hydrogen bond acceptor and donor groups could reinforce the interactions with hinge region of the ATP binding site (FIGURE 28).
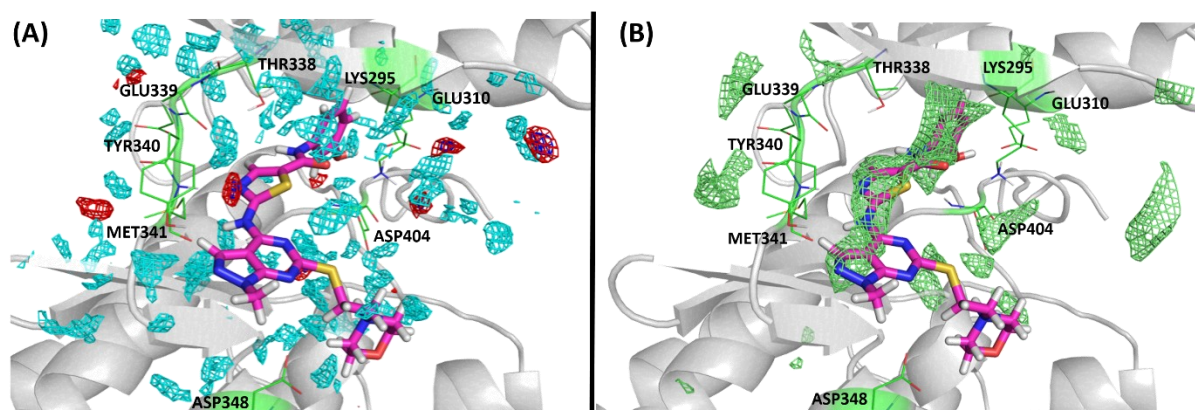
*Figure 28. Cyan, red, and blue meshes represent the isosurfaces for the interaction of HD, OA, and NA probe atoms with the ATP binding site (A). Lime meshes represent the isosurface for the interaction of A probe atom (B). In both the images, ATP binding site residues are reported in green lines, while **ALM4-23** in magenta sticks. The contour level used to produce the isosurfeces is reported in the text.*

Compound **5** was designed based on the same observation.

Compound **6**, **7**, **9** and **8** are characterized respectively by the addition of a bromine, chlorine, fluorine atom and a methyl group in para position with respect to the phenol OH, as suggested by their isosurfaces at contour level of -1.1 kcal/mol, -1.0 kcal/mol, -0.5 kcal/mol, and -0.7 kcal/mol (FIGURE 29).
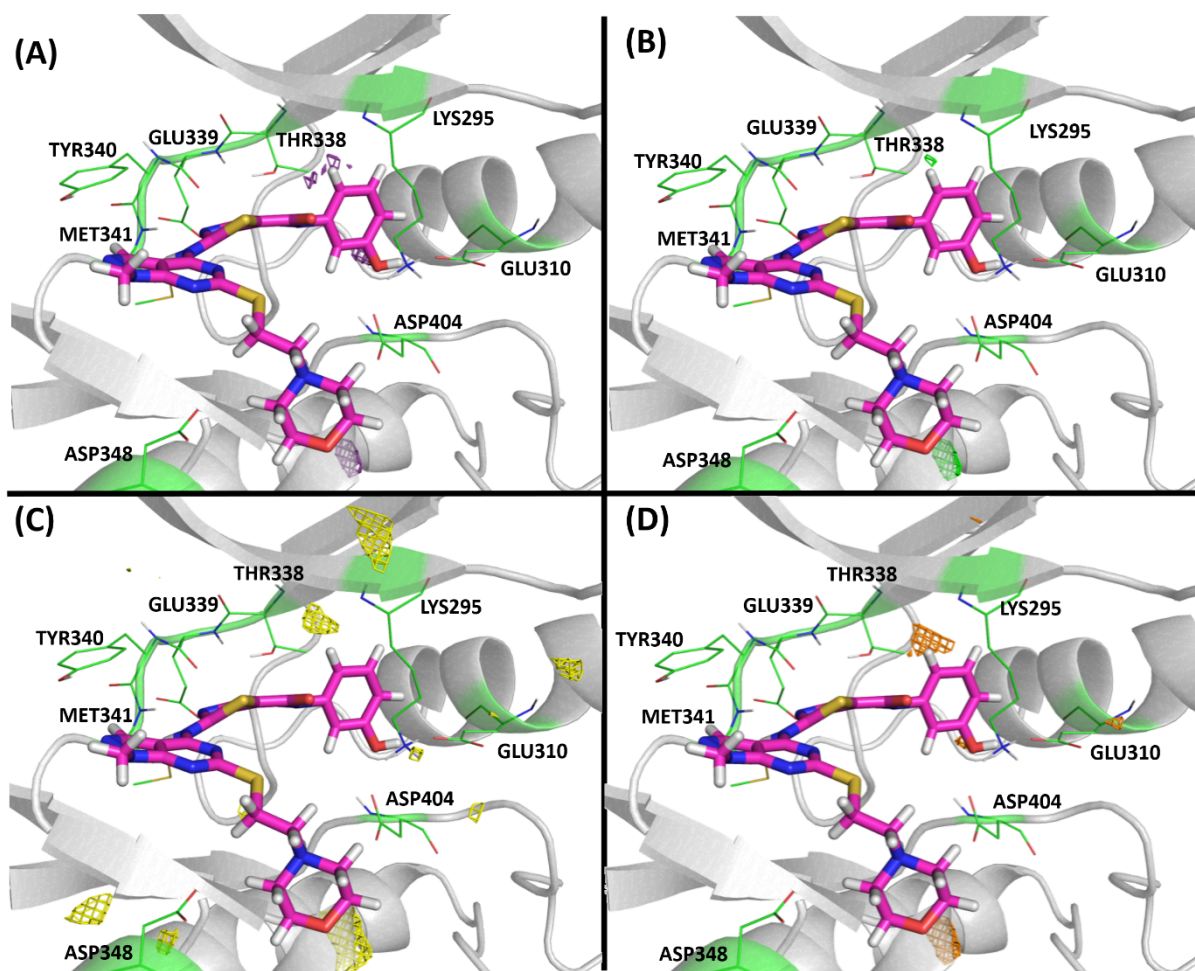
Figure 29. *Violet purple, green, yellow, and orange meshes represent the isosurfaces for the interaction of bromine (A), chlorine (B), fluorine (C), and C probe atom (D), respectively, with the ATP binding site. In all the images, ATP binding site residues are reported in green lines while **ALM4-23** in magenta sticks. The contour level used to produce the isosurfaces is reported in the text.*

Interestingly, dasatinib has the same substitution proposed in compound **8** which further supports the reliability of the used hit optimization computational approach.

Compound **10** is characterized by the addition of a methoxy group in ortho position of phenol OH, as suggested by the isosurface at contour level of -1.2 kcal/mol for the interaction of OA probe atom with the ATP binding site, that highlights how a hydrogen bond acceptor group could reinforce the interaction with Asp404 (FIGURE 30).
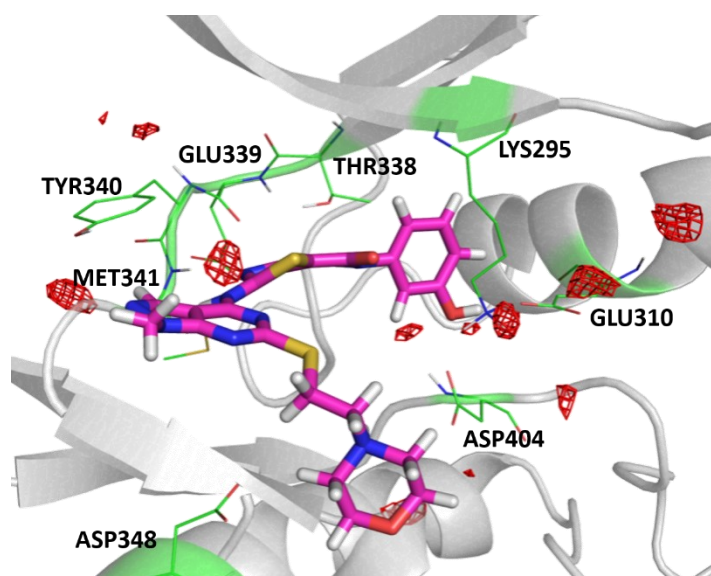
Figure 30. *Red meshes represent the isosurfaces for the interaction of OA probe atom with the ATP binding site at contour level reported in the text. ATP binding site residues are in green, while **ALM4-23** is represented by magenta sticks.*

In addition to these observations, decoration of the phenol ring with electron donor groups such as a methyl and a methoxy substituent (compound **8** and **10**) should improve the cation-π interaction with Lys295.

3.5.5 Docking studies of ALM4-23 derivatives

The proposed compounds were sketched, optimized in terms of tautomeric and ionization states at pH 7.4 and finally docked in the ATP binding site of the kinase using the previously validated docking procedure. The receptor used for docking calculations was obtained from the representative frame of the most populated MD cluster.

Binding mode of the proposed derivatives superimposed to the isosurfaces that have suggested their design are reported in FIGURE 31. All compounds maintain the same binding mode of **ALM4-23** and their substituents match the areas energetically most favoured for the interactions with the ATP binding site.
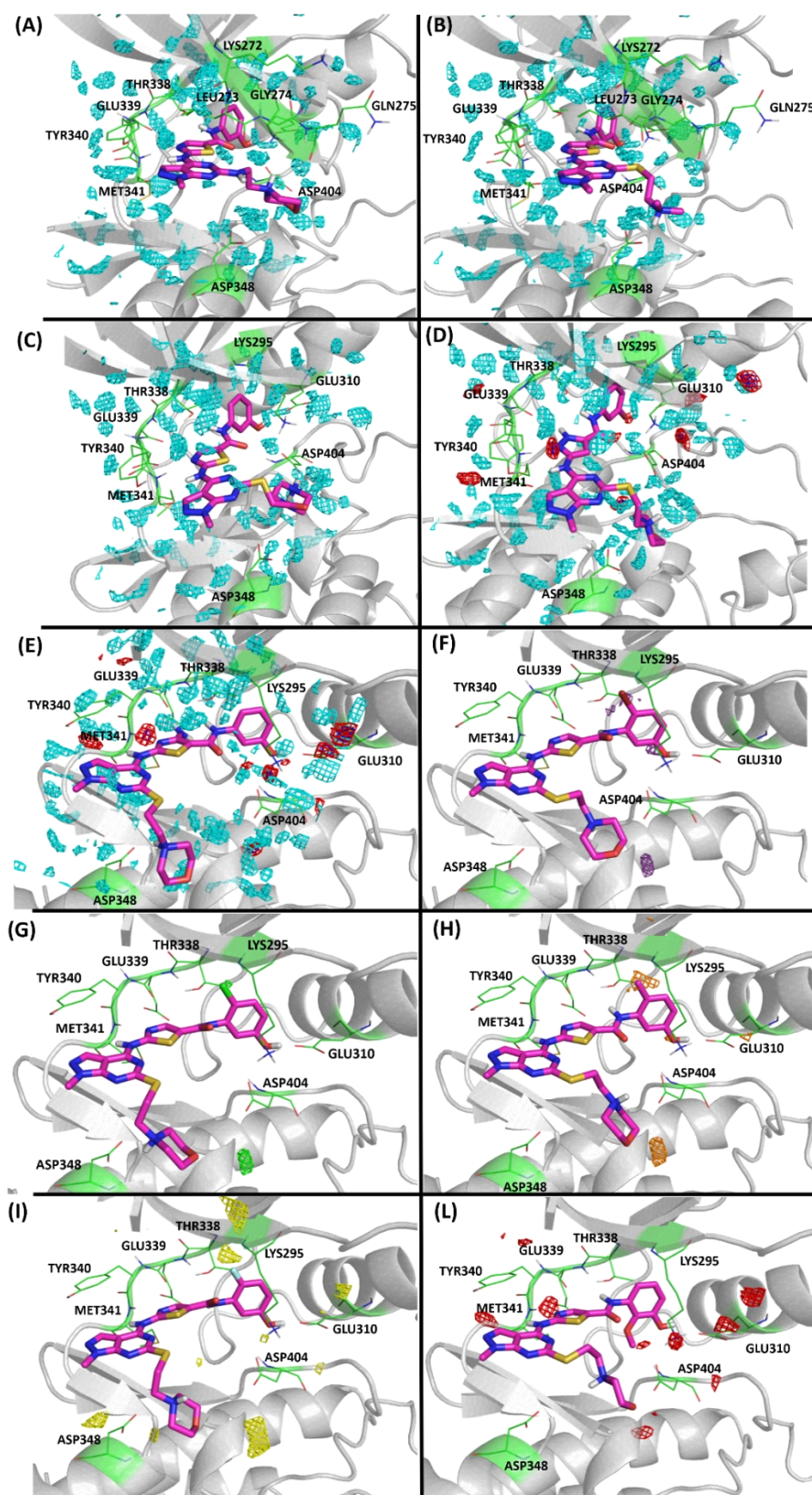
Figure 31. *Binding mode of compound **1** (A) compound **2** (B) compound **3** (C) compound **4** (D) compound **5** (E) compound **6** (F) compound **7** (G) compound **8** (H) compound **9** (I) and compound **10** (L) in the kinase ATP binding site. Ligands are represented in magenta sticks while ATP binding site residues are shown with green lines. The isosurfaces for the interaction of HD, OA, NA, Br, Cl, C, and F probe atoms with the ATP binding site at the contour levels reported before are represented by cyan, red, blue, violet purple, green, orange and yellow meshes, respectively.*

3.5.6 GaMD studies of the complexes of ALM4-23 derivatives, cluster analysis and evaluation of ΔG_{bin}

The ten complexes resulting from docking simulations were subjected to GaMD simulations to evaluate their stability. The same preparation and simulation protocol used for the **ALM4-23** complexes was employed, with the only exception represented by the production time that was set to 150 ns for each simulation. Visual inspection of GaMD trajectories showed great stability for all the complexes. In this regard, the evolution of RMSD of the ligands during the GaMD production time with respect to the docking pose in each complex is reported in FIGURE 32. Overall, ligands RMSD is stable in all the systems at low values (around 2 Å and 3 Å) with some cases (particularly in the complexes with compounds **5**, **6**, **8**, **9** and **10**) of temporary increment at slightly higher values. These increments are due to the natural fluctuations of substituents in C6 of pyrazolo[3,4-*d*]pyrimidine scaffold that are solvent exposed. In conclusion, these plots highlight the great binding stability of all designed derivatives to the ATP binding site of kinase.
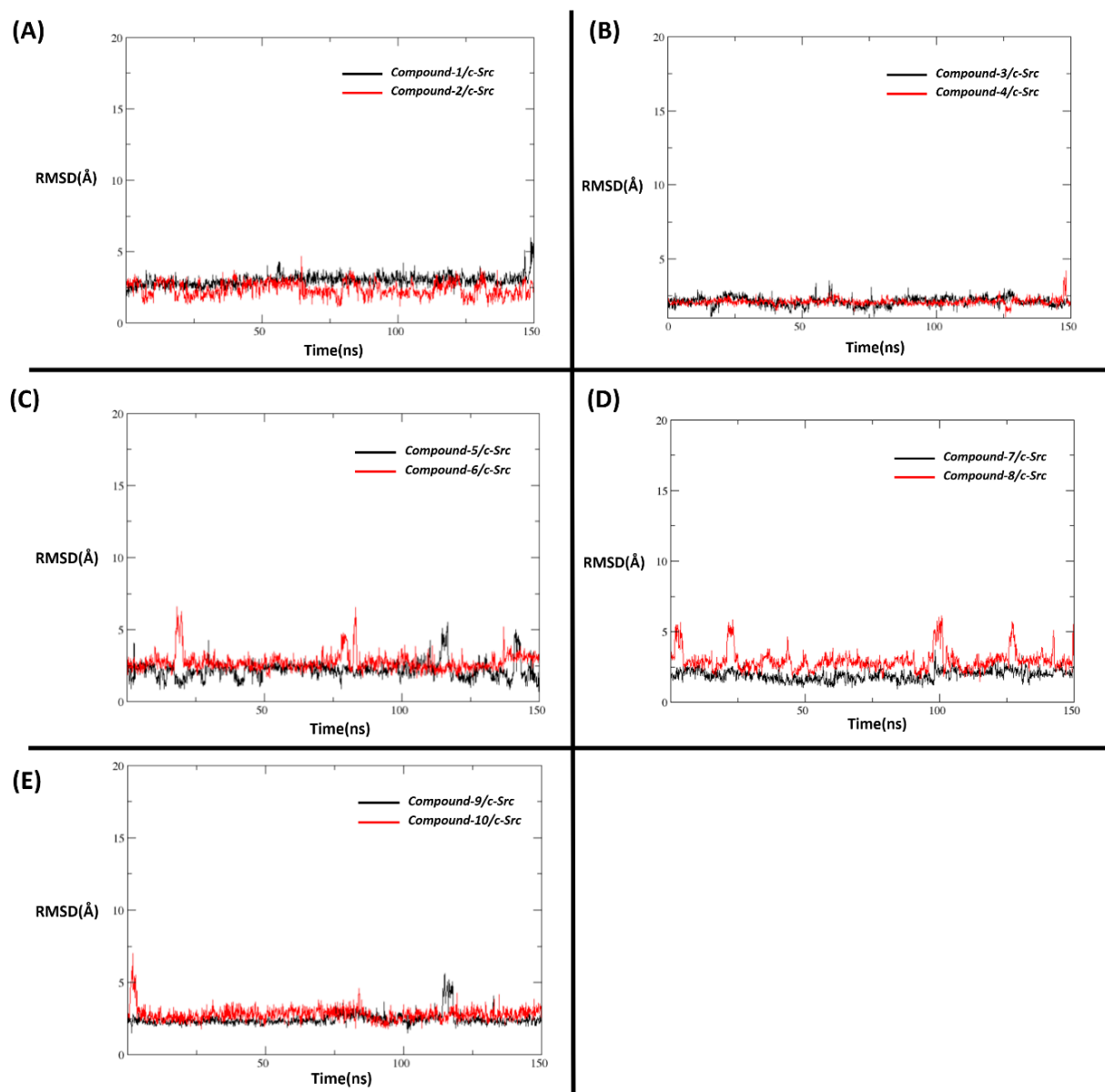
Figure 32. *RMSD of compound 1 (black) and 2 (red) (A), 3 (black) and 4 (red) (B), 5 (black) and 6 (red) (C), 7 (black) and 8 (red) (D), 9 (black) and 10 (red) (E) in complex with c-Src ATP binding site.*

The production phase of each GaMD trajectory was then submitted to cluster analysis to evaluate the most persistent interactions made by compounds within the ATP binding site. In each clustering process, the RMSD of the atoms of the amino acid residues around 6 Å from the ligand was used to evaluate the similarity between the frames. The percentage of the total population found in the clusters of the analysed trajectories is reported in TABLE 9.

COMPLEX	C0	C1	C2
Compound **1**/c-Src	99.5 %	0.5%	
Compound **2**/c-Src	99.9 %	0.1%	
Compound **3**/c-Src	67.8 %	20.7 %	11.5 %
Compound **4**/c-Src	99.8 %	0.2 %	
Compound **5**/c-Src	98 %	2 %	
Compound **6**/c-Src	99.9%	0.1 %	
Compound **7**/c-Src	89.8 %	10.1 %	0.1 %
Compound **8**/c-Src	96.1 %	3.9 %	
Compound **9**/c-Src	99.6 %	0.4%	
Compound **10**/c-Src	99.8 %	0.2 %	

Table 9. *C0, C1 and C2 represent the first, the second and the third most populated cluster of the trajectory related to a given complex.*

Finally, the MM-GBSA approach was applied to calculate the ΔG_{bin} of the complexes to evaluate the potential increase in affinity against c-Src of the proposed derivatives (TABLE 10). For these calculations, we used 200 frames around the representative one of the most populated cluster of each trajectory.

COMPLEX	ΔG_{bin} (kcal/mol)
ALM4-23/c-Src	-60.5092 ± 4.6782
Compound **1**/c-Src	-54.5552 ± 5.0053
Compound **2**/c-Src	-55.4212 ± 4.6790
Compound **3**/c-Src	-57.3972 ± 6.1196
Compound **4**/c-Src	-59.0758 ± 4.9568
Compound **5**/c-Src	-62.1696 ± 4.5874
Compound **6**/c-Src	-62.5038 ± 5.6756
Compound **7**/c-Src	-62.7515 ± 5.0422
Compound **8**/c-Src	-61.8230 ± 5.1936
Compound **9**/c-Src	-62.2246 ± 4.0046
Compound **10**/c-Src	-62.9063 ± 5.7363

Table 10. *Free energy of binding of the complexes between c-Src and **ALM4-23** congeneric compounds.*

The computational results on compounds **5**, **6**, **7**, **8**, **9**, **10** confirm a potential increase in affinity against c-Src contrary to compounds **1**, **2**, **3**, **4** that should not lead to an improvement of this property.

Moreover, these data suggest a preliminary SAR, where the introduction in para position of the phenol OH of halogens or a methyl group, could improve the potency against c-Src, as demonstrated by compounds **6-9**. Similarly, the addition in position 4 of the thiazole ring of a hydrogen bond donor group, like an amine (compound **5**), and the addition in ortho position of phenol OH of a hydrogen bond acceptor group, like a methoxy substituent (compound **10**), could improve the potency.

Conversely, the replacement of thiazole with a heterocyclic ring endowed with adjacent hydrogen bond acceptor and donor groups, like a pyrazole (compound **4**), should not lead to an improvement of activity. This last result is quite surprising because for compound **5**, which has the above-mentioned hydrogen bond acceptor-donor motif in its 4-amino-thiazole moiety, ΔG_{bin} improved with respect to the hit compound. A possible explanation can be based on the different distance between the hydrogen bond donor group and the acceptor oxygen of Glu339, that is higher for compound **4** in comparison to compound **5** (2.3 vs 2.0 Å, FIGURE 33). This could lead to the formation of a stronger hydrogen bond for compound **5**, thus explaining its most favourable ΔG_{bin}.
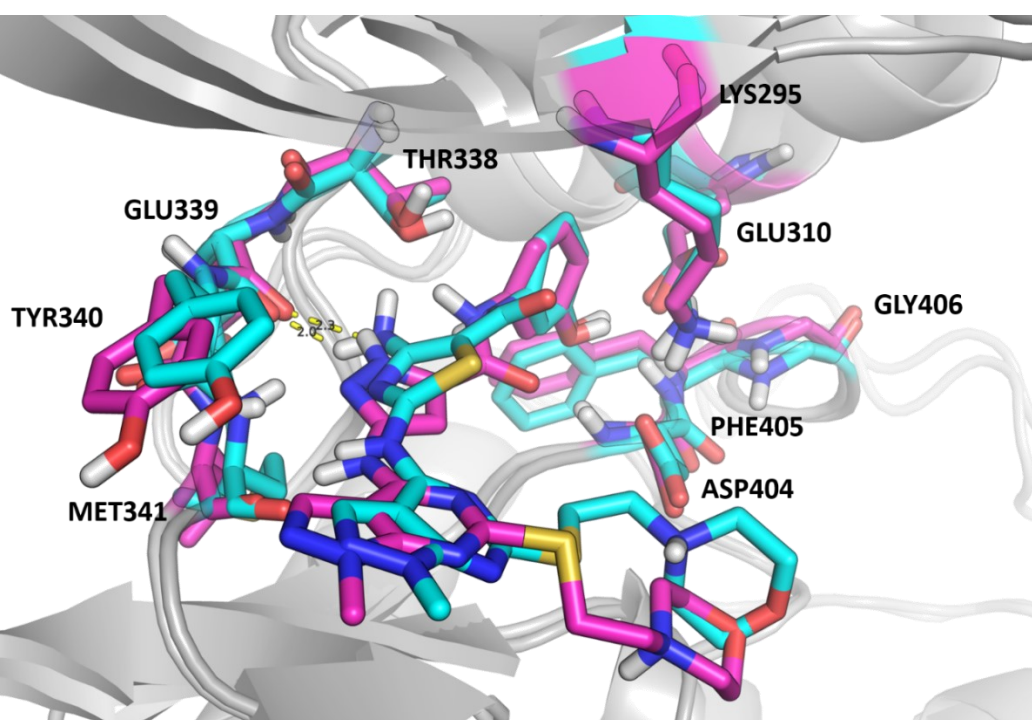
***Figure 33**. Overlapping of representative frames of the most populated clusters related to trajectories of compound 4 (magenta) and compound 5 (cyan).*

Finally, also the different substituents introduced in C6 of pyrazolo[3,4-*d*]pyrimidine scaffold, should not lead to an improvement of activity against c-Src, as demonstrated by compounds **1**,

2, **3**. A possible explanation of this negative trend could be the fact that C6 substituents are exposed to the solvent and therefore are not able to stably interact with the binding site.

3.6 Conclusions and future perspectives

In summary, the application of computational methods such as molecular docking and GaMD simulations gave suggestions on the possible mode of action of the pyrazolo[3,4-*d*]pyrimidine c-Src inhibitor **ALM4-23**. The results show that **ALM4-23** stabilizes the active conformation of the protein, thus acting as a potential type I kinase inhibitor. Starting from this observation, a hit optimization protocol, comprised of cluster analysis, affinity grid maps, molecular docking, GaMD simulations and MM-GBSA, was also applied. As a result, ten derivatives were identified that form very stable complexes with the active conformation of the kinase. Six of them have been predicted with a potential increase in activity against c-Src. They all will be synthesized and tested to confirm the reliability of the computational approach and then rationalize its future application or implementation.

To date only the compounds **6** and **8** have been synthesized ant tested against c-Src, showing IC_{50} values of 0.74 ± 0.11 nM and 257.7 ± 25.44 nM respectively. Enzymatic assay confirms the computational prediction for compound **6** which is characterized by an activity against c-Src ~32 fold higher than **ALM4-23** and comparable to that of dasatinib, differently from compound **8** for which we obtained a worsening of inhibitory activity compared to **ALM4-23.** **ALM4-23** showed $IC_{50} > 50$ µM on hepatocarcinoma HEPG2 cell line. In the attempt to justify the unexpected inactivity on this assay and the significant activity found in the enzyme assay, ADME properties of **ALM4-23** were determined. In particular, the aqueous solubility, metabolic stability, passive membrane permeability towards gastrointestinal membrane, and membrane retention were evaluated (TABLE 11).

	Aqueous solubility ($\mu g \cdot mL^{-1}$)	Metabolic stability (%)	Gastrointestinal P_{app} (10^{-6} cm·s^{-1}) (Membrane retention) %
ALM4-23	0.03	99.4	0.1 (3.1)

Table 11. *In vitro ADME profile of compound **ALM4-23.***

The low water solubility and poor passive membrane permeability led to the hypothesis that cellular inactivity could be due to compound inability to reach the intracellular target. These results strongly suggest that an improvement of the pharmacokinetic properties of **ALM4-23** and its derivatives is needed, by the functionalization of the solvent exposed N1 and C6 positions of the pyrazolo[3,4-*d*]pyrimidine scaffold (FIGURE 34).
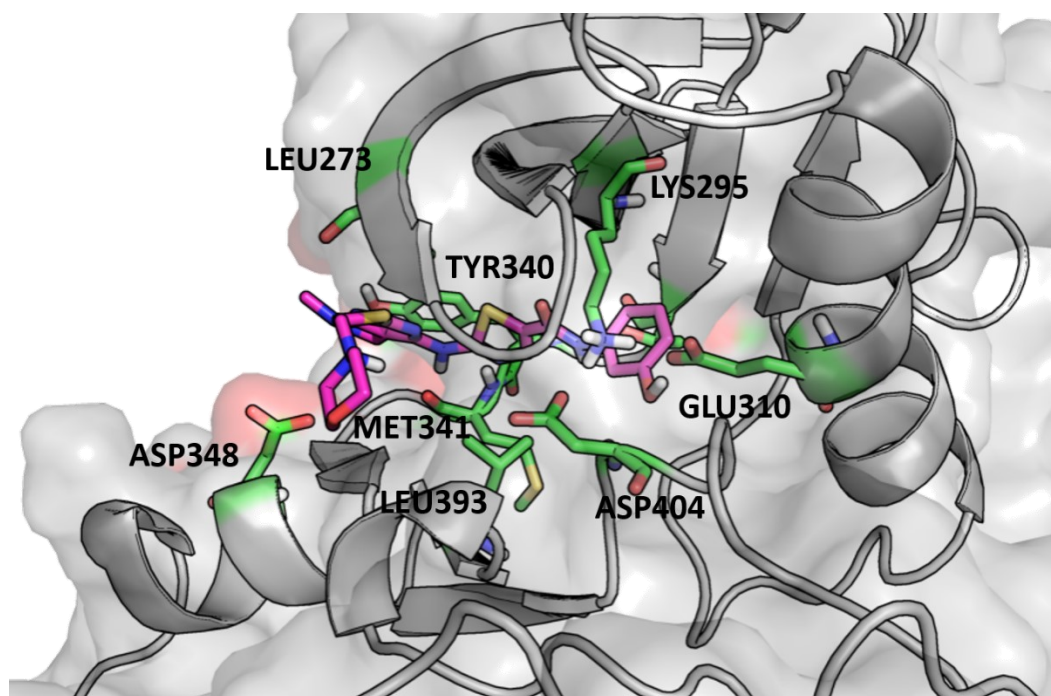
Figure 34. *Binding mode of **ALM4-23** in the representative frame of the most populated cluster. Ligand is reported in magenta sticks while interacting residues are reported in green sticks. The protein is reported as cartoon and surface.*

4 SBVS FOR THE IDENTIFICATION OF NEW C-SRC TYPE-II INHIBITORS

4.1 State of the art and aim of the project

As previously reported, c-Src represents an interesting pharmacological target for the treatment of many forms of cancer but it is also a protein kinase structurally very plastic due to its ability to assume an active and two inactive (CHO and DFG-out) conformational states.

Starting from these observations, the aim of project was to identify new c-Src inhibitors through the application of a SBVS procedure to a database of commercially available compounds. In this VS procedure, the 3D structure of the kinase domain of c-Src in DFG-out conformation was used with the further aim of exploiting the potential advantages, in terms of selectivity and drug target residence time, that could derive from targeting an inactive conformation.

4.2 Workflow

The SBVS procedure consists of the following steps:

- Selection and preparation of the crystallographic structure of the c-Src kinase domain in DFG-out conformation.

- Preparation of the Molport "All stock compounds" database of commercially available compounds, consisting of about 7 million molecules.

- Construction of a structure-based pharmacophore model that was used as a filter to virtually screen the Molport database, to identify a subset of compounds with the required steric and electronic features for the interaction with the ATP binding site of c-Src in DFG-out conformation. As a result, the pharmacophore-based filtration was passed by 344,991 molecules.

- Molecular docking of the remaining compounds within the ATP binding site and screening of the obtained docking poses against the previously build pharmacophore model. This second pharmacophore-based filtration was passed by 12,089 compounds.

- Once evaluated, through an enrichment simulation, the ability of the docking scoring function to classify active compounds at the beginning of its ranking list, the remaining 12,089 molecules were further reduced by setting a score threshold. As a result, 174 molecules were retained.

- Visual inspection of the docking poses of the remaining 174 molecules, to select a subset of compounds to purchase and submit to the biological assays.

4.3 Materials and methods

4.3.1 Protein preparation

Chain A of the X-ray structure 2OIQ was prepared with the same procedure reported in section 3.4.1, by using the Protein Preparation Wizard workflow of the Maestro Suite.

4.3.2 Database preparation

The compounds of the Molport Database were prepared in terms of 3D structures, tautomeric and ionization states with the same procedure reported in section 3.4.2.2.

4.3.3 Structure-based pharmacophore model generation and validation

The five feature structure-based pharmacophore model was manually built on the previously prepared complex by using the Phase software.[72] In detail, it is constituted by two directional hydrogen bond acceptor groups (A1 and A2), one directional hydrogen bond donor group (D3), one aromatic ring (R5), and one hydrophobic group (H4). All the pharmacophore features have a matching tolerance of 2 Å apart from A2 which has a value of 2.2 Å. The pharmacophore model is completed by excluded volumes that were generated:

- Ignoring the receptor atoms whose surface was within 3.2 Å of the ligand surface;
- Using a fixed van der Waals radii scaling factor of 0.05;
- Limiting the thickness of the excluded volumes shell, through the inclusion of the sphere only for receptor atoms that are within 4.5 Å distance from the ligand.

The evaluation of the sensitivity of the pharmacophore model in recognizing of c-Src Type-II ligands having the required chemical features was made as follows:

- Once obtained the bioactive conformations of the fifteen ligands of the test set from the corresponding crystallographic complexes, hydrogen atoms were added and ionization states were generated by using Applyhtreat and Ionizer at pH 7.4.
- The prepared ligands were submitted to conformational analysis by using the 'rapid' conformational sampling method. In detail, up to a maximum of 100 conformers per ligand (retaining the bioactive one) were generated with a conformational energy window of 1 kcal/mol. The choice to generate, starting from the bioactive conformation,

a set of conformers energetically assimilable, was made with the aim of giving a slight conformational freedom to the test set ligands. This should avoid the potential clashes that would occur in the next step, by screening only the bioactive conformations against a pharmacophore model generated on a receptor conformationally adapted to another ligand (imatinib). The conformational analysis was performed by using phase_database tool.

- The conformers library of the test set ligands was screened against the pharmacophore model by imposing that all the pharmacophore features were matched to pass the screen and reporting one hit per molecule as the final result. This step was performed by using phase_screen tool.

4.3.4 Database conformational analysis

The previously prepared Molport Database was submitted to conformational analysis by using the 'rapid' conformational sampling method. In detail, up to a maximum of 300 conformers per ligand (retaining the starting conformation) were generated within an energy window of 30 kcal/mol.

The ability of the conformational analysis procedure to generate the hypothetical bioactive conformation of a potential c-Src Type-II inhibitor was evaluated as follows:

- The structure of the fifteen ligands of the test set were sketched in Maestro and prepared as reported in section 3.4.2.2.
- The prepared ligands were submitted to the same conformational analysis reported for the Molport database compounds.
- The conformers library was screened against the previously generated pharmacophore model, by imposing that all the pharmacophore features were matched to pass the screen and reporting one hit per molecule as the final result.

4.3.5 First pharmacophore-based virtual screening

The pharmacophore model was used as filter to screen the library of conformers produced for the ligands of Molport database, by requiring that all the pharmacophore features were matched to pass the filtration and reporting as result one hit per molecule.

4.3.6 Docking simulations

4.3.6.1 Validation of docking procedure

Docking simulations were performed by using the standard precision (SP) procedure of the Glide software. The seven c-Src Type-II ligands were first sketched in Maestro and prepared as reported in section 3.4.2.2. Consequently, the ligands were subjected to the previously validated conformational analysis and screened against the pharmacophore model. Because two out of seven compounds (ligands of 2OIQ and 3G6G) are characterized by four out five chemical features defined in the pharmacophore model, a matching of four out of five features was imposed in this case to pass the screening.

The alignment of α carbons of 2OIQ (chain A), 3G6H (chain B), 3G6G (chain A), 3EL8 (chain A), 4YBJ (chain A), 4AGW (chain B), 3F3V (chain A) shows a great flexibility of the ATP biding site, that allows it to adapt differently to ligands. Based on these observations, the receptor grid was generated from the previously prepared crystallographic complex (chain A of 2OIQ) by slightly decreasing the van der Waals radii of the receptor non-polar atoms (scaling factor set to 0.9), to facilitate the docking of the six c-Src Type-II ligands that were not co-crystallized in the protein structure used for simulations. Moreover, the receptor grid was centred on the co-crystallized ligand and the protein was kept rigid. Finally, the conformations of the seven ligands that passed the pharmacophore-based filtration were used as input for docking simulations, with flexible ligand sampling.

RMSD calculation between the docked and the experimental poses was performed with smart_rms of Gold suite[59] considering only the ligand heavy atoms.

4.3.6.2   Molecular Docking of Molport compounds

The 344,991 remaining compounds of the Molport database were docked in the receptor grid produced in the validation step and by using the same docking procedure.

4.3.7 Second pharmacophore-based virtual screening

Only the pharmacophore sites were generated for the docking poses of the remaining 344,991 compounds. The pharmacophore filtration was set to find and to score matches without aligning the docking poses to the pharmacophore model, with the aim of maintaining the docked orientation of the ligands within the binding site. Moreover, compounds must match all the pharmacophore features to pass the filtration.
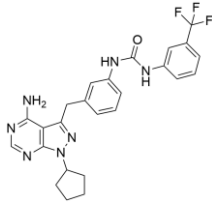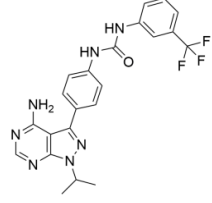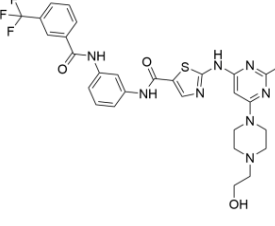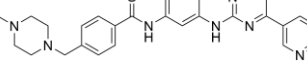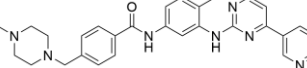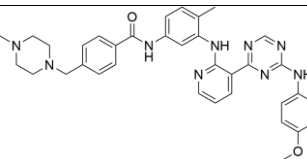
4.3.8 Enrichment assessment of Glide SP scoring function
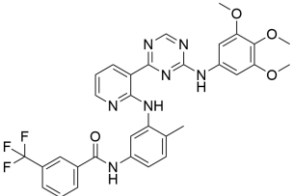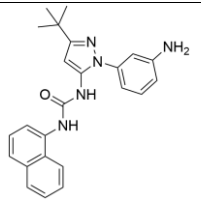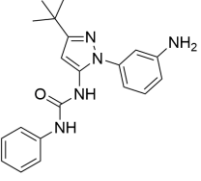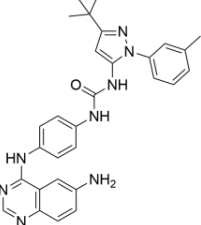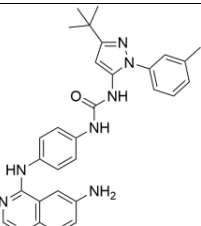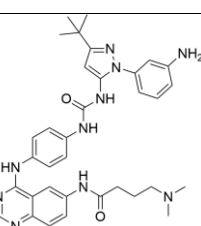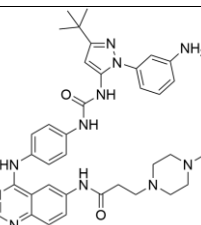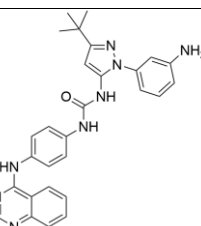
SMILES of the three active ligands of 3G6H, 3G6G, and 3EL8 were used as input to the online automated tool of DUD-E[73] which produced 50 decoys for each ligand.[74] Decoys were prepared by adding hydrogen atoms and minimized with OPLS_2005 forcefield, and then docked in the receptor grid. Finally, the resulting test set of actives and decoys was ranked by Glide score (SP procedure).

4.4 Results and discussion

4.4.1 Selection of the crystallographic structure

Necessary requirement to carry out the SBVS procedure was the selection of a crystallographic structure representative of the c-Src kinase domain in DFG-out conformation. At the time of this study, the sixteen available structures were all from chicken (TABLE 12). As reported in Section 3, in comparison to the kinase domain of the human c-Src, they differ by only two amino acid sited in solvent exposed regions and away from the ATP binding site. Based on these observations, a chicken c-Src kinase domain was selected for the SBVS procedure aimed at identifying inhibitors of the human protein.

PDB	CO-CRYSTALLIZED LIGAND	CHAINS	COMPLEX DEFINED IN CHAIN	RESOLUTION	MUTATION
3EL7		A	A = yes	2.8 Å	
3EL8		A-B	A = yes B = no	2.3 Å	
4YBJ		A-B	A = yes B = yes	2.61 Å	
2OIQ		A-B	A = yes B = no	2.07 Å	
3OEZ		A-B	A = yes B = yes	2.4 Å	L317I
3G6G		A-B	A = yes B = yes	2.31 Å	

3G6H		A-B	A = no B = yes	2.35 Å	T338I
3F3T		A-B	A = yes B = no	2.5 Å	S345C
3F3U		A-B	A = yes B = yes	2.5 Å	S345C
3F3V		A-B	A = yes B = yes	2.6 Å	S345C
3F3W		A-B	A = yes B = yes	2.6 Å	S345C T338M
3TZ7		A-B	A = yes B = yes	3.3 Å	S345C
3TZ8		A-B	A = yes B = yes	2.7 Å	S345C
3TZ9		A-B	A = yes B = yes	3.1 Å	S345C
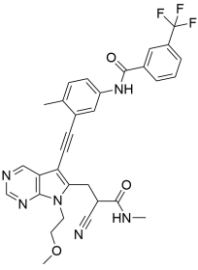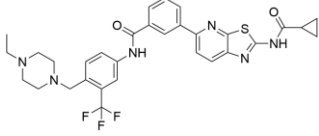
5SYS		A-B	A = yes B = yes	2.8 Å	V281C
4AGW		A-B	A = yes B = yes	2.6 Å	

Table 12. 3D structures of the c-Src kinase domain in DFG-out inactive conformation available on PDB.

Among them, the attention was focused on the X-ray crystal structures where the catalytic domains are co-crystallized with ligands able to induce the creation of a large DFG allosteric pocket, such as 2OIQ, 3G6G and 4AGW. The selection of a structure with a well exposed allosteric pocket should facilitate its targeting in the steps of the VS procedure, in which computational methods were used that maintain the protein in a locked conformation. Among these three structures, chain A of 2OIQ was selected as starting point of the following computational studies due to its higher resolution. The selected crystallographic complex was then optimized.

4.4.2 Database preparation

Another necessary requirement for a VS procedure is a database of compounds to be virtually screened. The Molport database "All stock compounds", consisting of 7,608,298 molecules, was selected to perform the SBVS procedure. Before the screening, compounds were optimized in terms of 3D structure, tautomeric and ionization states at pH 7.4.

4.4.3 Structure-based pharmacophore model generation

A structure-based pharmacophore model was built on the previously selected and prepared crystallographic complex between c-Src and imatinib (chain A of 2OIQ), by using the Phase software. The visual inspection of eleven complexes where c-Src is in DFG-out conformation and co-crystallized with Type-II and Type-III binders allowed to identify the main chemical features responsible for the interaction between the ligands and the ATP binding site (FIGURE 35).

It is worth noting that in the set of analysed structures there are six complexes (3EL8, 3EL7, 4YBJ, 2OIQ, 3G6G, and 4AGW) with the chicken wild-type kinase domain and not with the human sequence that will be the target of this study. Despite this, as previously mentioned, the amino acid differences between these two wild-type forms of the kinase domain are sited in solvent exposed regions and away from the ATP binding site. This makes reasonable to believe, as is generally reported in the literature, that the interactions of the ligands with chicken crystallographic c-Src and the consequent activity data [34,48,54,75–78] (TABLE 13) could be the same with the human sequence.

PDB LIGAND	ACTIVITY ON C-SRC
3EL8	IC_{50}: 0.025 µM
3EL7	IC_{50}: 0.480 µM
4YBJ	K_d: 0.53 nM
2OIQ	IC_{50}: 24.36 µM
3G6G	IC_{50}: 2.8 nM
4AGW	Reported as 'Src exhibited potent binding to the ligand'
3F3T	IC_{50}: 64.1 µM
3F3V	IC_{50}: 0.021 µM
3TZ7	IC_{50}: 1.1 nM
3TZ9	IC_{50}: 38 nM
3G6H	IC_{50} wild type: 4.6 nM IC_{50} (T338I): 6.4 nM

Table 13. Activity data of co-crystallized ligands against c-Src

Similarly, the ligands of additional four complexes (3F3T, 3F3V, 3TZ7 and 3TZ9) are co-crystallized with a chicken kinase domain having, in respect to the wild type form, a point mutation (S345C) that is outside the ATP binding site and solvent exposed. Finally, in the last complex 3G6H, the ligand is co-crystallized with the chicken kinase mutated at the gatekeeper residue (T338I) within the ATP binding site. This complex was included in the set of analysed structures because its ligand shows a similar binding mode with the structurally related ligand of 3G6G, that is co-crystallized with the chicken wild type form. This suggests that the binding mode of the 3G6H ligand within the ATP binding site of the wild type form should be the same

observed in the complex 3G6H, as further highlighted by the comparable efficacy of the compound against the wild type and mutated form of the enzyme (TABLE 13).

From the visual analysis of these crystallographic structures, several common features can be found:

- Eleven complexes have a hydrogen bond donor group able to interact with Glu310 of αC helix.
- Eleven have a hydrogen bond acceptor group able to interact with Asp404 of DFG motif.
- Eleven have a substituted aryl moiety that occupies the DFG allosteric pocket.
- Ten have a hydrogen bond acceptor group able to interact with Met341 of hinge region (the Type-III ligand of the 3F3T complex lacks this feature).
- Nine have a hydrophobic group located in the additional DFG pocket.
- Four have a hydrogen bond donor group able to interact with Met341 of hinge region.
- Three have a hydrogen bond donor group able to interact with Val383.
- Two have a hydrogen bond donor group able to interact with the gatekeeper residue of hinge region Thr338.
- Two have a hydrogen bond donor group able to interact with Glu339.
- One has a hydrogen bond donor group able to interact with Asp348.
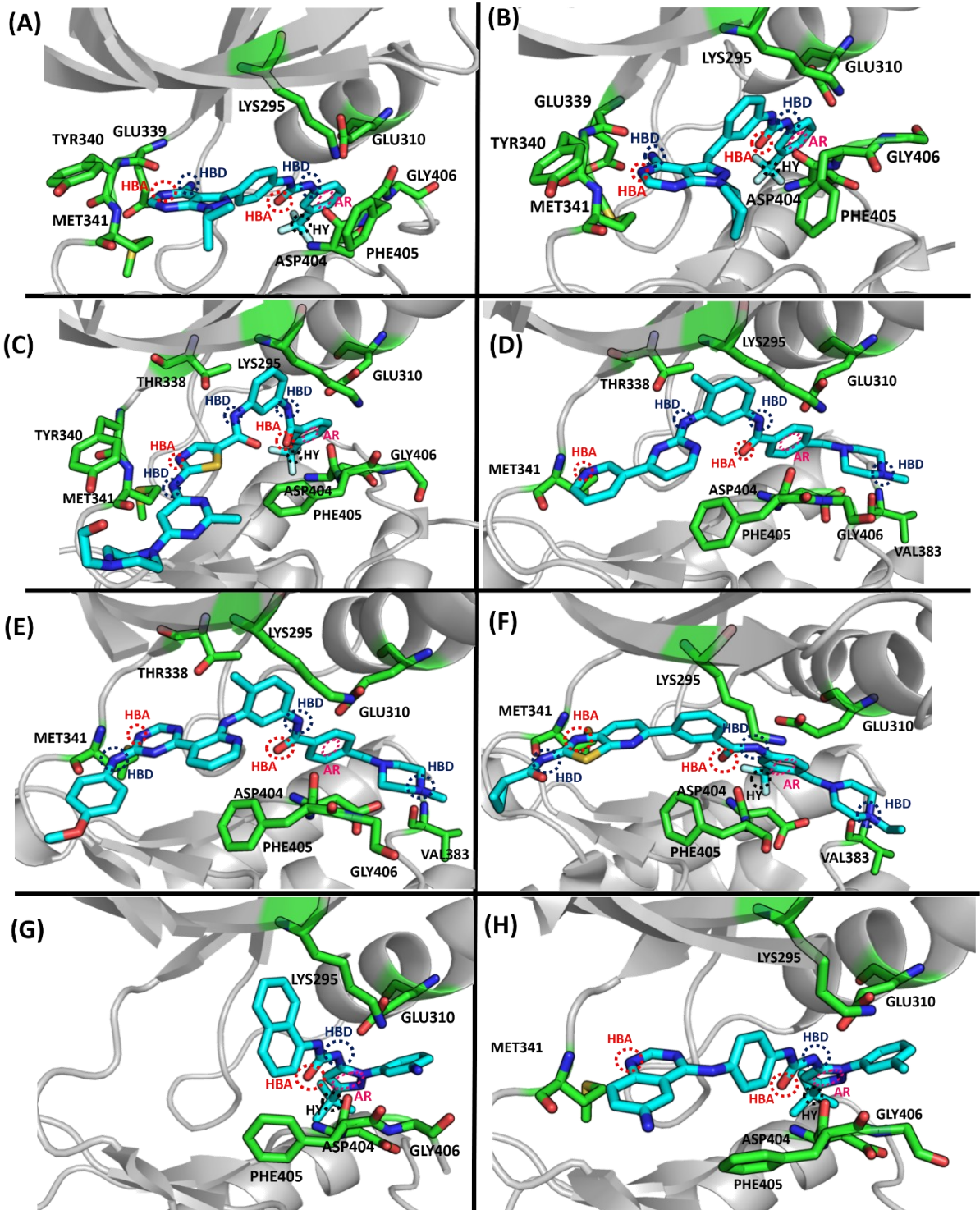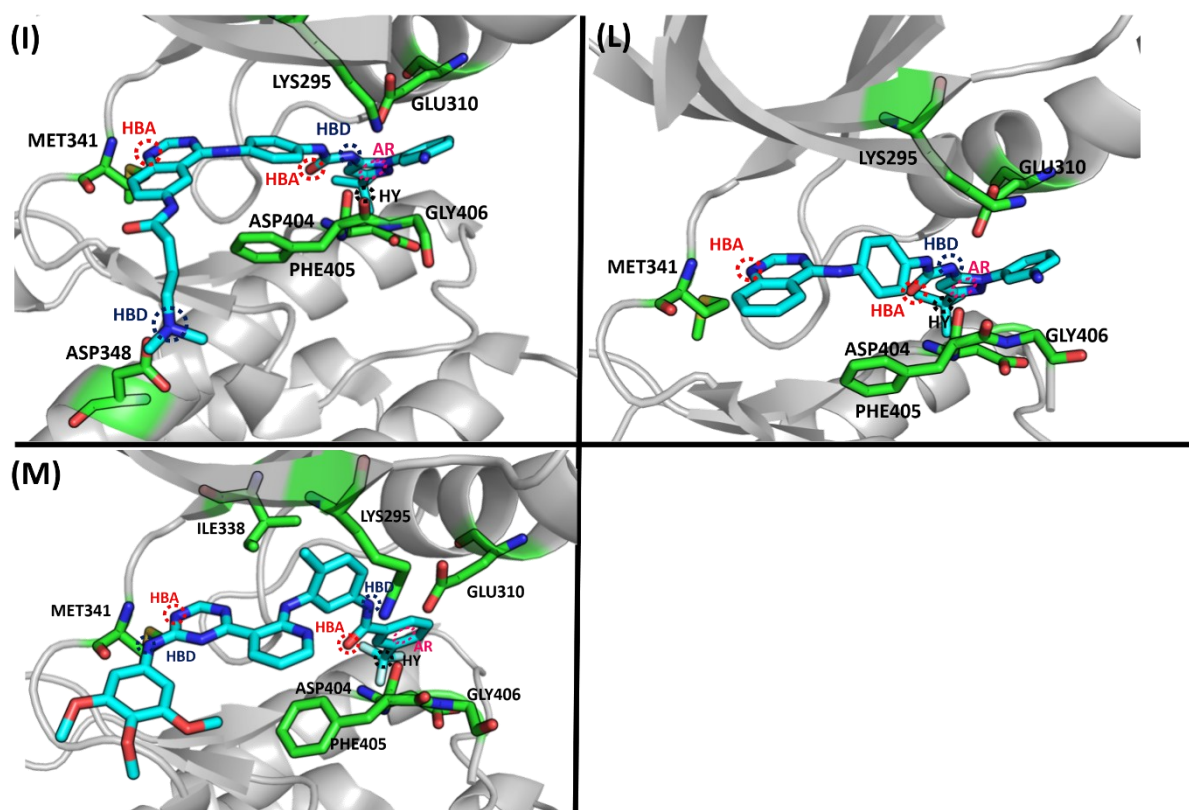
Figure 35. *Complexes from 3EL8 (chain A) (A), 3EL7 (chain A) (B), 4YBJ (chain A) (C), 2OIQ (chain A) (D), 3G6G (chain A) (E), 4AGW (chain B) (F), 3F3T (chain A) (G), 3F3V (chain A) (H), 3TZ7 (chain A) (I), 3TZ9 (chain A) (L) and 3G6H (chain B) (M). The chemical features responsible for the interaction between these compounds and the ATP binding site are reported as circles. In particular, hydrogen bond donor groups (HBD) are in blue, hydrogen bond acceptor groups (HBA) are in red, aryl moieties (AR) are in pink, hydrophobic groups (HY) are in black. Ligands are reported in cyan sticks while interacting residues and those representative of the conformational state of the protein are reported in green.*

Based on these observations, the five most conserved chemical features were included in the resulting structure-based pharmacophore model (FIGURE 36), that was further supported by a recent study in which the first three most conserved chemical features are considered necessary for the development of Type-II kinase inhibitors.[31]
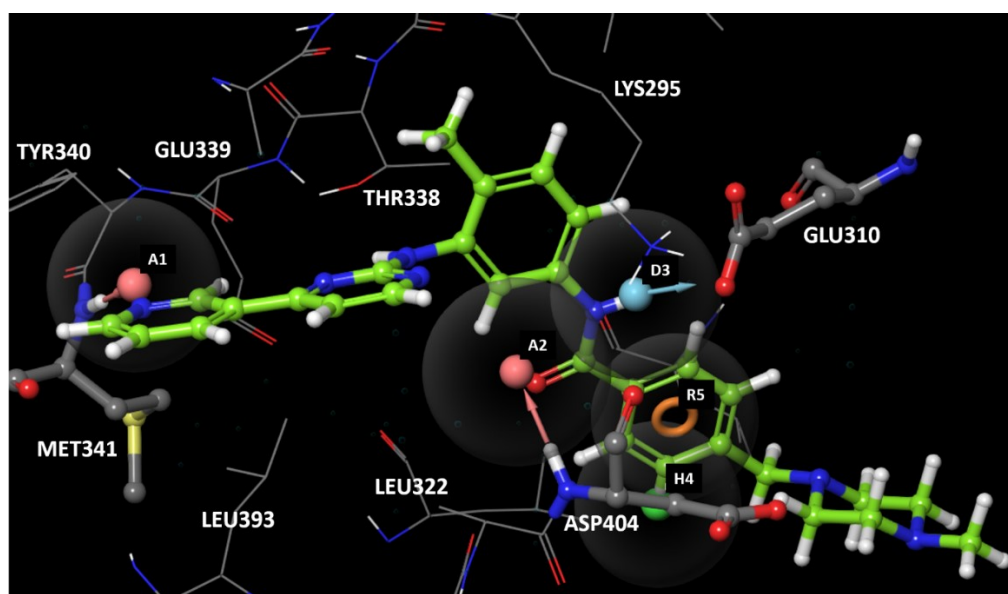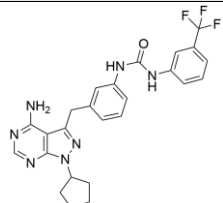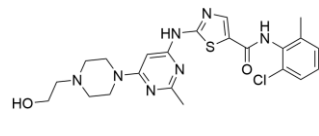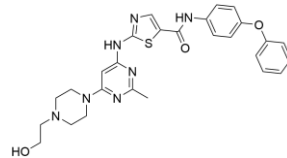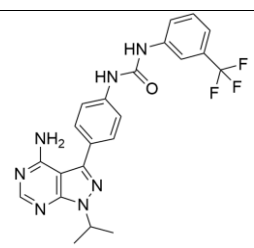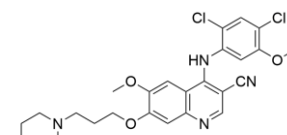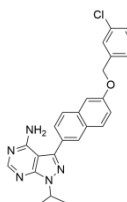
Figure 36. *Five feature structure-based pharmacophore model constituted by two hydrogen bond acceptor groups (A1 and A2), one hydrogen bond donor group (D3), one aromatic ring (R5), and one hydrophobic group (H4). They map the co-crystallized ligand (imatinib) respectively, the pyridine nitrogen atom, which interacts with Met341, the amide oxygen atom, which interacts with Asp404, the amide hydrogen atom, which interacts with Glu310, the phenyl group of the benzyl-piperazine moiety, which is sited in the additional DFG allosteric pocket, and one of its ortho hydrogen atoms (which is replaced by hydrophobic groups in the studied complexes). Finally, the small blue spheres represent the excluded volumes, that are the regions of the space representative of the receptor atoms and that can not be occupied by the molecules that will be aligned to the pharmacophoric model.*

The reliability of the pharmacophore model in the recognition of Type-II ligands with the required chemical features was evaluated by screening a set of c-Src Type-II, I and I½ binders (TABLE 14) in their bioactive conformations.

TYPE-II BINDERS	TYPE-I BINDERS	TYPE-I½ BINDERS
3EL7	3G5D	4YBK
3EL8	4MXO	4DGG
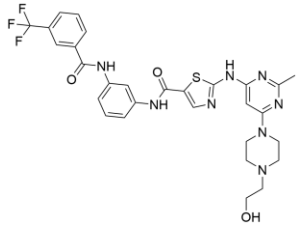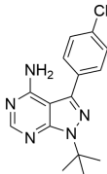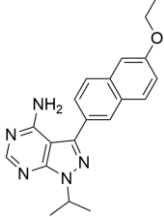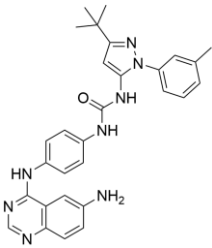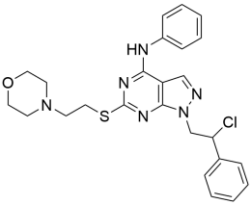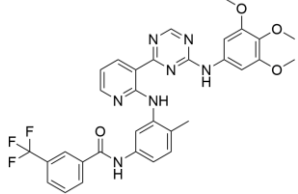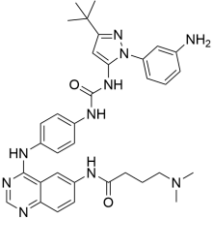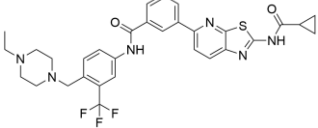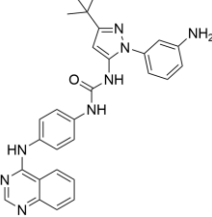
4YBJ	3GEQ	3UQF
3F3V	4O2P	
3G6H		
3TZ7		
4AGW		
3TZ9		

Table 14. *Chemical structures of the test set ligands and PDB IDs from which the ligand bioactive conformations were obtained.*

The pharmacophore model was able to retrieve seven out of the eight Type-II ligands discarding all the remaining compounds (FIGURE 37).
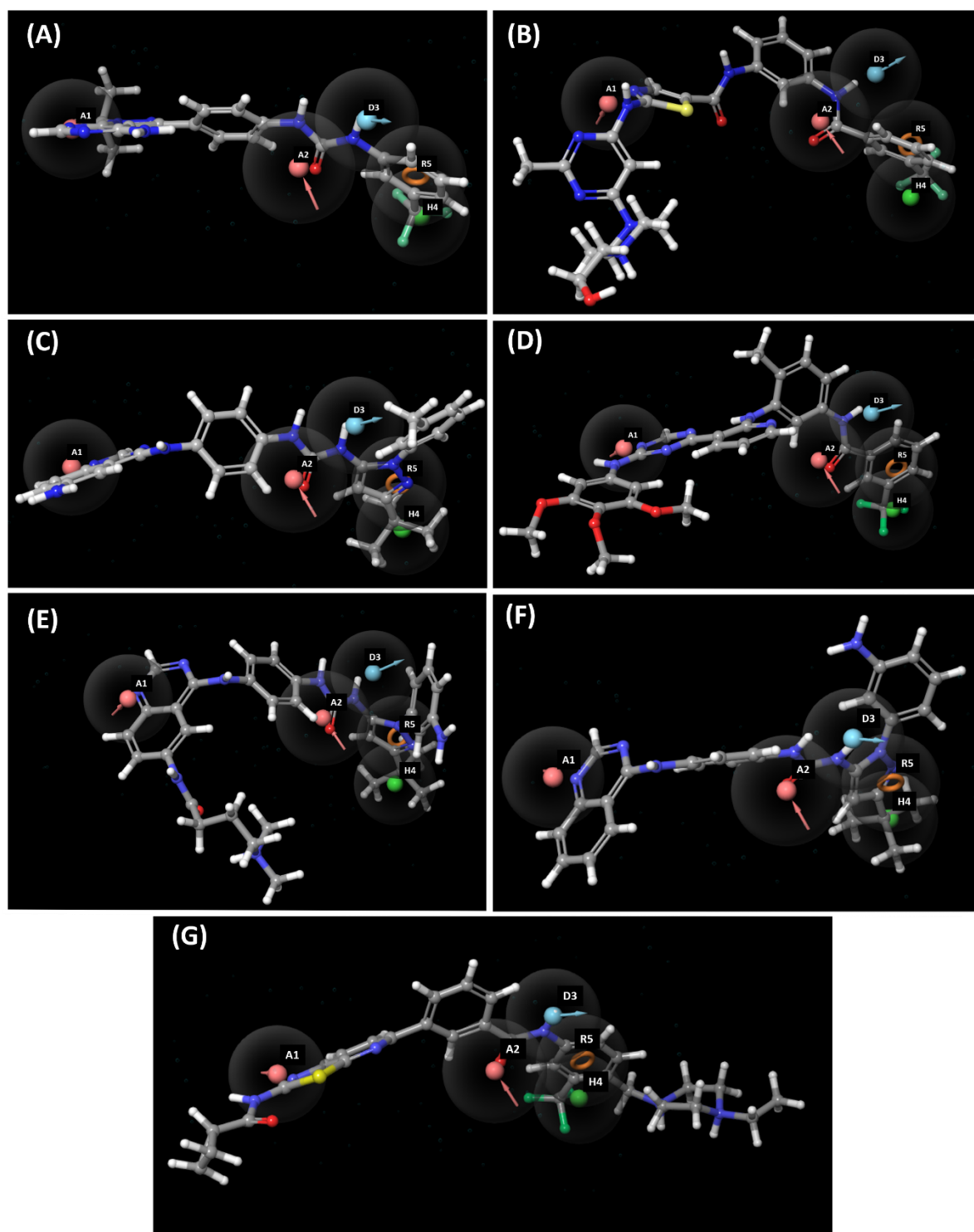
Figure 37. *Test set ligands that passed the pharmacophore-based virtual screening superimposed on the pharmacophore model: 3EL8 ligand (A), 4YBJ ligand (B), 3F3V ligand (C), 3G6H ligand (D), 3TZ7 ligand (E), 3TZ9 ligand (F), 4AGW ligand (G).*

73

4.4.4 Database conformational analysis

Following the construction and validation of the pharmacophore model, ligands were subjected to conformational analysis, producing up to a maximum of 300 conformers per molecule. The ability of the conformational analysis to generate the hypothetical bioactive conformation of a potential c-Src Type-II inhibitor was evaluated by producing the conformers library for the ligands of the previously used test set and by screening it against the pharmacophore model. Out of the fifteen c-Src binders, nine passed the pharmacophore filtration: the eight c-Src Type-II binders and the Type-I½ ligand of 4YBK. All the Type-II binders passed the screening in a conformation very similar to their bioactive conformer (FIGURE 38), validating the reliability of the conformational analysis.
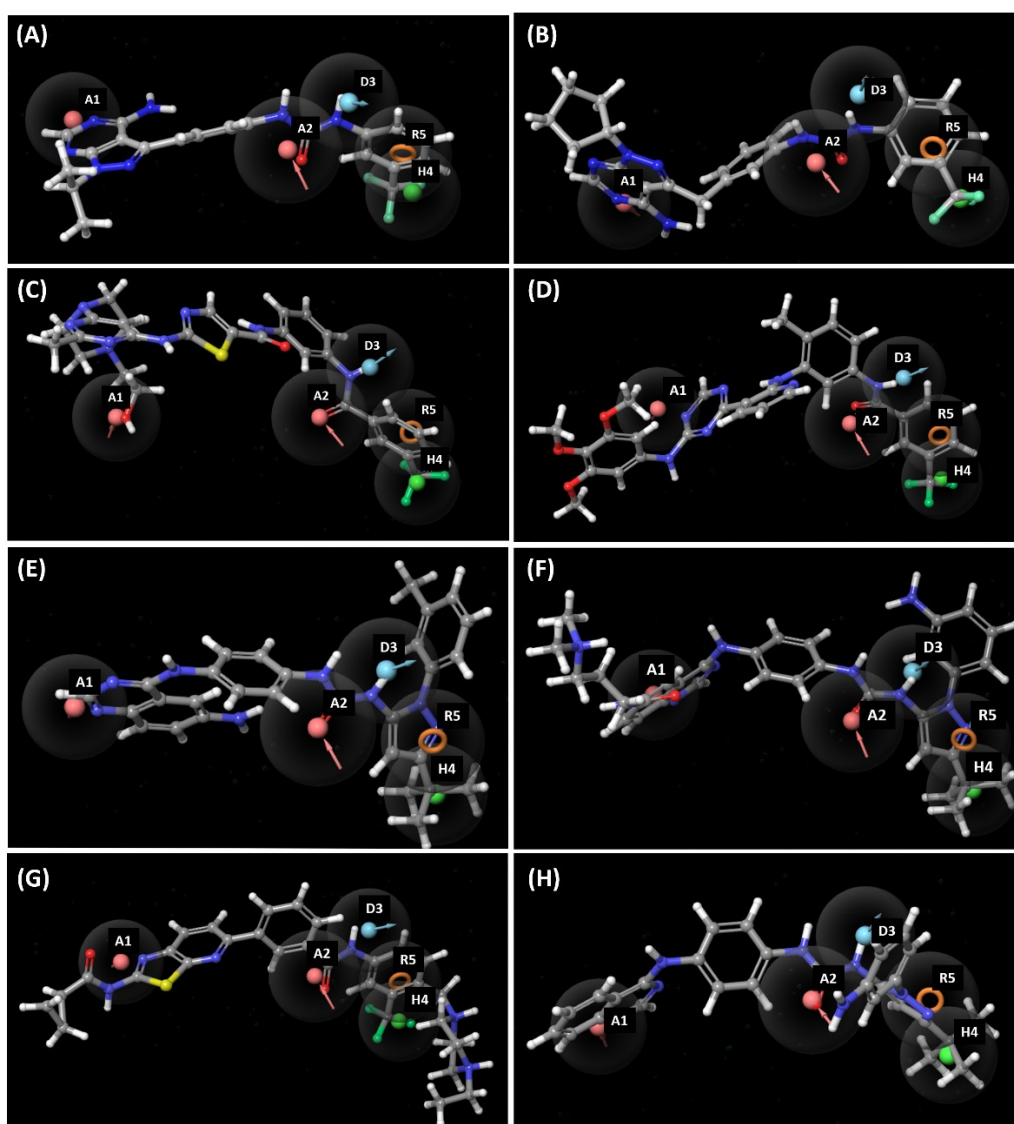
Figure 38. *Conformations of the test set ligands that passed the pharmacophore-based virtual screening superimposed on the pharmacophore model: 3EL8 ligand (A), 3EL7 ligand (B), 4YBJ ligand (C), 3G6H ligand (D), 3F3V ligand (E), 3TZ7 ligand (F), 4AGW ligand (G), 3TZ9 ligand (H).*

4.4.5 First pharmacophore-based virtual screening

The pharmacophore model was then used as a filter to screen the library of conformers produced for the ligands of Molport database, to identify a subset of compounds with the required steric and electronic features for the interaction with the ATP binding site of c-Src in DFG-out conformation. As a result, out of 7,608,298 compounds, 344,991 passed the pharmacophore-based filtration.

4.4.6 Molecular docking

The 344,991 remaining compounds were then subjected to docking calculations. The use of molecular docking in this step of SBVS procedure was considered necessary to obtain a more accurate evaluation of the binding modes of these molecules. In fact, the excluded volumes of the structure-based pharmacophore model were defined only for a small number of receptor atoms and by scaling their van der Waals radii. Consequently, the conformations of the molecules that passed the pharmacophore filtration could not coincide perfectly with the effective binding mode of such compounds. For this reason, application of molecular docking is expected to produce more reliable binding modes than those found within the pharmacophore-based filtration.

In detail, the 344,991 remaining compounds were docked in the ATP binding site of 2OIQ, by using Glide software. The reliability of the applied computational docking protocol was verified by simulating the ligand-protein interactions of seven crystallographic complexes (2OIQ, 4YBJ, 4AGW, 3G6H, 3G6G, 3F3V, and 3EL8) in which c-Src is co-crystallized with Type-II binders. As a result, the docking procedure was able to reproduce the experimental binding mode of five out of seven ligands (RMSD between the docking pose and the experimental binding mode ≤ 2 Å for all the five ligands) (FIGURE 39).
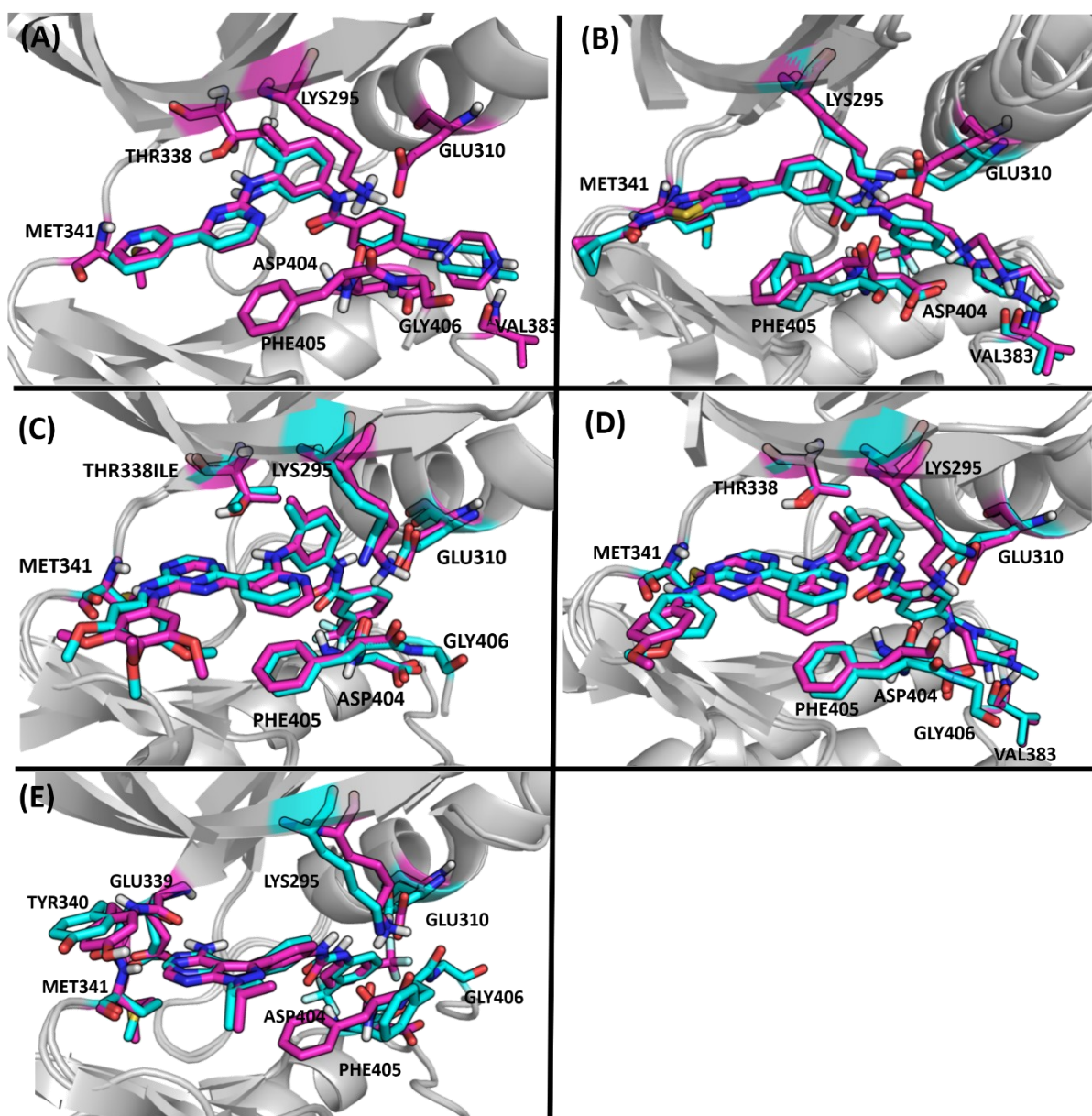
Figure 39. *Overlapping between docking (magenta sticks) and experimental complexes (cyan sticks) 2OIQ (A), 4AGW (B), 3G6H (C), 3G6G (D) and 3EL8 (E).*

4.4.7 Second pharmacophore-based virtual screening

The docking poses produced for the 344,991 remaining compounds were virtually screened against the pharmacophore model to retain only the molecules that preserved the required chemical features also in their docking pose. As a result, 12,089 compounds passed this second pharmacophore-based filtration.

4.4.8 Enrichment assessment and prioritization of compounds

Usually, docking is applied for its ability to discriminate between active and inactive compounds. To check for the reliability of the docking protocol, enrichment assessment is performed on a test set composed by experimentally confirmed active compounds and compounds presumed to be inactive against the examined target (decoys).[79]

Therefore, with the final aim of reducing in a rational way the number of compounds to analyse in the next step of the SBVS procedure, an enrichment experiment with Glide SP scoring function was carried out, creating a test set of active compounds and decoys.

To avoid bias deriving from inappropriate docking poses the ligands of 3G6H, 3G6G, and 3EL8 were chosen as active compounds. The docking procedure was able to perfectly reproduce their experimental binding modes. The three active ligands were then used as input for DUD-E to generate a set of 150 decoys (50 per active ligand) that were docked in the ATP binding site of the enzyme. Docking poses were found only for 83 decoys. Finally, the test set of 86 molecules (actives and decoys) was ranked by Glide score. The three active ligands corresponded to the first three positions of the ranking list (TABLE 15).

POSITION	COMPOUND	GLIDE SCORE (kcal/mol)
1	3G6H ligand	-12.7
2	3G6G ligand	-12.1
3	3EL8 ligand	-11.3
4	Decoy-1	-10.1
5	Decoy-2	-10
6	Decoy-3	-9.8
7	Decoy-4	-9.8
8	Decoy-5	-9.6
9	Decoy-6	-9.6
10	Decoy-7	-9.6
11	Decoy-8	….
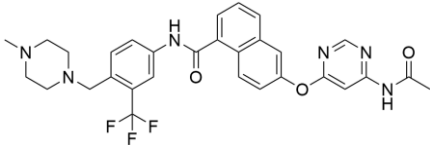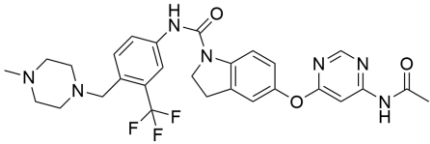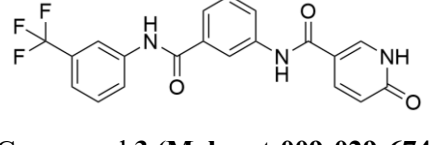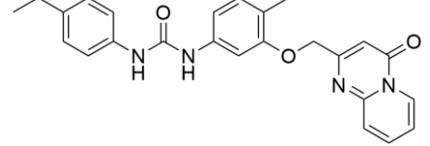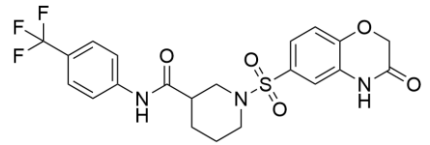
Table 15. *First 10 positions of the Glide score ranking list obtained for the test set of active compounds and decoys.*

Once verified the ability of the Glide SP scoring function to sharply separate the active compounds from decoys, the remaining 12,089 compounds of the Molport database were ranked with Glide score. The ranking list was then cut by using a Glide score threshold of -11.3

kcal/mol, that corresponded to the score of the last active compound calculated in the previous enrichment simulation. As a result, 174 molecules, were selected for the next step of SBVS procedure.

4.4.9 Final selection

The visual inspection of the docking poses of the resulting 174 compounds allowed the selection of nine compounds that were purchased (TABLE 16). It is worth noting that compounds **1** and **2** are inhibitors of several protein kinases,[80–82] but to the best of our knowledge, they were never tested against c-Src.

SELECTED COMPOUNDS	GLIDE SCORE (kcal/mol)
Compound **1 (Molport-046-594-284)**	-12.6
Compound **2 (Molport-046-417-231)**	-12.3
Compound **3 (Molport-009-029-674)**	-11.8
Compound **4 (Molport-007-879-128)**	-11.7
Compound **5 (Molport-007-812-255)**[*]	-11.5
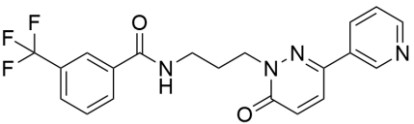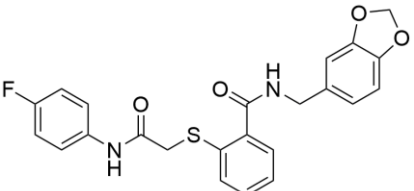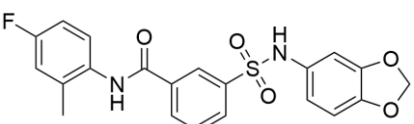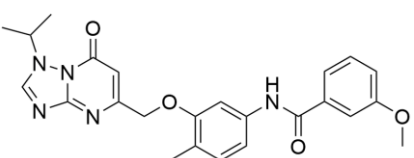
Compound **6 (Molport-005-299-712)**	-11.3
Compound **7 (Molport-005-516-038)**	-11.3
Compound **8 (Molport-009-512-929)**	-11.3
Compound **9 (Molport-007-706-790)**	-11.3

Table 16. *Chemical structures of the nine selected compounds and Glide score values associated to their predicted binding mode in the ATP binding site of c-Src.*
*Commercialized as racemic mixture. The Glide score value is related to S-enantiomer.

Moreover, among the 174 remaining compounds, several molecules are already known as c-Src inhibitors,[83–88] further supporting the reliability of the applied SBVS procedure in identifying the inhibitors of this protein kinase (TABLE 17).
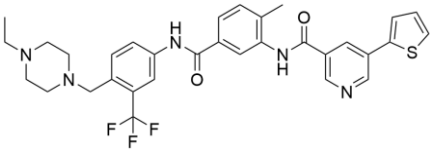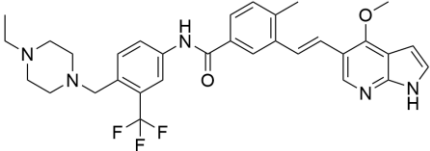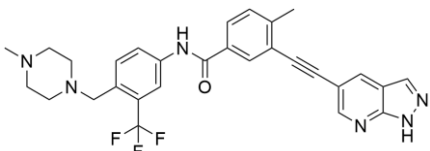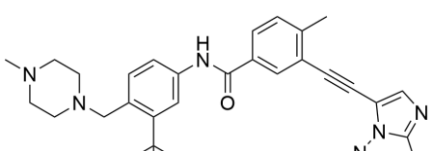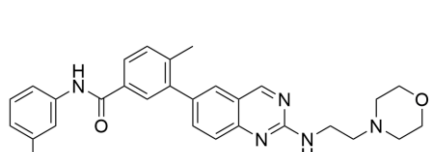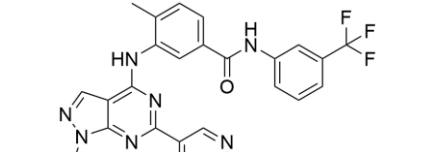
COMPOUND	GLIDE SCORE (kcal/mol)
ALW-II-41-27 (Molport-046-417-199)	-13.5
HG6-64-1 (Molport-039-139-670)	-13.4
GZD824 (Molport-042-676-166)	-12.9
Ponatinib (Molport-009-679-483)	-12.8
Compound-47 (Molport-021-804-913)	-12.6
NVP-BHG712 (Molport-016-633-311)	-12.4

Table 17. Chemical structures of the already known c-Src inhibitors prioritized by the SBVS procedure and Glide score values associated to their predicted binding mode in the ATP binding site of c-Src in DFG-out conformation.

The predicted binding mode of the nine selected compounds is shown in FIGURE 40. All of them map the five chemical features of the pharmacophore model. Moreover, additional interactions were found with the ATP binding site. In detail:

- Compound **1**: the acetamide moiety forms a second hydrogen bond with Met341, the naphthalene moiety made a cation-π interaction with Lys295, the di-protonated methyl-piperazine moiety provides two additional hydrogen bonds with His384 and Val383, respectively.

- Compound **2**: the acetamide moiety forms a second hydrogen bond with Met341, and the di-protonated methyl-piperazine moiety made a salt bridge with the side chain of Asp404.

- Compound **3**: the pyridone-carboxamide moiety forms hydrogen bonds with Glu339 and Thr338, and a cation-π interaction is found between the phenyl group of the benzodiamide moiety and Lys295.

- Compound **4**: the tolyl moiety forms a cation-π interaction with Lys295.

- Compound **5**: the morpholinone moiety forms a hydrogen bond with Glu339.

- Compound **7**: the sulfur atom forms a hydrogen bond with Lys295.

- Compound **8**: the sulfonamide group forms a hydrogen bond with Thr338, and the phenyl group of the benzamide moiety provides a cation-π interaction with Lys-295.

- Compound **9**: the phenyl group of the tolyl moiety forms a cation-π interaction with Lys295.
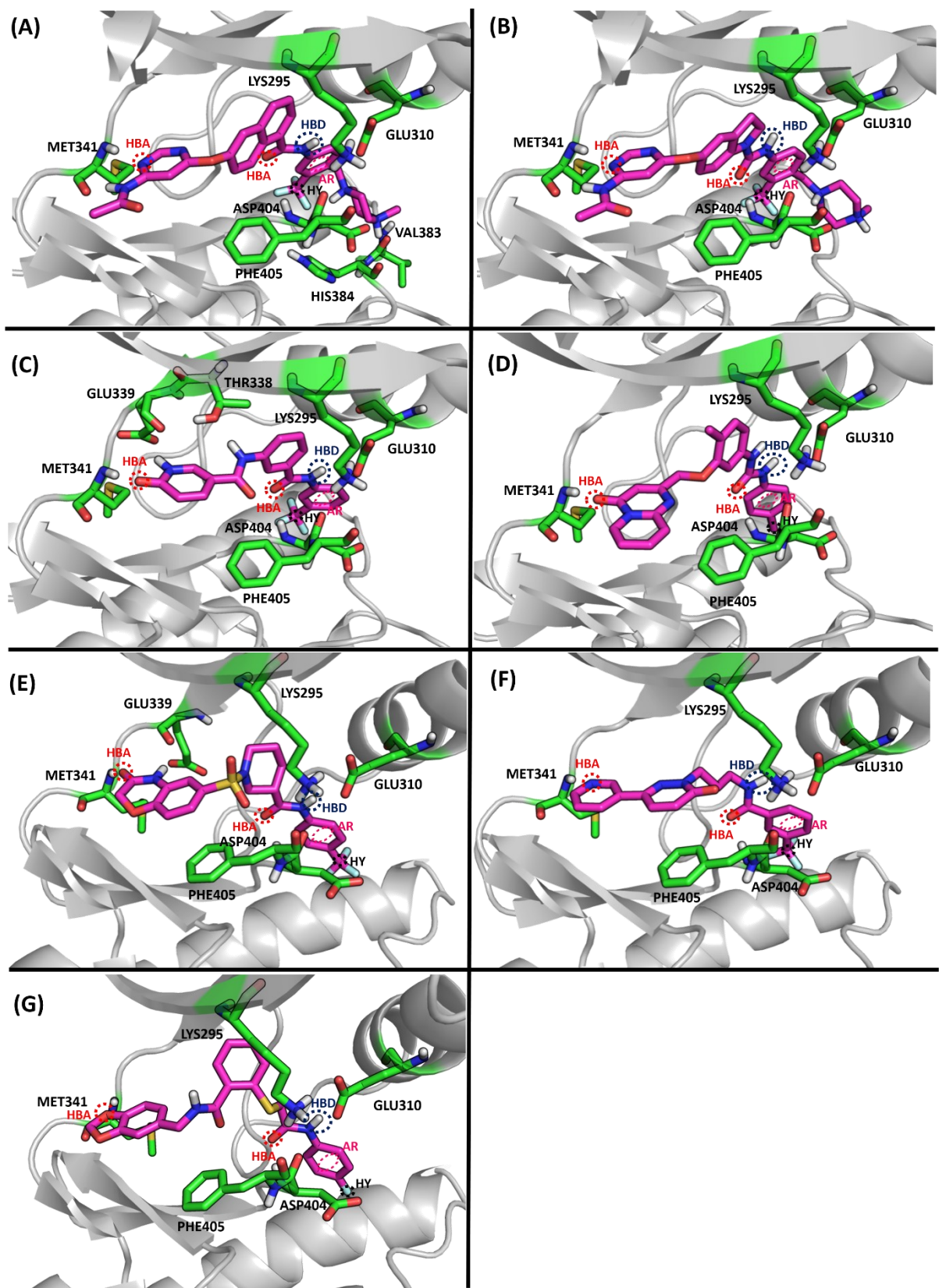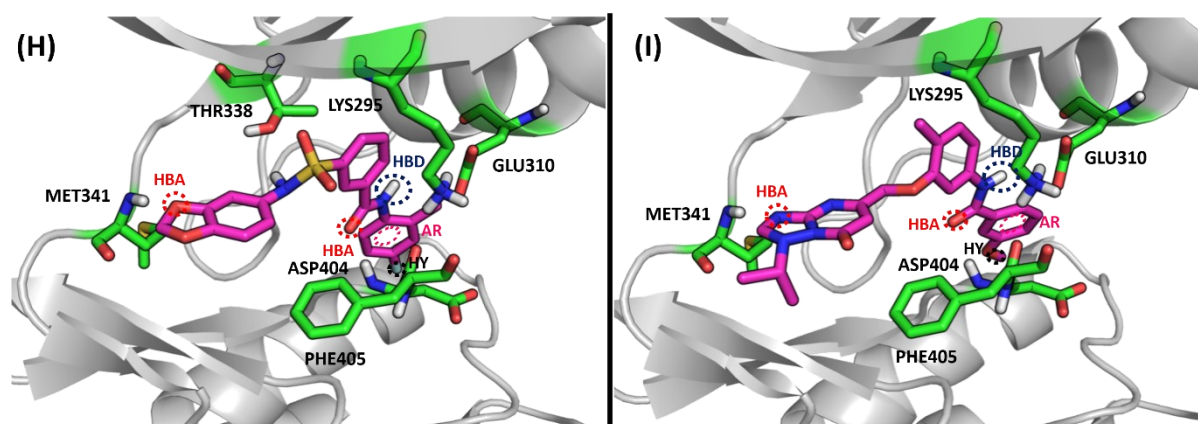
Figure 40. *Binding mode of compound **1** (A), compound **2** (B), compound **3** (C), compound **4** (D), compound **5** (E), compound **6** (F), compound **7** (G), compound **8** (H), compound **9** (I) within the c-Src ATP binding site in DFG-out conformation. The five chemical features defined in the pharmacophore model are shown as circles. In detail, the hydrogen bond donor group (HBD) able to interact with Glu310 is in blue, the hydrogen bond acceptor groups (HBA) able to interact respectively with Met341 and Asp404 are in red, the aryl group (AR) sited in the DFG pocket is in pink, and the hydrophobic substituent (HY) of the aryl group that further occupies the DFG pocket is in black. Ligands are reported in magenta sticks while interacting residues and those representative of the conformational state of the protein are reported in green.*

4.5 Conclusions and future perspectives

In this work, a SBSV procedure consisting of several computational filters was applied to a database of more than 7 million compounds to identify new Type-II c-Src inhibitors. The SBVS procedure led to the selection of nine compounds and prioritized several known c-Src inhibitors, confirming its reliability in the identification new inhibitors of this kinase. In the next future the selected compounds will be submitted to the biological tests to evaluate the c-Src inhibitory activity, the ability to stabilize the DFG-out inactive conformation, and the selectivity profile.

5 IDENTIFICATION OF A NEW FAMILY OF PYRAZOLO[3,4-*d*]PYRIMIDINE DERIVATIVES AS MULTITARGET FYN-BLK-LYN INHIBITORS ACTIVE ON B- AND T-LYMPHOMA CELL LINES

5.1 State of the art and aim of the project

As previously reported, Fyn, Lyn, and Blk represent interesting pharmacological targets for the treatment of some forms of lymphomas derived from B- and T-cells.

The research group of Professor Maurizio Botta works from long time to the design of SFK inhibitors endowed with pyrazolo[3,4-*d*]pyrimidine scaffold.[39,89–91] In this context, compound **Si308** (FIGURE 43A) emerged for its significant inhibitory action towards Fyn and interesting activity toward Lyn and Blk in a preliminary assay. Treatment with 10 μM **Si308** decreases the kinase activity of Lyn by 85% and Blk by 81%.[90]

Starting from these observations, the first aim of the project was to further investigate on the action of **Si308** towards Lyn and Blk. Once confirmed the SFKs multitarget inhibitory ability of **Si308**, a small library of its derivatives was synthesized to discover new pyrazolo[3,4-*d*]pyrimidines with anti-tumor activity on lymphoma cell lines.

5.2 Workflow

The ability of **Si308** to act also as a Lyn and Blk inhibitor was further investigated through the following steps:

- Molecular docking of the compound in the ATP binding site of the kinases.
- MD simulations of the docking complexes and evaluation of their theoretical ΔG_{bin}.
- Confirmation of the molecular modelling results by enzymatic assays.

Once achieved the first objective of the work, a small library of **Si308** derivatives was synthesized and tested against human B- and T-lymphoma cell lines.

5.3 Materials and methods

5.3.1 Protein preparation

The X-ray structures of Fyn and Lyn (2DQ7[92] and 5XY1,[93]respectively) were prepared by using the Protein Preparation Wizard workflow. In particular, crystallographic water molecules and ligands were removed, while hydrogen atoms, side chains and missing loops were added (using Prime), and bond orders were assigned. Finally, the network of hydrogen bonds was optimized at pH 7.4 and the systems were subjected to energy minimization with the OPLS3 force field (convergence was reached when the RMSD of the heavy atoms was below 0.30 Å).

5.3.2 Homology modelling

The amino acid sequence of human Blk kinase domain was acquired from UNIPROT database[94] (entry P51451) and used to perform a sequence similarity search with BLAST[95] in the Protein Data Bank. The X-ray structure of human Hck kinase domain (2HK5)[96] was chosen as template for the creation of the homology model. The loop refinement task was applied only for the loops not derived from the template, employing the serial loop sampling method and selecting default level of accuracy (recommended for loops up to 5 residues). Moreover, the side chain conformation of residues not derived from the template was optimized using the default sampling algorithm. Finally, the system was subjected to optimization of the of hydrogen bond network and to energy minimization as previously described for Fyn and Lyn structures.

The reliability of the model was evaluated by analysis of the Ramachandran plot (FIGURE 41).
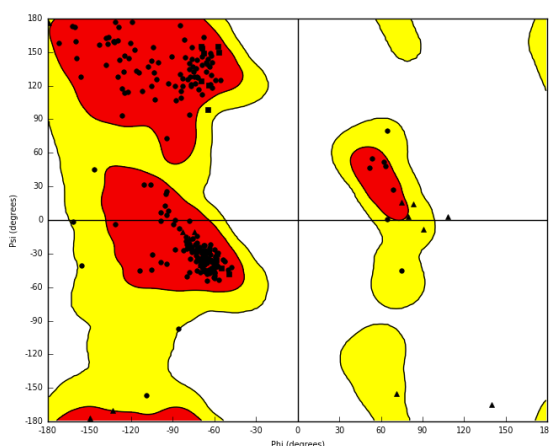
Figure 41. *Ramachandran plot of the homology structure of the kinase domain of human Blk. The different coloured areas indicate the most favoured region (red), the allowed region (yellow), and disallowed (white) regions.*[28]

Moreover, the energy of the model was evaluated by using ProSA-web tool[97,98], that calculate the z-score (a measure of the deviation of the total energy of the structure under study in comparison to the energy associated with a random conformation). FIGURE 42 showed the z-score of the produced model and the scores of all the experimentally determined proteins available in the Protein Data Bank. The energy value of the model is within the range of the scores calculated for the structures with a similar number of amino acid residues.
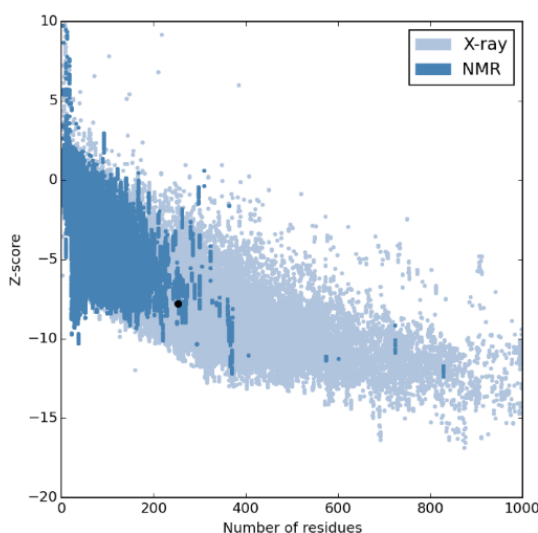
Figure 42. Z-score of homology model (black dot) and the proteins stored within the Protein Data Bank.[28]

5.3.3 Docking simulations

Docking simulations were performed by using Glide software (SP procedure). The receptor grids were generated from the previously prepared protein structures, centred on the corresponding co-crystallized ligands, and treating the protein as rigid. The ligands to be docked were respectively sketched and energy minimized through Maestro and MacroModel and finally docked within the receptor grids with flexible ligand sampling.

5.3.4 MD simulations

The simulation protocol applied to each complex was the same described in section 3.4.3.1, with the only exception that 50 ns of conventional MD simulation at constant temperature (300 K) and constant pressure (1 atm) were performed once obtained density equilibration.

5.3.5 Cluster analysis

Cluster analysis of the trajectories was performed as described in section 3.4.4.

The cluster count used to cluster the MD production phase and the complex of related trajectory are reported in TABLE 18.

COMPLEX	CLUSTER COUNT
Fyn-**Si308**	5
Lyn-**Si308**	3
Blk-**Si308**	2

Table 18. Number of clusters for each complex.

5.3.6 MM-GBSA analysis

The same protocol used in section 3.4.5 was applied to evaluate ΔG_{bin} of the three complexes and its pairwise decomposition.

5.4 Results and discussion

5.4.1 Molecular modelling and enzymatic evaluation of Si308

To evaluate the putative binding mode of **Si308** on the SFKs under investigation, docking studies were performed by using Glide software. The reliability of the applied computational docking protocol was assessed by docking compound **1a** (FIGURE 43A), a precursor of **Si308**, in the ATP binding site of Fyn (2DQ7). The results showed a binding mode analogous to that already published (FIGURE 43B).[90]
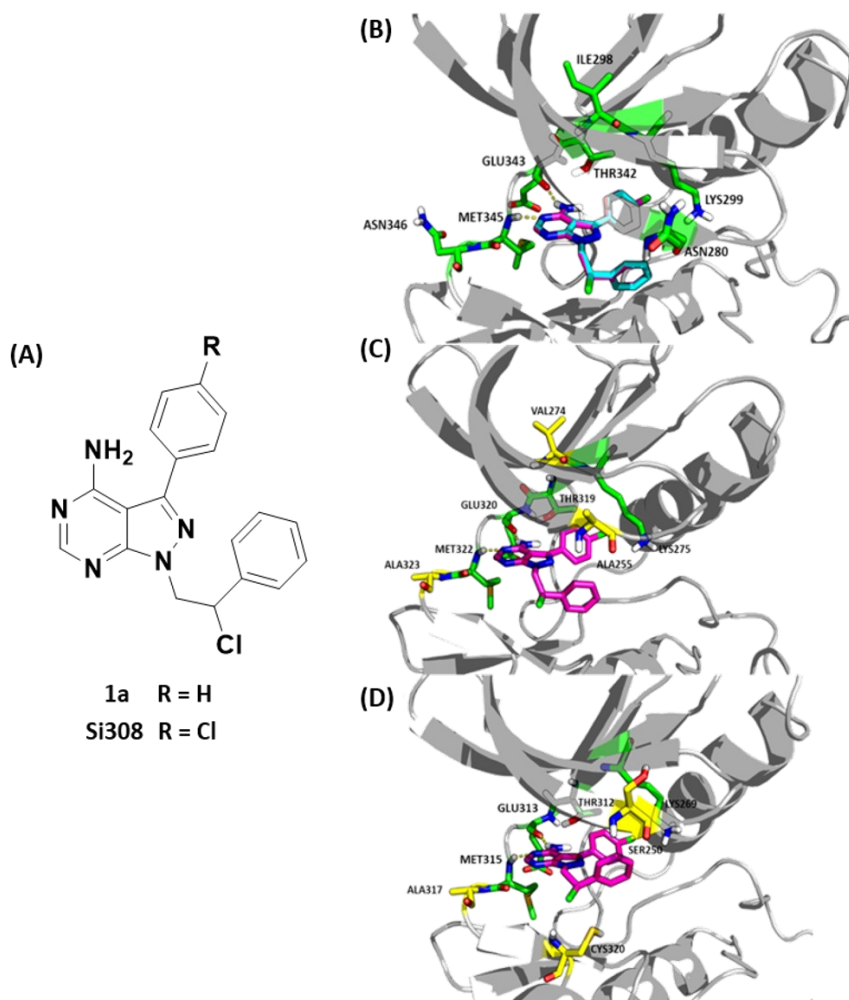
Figure 43. *Chemical structures of compounds **1a** and **Si308** (A). Predicted binding mode of compound **1a** and **Si308** in the ATP binding pocket of Fyn (B), Lyn (C), and Blk (D). Ligands are shown as cyan and magenta sticks, respectively. The yellow amino acids in (C) and (D) represent the residues of the ATP binding site of Lyn and Blk that differ from those found in Fyn.*[28]

It is important to note that the (R)-enantiomer of both compound **1a** and the pyrazolo[3,4-d]pyrimidines were used for docking simulations. The rationale behind this choice is based on a previous complex between c-Src and a structurally-related compound that was crystallized in its (R)-enantiomeric form.[39]

Compound **Si308** was then docked in the ATP binding site of Fyn, Lyn and Blk, which are characterized by high percentage of identity within the kinase domain. Remarkably, the residues of the ATP binding site are highly conserved (FIGURE 44), thus suggesting that the binding mode of **Si308** in the three kinases could be comparable.
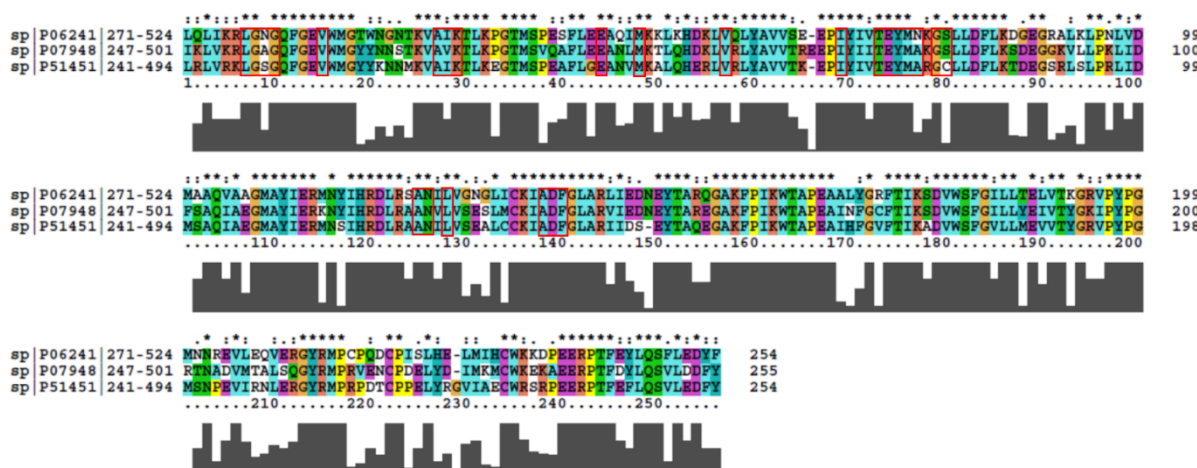
Figure 44. *Sequence alignment of the catalytic domain of Fyn, Blk, and Lyn (UniProtKB codes: P06241, P51451, and P07948, respectively). Bars below the sequence alignment correspond to the degree of amino acid conservation among the sequences (full bar: residues identity; empty bar: completely different residues). The red boxes show the residues of the ATP binding site.*[28]

For docking simulations, the X-ray structure 5XY1 of the human Lyn kinase was selected due to its highest resolution among available structures. On the contrary, no 3D structure of the kinase domain of Blk was available in Protein Data Bank at the time of the study, prompting us to build it by homology modelling techniques. For this purpose, a BLAST search in the Protein Data Bank supplied the human tyrosine kinase Hck as the sequence with the highest identity (75%) with respect to the query. The X-ray structure of the Hck kinase domain in active conformation (2HK5) was then selected as template for the generation of the Blk homology model.

Results of docking calculations showed the same binding mode of **Si308** in Fyn, Lyn, and Blk (FIGURE 43), as expected based on the high conservation of the ATP binding site residues. In all the complexes, **Si308** makes hydrogen bonds with the glutamate and methionine residues of the hinge region while the C3 phenyl group fits in the hydrophobic region of the ATP binding site.

To assess the stability of the complexes and to obtain further information about ligand-protein interactions, MD simulations were performed by using Amber16. Before simulations, docking complexes were solvated in a box of explicit water molecules containing counterions to neutralize the charge. Subsequently, the systems were relaxed through energy minimization, heated to 300 K, and density equilibrated to finally produce a MD trajectory of 50 ns for each complex. FIGURE 45A shows the evolution of the RMSD calculated on all the heavy atoms of the complexes during MD simulation time with respect to the last frame of minimization. In all the complexes, an initial increase of the RMSD in the equilibration steps with a subsequent stabilization of these values around 2 Å during the production times was noticed, suggesting a great stability of the complexes. The evolution of ligand RMSD during the MD simulation time with respect to the docking pose in each complex is reported in FIGURE 45B. In all the complexes, these values are stable around 2-3 Å, which is representative of the binding stability of **Si308** within the ATP binding site of these kinases.
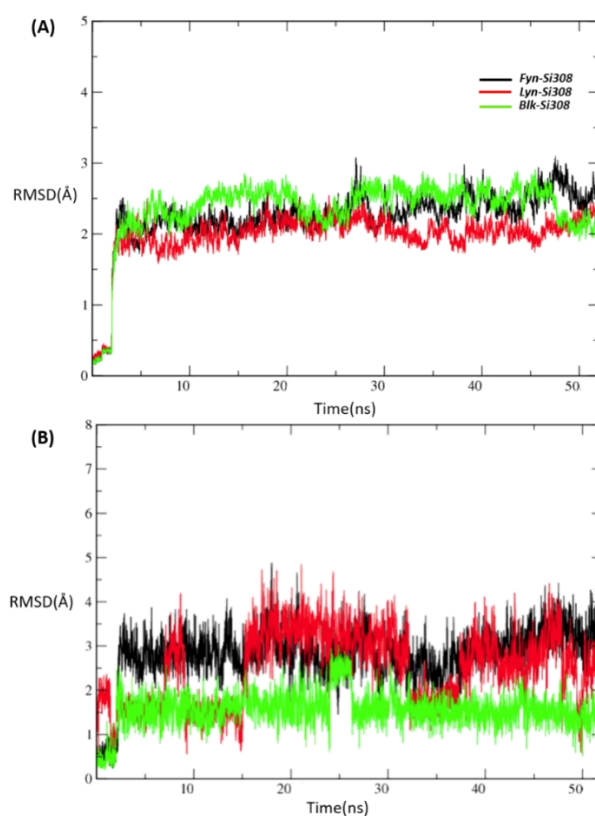
Figure 45. *RMSD of the heavy atoms (A) and of the ligand atoms (B) in the three complexes.*[28]

The production phases of the MD trajectories were then submitted to cluster analysis to evaluate the most persistent interactions made by the compound within the ATP binding sites. In this clustering process, the RMSD of the atoms of the amino acid residues around 6 Å from the

ligand was used to evaluate the similarity between the frames. The percentage of the total population found in the clusters of the analysed trajectories is reported in TABLE 19.

COMPLEX	C0	C1	C2	C3	C4
Fyn-**Si308**	70.7 %	29 %	0.3 %		
Lyn-**Si308**	90.7 %	9.3 %			
Blk-**Si308**	97.5 %	2.5 %			

Table 19. *C0, C1, C2, C3, and C4 represent the 5 most populated clusters of the trajectory related to a given complex.*

Subsequently, by using 200 frames around the representative frame of the most populated cluster, the pairwise decomposition of ΔG_{bin} of the complexes, implemented in the MM-GBSA approach, was carried out to obtain an energetic estimation of ligand-protein interactions (TABLE 20). The energetically most relevant interactions (lower than -2.5 kcal/mol) are made by **Si308** with Glu343, Met345, Tyr344 of the hinge region, the catalytic Lys299, and with two hydrophobic residues of the binding site (Val285 and Leu397) (residues numbering is related to Fyn sequence, the corresponding residues in Lyn are Glu320, Met322, Tyr321, Lys275, Val261 and Leu374, while in Blk are Glu313, Met315, Tyr314, Lys269, Val255 and Leu367).

Residue in Fyn(Lyn,Blk)	ΔG_{bin} (kcal/mol)		
	Si308/Fyn	**Si308/Lyn**	**Si308/Blk**
Val285(*261,**255)	-2.903 ± 0.509	-2.861 ± 0.559	-3.485 ± 0.565
Ala297(*273,**267)	-2.245 ± 0.428	-1.810 ± 0.299	-1.836 ± 0.342
Lys299(*275,**269)	-3.375 ± 0.563	-3.116 ± 0.500	-3.869 ± 0.564
Glu314(*290,**284)	-0.304 ± 0.155	-0.054 ± 0.021	-0.599 ± 0.285
Met318(*294,**288)	-0.583 ± 0.136	-0.634 ± 0.195	-0.533 ± 0.130
Val327(*303,**297)	-1.201 ± 0.285	-1.276 ± 0.316	-0.942 ± 0.208
Thr342(*319,**312)	-2.379 ± 0.648	-2.221 ± 0.526	-2.254 ± 0.577
Glu343(*320,**313)	-3.200 ± 0.520	-3.304 ± 0.443	-3.050 ± 0.469
Tyr344(*321,**314)	-3.359 ± 0.492	-3.175 ± 0.495	-2.887 ± 0.436
Met345(*322,**315)	-3.694 ± 0.471	-3.736 ± 0.523	-3.900 ± 0.466
Leu397(*374,**367)	-3.475 ± 0.389	-4.919 ± 0.686	-3.582 ± 0.343
Ala407(*384,**377)	-0.586 ± 0.136	-2.020 ± 0.446	-0.677 ± 0.152
Asp408(*385,**378)	-0.587 ± 0.324	-0.859 ± 0.344	-1.196 ± 0.374
Phe409(*386,**379)	-0.082 ± 0.028	-0.088 ± 0.056	-0.116 ± 0.036

Table 20. *Energetic values of ligand-protein interactions. *numbering of the same amino acid in Lyn **numbering of the same amino acid in Blk.*

The MM-GBSA approach was also applied to calculate the ΔG_{bin} of the complexes under study, by using the same 200 frames employed before (TABLE 21). The results suggest that **Si308** might bind Fyn, Lyn, and Blk with analogous affinity, confirming it as potential SFKs multitarget inhibitor.

COMPLEX	ΔG_{bin} (kcal/mol)
Fyn-**Si308**	-37.55 ± 2.59
Lyn-**Si308**	-37.84 ± 2.70
Blk-**Si308**	-37.28 ± 2.48

Table 21. *Free energy of binding of the three complexes.*

Finally, enzymatic assays were performed to evaluate the inhibitory activity of **Si308** against Fyn, Lyn and Blk. The compound is characterized by comparable values of inhibition against the three kinases (TABLE 22), confirming the computational results.

	Fyn K_i (µM)	Lyn K_i (µM)	Blk K_i (µM)
Si308	0.16 ± 0.01	0.19 ± 0.04	0.19 ± 0.05

Table 22*. Inhibitory activity of **Si308** toward Fyn, Lyn and Blk.*

5.4.2 Design of Si308 derivatives and biological evaluation

Once confirmed the activity of **Si308** against Fyn, Lyn and Blk, a library of twelve derivatives (**2a-l**) was synthesized to discover new compounds with anti-tumor activity in lymphoma cell lines. Some derivatives are characterized by different substituents and substitution pattern on the C3 phenyl ring, which is predicted to occupy part of the ATP binding pocket. Other derivatives, in addition to have para-substituents at the C3 phenyl ring, are characterized by replacement of the chloride group of the solvent-exposed molecular portion with H, Me, and OH substituents (FIGURE 46).
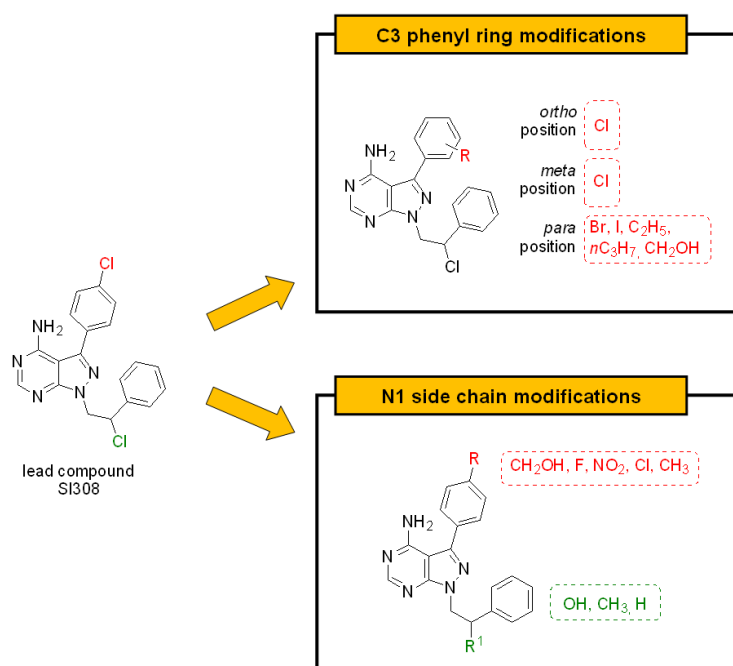
Figure 46. *General plan for the design of **Si308** derivatives. C3 phenyl ring point modifications (upper panel) and N1 side chain modifications combined with C3 phenyl ring modifications (lower panel) are shown.*[28]

The synthesized library of **Si308** derivatives was tested on GCB-DLBCL cell lines (DOHH2, FARAGE, SU-DHL-8, and SU-DHL-6), CTCL cell lines (HUT-78, H9, and HH), and a PTCL cell line (FE-PD). Compound **1b** ($K_i = 95$ nM against Fyn)[90] was added to the library. Cellular assays evidenced compound **2h** as the most promising of the series with $IC_{50} \leq 10$ μM in all cell lines (TABLE 23).
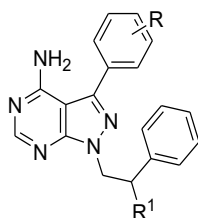
			IC$_{50}$ (μM)							
			B-cell lymphoma cell line				**T-cell lymphoma cell line**			
Cpd	R	R^1	DOHH2	FARAGE	SUDHL8	SUDHL6	HUT-78	FEPD	H9	HH
Si308	4-Cl	Cl	11.41	14.94	8.87	21.52	> 100	15.76	13.58	13.9
1b	4-CH$_3$	Cl	11.77	19.25	12.14	23.43	11.12	> 100	> 100	14.82
2a	4-Br	Cl	10.99	13.61	11.85	45.64	21.05	21.7	21.7	12.44
2b	4-I	Cl	3.68	12.64	11.46	13.17	12.03	20.46	11.61	5.24
2c	4-C$_2$H$_5$	Cl	21.18	22.74	17.47	29.68	58.00	~21.53	24.17	21.08
2d	4-nC$_3$H$_7$	Cl	4.38	29.36	6.72	31.11	9.90	29.15	4.74	48.53
2e	2-Cl	Cl	26.58	51.45	18.39	> 100	> 100	38.86	38.86	> 100
2f	3-Cl	Cl	27.35	30.9	13.69	> 100	> 100	24.89	~22.07	43.12
2g	4-CH$_2$OH	OH	12.09	26.96	27.18	34.20	> 100	26.49	~22.26	17.31
2h	4-CH$_2$OH	Cl	2.44	7.02	6.14	9.29	4.50	10.84	3.42	2.40
2i	4-F	CH$_3$	21.26	22.8	12.08	~22.92	> 100	> 100	18.76	11.81
2j	4-NO$_2$	CH$_3$	> 100	> 100	> 100	> 100	> 100	> 100	> 100	> 100
2k	4-Cl	H	11.81	18.6	35.96	> 100	> 100	31.1	> 100	18.17
2l	4-CH$_3$	H	9.74	15.98	37.03	> 100	> 100	~21.54	> 100	17.26

Table 23. *Anti-tumor activity of **Si308**, **1b**, and **2a-l** on lymphoma cell lines.*

5.4.3 Molecular modelling and enzymatic evaluation of 2h

The putative binding mode of **2h** in the ATP binding site of the three target kinases was evaluated by docking simulations. As result, **2h** showed an analogous binding mode within the three kinases and similarly to **Si308** makes hydrogen bonds with the glutamate and methionine residues of the hinge region. Moreover, compound **2h** makes hydrogen bonds with the catalytic glutamate of the αC helix and aspartate of the DFG motif (FIGURE 47).

Figure 47. *Chemical structure of* ***2h*** *(A). Predicted binding mode of* ***2h*** *in the ATP binding pocket of Fyn (B), Lyn (C), and Blk (D). Ligand* ***2h*** *is shown as orange sticks. The amino acids in yellow sticks in (C) and (D) represent the residues of the ATP binding site of Lyn and Blk that differ from those of Fyn.*[28]

Enzymatic assays were performed to confirm the activity of **2h** on Fyn, Lyn, and Blk, and resulted in sub-micromolar K_i values (TABLE 24).

	Fyn K_i (μM)	Lyn K_i (μM)	Blk K_i (μM)
2h	0.12 ± 0.03	0.19 ± 0.02	0.17 ± 0.08

Table 24. *Inhibitory activities of **2h** toward Fyn, Lyn and Blk.*

Remarkably, the enzymatic results obtained for **2h** are assimilable to those showed by the parent compound **Si308** even if the latter is characterized by a lower activity on cell lines.

5.4.4 *In vitro* ADME studies

To clarify the different cellular activity of **Si308** and **2h**, *in vitro* ADME assays were carried out. In detail, aqueous solubility, metabolic stability, passive membrane permeability, and membrane retention were evaluated (TABLE 25). Water solubility of **2h** is about an order of magnitude higher than **Si308**, which is explained by the presence of an alcohol polar group in the molecule. Both compounds are characterized by a comparable metabolic stability. Compound **2h** presents a slightly higher value of passive membrane permeability than **Si308**. Remarkably, **Si308** is characterized by a membrane retention percentage ~5 folds higher than **2h**, which could clarify the reasons of the higher cellular activity of **2h** than **Si308**.

	Aqueous solubility (μg·mL^{-1})	Metabolic stability (%)	Gastrointestinal P_{app} (10^{-6} cm·s^{-1}) (Membrane retention) %
Si308	0.01	85.2	9.95 (70.4)
2h	0.09	85.0	10.80 (15.2)

Table 25. *In vitro ADME properties of **Si308** and **2h**.*

5.5 Conclusions

The three members of SFKs, Fyn, Lyn and Blk, represent interesting molecular targets for the treatment of B- and T-cell lymphomas.[23,25–27,99] Herein, molecular modelling studies coupled with enzymatic assays were performed to investigate the ability of Fyn inhibitor **Si308** to act as a multitarget SFKs inhibitor. As result, the compound acts as Lyn and Blk inhibitor with a comparable activity to that showed toward Fyn. Subsequently, a small library of **Si308** derivatives was synthesized and tested against B- and T-lymphoma cell lines. Derivative **2h**, with IC_{50} values below 10 μM on all the tumor cell lines and K_i values in the sub-micromolar range against the three targets kinases, resulted as the most promising compound.

Interestingly, the parent compound **Si308**, even if has K_i values against Fyn, Lyn and Blk, comparable to those of compound **2h**, resulted less active on cell lines. To this end, *in vitro* ADME studies highlighted that this difference in biological activity could be due to higher percentage of **Si308** that is locked in the cell membrane than **2h**.

Finally, this work allowed to identify a potent multitarget Fyn-Lyn-Blk inhibitor characterized by an interesting antitumor activity on several human lymphoma cell lines and promising pharmacokinetic properties. These encouraging results open the way to develop the class of pyrazolo[3,4-*d*]pyrimidines as agents for the treatment of lymphomas.

All this work was published in 2019 in the European Journal of Medicinal Chemistry.[28]

6 BIBLIOGRAPHY

(1)   Duong-Ly, K. C.; Peterson, J. R. The Human Kinome and Kinase Inhibition as A Therapeutic Strategy. *Curr. Protoc. Pharmacol.* **2013**.

(2)   Roskoski, R. Src Protein–tyrosine Kinase Structure and Regulation. *Biochem. Biophys. Res. Commun.* **2004**, *324* (4), 1155–1164. https://doi.org/10.1016/j.bbrc.2004.09.171.

(3)   Tong, M.; Seeliger, M. A. Targeting Conformational Plasticity of Protein Kinases. *ACS Chem. Biol.* **2015**, *10* (1), 190–200. https://doi.org/10.1021/cb500870a.

(4)   Rossari, F.; Minutolo, F.; Orciuolo, E. Past, Present, and Future of Bcr-Abl Inhibitors: From Chemical Development to Clinical Efficacy. *J. Hematol. Oncol.* **2018**, *11* (1), 84. https://doi.org/10.1186/s13045-018-0624-2.

(5)   Thomas, S. M.; Brugge, J. S. Cellular Functions Regulated by Src Family Kinases. *Annu. Rev. Cell Dev. Biol.* **1997**, *13* (1), 513–609. https://doi.org/10.1146/annurev.cellbio.13.1.513.

(6)   Superti-Furga, G.; Courtneidge, S. A. Structure-Function Relationships in Src Family and Related Protein Tyrosine Kinases. *BioEssays* **1995**, *17* (4), 321–330. https://doi.org/10.1002/bies.950170408.

(7)   Aleshin, A.; Finn, R. S. SRC: A Century of Science Brought to the Clinic. *Neoplasia* **2010**, *12* (8), 599–607. https://doi.org/10.1593/neo.10328.

(8)   Parsons, S. J.; Parsons, J. T. Src Family Kinases, Key Regulators of Signal Transduction. *Oncogene* **2004**, *23* (48), 7906–7909. https://doi.org/10.1038/sj.onc.1208160.

(9)   Boggon, T. J.; Eck, M. J. Structure and Regulation of Src Family Kinases. *Oncogene* **2004**, *23* (48), 7918–7927. https://doi.org/10.1038/sj.onc.1208081.

(10)   Xu, W.; Doshi, A.; Lei, M.; Eck, M. J.; Harrison, S. C. Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Mol. Cell* **1999**, *3* (5), 629–638. https://doi.org/10.1016/S1097-2765(00)80356-1.

(11)   Roskoski, R. Src Protein-Tyrosine Kinase Structure, Mechanism, and Small Molecule Inhibitors. *Pharmacol. Res.* **2015**, *94*, 9–25. https://doi.org/10.1016/j.phrs.2015.01.003.

(12)   Yeatman, T. J. A Renaissance for SRC. *Nat. Rev. Cancer* **2004**, *4* (6), 470–480. https://doi.org/10.1038/nrc1366.

(13)   Irby, R. B.; Yeatman, T. J. Role of Src Expression and Activation in Human Cancer. *Oncogene* **2000**, *19* (49), 5636–5642. https://doi.org/10.1038/sj.onc.1203912.

(14)   Wheeler, D. L.; Iida, M.; Dunn, E. F. The Role of Src in Solid Tumors. *Oncologist* **2009**, *14* (7), 667–678. https://doi.org/10.1634/theoncologist.2009-0009.

(15)   Lynch, R. C.; Gratzinger, D.; Advani, R. H. Clinical Impact of the 2016 Update to the

WHO Lymphoma Classification. *Curr. Treat. Options Oncol.* **2017**, *18* (11), 45. https://doi.org/10.1007/s11864-017-0483-z.

(16) Swerdlow, S. H.; Campo, E.; Pileri, S. A.; Harris, N. L.; Stein, H.; Siebert, R.; Advani, R.; Ghielmini, M.; Salles, G. A.; Zelenetz, A. D.; et al. The 2016 Revision of the World Health Organization Classification of Lymphoid Neoplasms. *Blood* **2016**, *127* (20), 2375–2390. https://doi.org/10.1182/blood-2016-01-643569.

(17) https://www.cancer.org/cancer/lymphoma.html.

(18) Ekström-Smedby, K. Epidemiology and Etiology of Non-Hodgkin Lymphoma – a Review. *Acta Oncol.* **2006**, *45* (3), 258–271. https://doi.org/10.1080/02841860500531682.

(19) Chiu, B. C.-H.; Weisenburger, D. D. An Update of the Epidemiology of Non-Hodgkin's Lymphoma. *Clin. Lymphoma* **2003**, *4* (3), 161–168. https://doi.org/10.3816/CLM.2003.n.025.

(20) Matasar, M. J.; Zelenetz, A. D. Overview of Lymphoma Diagnosis and Management. *Radiol. Clin. North Am.* **2008**, *46* (2), 175–198. https://doi.org/10.1016/j.rcl.2008.03.005.

(21) Pasqualucci, L.; Dalla-Favera, R. Genetics of Diffuse Large B-Cell Lymphoma. *Blood* **2018**, *131* (21), 2307–2319. https://doi.org/10.1182/blood-2017-11-764332.

(22) Van Arnam, J. S.; Lim, M. S.; Elenitoba-Johnson, K. S. J. Novel Insights into the Pathogenesis of T-Cell Lymphomas. *Blood* **2018**, *131* (21), 2320–2330. https://doi.org/10.1182/blood-2017-11-764357.

(23) Battistello, E.; Katanayeva, N.; Dheilly, E.; Tavernari, D.; Donaldson, M. C.; Bonsignore, L.; Thome, M.; Christie, A. L.; Murakami, M. A.; Michielin, O.; et al. Pan-SRC Kinase Inhibition Blocks B-Cell Receptor Oncogenic Signaling in Non-Hodgkin Lymphoma. *Blood* **2018**, *131* (21), 2345–2356. https://doi.org/10.1182/blood-2017-10-809210.

(24) Scuoppo, C.; Jiguang, W.; Persaud, M.; Pasqualucci, L.; Rabadan, R.; Grandori, C.; et al. Repurposing Dasatinib for Ibrutinib-Resistant Diffuse Large B-Cell Lymphoma. *Blood* **2017**, *130*, 3843. https://doi.org/10.1182/blood.V130.Suppl_1.3843.3843.

(25) Palomero, T.; Couronné, L.; Khiabanian, H.; Kim, M.-Y.; Ambesi-Impiombato, A.; Perez-Garcia, A.; Carpenter, Z.; Abate, F.; Allegretta, M.; Haydu, J. E.; et al. Recurrent Mutations in Epigenetic Regulators, RHOA and FYN Kinase in Peripheral T Cell Lymphomas. *Nat. Genet.* **2014**, *46* (2), 166–170. https://doi.org/10.1038/ng.2873.

(26) Krejsgaard, T.; Vetter-Kauczok, C. S.; Woetmann, A.; Kneitz, H.; Eriksen, K. W.; Lovato, P.; Zhang, Q.; Wasik, M. A.; Geisler, C.; Ralfkiaer, E.; et al. Ectopic Expression

of B-Lymphoid Kinase in Cutaneous T-Cell Lymphoma. *Blood* **2009**, *113* (23), 5896–5904. https://doi.org/10.1182/blood-2008-09-181024.

(27) Petersen, D. L.; Krejsgaard, T.; Berthelsen, J.; Fredholm, S.; Willerslev-Olsen, A.; Sibbesen, N. A.; Bonefeld, C. M.; Andersen, M. H.; Francavilla, C.; Olsen, J. V; et al. B-lymphoid Tyrosine Kinase (Blk) Is an Oncogene and a Potential Target for Therapy with Dasatinib in Cutaneous T-Cell Lymphoma (CTCL). *Leukemia* **2014**, *28* (10), 2109–2112. https://doi.org/10.1038/leu.2014.192.

(28) Fallacara, A. L.; Passannanti, R.; Mori, M.; Iovenitti, G.; Musumeci, F.; Greco, C.; Crespan, E.; Kissova, M.; Maga, G.; Tarantelli, C.; et al. Identification of a New Family of Pyrazolo[3,4-d]Pyrimidine Derivatives as Multitarget Fyn-Blk-Lyn Inhibitors Active on B- and T-Lymphoma Cell Lines. *Eur. J. Med. Chem.* **2019**, *181*, 111545. https://doi.org/10.1016/j.ejmech.2019.07.048.

(29) Roskoski, R. Properties of FDA-Approved Small Molecule Protein Kinase Inhibitors. *Pharmacol. Res.* **2019**, *144*, 19–50. https://doi.org/10.1016/j.phrs.2019.03.006.

(30) Huse, M.; Kuriyan, J. The Conformational Plasticity of Protein Kinases. *Cell* **2002**, *109* (3), 275–282. https://doi.org/10.1016/S0092-8674(02)00741-9.

(31) Ung, P. M.-U.; Rahman, R.; Schlessinger, A. Redefining the Protein Kinase Conformational Space with Machine Learning. *Cell Chem. Biol.* **2018**, *25* (7), 916–924.e2. https://doi.org/10.1016/j.chembiol.2018.05.002.

(32) Kornev, A. P.; Taylor, S. S.; Ten Eyck, L. F. A Helix Scaffold for the Assembly of Active Protein Kinases. *Proc. Natl. Acad. Sci.* **2008**, *105* (38), 14377–14382. https://doi.org/10.1073/pnas.0807988105.

(33) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat. Commun.* **2014**, *5* (1), 3397. https://doi.org/10.1038/ncomms4397.

(34) Simard, J. R.; Klüter, S.; Grütter, C.; Getlik, M.; Rabiller, M.; Rode, H. B.; Rauh, D. A New Screening Assay for Allosteric Inhibitors of cSrc. *Nat. Chem. Biol.* **2009**, *5* (6), 394–396. https://doi.org/10.1038/nchembio.162.

(35) Fang, Z.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8* (1), 58–70. https://doi.org/10.1021/cb300663j.

(36) Casini, N.; Forte, I. M.; Mastrogiovanni, G.; Pentimalli, F.; Angelucci, A.; Festuccia, C.; Tomei, V.; Ceccherini, E.; Di Marzo, D.; Schenone, S.; et al. SRC Family Kinase (SFK) Inhibition Reduces Rhabdomyosarcoma Cell Growth in Vitro and in Vivo and Triggers p38 MAP Kinase-Mediated Differentiation. *Oncotarget* **2015**, *6* (14), 12421–12435.

https://doi.org/10.18632/oncotarget.3043.

(37) Ceccherini, E.; Indovina, P.; Zamperini, C.; Dreassi, E.; Casini, N.; Cutaia, O.; Forte, I. M.; Pentimalli, F.; Esposito, L.; Polito, M. S.; et al. SRC Family Kinase Inhibition Through a New Pyrazolo[3,4-d]Pyrimidine Derivative as a Feasible Approach for Glioblastoma Treatment. *J. Cell. Biochem.* **2015**, *116* (5), 856–863. https://doi.org/10.1002/jcb.25042.

(38) Schenone, S.; Bruno, O.; Ranise, A.; Bondavalli, F.; Brullo, C.; Fossa, P.; Mosti, L.; Menozzi, G.; Carraro, F.; Naldini, A.; et al. New Pyrazolo[3,4-d]Pyrimidines Endowed with A431 Antiproliferative Activity and Inhibitory Properties of Src Phosphorylation. *Bioorg. Med. Chem. Lett.* **2004**, *14* (10), 2511–2517. https://doi.org/10.1016/j.bmcl.2004.03.013.

(39) Tintori, C.; Fallacara, A. L.; Radi, M.; Zamperini, C.; Dreassi, E.; Crespan, E.; Maga, G.; Schenone, S.; Musumeci, F.; Brullo, C.; et al. Combining X-Ray Crystallography and Molecular Modeling toward the Optimization of Pyrazolo[3,4-d]Pyrimidines as Potent c-Src Inhibitors Active in Vivo against Neuroblastoma. *J. Med. Chem.* **2015**, *58* (1), 347–361. https://doi.org/10.1021/jm5013159.

(40) Vignaroli, G.; Iovenitti, G.; Zamperini, C.; Coniglio, F.; Calandro, P.; Molinari, A.; Fallacara, A. L.; Sartucci, A.; Calgani, A.; Colecchia, D.; et al. Prodrugs of Pyrazolo[3,4-d]Pyrimidines: From Library Synthesis to Evaluation as Potential Anticancer Agents in an Orthotopic Glioblastoma Model. *J. Med. Chem.* **2017**, *60* (14), 6305–6320. https://doi.org/10.1021/acs.jmedchem.7b00637.

(41) Spreafico, A.; Schenone, S.; Serchi, T.; Orlandini, M.; Angelucci, A.; Magrini, D.; Bernardini, G.; Collodel, G.; Di Stefano, A.; Tintori, C.; et al. Antiproliferative and Proapoptotic Activities of New Pyrazolo[3,4-d]Pyrimidine Derivative Src Kinase Inhibitors in Human Osteosarcoma Cells. *FASEB J.* **2008**, *22* (5), 1560–1571. https://doi.org/10.1096/fj.07-9873com.

(42) Angelucci, A.; Schenone, S.; Gravina, G. L.; Muzi, P.; Festuccia, C.; Vicentini, C.; Botta, M.; Bologna, M. Pyrazolo[3,4-d]Pyrimidines c-Src Inhibitors Reduce Epidermal Growth Factor-Induced Migration in Prostate Cancer Cells. *Eur. J. Cancer* **2006**, *42* (16), 2838–2845. https://doi.org/10.1016/j.ejca.2006.06.024.

(43) Delle Monache, S.; Sanità, P.; Calgani, A.; Schenone, S.; Botta, L.; Angelucci, A. Src Inhibition Potentiates Antitumoral Effect of Paclitaxel by Blocking Tumor-Induced Angiogenesis. *Exp. Cell Res.* **2014**, *328* (1), 20–31. https://doi.org/10.1016/j.yexcr.2014.08.002.

(44) Molinari, A.; Fallacara, A. L.; Di Maria, S.; Zamperini, C.; Poggialini, F.; Musumeci,

F.; Schenone, S.; Angelucci, A.; Colapietro, A.; Crespan, E.; et al. Efficient Optimization of Pyrazolo[3,4-d]Pyrimidines Derivatives as c-Src Kinase Inhibitors in Neuroblastoma Treatment. *Bioorg. Med. Chem. Lett.* **2018**, *28* (21), 3454–3457. https://doi.org/10.1016/j.bmcl.2018.09.024.

(45) Indovina, P.; Giorgi, F.; Rizzo, V.; Khadang, B.; Schenone, S.; Di Marzo, D.; Forte, I. M.; Tomei, V.; Mattioli, E.; D'Urso, V.; et al. New Pyrazolo[3,4-d]Pyrimidine SRC Inhibitors Induce Apoptosis in Mesothelioma Cell Lines through P27 Nuclear Stabilization. *Oncogene* **2012**, *31* (7), 929–938. https://doi.org/10.1038/onc.2011.286.

(46) Rossi, A.; Schenone, S.; Angelucci, A.; Cozzi, M.; Caracciolo, V.; Pentimalli, F.; Puca, A.; Pucci, B.; La Montagna, R.; Bologna, M.; et al. New Pyrazolo-[3,4-d]-pyrimidine Derivative Src Kinase Inhibitors Lead to Cell Cycle Arrest and Tumor Growth Reduction of Human Medulloblastoma Cells. *FASEB J.* **2010**, *24* (8), 2881–2892. https://doi.org/10.1096/fj.09-148593.

(47) Morisi, R.; Celano, M.; Tosi, E.; Schenone, S.; Navarra, M.; Ferretti, E.; Costante, G.; Durante, C.; Botta, G.; D'Agostino, M.; et al. Growth Inhibition of Medullary Thyroid Carcinoma Cells by Pyrazolo-Pyrimidine Derivates. *J. Endocrinol. Invest.* **2007**, *30* (10), RC31-RC34. https://doi.org/10.1007/BF03349220.

(48) Getlik, M.; Grütter, C.; Simard, J. R.; Klüter, S.; Rabiller, M.; Rode, H. B.; Robubi, A.; Rauh, D. Hybrid Compound Design To Overcome the Gatekeeper T338M Mutation in cSrc #. *J. Med. Chem.* **2009**, *52* (13), 3915–3926. https://doi.org/10.1021/jm9002928.

(49) Seeliger, M. A.; Nagar, B.; Frank, F.; Cao, X.; Henderson, M. N.; Kuriyan, J. c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty. *Structure* **2007**, *15* (3), 299–311. https://doi.org/10.1016/j.str.2007.01.015.

(50) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(51) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-Chain Conformations. *J. Mol. Biol.* **2002**, *320* (3), 597–608. https://doi.org/10.1016/S0022-2836(02)00470-9.

(52) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A Hierarchical Approach to All-Atom Protein Loop Prediction. *Proteins Struct. Funct. Bioinforma.* **2004**, *55* (2), 351–367. https://doi.org/10.1002/prot.10613.

(53) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the

Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput. Aided. Mol. Des.* **2010**, *24* (6–7), 591–604. https://doi.org/10.1007/s10822-010-9349-1.

(54) Dar, A. C.; Lopez, M. S.; Shokat, K. M. Small Molecule Recognition of c-Src via the Imatinib-Binding Conformation. *Chem. Biol.* **2008**, *15* (10), 1015–1022. https://doi.org/10.1016/j.chembiol.2008.09.007.

(55) Krishnamurty, R.; Brigham, J. L.; Leonard, S. E.; Ranjitkar, P.; Larson, E. T.; Dale, E. J.; Merritt, E. A.; Maly, D. J. Active Site Profiling Reveals Coupling between Domains in SRC-Family Kinases. *Nat. Chem. Biol.* **2013**, *9* (1), 43–50. https://doi.org/10.1038/nchembio.1118.

(56) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

(57) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. https://doi.org/10.1021/jm030644s.

(58) Schrödinger Release 2017-2: MacroModel, Schrödinger, LLC, New York, NY, 2017.

(59) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking [1]. *J. Mol. Biol.* **1997**, *267* (3), 727–748. https://doi.org/10.1006/jmbi.1996.0897.

(60) Schrödinger Release 2017-2: LigPrep, Schrödinger, LLC, New York, NY, 2017.

(61) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.

(62) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.

(63) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(64) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(65) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7),

3084–3095. https://doi.org/10.1021/ct400341p.

(66) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314–3321. https://doi.org/10.1021/ct300418h.

(67) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51* (1), 69–82. https://doi.org/10.1021/ci100275a.

(68) http://ambermd.org/tutorials/advanced/tutorial3/.

(69) Seeliger, D.; de Groot, B. L. Ligand Docking and Binding Site Analysis with PyMOL and Autodock/Vina. *J. Comput. Aided. Mol. Des.* **2010**, *24* (5), 417–422. https://doi.org/10.1007/s10822-010-9352-6.

(70) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(71) Miao, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. *Annual Reports in Computational Chemistry* **2017**, *13,* 231–278. https://doi.org/10.1016/bs.arcc.2017.06.005.

(72) Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided. Mol. Des.* **2006**, *20* (10–11), 647–671. https://doi.org/10.1007/s10822-006-9087-6.

(73) http://dude.docking.org/generate.

(74) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55* (14), 6582–6594. https://doi.org/10.1021/jm300687e.

(75) Kwarcinski, F. E.; Brandvold, K. R.; Phadke, S.; Beleh, O. M.; Johnson, T. K.; Meagher, J. L.; Seeliger, M. A.; Stuckey, J. A.; Soellner, M. B. Conformation-Selective Analogues of Dasatinib Reveal Insight into Kinase Inhibitor Binding and Selectivity. *ACS Chem. Biol.* **2016**, *11* (5), 1296–1304. https://doi.org/10.1021/acschembio.5b01018.

(76) Seeliger, M. A.; Ranjitkar, P.; Kasap, C.; Shan, Y.; Shaw, D. E.; Shah, N. P.; Kuriyan, J.; Maly, D. J. Equally Potent Inhibition of c-Src and Abl by Compounds That Recognize Inactive Kinase Conformations. **2009**, *69* (6), 2384–2393. https://doi.org/10.1158/0008-5472.CAN-08-3953.

(77) Weisberg, E.; Choi, H. G.; Ray, A.; Barrett, R.; Zhang, J.; Sim, T.; Zhou, W.; Seeliger, M.; Cameron, M.; Azam, M.; et al. Discovery of a Small-Molecule Type II Inhibitor of Wild-Type and Gatekeeper Mutants of BCR-ABL , PDGFRα , Kit , and Src Kinases: Novel Type II Inhibitor of Gatekeeper Mutants. **2010**, *115* (21), 4206–4216. https://doi.org/10.1182/blood-2009-11-251751.

(78) Richters, A.; Ketzer, J.; Getlik, M.; Grütter, C.; Schneider, R.; Heuckmann, J. M.; Heynck, S.; Sos, M. L.; Gupta, A.; Unger, A.; et al. Targeting Gain of Function and Resistance Mutations in Abl and KIT by Hybrid Compound Design. *J. Med. Chem.* **2013**, *56* (14), 5757–5772. https://doi.org/10.1021/jm4004076.

(79) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the Performance of 3D Virtual Screening Protocols: RMSD Comparisons, Enrichment Assessments, and Decoy Selection—What Can We Learn from Earlier Mistakes? *J. Comput. Aided. Mol. Des.* **2008**, *22* (3–4), 213–228. https://doi.org/10.1007/s10822-007-9163-6.

(80) Jones, K.; McHardy, T. Cancer Drug Discovery 2010: From Molecules to Medicine. *Expert Rev. Clin. Pharmacol.* **2010**, *3* (5), 613–615. https://doi.org/10.1586/ecp.10.50.

(81) Alexander, L. T.; Möbitz, H.; Drueckes, P.; Savitsky, P.; Fedorov, O.; Elkins, J. M.; Deane, C. M.; Cowan-Jacob, S. W.; Knapp, S. Type II Inhibitors Targeting CDK2. *ACS Chem. Biol.* **2015**, *10* (9), 2116–2125. https://doi.org/10.1021/acschembio.5b00398.

(82) Andraos, R.; Qian, Z.; Bonenfant, D.; Rubert, J.; Vangrevelinghe, E.; Scheufler, C.; Marque, F.; Régnier, C. H.; De Pover, A.; Ryckelynck, H.; et al. Modulation of Activation-Loop Phosphorylation by JAK Inhibitors Is Binding Mode Dependent. *Cancer Discov.* **2012**, *2* (6), 512–523. https://doi.org/10.1158/2159-8290.CD-11-0324.

(83) Amato, K. R.; Wang, S.; Hastings, A. K.; Youngblood, V. M.; Santapuram, P. R.; Chen, H.; Cates, J. M.; Colvin, D. C.; Ye, F.; Brantley-Sieders, D. M.; et al. Genetic and Pharmacologic Inhibition of EPHA2 Promotes Apoptosis in NSCLC. *J. Clin. Invest.* **2014**, *124* (5), 2037–2049. https://doi.org/10.1172/JCI72522.

(84) Matthews, J. M.; Bhatt, S.; Patricelli, M. P.; Nomanbhoy, T. K.; Jiang, X.; Natkunam, Y.; Gentles, A. J.; Martinez, E.; Zhu, D.; Chapman, J. R.; et al. Pathophysiological Significance and Therapeutic Targeting of Germinal Center Kinase in Diffuse Large B-Cell Lymphoma. *Blood* **2016**, *128* (2), 239–248. https://doi.org/10.1182/blood-2016-02-696856.

(85) Ren, X.; Pan, X.; Zhang, Z.; Wang, D.; Lu, X.; Li, Y.; Wen, D.; Long, H.; Luo, J.; Feng, Y.; et al. Identification of GZD824 as an Orally Bioavailable Inhibitor That Targets Phosphorylated and Nonphosphorylated Breakpoint Cluster Region-Abelson (Bcr-Abl)

Kinase and Overcomes Clinically Acquired Mutation-Induced Resistance against Imatinib. *J. Med. Chem.* **2013**, *56* (3), 879–894. https://doi.org/10.1021/jm301581y.

(86) Tan, F. H.; Putoczki, T. L.; Stylli, S. S.; Luwor, R. B. Ponatinib: A Novel Multi-Tyrosine Kinase Inhibitor against Human Malignancies. *Onco. Targets. Ther.* **2019**, *12*, 635–645. https://doi.org/10.2147/OTT.S189391.

(87) DiMauro, E. F.; Newcomb, J.; Nunes, J. J.; Bemis, J. E.; Boucher, C.; Buchanan, J. L.; Buckner, W. H.; Cee, V. J.; Chai, L.; Deak, H. L.; et al. Discovery of Aminoquinazolines as Potent, Orally Bioavailable Inhibitors of Lck: Synthesis, SAR, and in Vivo Anti-Inflammatory Activity. *J. Med. Chem.* **2006**, *49* (19), 5671–5686. https://doi.org/10.1021/jm0605482.

(88) Martiny-Baron, G.; Holzer, P.; Billy, E.; Schnell, C.; Brueggen, J.; Ferretti, M.; Schmiedeberg, N.; Wood, J. M.; Furet, P.; Imbach, P. The Small Molecule Specific EphB4 Kinase Inhibitor NVP-BHG712 Inhibits VEGF Driven Angiogenesis. *Angiogenesis* **2010**, *13* (3), 259–267. https://doi.org/10.1007/s10456-010-9183-z.

(89) Laurenzana, I.; Caivano, A.; La Rocca, F.; Trino, S.; De Luca, L.; D'Alessio, F.; Schenone, S.; Falco, G.; Botta, M.; Del Vecchio, L.; et al. A Pyrazolo[3,4-d]Pyrimidine Compound Reduces Cell Viability and Induces Apoptosis in Different Hematological Malignancies. *Front. Pharmacol.* **2016**, *7*. https://doi.org/10.3389/fphar.2016.00416.

(90) Tintori, C.; La Sala, G.; Vignaroli, G.; Botta, L.; Fallacara, A. L.; Falchi, F.; Radi, M.; Zamperini, C.; Dreassi, E.; Dello Iacono, L.; et al. Studies on the ATP Binding Site of Fyn Kinase for the Identification of New Inhibitors and Their Evaluation as Potential Agents against Tauopathies and Tumors. *J. Med. Chem.* **2015**, *58* (11), 4590–4609. https://doi.org/10.1021/acs.jmedchem.5b00140.

(91) Tintori, C.; Laurenzana, I.; La Rocca, F.; Falchi, F.; Carraro, F.; Ruiz, A.; Esté, J. A.; Kissova, M.; Crespan, E.; Maga, G.; et al. Identification of Hck Inhibitors As Hits for the Development of Antileukemia and Anti-HIV Agents. *ChemMedChem* **2013**, *8* (8), 1353–1360. https://doi.org/10.1002/cmdc.201300204.

(92) Kinoshita, T.; Matsubara, M.; Ishiguro, H.; Okita, K.; Tada, T. Structure of Human Fyn Kinase Domain Complexed with Staurosporine. *Biochem. Biophys. Res. Commun.* **2006**, *346* (3), 840–844. https://doi.org/10.1016/j.bbrc.2006.05.212.

(93) Kim, H. T. Crystal structure of Lyn kinase domain in complex with N-(1H-indazol-6-yl)-8-(piperidin-4-yloxy)-6-propylquinazolin-2-amine. https://www.rcsb.org/structure/5XY1.

(94) UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45* (D1), D158–D169. https://doi.org/10.1093/nar/gkw1099.

(95)   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

(96)   Sabat, M.; VanRens, J. C.; Laufersweiler, M. J.; Brugel, T. A.; Maier, J.; Golebiowski, A.; De, B.; Easwaran, V.; Hsieh, L. C.; Walter, R. L.; et al. The Development of 2-Benzimidazole Substituted Pyrimidine Based Inhibitors of Lymphocyte Specific Kinase (Lck). *Bioorg. Med. Chem. Lett.* **2006**, *16* (23), 5973–5977. https://doi.org/10.1016/j.bmcl.2006.08.132.

(97)   Wiederstein, M.; Sippl, M. J. ProSA-Web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. https://doi.org/10.1093/nar/gkm290.

(98)   Sippl, M. J. Recognition of Errors in Three-Dimensional Structures of Proteins. *Proteins Struct. Funct. Genet.* **1993**, *17* (4), 355–362. https://doi.org/10.1002/prot.340170404.

(99)   Phelan, J. D.; Young, R. M.; Webster, D. E.; Roulland, S.; Wright, G. W.; Kasbekar, M.; Shaffer, A. L.; Ceribelli, M.; Wang, J. Q.; Schmitz, R.; et al. A Multiprotein Supercomplex Controlling Oncogenic Signalling in Lymphoma. *Nature* **2018**, *560* (7718), 387–391. https://doi.org/10.1038/s41586-018-0290-0.

7 APPENDIX

7.1 Publications

- Fallacara, A. L.; Passannanti, R.; Mori, M.; Iovenitti, G.; Musumeci, F.; Greco, C.; Crespan, E.; Kissova, M.; Maga, G.; Tarantelli, C.; et al. Identification of a New Family of Pyrazolo[3,4-*d*]Pyrimidine Derivatives as Multitarget Fyn-Blk-Lyn Inhibitors Active on B- and T-Lymphoma Cell Lines. *Eur. J. Med. Chem.* **2019**, *181*, 111545. https://doi.org/10.1016/j.ejmech.2019.07.048.