UNIVERSITA' DEGLI STUDI DI SIENA
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 IDENTIFICATION OF NEW DDX3X INHIBITORS

SETTORE SCIENTIFICO-DISCIPLINARE: CHIM/08

DOTTORANDA:

Claudia Immacolata Trivisani

TUTOR:

Prof. Fabrizio Manetti

ANNO ACCADEMICO: 2019/2020

ACKNOWLEDGMENTS

I would first like to thank my supervisor, Professor Fabrizio Manetti, whose expertise was invaluable in formulating the research questions and methodology. Many thanks for the support, encouragement, patience, and for having sharpened my scientific thinking.

Special thanks to Dr. Annalaura Brai. Laura, who shared with me this project of research, is a scientific and personal reference point. Thank you for the answers you gave to all my questions and for the trust you have placed in me. Many thanks also to our research group.

I would also like to thank the Institute of Molecular Genetics (IGM-CNR) of University of Pavia. I would like to thank Professor Giovanni Maga and Dr. Valentina Riva for the enzymatic and kinetic studies on small molecules as inhibitors of the helicase site and unique motif of DDX3X.

Thanks to Professor Simone Giannecchini (University of Florence) and Professor Maurizio Zazzi (university of Siena) for the in vitro evaluation of the antiviral activities of the compounds.

I would like to acknowledge my supervisor at the University of Dundee, Professor Alessio Ciulli. I want to thank you for your support and for the opportunity to work in such an inspiring environment. Many thanks also to Dr. Nikolai Makukhin, Dr. Sarath Ramachandran, and Dr. Ryan Casement with who I have had the opportunity to work.

My biggest thanks go to Professor Maurizio Botta, who believed in me and gave me the opportunity to follow my dreams.

I would like to thank my mother, the strongest and most courageous person I have ever known. Thanks also to my sister, to stay near me even if I am so far from you. Thanks to my aunt, to be always by my side, to be my friend. Thanks to Federico, my lighthouse. Thanks to you and to my family for allowing me to return home when all seemed lost. Thanks also to Tub, I will never forget your white jumpsuit at the airport. I also thank all the people who have become an important part of my family.

Thanks to my friends. Alex, there will never be too many words to thank you. Thanks also to Anna and Angelica, that are always present in my life. Thanks to Francesca and Luca, who started with me this journey. Thanks to my special roommate, Giulia, for all the moments together. Thanks to Federica, Vladimir, Elisabetta, Angelo, Roberta, Maddalena, Angelo, Mario, Alessandro, and all the people that are in my life.

Finally, I would like to thank my father. You are always close to me.

Summary

ABSTRACT

The search for new antiviral drugs for the treatment of clinical and emerging viruses is a delicate task. Drugs available today are developed to target a specific virus or viral strain, and only few prophylactics show a broad-spectrum activity that can be used to treat drug-resistant infections or in case of emergencies. Among the strategies that can be pursued to search for new antiviral drugs, there is the inhibition of a host protein involved in the viral replication cycle. The inhibition of DDX3X, a human ATP-dependent RNA helicase, allowed to discover the first broad-spectrum antiviral compound able to inhibit the replication of HIV resistant strains, HCV and of emerging viruses like West Nile Virus, Japanese Encephalitis Virus, Dengue Virus.

In this project, several computational strategies have been applied to improve the biodistribution and pharmacokinetic properties of this compound and a fluorescent inhibitor was designed to understand the mode of action of DDX3X inhibitors in DENV infected cells.

The selective inhibition of DDX3X can be pursued targeting a small pocket, peculiar to the human protein, called unique motif (UM). The study of the interactions established by the first active compound within UM, allowed to identify the amino acids responsible of its activity. Considering these findings, a small library of derivatives able to establish the fundamental interactions with the UM was designed. Moreover, a pharmacophore-based virtual screening procedure allowed to discover new compounds that will be biologically evaluated as new UM inhibitors.

PROTAC is instead a potent strategy to target protein degradation. A PROTAC molecule, that is constituted by two active moieties hold together by a linker, allow the selective ubiquitination and degradation of the protein of interest by the proteasome. In this contest, the application of computational procedures on known SOCS2 binders, allowed to establish a SAR that will be used to design novel derivatives that can be used both as inhibitors of the protein and as binders of the E3 ligase. A pharmacophore-based virtual screening performed on the Elongin C allowed to discover new compounds that will be biologically evaluated to establish if they can be used as E3 ligase binders.

Introduction

Viral infections represent the greatest pandemic threat of the modern era. Relatively few prophylactics or therapeutics are available to treat clinical viruses such as the Human Immunodeficiency Virus (HIV), hepatitis viruses, and emerging viruses like SARS, the avian and swine influenza strains, and viruses that may be used for bioterrorism.

Antiviral drugs available today are divided into three macro-categories. The first category includes specific compounds designed to target a viral protein, such as, for example, the HIV inhibitors. These compounds are developed to target a specific virus or viral strain avoiding the interference with host cellular functions, but the high rate of mutation that characterizes viruses often led to the development of drug resistance and to unforeseen adverse effects.

In the second category, there are vaccines. Also in this case, there is the need to develop new vaccines for each virus or viral strain. To be effective, a vaccine must be administered before or soon after the viral exposure and is not immediately available in case of emerging viral threats. Moreover, the development of a vaccine could be difficult for some pathogens, such as HIV.

The last category includes interferons and other pro- or anti-inflammatories that are useful only against specific viruses and that are responsible of serious adverse effects since they directly interact with the host immune and endocrine systems.

Today, only few antiviral compounds are characterised by a broad-spectrum profile. Favipiravir (**Figure 1**) is an influenza RNA-dependent RNA polymerase inhibitor available in Japan that is characterized by broad spectrum RNA antiviral activity and is effective against the polymerase of Ebola and Lassa fever. [1, 2]

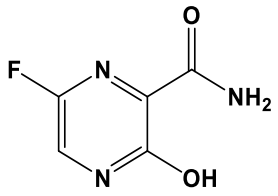


Figure 1. Structure of Favipiravir.

Cidofovir (Figure 2a) is a nucleotide analogue, with a broad-spectrum activity directed against DNA viruses including Herpes virus, Polyomavirus, Adenovirus, and Poxvirus, but it is characterized by high toxicity. [3] Brincidofovir (Figure 2b), a derivative of Cidofovir, is an antiviral drug in phase of development. It is characterized by an *in vitro* activity directed against DNA and RNA viruses, but no human efficacy was noted, and it is unclear if *in vitro* findings were the result of cell toxicity or of a true antiviral effect. [4]

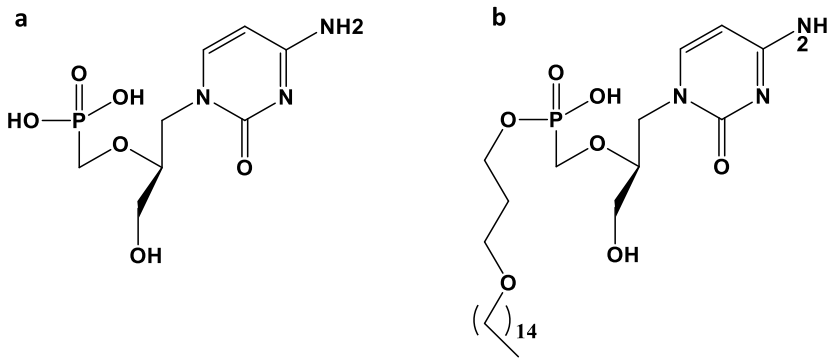


Figure 2. (a.) Structure of Cidofovir. (b.) Structure of Brincidofovir.

Ribavirin (Figure 3) is a nucleoside analogue able to inhibit viral polymerase enzymes, and is characterized by a broad spectrum activity against DNA viruses such as adenoviruses, and RNA viruses, such as Respiratory Syncytial Virus (RSV), Hepatitis C Virus (HCV), human influenza A and B viruses, parainfluenza viruses (HPIV), Hepatitis E Virus (HEV), metapneumovirus (hMPV), Crimean-Congo Haemorrhagic Fever (CCHF), and New and Old-World Haemorrhagic Arenaviruses. [5] It was also unsuccessfully used during the outbreak of the RNA coronavirus SARS in 2002. Unfortunately, ribavirin is associated with serious toxicity because of the development of hemolytic anemia that can result in worsening of cardiac diseases, leading to fatal and nonfatal myocardial infarctions. [6]

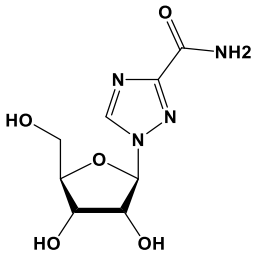


Figure 3. Structure of Ribavirin.

Remdesivir is a prodrug of a nucleotide analogue that is intracellularly metabolized to an analogue of adenosine triphosphate able to inhibit the viral RNA polymerases (Figure 4). It is effective against both filoviruses and coronaviruses [7] and in 2020 it has been used for the compassionate use in patients with severe SARS-CoV-2 infection, although it did not show significant clinical benefits. [8, 9]

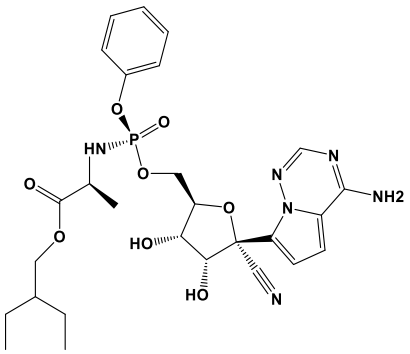


Figure 4. Structure of Remdesivir.

The drug repurposing strategy is pursued to search for new antiviral drugs, trying to identify new therapeutic uses for approved or investigational drugs. Although antiviral compounds are developed to be directed against a specific target, some drugs have been repositioned to be used in other viral infections. Examples of drug repurposing are Ganciclovir or Valganciclovir, reported in Figure 5. These two molecules are Citomegalovirus (CMV) antiviral agents that show activities against Adenovirus and Hepatitis B Virus. [10, 11]

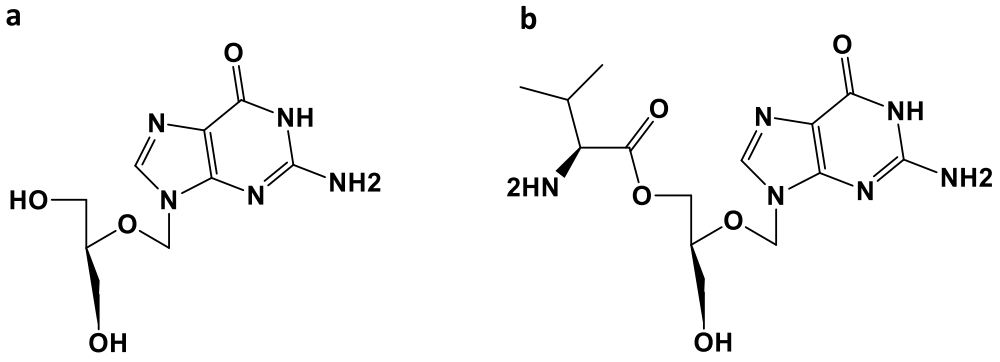


Figure 5. (a) Structure of Ganciclovir and (b) Valganciclovir.

Tenofovir is an agent employed in the treatment of HIV and HBV infections that is active against the DNA polymerase of Herpes Simplex Virus (HSV). [12] Foscarnet is another antiviral compound used in resistant herpesvirus infections also able to inhibit the reverse transcriptase (an RNA dependent DNA polymerase) of HIV-1 and HIV-2. [13] Nitazoxanide (Figure 6) is an anti-parasitic and anti-bacterial medication that shows an antiviral activity, using distinct mechanisms, against Hepatitis B, Hepatitis C, Norovirus, Rotavirus, Dengue, HIV, Yellow Fever, Japanese Encephalitis, and Influenza viruses. [14]

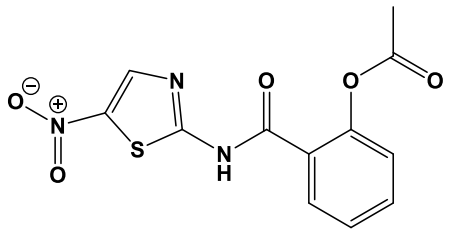


Figure 6. Structure of Nitazoxanide.

Despite all the antiviral compounds commercially available and the drug repurposing strategy, the lack of broad-spectrum antiviral agents creates a big chasm in fighting viral infectious emergencies. Today, the antiviral drug discovery process is focused on finding compounds that are able to interact with individual viruses and not with the whole family. Several antiviral agents have been repurposed and some of that have shown broad-spectrum properties, but it is urgent the development of a broad-spectrum antiviral agent that can be used in combination to the existing therapies or

to the monoclonal antibodies.[15] Moreover, it is necessary to balance the broad-spectrum activity with the toxicity against the host.

In this contest, the research group of Professor Botta discovered the first broad-spectrum antiviral compound able to inhibit the replication of several viruses (HIV, HCV, WNV, JEV).[16] The peculiarity of this compound is that it is directed against a host protein, the DEAD-box ATP-dependent RNA helicase DDX3X. The helicase is involved in several aspects of RNA metabolism and is a fundamental cofactor of the viral replication. Knockdown studies for DDX3X showed that the helicase does not impair the cell viability, while its inhibition is able to avoid the viral replication. Moreover, targeting a host cofactor to develop an antiviral compound it is possible to overcome the drug resistance problem.[17] Hence, this compound represents the starting point for the application of some computational approaches aiming to develop broad-spectrum antiviral agents as a new strategy to fight viral infections.

CHAPTER 1

HUMAN DEAD-BOX HELICASE DDX3X

DEAD-box RNA helicases are involved in several aspects of the RNA metabolism, including transcription, RNA splicing and stability, mRNA export and translation, and mitochondrial gene expression. DDX3X is a host cofactor that is widely used both by HIV, Hepatitis C Virus (HCV) and Poxviruses, among the responsible for the major global health threats, and by West Nile Virus (WNV), Dengue Virus (DENV) and Japanese Encephalitis Virus (JEV), involved in emerging rare diseases. The inhibition of host cofactors essential for viral replication is an emerging strategy that can be pursued to develop novel compounds, also avoiding the drug-resistance induced by high rate of mutations that characterizes viruses.

1.1 HUMAN DEAD-BOX HELICASES

DNA and RNA helicases are classified in five superfamilies (**Figure 7**) based on the sequence conservation of the helicase core. [18, 19]

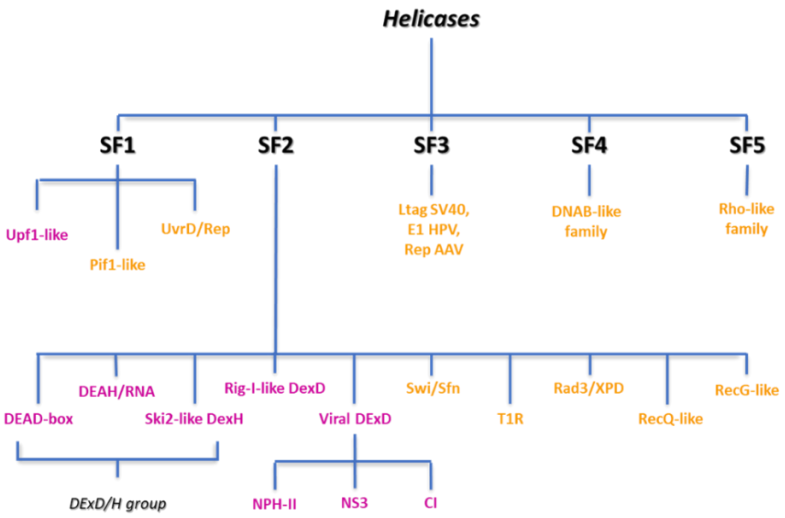


Figure 7. Classification of helicases into superfamilies (SF1-5) and families. RNA helicases are represented in magenta. In orange, the DNA helicases. SF1 and SF2 are further divided into families and groups based on the conservation of the helicase core. SF2 is subdivided into five families of RNA helicases and five families of DNA helicases. Within the RNA helicase family, the DExD/H group is comprised of the DEAD-box proteins (characterized by the conserved amino acid sequence Asp-Glu-Ala-Asp), the DEAH proteins (characterized by the amino acid sequence Asp-Glu-Ala-His), the RNA-like and Ski2-like proteins. On the other hand, the viral DExD family is divided into three groups: nucleoside triphosphate phosphohydrolase-II (NPH-II)-like helicases of Poxviridae, nonstructural protein 3 (NS3)-like helicases of Flaviviridae, and cytoplasmic inclusion (CI)-like helicases of Potyviridae. Finally, SF3 contains DNA helicases found in viruses, while SF4 and SF5 are DNA helicases of bacterial origin.

All the eukaryotic RNA helicases are included in SF1 and SF2. They are characterized by two globular domains (domains 1 and 2) connected by a short flexible linker. Since the helicase core is structurally related to the bacterial recombinase A protein (RecA), the domains are known as RecA-like. The motifs that form the helicase core are located at a defined position and are involved in ATP binding and hydrolysis, RNA binding, and in the coupling of ATP hydrolysis to unwinding. Motifs are not always present in all the members of the helicase family. In fact, the level of sequence conservation decreases among different families. [20] SF2 contains most of RNA helicases that are further divided into subfamilies and distinguished based on the amino acid sequence (**Figure 8**) of the conserved helicase motif II (DEAD, DEAH, DExH, and DExD helicases). [21]

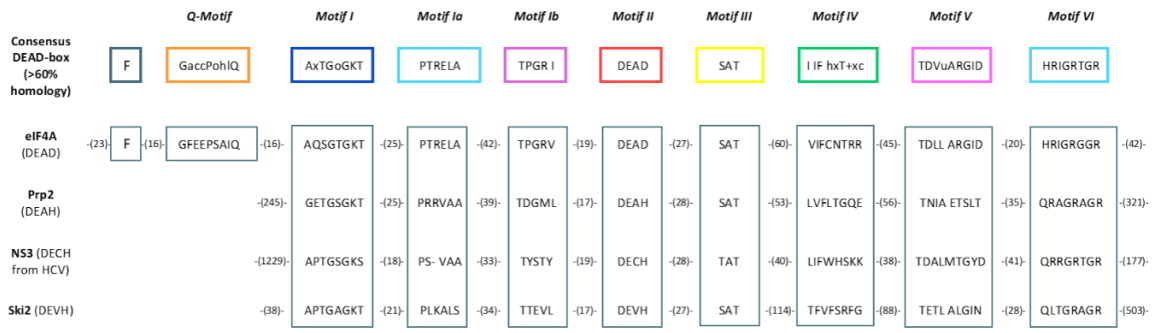


Figure 8. Conserved motifs in the DEAD-box and DExD/H-box families. Sequences of the conserved motifs from S. cerevisiae eIF4A (DEAD-box protein), Prp2 (DEAH-box protein), NS3 (DECH helicase from the hepatitis C virus), and Ski2 (DExH, Ski2 family).

RNA helicases are involved in the RNA duplex separation and strand exchange, strand annealing, and RNA assembly/disassembly, working in every cellular process associated with RNA metabolism, including pre-mRNA processing, ribosome biogenesis, RNA turnover, export, translation, surveillance, storage, and decay. RNA helicases can interact with other proteins to form multi-protein complexes playing additional roles in the cell and integrating different RNA metabolism processes.

The DEAD-box proteins are members of SF2 and can be distinguished by other RNA helicases by the variations within their conserved motifs. [22] They are present in all

the domains of life, playing a role in the transcription, translation, and degradation of prokaryotic, archaea, eukaryotic genomes, including the viral ones, [23] accomplishing regulatory functions essential for cell viability. Like the other helicases, they are involved in RNA metabolism and ribonucleoproteins (RNPs) remodeling. It is the largest RNA helicase family and is characterized by nine conserved motifs that form the ATP and RNA binding sites. The conserved motifs of the DExD/H helicases are clustered in a central core region of about 400 amino acids. At the same time, the N- and C-terminal extensions are highly variable in size and composition, reflecting the organization of the SF1 and SF2 superfamilies. [24]

The DEAD-box family contains more than 500 proteins that have been found in all eukaryotes and several prokaryotes. [25] The name of the family is due to the amino-acid sequence D-E-A-D (Asp-Glu-Ala-Asp) of its Walker B (or Motif II) motif.

DEAD-box proteins use the energy of the ATP hydrolysis to rearrange inter- or intra-molecular RNA structures or dissociate protein complexes from the RNA. In addition to the role played in RNA metabolism, two other functional features appear to be present in this family, such as the involvement in recurrent chromosomal translocations (DDX6, DDX10) or overexpression (DDX1, DDX4, DDX5, DDX6) in cancerous cells. However, despite the involvement of DEAD-box proteins in many diverse mechanisms, their precise roles, functions, and regulation largely remain to be elucidated.

1.2 HUMAN DEAD-BOX HELICASE DDX3X

DDX3X (or DBX) is an ATP-dependent RNA-helicase, member of superfamily 2. It is constituted by 12 conserved domains that fold together into two RecA-like domains hold together by a short flexible linker. The protein was identified in 1997 as one of the five X-chromosomal genes that have homologues in the non-recombining region of the Y-chromosome (DBY, DDX3Y). [26] The DDX3X gene escapes the X-inactivation and is ubiquitously expressed in a wide range of tissues, while the Y homolog gives an essential contribution in the early male germ cell development and maintenance. [27]

DDX3X is involved in various steps of the RNA metabolism, including ribosome biogenesis, mRNA transcription, pre-RNA splicing and stability, mRNA-export and translation, and mitochondrial gene expression. Several viruses require DDX3X for their replication process, and on this basis, the inhibition of DDX3X could be a novel therapeutic approach for the development of antiviral drugs. It has also been demonstrated that DDX3X is involved in the induction of antiviral mediators, and that plays an important role in cell cycle control, apoptosis and cancer. [28]

DDX3X is a shuttle protein between the cytoplasm and the nucleus. The export of DDX3X could be mediated both by the shuttle protein Chromosomal Maintenance 1 (CRM1 or Exportin 1) and via the TAP-dependent export pathway, [29] that is responsible for the nuclear export of mRNAs.

The DDX3X is predominantly localized in the cytoplasm. Nowadays, we know that it is mainly localized in the nucleus of healthy primary epidermis cells, while its localization is mostly cytoplasmic in skin tissue from cutaneous squamous cell carcinomas, suggesting a difference between transformed and non-transformed cells. According to this information, it is possible to consider the nuclear import and export of DDX3X as a highly regulated process, but the mechanisms at the base of its localization remain to be elucidated. [30]

1.2.1 DDX3X IN RNA METABOLISM

DDX3X is a transcriptional regulator able to activate the promoter of specific genes, [31] and is involved in the transcriptional processes using different pathways. DDX3X is a component of the spliceosome and the exon junction complex (EJC), but its role in mRNA metabolism is still poorly characterized. DDX3X is also involved in protein synthesis, as demonstrated by the helicase association to the translation initiation machinery and different cytoplasmic mRNA granules. [28]

1.2.2 DDX3X IN NUCLEAR EXPORT PATHWAY

The nuclear pore complex is responsible for exporting proteins and RNAs managed by several nuclear transport receptors. DDX3X interacts with CRM1, the primary mammalian export receptor that transports proteins containing a leucine-rich NES, [32] and TAP (Transporter associated with antigen processing) protein complex, [29] the main mRNA exporter. CRM1 is involved in the export of ribosomal and small nuclear RNAs, while TAP is recruited to export spliced mRNAs via the EJC. The interaction between DDX3X and TAP occurs via the C-terminus of DDX3X, with the helicase that could be bound to poly(A)-mRNAs in both nuclear and cytoplasmic fractions. DDX3X seems not to be required for general mRNA export, while it may use the TAP complex to export a specific mRNAs subset.

It has been found that DDX3X is incorporated in at least two distinct mRNPs that export different transcripts through the CRM1-dependent pathway. In the first mRNP, DDX3X is involved in the export of the cyclin D1, E1, A2, and B1 transcripts characterized by the presence of a structurally conserved RNA element in their 3'-UTR that is sensitive to the translation initiation factor eIF4E. The presence of this element allows these mRNAs to be exported into the cytoplasm using the eIF4E/CRM1-dependent mechanism in a cap-dependent but translation-independent manner. DDX3X is also a component of the eIF4E-dependent export mRNP present in the nuclear fraction, with an association not mediated by the transported mRNA. [33]

In the second complex, DDX3X is involved in the nuclear export of intron-containing HIV-1 mRNAs (**Figure 9**).

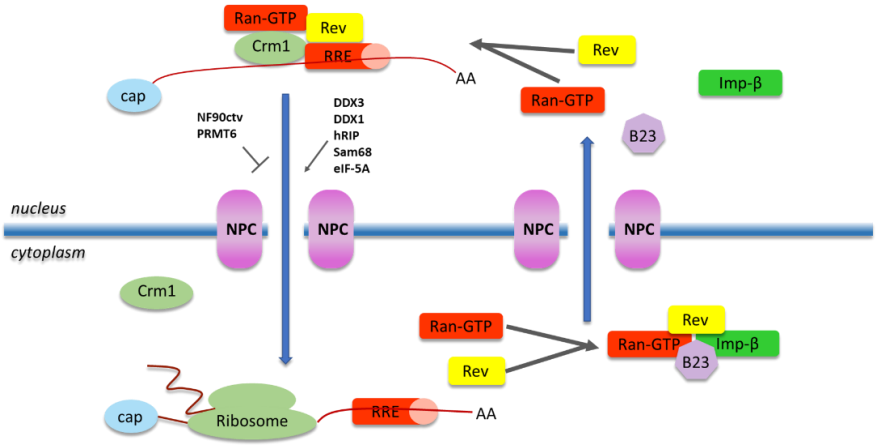


Figure 9. Export of the HIV-1 RNA to the cytoplasm. 9 and 4 kb viral RNAs contain RRE sequences necessary to interact with the Rev/Crm1/RanGTP complex and subsequent export. The DDX1 and DDX3 helicases facilitate the translocation of the Rev-RNA complex to the cytoplasm through the nuclear pore complex (NPC). Host cell factors could have positive (hRIP, SAM68, eIF-5a) or negative (NF90ctv, PRMT6) effects on the Rev-dependent export of the viral RNA, but their role is not entirely understood. Importin-b, RanGDP, and B24 are required for the import of Rev into the nucleus.[32]

Tat, Rev (a sequence-specific nuclear mRNA-export factor), and Nef are regulatory proteins encoded by the fully spliced mRNAs of HIV. The incompletely spliced HIV-1 mRNAs are responsible for producing auxiliary (Vif, Vpr, Vpu) and structural proteins. The unspliced, singly spliced, and multiply spliced HIV RNAs need to be exported into the cytoplasm to be translated. If Rev is absent, the incompletely spliced HIV-1 mRNAs are blocked into the nucleus, while the fully spliced HIV-1 mRNAs are exported, including the mRNA encoding Rev, since it is independent by Rev function. The unspliced HIV-1 mRNAs are characterized by the REV response element (RRE), a cis-acting sequence that specifically binds Rev. The N-terminal sequence of Rev is responsible for the binding to the RRE, while the C-terminal portion of Rev contains a sequence of 10 leucine amino acids that represent the Rev nuclear export signal (NES). When this sequence binds CRM1, the nuclear export of Rev/RNA occurs.

CRM1 interacts with some Nuclear Pore Complex components, the largest protein complex that exchanges components between nucleus and cytoplasm, and is essential for CRM1-mediated nuclear RNA export. [28]

1.2.3 DDX3X IN VIRUSES

DDX3X has been associated with the replication of several viruses that have a significant impact on human health, playing a dual function in the replication steps: first, DDX3X can be a cofactor of viral replication, and second, it is a mediator of the innate immune system. [34]

The helicase is a fundamental cofactor for the replication of HIV and other viruses. The interaction between DDX3X and the core protein of the hepatitis C virus (HCV) [35] allows the control of the infB promoter induction mediated by the helicase in response to the viral infection. To avoid that DDX3X stimulates the expression of INF-β, HCV uses the IPS-1 pathway to block the INF-β induction, while vaccinia virus and HBV interfere with the TBK1/IKKε pathway.

Other viruses employ DDX3X to facilitate their replication: Japanese Encephalitis Virus (JEV), murine Norovirus (NV), Dengue Virus (DENV), and West Nile Virus (WNV), even if the role played by the helicase during viral replication remain to be elucidated.

Given the important role played by DDX3X in the replication cycle of viruses, it has been considered a potential target for the development of novel antiviral compounds. Despite the fact that drugs directed against a viral protein are highly specific, targeting a host cofactor can have several advantages such as creating a high barrier to resistance, provide broad coverage of different genotypes/serotypes and multiple viruses, and expanding the list of potential targets, especially when druggable viral targets are limited.

1.2.4 DDX3X IN CANCER

DDX3X is involved in cell cycle regulation, cellular transformation, and apoptosis, playing oncogenic and tumour suppressor roles with a mechanism not yet elucidated. Several DDX3X somatic truncating mutations, including nonsense mutations and frameshift indels, have been found in cancer patients. The accumulation of these somatic changes may cause cancer development, progression, drug resistance, and cell proliferation.

However, DDX3X is a multifunctional protein that can have both positive and negative effects on cell proliferation based on cell type and environmental conditions. Recent studies have shown that DDX3X is a direct downstream target of HIF-1α (the master regulator of the hypoxia response) in breast cancer cell lines. [36] Further studies have demonstrated that DDX3X is responsible for repressing the expression of Kruppel-like factor 4 (KLF4), a zinc-finger containing transcription factor and cell cycle repressor associated with growth arrest. [37] In addition, as proof of the oncogenic role of DDX3X, its activation by benzo[*a*]pyrene diol epoxide (BPDE), the active metabolite of benzo[*a*]pyrene present in tobacco smoke, can promote growth, proliferation and neoplastic transformation of breast epithelial cells. [38]

1.3 DDX3X: THE PROTEIN STRUCTURE

Twelve conserved motifs typical of the DEAD-box helicases family constitute the helicase core of DDX3X. In particular, Motifs Q, I (Walker A containing the phosphate-binding P-loop), II (Walker B, DExD-box), Ia, Ib, and III constitute the RecA-like domain 1 of DDX3X and are responsible for the ATP binding and hydrolysis, while motifs IV, V and VI are involved in the formation of RecA-like domain 2 (**Figure 10**) and RNA binding.

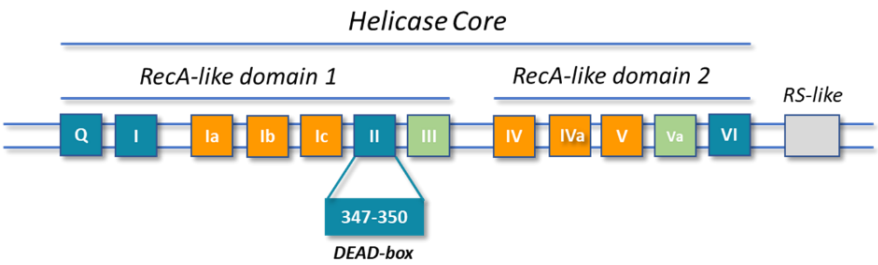


Figure 10. Schematic representation of the conserved motifs of DDX3X. These motifs fold together to form the two I RecA-like domains. These domains constitute the helicase core of the protein. Motif II contains the DEAD sequence, from which the proteins take their name.

The name of the two domains derives from their three-dimensional structure that resembles the one of the bacterial Recombinase A, with five β-strands surrounded by five α-helices.

The tails of DDX3X are largely unstructured and are constituted by various motifs with different functions. The N-terminal tail presents a CRM1-dependent NES [39] and an eIF4E-binding motif, [40] while the C-terminal is a low complexity region that contains conserved sequences essential for oligomerization. [41, 42]

DDX3X presents an ATP-dependent unwinding activity directed against RNA duplex of 50 nucleotides. Mutants with a total (K230E) or partial (S382L) loss of the ATPase activity do not show the unwinding activity. The presence of RNA or DNA can enhance the Mg^{+2}-dependent ATPase activity. The N- and C-terminal regions of DDX3X are also necessary for full ATPase activity *in vitro* and in the regulation of the catalytic properties. Residues 250-259 located between motifs I (ATP-binding) and Ia (RNA-binding) can play a role in positioning the 3'-OH of the RNA substrate close to a region of positively charged groups. Some functional *in vitro* and *ex vivo* assays showed that these residues were involved in RNA and ATP binding and RNA duplex unwinding.

Thus, DDX3X can bind and hydrolyse ATP and unwind RNA duplexes, consistent with its function as an mRNP remodeling factor.

1.4 THE MECHANISM OF RNA DUPLEX UNWINDING

In 2019 the research group of Ji X. has explained the functional cycle of RNA duplex unwinding of DDX3X. [43]

This cycle (**Figure 11**) is characterized by dramatic conformational changes in the helicase core. In particular, the protein can assume:

- The apo form (PDB ID: 5E7I);
- The pre-unwound form (PDB ID: 6O5F), in which there is the formation of the ternary complex DDX3X:dsRNA:DDX3X and the helicase binds the ATP;

- The post-unwound form (represented by our homology model built based on VASA DEAD-box helicase), in which the helicase binds the ATP and a ssRNA.
- The post-release form (PDB ID: 4PXA) in which the protein binds the ADP.

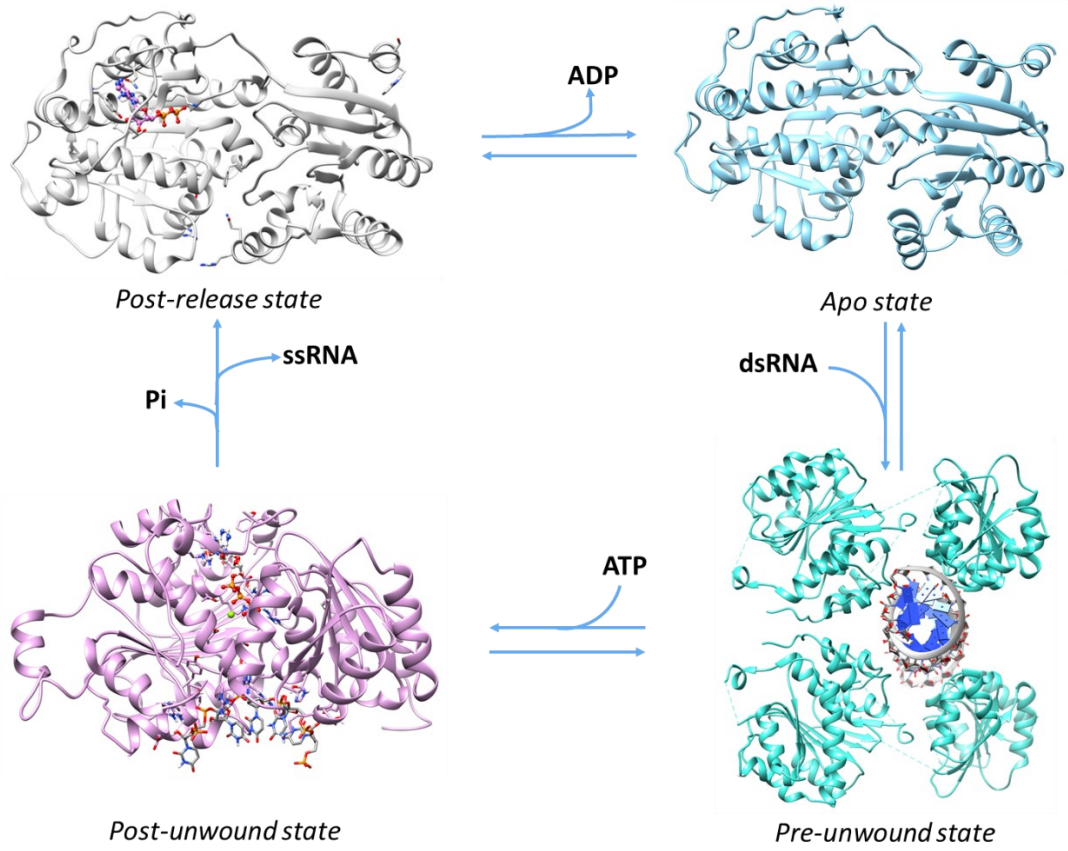


Figure 11. The functional cycle of DDX3X and representation of the conformational changes that characterize the unwinding process.

To switch from the apo to pre-unwound state, Domain 2 of DDX3X shifts of about 60 Å and rotates by 180°. The post-unwound state is reached through the shift of Domain 2 of about 35 Å and the rotation of 180°. To return in the apo state, Domain 2 shifts of about 25 Å without obvious rotation.

1.5 STATE OF ART

The research group of Professor Botta has hardly worked in targeting both the ATPase and RNA binding region of DDX3X, leading to the identification of the first

and second generation small molecules designed to inhibit the ATPase activity of the protein, while biological screening of known NTPase/helicase inhibitors led Yedavalli et al. to the identification of ring-expanded nucleosides (REN) as potential DDX3X inhibitors. [44] Subsequently, in the laboratory of Professor Botta, the first inhibitor specifically designed to target the helicase binding site was discovered, [45] laying the foundations for the development of new molecules with antiviral activity.

1.5.1 DDX3X ATPASE INHIBITORS

The first molecule able to inhibit the ATPase activity of DDX3X is FE15 (K_i = 5.4 μM, **Figure 12**) that was discovered in 2008. This compound is also able to inhibit the replication of HIV-1 in MT4-cells with an EC_{50} = 86.7 μM, without showing cytotoxicity (CC_{50} > 200 μM in MOLT-4 T-lymphocytic). [46]

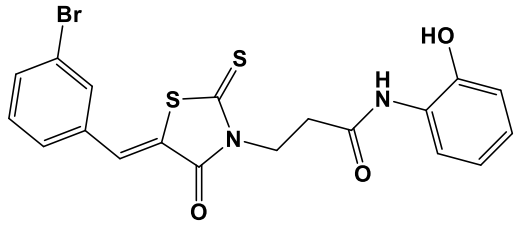


Figure 12. Structure of FE15, the first DDX3X ATPase inhibitor.

In the same year, a biological screening on a library of known NTPase/helicase inhibitors led Yedavalli et al. to the discovery of REN (ring expanded nucleosides) as inhibitors of the ATP-dependent activity of DDX3X, suppressing HIV-1 replication in T cells and monocyte-derived macrophages. [44]

In 2011, FE109 (**Figure 13a**), a second generation DDX3X inhibitor with an improved activity profile (K_i = 0.2 μM), was discovered. Additional inhibitors with a triazine scaffold were identified. Among them, the best biological profile was shown by FE87 (**Figure 13b**), with a K_i of 0.1 μM on DDX3X, an EC_{50} value of 2.0 μM in the inhibition of viral load of peripheral blood mononucleated cells (PBMCs) infected with HIV and cytotoxicity of 20 μM in HeLa cells (Selectivity Index = 10). [47]

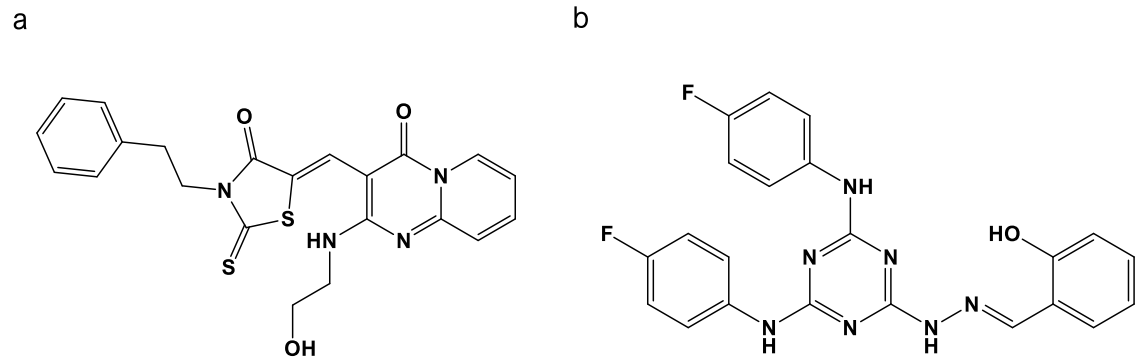


Figure 13. Structure of FE109 (a) and FE107 (b), both belonging to the second generation DDX3X inhibitors.

These compounds showed low selectivity in *in vivo* treatments, and, for this reason, next generation compounds able to inhibit the helicase activity of DDX3X was investigated.

1.5.2 DDX3X HELICASE INHIBITORS

The development of the helicase inhibitors is founded on the mechanism proposed by Schütz [48] who explains the opening of the RNA binding site. This mechanism is coupled with the presence of a conserved residue essential for the helicase activity and suggests that an inhibitor able to target this site could lock the DDX3X helicase in a catalytically inactive conformation. [49] In 2012, the first inhibitors of HIV-1 replication specifically designed to interfere with the RNA binding on DDX3X by interacting with the closed conformation of the enzyme, were discovered. [45]

The development of the homology model of DDX3X in the closed conformation and a high-throughput docking (HTD) approach applied to the RNA binding site led to the identification of high-affinity hit compounds within a commercial database. These studies allowed the identification of 3 compounds as promising hits that inhibit the helicase activity of DDX3X at micromolar concentration, and one of them has been chosen for further optimization, taking in account its easier synthetic accessibility and functionalization. Among the synthesized compounds, the most potent anti-helicase compound **6** (IC_{50} = 1 μM, **Figure 14**) significantly inhibited also the DDX3X ATPase activity (IC_{50} = 20 μM, respectively). No inhibition was observed for the

human DDX1 and the HCV NS3 RNA helicase, indicating a selective binding to DDX3X. Moreover, **6** was able to suppress HIV replication in PBMCs with an EC_{50} = 10 μM.

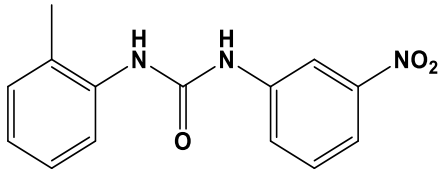
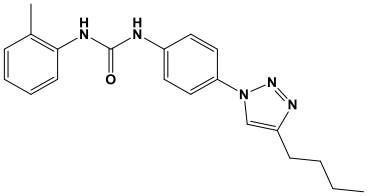


*Figure 14. Structure of **6**, a DDX3X helicase inhibitor.*

The first compound characterised by broad-spectrum antiviral activity (HIV, HCV, DENV, and WNV) in infected cells was **16d** (**Table 1**), that was discovered through optimization studies of **6**. [50]

*Table 1. Structure and antiviral activity of **16d**.*



Compound	DDX3X	HIV[a]	HCV[b]	JEV[c]	WNV[c]	Dengue[c]
16d	0.3	1.1	0.97	20	16.05	2.55

Activity, expressed as micromolar concentration, has been evaluated in [a]PBMC (Peripheral Blood Mononuclear Cells), [b]LUNET (LucUbiNeo-ET cells), and [c]Huh7 (Hepatocellular carcinoma cells).

16d was active against HIV-1 drug-resistant strains (**Table 2**), suggesting that DDX3X targeting agents may treat HIV/HCV coinfections, patients harbouring drug-resistant viruses, and emerging viral diseases, for which no specific drugs are available.

*Table 2. Compound **16d** was tested against several resistant strains of HIV, demonstrating that it inhibits their replication.*

HIV-1 STRAIN [a]	DRUG RESISTANCE CLASS [b]	IC_{50} [95% CI] (μM)	Fold Change [c]
114[d]	wild type	1.11 [0.31-3.90]	/
11808	PIs	0.23 [0.08-0.65]	0.2
7406	NRTIs	0.33 [0.13-0.87]	0.3
7404	NRTIs	0.22 [0.11-0.47]	0.2
12227	NNRTIs	0.94 [0.21-1.34]	0.8
12235	NNRTIs	0.36 [0.15-0.87]	0.3
11845	INIs	0.37 [0.26-0.52]	0.3

[a] NIH AIDS Reagent Program catalogue number (www.aidsreagent.org). [b] PIs: Protease Inhibitors; NRTIs: nucleos(t)ide reverse transcriptase inhibitors; NNRTIs: non nucleos(t)ide reverse transcriptase inhibitors; INIs: integrase inhibitors. [c] Ratio between IC_{50} toward resistant strain and IC_{50} toward wild type strain. [d] NL4-3 HIV-1 wild type reference strain.

Moreover, the good toxicity profile confirmed by *in vivo* studies suggests that the DDX3X activity, although essential for viral replication, may be dispensable to the cell as shown by preclinical studies.

1.6 AIM OF WORK

Considering all these findings and as part of our continuing efforts to identify novel antiviral agents, several computational approaches were used to improve activity, selectivity, and chemical-physical proprieties of DDX3X helicase inhibitors. To further improve the selectivity of these inhibitors, great attention was focused on an amino acidic sequence unique for DDX3X, that will be discussed later. The study of the binding pocket together with the application of several computational approaches allowed to rationally design a library of compounds able to establish the

fundamental interactions with the pocket and to project a more selective virtual screening protocol to search for new potentially active chemical scaffolds.

CHAPTER 2

HELICASE BINDING SITE: RESULTS AND DISCUSSION

This research project is based on the mechanism that regulates the RNA binding proposed by Schütz [48] in 2010. Schütz explains that DDX3X switches through three different conformational states, reported in **Figure 15**, to carry out its functions. In the open conformation, DDX3X binds the ADP, and the protein is completely incapable of RNA binding. After the ATP binding, the α-helix 8 starts to move out from RNA-pocket, and DDX3X assumes the pre-RNA binding conformation. Finally, in the closed conformation, the α-helix 8 rotates out of the RNA binding site allowing the RNA binding. This last conformation is stabilized by a salt bridge that is formed between the Asp350 of the α-helix 8 and the Arg503 of motif V.

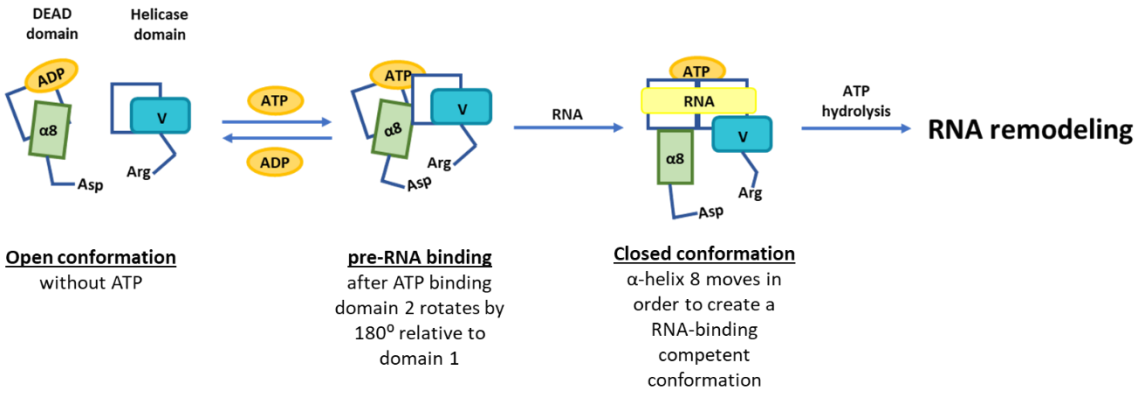


Figure 15. Schematic representation of the mechanism of RNA binding proposed by Schüz.

The ATP-mimetic inhibitors previously identified and able to bind the open conformation of DDX3X showed a lack of selectivity for *in vivo* treatments. Moreover, since the high homology between RNA helicases, the research of selective ATPase inhibitors of DDX3X is a challenging task. For this reason, targeting the RNA binding site present only in the closed conformation of DDX3X, it is possible to develop a more selective inhibitor able to lock the protein in a catalytically inactive state. Since no crystal structures of this conformation are still available on the PDB, a homology model was built [51] and used as a starting point for applying several computational protocols to discover new active molecules.

In 2019, Song and Ji released the crystal structure of a ternary complex constituted by DDX3X:dsRNA:DDX3X (PDB ID: 6O5F), [43] explaining a new mechanism at the base of RNA binding and unwinding. DDX3X switches among four different states, already discussed in paragraph 1.4, and our homology model has been represented as the post-unwound state of DDX3X.

2.1 THE HOMOLOGY MODEL

The homology model of the closed conformation, or post-unwound state of DDX3X has been built using Prime software, [52, 53, 54] to target the helicase site of the protein. Two templates were chosen to develop the model. The first one is the crystal structure of the closed conformation of Drosophila Vasa DEAD-box in complex with RNA (PDB ID: 2DB3). This template has been chosen since it has a sequence identity of 44% with the human DDX3X. The second template is the crystal structure of the open conformation of human DDX3X (PDB ID: 2I4I), used to model the missing parts of DDX3X in the Drosophila Vasa structure, such as the unique motif.

The comparison between the crystal structure of the DEAD-box protein VASA bound to poly(U) and the *in silico* modelled RNA-bound closed conformation of DDX3X, shows that all amino acids involved in RNA binding of VASA are conserved in DDX3X. This finding suggests a conserved mechanism of RNA binding in the DEAD-box helicase family that has been further confirmed in 2020, [55] with the release of the crystal structure of the post-unwound state of DDX21 (PDB ID: 6L5N).

As it is possible to observe in **Figure 16**, the comparison of the three structures highlights that the residues involved in the interaction with RNA are conserved among them.

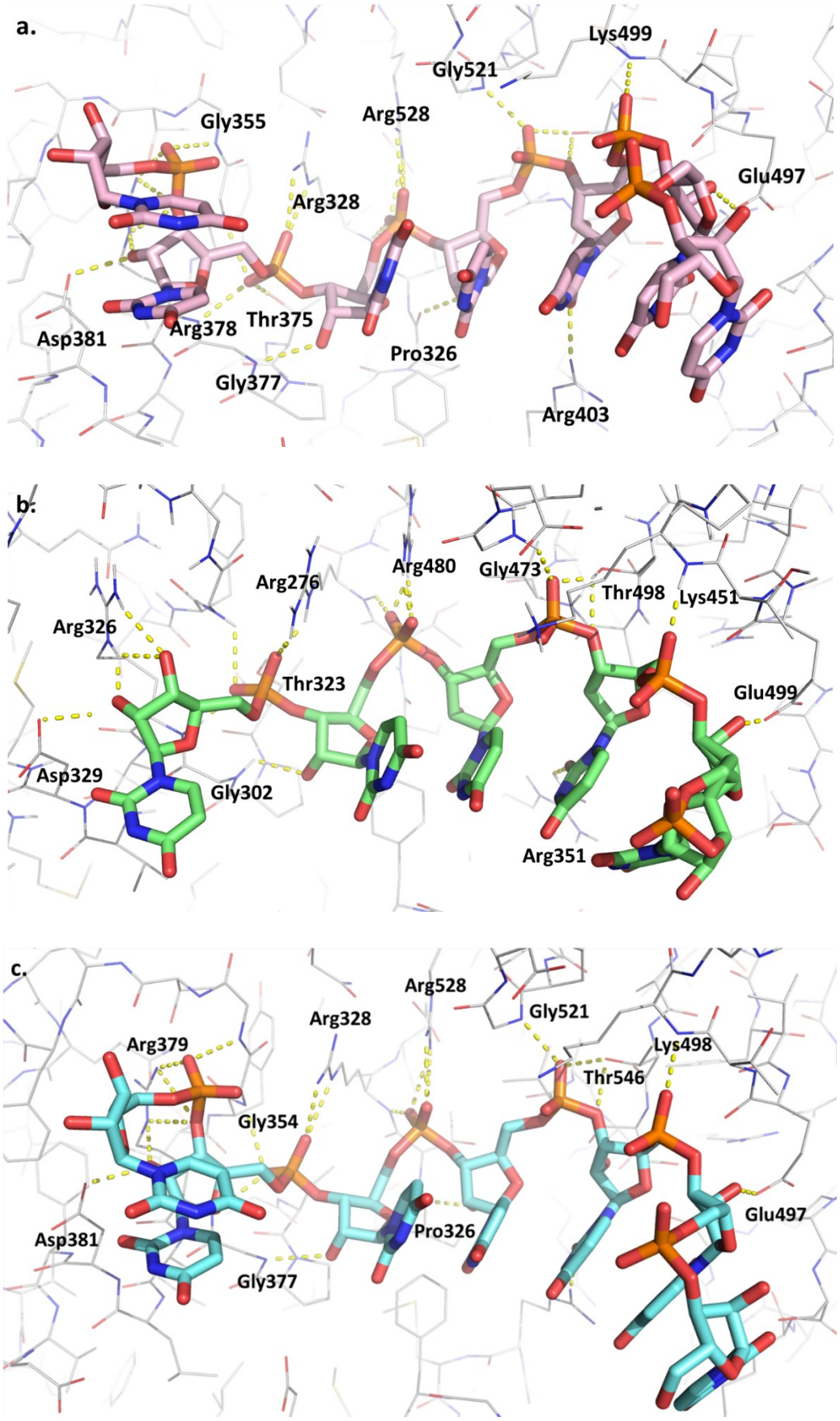


Figure 16. Analysis of residues involved in RNA binding. (a.) Representation of the interactions that involve VASA DEAD-box helicase and the poly(U) RNA strand. (b.) Representation of the interactions among the modelled

structure of the closed conformation of DDX3X and the poly(U) RNA strand. ***(c.)*** *Representation of the interactions between DDX21 and the poly(U) RNA strand.*

This evidence, that suggests that the RNA binding is a conserved mechanism in the DEAD-box helicase family, provides support for the reliability of our homology model.

2.2 DESIGN OF DDX3X INHIBITORS TO OVERCOME HIV-1 DRUG-RESISTANCE

Human immunodeficiency virus (HIV) is the causative agent of the acquired immunodeficiency syndrome (AIDS) which is responsible for over 1 million deaths per year worldwide. The efforts that the scientific community has made over the past 30 years have allowed the FDA to approve more than 25 drugs [56] capable of inhibiting viral replication. However, as the virus can latently persist in infected cells, these drugs are unable to eradicate this pathogen. [57] Today, the viral infection is treated using HAART (Highly Active Anti-Retroviral Therapy), [58] that consists in combination of three or more drugs able to halt the viral replication in most of treated patients. The HAART is life-long, and its efficacy can be compromised by the emergence of toxicities, comorbidities, and the selection of resistance mutations, which ultimately lead to the progression of the disease. For all these reasons, it is urgent to find novel drugs that can be used in the treatment of HIV-1 infection. DDX3X allows the design of indirect-acting antiviral agents (IAAs), an interesting class of drug characterized by a great genetic barrier that makes them less susceptible to resistance. [57] Since **16d** is able to inhibit the replication of HIV-1 wild type and drug-resistant strains, the previously built homology model of DDX3X has been used to rationally design a series of novel derivatives of **16d** to improve the biodistribution (BD) and pharmacokinetic (PK) properties.

The compound has been modified on the side chain at the triazole (C4) position, inserting polar groups and adding several hydrogen bond donors on the tolyl ring. Twenty-five molecules able to retain the **16d** main interactions (**Figure 17**) with the helicase binding site and characterized by a good *in silico* predicted aqueous solubility value, were selected to be synthesized and validated as DDX3X inhibitors.

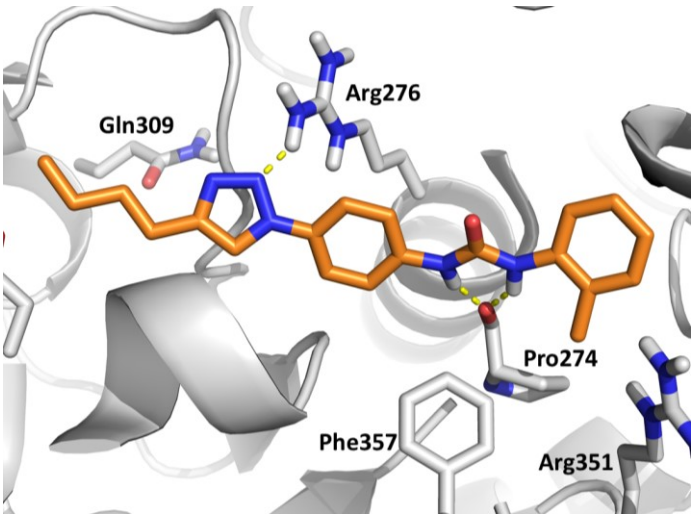


*Figure 17. Docking pose of **16d**. The compound was docked into the RNA binding site of the closed conformation of DDX3X. The binding mode analysis shows that the ureidic group of **16d** is involved in hydrogen bond interactions with the backbone of Pro274, while the triazole ring takes contacts with Arg276.*

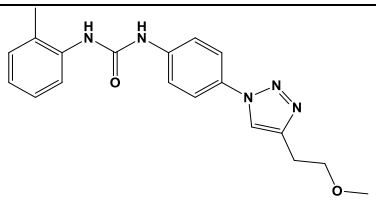
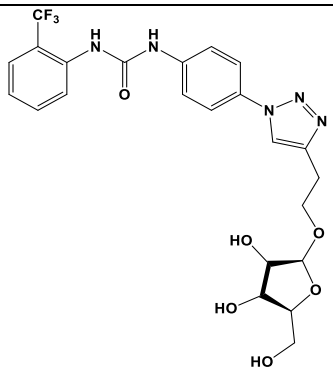
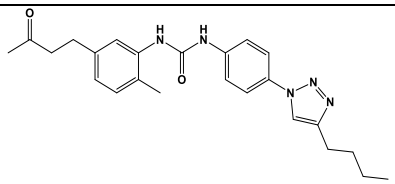
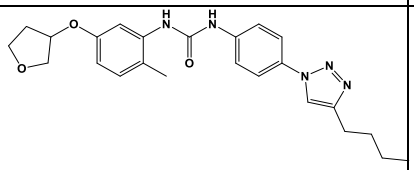
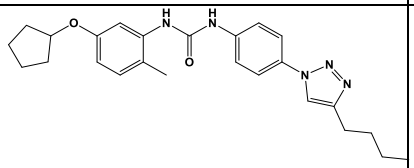
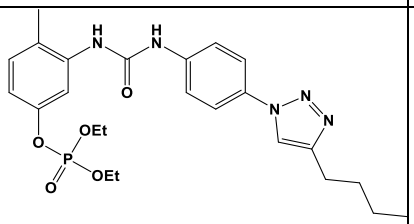
Among these compounds, 21 novel DDX3X inhibitors were found, and their ADME properties were experimentally confirmed *in vitro*. Eleven compounds were evaluated against HIV-1 infected cells, revealing promising antiviral activity and selectivity index (SI). Based on the ADME parameters, antiviral activity, and SI, **6b** was selected to evaluate its activity against a panel of HIV-1 resistant strains carrying mutations to currently approved direct antiviral agents (DAAs). BD, PK, and sub-chronic toxicity experiments were performed in mice and revealed an optimal tolerability and improved BD of the compound. [57]

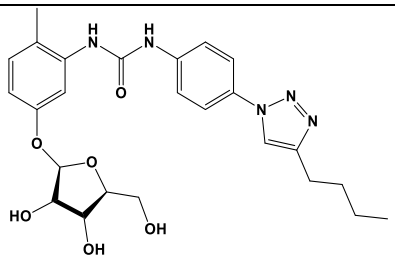
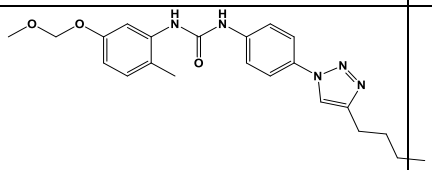
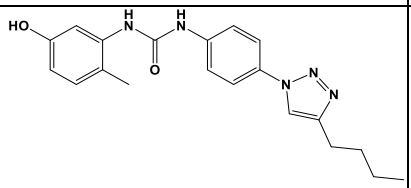
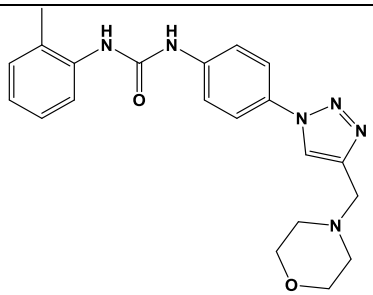
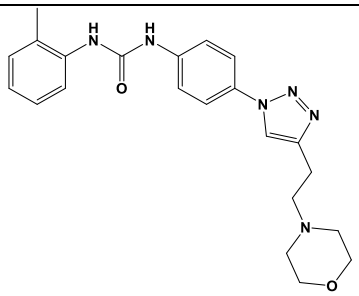
2.2.1 MOLECULAR MODELLING AND SAR STUDIES

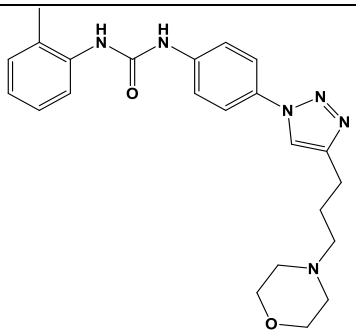
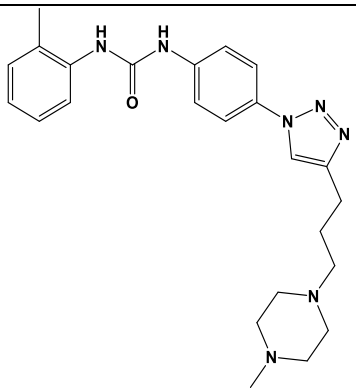
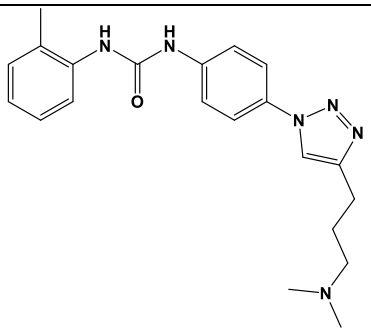
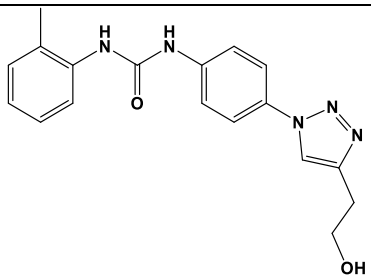
Table 3 summarizes the 25 novel derivatives of **16d**, together with their docking score obtained using Gold version 5.2, [59] the predicted aqueous solubility value calculated using QikProp, [60] and their anti-enzymatic activity. The activity was determined against the helicase activity of DDX3X using the fluorescence resonance energy transfer (FRET)-based assay [16] and is reported as half-maximal inhibitory concentration (IC_{50}).

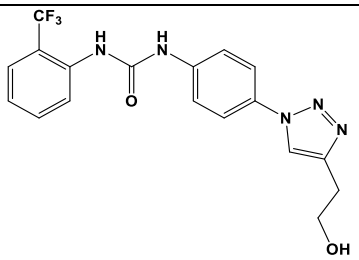
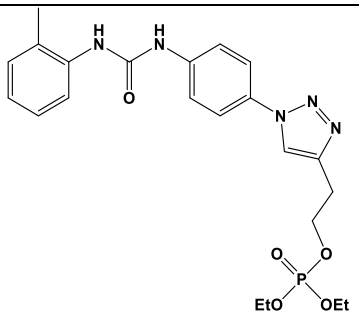
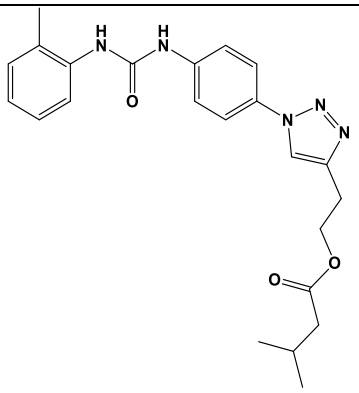
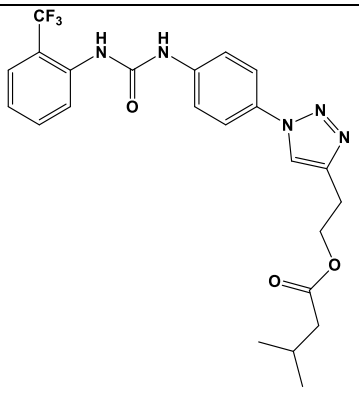
Table 3. Docking score, predicted aqueous solubility, and anti-enzymatic activity of the compounds against the DDX3X helicase.

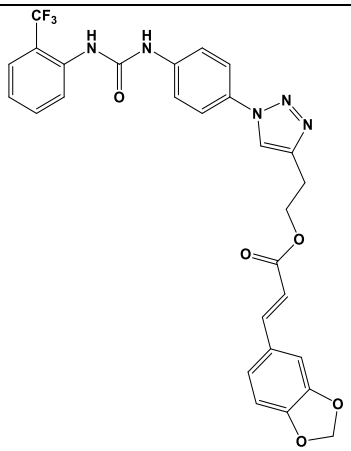
Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
16d		27.23	-6.40	0.30±0.16
6a		28.56	-6.39	0.15±0.09
6b		29.66	-6.78	0.12±0.10
7c		23.15	-5.91	n.d.
7g		23.54	-7.02	2.02±0.7
8d		30.02	-7.07	1.51±0.8

Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
9e		26.90	-6.00	1.00±0.5
9f		20.01	-4.49	n.a.
10a		30.79	-7.55	2.49±0.7
15a		31.01	-7.71	0.10±0.07
15b		31.63	-7.89	0.14±0.05
15c		29.33	-7.10	0.94±0.12

Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
15d		23.17	-5.82	n.a.
15e		25.87	-6.29	0.82±0.12
15f		24.89	-6.24	0.49±0.10
20a		27.31	-4.29	n.a.
20b		27.53	-4.42	59.68±3.0

Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
20c		28.01	-4.42	53.80±1.1
21c		28.38	-4.25	40.06±1.7
22c		26.83	-5.25	2.90±0.2
23a		23.79	-4.75	13.00±3.1

Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
23b		21.88	-5.46	11.20±1.2
24		26.36	-5.53	0.89±0.6
25		25.73	-7.60	1.29±2.1
26		23.58	-8.21	0.40±0.5

Cmpd ID	Structure	Docking Score	QPlogS [a]	IC_{50} [b] ±SD (μM)
27		25.86	-8.60	5.09±0.8

Data represent mean values of at least two experiments, each performed in duplicate ± S.D. [a] predicted aqueous solubility with acceptable values in the range of -6 and -1. [b] IC_{50} half-maximal inhibitory concentration. ***n.a.****: not active;* ***n.d.****: not determined, compound precipitated from the medium.*

As it is possible to see in **Table 3,** the introduction of an isoquinoline ring in **6a** (IC_{50}=0.15 μM) and **6b** (IC_{50}=0.12 μM) is well tolerated. The introduction of the isoquinoline ring allow the formation of a hydrogen bond with Arg351, and the establishment of hydrophobic interactions with Val500 and Ala499. The binding mode of these compounds is reported in **Figure 18**.

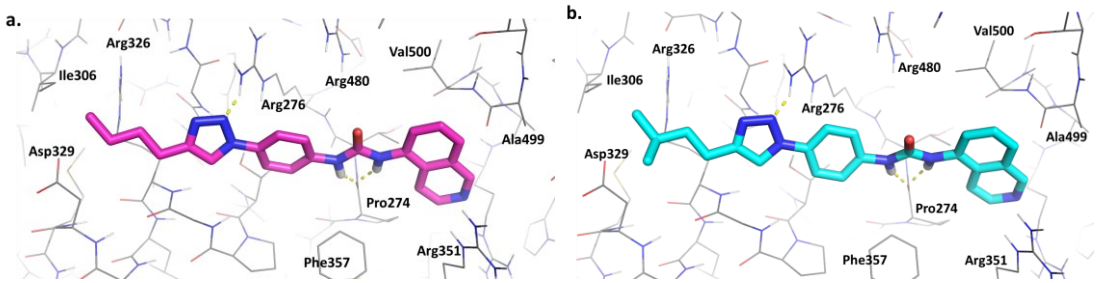


Figure 18. Binding mode of ***6a (a.)*** *and* ***6b (b.)****. Compounds retain* ***16d*** *main interaction. The isoquinoline ring is accommodated into a hydrophobic pocket constituted by Val500 and Ala499. The nitrogen atom takes contacts with Arg351.*

Introduction of the ester in **7g** allows to maintain a good inhibitory activity (IC_{50}= 2.02 μM), while its corresponding carboxylic acid **7c** precipitated during the assays.

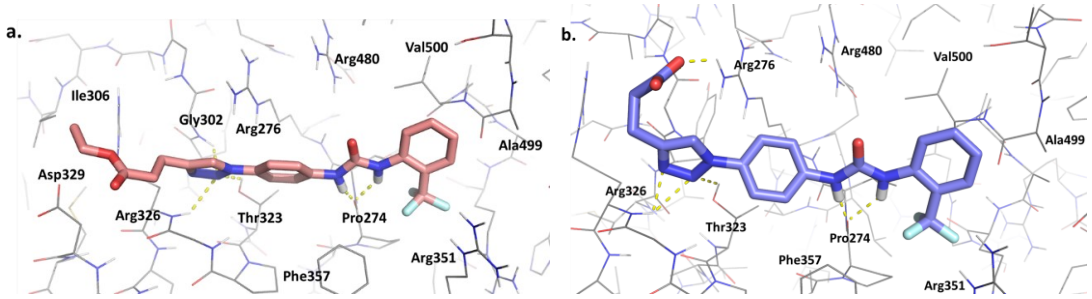


*Figure 19. (a.) Binding mode of **7g**. The triazole ring of **7g** interacts with Arg326 and the backbone of Gly302. The ureidic group is involved in hydrogen bonds with Pro274. (b.) Binding mode of **7c**. The carboxylic group of the compound interacts with Arg276. The triazole ring interacts with Arg326 and Thr323. The ureidic group interacts with Pro274.*

Ether derivatives **8d** and **9e** retain satisfactory activities of 1.51 and 1.00 μM respectively, while the introduction of ribose in the sidechain abolishes the activity (**9f**), probably due to the presence of a bulky group on the alkyl chain (**Figure 20c**).

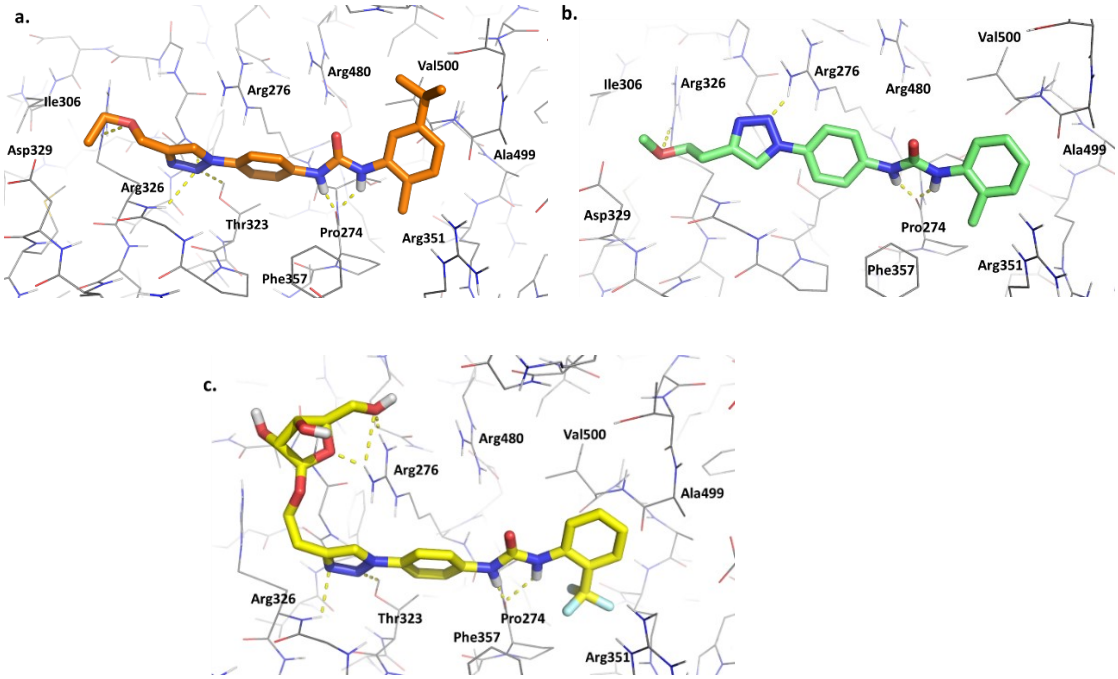
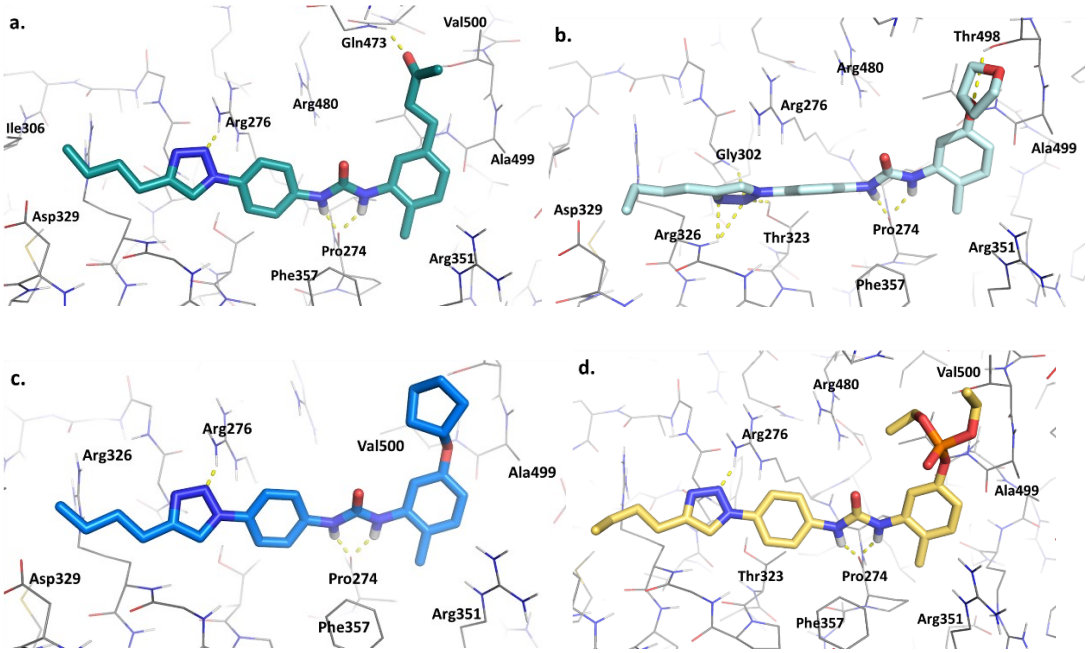


*Figure 20. (a.) Binding mode of **8d**. The triazole ring interacts with Arg326 and Thr323. The uredic group interacts with the backbone of Pro274. The ether oxygen interacts with Arg276. (b.) Binding mode of **9e**. Compound retains the same interactions of **8d**. (c.) Binding mode of **9f**. The ribose ring interacts with Arg276; the triazole ring interacts with Arg326 and Thr323. The ureidic group forms hydrogen bonds with Pro274.*

Several substitutions have been made on **16d** to evaluate the effect of hydrogen bond acceptor on tolyl ring. In particular, despite the additional interaction of the carbonyl group of **10a** with the backbone of Gln473, the compound presents an 8-fold reduction of the activity, probably due to the pocket constraints. The tetrahydrofuran derivative **15a** reaches an IC_{50} of 0.10 μM. **15a** retains all the key interactions of **16d**, and in addition, the ether-oxygen establishes a hydrogen bond with Thr498. Cyclopentyl derivative **15b**, phosphoric ester **15c** and 5-methoxymethoxyl derivative **15e** show high inhibitory activities of 0.14, 0.94 and 0.82 μM. All compounds form the key interactions with Gly302, Arg326 and Thr323, while the phenyl ring of **15c** establishes a cation-π interaction with Arg351, and the ether-oxygen of **15e** forms a hydrogen bond with Thr498. Phenol **15f** maintains a good inhibitory activity of 0.49 μM, while the introduction of a ribose ring causes a reduction of the activity, probably due to the high steric hindrance (**15d**). Binding modes of the compounds are reported in **Figure 21**.



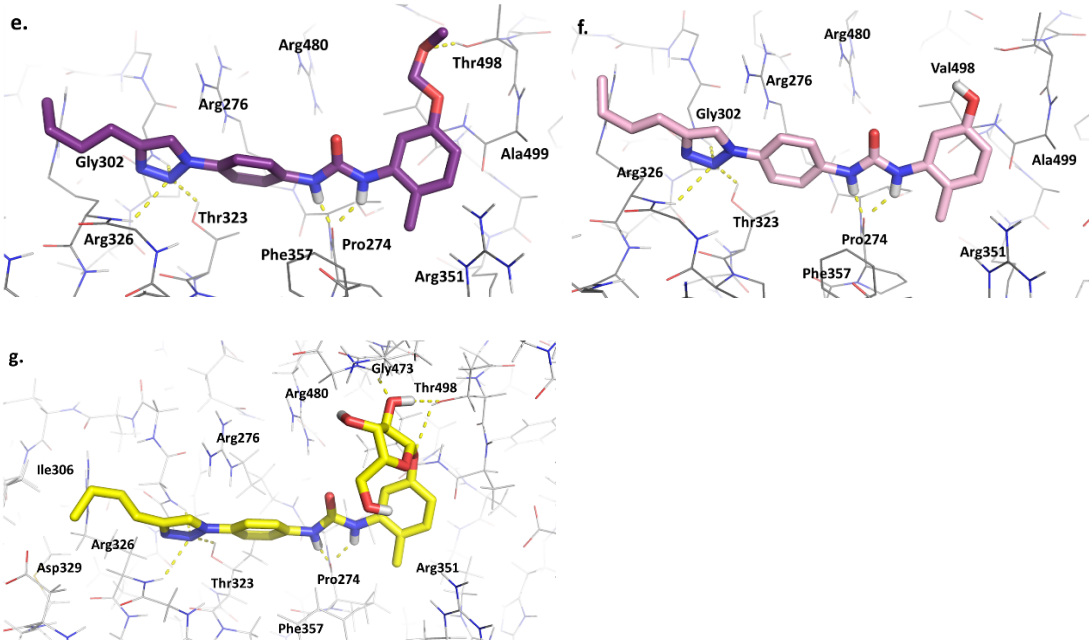


*Figure 21. (a.) Binding mode of **10a**. The triazole ring forms a hydrogen bond with Arg276. The ureidic group binds Arg480. The carbonyl group interacts with Gln473. **(b.)** Binding mode of **15a.** The triazole ring of **15a** interacts with Arg326, while the ureidic group interacts with Pro274. The ether-oxygen establishes a hydrogen bond with Thr498. **(c.)** Binding mode of **15b**. The triazole ring of the compound interacts with Arg276. The ureidic group interacts with the backbone of Pro274. **(d.)** Binding mode of **15c**. Compound retains the same interactions of **16d**. **(e.)** Binding mode of **15e**. The triazole ring interacts with the backbone of Gly302 and Arg326, and with the side chain of Thr323. The ureidic group binds Pro274. The oxygen of the para methoxymethyl-hydroxy group is involved in h-bond interactions with Thr498. **(f.)** Binding mode of **15f**. The triazole ring of the compound is involved in h-bond interactions with the side chain of Thr323 and the backbone of Arg326 and Gry302. The ureidic group binds the backbone of Pro274. **(g.)** Binding mode of **15d**. Thre triazole ring of the compouns interacts with Thr323 and with the backbone of Arg326. The ureidic group interacts with the backbone of Pro274. The ribose ring establishes hydrogen bond interactions with Thr498 and with the backbone of Gly473.*

The alkyl chain at the triazole (C4) position was explored by introduction of amines. Three different morpholine derivatives were evaluated, increasing the chain length from one to three methylenes (derivatives **20a-c**). The morpholine was replaced by the methylpiperazine (**21c**) and N,N-dimethylamine (**22c**). As reported in **Table 3**, the introduction of bulky and hindered groups causes a loss of activity, confirming that the linear and long sidechain are preferred. [16, 61]

This result has also been confirmed by the moderate activity of alcohols **23a** and **23b**, probably due to their lower occupancy within the pocket. Phosphate ester **24** retains a good inhibitory value, confirming that polar atoms are well tolerated in the presence of linear and flexible substituents. In particular, the phosphate ester of **24** establishes a hydrogen bond with Arg276, and the phenyl ring makes a cation-π interaction with Arg480. Among derivatives characterized by the introduction of

polar groups on the triazole sidechain, the best results were obtained with ester derivatives **25** and **27**, which make an additional hydrogen bond between the carbonyl ester and Arg276. In fact, **25** has an activity of 1.29 μM, and the corresponding trifluoromethyl derivative **26** shows an IC_{50} of 0.40 μM. Ester **27**, due to the presence of the bulkier cinnamoyl moiety, is less active with an IC_{50} of 5.09 μM, which confirmed our previous findings again.

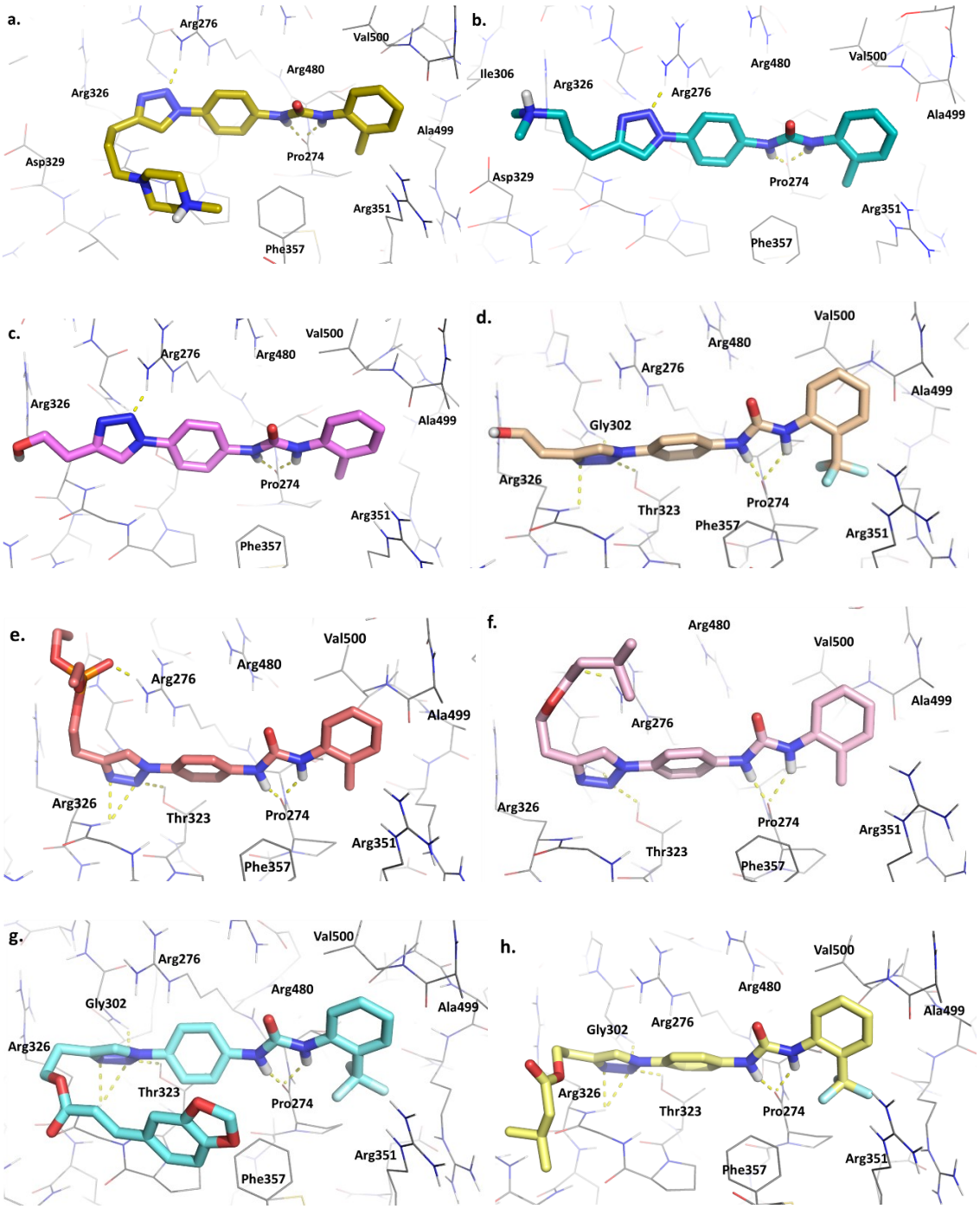


*Figure 22. (a. b. c.) Binding modes of **21c**, **22c,** and **23a**. Compounds retain the same interactions of **16d**. (**d**.) Binding mode of **23b.** The triazole ring interacts with Arg326 and Thr323. The ureidic group interacts with the backbone of Pro274. (**e.**) Binding mode of **24**. The ureidic group of 24 interacts with Pro274. The triazole ring interacts with Arg326 and Thr323. The phosphate ester forms hydrogen bonds with Arg276. (**f.**) Binding mode of **25**. The ureidic group and the triazole ring of 25 retain the same interactions of 26d, while the carbonyl ester interacts with Arg276. (**g**.). Binding mode of **27**. The triazole ring of **27** interacts with the backbone of Arg326, while the ureidic group interacts with Pro274. (**h.**) Binding mode of **26**. Compound interacts with Arg326 and Pro274.*

2.2.2 EVALUATION OF HIV ANTIVIRAL ACTIVITY AND CYTOTOXICITY

Nine compounds with the most promising DDX3X inhibitory activities, together with **23b** that showed an inhibitory activity below average, and **15d** that was found inactive against DDX3X and used as negative control, were evaluated for their antiviral activity against HIV. The antiviral activities of the compounds are reported in **Table 4**, together with the half maximal inhibitory concentration (CC_{50}) values. Compounds **6a**, **6b,** and **15b** show the best values and selectivity indexes and, in particular, the IC_{50} value of **15b** is 5-fold higher than that of **16d**.

Table 4. Antiviral activity of selected compounds against HIV-1 infected cells.

Cmpd ID [a]	IC_{50}[b]±SD (μM)	CC_{50}[c] (μM)	SI[d]
16d	1.11±0.5	90	81
6a	2.2±1.5	80	36
6b	2.1±1.0	86	41
7g	>50	125	-
8d	3.8±1.1	90	24
15a	2.6±1.0	16	6
15b	0.2±0.02	40	200
15c	5.5±1.6	16	2.9
15d	>50	100	-
23b	>50	80	-
24	>50	50	-

Cmpd ID [a]	IC_{50}[b]±SD (μM)	CC_{50}[c] (μM)	SI[d]
26	>50	95	-

[a] Data represent mean ± standard deviation of at least two experiments; [b] IC_{50}: half maximal inhibitory concentration; [c] CC_{50}: Half maximal cytotoxic concentration; [d] Selectivity Index (SI): CC_{50} to IC_{50} ratio.

Compound **8d** retains appreciable activity and low cytotoxicity, while **15a** and **15c** were found toxic at the concentration of 16 μM. The negative control **15d** was found inactive at the maximum concentration tested (50 μM), as well as **23** that does not show any significant activity up to 50 μM, probably due to its low inhibiting concentration. **24** and **26** were found inactive since their susceptibility to hydrolysis, as reported in paragraph 2.2.3.

Since **15b** is characterized by a low aqueous solubility (see paragraph 2.2.3), lower than **16d** and outside the recommended range reported for a drug candidate, [62] **6b** was chosen to be evaluated against a panel of HIV-1 strains carrying mutations that cause resistance to the major antiviral classes reported in **Table 5**. IC_{50} values of **6b** are comprised between 1.5 and 2.3 μM, resulting in fold change values ranging from 0.7 to 1.1, confirming both that resistant strains are fully susceptible to DDX3X inhibitor **6b** and that the Indirect Acting Antivirals (IAAs) are effective in overcoming viral resistance.

Table 5. *Antiviral activity of compound **6b** against HIV-1 resistant strains.*

Virus[a]	Drug class resistance[b]	Resistance mutations	IC_{50}[c]±SD (μM)	FC[d]
NL4-3	None	None	2.1±0.8	-
11808	PIs	Majior: V32I, I54V, I84V, L90M Accessory: L10F, V11I,	2.3±0.1	1.1

Virus[a]	Drug class resistance[b]	Resistance mutations	IC_{50}[c]±SD (μM)	FC[d]
		K20T, L33F, E35G, A71I, G73S, L89V		
7400	NRTIs	M41L, E44D, D67N, T69D, M184V, L210W, T215Y	1.8±0.8	0.8
11847	INIs	G140S, Q148H	1.5±0.1	0.7
12231	NNRTIs	K103N, V179F, Y181C	1.8±0.1	0.8

[a] NIH AIDS Reagent Program catalogue number (www.aidsreage3nt.org). [b] PIs: protease inhibitors; NRTIs: nucleos(t)ide reverse transcriptase inhibitors; NNRTIs: non-nucleoside reverse transcriptase inhibitors; INIs: integrase inhibitors; [c] Antiviral activity calculated with a two round of infection assay; [d] Fold change values indicate the ratio between IC_{50} values from resistant and wild type strains.

2.2.3 *IN VIVO* EVALUATION OF BIODISTRIBUTION, PHARMACOKINETICS AND TOXICITY

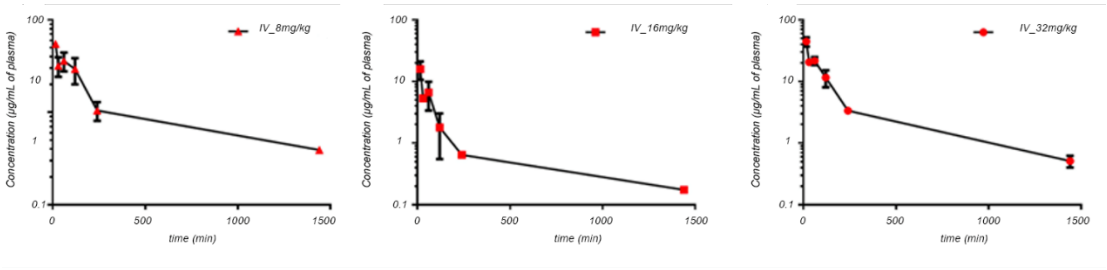
6b, 10-fold more soluble than **16d** and characterised by non-limiting value of AppP (**Table 6**), was selected as a preclinical candidate for the *in vivo* evaluation of the biodistribution, pharmacokinetics and toxicity.

Table 6. In vitro ADME studies of selected compounds.

Cmpd ID	AppP[a]	LogS[b]	HLM Stability[c]
16d	$1.86 \cdot 10^{-6}$	-7.0	99
6b	$0.18 \cdot 10^{-6}$	-5.7	89.7

[a] Apparent permeability reported in $cm \cdot s^{-1}$; [b] Aqueous solubility expressed as log of molar concentration; [c] Human liver microsomal metabolic stability expressed as percentage of unmodified parent drug.

The half-life elimination and the plasmatic clearance values denoted that **6b** was slowly eliminated after intravenous administration (**Figure 23**).



Intravenous Administration				
Parameter	Unit	8 mg/Kg	16 mg/Kg	32 mg/Kg
t1/2	min	538.1	465.4	280.9
Cmax	µg/mL	5.5	15.8	44.4
AUC o-t	µg/mL·min	1536.1	1422.1	5907.2
AUC 0-inf_obs	µg/mL·min	1839.8	1539.3	6114.4
Vz obs	L/Kg	3.4	7.0	207.1
Cl obs	mL/min/Kg	4.3	8.0	48.0

*Figure 23. Pharmacokinetic analysis. Four groups of 18 Balb/c mice were treated with compound **6b**, at the dose of 8 mg/kg, 16 mg/kg, or 32 mg/kg. Mice were all injected by i.v. administration via tail vein, and blood samples were collected at 15 and 30 minutes and 1, 2, 4, 6, and 24 hours. Samples were processed and analyzed by HPLC-MS.*

The analysis of tissue samples showed that **6b** reached the maximum concentration after 1 hour and was completely eliminated after 24 hours, as shown in **Figure 24**.

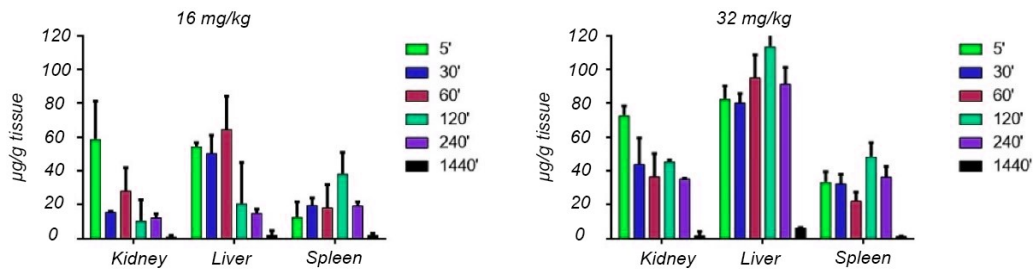


*Figure 24. Biodistribution of **6b** in Balb/c mice. Concentration levels of the compound in mice tissues at 5, 30, 60, 120, 240, and 1440 min of **6b** at the dose of 16 mg/kg and 32 mg/kg i.v.*

Since no trace of **6b** was found in the brain, the compound is not able to cross the blood-brain barrier. The presence of **6b** in the spleen suggests also that the compound could be used in the reservoir reduction. [63] Moreover, **6b** reaches good plasma concentration, major than its IC_{50}. In fact, after 240 minutes, its plasmatic concentration was around 10 μM, 5-fold higher than its IC_{50}. After 24 hours, the plasmatic concentration decreased to 2 μM, suggesting that a daily administration will be necessary in a future efficacy test.

Toxicity studies showed that no histological alterations were found at the tested dose, as highlighted in **Figure 25**, demonstrating that **6b** is well tolerated.

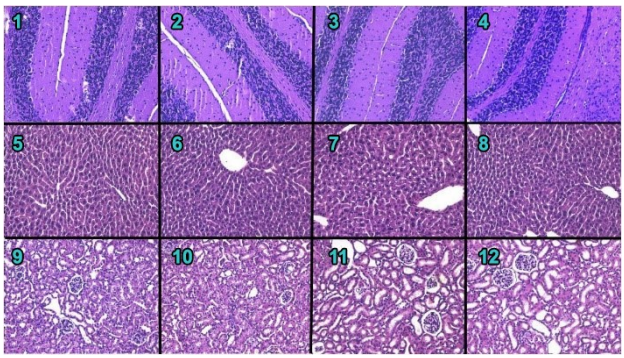


Figure 25. Representative images of the histological examination of HE-stained sections of ***Brains (A, B, C, D)****,* ***A-B****: Treated mice,* ***C****: Vehicle group,* ***D****: Wilde Type group;* ***Liver (E, F, G, H)****,* ***E-F****: Treated mice,* ***G****: Vehicle group,* ***H****: Wilde Type group;* ***Kidneys (I, L, M, N)****,* ***I-L****: Treated mice,* ***M****: Vehicle group,* ***N****: Wilde Type group. Treated mice do not exhibit abnormal histopathological changes compared with control groups.*

2.2.4 SUMMARY AND CONCLUSIONS

The homology model of the closed conformation (or post-unwound state) of DDX3X has been used to rationally design a series of derivatives of hit compound **16d** to ameliorate the biodistribution and pharmacokinetic properties. Twenty-five novel potential inhibitors of the helicase activity of the human protein DDX3X were selected based on their ability to retain compound **16d** main interactions within the binding site, and on their *in silico* predicted aqueous solubility. These compounds were synthetized and evaluated for their ability to inhibit the helicase activity of DDX3X. Twenty-one novel DDX3X inhibitors were found, with inhibitory activities ranging from 0.1 to 60 μM. The evaluation of 11 compounds against HIV-1 infected

cells revealed activities ranging from the low micromolar to submicromolar range, while a preliminary ADME analysis allowed to select the most promising compound for further *in vivo* assays. Compound **6b**, 10-fold more soluble than hit compound **16d**, was selected to perform additional studies. The antiviral activity of **6b** was evaluated against a panel of HIV-1 resistant strains carrying resistance mutations in the viral proteins targeted by the currently approved DAAs, with fold changes comprised between 0.8 and 1.1 that demonstrated the full susceptibility to the drug. PK, BD, and subchronic toxicity studies performed in mice revealed optimal tolerability and improved biodistribution with respect to **16d**, identifying **6b** as a novel promising DDX3X inhibitor with antiviral activity against HIV-1 wild type and resistant strains. Since i*n vivo* studies highlighted the optimal tolerability of the compound, **6b** is considered as a good candidate for *in vivo* experiments alone or in combination with other drugs with a different mechanism of action as a novel strategy to overcome the drug resistance associated with other DAAs, and it could be used in the treatment of the reservoirs. These results demonstrate that it is possible to use host targeting antivirals such as DDX3X inhibitors against viral diseases, in the treatment of resistant HIV strains that are no more susceptible to the DAAs currently available on the market. Moreover, the SAR analysis of analogues and congeneric compounds will be taken in consideration to rationally design novel back-up derivatives with improved ADME properties.

2.3 IDENTIFICATION OF NOVEL DENV-2 INHIBITORS

To search for novel Dengue Virus inhibitors, **16d** and a previously discovered sulfonamide series of WNV inhibitors able to inhibit the helicase activity od DDX3X, [61] were biologically evaluated to verify their activity against DENV-2. The assay allowed to discover that most of the compounds showed antiviral activity in the low micromolar range, with half-maximal inhibitory concentration (IC_{50}) values that are comparable to or lower than that of **16d**. Results are reported in **Figure 26**.

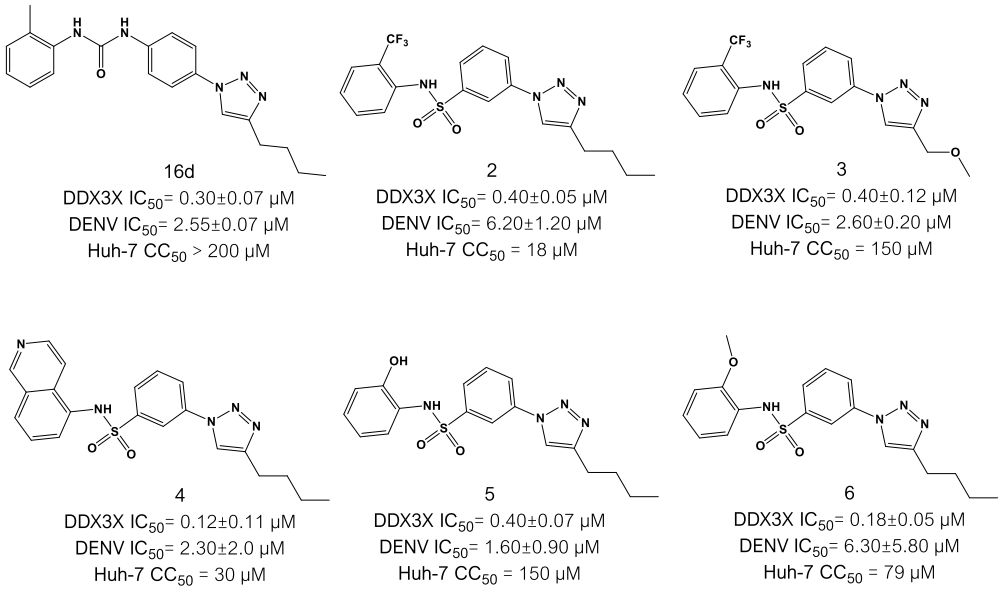


Figure 26. Structures of known DDX3X inhibitors, anti-enzymatic activities expressed as half-maximal inhibitory concentration (IC_{50}) ± standard deviation (SD), antiviral activity against DENV-2, and cytotoxicity on Huh-7 Cells.

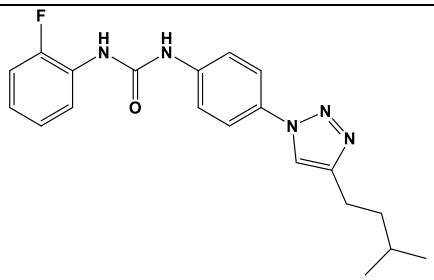
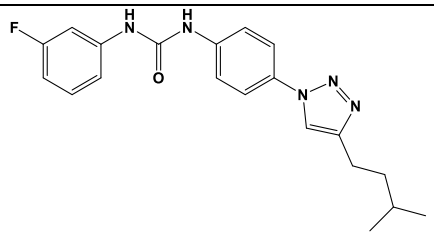
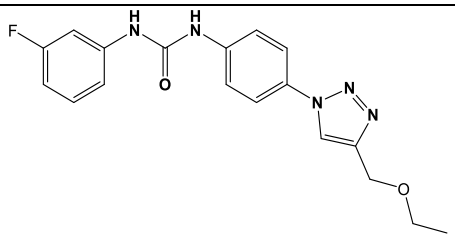
Toxicity studies showed that, among known DDX3X inhibitors (**Figure 26**), **5** is the most active compound against DENV infection with an IC_{50} of 1.6 μM and low cytotoxicity (SI = 94). [61] The trifluoromethyl derivative **2** has the lower anti-DENV activity and higher toxicity. **3** and **4** are characterized by antiviral activities in the low micromolar range, and the toxicity is comparable to that of **5**. The methoxy derivative **6** presents low antiviral activity and higher standard deviation, probably due to its lower solubility.

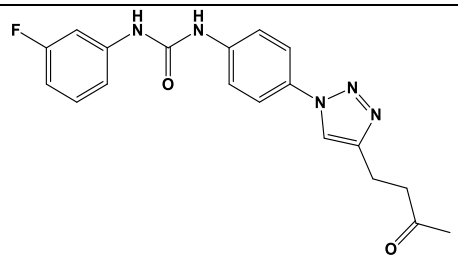
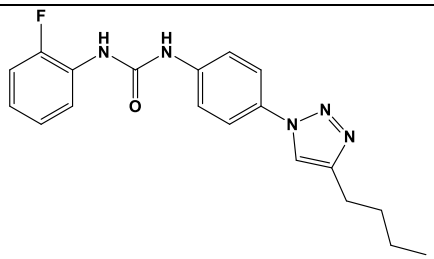
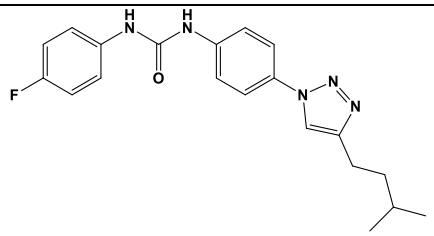
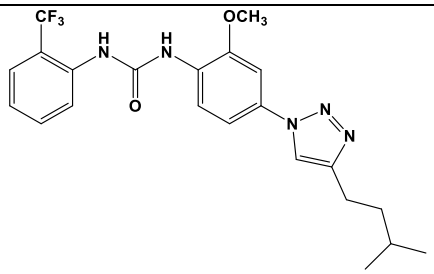
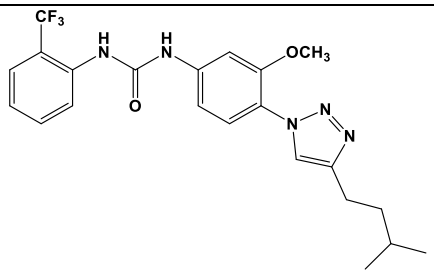
2.3.1 MOLECULAR MODELING AND SAR ANALYSIS

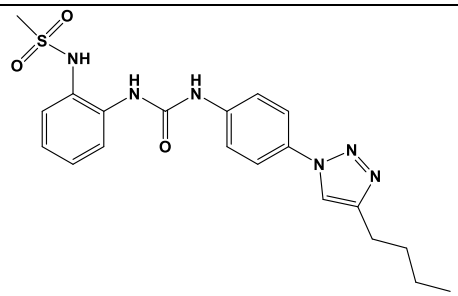
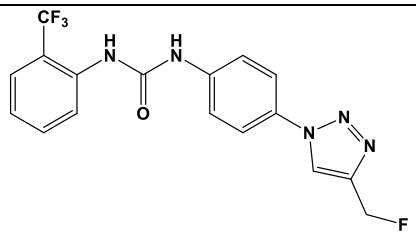
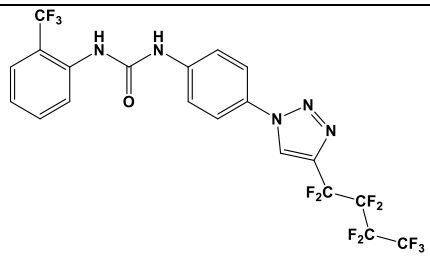
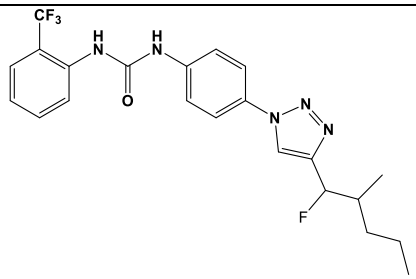
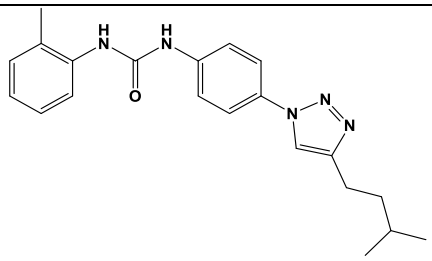
Taking in account these results, docking and molecular dynamics studies were carried out to design another small series of derivatives, introducing electron-withdrawing groups in the urea series and polar groups in the sulfonamide ones. The novel urea and sulfonamide compounds **7-24** were then synthesized and tested for their ability to inhibit the helicase activity of DDX3X using the FRET-based biochemical assay. [16, 51, 61]

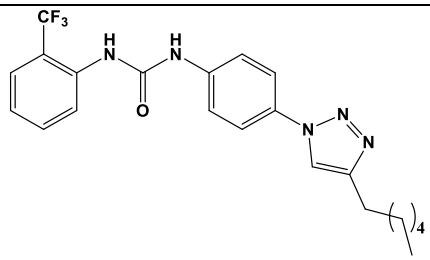
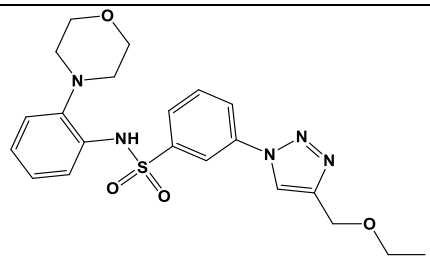
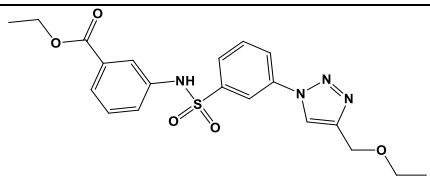
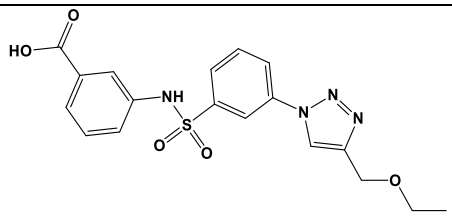
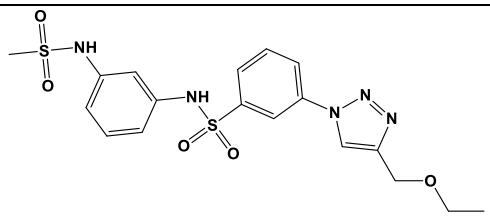
As reported in **Table 7**, several derivatives showed activities from the low micromolar to the submicromolar range.

Table 7. Structures of novel DDX3X inhibitors. The table below reports the docking score value obtained using Gold version 5.2, and the anti-enzymatic activities expressed as half-maximal inhibitory concentrations (IC_{50}) calculated using a FRET-based assay.

Cmpd ID	Structure	Docking Score	DDX3X IC_{50} [a] ±SD (μM)
7		26.24	0.12±0.10
8		25.33	0.20±0.09
9		24.81	21.2±1.50

Cmpd ID	Structure	Docking Score	DDX3X IC_{50} [a] ±SD (μM)
10		24.20	2.4±0.90
11		25.55	0.60±0.12
12		25.83	1.9±0.07
13		22.72	1.05±0.91
14		23.22	0.30±0.09

Cmpd ID	Structure	Docking Score	DDX3X IC_{50} [a] ±SD (μM)
15		22.05	0.06±0.16
16		26.01	2.2±1.01
17		21.01	5.97±1.24
18		24.86	1.05±0.91
19		27.40	0.17±0.12

Cmpd ID	Structure	Docking Score	DDX3X IC_{50} [a] ±SD (μM)
20		25.23	0.30±0.24
21		29.26	0.06±0.04
22		28.96	0.0005±0.07
23		27.95	n.d.
24		28.68	0.08±0.05

As suggested by docking studies, the replacement of the methyl substituent with small and electron-withdrawing groups such as fluorine is well tolerated in ortho (**7** and **11**, **Figure 27 a** and **b**) and meta positions (**8**, **9**, and **10**, **Figure 27 c, d, e**), while the para-substituted derivative **12** (**Figure 27f**) is about ten-fold less active. In contrast with the sulfonamide series, introduction of the ethoxymethyl side chain was

less tolerated than the other substitutions, with compound **9** being about 100-fold less active. When the butyl side chain on the triazole ring was replaced with other linear and not hindered substituents, such as an isopentyl (**11** and **8**) and a 3-oxobutyl chain (**10**), the resulting compounds showed comparable or slightly decreased activity.

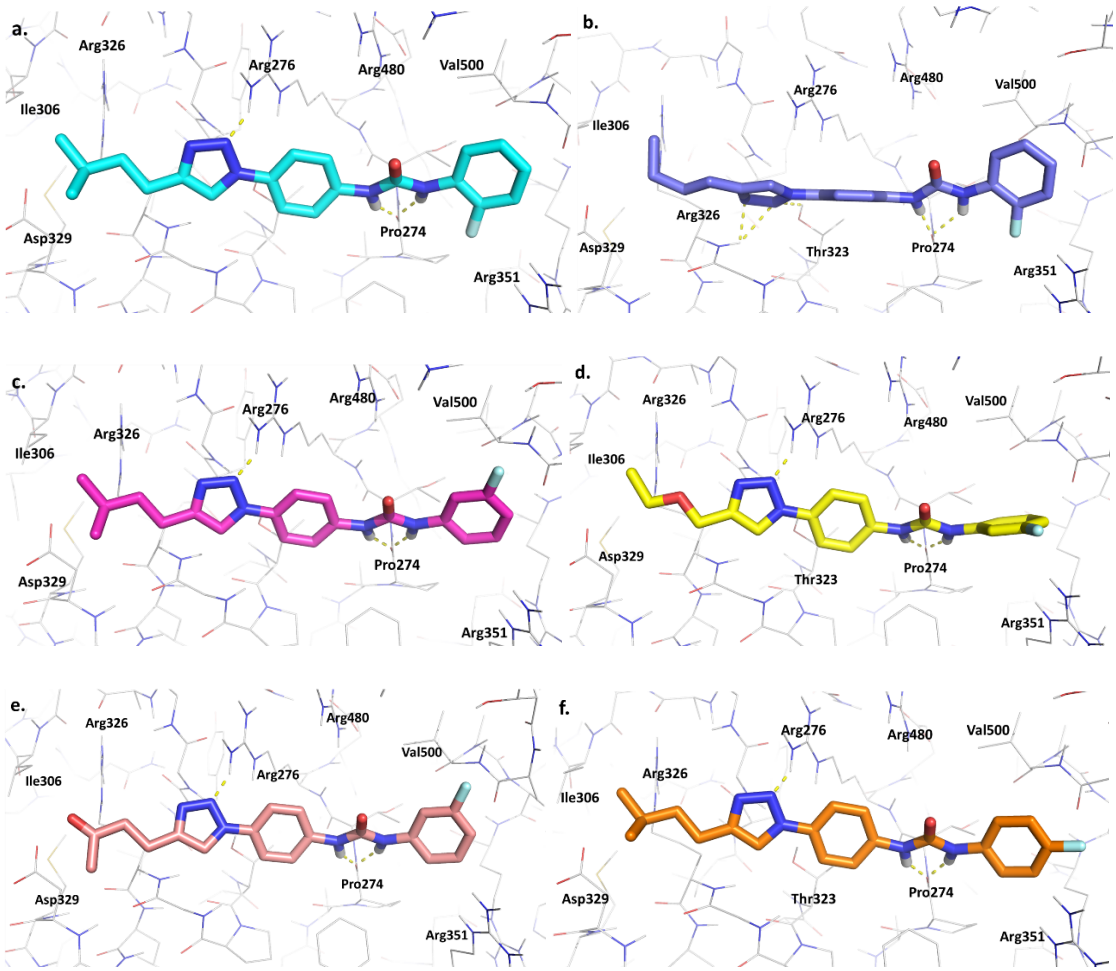


*Figure 27. Binding mode of **7**, **11**, **8**, **9**, **10**, and **12**. (**a.**) Binding mode of **7**. The triazole ring interacts with Arg276, while the ureidic group takes hydrogen bond interactions with the backbone of Pro274. (**b.**) Binding mode of **11**. The triazole ring forms two hydrogen bonds with the backbone of Arg276. The ureidic group interacts with the backbone of Pro274. (**c.**) Binding mode of **8**. The triazole ring forms hydrogen bonds with the side chain of Arg276 while the ureidic group interacts with the backbone of Pro274. (**d.**) Binding mode of **9**. The triazole ring interacts with Arg276 while the ureidic group interacts with Pro274. (**e.**) Binding mode of **10**. The triazole ring forms hydrogen bonds with Arg276, while the ureidic group takes interactions with Pro274. (**f.**) Binding mode of **12**. Compound interacts with Arg276 and Pro274 with its triazole and ureidic group, respectively.*

Replacement of the methyl moiety with the bioisosteric trifluoromethyl group was well tolerated (13, 14, 16, 17, 18, and 20), as already observed within the sulfonamide series. Figure 28 shows the binding mode of the last compounds.

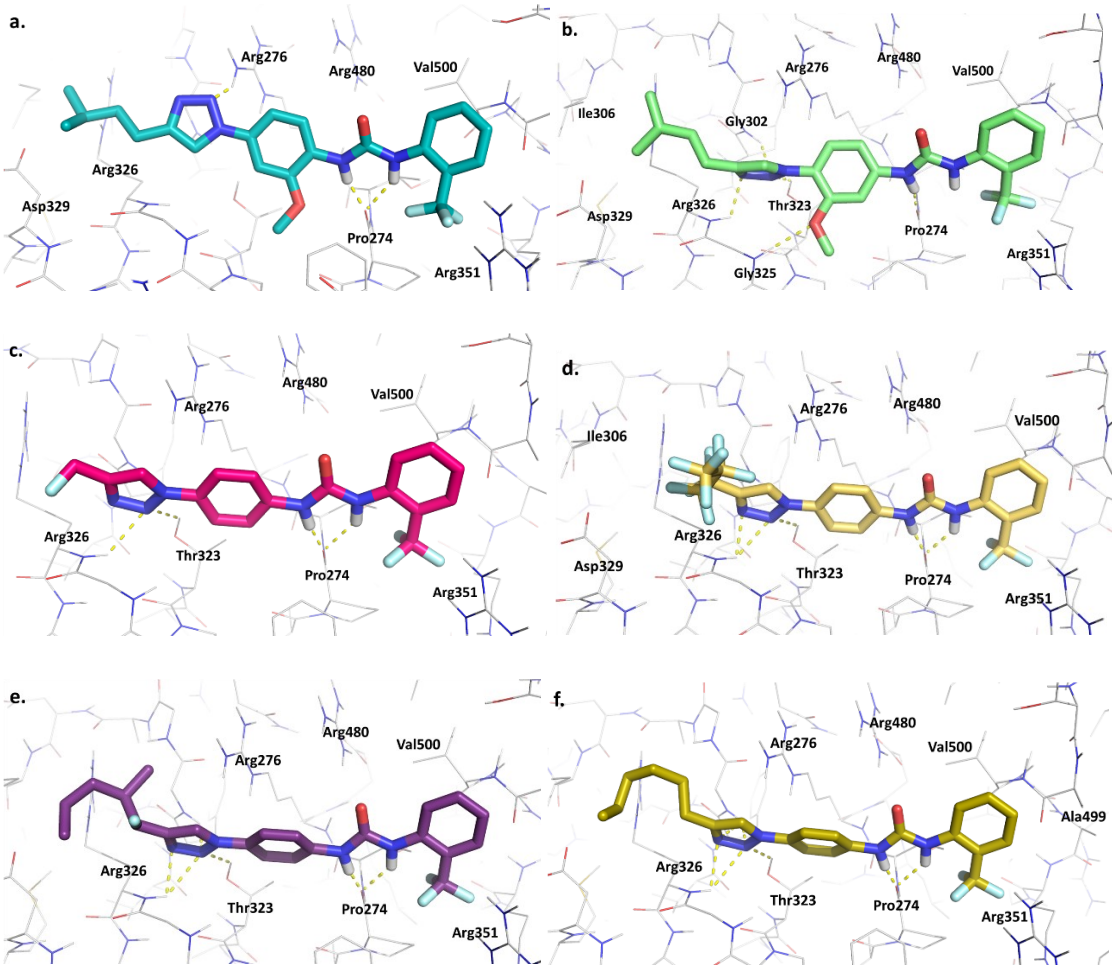


Figure 28. Binding modes of 13, 14, 16, 17, 18, and 20. (a.) Binding mode of 13. The ureidic group of the compound takes interactions with the backbone of Pro274, while the triazole ring binds the Arg276. (b.) Binding mode of 14. The ureidic group interacts with the backbone of Pro274. The methoxy group interacts with the backbone of Gly325. The triazole ring interacts with the backbones of Gly302 and Arg326. (c.) Binding mode of 16. The ureidic group of 14 interacts with the backbone of Pro274, while the triazole ring forms hydrogen bonds with the backbone of Arg326 and the side chain of Thr323. (d.) Binding mode of 17. The ureidic group of the compound forms hydrogen bonds with the backbone of Pro274. The triazole ring interacts with the backbones of Arg326 and Thr323 and with the side chain of Arg276. (e.) Binding mode of 18. The ureidic group of the compound interacts with the backbone of Pro274, while the triazole ring interacts with the backbone of Arg326 and the side chain of Thr323. (f.) Binding mode of 20. The ureidic group of the compound takes interactions with Pro274, while the triazole ring interacts with the backbone of Arg326 and the side chains of Thr323 and Arg276.

Among the trifluoromethyl derivatives, 13, with a methoxy group in the meta position with respect to the triazole ring, was less active than the corresponding ortho

derivative **14**, with **14** being able to form an additional hydrogen bond between the oxygen of the methoxy group and Gly325. Since bulky electron-donating groups were well accepted in the sulfonamide series, [61] an *o*-methylsulfonamido derivative in which ethoxymethyl was replaced with a butyl side chain was synthesized. As a result, **15** showed high inhibitory activity with an IC_{50} value of about 0.1 μM. Compound **16** with a fluoro-methyl substitution at C4 of the triazole was about ten-fold less active than derivatives **19** (**Figure 29**) and **20**, confirming that longer lipophilic side chains are preferred. In contrast, **17** was about 20-fold less active, probably due to the perfluorobutyl side chain, which according to docking analysis is accommodated outside the binding site.

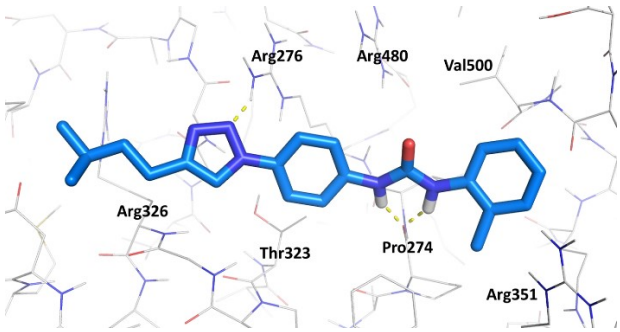


*Figure 29. Binding mode of **19**. The ureidic group of the compound interacts with the backbone of Pro274 while the triazole ring interacts with the side chain of Arg276.*

The sulfonamide series demonstrates promising anti-enzymatic activity, with **21**, **22**, and **24** (**Figure 30 a**, **b**, and **c**, respectively) being characterized by IC_{50} values of 0.06, 0.005, and 0.08 μM. Morpholino derivative **21** (**Figure 30 d**) maintains all the fundamental interactions of the sulfonamide series, particularly a hydrogen bond between triazole and Arg276 and two hydrogen bonds between sulfonamide and Arg480 and Arg276, and establishes an additional hydrogen bond between the morpholine oxygen and Arg351. Ester derivative **22** was extremely active; however, the corresponding carboxylic acid **23** was not evaluated due to its precipitation during the assays. Finally, the meta methylsulfonamido derivative **24** showed an activity value of 0.08 μM, comparable to that of urea compound **15** (**Figure 30 d**).

Even in this case, **24** maintains all the key interactions and forms an additional hydrogen bond between the oxygen of the methyl sulfonamide and Arg351.

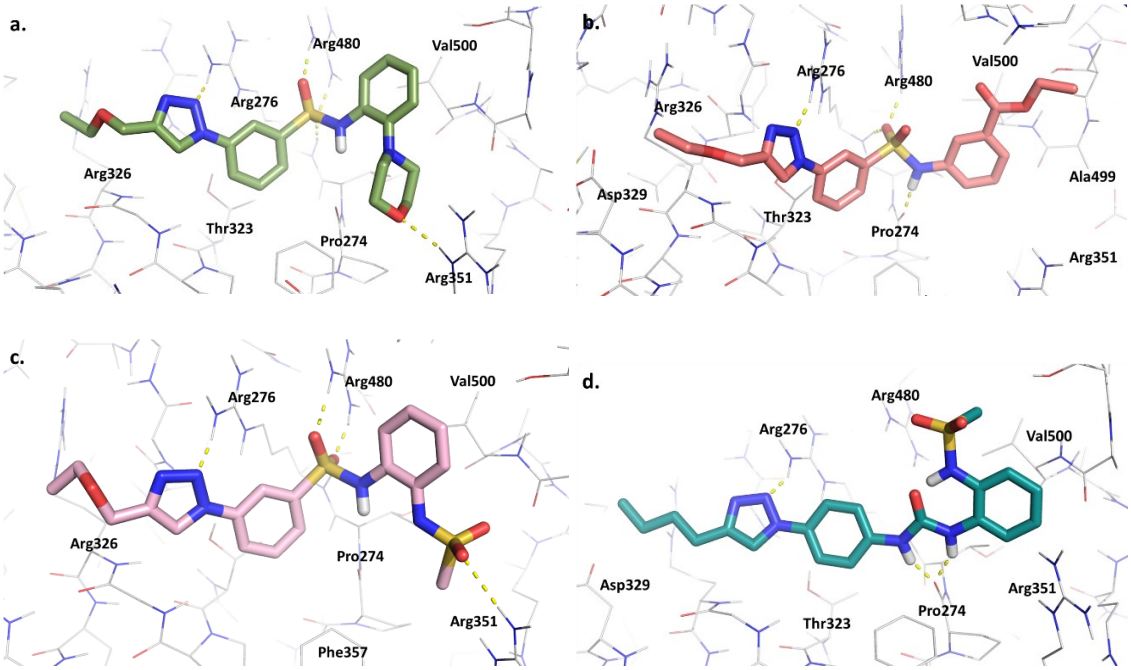


*Figure 30. Binding mode of **21**, **22**, **24**, and **15**. (**a.**) Binding mode of **21**. The oxygen of the morpholine ring interacts with the side chain of Arg351. The sulfonamide establishes hydrogen bonds with Arg480. The triazole ring interacts with the side chain of Arg276. (**b.**) Binding mode of **22**. The sulfonamide interacts with the backbone of Pro274 and the side chain of Arg480. The triazole ring interacts with Arg276. (**c.**) Binding mode of **24**. The two sulfonamide groups interact with the side chains of Arg351 and Arg480. The triazole ring interacts with the side chain of Arg276. (**d.**) Binding mode of **15**. The ureidic group makes interactions with the backbone of Pro274, while the triazole ring interacts with the side chain of Arg276.*

2.3.2 EVALUATION OF THE ANTIVIRAL ACTIVITY AND CYTOTOXICITY

Twelve selected compounds were evaluated for their antiviral activity and cytotoxicity using the immunodetection assay. Their IC_{50} values ranged from the low micromolar to submicromolar and were comparable to or lower than those of ribavirin and sofosbuvir, two broad-spectrum antivirals used as reference compounds.[64] As reported in **Table 8**, **7** shows the best results, with an IC_{50} of 0.9 μM, low cytotoxicity, and a selectivity index of 222. **9** was inactive probably due to its low DDX3X inhibition, while **10** had a very promising IC_{50} value of 0.3 μM and a selectivity index of 64.7. Fluorinated derivative **11** showed a low selectivity index, but it was about four-fold less active than the corresponding isopentyl derivative **8**.

Table 8. Antiviral activity of selected compounds against DENV-2 infected cells.

Cmpd ID	IC_{50}[a]±SD (μM)	CC_{50}[b] (μM)	SI[c]
7	0.9±0.1	200±12	222.2
8	2.5±0.8	200±23	80.0
9	>145	145±18	-
10	0.3±0.1	20±4	64.7
11	10.5±9.1	200±18	19.0
15	2.5±0.2	35±6	14.0
19	1.4±0.8	170±24	121.4
20	2.7±0.1	170±19	62.9
21	5.3±5.0	100±9	18.9
22	4.3±2.3	100±14	23.3
23	>7	7.0±2	-
24	8.3±3.7	87±11	10.5
Ribavirin	4.0±0.6	100±14	25.0
sofosbuvir	3.8±1.1	200±17	52.5

[a] IC_{50}: mean ± standard deviation of half maximal inhibitory concentration calculated in Huh7 cells from at least two experiments; [b] CC_{50}: half-maximal cytotoxic concentration, evaluated on Huh-7 cells; [c] SI: Selectivity index, calculated as the ratio between CC_{50} and IC_{50}.

19 and 20 had favourable antiviral activities and cytotoxicity, comparable to those of the fluorinated compound 8. Sulfonamides 21 and 22, despite their potent DDX3X inhibitory concentrations reported in Table 7, showed activities of 5.3 and 4.3 μM, respectively. The low activity of 22 is probably due to its cellular hydrolysis into the corresponding carboxylic acid 23, which was toxic and inactive. Sulfonamide 24 and urea 15 showed IC_{50} of 2.5 and 8.3 μM, respectively, and their selectivity indexes were lower than those of the other compounds of the series.

2.3.3 MODE OF ACTION OF DDX3X INHIBITORS IN DENV INFECTED CELLS

To better understand the mode of action of our compounds in DENV infected cells, a fluorescent inhibitor was designed and synthetized to be used as a probe in time-course confocal microscopy experiments. , inserting the FITC on the C4 triazole side chain, obtaining compound **25** (**Figure 31**).

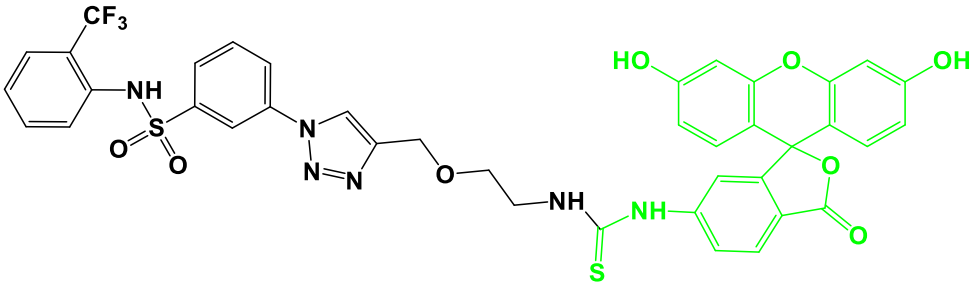


Figure 31. Structure of the fluorescent inhibitor 25.

Compound **25** was tested for its ability to inhibit DENV-2 replication in Huh-7 cells. Results showed that its IC_{50} was 28.2 ± 4.6 μM while its CC_{50} was 140.0 μM. The time-course localization of the fluorescent probe was studied by investigating its colocalization with DDX3X and viral protein NS5, discovering that DDX3X is primarily localized into the cytoplasm, while during viral infections, DDX3X is recruited to perinuclear spots, particularly between 6 and 24 hours post-infection (**Figure 32** and **Figure 33**).

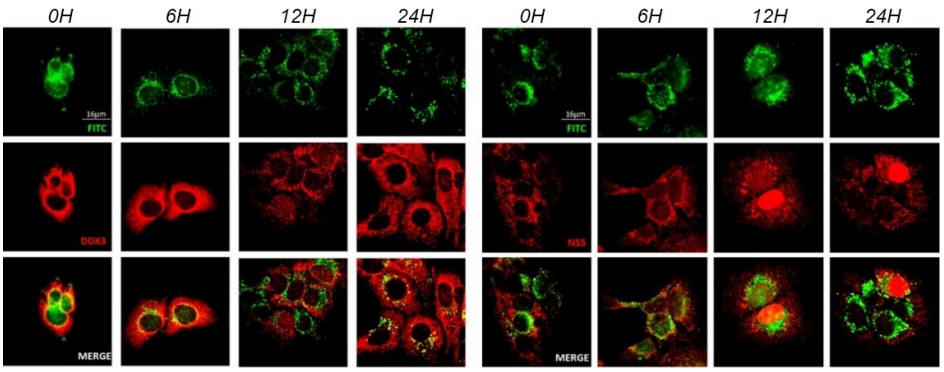


Figure 32. Immunofluorescence analysis. DENV-2 infected cells were treated with DDX3X fluorescent inhibitor 25 at a concentration of 20 μM. Cells were stained at different time-points with a DDX3 antibody (detected using Alexa Fluor 568-labeled secondary antibody, left panels) and DENV NS5 antibody (detected using Alexa Fluor

568-labeled secondary antibody, right panels) and examined by confocal microscopy. Individual antibody stained, as well as merged images, are shown as indicated. Each experiment was repeated at least two times.

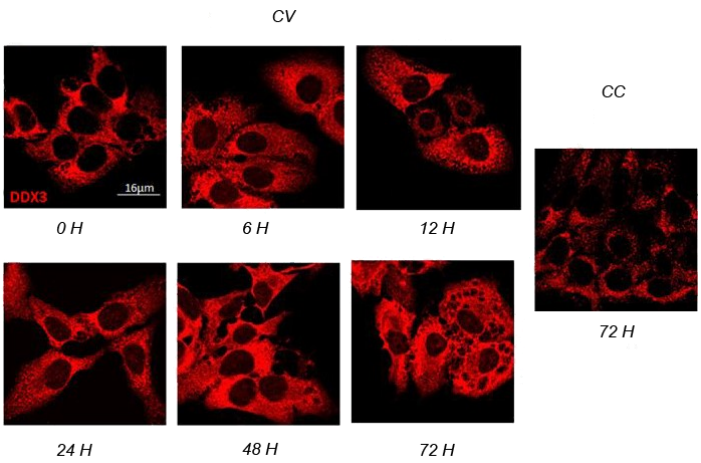


Figure 33. DDX3X expression and localization in uninfected Huh-7 cells (CC) and in DENV-2 infected Huh-7 cells (CV).

The recruitment of DDX3X to these structures could be a potential shared strategy employed by RNA viruses, such as HCV and WNV, that exploit the DDX3X function. [65, 66, 67] This study highlight that DDX3X increases its expression during infection, with a maximum expression at 48 and 72 hours (CV panels in **Figure 33**). **25** colocalized with DDX3X at the perinuclear region in the first 6 hours of treatment (**Figure 32** left panels) and then induced DDX3X cytoplasmatic localization as in uninfected Huh7 cells (**Figure 33**, **CC** panel). The same experiment was conducted to analyse the cellular localization of the viral protein NS5, discovering that in the first 6 hours, NS5 had a cytoplasmatic localization and was colocalized with **25** and then moved in the nucleus. As shown in **Figure 34** and **Figure 35**, **25** was associated with an increased number of viable cells and with a significant reduction of the total number of NS5 positive cells at 48 and 72 hours.

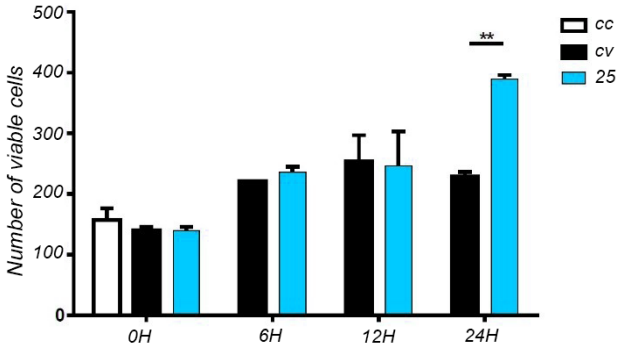


*Figure 34. Effect of **25** on the number of viable cells. Cells were stained with DAPI and counted at a magnification of 40x in 5 random fields per well. **25** did not affect cell viability in the first 12 h but showed a significant protective effect at 24 h. White bar, (CC) uninfected cells. Black bars, (CV) cells infected with DENV-2 (100 $TCID_{50}$). Green bars, (25) cells infected with DENV-2 (100 $TCID_{50}$) were treated with **25**.*

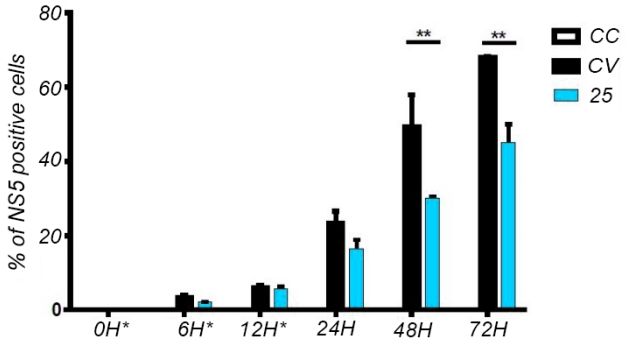


*Figure 35. Effect of **25** on the expression of NS5 viral protein. NS5 expression was calculated in cells counted at a magnification of 40x in 5 random fields per well. NS5 was expressed in the cytoplasm. **25** did not affect cell viability in the first 12 h but showed a significant protective effect at 24 h. White bar, (CC) uninfected cells. Black bars, (CV) cells infected with DENV-2 (100 $TCID_{50}$). Green bars, (**25**) cells infected with DENV-2 (100 $TCID_{50}$) were treated with **25**.*

2.3.4 SUMMARY AND CONCLUSIONS

In this study, the attention was focused on expanding the SAR considerations around the two series of already discovered DDX3X inhibitors, concentrating on searching novel promising compounds active against DENV-2 infection. Novel DDX3X helicase inhibitors with improved antiviral activity were discovered, and their activity is comparable to or lower than those reported for known broad-spectrum antivirals such as ribavirin or sofosbuvir. Remarkably, the most promising derivative, compound **10**, is about 9-fold more active than **16d**.

The mechanism of action of the designed compounds was evaluated in infected cells using the fluorescent derivative **25**. Immunofluorescence analysis confirms that **25**,

during the first hours of DENV infection, colocalized with DDX3X, promoting the reduction of NS5 positive cells and recovering the cell number over time (until 72 hours).

The low cytotoxicity of compounds, evaluated by measuring ATP concentration, indicates that they are characterized by high cellular tolerability.

These results confirm that DDX3X inhibitors are a safe and promising antivirals class, supporting their evaluation in an animal model of DENV infection. Moreover, the development of DDX3X inhibitors offers the advantage of fighting different viruses, including novel emerging ones, with a unique molecule, reducing the risk of developing drug resistance since human genes coding for proteins involved in viral replication are minimally vulnerable to mutations induced by drug administration.

2.4 Materials and Methods

Ligand Preparation

Ligands were designed by means of Maestro (version 11.2.014) graphical interface [68] and pre-treated with LigPrep (version 2014:4) [69] tool at default settings, checking the ionization state at physiological pH with Epik. [70, 71, 72]

Docking Studies

The docking procedure was performed using Gold, version 5.2. [59] All poses were subjected to a consensus scoring procedure, choosing Chemscore as fitness function and Goldscore as rescoring fitness function. The pocket under investigation was inserted into a grid box centered on Arg276, enclosing residues lying within 10 Å from such amino acid. The genetic algorithm parameter settings were employed using a 100% search efficiency, and 100 runs were carried out for each ligand. The first ranked solution of each ligand was selected for further analysis.

ADME Prediction

QikProp [60] is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches. ADME molecular properties of all selected compounds were predicted with Qikprop, using the Fast mode option. Predicted aqueous solubility (QPlogS), where S is expressed in mol dm^{-3} and represents the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid. The recommended range is -6.5 to -0.5.

Pictures

Pictures of the modeled ligand-enzyme complexes, together with graphic manipulations, were rendered with the PyMOL package [73] (version 1.8.4.0, http://www.pymol.org/).

CHAPTER 3

UNIQUE MOTIF: RESULTS AND DISCUSSION

DEAD-box proteins of human, insect, yeast, and archaeal are constituted by an insertion of about 30 amino acids located between motifs I and Ia (Figure 36). Among all eukaryotic DEAD-box proteins, only three members appear to have longer insertions: DDX3X (42 aa), DDX42 (150 aa), and DDX1 (240 aa).

Sequence alignment of this region of DDX3X and other DEAD-box proteins revealed a high degree of identity for the sequences immediately flanking the motifs I and Ia, while the central part is more variable.

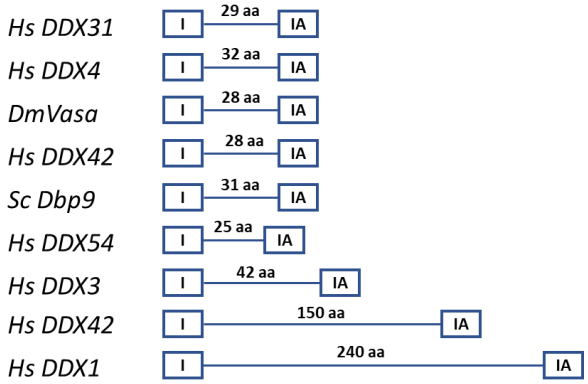


Figure 36. Schematic representation of the length of the insertion sequences between motifs I and Ia of DDX3 and different DEAD-box proteins. The number of amino acids in each insertion is indicated.

Sequence analysis also showed that the human proteins present an invariant -[DG]- motif in common with DDX3X, while proteins with a more distant phylogenetic relationship (*Drosophila* Vasa, *T. thermophilus* Hera, *S. cerevisiae* Dbp9) showed a -[ExG]- conserved box. Moreover, the Vasa and Hera proteins also have in common with DDX3X an additional basic amino acid -[ExGR]- in this box, while DDX3X is unique in having an additional positively charged extension -[YGRRK]-.

The central 10 residues of the DDX3X insertion (aa 250–259) are important for the ATPase activity, nucleic acid binding and unwinding and are involved in HIV-1 RNA binding. A specific peptide ligand to this region reduces the ability of DDX3X to support HIV-1 replication in infected cells. [74]

The unique insertion of DDX3X is called "unique motif" (sequence ALRAMKENGRYGRRK, aa 250–264) and is present in the homology model of the closed conformation of DDX3X.

Using pocketPicker, [75] a Pymol [73] plugin, a little pocket has been identified and used to perform a preliminary screening to search theoretically active scaffolds (Figure 37).

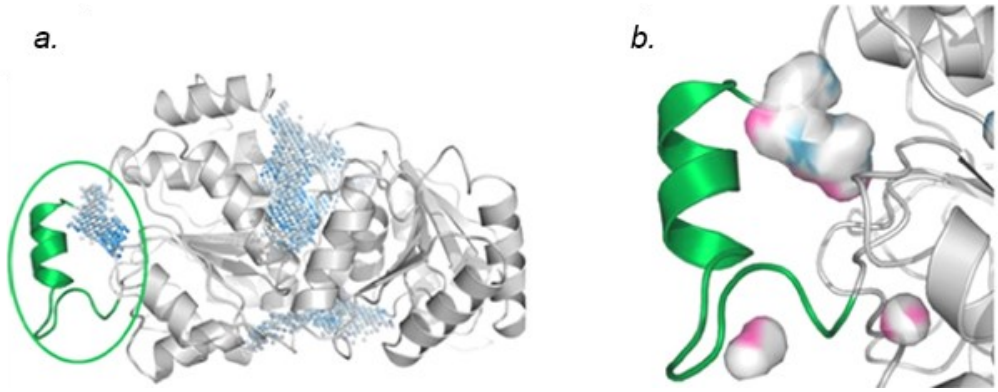
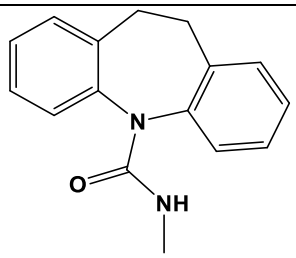
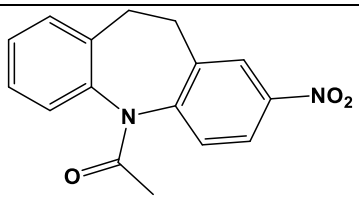
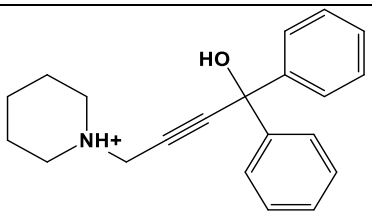
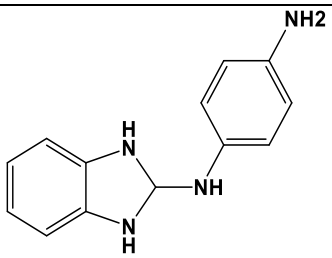
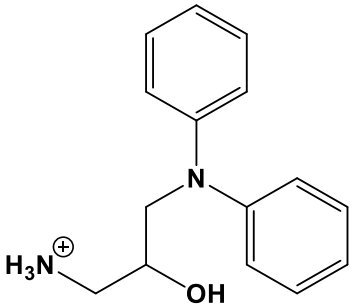


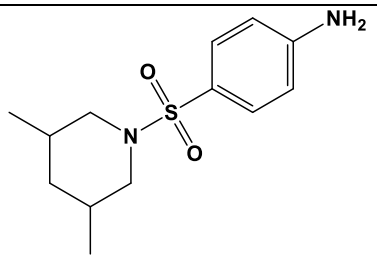
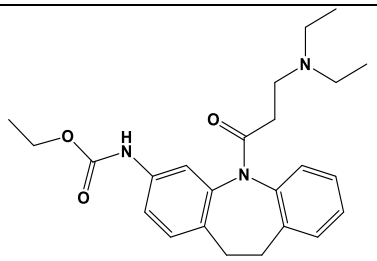
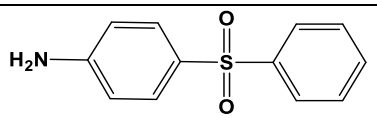
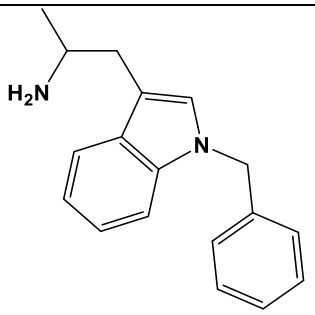
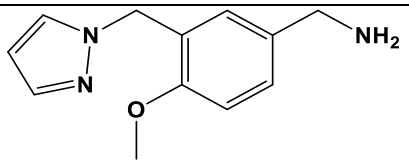
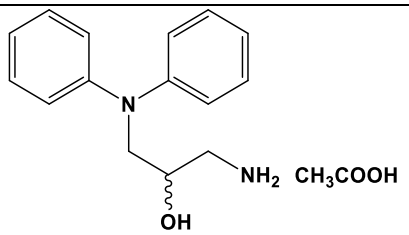
*Figure 37. Molecular modeling of the inhibitor-binding pocket. **(a.)** Pockets identified by PocketPicker are represented in dots. The unique motif UM (cyan) with the adjacent pocket is circled. **(b.)** Zoomed-in view of the little pocket around the unique motif (cyan).*

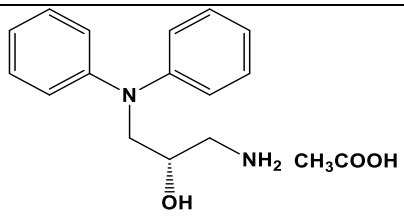
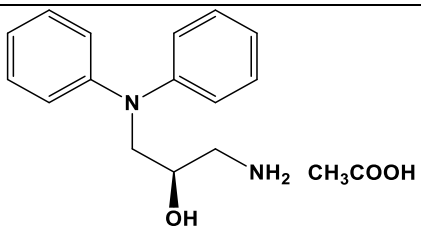
3.1 THE UNIQUE MOTIF TO DEVELOP SELECTIVE INHIBITORS OF DDX3X

The *in silico* preliminary screening was performed using two commercial libraries (Asinex Gold and Platinum consisting of 583 040 molecules) and allowed to obtain ten compounds that were selected based on their chemical structures and their polar interactions into the active site. The molecules were tested to evaluate their ability to inhibit DDX3X RNA helicase activity in enzymatic assays. Selected compounds, their structures, docking scores, and anti-enzymatic activities are reported in Table 9.

Table 9. Inhibitory activity and docking score of selected compounds targeting the UM of DDX3X. [a] ID_{50}, inhibitor concentration reducing by 50% the enzyme activity. Values are the mean of three independent replicates ± SD.

Cmpd ID	Chemical Structure	Activity [a] (μM)	Docking Score (ChemScore)	QPlogS
1		n.a	25.22	-3.52
2		n.a.	24.76	-3.85
3		n.a.	25.15	-4.28
4		n.a.	24.83	-2.90
5		0.7±0.2	24.68	-2.13

Cmpd ID	Chemical Structure	Activity [a] (μM)	Docking Score (ChemScore)	QPlogS
6		n.a.	25.82	-3.36
7		n.a.	27.27	-5.33
8		200±30	24.94	-2.42
9		n.a.	25.68	-3.44
10		n.a.	25.86	-1.63
11		0.002±0.0005	24.68	-2.13

Cmpd ID	Chemical Structure	Activity [a] (μM)	Docking Score (ChemScore)	QPlogS
12		0.007±0.0001	24.28	-2.13
13		0.006±0.0001	21.65	-2.13

The analysis of the binding mode of these compounds allowed to formulate hypotheses on the structural features that are important for their activity. **8** makes polar interactions with Arg262 and Pro257 (**Figure 38a**) while **10** interacts with His318, Arg262, Glu256 and Cys317 (**Figure 38b**).

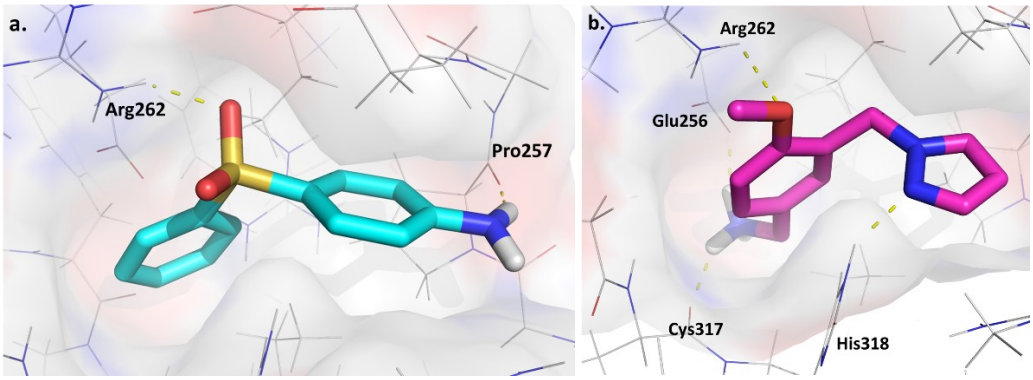


*Figure 38. (a.) Binding mode of **8**. Compound interacts with Arg262 and Pro257. (b.) Binding mode of **10**. The methoxy group interacts with the side chain of Arg262. The pyrazole ring interacts with His318. The terminal amino group interacts with Glu256 and Cys317.*

1, **2**, **7** are involved in hydrophobic interactions with Arg262, and Glu249 (**Figure 39 a,b,c,** respectively).The phenyl ring of **1** establish pi-pi stacking interactions with

His318 while the carbonyl oxygen is involved in hydrogen bond interactions with Arg262. The nitrobenzene of **2** is involved in pi-pi stacking interactions with His318 while the carbonyl oxygen interacts with the side chain of Arg262.

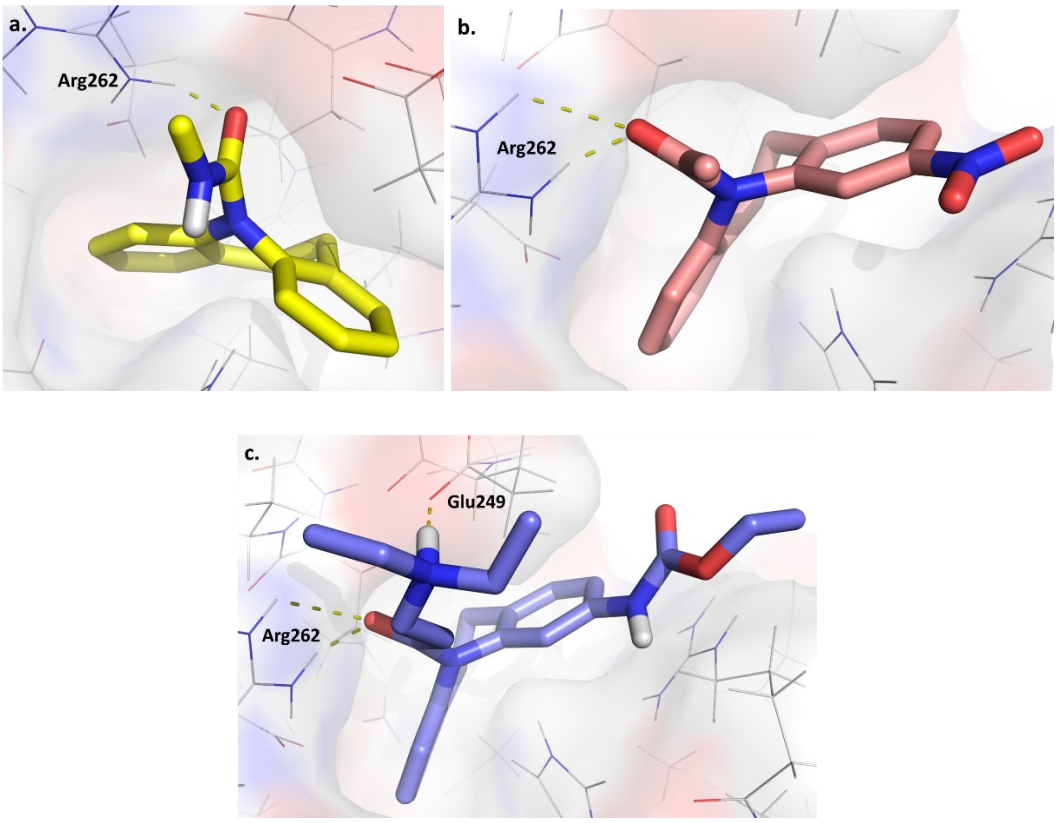


*Figure 39. (a.) Binding mode of **1**. The compound interacts with Arg262 while both the phenyl rings establish pi-pi interactions with His318. (b.) Binding mode of **2**. The carbonyl oxygen of 2 forms hydrogen bonds with Arg262, while the nitrobenzene is involved in pi-pi interactions with His318. (c.) Binding mode of **7**. The compound interacts with Arg262 and Glu249.*

5 is characterized by an interesting shape that fits into the pocket like a "cap" and is the only one that presents a hydroxyl group able to interact with Glu256 and Cys317(**Figure 40**). The compound establishes polar interactions with Glu256, Cys317, Tyr266, and hydrophobic interactions with Arg262 and His318.

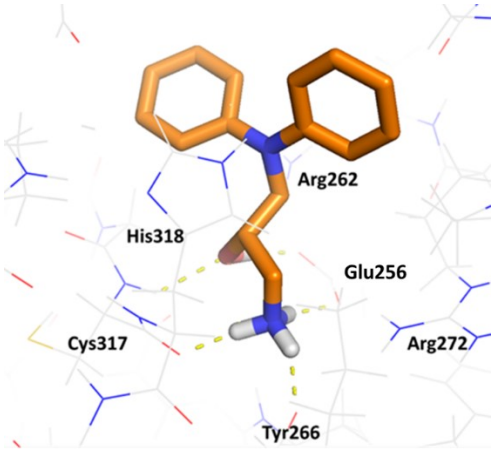


Figure 40. Binding mode of 5. Compound interacts with Cys317, Tyr266, and Glu266.

Even though a high docking score characterizes 1 and 7, the molecules were found inactive. The analysis of the binding mode of these compounds is reported in Figure 39a and c, and shows that these molecules are not able to interact with Cys317, Tyr266, and Glu256. Conversely, these interactions characterize the binding mode of 5, 11, and 13 and are probably responsible for their activity. To confirm the activity of 5, the compound was resynthesized and converted into the water-soluble acetate salt 11 (Table 9). Moreover, the R and S enantiomers 12 and 13 (Table 9) were synthesized. To experimentally validate these findings and understand the role of Glu256 in the binding of 11, 12, and 13, they were further investigated.

3.1.1 THE IMPORTANCE OF GLU256 IN THE RNA HELICASE ACTIVITY

The molecular modelling predictions were confirmed by mutagenesis studies, where the effect of the variation of Glu256 and Arg262 of DDX3X with Alanine were evaluated. Two single E256A and R262A mutants, together with the double E256A/R262A mutant, were tested for their RNA helicase activities in a FRET-based assay to evaluate the impact of these mutations on the activity of the enzyme, and the kinetic parameters for the reaction are reported in Table 10.

Table 10. Kinetic parameters for the RNA helicase activity of DDX3X wild type and the DDX3X(E256A), DDX3X(R262A), and DDX3X(E256A/R262A) mutants.

Enzyme	K_m(RNA)[a] (μM)	K_{cat}[a] (FU[b]/min)	k_{cat}/K_M (μM^{-1}, FU/min)	fold reduction[c]
Wild type	0.5±0.1	1146±22	2278	1
E256A	0.6±0.2	126±10	210	10.8
R262A	9.0±2.0	4794±62	522	4.4
E256A/R262A	0.6±0.25	148±70	307	7.4

[a] Kinetic parameters K_m and k_{cat} represent the mean of three independent experiments ± S.D; [b] FU, arbitrary fluorescence emission units; [c] Ratio k_{cat}/K_M(wt)/k_{cat}/K_M(mut).

The E256A mutation causes a 10.8-fold reduction of the k_{cat}/K_M value for RNA unwinding with respect to wild-type DDX3X, while the R262A mutation also reduced the k_{cat}/K_M value by 4.4-fold with respect to the wild-type enzyme. The double mutation did not result in an additive effect, showing an intermediate phenotype (7.4-fold reduction in the k_{cat}/K_M value with respect to the wild type), suggesting that the effect of the E256A mutation did not change by the presence of the additional R262A substitution. The comparison of the single mutants for the ATPase activities showed that the E256A mutant was characterised by comparable activity with respect to the wild-type enzyme, while the ATPase activity of the R262A mutant was reduced, similarly to the double E256A/R262A mutant. These studies demonstrate that Glu256 is essential for the RNA helicase but not for the ATPase activity of DDX3X, while Arg262 plays an auxiliary role, coupling ATP hydrolysis to RNA unwinding.

3.1.2 GLU256 IS RESPONSIBLE OF THE INTERACTION OF THE UM COMPOUNDS WITH DDX3X

11 and **13**, designed based on the *in silico* docking studies to interact with the mutated residues, were tested for their ability to inhibit the RNA helicase activity of DDX3X wild-type and mutant proteins. Even though the racemic compound **11** and its two enantiomers (**12** and **13**) showed comparable activities, as reported in **Table**

9, 11 was selected for further analysis. In Table 11 is shown that the E256A substitution caused a 36500- and 43-fold reduction of the inhibitory potencies of 11 and 13, respectively, while the mutation R262A did not significantly change the inhibitory potencies of the compounds with respect to the wild-type enzyme. Thus, Glu256 seems to be the major determinant for inhibitor binding.

Table 11. Inhibition of the RNA helicase activity of DDX3X wild type and the DDX3X(E256A) and DDX3X(R262A) mutants.

Cmpd ID	wild type ID_{50}, (μM) [a]	E256A ID_{50}, (μM) (fold resistance)[b]	R262A ID_{50}, (μM) (fold resistance)
11	0.002±0.0005	73±10 (36500)	0.001±0.0003 (0.5)
13	0.008±0.0001	0.3±0.1 (43)	0.004±0.002 (0.6)

[a] ID_{50}, concentration of the compounds giving 50% inhibition of the activity measured for the respective enzyme in the absence of an inhibitor. Values were calculated as described in the Experimental Section and represent the mean of three independent experiments ±S.D.; [b] Ratio ID_{50}(mut)/ID_{50}(wt).

These results agree with the molecular modeling and the kinetic analysis (Table 2), indicating that Glu256 establishes key interactions with the RNA, and it is possible to hypothesise that the binding of the inhibitor to this residue can disrupt the interaction with the RNA.

3.1.3 SELECTIVITY AND ANTIVIRAL ACTIVITY OF UM COMPOUNDS

11 and 13 were evaluated also for their specificity of inhibition DDX3X. The inhibition assays were performed with human DDX1, plant STRS2 from *Arabidopsis thaliana*, and the DExH-family viral NS3 RNA helicase of HCV.

In Table 12, it is possible to see that all the compounds tested inhibited DDX1 with potencies almost four orders of magnitude lower than toward DDX3X (selectivity ≈9000-fold) while they were completely inactive against STRS2 and NS3, confirming the high selectivity of inhibition for DDX3X.

Table 12. Inhibition of the RNA helicase activity of human DDX1, Arabidopsis STRS2, and HCV NS3.

Cmpd ID	DDX1 ID_{50}, (μM)[a] (-fold selectivity)[b]	STRS2 ID_{50}, (μM) (-fold selectivity)	NS3 ID_{50}, (μM) (-fold selectivity)
11	18±1.202 (9000)	>100 (>50000)	>100 (>50000)
13	58±4.5 (8285)	>100 (>10000)	>100 (>10000)

[a] ID_{50}, 50% inhibitory concentration. Values represent the mean of three independent experiments ±S.D.; [b] Ratio ID_{50}/ID_{50} (wt DDX3X) and the ID_{50} (wt DDX3X) values used were those reported in Table 9.

11, 12, and 13 were also tested for their ability to reduce WNV and DENV replication on Huh-7 cells. As reported in Table 13, 11 and its pure enantiomers 12 and 13 showed promising antiviral activity in the micromolar range, resulting in IC_{50} values comparable to (DENV) or significantly lower (WNV) than those shown by the broad-spectrum antiviral ribavirin.

Table 13. Antiviral activity and cytotoxicity of selected compounds against DENV and WN. Data represent mean ± S.D. of three experiments.

Cmpd ID	IC_{50}[a] (DENV) (μM)	IC_{50}[a] (WNV) (μM)	CC_{50}[b,c] (μM)	CC_{50}[b,c] (μM)
11	7.9±3.3	2.3±0.6	100±11.3	>100
12	12.6±5.4	0.90±0.20	120±18.0	>100
13	6.3±5.8	0.88±0.10	79±11.8	>100
Ribavirin	4.0±0.6	91.5±5	>100	>100

[a] IC_{50}: half-maximal inhibitory concentration; [b] CC_{50}: half-maximal cytotoxic concentration; [c] Calculated using the CellTiter-Glo kit; [c] Calculated using the MTT kit; ribavirin was used as a reference compound.

3.1.4 DISCUSSION

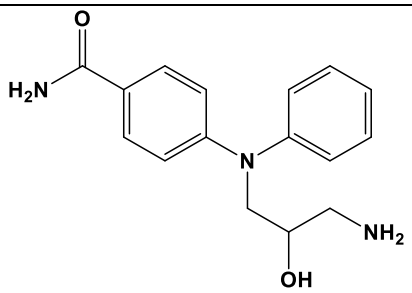
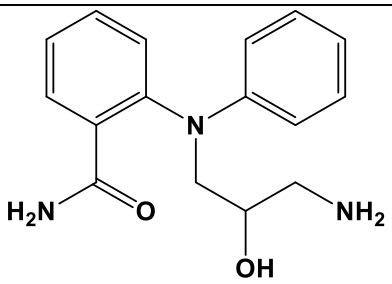
The homology model of the closed conformation of human DDX3X has been used to target the unique structural domain of the protein, allowing to discover the first class of inhibitors of the human helicase DDX3X that have been rationally designed

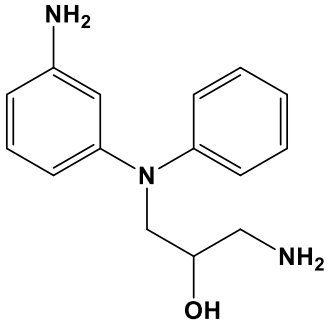
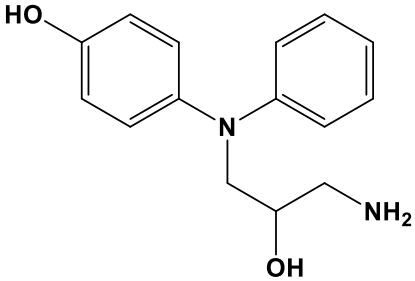
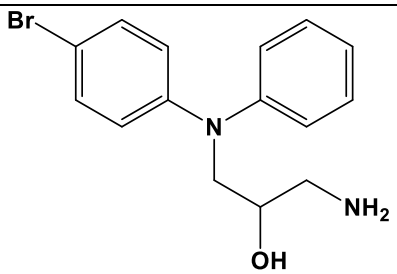
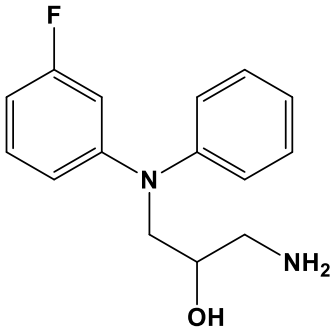
to recognize the unique motif (sequence ALRAMKENGRYGRRK, aa 250-264) of DDX3X. Molecules showed no toxicity in cells and a considerable antiviral effect, suppressing the replication of WNV and DENV-2 viruses in infected cells. Since DDX3X is involved in the replication cycle of several viruses, it is possible to hypothesize that the unique domain of DDX3X could be used as a target for broad-spectrum antiviral drugs endowed with maximal selectivity. **5** was used to design a library of derivatives able to target this domain, to expand SAR studies and fully characterize their mode of interaction. Moreover, a more accurate 3D pharmacophore-based screening was performed to find new molecules to purchase and submit to biological assays.

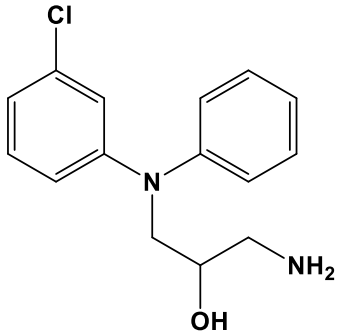
3.2 DESIGN OF DERIVATIVES OF COMPOUND 5

A first strategy to find new molecules able to bind the UM pocket of DDX3X and with an increased molecular weight has been focused on modifying the scaffold of **5**. The first series of compounds is characterized by small changes that have been reported on one of the phenyl rings. All molecules were docked into the UM binding site and have been visually analysed to discard molecules that were docked into the binding site or that were not able to retain the main interaction of **5**. Selected compounds are reported in **Table 14** together with their molecular weights (MW), docking scores and QPlogS values calculated using QikProp. [60]

*Table 14. First series of derivatives of **5**. Molecules have been selected based on their binding mode and their ability to retain the interactions of **5**.*

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14		285.35	21.24	-1.04
15		285.35	22.30	-0.98

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
16		257.34	22.30	-0.99
17		258.32	21.97	-1.25
18		321.22	22.58	-2.29
19		260.31	21.70	-2.24

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
20		276.76	21.87	-2.34

The benzamide ring of **14** interacts with Gly316. The -OH group interacts with Glu256 while the NH_2 is involved in hydrogen bond interactions with Glu256 and the side chain of Cys317. The hydroxy and the ammine group of **15** gave the same interactions of **14**. The amide group makes interactions with Glu249 and Arg262. The aniline moiety of **16** is involved in hydrogen bond interactions with Glu249. The hydroxyl group interacts with Glu256, while the terminal ammine interacts with the side chain of Glu256 and the backbone of Cys317. **17**, **21,** and **22** are well accommodated into the binding pocket and establish interactions with Glu256 and the side chain of Cys317. **20** interacts with Glu256, Cys317, and the side chain of Gln265.

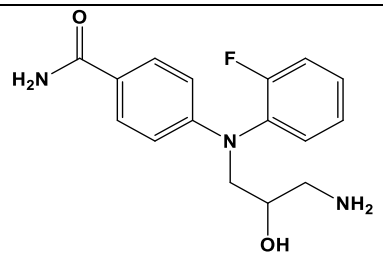
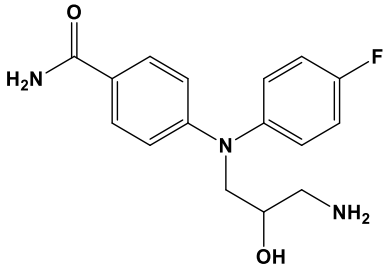
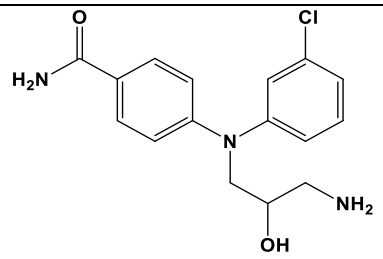
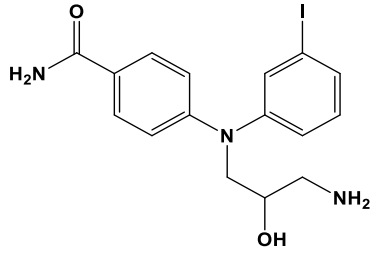
These compounds represent the starting point for the design of second-generation derivatives in which modifications have been made on the second phenyl ring. Molecules were docked into the UM pocket, and only those correctly docked into the binding site and able to retain the fundamental interactions of **5** were retained.

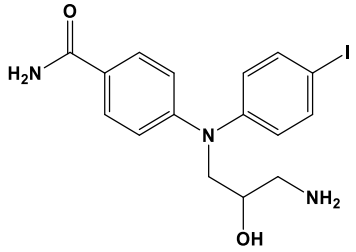
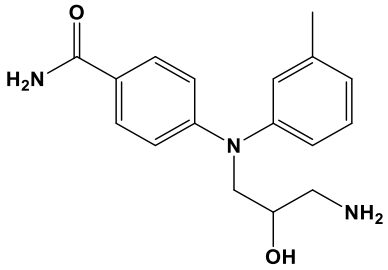
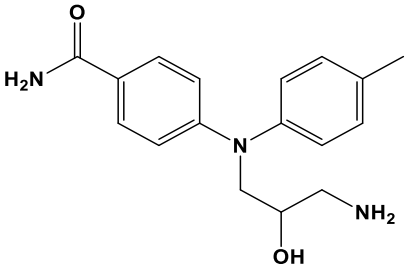
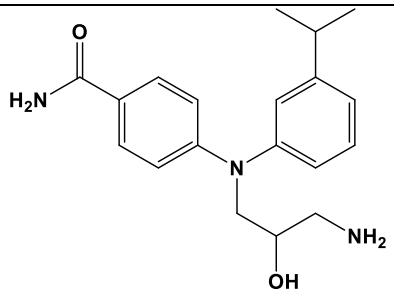
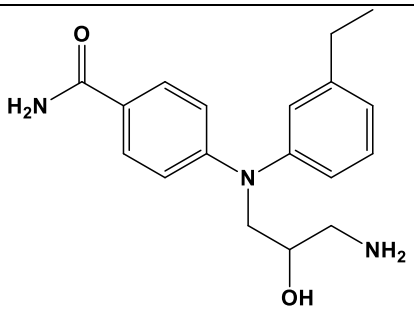
3.2.1 DERIVATIVES OF 14

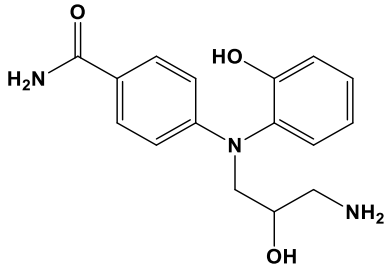
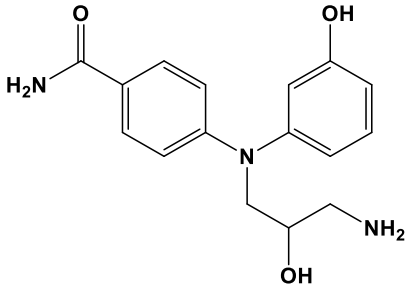
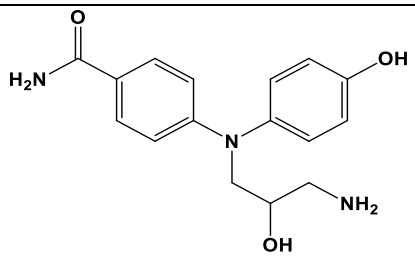
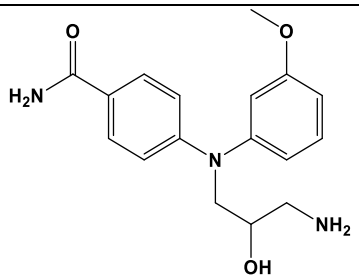
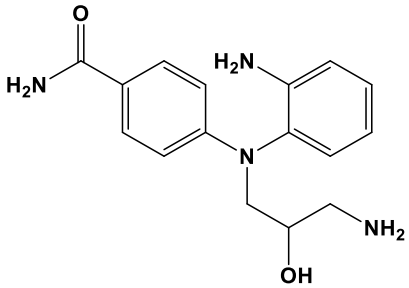
Derivatives of **14** are reported in table 16. The introduction of a halogen, a methyl, or an isopropyl substituent on the second phenyl ring is well tolerated, and the benzamide ring establishes interactions with the backbone of Arg315. The only exceptions are compounds **14b** and **14h** that lost this interaction, probably due to

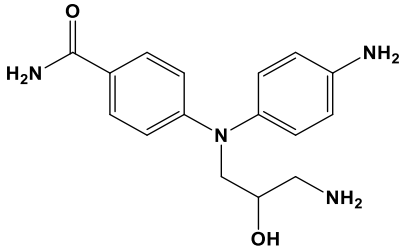
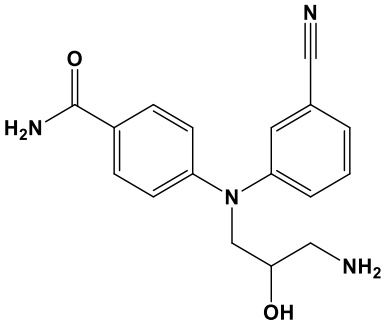
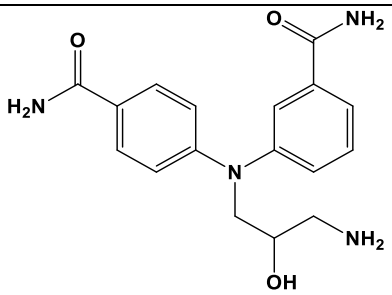
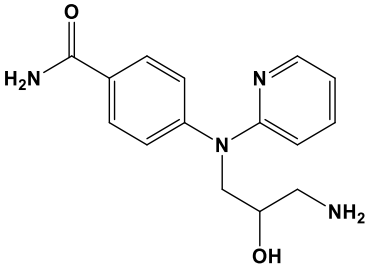
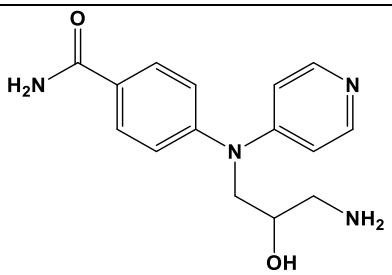
the high steric hindrance of the substituents. The introduction of a phenolic OH substituent in ortho and para positions (**14j**, **14l**) allows the interaction with Glu249, while the benzamide interacts with Arg315. The introduction of a -NH_2 group in ortho and para positions (**14n**, **14o**) allows to establish the same interactions. The substitution of the phenyl ring with a pyridine (**14s**), 2,3-diazabenzene (**14t**), and 2,5-diazabenzene (**14y**) allow gaining an interaction with His318.

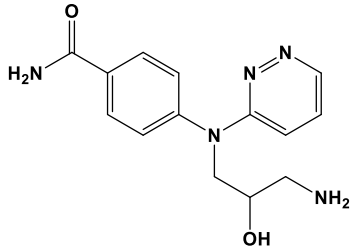
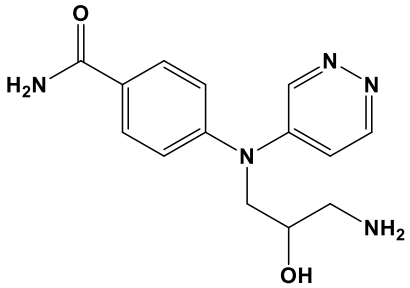
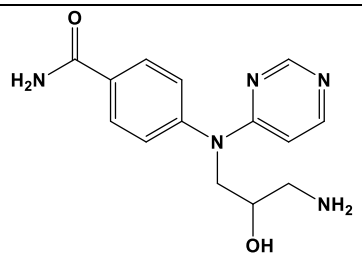
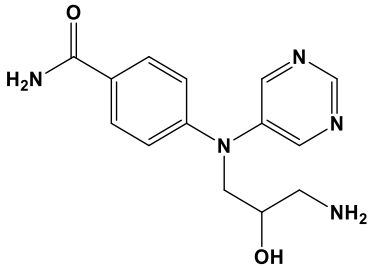
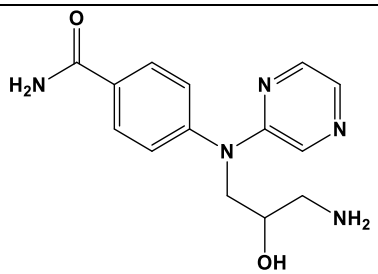
Table 15. Derivatives of 14. Molecular weight, docking score values, and the predicted solubility are reported.

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14a		303.34	21.81	-1.76
14b		303.34	21.00	-1.77
14c		319.79	24.34	-1.87
14d		411.24	22.04	-2.20

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14e		411.24	22.51	-2.06
14f		299.37	23.20	-2.04
14g		299.37	24.50	-1.82
14h		327.43	20.70	-2.18
14i		313.40	21.65	-1.67

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14j		301.35	19.97	-1.11
14k		301.35	23.07	-1.03
14l		301.35	24.57	-1.23
14m		315.37	21.49	-1.55
14n		300.36	22.95	0.95

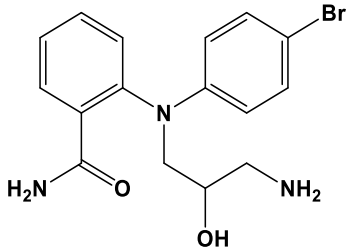
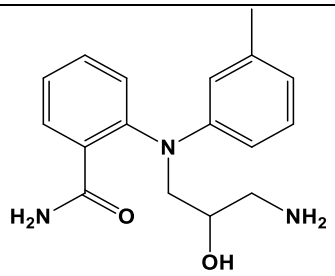
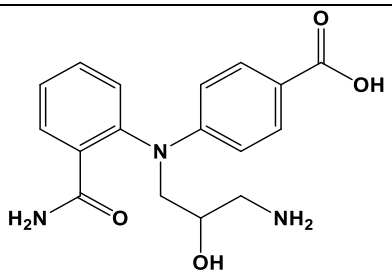
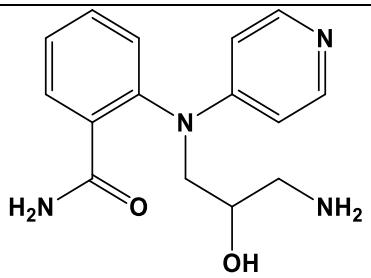
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14o		310.36	27.15	-1.10
14p		328.37	22.40	-2.65
14q		386.33	21.94	-1.59
14r		287.32	22.98	-1.41
14s		287.32	21.46	-0.92

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
14t		287.32	22.82	-1.07
14u		287.32	21.29	-1.08
14v		287.32	19.96	-1.06
14x		287.32	18.65	-0.76
14y		287.32	20.84	-0.96

3.2.2 DERIVATIVES OF 15

The introduction of a bromine atom in **15a** or a methyl group in **15b** is well tolerated. These compounds can establish a hydrogen bond interaction with the side chain of Glu249. Compound **15c** is not able to establish further interactions with the UM binding pocket. The amide of compound **15d** interacts with Arg262 and Glu249.

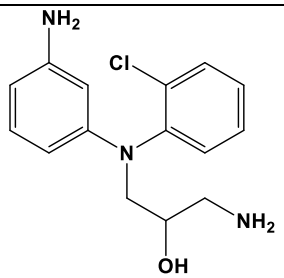
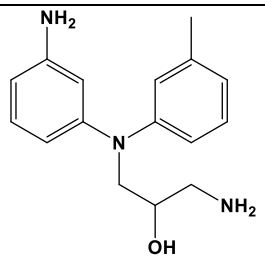
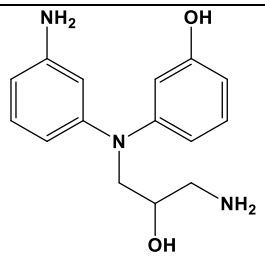
Table 16. Derivatives of 15. Molecular weight, docking score, and the predicted logS of the compounds are reported.

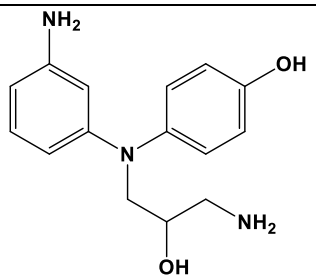
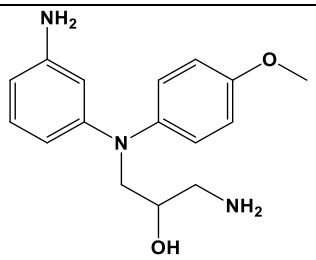
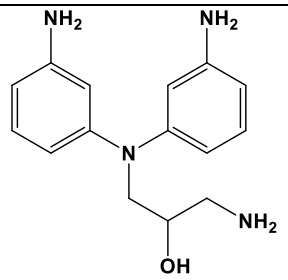
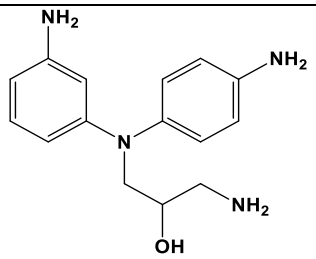
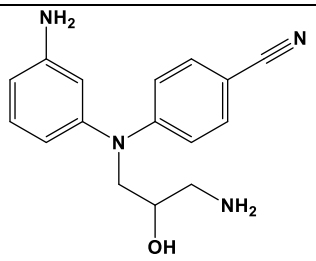
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
15a		321.22	21.38	-2-20
15b		299.37	23.14	-1.42
15c		328.37	23.75	-1.06
15d		286.33	22.08	-0.63

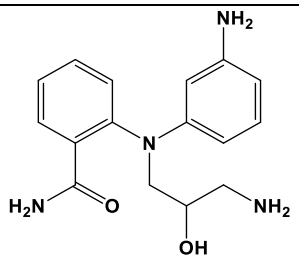
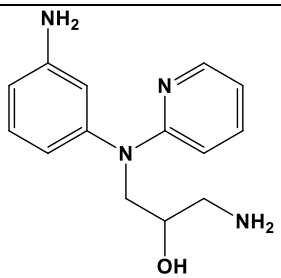
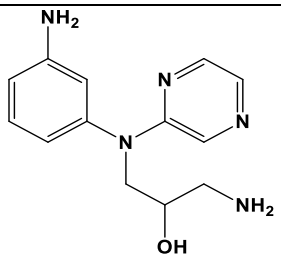
3.2.3 DERIVATIVES OF 16

Selected derivatives of **16** are reported in **Table 17**. The introduction of a hydrogen bond donor allows gaining further interaction with Glu249. The introduction of the pyridine ring in **16k** and of the pyrazine in **16l** allows the interaction with His 318. The introduction of the chlorine atom (**16a**), as well as the substitution of the benzyl ring with the aniline (**16d**, **16h**), that establish interactions with Glu249, are well tolerated.

Table 17. Derivatives of 16. Molecular weight, docking score, and predicted logS are reported.

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
16a		291.78	22.09	-2.81
16b		271.36	25.23	-2.81
16d		273.33	23.28	-2.16

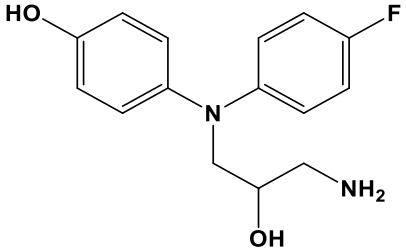
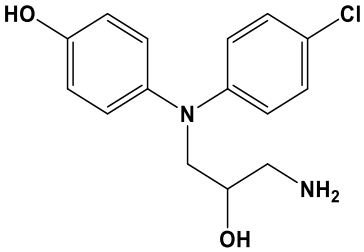
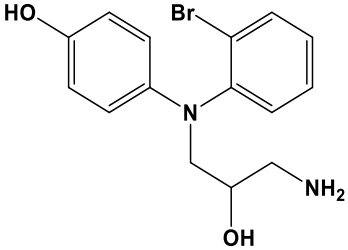
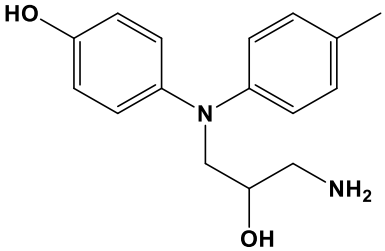
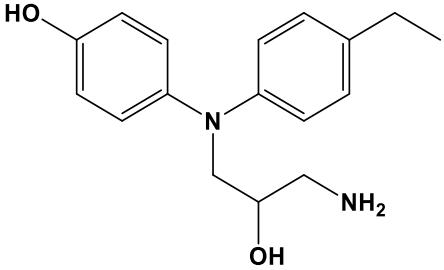
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
16e		273.33	23.89	-2.16
16f		287.36	21.04	-2.52
16d		272.35	23.21	-2.03
16h		272.35	24.73	-2.03
16i		282.34	22.66	-3.22

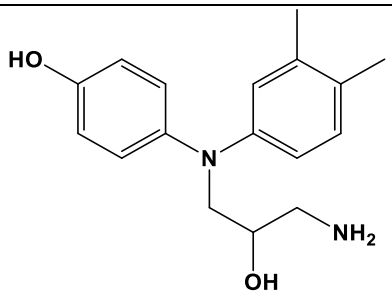
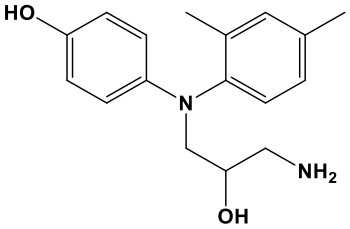
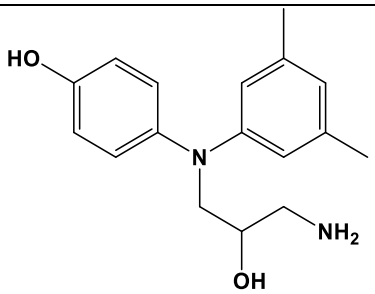
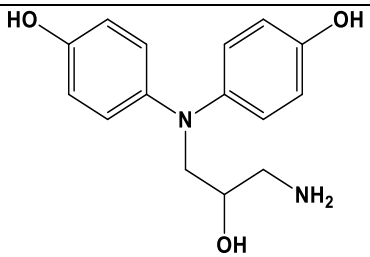
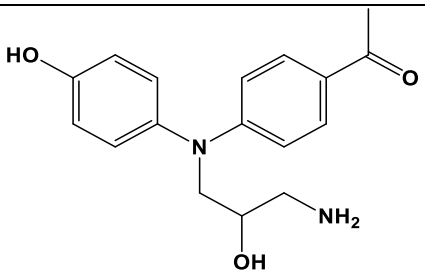
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
16j		300.36	22.00	-2.07
16k		258.32	25.69	-1.94
16l		259.31	23.12	-1.56

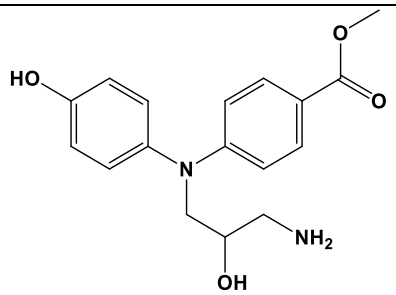
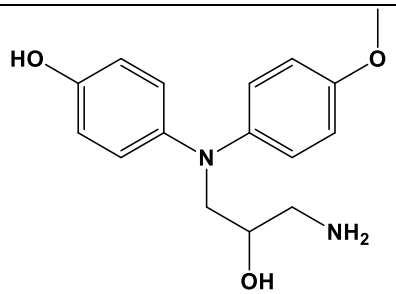
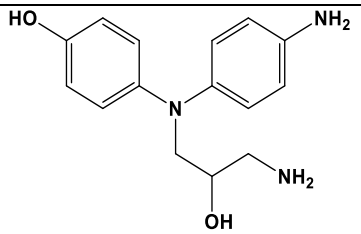
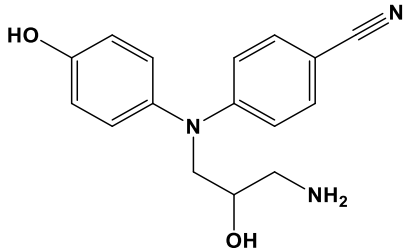
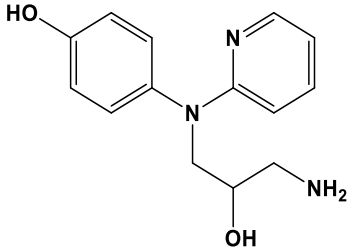
3.2.4 DERIVATIVES OF 17

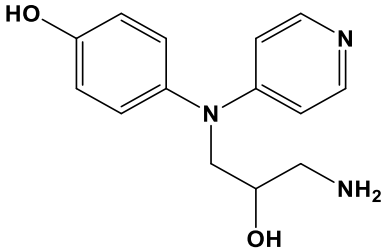
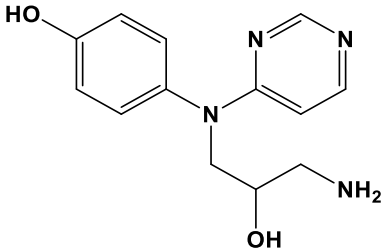
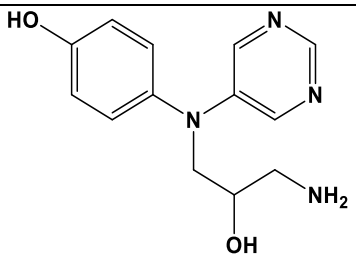
Derivatives of **17** are reported in table 20. The hydroxy group is almost always involved in hydrogen bonds with the backbone of Gly248. Compound **17i**, which presents an additional hydroxyl group on the second phenyl ring, can establish a hydrogen bond with the backbone of Gly316. The same interactions have been found for compound **17m**. The substitution of the phenyl ring with the pyridine as in **17p** allows the interaction with His318.

Table 18. Derivatives of 17. Molecular weight, docking score, and predicted logS are reported.

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
17a		276.31	21.57	-1.75
17b		292.764	23.45	-1.87
17c		337.215	21.52	-1.80
17d		272.346	23.10	-1.71
17e		286.37	22.34	-2.11

Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
17f		286.37	22.66	-2.40
17g		286.37	24.53	-1.88
17h		286.37	22.14	-2.22
17i		274.32	24.36	-0.82
17j		300.58	22.00	-1.68

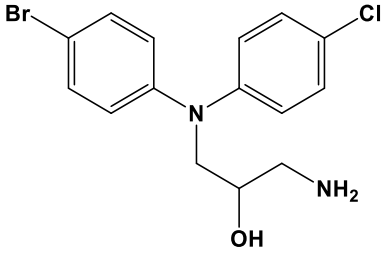
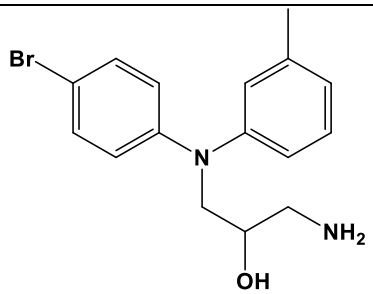
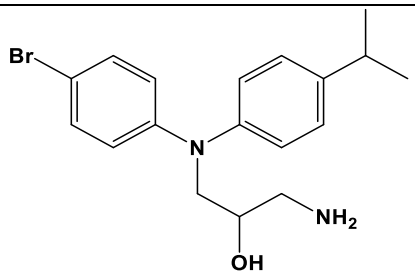
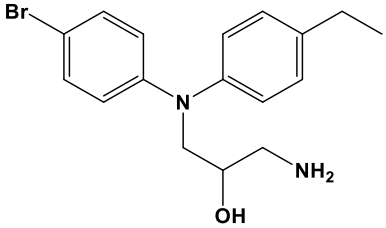
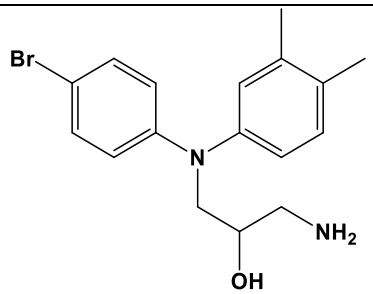
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
17k		316.36	23.18	-2.17
17l		288.35	20.30	-1.73
17m		273.33	26.65	-0.85
17n		282.34	23.47	-2.33
17o		259.31	20.52	-0.98

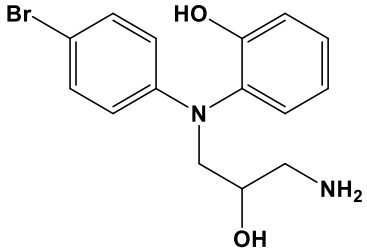
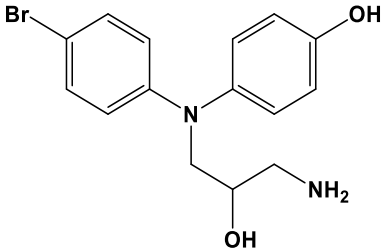
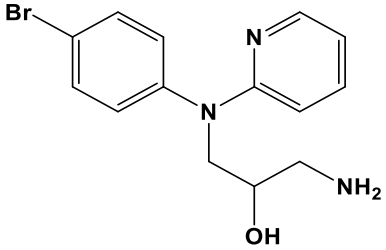
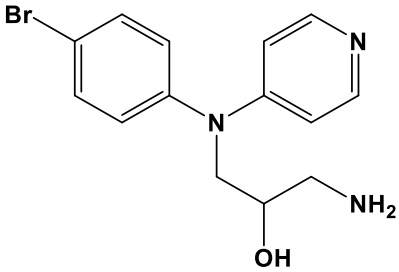
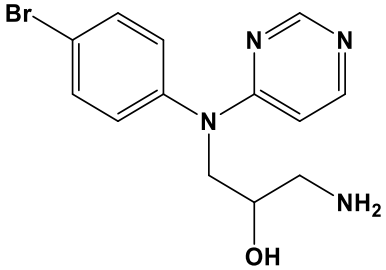
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
17p		259.31	21.26	-0.76
17q		260.30	20.81	-1.17
17r		260.30	20.61	0.57

3.2.5 DERIVATIVES OF 18

The introduction of the chlorine atom and of hydrophobic groups is well tolerated and compounds retain the main interactions of **5**. The hydroxy group of **18f** and **18g** in meta and para position (Table 22), respectively, can establish hydrogen bond interactions with Glu249. The substitution of the phenyl ring with pyridine (**18h**) or pyrimidine (**18j**) allows the interaction of the nitrogen atom with His318.

Table 19. Derivatives of 18. Molecular weight, docking score and predicted logS are reported.

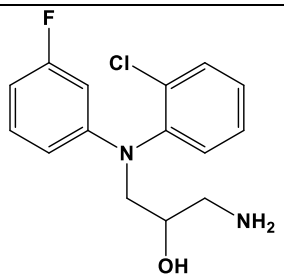
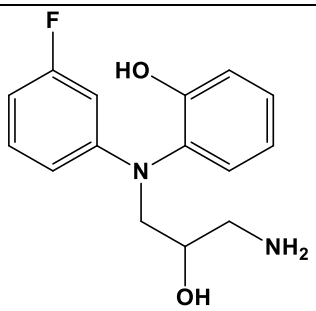
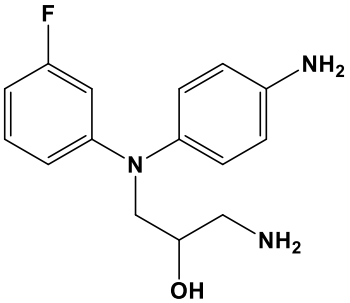
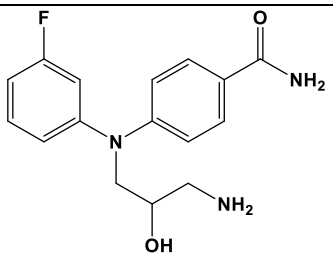
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
18a		355.66	22.14	-2.86
18b		335.24	22.82	-2.60
18c		363.30	21.30	-3.40
18d		349.27	21.40	-2.93
18e		349.27	21.01	-3.20

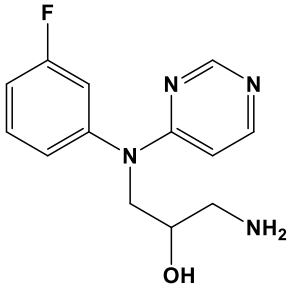
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
18f		337.215	22.71	-1.95
18g		337.215	24.24	-1.93
18h		322.204	22.36	-2.19
18i		322.204	19.69	-2.05
18j		323.192	19.35	-1.81

3.2.6 DERIVATIVES OF 19

Derivatives of **19** are reported in table 22. The insertion of chlorine (**19a**) is well tolerated. **19c** interacts with Glu249, while the **19d** makes interactions with the backbone of Gly316.

Table 20. Derivatives of 19. Molecular weight, docking score, and predicted logS are reported.

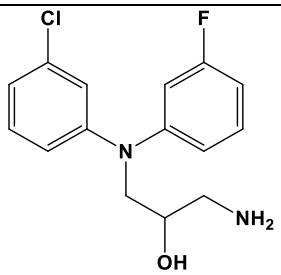
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
19a		294.76	21.10	-2.55
19b		276.31	22.46	-1.52
19c		257.33	24.41	-1.49
19d		303.34	23.45	-1.95

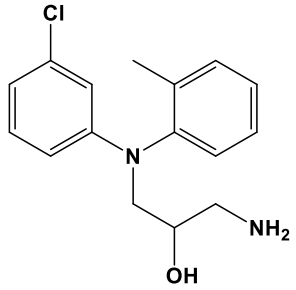
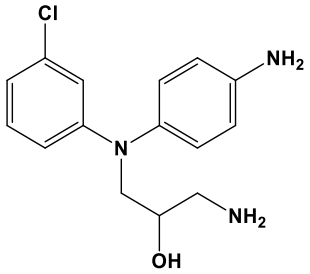
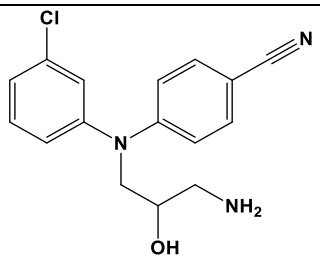
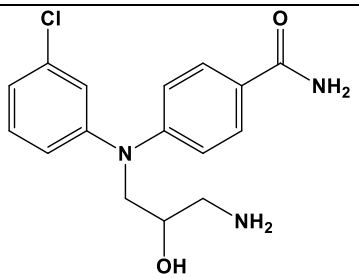
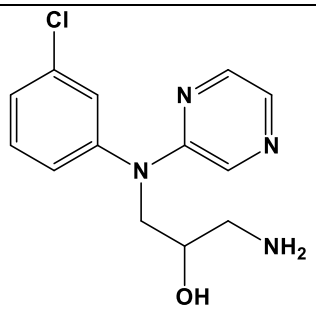
Cmpd ID	Structure	Molecular Weight	Docking Score	QPlogS
19e		262.29	18.58	-1.26

3.2.7 DERIVATIVES OF 20

Derivatives of **20** are reported in table 24. The introduction of a fluorine atom in the meta position, like the methyl group in the ortho position, is well tolerated. The aniline moiety of **20c** allows establishing interaction with Glu249. The benzamide group of **20e** establishes the same interaction. The substitution of the phenyl ring with the pyrazine ad in **20f** allows a hydrogen bond interaction with His153.

Table 21. Derivatives of 20. Molecular weight, docking score, and predicted logS are reported.

Cmpd ID	Structure	Docking Score	QPlogS	Molecular Weight
20a		22.12	-2.37	294.76

Cmpd ID	Structure	Docking Score	QPlogS	Molecular Weight
20b		21.30	-2.92	290.79
20c		23.48	-1.65	292.78
20d		21.69	-3.47	301.78
20e		24.07	-1.79	319.79
20f		21.66	-1.59	278.74

3.3 Materials and Methods

Ligands Preparation

Ligands were designed by means of Maestro (version 11.2.014) graphical interface [68] and pre-treated with LigPrep (version 2014:4) [69] tools at default settings, checking the ionization state at physiological pH with Epik. [70, 71, 72]

Docking Studies

All the selected molecules were docked within the UM pocket using Gold 5.2 software. [59] The pocket under investigation was inserted into a grid box centered on Arg292 and enclosing residues lying within 9 Å from such amino acid. The genetic algorithm parameter settings were employed using the search efficiency set at 100%, and 100 runs were carried out for each ligand. Chemscore was chosen as the fitness function, and asp as rescore fitness function. For each inhibitor, the first ranked solution was selected for further analysis.

ADME PREDICTION

ADME molecular properties of all selected compounds were predicted with Qikprop, using the Fast mode option. QikProp [60] is a quick, accurate, easy-to-use ADME prediction program. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches.

Pictures

Pictures of the modeled ligand-enzyme complexes, together with graphic manipulations, were rendered with the PyMOL package [73] (version 1.8.4.0 [http://www.pymol.org/]).

3.4 A PHARMACOPHORE-BASED VIRTUAL SCREENING TO FIND NOVEL CHEMICAL SCAFFOLDS

Considering the results obtained with **11** and the difficulties in designing novel derivatives of **5** able to bind the UM pocket, a more accurate 3D pharmacophore-based screening was performed to find new possible scaffolds to purchase and submit to biological assays.

For this purpose, a 3D pharmacophore structure-based model has been generated using LigandScout software. [76] Pharmacophores allow the definition of the essential features of one or more molecules with the same biological activity; in this case, the essential features identified were used to research molecules that share the same features arranged in the same relative orientation. LigandScout is a user-friendly integrated platform for molecular design and accurate virtual screening that generate both ligand- and structure-based pharmacophore models.

In this protocol, a structure-based modelling procedure takes into account the ligand-binding, the pocket proprieties and allows to create pharmacophore models starting from the complex structure and including excluded volume spheres that represent the areas occupied by target amino acids and thus forbidden to any potential ligand.

The 3D structure of the homology model of DDX3X in closed conformation and in complex with **5** was chosen as input for the generation of the first structure-based pharmacophore. A 3D pharmacophore hypothesis representing the main interactions between the enzyme and the inhibitor has been obtained through the automated structure-based model generation protocol. In **Figure 41a** is reported the generated model for docked complex DDX3-**5**. It consists of five features, including:

- a hydrogen bond acceptor (red arrow) that represents the interaction among the hydroxyl group and Arg262;
- a hydrogen bond donor (green arrows) that represents the interaction between the hydroxyl group and Glu256;
- two hydrogen bond donors that represent the interactions between the amine hydrogens and the backbone of Cys317 and Thr266;

- a hydrophobic region (yellow sphere) that interacts with Ile268;
- twelve excluded volumes.

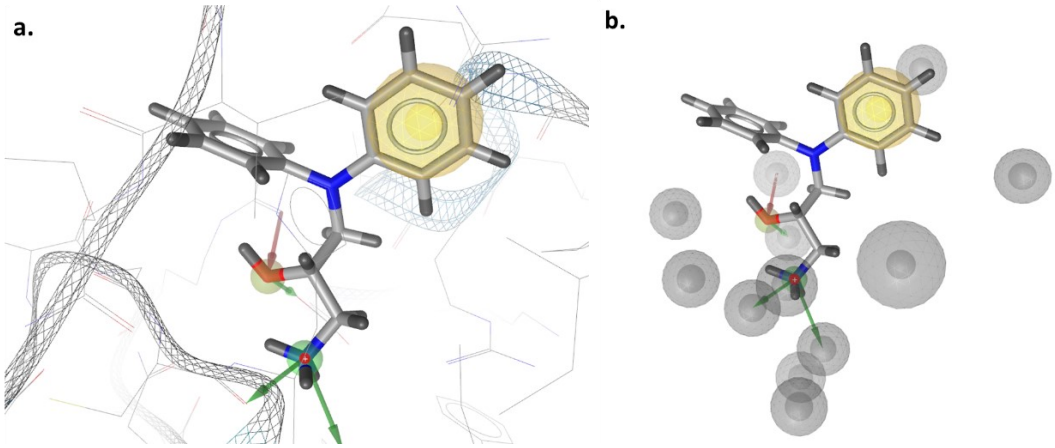


*Figure 41. (a.) Structure-based pharmacophore model generated using LigandScout from homology model of DDX3 close conformation in complex with **5**. Hydrogen bond donor (red arrow) and acceptors (green arrows) are shown together with the hydrophobic feature. **(b.)** Representation of the excluded volumes that represent region of space occupied by the protein side chain and that are used to characterize the inaccessible areas for any potential ligand.*

In the next step, the MolPort commercial database, available online and constituted by 7000000 molecules, was screened using the iscreen protocol integrated in LigandScout tools. At the end of the filtration step, 2795 molecules were found.

These compounds were further filtered by docking using the GOLD software. [59] Docked compounds were prioritized considering the docking score value and the visual evaluation of the docked pose agreement with their pharmacophore mapping pose. All molecules with a docking score lower than that of **5**, which has a value of 24.68, were discarded, obtaining 610 molecules.

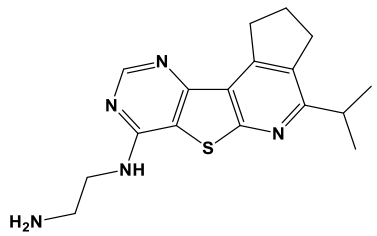
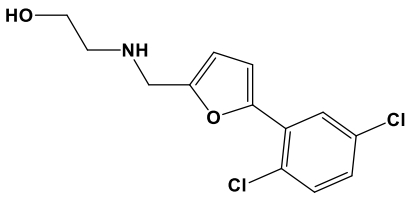
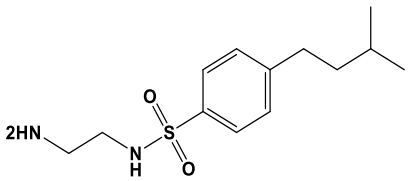
As an additional filter, the webserver FafDrugs4 [77] was used to discard PAINS (Pan Assay Interference Compounds). No pains were found, but molecules identified as intermediates (molecules with a low-risk structural alert) were abandoned. The number of molecules accepted by FafDrugs4 is 501.

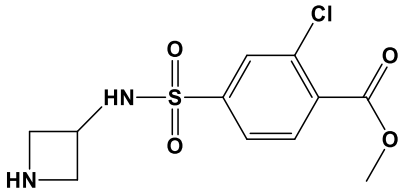
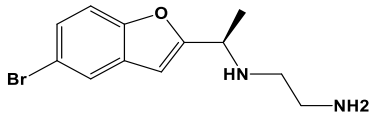
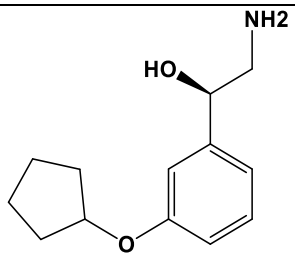
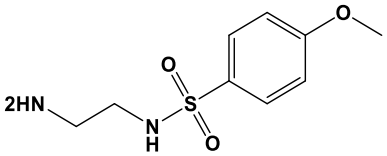
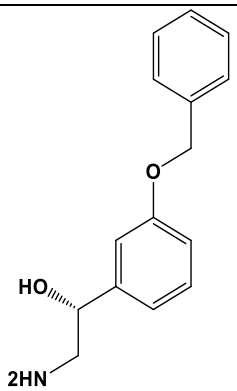
These molecules were docked again in the UM pocket of DDX3X using GOLD. Configuration settings are reported in Materials and Methods section. Visual

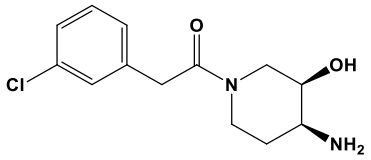
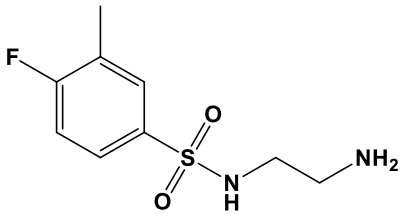
inspection allowed to discard all the compounds docked outside the UM pocket and that could not establish the fundamental interactions with the UM pocket.

The resulting 22 compounds were evaluated for their ΔG binding using Prime MM-GBSA to calculate the binding energies for all the ligands and DDX3X receptor. Finally, compounds have been prioritized based on their binding affinity value, synthetic feasibility, and commercial availability. The first 10 compounds are reported in table 25 together with their docking score, QPlogS values and commercial ID.

Table 25. Chemical structures, docking scores, MMGBSA values, predicted aqueous solubility values and commercial ID of the first ten prioritized compounds obtained from the screening of MolPort database.

Cmpd ID	Structure	Docking Score	MMGBSA	QPlogS	MolPort ID
MP01		21.31	-40.84	-3.38	MolPort-002-658-444
MP02		21.32	-38.15	-2.86	MolPort-000-866-004
MP03		22.81	-37.37	-1.65	MolPort-038-957-913

Cmpd ID	Structure	Docking Score	MMGBSA	QPlogS	MolPort ID
MP04		22.68	-37.28	-3.53	MolPort-038-998-749
MP05		21.89	-35.22	-0.88	MolPort-013-051-042
MP06		25.15	-34.84	-0.87	MolPort-008-549-057
MP07		21.61	-33.32	0.50	MolPort-002-090-640
MP08		23.55	-33.00	-1.78	MolPort-008-568-306

Cmpd ID	Structure	Docking Score	MMGBSA	QPlogS	MolPort ID
MP09		23.38	-29.70	-1.13	MolPort-023-323-129
MP10		23.59	-28.35	-1.54	MolPort-028-767-230

All these compounds are predicted to retain the main interactions found for **5**, especially with Glu256, the principal responsible of the interaction of the UM compounds with DDX3X. Hence, the compounds have been selected to be purchased or synthesised and then to be submitted to the biological assay. If compounds will be active, it will be possible to design a novel library of derivatives to rule preliminary SAR considerations. If the compounds will be not active, they can be used in a new virtual screening procedure, avoiding this kind of scaffolds.

3.5 Materials and Methods

Pharmacophore generation

The pharmacophore model was generated with LigandScout,[67] version 4.3, based on DDX3-5 docked complex and using the simplified pharmacophore generation mode. The pharmacophore is constituted of five features (three hydrogen bond donors, one hydrogen bond acceptor, and one hydrophobic feature, that account for the interactions between 5 and DDX3) and twelve excluded volumes.

Ligands preparation

Ligands were designed with Maestro (version 11.2.014) graphical interface [68] and pre-treated with LigPrep (version 2014:4) [69] tools at default settings, checking the ionization state at physiological pH with Epik. [70, 71, 72]

Library preparation

MolPort commercial library is available online and was used for the pharmacophore-based virtual screening. The library has been downloaded in sdf format and then converted in LigandScout input file format (.LDB files). The databases that LigandScout uses for internal virtual screening have been created using a command-line tool called ibgen with the following parameters: confgen-type “icon-best” and num-confs “50” while other parameters were set as default value.

LigandScout uses the iCon conformer generator implemented starting from LigandScout 4.0. The iCon-fast option generates a maximum of 25 conformers for molecule, while the iCon-best option generates 500 conformers; it is possible to define the exact number of conformers to be generated with the *num-confs* options, that has been set to 50.

Virtual screening

The pharmacophore model was used to screen MolPort commercial database using LigandScout Iscreen module (verson 4.3). [76] The multi-conformational database has been screened by applying the BEST Flexible search to retrieve the compounds with novel scaffolds and with desired chemical features. Maximum Omitted Features option was

chosen as 0 for the first screening and it was changed into 1 for the second screening because mapping all features present in the structure-based hypothesis will reduce the hit rate. A total of 2795 molecules were obtained.

Docking Studies

All the selected molecules were docked within the UM pocket using GOLD 5.2 software. [59] The pocket under investigation was inserted into a grid box centred on Arg292 and enclosing residues lying within 9 Å from such amino acid. The genetic algorithm parameter settings were employed using the search efficiency set at 100%, and 30 and, in the second docking protocol, 100 runs were carried out for each ligand. Chemscore was chosen as the fitness function, and asp as rescore fitness function. For each inhibitor, the first ranked solution was selected for further analysis.

PAINS Analysis

FAF-Drugs4 [77] (Free ADME-Tox Filtering Tool) is a web-server able to predict some ADME-Tox properties to help the hit selection and discard PAINS [78] before chemical synthesis or ordering using pre-defined filters.

MMGBSA analysis

Prime MM-GBSA [54] is a tool that allows the calculation of the binding energies for a protein-ligand complex. Calculations were performed using the OPLS3 force field and the VSGB solvation model. [79] All amino acids that are 5 Å distant from the ligand have been relaxed in the calculation, together with the ligands. Default settings are left for the sampling method.

ADME prediction

ADME molecular properties of all selected compounds were predicted with Qikprop, using the Fast mode option. QikProp [60] is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches.

Pictures

Pictures of the modeled ligand-enzyme complexes, together with graphic manipulations, were rendered with the PyMOL package [73] (version 1.8.4.0 [http://www.pymol.org/]).

CHAPTER 4

THE PROTAC PROJECT

PROTAC, or Proteolysis-targeting Chimera, is a small-molecule strategy to targeted protein degradation.

A PROTAC molecule is constituted by two active domains and a linker. The first domain in constituted by a small molecule that binds to a Protein of Interest (POI), while the second domain is constituted by a small-molecule ligand for an E3 ubiquitin ligase (E3). Upon binding the two proteins simultaneously, the PROTAC beings the two proteins in proximity of each other, and as a result the E3 can ubiquitinate the target protein, which then gets selectively degraded (Figure 42). The degradation of the proteins happens by endogenous 26S proteasome.

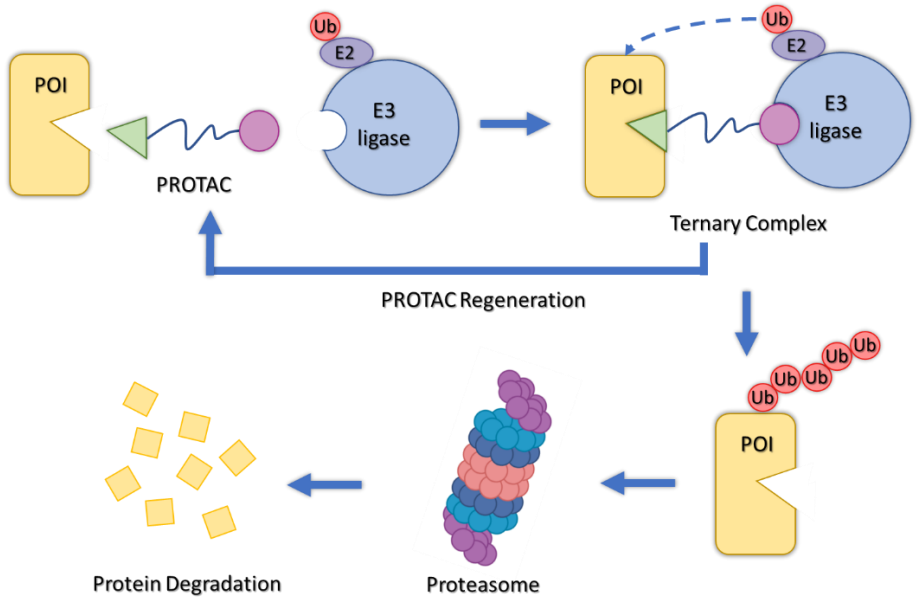
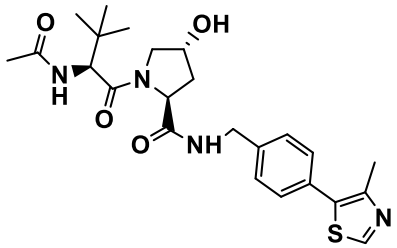
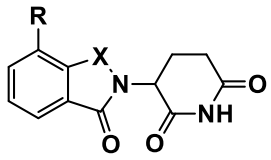


Figure 42. Mechanism of PROTAC-mediated protein degradation.

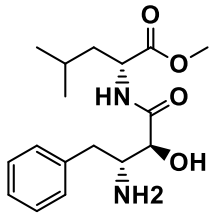
Although there are more than 600 ubiquitin E3 ligases, the most popular E3 ligases are Skp1-Cullin-F box complex containing Hrt1 (SCF), [80] Von Hippel-Lindau tumour suppressor (VHL), [81] Cereblon (CRBN), [82] inhibitor of apoptosis proteins (IAPs), [83] and mouse double minute 2 homolog (MDM2) [84] have been used for the design of PROTACs. Representative ligands of these E3 ligases are reported in Figure 43.



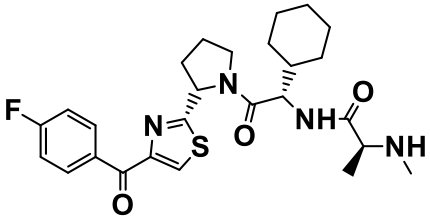
VHL E3 ligand:
VHL1/VH032 (**VHL**)
CAS No:1148188-62-2



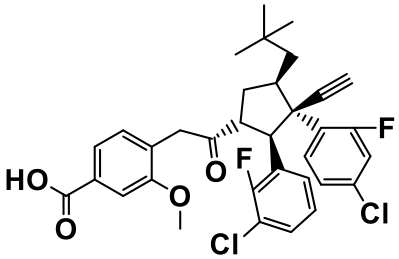
CRBN E3 ligand:
R=H, X=CO, Thalidomide(**THA**)
CAS No:50-35-1
R=NH_2, X=CO, Pomalimide (**POMA**)
CAS No: 19171-19-8
R=NH_2, X=CH_2, Lenalidomide (**LENA**)
CAS No: 191732-72-6



IAP E3 ligand:
Methyl bestatin (**MeBS**)
CAS No:65322-89-6



IAP E3 ligand:
LCL161 derivative (**LCL161**)
CAS No:1005342-46-0



MDM2 E3 ligand:
RG7338 (**RG**)
CAS No:1229705-06-9

Figure 43. Small-molecule ligands of E3s used for PROTAC.

One of the advantages of PROTAC technology is that these molecules work catalytically. A PROTAC molecule can lead to the degradation of multiple molecules of POI and can act sub-stoichiometrically at fractional occupancy of the POI. Consequently, PROTACs often show higher POI degradation than expected based on their binding affinity to the POI alone. [85]

The development of PROTACs have also opened a new chapter for developing drugs and novel chemical knockdown tools in cancer therapy, allowing overcoming drug resistance. For example, although kinase inhibitors are very effective in cancer therapy, they can lead to drug resistance and disease recurrence. [86] Using PROTACs, able to degrade the whole protein, it is possible to treat drug-resistant cancers. Kinase degradation allows the elimination of both the enzymatic and nonenzymatic functions of these proteins, expanding the druggable space of the

existing targets and regulating proteins that are difficult to control by traditional small molecule inhibitors.

PROTACs allow the degradation of undruggable proteins that lack catalytic activity and/or have independent catalytic functions. STAT3 (Signal Transducer and Activator of Transcription 3) is an attractive target involved in several signalling pathways that lack a druggable site on its surface, limiting the development of inhibitors. In November 2019, Shaomeng Wang's group first reported a potent PROTAC targeting STAT3 with potent biological activities *in vitro* and *in vivo*, [87] confirming the key potential of PROTAC technology.

Finally, PROTAC represents a fast and reversible strategy to obtain chemical knockdown *in vivo*. Traditional genetic protein knockout technologies, zinc-finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), or CRISPR-Cas, usually have a long cycle, irreversible mode of action, and high cost, which brings much inconvenience for research, especially in non-human primates. Besides, these genetic animal models sometimes produce phenotypic misunderstanding due to potential gene compensation or gene mutation. More importantly, the traditional genetic method cannot be used to study the function of embryonic-lethal genes *in vivo*. Unlike DNA-based protein knockout technology, PROTACs knockdown target proteins directly, allowing the knockdown of a target protein at specific time points and enabling the recovery of the target protein after the drug treatment. This new reversible knockdown method can be used as supplement to the existing genetic tools. [88]

4.1 THE SOCS2 PROTEIN

Cytokines belongs to a large family of glycoproteins and are important mediators of cell-cell communication. These proteins regulate critical biological processes such as proliferation, differentiation, immunity and haematopoiesis. [89] They are secreted by cells as a result of environmental stimuli and are able to transmit information to neighbouring cells through the binding to the appropriate receptor on their surface. The message is rapidly transferred to the nucleus using different signalling cascades activating the Janus kinase and signal transducer and activator of the transcription (JAK-STAT) pathway.

Cytokine actions are strongly controlled both in magnitude and duration. In fact, aberrant cytokine signalling has been associated with many diseases, including cancers, disorders in haematopoiesis, and autoimmune diseases. Among the regulators of cytokines, there are the protein inhibitors of activated STATs (PIAS), the Src-homology 2 (SH2) containing protein tyrosine phosphatases (SHPs) and the protein family of suppressors of cytokine signalling (SOCS). [90] SOCS proteins are rapidly induced upon JAK/STAT signalling to negatively regulate cytokine signalling via a classical feedback loop. [91]

The SOCS family consists of 8 proteins, SOCS1-SOCS7, and the cytokine-inducible SH2-containing protein (CIS). They are characterised by a central SH2 domain, an N-terminal domain of variable length and sequence, and a C-terminal 40 amino-acid region called SOCS box (Figure 44).

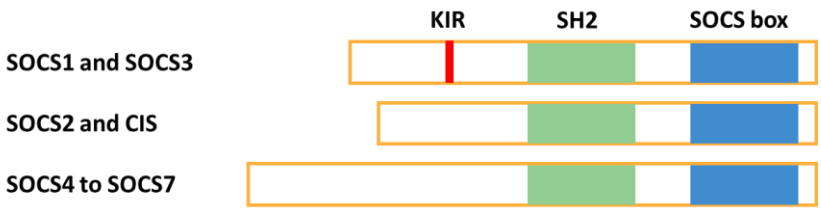


Figure 44. Schematic representation of SOCS structure. KIR is the kinase inhibitory region. SH2 domain mediates the recruitment of phosphorylated tyrosine-containing sequence of the substrate. SOCS box domain binds the adaptor EloBC and is involved in protein ubiquitination.

SOCS proteins can interact with a series of signalling intermediates by binding their SH2 domain to phosphorylated tyrosine residues, particularly those on cytokine receptors and JAKs, leading to the blockade of the signal. [92] The SOCS box allow them to acts as ubiquitin ligases for associated proteins and target them for proteasomal degradation. All SOCS proteins are able to bind EloBC and Cullin5, allowing the formation of different SOCS-EloBC-Cullin5-Rbx2 ($CRL5^{SOCS}$) E3 ligases complexes. In this paragraph it will be analysed the E3 ligase complex constituted by Rbx1/2, Cullin5, ElonginB, ElonginC and SOCS2, with SOCS2 that function as the substrate recognizing component. [93, 94, 95]

SOCS2 is induced by GH, PRL, EPO, GM-CSF, G-CSF, IL-1, IL-2, IL-3, IL-4, IL-6 IL-15, CNTF, IFN-alpha, IFN-gamma, LIF and insulin [96] cytokines, but it can regulate also several signalling pathways, such as those induced by GH, PRL, LIF, IL-2, IL-3, IL-6, EGF and IGF-1. [97] It is also involved in the ubiquitination of target proteins, including receptors such as GHR and diverse signalling proteins. SOCS2 is the primary suppressor of growth hormone (GH) pathway where a gigantism phenotype was observed in a $SOCS2^{-/-}$mice. [98] However, also the overexpression of SOCS2 in transgenic mice led to the same phenotype. In, fact, the SOCS2 overexpression is associated with growth hormone (GH) and is so involved in cell growth. [99]

The attenuation of GHR signalling is based on two GHR phosphorylation sites recognized by SOCS2. [100, 101] The SOCS2-EloBC-Cul5-Rbx2 complex, or $CRL5^{SOCS2}$ E3 ligase, interacts with the pY487 site of GHR to target the receptor for ubiquitination and proteasomal degradation. [102] The pY595 site of GHR is involved in interactions with SOCS2, STAT5b and SHP2 (SH2 domain-containing phosphatase 2). If SOCS2 binds this site, the receptor is no more able to interact with STAT5b, causing the signalling inhibition. [100] If both sites are deleted, SOCS2 is not able to inhibit the GH signalling. [101, 102]

SOCS2 interacts also with other substrates, such as the erythropoietin receptor (EpoR) that interact at the site pY426, [103] the leptin receptor that interact at the site pY107, [104] the epidermal growth factor receptor, [105] and the insulin-like growth factor-I receptor. [106]

SOCS2 is downregulated in breast, lung, liver, and ovarian cancer. It is involved in regulating the NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) pathway that is implicated in the immune and inflammatory responses. SOCS2 targets NDR1 for the degradation and to negatively regulate the production of TNFα induced by NF-kB. If NF-kB is downregulated, there are increased levels of NDR1 that lead to an aggressive behaviour of PC3 prostate cancer cells. [107]

As a result of these roles and functions, SOCS2 is involved in the development of cancer, diabetes, neurological and inflammatory diseases, and for this reason, it is potentially an attractive therapeutic target. [92]

4.2 STATE OF THE ART

The research group of Professor Ciulli has solved the SOCS2-ElonginB-ElonginC (SBC) crystal structure in complex with phosphorylated epitope peptides derived from its physiological targets GHR and EpoR [108] that have been used to design SOCS2 ligands as potential therapeutics. The design of SOCS2 binders at the pY pocket can be used as inhibitors of the $CRL5^{SOCS2}$, preventing the degradation of target substrate receptors and prolonging the activity of cytokine signaling pathway. Moreover, a SOCS2 binder can provide the starting point to design novel PROTAC degraders to hijack SOCS2 CRL activity towards target proteins inside cell. [109]

In particular, the analysis of the SBC in complex with EpoR_pY426 peptide (SBC-EpoR, PDB ID: 6I4X) showed that the pY residue is involved in hydrogen-bonds with residues of Arg73, Ser75, Ser76, Thr83, and Arg96 of SOCS2, while the backbone of EpoR_pY426 peptide from Glu(−1) to Leu(+3) interacts with Thr93, Asn94, Asp107 (**Figure 45**). Hydrophobic interactions involve the C-terminal residues of EpoR_pY426, Ile(+2), Leu(+3), and Pro(+5), and residues of Leu95, Leu106, Ser108, Ile109, Val112, Leu116, and Leu150 of SOCS2.

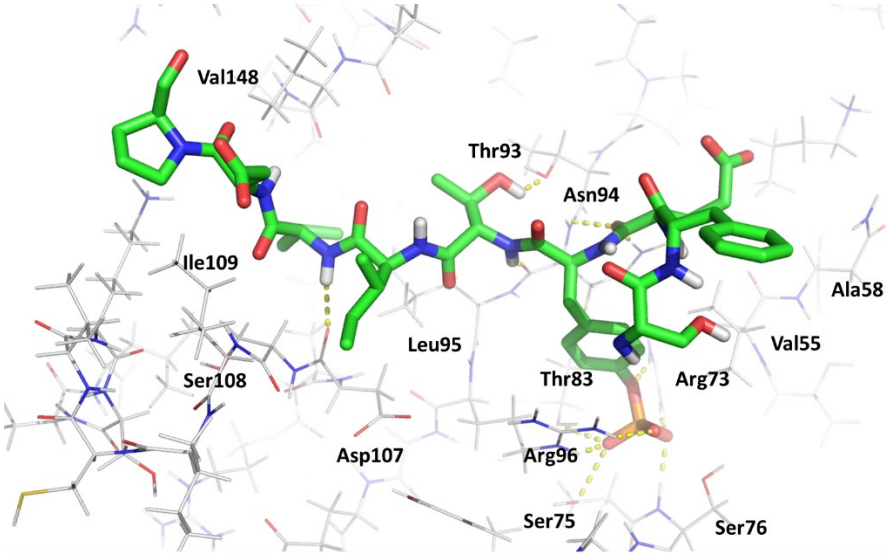


Figure 45. Structure of SBC in complex with EpoR_pY426. The phosphate group is involved in hydrogen bond interactions with Arg96,Thr83, Ser75 and Ser76. The backbone of EpoR_pY426 peptide from Glu(−1) to Leu(+3) interacts with Thr93, Asn94, Asp107, while the C-terminal residues of EpoR_pY426, Ile(+2), Leu(+3), and Pro(+5), and residues of Leu95, Leu106, Ser108, Ile109, Val112, Leu116, and Leu150 (LHS). Glu(-1), Phe(-2) and Ser(-3) interacts with Val55, Ala58 andArg73 (RHS).

The SBC-GHR complex presents two anti-parallel copies of GHR_pY595 peptides (called peptide A and B) that bind the SH2 domain of SOCS2 (PDB ID: 6I5J). [109] The pY of peptide A binds the pY pocket of SOCS2 in the same way of EpoR peptide while the pY of peptide B is instead solvent exposed and establish hydrogen bonds interactions with the His149 of SOCS2 (Figure 46). Both peptides form hydrogen bonds with Leu95, Leu106, Ser108, Leu116, and Leu150 of SOCS2, while hydrophobic interactions are formed with Leu95, Leu106, Ser108, Leu116, and Leu150 (Fig. 2d). Other hydrophobic interactions are formed between Val(−3) of peptide A of SBC-GHR and Thr88, Ala90, Thr93, Leu95, and Val148 of SOCS2.

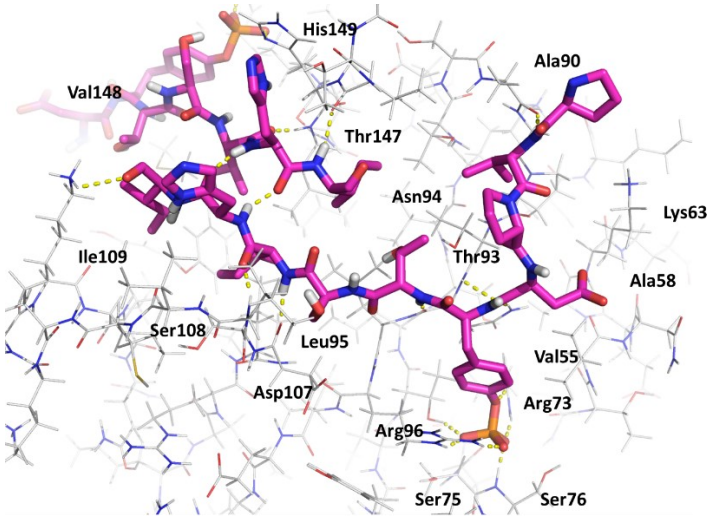


Figure 46. Structure of SBC-GHR complex. Two copies of GHR_pY595 peptides binsd SOCS2. Peptide A establish the same interaction of the EpoR peptide reported in Figure 45. The pY of peptide B is solvent exposed and establish hydrogen bonds interactions with the His149 of SOCS2. Hydrogen bonds are formed with Leu95, Leu106, Ser108, Leu116, and Leu150 of SOCS2, while hydrophobic interactions are formed with Leu95, Leu106, Ser108, Leu116, Thr88, Ala90, Thr93, Leu95, and Val148 of SOCS2.

Starting from all these findings, a series of SOCS2 binders have been designed in a structure-guided manner and synthetized and their activities have been tested.

4.3 AIM OF THE WORK

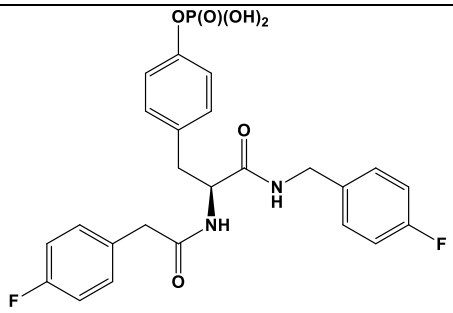
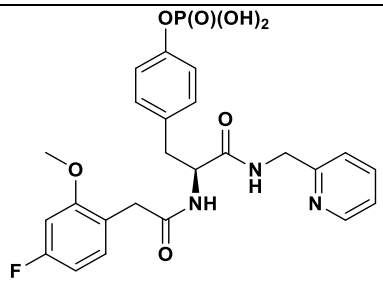
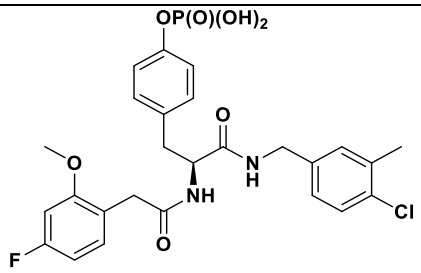
This project aims to establish a Structure-Activity Relationship (SAR) of the SOCS2 binders and develop a QSAR model that will be used to predict the binding affinity of novel compounds to synthesise. To this purpose, the co-crystal structures of SOCS2 with its binders have been analysed and used to perform molecular dynamics simulations and calculate the ΔG of the complexes. The compounds have also been used to develop a QSAR model using the AutoQSAR tool [110, 111] in the Schrödinger suite.

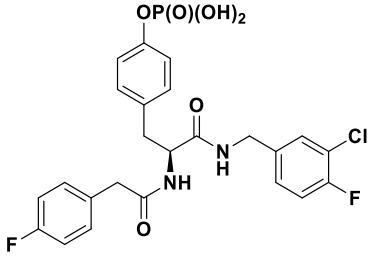
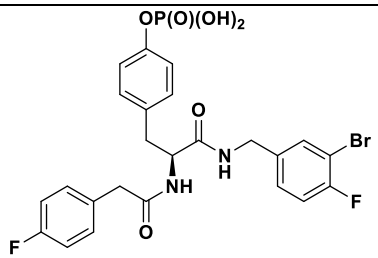
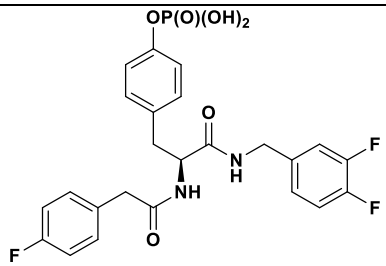
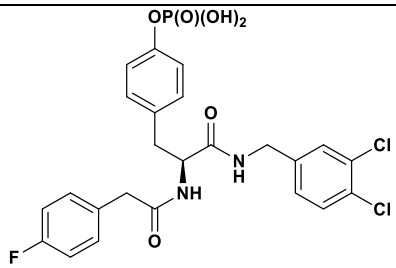
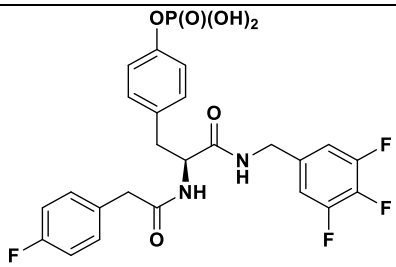
4.4 MOLECULAR DYNAMICS SIMULATION STUDIES OF SOCS2 INHIBITORS

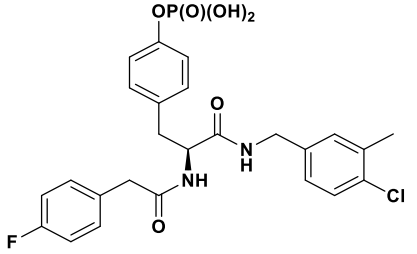
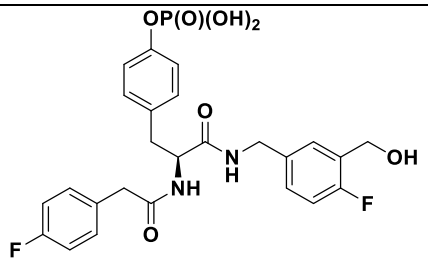
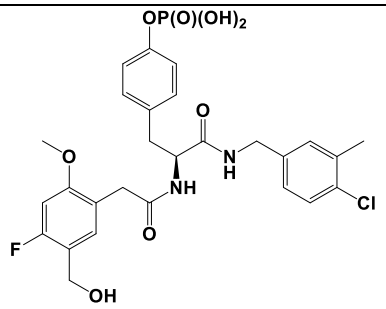
In Table 22 are reported the structures and the binding affinity values of the SOCS2 binders together with their MMGBSA (ΔG) values. The activity has been reported as K_d and was measured using the Isothermal Titration Calorimetry (ITC) assay. All the compounds have been subjected to a Molecular Dynamics (MD) simulation studies.

Each complex was subjected to post-processing trajectory analysis that include the calculation of the root-mean-square deviation (RMSD) value, that represents the measure of the average distance between the atoms and is employed to study the configuration of ligands when bound to macromolecules; the hydrogen bond occupancy value, to determine the percentage of H-bond occupancy between protein and ligand; the ΔG value using the MMGBSA method. The ΔG value is expressed in kcal/mol and estimates the free energy of the binding of small ligands to biological macromolecules. [112]

Table 22. Structure, binding affinity values and ΔG values of the SOCS2 binders.

Cmpd ID	Structure	Binding Affinity (μM)	MMGBSA (ΔG) (kcal/mol)
MN352		2.6	-71.70
MN448		12	-28.52
MN416		0.78	-74.12

Cmpd ID	Structure	Binding Affinity (μM)	MMGBSA (ΔG) (kcal/mol)
MN356		0.81	-79.73
MN357		0.78	-70.73
MN366		1.3	-70.93
MN367		1,0	-51.90
MN368		1,2	-66.87

Cmpd ID	Structure	Binding Affinity (μM)	MMGBSA (ΔG) (kcal/mol)
MN402		0.35	-71.95
MN403		4.5	-76.93
MN442		1.1	-71.01

4.4.1 MN352

The analysis of the SOCS2-MN352 complex, represented in **Figure 47**, shows that the ligand is accommodated into the pY binding pocket and interacts with the protein through a dense network of hydrogen bonds. The phosphate group binds residues of Thr83, Arg96, Arg73, Ser75 and Ser76 and is also involved in a hydrogen bond with a water molecule. The amide groups are involved in hydrogen bonds with the backbone and the side chain of Asn94 and Thr93. A fluorobenzyl ring is accommodated into a hydrophobic pocket constituted by Leu150, Ile110, and Val86 while the other one takes contacts with residues of Pro92, Ala58, and Val55.

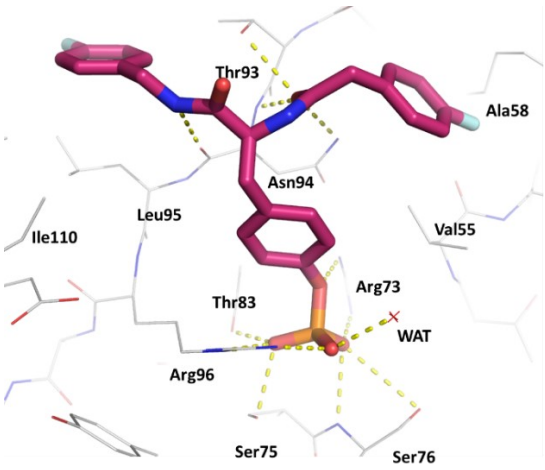


Figure 47. Crystal structure of the complex SOCS2-MN352.

The analysis of the MD simulation of the MN352-SOCS2 complex shows that the compound is not subjected to great conformational changes, with fluctuations in the RMSD plot of the ligand that are due to the rotation of the fluorobenzyl ring in the RHS around the carbon atom.

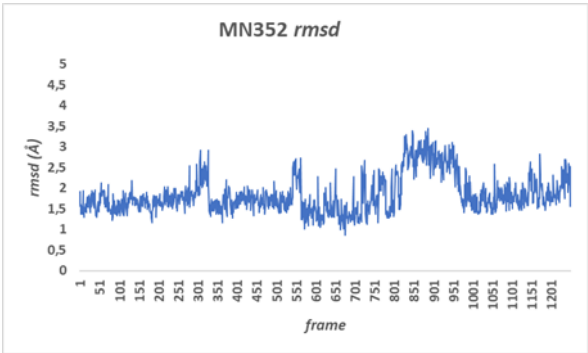


Figure 48. RMSD plot of MN352. Values below or equal to 3Å were observed, indicating a stable and well converged molecular dynamics simulation.

During the simulation strong hydrogen bonds are formed, in particular with residues of Arg96, Arg73, Ser76, Ser75, Thr83, and Asn94, as reported in Table 23.

Table 23. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN352. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	140
Arg73 (side chain)	132
Ser76 (side chain)	94
Thr83 (side chain)	85
Asn94 (main chain)	83
Asn94 (side chain)	83
Ser75 (side chain)	79
Ser74 (main chain)	68
Asn94 (main chain)	33

The trajectory analysis allows the identification of other interactions that involve the amide and phosphate groups of the ligand and water molecules of the solvent.

The estimate of the free energy of binding of MN352 to SOCS2 shows a value of -71.70 kcal/mol. This negative value indicates a strong binding among the ligand and the macromolecule.

4.4.2 MN448

Compound MN448 is characterized by the presence of a pyridine ring and a fluoro-methoxybenzene. The pyridine is positioned into a hydrophobic pocket constituted by Leu150, Ile110, Val148, and Leu95 while the fluoro-methoxybenzene takes contacts with Lys59, Lys63 and Pro92. The phosphate group strongly interacts with residues of Thr83, Arg96, Arg73, Ser75 and Ser76. The amide groups interact with Asn94 and water molecules.

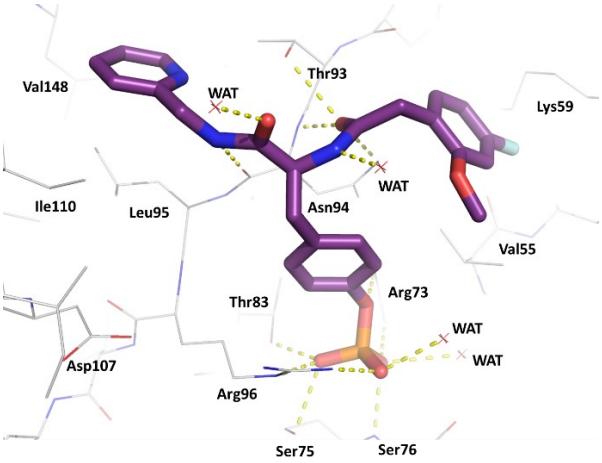


Figure 49. Crystal structure of the complex SOCS2-MN448.

The analysis of the MD simulation shows that the nitrogen of the pyridine interacts with water molecules of the solvent. At the same time, the fluoro-methoxybenzene changes its conformation due to the free rotation of the carbon atom in single bond, establishing interactions with the side chain of Lys59 and water molecules. The amide groups of MN448 are involved in H-bond interactions with Asn94 and water molecules. The phosphate group is involved in hydrogen bond interactions with Arg96 and lose all the other interactions with the binding site.

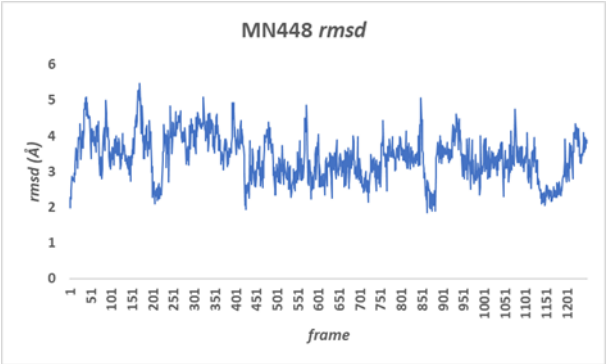


Figure 50. RMSD plot of MN448. Values reach values major than 5Å, indicating that the compound lose its initial conformation during the simulation.

During the simulation, the phosphate group binds water molecules that enter the binding pocket, and this do not allow the interactions of the ligand with Ser75 and Ser76, as it is possible to see in table below.

Table 24. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN448. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	98
Arg73 (side chain)	91
Asn94 (main chain)	76
Thr93 (side chain)	69
Ser76 (side chain)	55
Lys59 (side chain)	32
Ans94 (main chain)	6
Thr93 (side chain)	3
Thr83 (side chain)	2
Ser75 (side chain)	1

The estimate of the free energy of binding of MN448 to SOCS2 shows a value of -28.52 kcal/mol. This value agrees with both the behaviour of the complex during the MD simulation, were the compound loses the hydrogen bond network that characterize its binding mode and the binding affinity value of the compound (12 μM).

4.4.3 MN416

The chloromethyl benzene of MN416 is accommodated into a hydrophobic pocket constituted by Leu150, Leu95, and Val86, while the fluoro-methoxybenzene is in

contact with Lys59, Val55 and Pro92. The amides groups are involved in hydrogen bonds interactions with Thr93 and Asn94. The phosphate group is involved in hydrogen bond interactions with Arg96, Ser75, Ser76 and Arg73.

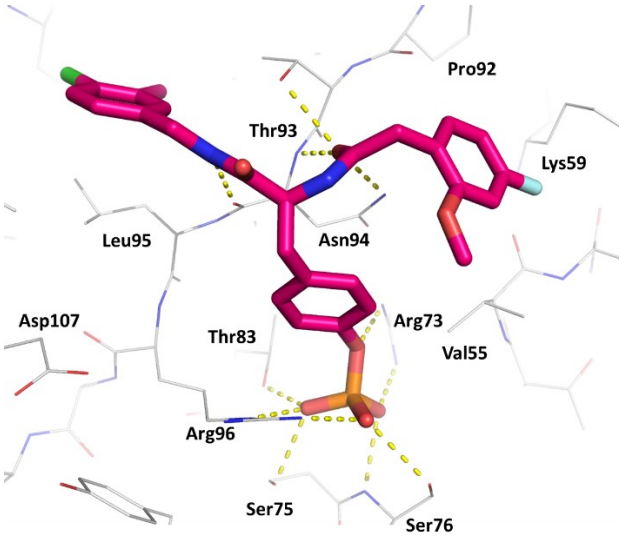


Figure 51. Crystal structure of the complex SOCS2-MN416.

The analysis of the trajectory of the MN416-SOCS2 complex shows that the chloromethyl benzene stably occupies the hydrophobic pocket of SOCS2. The fluoro-methoxybenzene changes its conformation during the simulation and is responsible of the fluctuations in the RMSD plot.

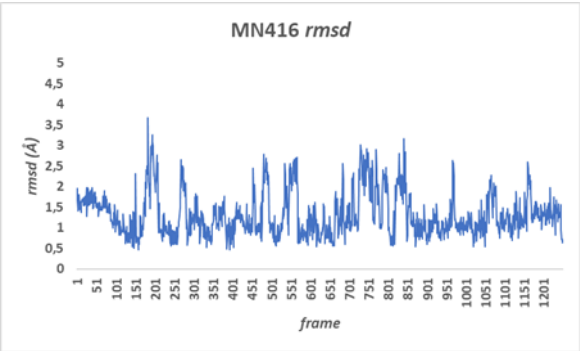


Figure 52. RMSD plot of MN416. Values below or equal to 3Å were observed, indicating a stable and well converged molecular dynamics simulation.

The oxygen atom of the methoxy group is involved in hydrogen bonds interactions with water molecules and Lys59. The amide groups interact with Asn94 and water molecules. The phosphate group stably interacts with Arg96, Ser75, and Ser76 and some water molecules, without to be displaced by the solvent.

Table 25. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN416. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	137
Arg73 (side chain)	121
Ser76 (side chain)	88
Asn94 (side chain)	88
Ser75 (side chain)	76
Ser76 (main chain)	69
Thr93 (side chain)	66
Asn94 (side chain)	39
Asn94 (main chain)	21
Lys59 (side chain)	1

The estimate the free energy of binding of MN416 to SOCS2 shows a value of -74.12 kcal/mol, suggesting a strong affinity between MN416 and the protein.

4.4.4 MN356

The analysis of the crystal structure of the complex SOCS2-MN356 shows that the chloro-fluoro-benzyl ring of the compound is positioned into the hydrophobic pocket constituted by Leu150, Leu95, and Val86. The fluorobenzyl ring takes contact with Pro92, Lys59 and Val55. The amide groups interact with Ans94, Thr93 and a water

molecule. The phosphate group is strongly involved in hydrogen bonds interactions with Thr83, Arg96, Ser75, Ser76 and Arg73.

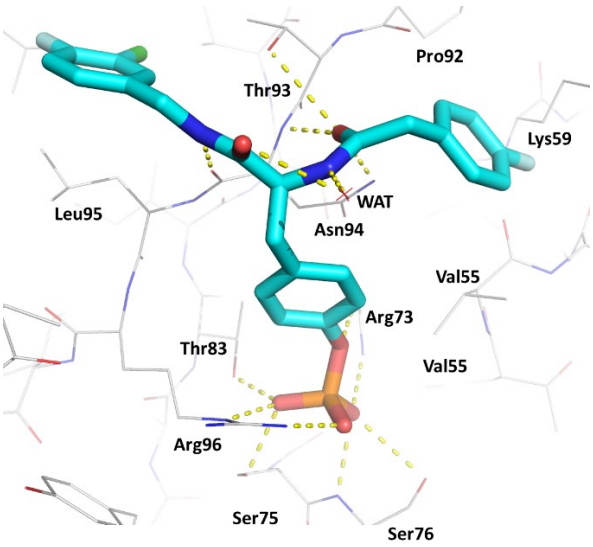


Figure 53 Crystal structure of the complex SOCS2-MN356.

During the simulation, the fluorobenzyl ring changes its conformation causing the movement of the whole molecule, as it is highlighted in the RMSD plot.

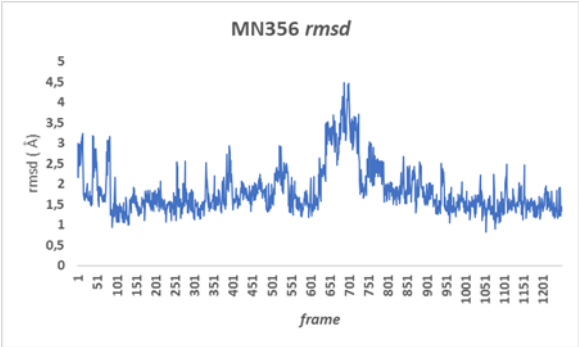


Figure 54. *RMSD plot of MN356. Values that reach 4.5Å show that the molecule changes its conformation during the simulation, especially at the beginning of the simulation and from the frame 650 to 800.*

All the hydrogen bonds are maintained during the simulation, especially between the amide groups and the side chain of Asn94 and between the phosphate group and

Arg96, Ser75, Ser76. Other interactions involve the amides and phosphate groups and water molecules.

Table 26. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN356. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	165
Arg73 (side chain)	113
Ser76 (side chain)	96
Ser75 (side chain)	91
Asn94 (side chain)	86
Ans94 (main chain)	83
Thr83 (side chain)	82
Ser76 (main chain)	57
Asn94 (main chain)	28
Thr93 (side chain)	3
His149 (main chain)	2

The strong interactions between the ligand and the protein is confirmed by the free energy of binding of MN356 to SOCS2, that has a value of -79.73 kcal/mol. This is the most negative value of the whole series of compounds here analysed.

4.4.5 MN357

The crystal structure of the complex SOCS2-MN357 shows that the bromo-fluoro-benzyl ring is in contacts with residues of Leu150, Leu95, and Val86 while the fluorobenzyl ring is in contact with Lys59, Pro92, Val55. The amide groups are

involved in H-bond interactions with Asn94, Thr93 and a structural water. The phosphate group interacts with Thr83, Arg96, Ser75, Ser76, Arg73.

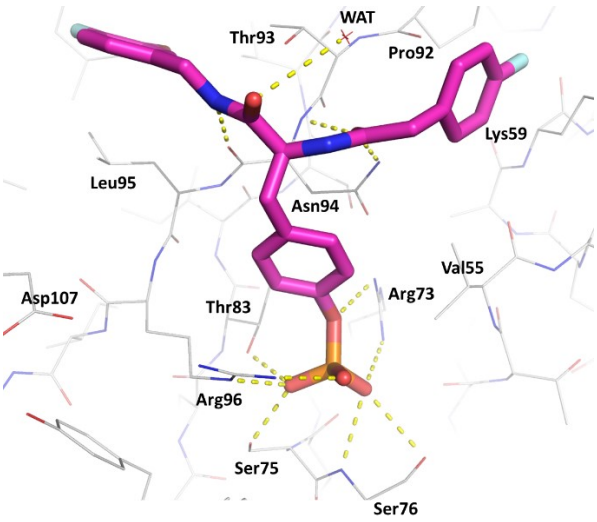


Figure 55. Crystal structure of the complex SOCS2-MN357.

The analysis of the MD simulation shows that the ligand stably interacts with the pY binding site. The only conformational variation concerns the fluorobenzyl ring that rotates together with the carbon atom. All the hydrogen bonds previously analysed strongly anchor the ligand to the binding pocket.

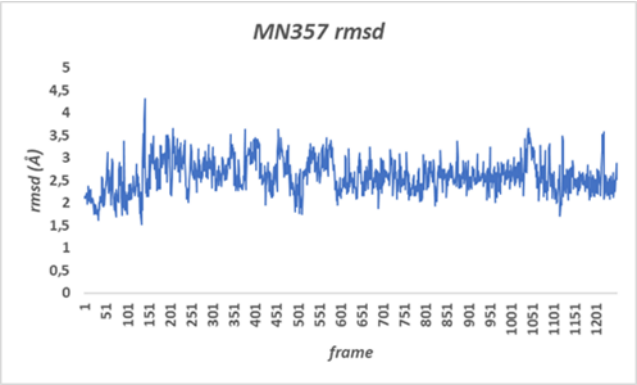


Figure 56. RMSD plot of MN357. Values below or equal to 4Å were observed, indicating a stable and well converged molecular dynamics simulation.

The rate of occupancy of hydrogen bonds is reported in the table below.

Table 27. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN357. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Asn73 (side chain)	138
Arg96 (side chain)	112
Ser76 (side chain)	91
Asn94 (main chain)	80
Asn94 (side chain)	75
Thr83 (side chain)	67
Ser47 (main chain)	65
Ser75 (side chain)	62
Asn94 (main chain)	29
Thr93 (side chain)	6
Lys59 (side chain)	4

The analysis of the free energy of binding of MN356 to the macromolecule shows a value of -70.73, indicating the formation of a stable complex.

4.4.6 MN366

The analysis of SOCS2-MN366 complex shows that the difluoro benzyl ring of MN366 establishes hydrophobic interactions with Leu150, Leu95, and Val86 while the fluorobenzyl ring is in contact with Pro92, Lys59 and Val55. The amide groups interact with Ans94 and Thr93. The phosphate group is instead involved in hydrogen bond interactions with Thr83, Arg96, Ser75, Ser76, Arg73 and a water molecule.

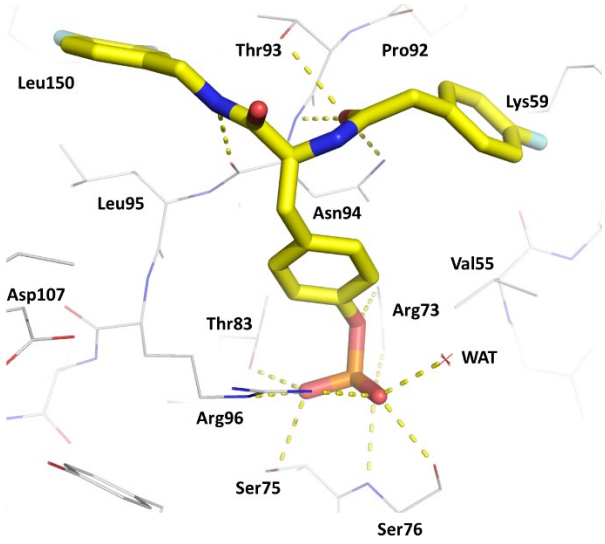


Figure 57. Crystal structure of the complex SOCS2-MN366.

During the simulation both the difluoro benzyl ring and the fluorobenzyl ring change their conformation and this movement is responsible for the fluctuations in the RMSD plot.

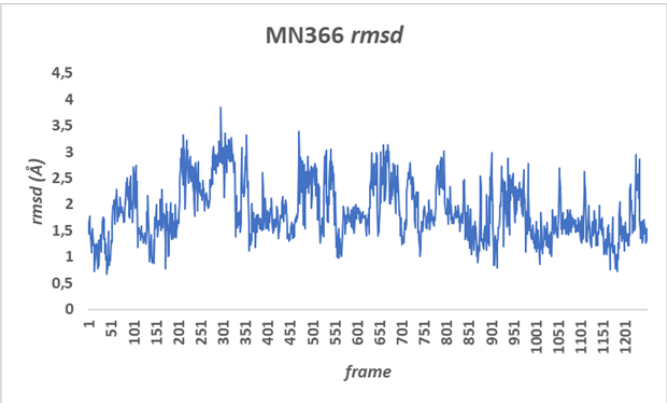


Figure 58. RMSD plot of MN366. This plot evidence that during the simulation the ligand moves into the binding site. Nevertheless, the compound is able to maintain the fundamental interactions with the binding site.

Both the amide groups and the phosphate group establish hydrogen bond interactions also with water molecules. However, the fundamental interactions that characterise the binding mode of MN366 are maintained for the whole simulation, as it is possible to see in the table below.

Table 28. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN366. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	125
Arg73 (side chain)	104
Ser76 (side chain)	93
Ser75 (side chain)	85
Asn94 (main chain)	80
Thr83 (side chain)	77
Asn94 (side chain)	77
Ser76 (main chain)	63
Thr93 (side chain)	31
Asn94 (main chain)	14

The estimate the free energy of binding of MN366 to SOCS2 shows a value of -70.93 kcal/mol, suggesting a strong affinity between MN366 and the protein.

4.4.7 MN367

The analysis of the SOCS2-MN356 complex shows that the dichloro-benzyl ring is positioned into the hydrophobic pocket constituted by Leu150, Leu95, and Val86 while the fluorobenzyl ring is accommodated into the pocket constituted by Pro92, Lys59 and Val55. Amide groups are involved in hydrogen bond interactions with Thr93, Asn94 and a structural water molecule. The phosphate group forms hydrogen bond with Thr83, Arg96, Ser75, Ser76, Arg73 and a water molecule.

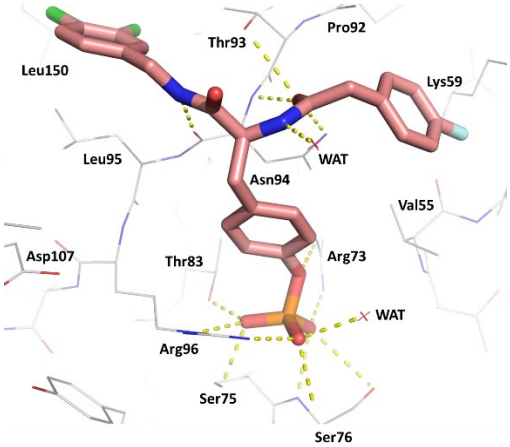


Figure 59. Crystal structure of the complex SOCS2-MN367.

The analysis of the MD simulations shows that the fluorobenzyl ring, that is exposed to the solvent, is subjected to the variation of its position due to the rotation of the carbon atom of the side chain, causing the fluctuation in the RMSD plot.

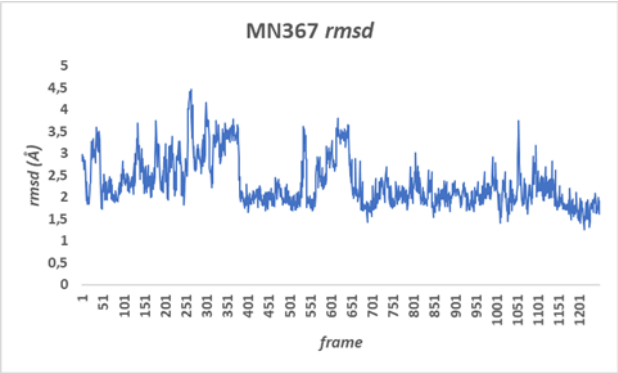


Figure 60. RMSD plot of MN367. During the simulation the compound interacts with water molecules and lose the interactions between the phosphate group and Ser75 and Ser76. Nevertheless, the compound is able to bind the pY pocket.

The amide groups are involved in H-bond interactions with Asn94 and the water molecules. The phosphate group partially lose the interactions that have been previously analysed. Water molecules enter the binding site of the ligand, and the phosphate group is no more able to interact with Ser75 and Ser76. This finding has been confirmed by the analysis of the occupancy of hydrogen bonds that is reported in Table 29.

Table 29. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN367. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg73 (side chain)	109
Arg96 (side chain)	79
Ans94 (main chain)	61
Ser76 (side chain)	57
Asn94 (side chain)	50
Asn94 (main chain)	43
Lys59 (side chain)	29
Thr83 (side chain)	12
Ser75 (side chain)	12
Ser76 (main chain)	10

The analysis of the free energy of binding of MN367 to SOCS2 shows a value of -51.90 kcal/mol, indicating a less strong binding.

4.4.8 MN368

The trifluoro-benzyl ring of MN368 is involved in hydrophobic interactions with Leu150, Leu95, and Val86 while the fluorobenzyl ring take contacts with Pro92, Lys59 and Val55. The amide groups interact with Asn94 and Thr92, while the phosphate group is involved in hydrogen bond interactions with Thr83, Arg96, Ser75, Ser76, Arg73, and a water molecule.

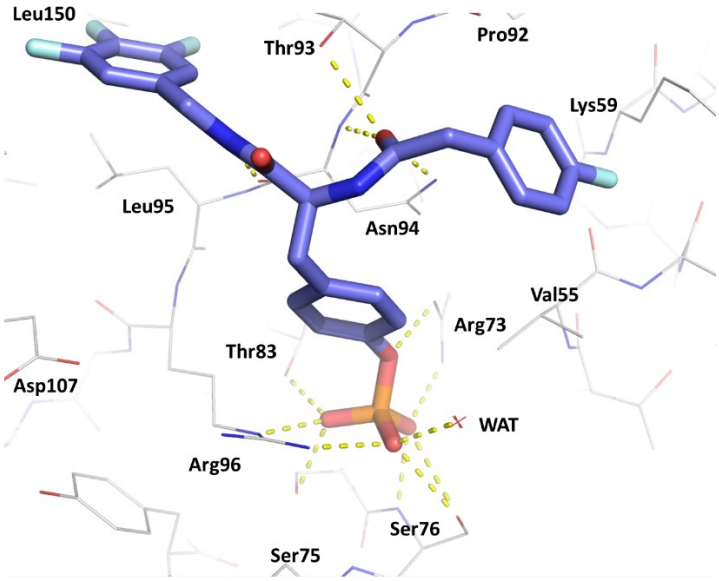


Figure 61. Crystal structure of the complex SOCS2-MN368.

During the MD simulation, the trifluoro-benzyl and the fluorobenzyl rings change their conformation, causing the fluctuation of the RMSD value reported in the plot.

After this initial movement, the molecule results to be stably bound to the amino acid residues of the pY pocket, as it is reported in the occupancy table of the MN368-SOCS2 complex.

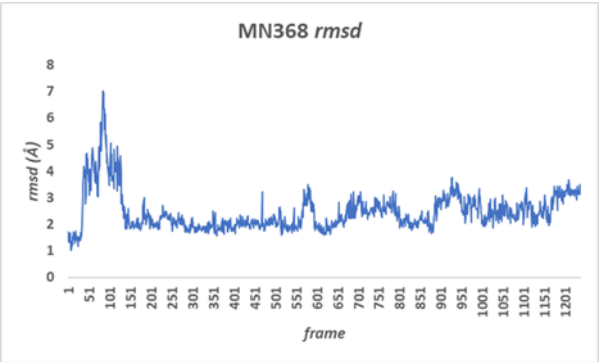


Figure 62. RMSD plot of MN368. At the beginning of the simulation, and until the frame 150, there is a movement of the molecule into the binding site. Then, the RMSD values reach value that indicate the normal movement of the molecule into the binding site.

During the simulation, both the amide groups and the phosphate group interact with water molecules.

Table 30. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN368. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	135
Arg73 (side chain)	108
Ser76 (side chain)	89
Ser75 (side chain)	88
Asn94 (main chain)	78
Asn94 (side chain)	68
Thr83 (side chain)	65
Ser76 (main chain)	64
Thr93 (side chain)	26
Asn94 (main chain)	21

The analysis of the free energy of binding of MN368 to SOCS2 shows a value of -66.87 kcal/mol.

4.4.9 MN402

The chloromethylbenzyl ring of MN402 takes hydrophobic contacts with Leu150, Leu95, and Val86. The fluorobenzyl ring is positioned into the pocket constituted by Pro92, Lys59 and Val55. The amide groups are involved in hydrogen bond interactions with Thr93, Asn94 and a structural water. The phosphate group binds Thr83, Arg96, Ser75, Ser76 and Arg73.

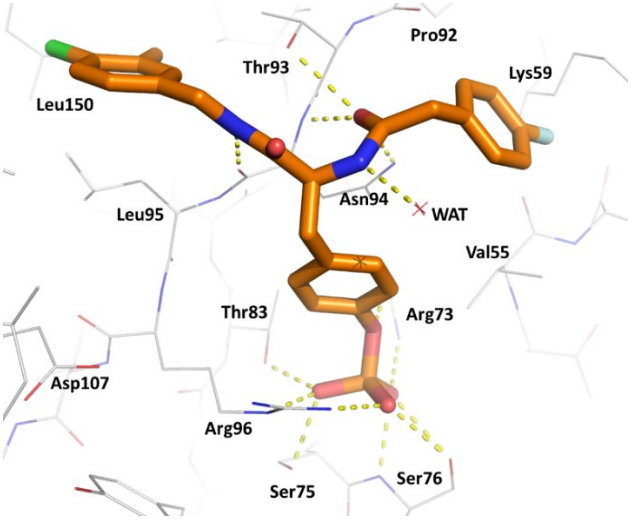


Figure 63. Crystal structure of the complex SOCS2-MN402.

During the MD simulation, the fluorobenzyl ring changes its conformation due to the rotation of the carbon atom. The amide groups are involved in hydrogen bond interactions with Asn94 and water molecules.

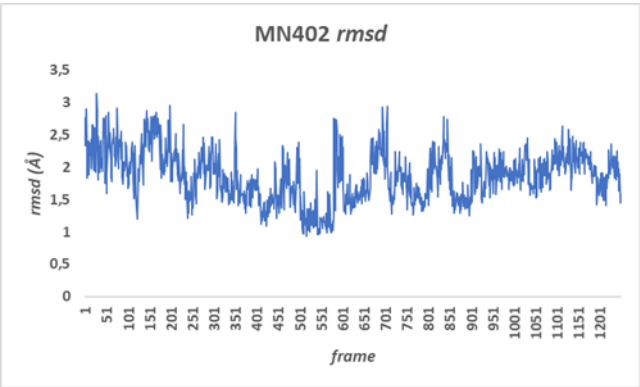


Figure 64. RMSD plot of MN402. Values below or equal to 3Å were observed, indicating a stable and well converged molecular dynamics simulation.

The phosphate group establish long-lasting interactions with Arg96, Ser75, and Ser76. Moreover, the phosphate group interacts with water molecules of the solvent.

Table 31. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN402. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	138
Arg73 (side chain)	118
Ser75 (side chain)	92
Ser76 (side chain)	88
Ser76 (main chain)	77
Asn94 (main chain)	69
Asn94 (side chain)	69
Thr83 (side chain)	56
Asn94 (main chain)	35
Thr93 (side chain)	22

The estimate the free energy of binding of MN402 to SOCS2 shows a value of -71.95 kcal/mol, suggesting a strong affinity between MN402 and the protein.

4.4.10 MN403

The fluorobenzyl alcohol of MN403 is accommodated into the pocket constituted by Leu150, Leu95, and Val86. In particular, the oxygen atom interacts with the side chain of Leu150 and with a structural water molecule. The fluorobenzyl ring takes contacts with Pro92, Lys59 and Val55. The amides groups are involved in hydrogen bond interactions with Thr93 and Asn94. The phosphate group interacts with Arg96, Ser75, Ser76, and Arg73.

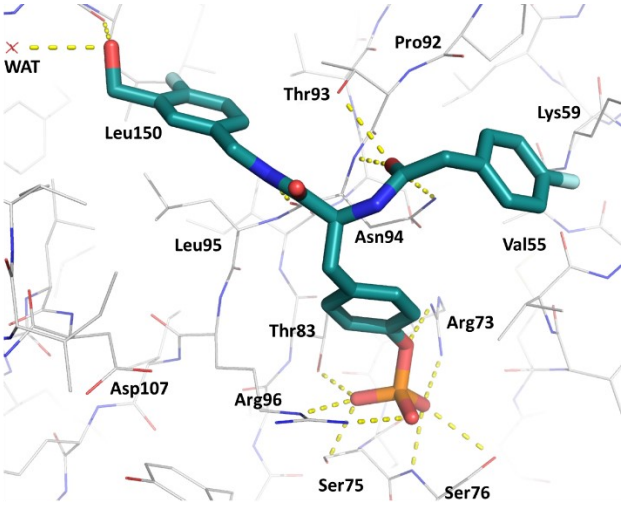


Figure 65. Crystal structure of the complex SOCS2-MN403.

During the simulation the alcohol group interacts with His249 and with the side chain of Leu150. Other interactions are established with water molecules. The movement of the fluorobenzyl ring is responsible for the fluctuations in the RMSD plot.

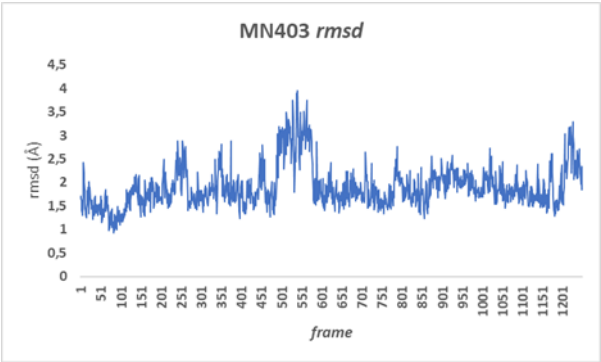


Figure 66. RMSD plot of MN403. The RMSD values indicate a compound that is stable into the binding site. The increment in the RMSD value at the frames 500-600 is due to the movement of the fluorobenzyl ring that is solvent exposed and changes its conformation.

Water molecules of the solvent interact with the amide and phosphate groups. The rate of occupancy of hydrogen bond for the complex SOCS2-MN403 is reported in Table 32.

Table 32. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN403. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	166
Arg73 (side chain)	119
Ser76 (side chain)	96
Ser75 (side chain)	90
Thr83 (side chain)	86
Asn94 (main chain)	83
Asn94 (side chain)	81
Ser76 (main chain)	62
Asn94 (main chain)	29
Thr93 (side chain)	6
His149 (main chain)	2

The estimate free energy of binding of MN403 to SOCS2 shows a value of -76.93 kcal/mol, indicating a strong affinity between MN403 and the protein.

4.4.11 MN442

Analysing the SOCS2-MN442 complex it is possible to see that the chloromethyl-benzene is accommodated into the hydrophobic pocket constituted by Leu150, Leu95, and Val86. The methoxy-fluorobenzyl alcohol interacts with the Lys59. The amide groups interact with Thr93, Asn94 and a water molecule. The phosphate group is involved in hydrogen bonds with Arg96, Ser75, Ser76, Arg73 and water molecules.

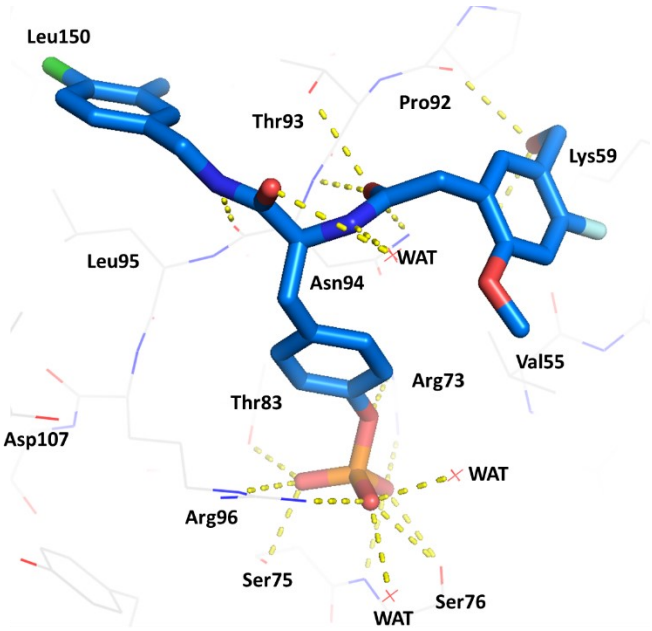


Figure 67. Crystal structure of the complex SOCS2-MN442.

During the simulation, the methoxy-fluorobenzyl alcohol interacts with the Lys59 and water molecules. Since it is exposed to the solved, it changes its conformation, causing the fluctuation in the RMSD plot.

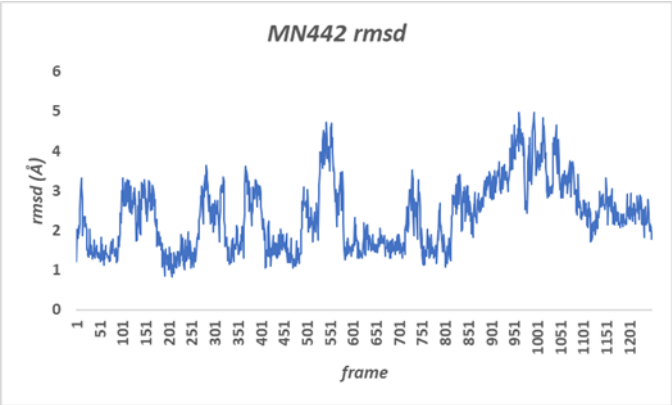


Figure 68. RMSD plot of MN442. Fluctuation in the RMSD plot are due to the methoxy-fluorobenzyl alcohol that interacts with water molecules of the solvent and rotates into the binding site.

The amide groups stably interact with Thr93 and Asn94, and other hydrogen bonds are formed among these groups and water molecules. The phosphate group forms hydrogen bonds with Arg73, Arg96, Ser75, Ser76 and water molecules.

Table 33. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN442. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg73 (side chain)	134
Arg96 (side chain)	117
Ser76 (side chain)	97
Asn94 (main chain)	80
Thr83 (side chain)	76
Ser75 (side chain)	76
Ser76 (main chain)	74
Asn94 (side chain)	70
Asn94 (main chain)	28
Pro63 (main chain)	12
Lys59 (side chain)	9
Asn94 (side chain)	7
Val55 (main chain)	2
Asn56 (side chain)	2

The analysis of the estimate the free energy of binding indicates that the complex SOCS2-MN403 is characterized by a strong affinity, with a MMGBSA value of -71.01 kcal/mol.

4.5 DISCUSSION

The crystal structures of the complexes constituted by SOCS2 and several binders have been analysed through the Molecular Dynamics simulation with the aim of establish a SAR.

MN352 (K_d = 2.6 μM), which presents two fluorobenzyl rings, is able to interact stably with the binding site, establishing long-lasting interactions with the pY pocket. The introduction of a chlorine atom in MN356 in one of the two benzyl rings is well tolerated, leading to improved activity (K_d = 0.81 μM). The complex constituted by SOCS2 and MN356 is characterized by the most negative value of free energy of binding, indicating that this is the most stable complex among those here analysed. The replacement of the chlorine atom with the bromine one does not affect the binding affinity (MN357, K_d=0.78 μM) and the stability of the complex (ΔG = -70.93 kcal/mol). The introduction of a dichlorobenzyl group does not impair the binding affinity of the compound (MN367, K_d= 1.0 μM), but the MD simulation highlight thatthe movement of this ring together with the one of the fluorobenzyl ring allows the water molecules to enter the pY binding site, with the phosphate group that partially lose its interactions. This explains the more positive value of predicted binding affinity of the compound that shows a value of -51.90 kcal/mol.

Also the introduction of a trifluorobenzyl ring does not significantly affect the activity of the compound (MN368, K_d=1.2 μM), but the affinity for the macromolecule slowly decreases. Instead, the introduction of an alcohol group in MN403 led to a diminution of the activity.

A little improvement of the K_d value is obtained substituting the fluorine atom of MN356 with a methyl group. MN402 shows a K_d value of 0.35 μM and a negative ΔG, indicating the formation of a stable protein-binder complex.

The substitution of the fluorobenzyl ring of MN402 with a fluoro-methoxybenzene is well tolerated (MN416, K_d=0.78 μM), as confirmed by the MD simulations studies and MMGBSA analysis, with the oxygen atom that can interact with Lys59. In addition, the introduction of the methoxy-fluorobenzyl alcohol is well tolerated (MN442, K_d=1.1 μM) and forms a stable complex with SOCS2.

MN448 (K_d=12 μM) is constituted by a fluoro-methoxybenzene and a pyridine. The presence of polar groups in both rings allows the interaction with water molecules that partially displace the molecule from the binding site. The movement of the whole molecule allows the water to enter the binding site, whit the compound that is no

more able to interact with the fundamental amino acid of the pY pocket. This is translated into a low affinity for SOCS2.

To further understand which of the two rings of MN448 influences more the behaviour of MN448, two derivatives have been designed.

4.5.1 MN448-a

The fluorobenzyl ring is accommodated into a hydrophobic pocket and takes contacts with Leu150, Ile 110, Val148, and Leu 95. The fluoro-methoxybenzene is in contact with Pro92, Lys59, and Val55. The amide groups are strongly involved in H-bond interactions with Asn94 and Thr93. The phosphate group binds Ser75, Ser76, Arg73, Leu82, and Thr83.

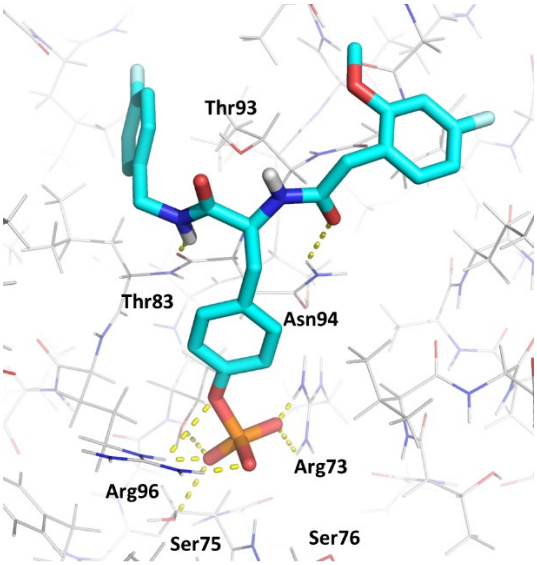


Figure 69. Binding mode of MN448-a. ($K_{d\,pred}$= 3.90 μM)

The fluorobenzyl ring is accommodated into a hydrophobic pocket and takes contacts with Leu150, Ile 110, Val148, and Leu 95. The fluoro-methoxybenzene is in contact with Pro92, Lys59, and Val55. The amide groups are strongly involved in H-bond interactions with Asn94 and Thr93. The phosphate group binds Ser75, Ser76, Arg73, Leu82, and Thr83.

During the simulation the methoxy group interacts with the side chain of Lys59 and with water molecules. Moreover, this ring rotates around the carbon atom, raising the RMSD to values of 6Å.

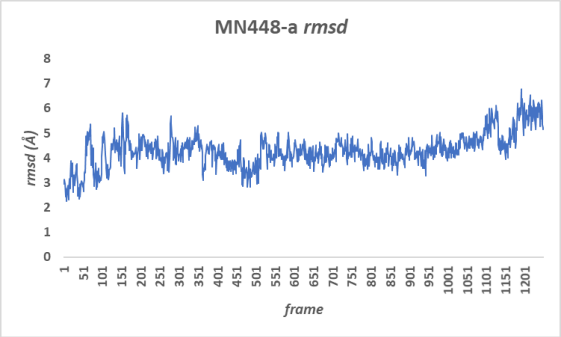


Figure 70. RMSD plot of MN448-a. The RMSD reaches values of 6Å. Despite the great value of RMSD, the compound stably interacts with the pY pocket, especially with the phosphate and amide groups.

The amide group strongly interacts with the side chain of Asn94, Thr93 and water molecules. Also the phosphate group stably interacts with the residues of the pY pocket and with the water molecules.

Table 34. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN448-a. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Asn96 (side chain)	149
Arg73 (side chain)	123
Ser76 (side chain)	94
Ser75 (side chain)	87
Asn94 (side chain)	78
Thr83 (side chain)	76
Ser76 (main chain)	66

Asn94 (main chain)	57
Lys59(side chain)	4
Thr93 (side chain)	3

The estimation of the free energy of binding indicates that the complex SOCS2-MN448-a has a value of -69.36 kcal/mol, indicating a strong affinity between the molecule and the protein.

4.5.2 MN448-b

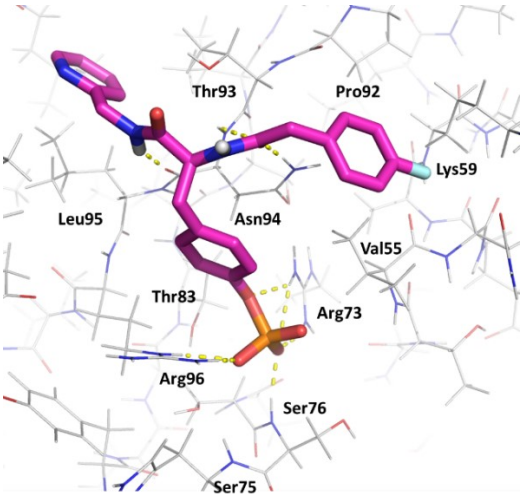


Figure 71. Binding mode of MN448-b. ($k_{d\ pred}$ = 2.40 μM)

The analysis of the binding mode of MN448-b shows that the pyridine ring is accommodated into the hydrophobic pocket constituted by Leu150, Ile 110, Val148, and Leu 95 while the fluorobenzyl ring is in contact with Pro92 and Lys59. The amide groups are involved in hydrogen bonds interactions with Asn94 and Thr93. The phosphate group interacts with Arg96, Ser75, Ser76 and Arg73.

During the MD simulation, the nitrogen of the pyridine ring forms hydrogen bonds with water molecules of the solvent, while the fluorobenzyl ring changes its position.

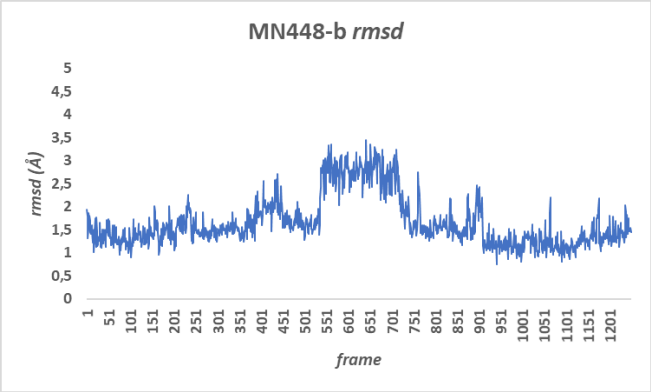


Figure 72. RMSD plot of MN448-b. Values below or equal to 3Å were observed, indicating that the pyridine ring is well tolerated than the fluoro-methoxybenzene of MN448-b.

Amide and phosphate groups establish long-lasting interactions with the amino acids into the pY pocket and interact also with water molecules of the solvent.

Table 35. Rate of occupancy of hydrogen bond interactions for the complex SOCS2-MN448-b. A high rate of occupancy indicates a great number of interactions occurred during the simulation.

Residue Number	Occupancy (%)
Arg96 (side chain)	148
Arg73 (side chain)	126
Ser76 (side chain)	95
Ser75 (side chain)	86
Asn94 (side chain)	85
Thr83 (side chain)	84
Asn94 (main chain)	82
Ser76 (main chain)	68
Asn94 (main chain)	25
Thr93 (side chain)	4

The analysis of the predicted binding affinity between SOCS2 and MN448-b shows a value of -71.20 kcal/mol.

The prediction of the binding affinity of MN448-a and MN448-b shows that the pyridine ring is well tolerated than the fluoro-methoxybenzene. This is confirmed both by the analysis of the MD trajectories, which show that the fluctuation of the fluoro-methoxybenzene causes the raise of the RMSD to 6Å, but also of the predicted ΔG, which shows that MN448-b is more stable in the binding site.

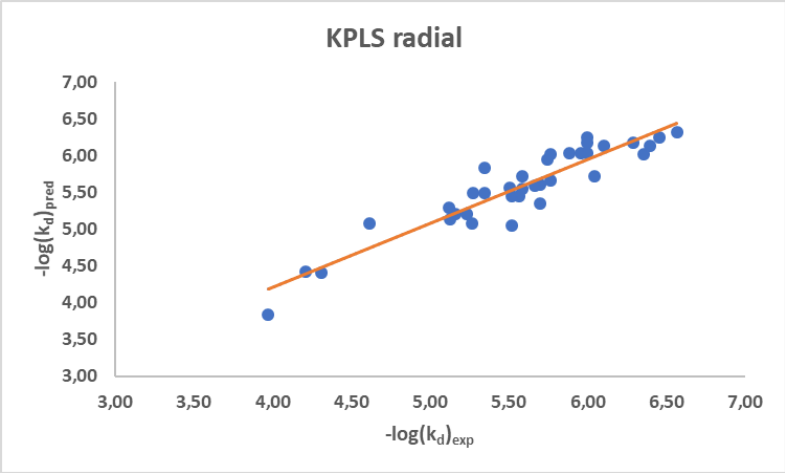
However, the behaviour of MN448 could be addressed to the presence of hydrophilic groups in both the aromatic rings. The interaction of these groups with the solvent could lead to the displacement of the compound from the binding pocket, resulting in a loss of activity.

4.6 DEVELOPMENT OF A QSAR MODEL

The prediction of the binding affinities of MN448-a and MN448-b have been carried out by developing a a QSAR model, using all the available data of SOCS2 binders and the AutoQSAR tool [110, 111] available in the Schrödinger suite. The dataset that has been used to build the model is composed of 49 compounds with binding affinities known as previously measured experimentally using biophysical methods in the Ciulli Lab. The set has been split randomly into a Training Set, constituted by the 75% of the compounds, and a Test Set, which included the 25% of the compounds.

The binding affinities of the learning set are reported as $pIC_{50}=-\log(K_d)$.

The best model obtained using the KPLS machine learning method showed a $r^2=0.86$ and $q^2= 0.77$. The model was obtained with molprint2D fingerprints.



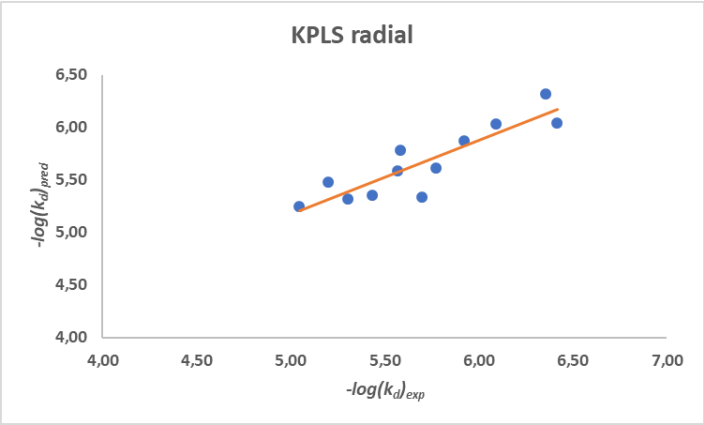
4.6.1 EXTERNAL VALIDATION SET:

To externally validate the model, 11 compounds with known binding affinities have been used. In the table below are reported the experimental and predicted activities of the compounds.

Table 36. Experimental versus Predicted binding affinities of the compounds.

Compound ID	$-\log(K_d)_{exp}$	$-\log(K_d)_{pred}$
MN416	6.10	5.98
MN441	5.28	5.30
MN69	5.28	5.19
MN408	5.46	5.66
MN171	5.56	5.25
MN405	5.62	6.11
MN484	5.74	6.15
MN406	5.82	6.04
MN370	6.00	5.82
MN448	4.92	5.05

Compound ID	$-\log(K_d)_{exp}$	$-\log(K_d)_{pred}$
MN482	6.36	6.06



The validated model has been used to predict the binding affinities of MN448-a, and MN448-b reported 4.5.1 and 4.5.2.

4.7 CONCLUSIONS

SOCS2 protein is an essential regulator of cellular responses to cytokines. It regulates several signalling pathways, and is involved in the ubiquitination of target protein, such as the Growth Hormone Receptor (GHR) and several signalling proteins. It is involved in the modulation of the innate and adaptive immune responses modulating the NF-kB pathway, while its downregulation is linked to the development of breast, lung, liver and ovarian cancer. . For all these reasons, SOCS2 represent an attractive target for the development of molecules able to inhibit its functions. Since SOCS2 is a component of an E3 ubiquitin ligase complex, a molecule able to bind this protein can potentially be used to develop a PROTAC, inducing the degradation of POI involved in the development of several pathologies. Some computational approaches have been applied on SOCS2-binders complexes previously identified, with the aim to establish a SAR that will be taken in

consideration to guide the synthesis of new derivatives with improved activities and pharmacokinetics proprieties. The analysis of the binding mode of the compounds, together with the post-processing analysis of the MD simulation, allowed to identify the amino acidic residues that are important for binding affinity and specificity. The analysis of MN448 together with Mol1 and Mol2 allowed understanding that the fluoro-methoxybenzene destabilize the binding mode of the compounds with respect to the pyridine ring. The introduction of two polar groups on both the aromatic rings of MN448 causes its displacement from the active site during the simulation, resulting in a loss of activity of the molecule. The development of validated QSAR model will allow the prediction of the binding affinities of new molecules to prioritise for the synthesis.

4.8 Materials and Methods

Protein and Ligands Preparation

Proteins were pre-treated with the Protein Preparation Wizard [113, 114] tool of Maestro 12.3 [115] suite, by assigning bond order, adding hydrogens and capping termini. In the refine module, the H-bond assignment procedure was performed using exhaustive sampling and sample water orientation options marked. Missing loops were filled using Prime.

Ligands were designed by means of Maestro (version 2020.1) graphical interface [115] and pre-treated with LigPrep [116] tools at default settings, checking the ionization state at physiological pH with Epik. [117, 118, 119]

Docking Studies

Compounds were docked using the GLIDE [120, 121, 122, 123] (version 86013) docking tools, with the SP (standard-precision) procedure, allowing a flexible ligand sampling and performing a post-docking minimization. All other settings have been left at their default values. The best three poses have been included and analysed.

Molecular Dynamics studies

All complexes have been subjected to molecular dynamics (MD) simulation studies using NAMD [124] software. The protein was described using the AMBER18 ff14SB forcefield. Parametrization of ligands was performed using antechamber 19.0 module of Amber 18. [125, 126] GAFF (Generalized Amber Force Field) [126] force field has been used to assign the atom type to the molecules. Charges were assigned using AM1-BCC method. [127] Complexes were solvated with a 11 Å octahedral box of TIP3P water molecules, and Cl^- ions were added to neutralize the system. Minimization and heating of the system have been performed using NAMD for 250,000 steps. The temperature was increased from 0 to 300 K, using Langevin dynamics and keeping constant the volume. [128] The equilibration of the system was performed for 250,000 steps, at constant temperature and pressure. The

Berendsen barostat [129] was used to control the pressure. Finally, were performed 50 ns of MD simulations at the temperature of 300K and at the pressure of 1 atm. The SHAKE algorithm was used to treat hydrogen-containing bonds. The whole simulation has been performed in Periodic Boundary Condition (PBC). VMD [130] was used to analyse the trajectories, calculate the occupancy of hydrogen bonds and to perform RMSD analysis.

MMGBSA analysis

The analysis of the stability of each SOCS2-ligand complex was carried out through the application of the MMPB(GB)SA method, available in AmberTools [131] The MMPBSA.py python script allows calculating the energy difference between the bound and unbound state of the complex, using the topology files of the complex, of the desolvated protein and ligand, of the solvated complex and of the production file dynamics. MMGB(PB)SA analysis was performed on a trajectory obtained through the cpptraj module of AmberTools that has been placed in saline solution with a molarity of 0.150M to reproduce the physiological conditions.

QSAR model

A total of 49 compounds with known activities have been used to build a QSAR model using the AutoQSAR tool available in the Schrödinger suite. [110, 111] The activities of the molecules are reported as pIC50 = -log(K_d).

Molecules were imported in the entry list, choosing to build a regression model in the AutoQSAR panel. The learning set was randomly split in training and test set, including in the first one 75% of the compounds, and in the second one 25%.

The best pre-trained model has been cross-validated. Finally, the model was used to predict an external set of compounds.

4.8 THE ELONGIN C PROTEIN

Cullins are a family of proteins that are conserved throughout bacteria, plants, and mammals. [132] Six cullin proteins (CUL1, 2, 3, 4A, and 4B) belong to this family [133] and are characterised by the presence of a carboxy-terminal cullin homology (CH) domain. The human genome encodes for other two atypical cullin proteins, named CUL7 and CUL9, constituted by two additional homology domains: the DOC1 (destruction of cyclin B) and the IBR (in-between ring). [133] Cullin proteins form the Cullin-RING ubiquitin ligases (CRLs) complexes by assembling with RING-box protein 1 (RBX1, also known as ROC1) or RBX2 (also known as ROC2 or SAG-sensitive to apoptosis gene). [134]

CRLs are therefore constituted by cullins, RINGs, adaptor proteins, and substrate recognition receptors. [135] The two RING-box components (RBX1 and RBX2) can bind two zinc atoms via a $C_3H_2C_3$ motif to form the RING finger domain required for the activity of CRLs. [136] Whit the except for CUL9, all the other cullin proteins interact with RBX1 or RBX2 through an evolutionarily conserved CH domain at the C-terminus. [135] The human genome also contains four adaptor proteins: the S-phase kinase-associated protein 1 (Skp1), Elongin B and C, and damaged DNA binding protein 1 (DDB1). [137] While Skp1 interacts with CUL1 and CUL7, Elongin B and C interact with CUL2 and CUL5 and DDB1 with CUL4A and CUL4B. [137]

CRLs are E3 ubiquitin ligases responsible for the ubiquitination of about 20% of intracellular proteins through the ubiquitin-proteasome system (UPS). [135] Protein degradation by the UPS happens through the ubiquitination step that adds mono- or poly-Ubiquitin tag (Ub-tag) to the targeted proteins and the proteasomal degradation of Ub-tagged proteins. Cancer cells use UPS to increase their growth and decrease apoptotic cell death. [138] Hence, cullin proteins play critical roles in cancer through various mechanisms, including DNA replication and repair, cell cycle control, apoptosis, oxidative stress, chromatin remodelling, hypoxia, regulation of oncogenes, and signal transduction involved in p53 and WNT.

Table 37. Intracellular pathways deregulated by Cullin proteins.

Cullin protein	Process Deregulated	Cellular Pathway
Cul1	Cell Cycle	p21, p27, cyclin D, cyclin E
	Signal Transduction	B-catenin, Notch, p53
	Gene Transcription	IkBα, c-Myc, c-Jun
Cul2	Hypoxia	HIF1α
Cul3	Apoptosis	Caspase-8
	Transduction	P53
	Oxidative Stress	Nrf2
Cul4A/B	Chromatin Remodeling	Histone 3 and 4
	DNA Repair	XPC
	DNA Replication	CDT1
Cul5	Signal Transduction	p53
Cul7	Signal Transduction	P53
	Gene Transcription	c-Myc

Cullin2 (Cul2) is the central scaffold of the VHL ligase that recruits the von Hippel-Lindau protein (pVHL) through an adaptor subunit constituted by Elongin B (EloB) and Elongin C (EloC) and by the C-terminal domain of the RING box protein, Rbx1.

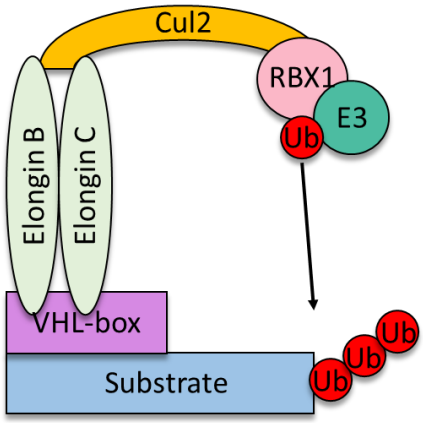


Figure 73. Cullin2 protein recruits the adaptors Elongin B and Elongin C, the receptor protein VHL-box and RING protein to form the CRL E3 ligase complex that promote ubiquitin to transfer from RBX1 to a protein substrate.

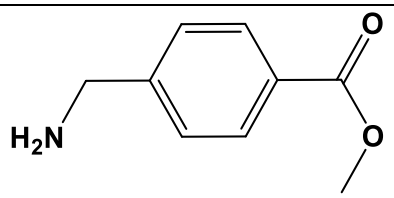
E3 ubiquitin ligases catalyse the ubiquitin transfer from an E2-conjugating enzyme to substrate proteins, inducing proteasomal degradation. CRLs have been targeted using small molecules, [133] and CRL binders can be converted into targeted protein degraders by conjugating them to a ligand of a protein of interest.

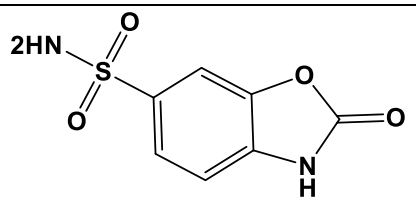
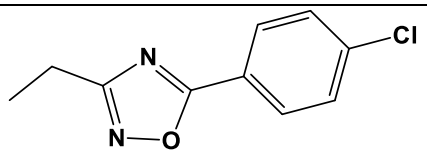
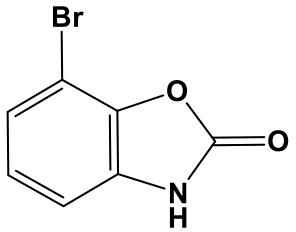
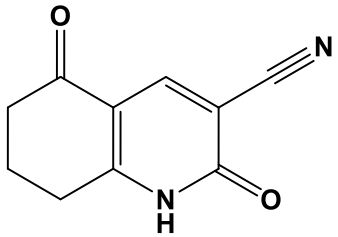
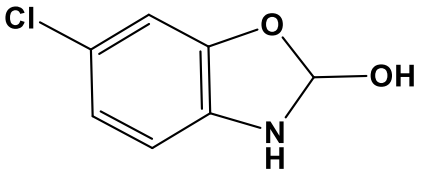
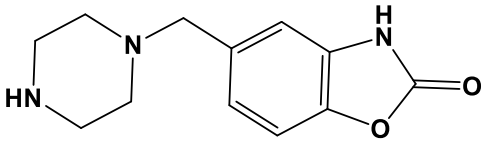
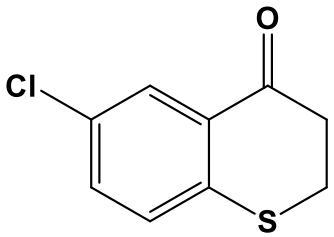
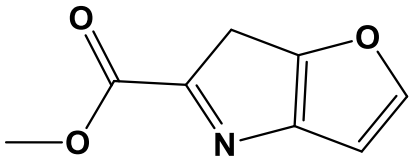
The significant roles of CRLs in several biological processes and human diseases have emerged in cancer, where the genes encoding for E3 ligase subunits and their native substrates are often found as oncogenes or tumour suppressors.

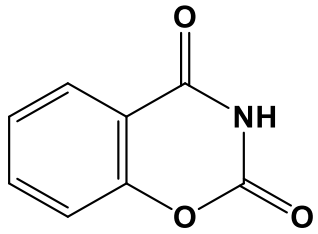
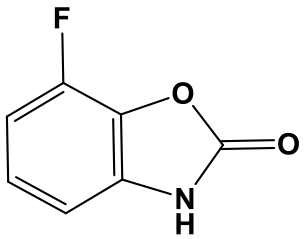
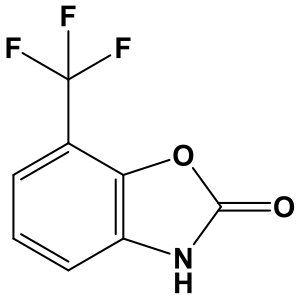
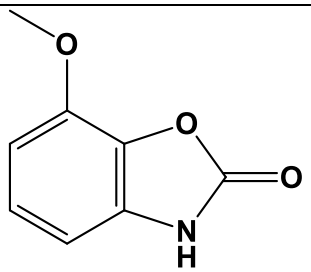
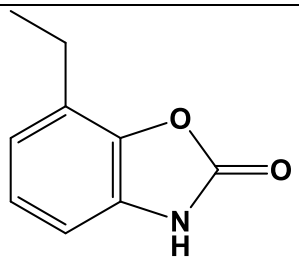
To develop potential anti-cancer agents, it is important to develop CRL inhibitors by targeting cullin neddylation and CRL-mediated ubiquitination.

A fragment screen was carried out in the Professor Ciulli laboratory against the receptor-adaptor trimeric subunit composed by pVHL, EloB and EloC (VBC) this led to the discovery of fragments able to bind a cavity at the EloC:Cul interface that has been called EloC pocket. [139] and that is constituted by residues of Glu64, Ile65, Pro66, Glu102, Met105, Ala106, and Phe109. While the VBC system was studied in this case, EloC is also the substrate adapter in CUL5 E3 ligase assemblies such as the SOCS family of proteins. This increases the application of any optimised chemical ligands and probes binding at this pocket. The analysis of the co-crystallised ligands shows that the aromatic portion of the compounds forms hydrophobic interactions with the pocket, while the carbonyl carbon of the fragments engages in a C═O···C═O contact with the backbone carbonyl oxygen of Glu64. Starting from these findings, novel fragments have been designed, and their structures are reported in the table below.

Table 38. Structure and binding affinity of fragments able to bind the EloC pocket.

Compound ID	Structure	Dissociation constant (Kd)
AT18		n.d.

Compound ID	Structure	Dissociation constant (Kd)
EN4		n.d.
F005		n.d.
FC1		210 μM
G14		n.d.
GC264		950 μM
GC271		n.d.
MB1200		>1 mM
MB235		>1 mM

Compound ID	Structure	Dissociation constant (Kd)
MG13		930 μM
RC154-1		540 μM
RC154-3		>1 mM
RC163		>1 mM
RC180		n.d.

In order to explore different chemical scaffolds, a classical computational protocol of pharmacophore-based virtual screening has been applied to the MolPort commercial database.

4.9 THE VIRTUAL SCREENING PROCEDURE

The three-dimensional crystal structures of the previously reported binders in complex with VBC had been previously solved in the Ciulli group and were here used to generate a pharmacophore that merges all the features that are important for the interactions and, therefore, for the compounds' binding. In Figure 74 is shown the merged pharmacophore model that consists of six features including: a hydrogen bond acceptor (red sphere) that represent the carbonyl group, a hydrogen bond donor (green sphere) that represents the NH group, one aromatic feature, a hydrophobic feature that corresponds to the bromine atom of FC1, necessary for the activity of the compounds, a positive ionizable area (blue star) with two green arrows that indicate two hydrogen bonds. Moreover, eleven excluded volumes have been found and represent the inaccessible area to any potential ligand.

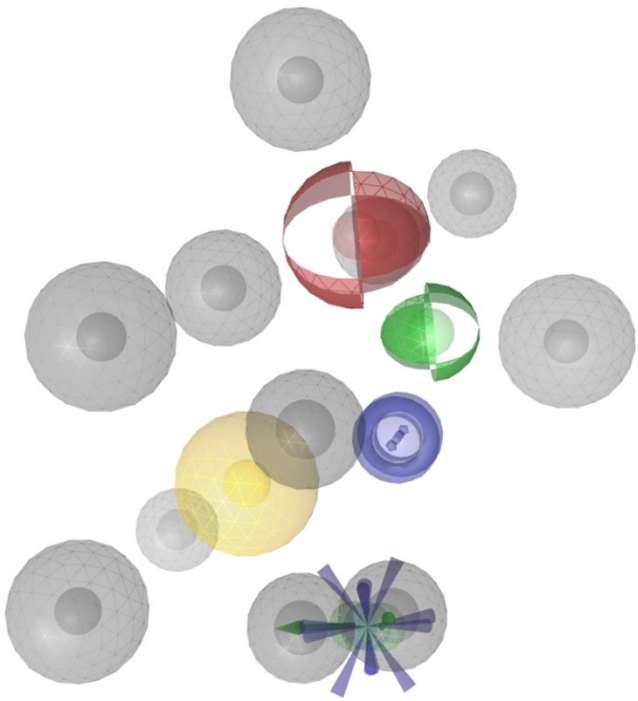
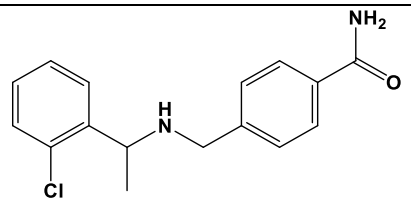
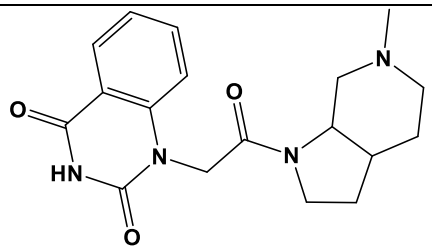
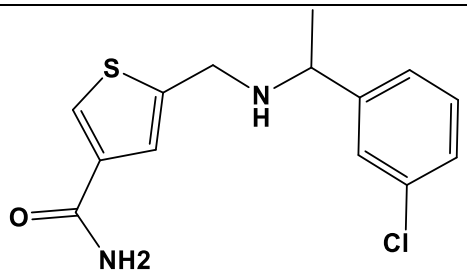
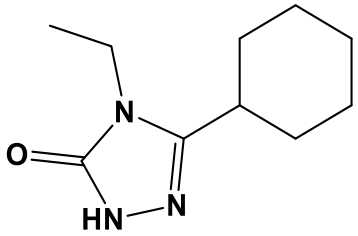
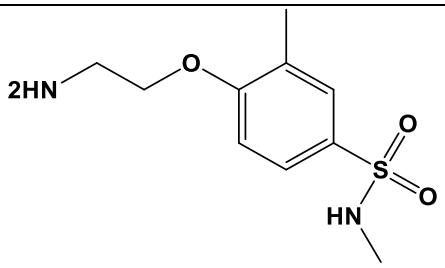


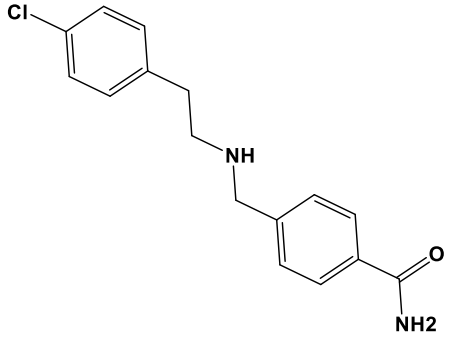
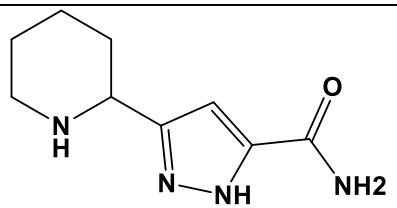
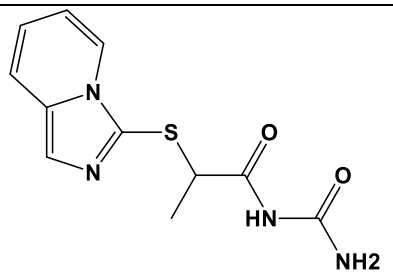
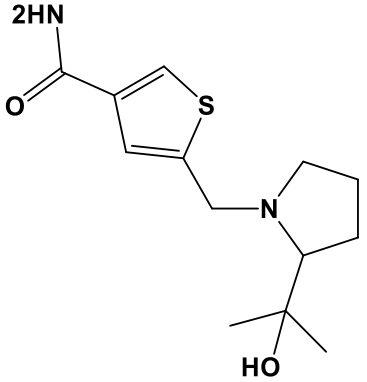
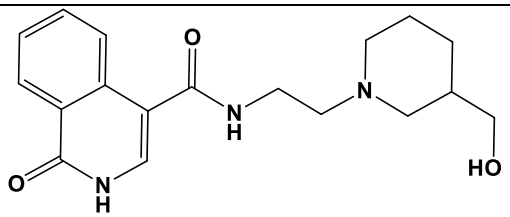
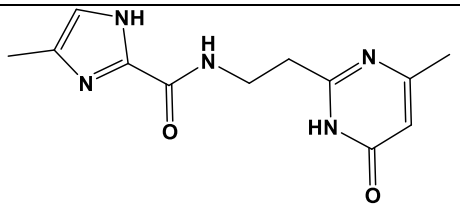
Figure 74. Structure-based pharmacophore model generated using LigandScout from the EloC pocket in complex with the fragments reported in Table 38. The pharmacophore consists of six features including: a hydrogen bond acceptor (red sphere), a hydrogen bond donor (green sphere) that represents the NH group, an aromatic feature, a hydrophobic feature, a positive ionizable area (blue star) with two green arrows that indicate two hydrogen bonds.

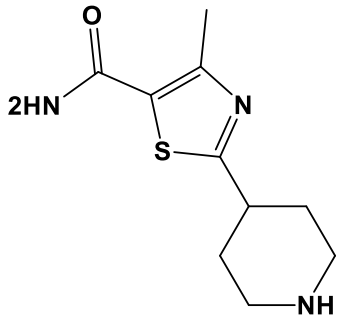
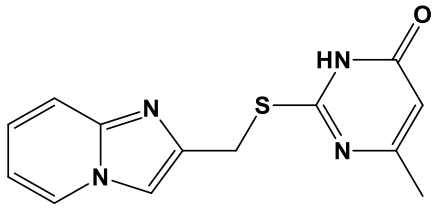
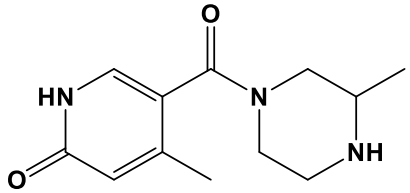
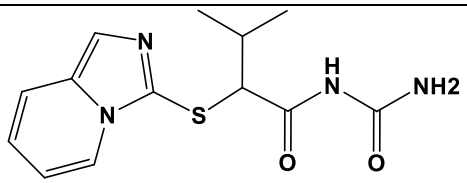
The MolPort commercial database was screened using the iscreen protocol integrated in LigandScout tools to search for new potential active binders. A large set of 7,000,000 compounds constituted the database, and at the end of the pharmacophore-based screening, a total amount of 15,425 compounds were selected based on their agreement with the pharmacophore model. The compounds were filtered by docking using the GLIDE docking tools [120, 121, 122, 123] using the parameters described in the material and methods section and obtaining 11270 compounds. Based on the docking score and on the capability of the compounds to match the pharmacophore in their mapping pose, 10% of the compounds have been subjected to MMGBSA analysis using the Prime module. A total of 783 molecules that showed a ΔG value major than FC1 have been selected for further analysis. In particular, the ADME properties were calculated using QikProp tool available in the Schrödinger suite. [140] Molecules were discarded based on their predicted aqueous solubility (QPlogS) values (the recommended range is less than -6 or greater than -1); predicted octanol/water (QP log Po/w) partition coefficient (range of recommended values = −2.0 – + 6.5); predicted apparent Caco-2 cell permeability (QPPcaco): (a value minor than 25 is poor), selecting 615 compounds. To avoid compounds that can interfere with the biological assays (Pan Assay Interference compounds or PAINS), the FafDrugs4 webserver have been used. No PAINS were found, but 37 molecules were identified as intermediates and were abandoned. The last 578 compounds were visually evaluated for their agreement with the developed pharmacophore and their ability to bind the EloC pocket.

The best 15 compounds are reported in the table below and are suitable to be purchased from vendors or synthesised and then tested in binding studies. In the table are reported the chemical structures, docking score, MMGBSA values and vendor codes of the selected compounds.

Table 39. Chemical structures, docking scores, MMGBSA values and commercial ID of the first fifteen compounds obtained from the screening of MolPort database.

Cmpd ID	Structure	Docking Score	MMGBSA (kcal/mol)	MolPort ID
MP01		-5.23	-44.95	MolPort-005-598-243
MP02		-5.00	-40.38	MolPort-046-133-770
MP03		-4.90	-55.22	MolPort-023-239-858
MP04		-5.89	-31.01	MolPort-023-244-231
MP05		-5.60	-53.91	MolPort-023-282-798

MP06		-4.84	-55.02	MolPort-009-147-414
MP07		-5.00	-32.33	MolPort-039-209-779
MP08		-4.80	-44.58	MolPort-027-642-200
MP09		-4.87	-38.61	MolPort-044-584-006
MP10		-4.95	-47.31	MolPort-030-061-417
MP11		-5.62	-39.21	MolPort-044-697-643

MP12		-5.02	-32.74	MolPort-046-673-749
MP13		-5.21	-44.00	MolPort-005-314-922
MP14		-5.15	-53.19	MolPort-042-687-557
MP15		-4.93	-42.24	MolPort-028-073-458

This study identifies fifteen small molecules as potential ligands of the EloC pocket by using pharmacophore based virtual screening, molecular docking, and MM-GBSA method. The results of this study provide fifteen potential ligands to be considered in the futher optimisation of compounds targeting this pocket.

4.10 CONCLUSIONS

Virtual screening methods play an important role in the drug discovery processes and have the advantages of reducing time and cost efficiencies to allow to search for small molecules in large compound databases to identify structures which are potentially bound to a therapeutic target. Starting from hit compounds previously identified for the EloC pocket, a pharmacophore-based virtual screening procedure was applied to find novel compounds to be validated as inhibitors of this protein.

The pharmacophore model was constructed based on the chemical features of the binders that interact with low binding affinity but high ligand efficiency and specificity to the EloC pocket. The candidate compounds were filtered through the pharmacophore-fit scores, docking energy scores, drug-likeness filters and ADMET properties. Moreover, the analysis of the predicted binding affinity of the selected compounds showed a high value in comparison to FC1, the co-crystallized ligand that presents the best binding affinity. The top fifteen candidate compounds, chosen considering their agreement with the pharmacophore model, their binding mode and MMGBSA value, were validated as potential EloC pocket binders and suggested for potential future purchases or synthesis.

4.11 Materials and Methods
Protein and Ligands Preparation

Proteins were pre-treated with the Protein Preparation Wizard [113, 114] tool of Maestro 12.3 suite, [115] by assign bond order, adding hydrogens and capping termini. In addition in the refine module, H-bond assignment procedure was performed using exhaustive sampling and sample water orientation options marked. Missing loops have been filled using Prime.

Ligands were designed by means of Maestro (version 2020.1) graphical interface [115] and pre-treated with LigPrep [116] tools at default settings, checking the ionization state at physiological pH with Epik.

Pharmacophore generation

The pharmacophore model was generated with LigandScout, [76] version 4.4, based on the compounds reported in Table 38 in complex with CRL, merging all features important for the activities of the compounds. The pharmacophore consists of six features including: a hydrogen bond acceptor (red sphere), a hydrogen bond donor (green sphere) that represents the NH group, an aromatic feature, a hydrophobic feature, a positive ionizable area (blue star) with two green arrows that indicate two hydrogen bonds.

Library preparation

MolPort commercial library is available online and was used for the pharmacophore-based virtual screening. The library has been downloaded in sdf format and then converted in LigandScout input file format (.LDB files). The databases that LigandScout uses for internal virtual screening have been created using a command-line tool called ibgen with the following parameters: confgen-type, icon-best; num-confs, 50 while other parameters were set as default value.

LigandScout uses the iCon conformer generator implemented starting from LigandScout 4.0. the iCon-fast option generates a maximum of 25 conformers for molecule, while the iCon-best option generates 500 conformers; it is possible to define the exact number of conformers to be generate with the *num-confs* options, that has been set as 50.

Virtual screening

The merged pharmacophore has been used as a search query for the virtual screening of the multi-conformational databases by applying the BEST Flexible search to retrieve the compounds with novel scaffolds and with desired chemical features. Maximum Omitted Features option was chosen as 0 for the first screening and it was changed into 1 for the second screening because mapping all features present in the structure-based hypothesis will reduce the hit rate.

Docking Studies

All the selected molecules were docked within the EloC pocket using GLIDE software [120, 121, 122, 123] included in the Maestro suite. The pocket under investigation was inserted into a grid box centred on FC1 ligand and enclosing residues lying within 15Å from such amino acid. Compounds were docked using the Standard Precision procedure and allow only the refinement of the conformation selected by the pharmacophore-based screening. A total of 11270 were docked into the binding site.

MMGBSA analysis

Prime MM-GBSA [54] is a tool that allows the calculation of the binding energies for a protein-ligand complex. Calculations were performed using the OPLS3 force field and the VSGB solvation model. [141] All amino acids that are 4Å distant from the ligand will be relaxed in the calculation, together with the ligands. Default settings are left for the sampling method.

ADME prediction

QikProp [60] is a quick, accurate, easy-to-use absorption, distribution, metabolism, and excretion (ADME) prediction program. QikProp predicts physically significant descriptors and pharmaceutically relevant properties of organic molecules, either individually or in batches. ADME molecular properties of all selected compounds were predicted with Qikprop, using the Fast mode option.

PAINS Analysis

FAF-Drugs4 [77] (Free ADME-Tox Filtering Tool) is a web-server able to predict some ADME-Tox properties (Adsorption, Distribution, Metabolism, Excretion, and Toxicity) to help the hit selection and discard PAINS [78] before chemical synthesis or ordering using pre-defined filters.

Pictures

Pictures of the modelled ligand-CRL complexes, together with graphic manipulations, were rendered with the PyMOL package [73] (version 1.8.4.0 [http://www.pymol.org/]).

Conclusions

DDX3X is a DEAD-box ATP-dependent RNA helicase DDX3X that is involved in several aspects of RNA metabolism and is a fundamental cofactor of the viral replication. The research group of Professor Botta discovered the first compound characterised by broad-spectrum antiviral activity able to act against DDX3X. The compound (**16d**) was able to inhibit the replication of several viruses (HIV, HCV, WNV, JEV, DENV) [16] and has represented the starting point for the application of hit-optimization procedures.

Starting from **16d** (IC_{50} = 0.30 μM), molecular modeling methods have been used to guide the synthesis of novel derivatives in order to find compounds with increased activity, improved pharmacokinetics proprieties and able to inhibit HIV resistant strains. The rational design of derivatives of **16d** allowed to discover twenty-one novel DDX3X inhibitors with inhibitory activities ranging from 0.1 to 60 μM. Eleven compounds were evaluated against HIV-1 infected cells revealing activity ranging from the low micromolar to submicromolar concentrations. Compound **6b** showed a significant antiviral activity against wt HIV-1 (HIV IC_{50} = 2.1 μM) and resistant strains (see **Table 5**), thus representing a good candidate for *in vivo* experiments. [57]

Compound **16d** (DENV IC_{50} = 2.55 μM) and a previously discovered series of sulfonamide compounds [61] were instead used as starting point for the design of novel compounds active against DENV infection. Novel DDX3X helicase inhibitors were discovered, with activities that are comparable to or lower than those reported for ribavirin or sofosbuvir. The design of a fluorescent inhibitor allowed to evaluate also the mechanism of action of the compounds, discovering that, during the first hours of DENV infection, the fluorescent derivative **25** colocalized with DDX3X, promoted the reduction of NS5 positive cells, and recovered the cell number over time. [74]

All these compounds, characterised by low toxicity, confirm that DDX3X inhibitors are safe and promising antiviral agents that offer the advantage of fighting different viruses, including novel emerging ones, with a unique molecule, reducing the risk of developing drug resistance.

To further improve the selectivity of DDX3X inhibitors, several studies have been conducted on derivatives of compound **5** (IC_{50} = 0.7±0.2 μM) the first inhibitor able to bind the unique motif of DDX3X. Mutational studies allowed the identification of amino acid residues responsible of the activity of the compound, [142] leading to the design of a small library of derivatives able to retain the main interactions found for **5**. Moreover, a more selective virtual screening performed on the DDX3X unique motif allowed to identify compounds predicted to bind the small pocket of the UM establishing interactions with residues that are responsible of the activity of **5**. These molecules will be synthesised or purchased and then submitted to biological assays. The biological evaluation will confirm the reliability of the binding mode predicted by computational studies, and, if the results will be positive, novel series of compounds will be designed with the aim of inhibiting DDX3X in a selectivity matter.

PROTAC is instead a useful and potent strategy to target protein degradation. PROTACs, that are constituted by two active moieties hold together by a linker, allow the selective degradation of POIs, recruiting the E3 ligase to the target protein. This results in selective ubiquitination and degradation of the POIs by the proteasome.

SOCS2 is a component of the E3 ubiquitin ligase complex, and it could be used to design a molecule that can be used to develop a PROTAC. [109] The application of some computational approaches on SOCS2-binders complexes previously identified, allowed to establish a SAR that will be taken in consideration to guide the synthesis of new derivatives with improved activities and pharmacokinetics proprieties. Moreover, the generated QSAR model will be used to predict the binding affinities of these new molecules to prioritise for the synthesis.

Elongin C (EloC) is another component of the CRL complex. It is a protein adaptor that is involved in the proteasomal degradation of target proteins via different E3 ubiquitin ligase complexes and that can be used to develop a degrader. [138] Since the necessity to find novel scaffolds able to bind the EloC pocket, the three-dimensional information of previously discovered EloC binders have been used to develop a pharmacophore model used to filter the MolPort commercial database. [139] The Virtual Screening procedure allowed to select 10 compounds that will be purchased and biologically evaluated as EloC binders.

References

[1] K. Rosenke, “Use of Favipiravir to Treat Lassa Virus Infection in Macaques,” *Emerg Infect Dis,* no. 9, pp. 1696-1699, 2018.

[2] Y. Furuta, “Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase,” *Proc Jpn Acad Ser B Phys Biol Sci,* vol. 93, no. 7, pp. 449-463, 2017.

[3] S. Safrin, “Clinical uses of cidofovir,” *Rev Med Virol,* vol. 7, no. 3, pp. 145-156, 1997.

[4] J. F. Camargo, “The use of brincidofovir for the treatment of mixed dsDNA viral infection,” *J Clin Virol,* vol. 83, pp. 1-4, 2016.

[5] F. Y. Aoki, “Antiviral Drugs for Influenza and Other Respiratory Virus Infections,” *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases.,* p. 531–545, 2015.

[6] [Online]. Available: https://www.drugs.com/pro/ribavirin.html. [Accessed November 2020].

[7] M. L. Agostini, “Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease,” *mBio,* vol. 9, no. 2, pp. e00221-18, 2018.

[8] J. Grein, “Compassionate Use of Remdesivir for Patients with Severe Covid-19,” *N Engl J Med,* vol. 382, no. 24, pp. 2327-2336, 2020.

[9] A. K. Singh, “Remdesivir in COVID-19: A critical review of pharmacology, pre-clinical and clinical studies,” *Diabetes Metab Syndr,* vol. 14, no. 4, pp. 641-648, 2020.

[10] K. Toth , “Valganciclovir inhibits human adenovirus replication and pathology in permissive immunosuppressed female and male Syrian hamsters,” *Viruses,* vol. 7, no. 3, pp. 1409-28, 2015.

[11] R. G. Gish, “Ganciclovir treatment of hepatitis B virus infection in liver transplant recipients,” *Hepatology,* vol. 1, no. 1-7, 1996.

[12] G. Andrei , “The Anti-Human Immunodeficiency Virus Drug Tenofovir, a Reverse Transcriptase Inhibitor, Also Targets the Herpes Simplex Virus DNA Polymerase,” *J Infect Dis,* vol. 217, no. 5, pp. 790-801, 2018.

[13] T. Delory , “Foscarnet, zidovudine and dolutegravir combination efficacy and tolerability for late stage HIV salvage therapy: A case-series experience,” *J Med Virol,* vol. 88, no. 7, pp. 1204-10, 2016.

[14] J.-F. Rossignol, “Nitazoxanide: a first-in-class broad-spectrum antiviral agent,” *Antiviral Res,* vol. 110, pp. 94-103, 2014.

[15] Y. Jin, “Human monoclonal antibodies as candidate therapeutics against emerging viruses,” *Front Med,* vol. 11, no. 4, pp. 462-470, 2017.

[16] A. Brai, “Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents,” *Proc Natl Acad Sci U S A,* vol. 113, no. 19, pp. 5388-93, 2016.

[17] M. Ishaq, “Knockdown of cellular RNA helicase DDX3 by short hairpin RNAs suppresses HIV-1 viral replication without inducing apoptosis,” *Mol Biotechnol,* vol. 39, no. 3, pp. 231-8, 2008.

[18] C.-Y. Chen, “Cellular RNA helicases and HIV-1: insights from genome-wide, proteomic, and molecular studies,” *Virus Res,* vol. 171, no. 2, pp. 357-65, 2013.

[19] P. Linder, “Dead-box proteins: a family affair--active and passive players in RNP-remodeling,” *Nucleic Acids Res,* vol. 34, no. 15, pp. 4168-80, 2006.

[20] J.-W. Shih , “Human DExD/H RNA helicases: emerging roles in stress survival regulation,” *Clin Chim Acta,* vol. 436, no. 45, pp. 45-58, 2014.

[21] S. E. Nunes-Düby, “Similarities and differences among 105 members of the Int family of site-specific recombinases,” *Nucleic Acids Res,* vol. 26, no. 2, pp. 391-406, 1998.

[22] O. Cordin, “The DEAD-box protein family of RNA helicases,” *Gene,* vol. 15, no. 367, pp. 17-37, 2006.

[23] P. Linder , “From unwinding to clamping - the DEAD box RNA helicase family,” *Nat Rev Mol Cell Biol,* vol. 22, no. 12, pp. 505-16, 2011.

[24] M. E. Fairman-Williams, “SF1 and SF2 helicases: family matters,” *Curr Opin Struct Biol,* vol. 20, no. 3, pp. 313-24, 2010.

[25] E. Silverman, “DExD/H-box proteins and their partners: helping RNA helicases unwind,” *Gene,* vol. 17, no. 312, pp. 1-16, 2003.

[26] S. H. Park, “Assignment of a human putative RNA helicase gene, DDX3, to human X chromosome bands p11.3-->p11.23,” *Cytogenet Cell Genet,* vol. 81, no. 3-4, pp. 178-9, 1998.

[27] A. A. Kotov, “Progress in understanding the molecular functions of DDX3Y (DBY) in male germ cell development and maintenance,” *Biosci Trends,* vol. 11, no. 1, pp. 46-53, 2017.

[28] R. Soto-Rifo , “The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism,” *Wiley Interdiscip Rev RNA,* vol. 4, no. 4, pp. 369-85, 2013.

[29] M.-C. Lai , “The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control,” *Mol Biol Cell,* vol. 19, no. 9, pp. 3847-58, 2008.

[30] M. Schröder, “Human DEAD-box protein 3 has multiple functions in gene regulation and cell cycle control and is a prime target for viral manipulation,” *Biochem Pharmacol,* vol. 79, no. 3, pp. :297-306, 2010.

[31] M. Schröder , “Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKKepsilon-mediated IRF activation,” *EMBO J,* vol. 27, no. 15, pp. :2147-57, 2008.

[32] A. Fröhlich, “DEAD-box RNA helicase DDX3 connects CRM1-dependent nuclear export and translation of the HIV-1 unspliced mRNA through its N-terminal domain,” *Biochim Biophys Acta,* vol. 1859, no. 5, pp. 719-30, 2016.

[33] M. Caputi, “The Regulation of HIV-1 mRNA Biogenesis,” *RNA Processing,* 2011.

[34] M. K. Kukhanova , “DEAD-box RNA Helicase DDX3: Functional Properties and Development of DDX3 Inhibitors as Antiviral and Anticancer Drugs,” *Molecules,* vol. 25, no. 4, p. 1015, 2020.

[35] N. Mamiya, “Hepatitis C virus core protein binds to a DEAD box RNA helicase,” *J Biol Chem,* vol. 274, no. 22, pp. 15751-6, 1999.

[36] G. M. Bol, “Expression of the RNA helicase DDX3 and the hypoxia response in breast cancer,” *PLoS One,* vol. 8, no. 5, p. e63548, 2013.

[37] E. Cannizzaro , “DDX3X RNA helicase affects breast cancer cell cycle progression by regulating expression of KLF4,” *FEBS Lett,* vol. 592, no. 13, pp. 2308-2322, 2018.

[38] M. Botlagunta, “Oncogenic role of DDX3 in breast cancer biogenesis,” *Oncogene,* vol. 27, no. 28, pp. 3912-22, 2008.

[39] V. S. R. K. Yedavalli, “Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function,” *Cell,* vol. 119, no. 3, pp. 381-92, 2004.

[40] A. Hilliker , “The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex,” *Mol Cell,* vol. 43, no. 6, pp. 962-72, 2020.

[41] D. Sharma , “The Ded1/DDX3 subfamily of DEAD-box RNA helicases,” *Crit Rev Biochem Mol Biol,* vol. 49, no. 4, pp. 343-60, 2014.

[42] S. N. Floor, “Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3,” *J Biol Chem,* vol. 291, no. 5, pp. 2412-2421, 2016.

[43] H. Song, “The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X,” *Nat Commun,* vol. 10, no. 1, p. 3085, 2019.

[44] V. S. R. K. Yedavalli , “Ring expanded nucleoside analogues inhibit RNA helicase and intracellular human immunodeficiency virus type 1 replication,” *J Med Chem,* vol. 51, no. 16, pp. 5043-51, 2008.

[45] M. Radi, “Discovery of the first small molecule inhibitor of human DDX3 specifically designed to target the RNA binding site: towards the next generation HIV-1 inhibitors,” *Bioorg Med Chem Lett,* vol. 22, no. 5, pp. 2094-8, 2012 .

[46] G. Maga, “Pharmacophore modeling and molecular docking led to the discovery of inhibitors of human immunodeficiency virus-1 replication targeting the human cellular aspartic acid-glutamic acid-alanine-aspartic acid box polypeptide 3,” *J Med Chem,* vol. 51, no. 21, pp. 6635-8, 2008.

[47] G. Maga, “Toward the discovery of novel anti-HIV drugs. Second-generation inhibitors of the cellular ATPase DDX3 with improved anti-HIV activity: synthesis, structure-activity relationship analysis, cytotoxicity studies, and target validation,” *ChemMedChem,* vol. 6, no. 8, pp. 1371-89, 2011.

[48] P. Schütz , “Comparative structural analysis of human DEAD-box RNA helicases,” *PLoS One,* vol. 5, no. 9, p. e12791, 2010.

[49] T. Sekiguchi, “Phosphorylation of threonine 204 of DEAD-box RNA helicase DDX3 by cyclin B/cdc2 in vitro,” *Biochem Biophys Res Commun,* vol. 356, no. 3, pp. 668-73, 2007.

[50] A. Brai, “Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents,” *Proc Natl Acad Sci U S A,* vol. 113, no. 19, pp. 5388-93, 2016.

[51] R. Fazi, “Homology Model-Based Virtual Screening for the Identification of Human Helicase DDX3 Inhibitors,” *J Chem Inf Model,* vol. 55, no. 11, pp. 2443-54, 2015.

[52] M. P. Jacobson , “A hierarchical approach to all-atom protein loop prediction,” *Proteins,* vol. 55, no. 2, pp. 351-67, 2004.

[53] M. P. Jacobson, “On the Role of Crystal Packing Forces in Determining Protein Sidechain Conformations,” *J Mol Biol,* vol. 320, no. 3, pp. 597-608, 2002.

[54] *Schrödinger Release 2020-4: Prime, Schrödinger, LLC, New York, NY, 2020.*

[55] Z. Chen, “Structural Basis of Human Helicase DDX21 in RNA Binding, Unwinding, and Antiviral Signal Activation,” *Adv Sci (Weinh),* vol. 7, no. 14, p. 2000532, 2020.

[56] [Online]. Available: https://www.poz.com/drug_charts/hiv-drug-chart.

[57] A. Brai, “DDX3X inhibitors, an effective way to overcome HIV-1 resistance targeting host proteins,” *Eur J Med Chem,* vol. 200, p. 112319, 2020.

[58] G. Antonelli, “Antiviral therapy: old and current issues. International Journal of Antimicrobial Agents 40, 95–102 (2012).,” *International Journal of Antimicrobial Agents,* vol. 95, pp. 95-102, 2012.

[59] G. Jones, “Development and validation of a genetic algorithm for flexible docking,” *J Mol Biol,* vol. 267, no. 3, pp. 727-48, 1997.

[60] *Schrödinger Release 2020-4: QikProp, Schrödinger, LLC, New York, NY, 2020..*

[61] A. Brai, “DDX3X Helicase Inhibitors as a New Strategy To Fight the West Nile Virus Infection,” *J Med Chem,* vol. 62, no. 5, pp. 2333-2347, 2019.

[62] P. A. Lipinski, “Drug-like properties and the causes of poor solubility and poor permeability,” *J Pharmacol Toxicol Methods,* vol. 44, no. 1, pp. 235-49, 2000.

[63] D. J. Nolan, “The Spleen Is an HIV-1 Sanctuary During Combined Antiretroviral Therapy,” *AIDS Res Hum Retroviruses,* vol. 34, no. 1, pp. 123-125, 2018.

[64] J. Chang, “Combination of α-glucosidase inhibitor and ribavirin for the treatment of dengue virus infection in vitro and in vivo,” *Antiviral Res,* vol. 89, no. 1, pp. 26-34, 2011.

[65] Y. Ariumi , “Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection,” *Front Genet,* vol. 5, p. 423, 2014.

[66] H. S. Chahar, “P-body components LSM1, GW182, DDX3, DDX6 and XRN1 are recruited to WNV replication sites and positively regulate viral replication,” *Virology,* vol. 436, no. 1, pp. 1-7, 2013.

[67] C. Li, “Cellular DDX3 regulates Japanese encephalitis virus replication by interacting with viral un-translated regions,” *Virology,* vol. 449, pp. 70-81, 2014.

[68] *Schrödinger Release 2020-4: Maestro, Schrödinger, LLC, New York, NY, 2020..*

[69] *Schrödinger Release 2020-4: LigPrep, Schrödinger, LLC, New York, NY, 2020..*

[70] J. R. Greenwood, “Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution,” *J Comput Aided Mol Des,* vol. 24, no. 6-7, pp. 591-604, 2010.

[71] J. C. Shelley, “Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules,” *J Comput Aided Mol Des,* vol. 21, no. 12, pp. 681-91, 2007.

[72] *Schrödinger Release 2020-4: Epik, Schrödinger, LLC, New York, NY, 2020..*

[73] *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC..*

[74] A. Brai, “Exploring the Implication of DDX3X in DENV Infection: Discovery of the First-in-Class DDX3X Fluorescent Inhibitor,” *ACS Med Chem Lett,* vol. 11, no. 5, pp. 956-962, 2020.

[75] M. Weisel,, “PocketPicker: analysis of ligand binding-sites with shape descriptors,” *Chem Cent J.,* vol. 1, p. 7, 2007.

[76] G. Wolber, “LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters,” *J Chem Inf Model,* vol. 45, no. 1, pp. 160-9, 2005.

[77] D. Lagorce , “FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects,” *BMC Bioinformatics,* vol. 9, p. 396, 2008.

[78] J. B. Baell , “New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays,” *J Med Chem,* vol. 53, no. 7, pp. 2719-40, 2010.

[79] J. Li, “The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling,” *Proteins,* vol. 79, no. 10, pp. 2794-812, 2011.

[80] R. J. Deshaies, “SCF and Cullin/RING H2-Based Ubiquitin Ligases,” *Annu.Rev.Cell Dev.Biol,* vol. 15, pp. 435-467, 1999.

[81] W. G. Kaelin, “The von Hippel-Lindau Tumor Suppressor Protein: Roles in Cancer and Oxygen Sensing,” *Cold Spring Harb Symp Quant Biol,* vol. 70, pp. 159-166, 2005.

[82] Q. Shi, “Cereblon: A Protein Crucial to the Multiple Functions of Immunomodulatory Drugs as well as Cell Metabolism and Disease Generation,” *Journal of Immunology Research,* vol. 2017, p. e9130608, 2017.

[83] J. Silke, “Inhibitor of Apoptosis (IAP) Proteins-Modulators of Cell Deatch and Inflammation,” *Cold Spring Harb Perspect Biol,* vol. 5, p. a008730, 2013.

[84] T. Iwakuma, “MDM2, an introduction,” *Molecular Cancer Research,* vol. 1, no. 14, 2003.

[85] T. Ishida, “E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones,” *SLAS Discovery,* vol. I, p. 19, 2020.

[86] “Drugs@FDA: FDA-Approved Drugs.,” [Online]. Available: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=021481..

[87] L. Bai, “Potent and Selective Small-Molecule Degrader of STAT3 Achieves Complete Tumor Regression In Vivo,” *Cancer Cell,* vol. e17, pp. 498-511, 2019.

[88] X. Sun, “A chemical approach for global protein knockdown from mice to non-human primates,” *Cell Discovery,* vol. 5, pp. 1-13, 2019.

[89] J. Lotem, “Cytokine control of developmental programs in normal hematopoiesis and leukemia,” *Oncogene,* vol. 21, no. 21, pp. 3284-94, 2002.

[90] F. Seif, “The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells,” *Cell Communication and Signaling,* vol. 15, no. 23, 2017.

[91] B. A. Croker, “SOCS Regulation of the JAK/STAT Signalling Pathway,” *Semin Cell Dev Biol,* vol. 19, no. 4, p. 414–422, 2008.

[92] K. Inagari-Ohara, “SOCS, inflammation, and cancer,” *JAK_STAT,* vol. 2, p. e24053, 2013.

[93] B. T. Kile, “The SOCS box: a tale of destruction and degradation,” *Trends in Biochemical Sciences,* vol. 27, pp. 235-241, 2002.

[94] T. Kamura, “VHL-box and SOCS-box domains determine binding specificity for Cul2-RBX1 and Cul5-Rbx2 modules of ubiquitin ligases,” *Genes Dev,* vol. 18, pp. 3055-3065, 2004.

[95] D. J. Hilton, “Twenty proteins containing a C-terminal SOCS box form five structural classes,” *PNAS,* vol. 95, pp. 114-119, 1998.

[96] R. Starr, “A family of cytokine-inducible inhibitors of signalling,” *Natyre,* vol. 387, pp. 917-921, 1997.

[97] J. Piessevaux, “Functional cross-modulation between SOCS proteins can stimulate cytokine signaling,” *J Biol Chem,* vol. 281, no. 44, pp. 32953-66, 2006.

[98] I. D. Dimitriou, “Putting out the fire: coordinated suppression of the innate and adaptive immune systems by SOCS1 and SOCS3 proteins,” *Immunological Reviews,* vol. 224, pp. 265-283, 2008.

[99] D. Metcalf, “Gigantism in mice lacking suppressor of cytokine signalling-2,” *Nature,* vol. 405, pp. 1069-1073, 2000.

[100] P. Ram, “SOCS/CIS Protein Inhibition of Growth Hormone-stimulated STAT5 Signaling by Multiple Mechanisms,” *J Biol Chem,* vol. 274, pp. 35553-35561, 1999.

[101] C. J. Greenhalgh , “SOCS2 negatively regulates growth hormone action in vitro and in vivo,” *J Clin Invest,* vol. 115, pp. 397-406, 2005.

[102] M. Vesterlund, “The SOCS2 ubiquitin ligase complex regulates growth hormone receptor levels,” *PLoS One,* vol. 6, no. 9, p. e25358, 2011.

[103] S. Eyckerman, “Design and Application of a cytokine-receptor-based interaction trap,” *Nature Cell Biology,* vol. 3, pp. 1114-1119, 2001.

[104] D. Lavnes, “A complex interaction pattern of CIS and SOCS2 with the leptin receptor.,” *Journal of Cell Science,* vol. 119, pp. 2214-2224, 2006.

[105] Y. Goldshmit, “SOCS2 Induces Neurite Outgrowth by Regulation of Epidermal Growth Factor Receptor Activation,” *J Biol Chem,* vol. 279, pp. 16349-16355, 2004.

[106] B. R. Dey, “Interaction of Human Suppressor of Cytokine Signaling (SOCS)-2 with the Insulin-like Growth Factor-I Receptor,” *J Biol Chem,* vol. 273, pp. 24095-24101, 1998.

[107] I. Paul, “The ubiquitin ligase Cullin5 SOCS2 regulates NDR1/STK38 stability and NF-κB transactivation,” *Scientific Report,* vol. 7, p. 42800, 2017.

[108] X. Lucass, “Recognition of substrate degrons by E3 ubiquitin ligases and modulation by small-molecule mimicry strategies,” *Current Opinion in Structural Biology,* vol. 44, pp. 101-110, 2017.

[109] W. W. Kung, “Structural Insights into Substrate Recognition by the SOCS2 E3 ubiquitin ligase,” *Nature Communication,* vol. 10, p. 2534, 2019.

[110] S. L. Dixon , “AutoQSAR: an automated machine learning tool for best-practice QSAR modeling,” *Future Med Chem,* vol. 15, no. 8, pp. 1825-1839, 2016.

[111] *Schrödinger Release 2020-4: AutoQSAR, Schrödinger, LLC, New York, NY, 2020.*

[112] S. Genheden, “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities,” *Expert Opin Drug Discov,* vol. 10, no. 5, pp. 449-461, 2015.

[113] G. M. Sastry, “Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments,” *J. Comput. Aid. Mol. Des.,* vol. 27, no. 3, pp. 221-234, 2013.

[114] *Schrödinger Release 2020-4: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2020; Impact, Schrödinger, LLC, New York, NY, 2020; Prime, Schrödinger, LLC, New York, NY, 2020..*

[115] *Schrödinger Release 2020-4: Maestro, Schrödinger, LLC, New York, NY, 2020..*

[116] *Schrödinger Release 2020-4: LigPrep, Schrödinger, LLC, New York, NY, 2020..*

[117] J. R. Greenwood, “Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution,” *J. Comput. Aided Mol. Des.,* vol. 24, pp. 591-604, 2010.

[118] J. C. Shelley, “Epik: a software program for pKa prediction and protonation state generation for drug-like molecules,” *J. Comp.-Aided Mol. Design,* vol. 21, pp. 681-691, 2007.

[119] *Schrödinger Release 2020-4: Epik, Schrödinger, LLC, New York, NY, 2020..*

[120] R. A. Friesner, “Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes,” *J Med Chem,* vol. 49, pp. 6177-6196, 2006.

[121] T. A. Halgren, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening,” *J Med Chem,* vol. 47, p. 1750–1759, 2004.

[122] R. A. Firesner, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy,” *J Med Chem,* vol. 47, p. 1739–1749, 2004.

[123] *Schrödinger Release 2020-4: Glide, Schrödinger, LLC, New York, NY, 2020.*

[124] J. C. Phillips, “Scalable molecular dynamics on CPU and GPU architectures with NAMD.,” *Journal of Chemical Physics,* 2020.

[125] J. Wang, “Automatic atom type and bond type perception in molecular mechanical calculations,” *Journal of Molecular Graphics and Modelling,* vol. 25, p. 247260, 2006.

[126] J. Wang, “Development and testing of a general AMBER force field,” *Journal of Computational Chemistry,* vol. 25, pp. 1157-1174, 2004.

[127] A. Jakalian, “Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation,” *J Comput Chem,* vol. 16, pp. 1623-41, 2002.

[128] J. Izaguirre, “. Langevin stabilization of molecular dynamics,” *J. Chem. Phys.,* vol. 114, p. 2090–2098, 2001.

[129] A. Mudi, “Effect of the Berendsen thermostat on the dynamical properties of water,” *Mol Phys,* vol. 2004, pp. 681-685, 2004.

[130] W. Humphrey , “VMD - Visual Molecular Dynamics,” *J. Molec. Graphics,* vol. 14, pp. 33-38., 1996.

[131] A. D. Case and e. al., *AMBER 2020, University of California, San Francisco.*

[132] A. Sarikas, “The cullin protein family,” *Genome Biol,* vol. 12, p. 220–232, 2011.

[133] P. Sharma, “CUL4A ubiquitin ligase: a promising drug target for cancer and other human diseases,” *Open Biol,* vol. 4, no. 2, p. 130217, 2014.

[134] D. Wei, “Small RING Finger Proteins RBX1 and RBX2 of SCF E3 Ubiquitin Ligases: The Role in Cancer and as Cancer Targets,” *Genes,* vol. 1, pp. 700-7, 2010.

[135] Y. Zhao, “Ligases as attractive anti-cancer targets.,” *Curr Pharm Des,* vol. 19, pp. 3215-25, 2013.

[136] Y. Sun, “Functionla characterization of SAG/RBX2/ROC2/RNF7, an antioxidant protein and an E3 ubiquitin ligase,” *Protein Cell,* vol. 4, pp. 103-16, 2013.

[137] S. Jackson, “CRL4s: the CUL4-RING E3 ubiquitin ligases,” *Trends Biochem Sci,* vol. 34, pp. 562-70, 2009.

[138] M. Shen, “Targeting the ubiquitin-proteasome system for cancer therapy,” *Expert Opin Ther Targets,* vol. 17, pp. 1091-108, 2013.

[139] L. Xavier, “Surface Probing by Fragment-Based Screening and Computational Methods Identifies Ligandable Pockets on the von Hippel–Lindau (VHL) E3 Ubiquitin Ligase,” *J Med Chem,* vol. 61, no. 16, pp. 7387-7393, 2018.

[140] “Schrödinger Release 2020-4: QikProp, Schrödinger, LLC, New York, NY, 2020.”.

[141] C. Mulakala, “Could MM-GBSA be accurate enough for calculation of absolute protein/ligand binding free energies?,” *Journal of Molecular Graphics and Modelling,* vol. 46, pp. 41-51, 2013.

[142] V. Riva , “Unique Domain for a Unique Target: Selective Inhibitors of Host Cell DDX3X to Fight Emerging Viruses,” *J Med Chem,* vol. 63, no. 17, pp. 9876-9887, 2020.

[143] F. Saladini, “Agreement between an in-house replication competent and a reference replication defective recombinant virus assay for measuring phenotypic resistance to HIV-1 protease, reverse transcriptase, and integrase inhibitors,” *J Clin Lab Anal,* vol. 32, no. 1, p. e22206, 2018.

[144] A. M. Ward, “Quantitative mass spectrometry of DENV-2 RNA-interacting proteins reveals that the DEAD-box RNA helicase DDX6 binds the DB1 and DB2 3' UTR structures,” *RNA Biol,* vol. 8, no. 6, pp. 1173-86, 2011.

[145] L. C. Reineke, “Diversion of stress granules and P-bodies during viral infection,” *Virology,* vol. 436, no. 2, pp. 255-67, 2013.

[146] S. A. N. McElhinny, “Genome instability due to ribonucleotide incorporation into DNA,” *Nat Chem Biol,* vol. 6, no. 10, pp. 774-81, 2010.

[147] S. A. N. McElhinny , “Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases,” *Proc Natl Acad Sci U S A,* vol. 107, no. 11, pp. 4949-54, 2010.

[148] K. W. Caldecott, “Molecular biology. Ribose--an internal threat to DNA,” *Science,* vol. 343, no. 6168, pp. 260-1, 2014.

[149] C. J. Potenski , “How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells,” *Nucleic Acids Res,* vol. 42, no. 16, pp. 10226-34, 2014.

[150] J. L. Sparks, “RNase H2-initiated ribonucleotide excision repair,” *Mol Cell,* vol. 47, no. 6, pp. 980-6, 2012.

[151] B. Hiller, “Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity,” *J Exp Med,* vol. 209, no. 8, pp. 1419-26, 2012.

[152] V. Riva, “From the magic bullet to the magic target: exploiting the diverse roles of DDX3X in viral infections and tumorigenesis,” *Future Med Chem,* vol. 11, no. 11, pp. 1357-1381, 2019.

[153] C.-H. Chan, “DNA Damage, Liver Injury, and Tumorigenesis: Consequences of DDX3X Loss,” *Mol Cancer Res,* vol. 17, no. 2, pp. 555-566, 2019.

[154] L. Zhao, “Multifunctional DDX3: dual roles in various cancer development and its related signaling pathways,” *Am J Cancer Res,* vol. 6, no. 2, pp. 387-402, 2016.

[155] G. Kellaris, “A hypomorphic inherited pathogenic variant in DDX3X causes male intellectual disability with additional neurodevelopmental and neurodegenerative features,” *Hum Genomics,* vol. 12, no. 1, p. 11, 2018.

[156] C.-Y. Chen , “Targeted inactivation of murine Ddx3x: essential roles of Ddx3x in placentation and embryogenesis,” *Hum Mol Genet,* vol. 25, no. 14, pp. 2905-2922, 2016.

[157] A. Vaisman, “Redundancy in ribonucleotide excision repair: Competition, compensation, and cooperation,” *DNA Repair (Amst),* vol. 29, pp. 74-82, 2015.

[158] N. Awano, “Escherichia coli RNase R has dual activities, helicase and RNase,” *J Bacteriol,* vol. 192, no. 5, pp. 1344-52, 2010.

[159] M. A. Kidwell , “Evolutionarily conserved roles of the dicer helicase domain in regulating RNA interference processing,” *J Biol Chem,* vol. 289, no. 41, pp. 28352-62, 2014.

[160] M. Figiel, “The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and Aicardi-Goutières syndrome defects,” *J Biol Chem,* vol. 286, no. 12, pp. 10540-50, 2011.

[161] E. Mentegari, “Ribonucleotide incorporation by human DNA polymerase η impacts translesion synthesis and RNase H2 activity,” *Nucleic Acids Res,* vol. 45, no. 5, pp. 2600-2614, 2017.

[162] D. Sharma , “Biochemical Differences and Similarities between the DEAD-Box Helicase Orthologs DDX3X and Ded1p,” *J Mol Biol,* vol. 429, no. 23, pp. 3730-3742, 2017.

[163] A. A. Putnam , “Division of Labor in an Oligomer of the DEAD-Box RNA Helicase Ded1p,” *Mol Cell,* vol. 59, no. 4, pp. 541-52, 2015.

[164] Y. Kim , “RNA Remodeling Activity of DEAD Box Proteins Tuned by Protein Concentration, RNA Length, and ATP,” *Mol Cell,* vol. 63, no. 5, pp. 865-76, 2016.

[165] A. Garbelli, “A motif unique to the human DEAD-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication,” *PLoS One,* vol. 6, no. 5, p. e19810, 2011.

[166] E. Crespan, “Impact of ribonucleotide incorporation by DNA polymerases β and λ on oxidative base excision repair,” *Nat Commun,* vol. 7, p. 10805, 2016.

[167] A. Garbelli, “How to win the HIV-1 drug resistance hurdle race: running faster or jumping higher?,” *Biochem J,* vol. 474, no. 10, pp. 1559-1577, 2017.

[168] M. R. Albertella, “The overexpression of specialized DNA polymerases in cancer,” *DNA Repair (Amst),* vol. 4, no. 5, pp. 583-93, 2005.

[169] C. Günther, “Defective removal of ribonucleotides from DNA promotes systemic autoimmunity,” *J Clin Invest,* vol. 125, no. 1, pp. 413-24, 2015.

[170] J.-W. Shih , “Human DExD/H RNA helicases: emerging roles in stress survival regulation,” *Clin Chim Acta,* vol. 436, pp. 45-58, 2014.

APPENDIX 1: LIST OF PUBLICATIONS

1. Riva V.; Garbelli A.; Casiraghi F.; Arena F.; Trivisani C.I.; Gagliardi A.; Bini L.; Schroeder M.; Maffia A.; Sabbioneda S.; Maga G. Novel alternative ribonucleotide excision repair pathways in human cells by DDX3X and specialized DNA polymerases. *Nucleic Acids Res.* 2020, 48(20):11551-11565.

2. Riva V.; Garbelli A.; Brai A.; Casiraghi F.; Fazi R.; Trivisani C.I.; Boccuto A.; Saladini F.; Vicenti I.; Martelli F.; Zazzi M.; Giannecchini S.; Dreassi E.; Botta M.; Maga G. Unique Domain for a Unique Target: Selective Inhibitors of Host Cell DDX3X to Fight Emerging Viruses. J Med Chem. 2020, 63(17):9876-9887.

3. Brai A.; Saladini F.; Zamperini C.; Trivisani C.I.; Giannini A.; Boccuto A.; Dreassi E.; Zazzi M.; Maga G.; Botta M. DDX3X inhibitors, an effective way to overcome HIV-1 resistance. *Eur J Med Chem*. 2020, 200:112319.

4. Brai A.; Boccuto A.; Monti M.; Marchi S.; Vicenti I.; Saladini F.; Trivisani C.I.; Pollutri A.; Trombetta C.M.; Montomoli E.; Riva V.; Garbelli A.; Nola E.M.; Zazzi M.; Maga G.; Dreassi E.; Botta M. Exploring the Implication of DDX3X in DENV Infection: Discovery of the First-in-Class DDX3X Fluorescent Inhibitor. *ACS Med Chem Lett.* 2020, 11(5):956-962.

5. Brai A.; Ronzini S.; Riva V.; Botta L.; Zamperini C.; Borgini M.; Trivisani C.I.; Garbelli A.; Pennisi C.; Boccuto A.; Saladini F.; Zazzi M.; Maga G.; Botta M. Synthesis and Antiviral Activity of Novel 1,3,4-Thiadiazole Inhibitors of DDX3X. *Molecules.* 2019, 24(21):3988.

APPENDIX 2: CONGRESS PARTICIPATIONS

- 19-24 May 2019, "XII European Workshop in Drug Design (EWDD)" - Certosa di Pontignano, Siena

- 16-20 September 2018, "22nd European Symposium on Quantitative Structure-Activity Relationships (EuroQSAR)" Thessaloniki, Greece. Poster Presentation: *Development of Broad-Spectrum Antiviral Agent Able to Inhibit Human DDX3 Protein*

- 20-24 May 2018, "VII European Workshop in Drug Synthesis (EWDSy) - Certosa di Pontignano, Siena

APPENDIX 3: CERTIFICATES AND AWARDS

Poster prize from 22nd European Symposium on Quantitative Structure-Activity Relationship (EuroQSAR 2018), 3rd placement.

Title of the poster: "*Development of Broad-Spectrum Antiviral Agent Able to Inhibit Human DDX3 Protein*".