UNIVERSITAT ROVIRA i VIRGILI

Development of a SAR-Guided Virtual Screening Strategy to Identify Allosteric Inhibitors of Zika Virus NS2B/NS3 Protease

Judith Montano Hernández

FINAL DEGREE PROJECT IN BIOTECHNOLOGY

Academic Tutor/Supervisor:	Dr. Gerard Pujadas Anguiano Department of Biochemistry and Biotechnology (URV) gerard.pujadas@urv.cat
In cooperation with:	Cheminformatics and Nutrition Research Group (URV)
Supervisor/s:	Dr. Santiago Garcia Vallvé, Dr. Aleix Gimeno Vives Department of Biochemistry and Biotechnology (URV) santi.garcia-vallve@urv.cat, aleix.gimeno@urv.cat

Data de convocatòria; Setembre, 2025

Jo, Judith Montano Hernández, amb DNI 41030494N, soc coneixedora de la guia de prevenció del plagi a la URV Prevenció, Detecció i Tractament del Plagi en la Docència: Guia per a Estudiants (aprovada el juliol de 2017) (http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formaciocompetencies-nuclears/plagi/) i afirmo que aquest TFG no constitueix cap de les conductes considerades com a plagi per la URV.

Tarragona, 3 de setembre de 2025

(signatura)

Table of Contents

1. Center Information

This thesis was developed during an internship at the Cheminformatics and Nutrition research group (QiN) in the Department of Biochemistry and Biotechnology at the Universitat Rovira i Virgili (URV). The QiN research group specializes in Computer-Aided Drug Discovery (CADD), and it is currently focused on the NEXT-PANDEMICS project. The NEXT-PANDEMICS project is designed to provide computational tools that support antiviral drug discovery against viruses posing a pandemic threat. This thesis was carried out as part of this project.

2. Abstract

Zika virus remains a global health concern because of its reemerging risk, severe complications, and lack of vaccines and antiviral drugs to treat it. Among Zika virus proteins, the NS2B/NS3 protease is an attractive target, as it is essential to process the viral polyprotein. However, active-site inhibition approaches have been challenging due to poor permeability and high homology between human serine proteases and the viral active center. In this work, a strategy focusing on allosteric inhibition is used to overcome those limitations and promote the inactive conformation of the NS2B/NS3 protease. To design an effective strategy, a structure-activity relationship interpretation was carried out to understand the inhibition mechanism of previously reported allosteric inhibitors. These insights were then integrated into a virtual screening workflow that combined protein-ligand docking, structure-based pharmacophore filtering, shape and electrostatic similarity filtering, and ADME evaluation. In total, twenty compounds were identified as promising allosteric inhibitors, which could be considered for future experimental evaluation.

Keywords: Zika Virus; NS2B/NS3 Protease; Allosteric Inhibitors; Virtual Screening; *In Silico* SAR.

3. Introduction

3.1. Zika Virus: an Overview of its Global Health Risk

The Zika virus (ZIKV) is a virus that forms part of the *Flaviviridae* family, closely related to Dengue virus (DENV), West Nile virus (WNV), Yellow Fever virus (YFV), and Japanese Encephalitis virus (JEV). It is a vector-borne virus, mainly spread by mosquitoes belonging to the Aedes genus. Nonetheless, the virus can also be transmitted through blood, vertical mother-to-child transmission, and sexual contact (Tajik et al., 2024). Whereas the first human transmission was reported in Nigeria in the 1950s, it was declared a global health risk by the World Health Organization (WHO) in 2016 when an epidemic affected the American continent. Most infected people are asymptomatic or develop mild symptoms (Mwaliko et al., 2021). However, it can cause severe neurological complications such as Guillain-Barré syndrome or transverse myelitis. In pregnancy, Zika virus infections can increase the risk of preterm birth and miscarriages, as well as Congenital Zika Syndrome (CZS), characterized by newborns presenting microcephaly and other congenital malformations. These neurological and pregnancy risks are the focus of concern (Rabe et al., 2025).

As of now, Zika virus incidence has declined globally, but its transmission remains present mainly across tropical and subtropical regions. At the same time, autochthonous cases have been reported in 92 countries, including previously non-endemic areas, reflecting the impact of climate change and transportation-driven spread. With more people now living in areas with competent *Aedes* and no prior ZIKV exposure, new outbreaks could occur (Rabe et al., 2025). Despite this situation, there is a lack of specific and effective vaccines or antiviral drugs approved to treat Zika virus infections. Therefore, the development of new treatments remains an urgent priority (Feng, 2024).

3.2. Genome and Infectious Cycle of Zika Virus

ZIKV is an enveloped, positive-sense single-stranded RNA virus (10.8 kb approximately) whose genome is formed by two flanking untranslated regions (UTR) and an open reading frame (ORF) that encodes for a single polyprotein. This polyprotein is then processed by viral and host proteases into three structural proteins (C, prM/M, and E) and seven non-structural proteins (NS1, NS2A, NS2B,

NS3, NS4A, NS4B, and NS5). The structural proteins form an icosahedral capsid surrounded by the lipidic envelope, which confers ZIKV a surprisingly high adaptability to hot climates. The non-structural proteins have roles in ZIKV replication as well as packaging of the genome (Feng, 2024; Sirohi & Kuhn, 2017).

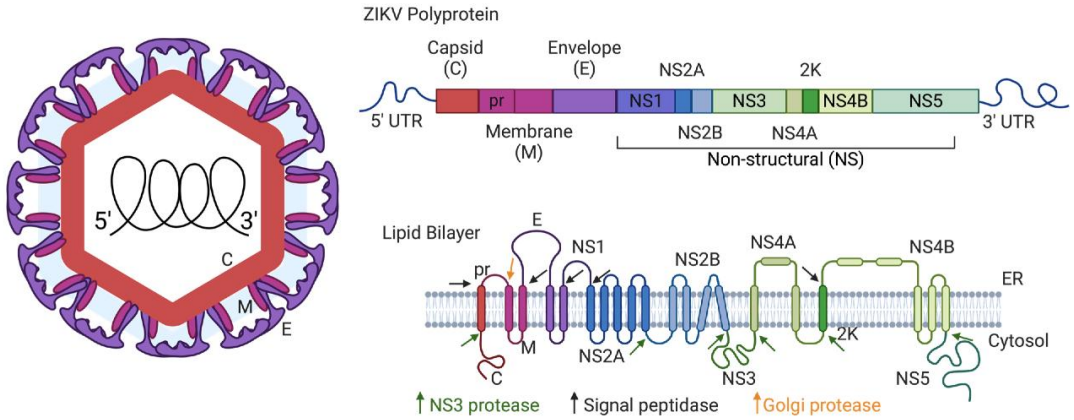


Figure 1. *Representation of ZIKV virion and genome structure. The lipidic envelope is not represented. Polyprotein cleavage sites in the endoplasmic reticulum are marked with colored arrows. Adapted from (Metzler & Tang, 2024).*

When a human is bitten by an *Aedes* mosquito infected with ZIKV, the virions released from its salivary glands penetrate the skin and initiate the infection process (Figure 2). The E protein binds to permissive receptors at the cell surface and then is internalized through clathrin-mediated endocytosis. Progressive acidification within the endosome triggers fusion between the viral membrane and the endosomal membrane, causing the RNA genome to be released into the cytoplasm. The positive-sense RNA is translated on the endoplasmic reticulum membranes as a polyprotein that is then cleaved by host peptidases and by the viral NS2B/NS3 protease to generate the ten proteins required for a complete replication process. These proteins reorganize the endoplasmic reticulum membranes into replication organelles where the RNA-dependent RNA polymerase NS5 synthesizes a short-lived negative-strand intermediate used to transcribe the new genomic RNAs. Viral assembly occurs at the endoplasmic reticulum membranes, where C protein associates with newly formed genomic RNA. This forms immature virions that also contain the prM and E proteins but are still non-infectious. As they travel through the trans-Golgi network, the acidic environment exposes the prM cleavage site, allowing host furin proteases to

process prM into M. The now mature virions are released by exocytosis (Metzler & Tang, 2024; Sirohi & Kuhn, 2017).

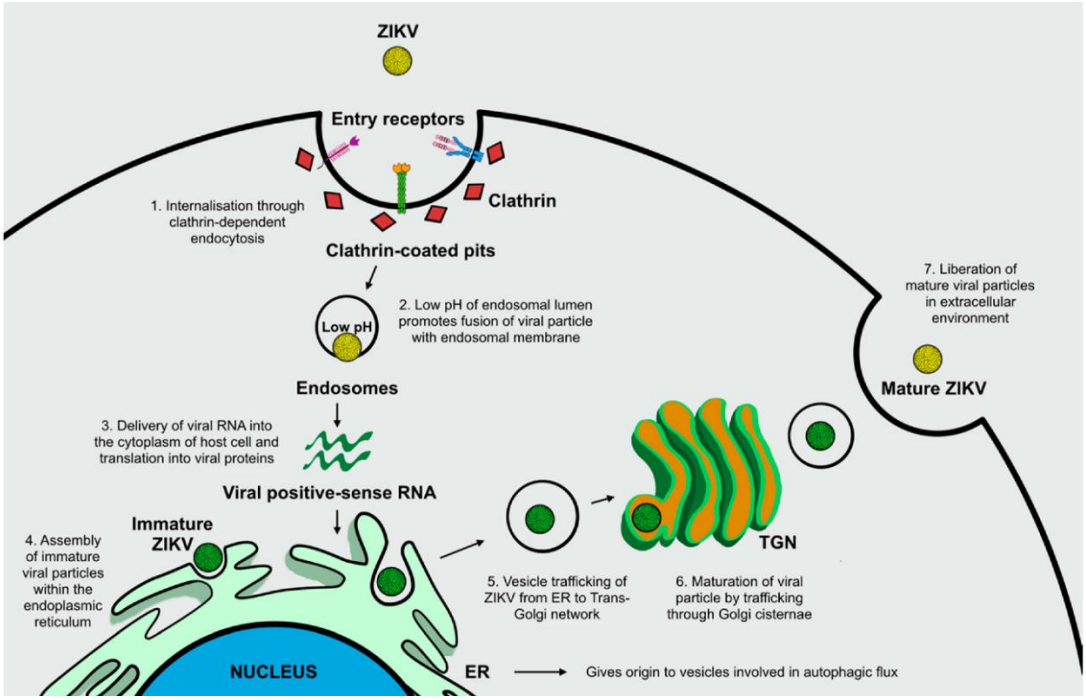


Figure 2. *Scheme of ZIKV replication cycle. Adapted from (Gratton et al., 2019).*

As described above, the NS2B/NS3 plays a central role in ZIKV replication cycle because it is responsible for processing the viral polyprotein. Without its activity, the structural and non-structural proteins required for genome replication and virion assembly would not be produced (Figure 1). That represents the main factor that makes NS2B/NS3 protease an interesting target for antiviral drug discovery (Meewan et al., 2023).

3.3. Functional and Structural Features of the NS2B/NS3 Protease

ZIKV NS3 protein is a multifunctional protein that contains an N-terminal trypsin-like serine protease (NS3pro) and a C-terminal helicase, which also shows RNA triphosphatase and NTPase activities. To adopt its catalytic activity, the NS3pro domain associates with a hydrophilic fragment of NS2B, thus forming the NS2B/NS3 protease (Shiryaev et al., 2023). The protease active site includes the catalytic triad formed by His51, Asp75, and Ser135 (Figure 3A, 3C) (Maus et al., 2023).

The NS2B/NS3 protease has been described to adopt two different conformations: closed and open (Figure 3A). In the closed state, NS2B wraps

around the NS3pro core. In the open state, NS2B is more loosely associated with the NS3pro core, which prevents the active site from adopting the correct folding. These conformational dynamics seem to be similar among flaviviruses, including DENV. Nonetheless, it has been noticed that a loop within residues 152-167 of the ZIKV NS3pro (open state) has a different position compared to the DENV protease. Also, the C-terminal of the NS2B protein presents a higher flexibility than its DENV counterpart (Maus et al., 2023). For that reason, some studies propose that the open conformation described for ZIKV might be a super-open conformation, where NS2B is further displaced from NS3pro and NS3pro C-terminal residues display a different folding (Shiryaev et al., 2023). Despite this lack of consensus, the general agreement is that the ZIKV NS2B/NS3 protease depends on an equilibrium between a closed, catalytically active state and an open, inactive state.

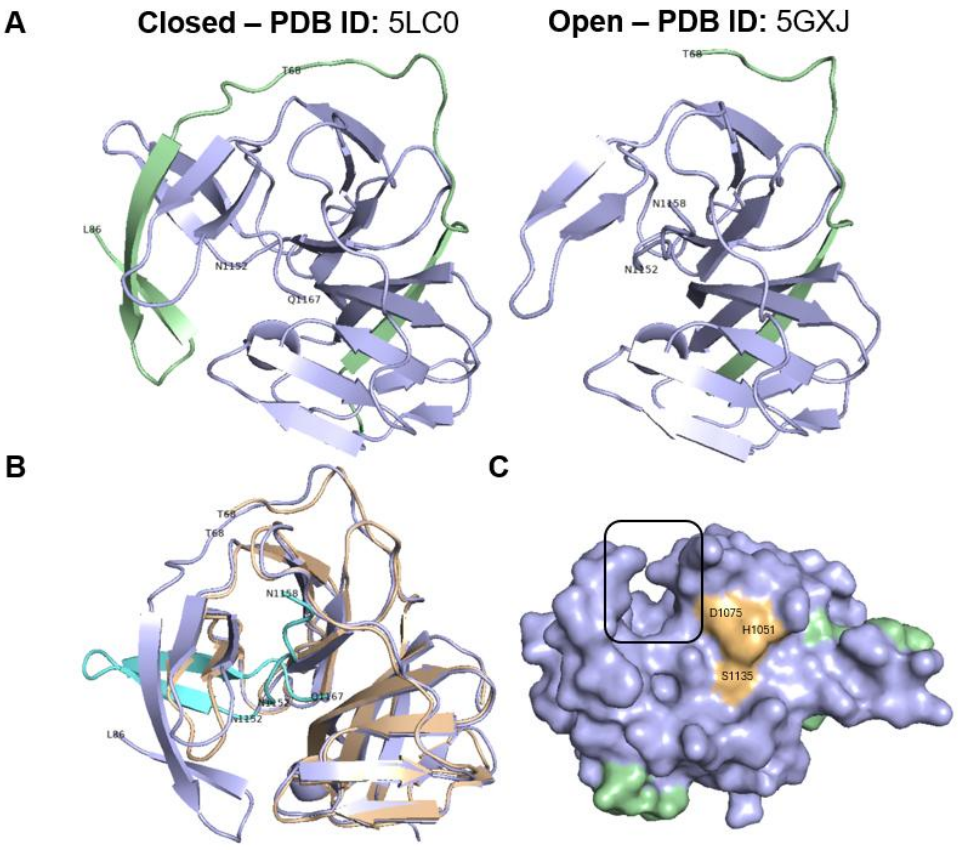


Figure 3. *Illustration of ZIKV NS2B/NS3 protease. (A) Conformational dynamics of the NS2B/NS3 protease. NS2B is illustrated in green, whereas NS3pro is illustrated in blue. 5GXJ (open) lacks the C-terminal region of NS2B, due to its intrinsic flexibility in the open conformation. (B) Structural superposition of 5LC0 (blue) and 5GXJ (orange). The conformational change in the loop 152-167 is shown in cyan. (C) Surface of the NS2B/NS3 protease. In yellow, the catalytic triad (His51, Asp75, and Ser135) is shown. The location of the allosteric pocket is outlined in black. Obtained with PyMol (Schrödinger, 2025e).*

Further structural studies revealed that two loops are responsible for the conformational switch described (Figure 3A, 3B). On NS2B, the loop found in residues 69-87 adopts different conformations between open and closed states. It has been described that this region helps to assemble a negatively charged region (S2) important for substrate recognition in the active site. On NS3pro, the loop formed by residues 152-167 is involved in the formation of another region (S1) necessary for substrate recognition and shows distinct conformations in both states. Deleting them results in the inhibition of the protease activity (Chen et al., 2016). Specifically, a mutational assay pointed to Asn152 as the molecular switch between the open and closed conformation (Brecher et al., 2017).

An allosteric pocket has been identified in the region around Asn152 (Figure 3C), which becomes accessible once NS2B becomes loosely associated with NS3pro (open state). This pocket is relatively conserved among flaviviruses and contains a hydrophobic region, thus representing a druggable target for small molecule inhibitors (Shiryaev et al., 2023).

3.4. Competitive versus Allosteric Approaches for NS2B/NS3 Protease Inhibition

Given the structural and functional characteristics of the NS2B/NS3 protease, two main strategies have been approached in drug discovery: competitive and allosteric inhibition.

Early efforts in this field focused on the development of competitive inhibitors directed at the active site of the protease. However, since the active center is highly conserved between flaviviruses and shares high homology with human serine proteases, there is a risk of off-target effects and low selectivity (Shiryaev et al., 2023). Furthermore, due to the active-site preference for polar molecules, identified inhibitors present low cell permeability and, in consequence, low bioavailability (Millies et al., 2019).

For that reason, attention has shifted towards the identification of non-competitive inhibitors. By doing this, the strategy focuses on stabilizing the inactive NS2B/NS3 conformation and preventing its activity. As the described allosteric pocket presents hydrophobic regions, non-competitive inhibitors have better

pharmacokinetic properties. Moreover, it has been shown that non-competitive inhibitors have no effect on host serine proteases, preventing potential cytotoxicity (Shiryaev et al., 2023). This work focuses on exploiting this strategy and aims to identify new small-molecule candidates through a virtual screening workflow based on an analysis of molecular mechanisms and features of known allosteric inhibitors.

3.5. Computational Approaches for Virtual Screening

Virtual screening is a computer-aided drug discovery (CADD) technique used to select molecules from a large library that are most likely to bind and have a particular activity against a target (Gimeno et al., 2019). In most cases, a virtual screening workflow consists of multiple steps to filter down the number of ligand candidates with the desired features (Gimeno et al., 2019; Vázquez et al., 2020). Compared to experimental high-throughput screening, virtual screening can screen millions of compounds at a remarkably lower cost and effort, making it a popular approach in the early stages of drug discovery. Virtual screening can be developed in two different approaches: ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS). In LBVS, known actives are employed as a reference to filter library compounds by molecular similarity using descriptors that capture 2D and 3D information, such as fields, shape, or ligand-based pharmacophores. In contrast, SBVS uses a 3D model of the target, generally obtained through protein-ligand docking or structure-based pharmacophores, to rank compounds based on receptor-ligand complementarity. Developing a mixed strategy with LBVS and SBVS approaches often improves hit retrieval (Vázquez et al., 2020).

To develop an effective virtual screening strategy, researching the target and prior SAR studies of known ligands is a must, as it helps select methodologies and interpret results. Also, it is crucial to prepare both the compound library and the protein structure. Protein preparation is necessary to correct problems in experimental structures. It usually consists of adding missing hydrogen atoms, assigning the protonation state at the desired pH, removing structural artifacts, applying terminal caps, and filling in missing side chains. For its part, ligand preparation generates 3D conformations and assigns protonation, tautomer, and stereochemistry at the working pH (Gimeno et al., 2019).

Before applying the virtual screening strategy to the compound library, the workflow must be validated. This is important because, like any prediction model, it is not error-free. Virtual screening validation is performed using a validation library that contains known active ligands and decoys (i.e., compounds structurally similar to the active ligands but presumed to be inactive). The validation library is then evaluated using specific metrics such as the enrichment factor (EF). This metric quantifies the extent to which a step increases the fraction of actives over the baseline (Gimeno et al., 2019).

3.5.1. ADME Filtering

ADME (Absorption, distribution, metabolism, and excretion) filtering is a key early step in virtual screening that allows to narrow the chemical space of large libraries, prioritizing compounds with drug-like properties before more costly computational steps are applied. In practice, it combines physicochemical properties (e.g., molecular weight, solubility, lipophilicity), pharmacokinetic properties (e.g., gastrointestinal absorption, permeability), and medicinal-chemistry alerts (e.g., pan-assay interference compounds or PAINS, structural alerts/reactive groups). It is necessary to point out that these filters serve as a guide, and strict rules or the blind use of alerts can eliminate potential drug candidates. A good way to minimize this limitation is to keep the rules flexible enough and adjust the criteria to the project context (Kralj et al., 2023). Additionally, ADME filtering can be used at the end of the process to further assess the properties of the selected compounds.

3.5.2. Protein-Ligand Docking

Protein-ligand docking is a computational technique designed to predict how a molecule binds to a protein binding site (Gimeno et al., 2019). To make this prediction, docking uses a protein structure and a candidate ligand, exploring the defined binding site for possible orientations, called binding poses, and scoring them to return the poses with the lowest predicted energy. Before docking, the binding site should be delimited in the form of a docking grid (Gimeno et al., 2019; Paggi et al., 2025). Also, constraints can be introduced to require key contacts (e.g., hydrogen bonds) or occupancy within the pocket (Gimeno et al., 2019).

As with any predictive model, the proposed poses and scores are not definitive proof of binding and may not accurately reflect the actual biological pose, especially since docking usually employs rigid protein structures. Despite that, docking has become an essential tool that provides a relatively fast way to screen large compound libraries and helps remove compounds that do not fit the binding site (Gimeno et al., 2019; Paggi et al., 2025).

3.5.3. Structure-Based and Ligand-Based Pharmacophore Modeling

A pharmacophore is a model that summarizes the essential three-dimensional features that must be present in a ligand to ensure its biological activity against a target. Typical features include hydrogen bond donors or acceptors, positive and negative ionic groups, hydrophobic or aromatic groups, as well as excluded volumes that represent the shape of the binding site or the space occupied by inactive compounds. These features are represented by spheres, vectors, or planes (Giordano et al., 2022).

There are two main methods to develop a pharmacophore model: structure-based and ligand-based. On the one hand, for the structure-based approach, features are derived from a protein-ligand complex or an apo protein structure, which collects information about the defined ligand conformation (protein-ligand complex), protein-ligand interactions (protein-ligand complex), and binding-site characteristics. On the other hand, in the ligand-based approach, a group of known active compounds is aligned into possible conformations to extract the features they share. This strategy is particularly useful when no protein-ligand complex is available, as it can provide relevant information about possible ligand binding mechanisms (Seidel et al., 2020).

Pharmacophore models are used for rapid screening of libraries, filtering poses after docking, and rationalizing structure-activity relationships (Seidel et al., 2020). However, it is good practice to critically assess the generated models and make manual modifications, as ligand-based models are sensitive to the quality of alignment/conformer generation, as well as the number of actives/inactives, and structure-based models may include irrelevant pharmacophoric features (Giordano et al., 2022; Seidel et al., 2020).

3.5.4. Electrostatic and Shape Similarity Filtering

In virtual screening, electrostatic and shape similarity filtering are used as a selection process that keeps molecules whose shape and electrostatic potential resemble those of a reference molecule (Gimeno et al., 2019).

In the case of electrostatic similarity, the objective is to select compounds that can match the electrostatic environment of the binding site based on the knowledge that electrostatic complementarity is an important mechanism for ligand binding. As for shape similarity, it is assumed that molecules with a similar shape may fit into the same binding site. As it only considers shape, this technique helps to find new scaffolds. Nonetheless, it may select molecules that lack the chemical features required to be active. Thus, by combining it with electrostatic similarity, this limitation can be diminished (Gimeno et al., 2019).

3.6. *In Silico* Interpretation of Structure-Activity Relationships

In silico interpretation of structure-activity relationships (SAR) is a computational approach that aligns a set of ligands and helps examine how structural and chemical characteristics relate to activity by analyzing three-dimensional molecular fields that represent electrostatic, steric, and hydrophobic features as field points. This field-based description can help rationalize how known ligands get recognized in a protein binding site (Cheeseright et al., 2006b). To carry out this analysis, activity cliffs are used, defined as pairs of similar molecules that present high activity differences (Dablander et al., 2023). Methods such as Activity Atlas from Flare (Cresset, 2025) create a 3D qualitative model that summarizes the average physicochemical features shared by ligands, as well as activity cliff patterns of favorable/unfavorable electrostatic, hydrophobic, and shape regions across aligned molecules (Attiq et al., 2022). Then, activity cliffs between pairs of ligands can be assessed through other methods, which helps to pinpoint specific physicochemical changes behind an activity shift. Together with protein binding site context and ligand-protein interactions, this analysis can help develop a shared molecular mechanism among known ligands, informing the development of a virtual screening strategy (Dablander et al., 2023).

4. Hypothesis and Objectives

This project hypothesizes that integrating ligand-based and structure-based knowledge can support the development of a selective virtual screening strategy to identify potential allosteric inhibitors for the Zika virus NS2B/NS3 protease.

With that purpose in mind, this project has the following objectives:

1. Analyze known active compounds to understand their mechanism of action via SAR interpretation.
2. Use protein-ligand docking, pharmacophore models, shape/electrostatic comparisons, and ADME filters to develop a rational virtual screening strategy.
3. Identify novel allosteric inhibitors for Zika virus NS2B/NS3 protease.

5. Materials and Methods

A virtual screening workflow combining ligand-based and structure-based methods was implemented to identify allosteric inhibitors of the ZIKV NS2B/NS3 protease. After selecting the protein structure and ligands, two ligand-based pharmacophores were generated to indirectly validate binding poses necessary to align molecules for SAR interpretation. This analysis was conducted to gain insights into the relevant features present in known allosteric inhibitors. With this pre-analysis complete, a virtual screening strategy was developed based on the insights gained. The steps followed were protein-ligand docking, structure-based pharmacophore filtering, and shape/charge similarity filtering. Ultimately, post-screening clustering, ADME assessment, and protein-ligand interaction analysis were performed for the remaining molecules to select the final hit compounds. A simplified version of the process is represented in Figure 4.

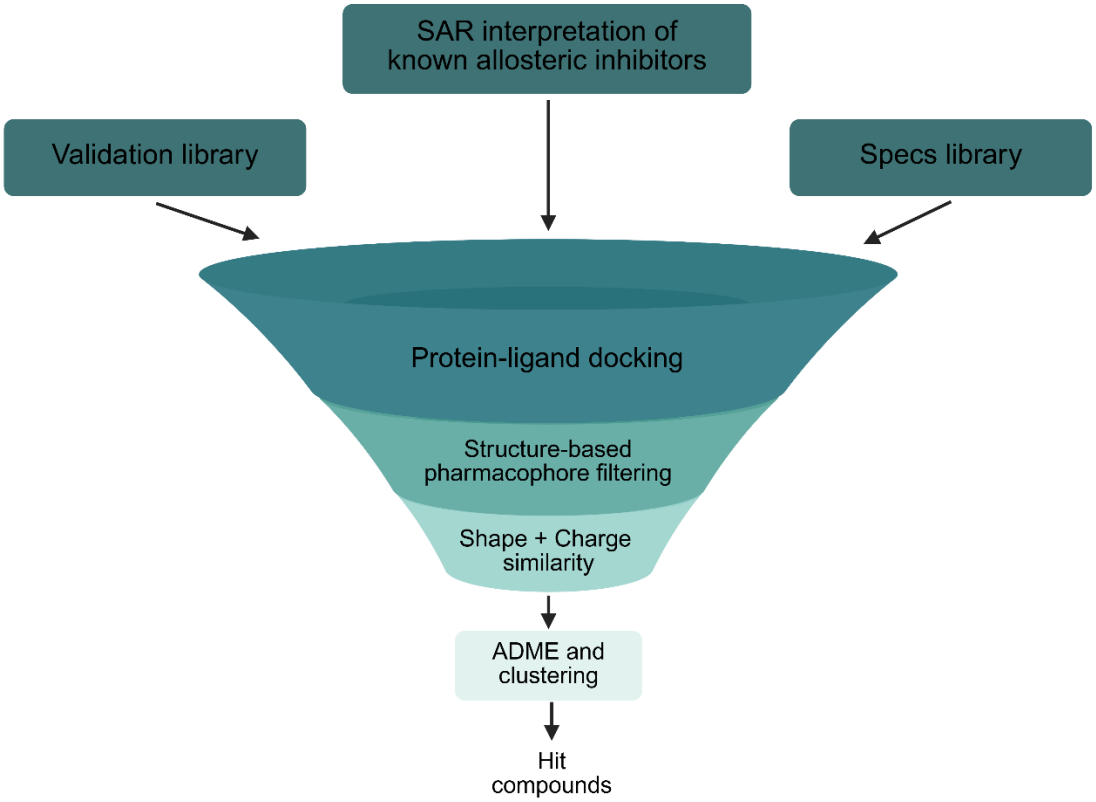


***Figure 4**. A simplified scheme of the virtual screening strategy followed. Created in https://BioRender.com.*

5.1. Protein Structure Selection and Preparation

The first step was to identify all available Zika virus NS2B/NS3 protease structures in an open conformation deposited in the Protein Data Bank (PDB) (Berman et al., 2000). The search was executed with the following query: “Source Organism Taxonomy Name (Full Lineage) - is - Zika Virus – AND - Structure Title - has exact phrase – NS2B-NS3 protease”. With this search, a total of 195 structures were obtained. To identify structures with an open conformation, the terms “apo”, “open”, “super-open” and “conformation” were manually searched in the structure titles and citation papers. A total of six structures were found (Table 1), of which only one was co-crystallized with an inhibitor (PDB ID: 7M1V).

PDB ID	Ligand PubChem CID	Resolution
7M1V	257767	1.60 Å
5GXJ	-	2.60 Å
5T1V	-	3.10 Å
5TFN	-	3.00 Å
5TFO	-	2.51 Å
6UM3	-	2.50 Å

Table 1. *Available Zika virus NS2B/NS3 protease with open conformation in the Protein Data Bank (PDB) (Berman et al., 2000).*

While the 7M1V structure is a mutant protein designed to stabilize the open conformation, it served as the working reference, as it has the best resolution (1.60 Å), and the bound inhibitor facilitates the identification of the allosteric site and the definition of docking grids. Moreover, a structural comparison analysis between 7M1V and 5GXJ proved that the mutations did not cause major conformational differences (Figure 5), with a Root Mean Square Deviation (RMSD) of 0.38 Å. For reference, RMSD is a measure of the average distance between atomic positions of two aligned structures, where a lower value indicates higher structural similarity. Generally, two structures are considered the same if their RMSD value is lower than ≤2.00 Å (Vittorio et al., 2024). The structural comparison was carried out in PyMOL (Schrödinger, 2025e). To ensure optimal structural geometry for the virtual screening, the corresponding refined and rebuilt model from the PDB-REDO (Joosten et al., 2014) database was used throughout the virtual screening process.

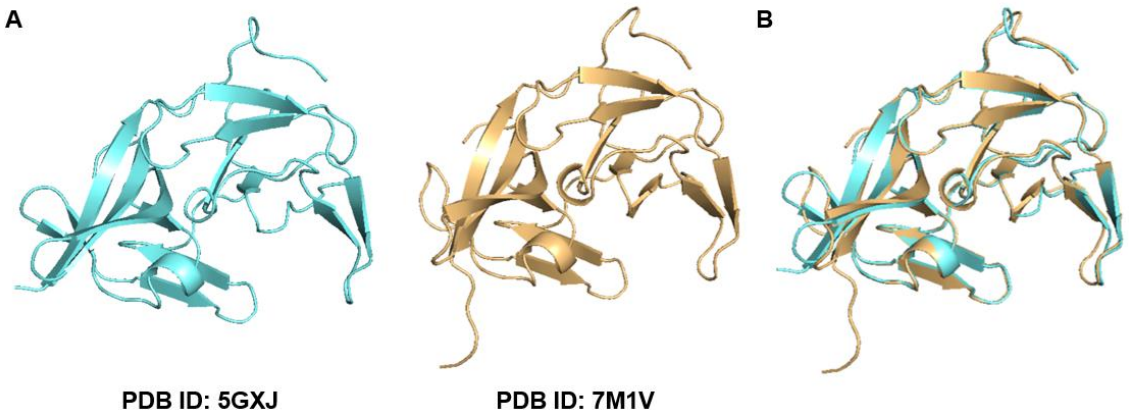


Figure 5. *Structural comparison between 7M1V and 5GXJ structures. (A) Side-to-side comparison. 5GXJ is represented in cyan, and 7M1V is represented in orange. (B) Superposition of both structures. Visually, no major conformational difference is seen, and the superposition has an RMSD value of 0.38 Å. Obtained with PyMOL (Schrödinger, 2025e).*

The primary protein preparation was performed in Flare (Cresset, 2025) with the following settings: the calculation method was set to normal, and intelligent capping was applied to cap the N- and C-termini. The protonation state was set at pH 7.00, allowing changes in protonation and tautomeric states. As the 7M1V biological assembly corresponds to a homodimer, only chain A (containing the bound inhibitor) was conserved, and chain B was removed. Additionally, all water molecules were removed to work with the dry protein, as they can be crystallographic artifacts, and increase the complexity of docking calculations.

Later in the virtual screening process, when the structure was imported into Maestro (Schrödinger, 2025c), preprocessing with the Protein Preparation Workflow (Madhavi Sastry et al., 2013) was performed to cap all termini, as they were incorrectly interpreted as missing atoms.

5.2. Ligand Selection and Preparation

Known active compounds against Zika virus NS2B/NS3 protease were identified from the available literature (Maus et al., 2021; Millies et al., 2019; Nie et al., 2021; Shiryaev et al., 2017; Yao et al., 2019) and then collected from PubChem (Kim et al., 2025), prioritizing those with reported inhibitory activity (IC50). To explore structurally related groups, molecules were clustered based on their 2D molecular fingerprints (binary digits representing structural features). These molecular fingerprints were used to calculate the Tanimoto similarity between compounds, ranging from 0 to 1, where 1 represents the maximum similarity. A Tanimoto similarity cutoff of 0.60 was established, and compounds with a

Tanimoto similarity greater than 0.60 formed a cluster. Lastly, a dendrogram was used to visualize the results. Two main series of active compounds (cluster 1 and cluster 2) were selected (see Results 6.1 for specific description).

To complete the validation library, a set of decoys was identified using compounds from the ZINC20 database (Irwin et al., 2020) and the DecoyFinder software (Cereto Massague et al., 2012). For each active molecule, 50 decoys were calculated, generating a total of 8,250 decoys.

Ligand preparation was divided into two groups, depending on the process for which the ligands were used. The first preparation, used for SAR interpretation, was performed in Flare (Cresset, 2025), applying rule-based protonation. The second preparation was used for ligand-based pharmacophore modeling and virtual screening validation. This preparation involved a two-step process. First, compounds were processed with Omega (Hawkins et al., 2010) to eliminate those with undefined chirality. Then, they were prepared using LigPrep (Schrödinger, 2025b). LigPrep was used to generate protonation states at pH 7.00 ± 1.00 using Epik (Johnston et al., 2023), to activate tautomer generation, and to generate stereoisomers that retain the specified chirality.

5.3. Ligand-Based Pharmacophore Modeling for Binding Pose Selection

Because docking poses for cluster 2 compounds were highly flexible (Millies et al., 2019), and no experimental binding poses exist for the two main clusters of active compounds to use as reference, the initial SAR analysis proved inconclusive. To overcome these limitations and try to indirectly define possible poses, a ligand-based pharmacophore strategy was developed using Phase (Dixon et al., 2006). To build these pharmacophores, a selection of compounds from each cluster of active compounds was made, including the most active and the least active compounds. The compounds with a pIC50 ≤5.00 were considered inactive and used to create the excluded volume shell. The pharmacophore hypothesis was constructed with the following settings: number of features 3 to 5 (with 4 being the preferred minimum), the hypothesis should match at least 90% of ligands, and for cluster 2, two acceptors were required. Then, the pharmacophore method was used to find the best alignment and common features, and 50 conformers were generated for each compound. The aligned

ligand conformations from the pharmacophore hypothesis with the best score (PhaseHypoScore) for each cluster were selected as the alignment reference.

The aligned ligand conformations were then used to superpose previously generated docking poses of the most active compounds (the same actives used for the pharmacophores). These docking poses were obtained using Flare, employing the crystallized ligand from the 7M1V structure to define the grid. To create as many accurate conformations as possible, the docking was performed using the "Very Accurate but Slow" calculation method with extra precision quality, and a maximum of 20 poses was generated for every selected ligand. The superposition was performed for all generated poses using the Superposition tool from Maestro (Schrödinger, 2025c), with "Atom Pairs" selected as the method. For cluster 1, three atom pairs were used, and for cluster 2, four atom pairs were selected by visual inspection (see Results 6.2 section for specific description). Afterwards, the poses with the best RMSD value for each cluster were chosen as the reference docking poses.

5.4. Interpretation of Structure-Activity Relationships

The SAR interpretation was performed separately on the two clusters of active compounds selected for the validation library. To analyze them, it is necessary to align each active compound with a reference. The references utilized were the predicted docking poses indirectly validated by the ligand-based pharmacophores. Using the component Conf Hunt and Align from Flare (Cresset, 2025), all actives from each cluster were aligned to the corresponding reference. The conformational hunt was run in the accurate but slow mode, and the alignment was performed applying both substructure and normal (shape + field) modes. The alignment mode that worked best for each cluster was selected by visual inspection: normal mode for cluster 2 and substructure for cluster 1.

With the aligned compounds, Activity Atlas (Cheeseright et al., 2006a) was applied to generate a 3D landscape that included the average features of the ligands, as well as activity cliff summaries of electrostatics, hydrophobicity, and shape.

5.5. Screening Library Preparation and ADME Filtering

The screening library used for this virtual screening was the Specs library (Wu et al., 2019), which consists of single synthesized, well-characterized and drug-like small molecules. The Specs database selected was “Specs_SC_10mg” from June 2025, containing up to 327,132 compounds. To prepare the Specs library for virtual screening, the same protocol used for the validation library was performed (see section 5.2).

After the preparation, the library was subjected to ADME filtering to eliminate compounds with undesirable physicochemical features. The chosen criteria were the following one: firstly, compounds with high risk of giving false positives were discarded using a PAINS filter present in Canvas (Schrödinger, 2025a). Secondly, compounds that had more than 1 violation of the Lipinski’s rule (Molecular Weight ≤ 500 g/mol, MLOGP ≤ 4.15, N or O ≤ 10, and NH or OH ≤ 5) were removed. This was performed using QikProp (Schrödinger, 2025d) and the ligand filtering component from Maestro (Schrödinger, 2025c). This ADME filtering reduced the number of compounds in the library from 327,132 to 230,624.

5.6. Protein-Ligand Docking Configuration

In this virtual screening workflow, docking was performed solely to discard molecules that could not fit the binding site, and to generate possible binding poses for each of them, but not to estimate affinity. This was based on the previous knowledge of unstable pose generation for cluster 2 of the validation library (Millies et al., 2019). For this reason, no validation was conducted, and only one grid was built.

To construct the docking grid, the predicted protein-ligand complex of reference (Figure 7) from cluster 1 (most stable) was employed instead of the crystallographic protein-ligand complex. This grid had the corresponding centroid coordinates of the predicted ligand pose (x=5.04, y=0.60, z=-10.30), an enclosing box of 15 Å, and a ligand diameter midpoint box of length x=9.00 Å, y=9.00 Å, and z=9.00 Å. The objective was to assemble a more restrictive grid that could moderate the ligand movement in the binding site. The docking was carried out with Glide SP (Friesner et al., 2004), generating 10 poses per ligand, so that at

least one of them could represent the intended pose before pharmacophore filtering.

5.7. Structure-Based Pharmacophore Modeling

Two independent structure-based pharmacophores were constructed using Phase (Dixon et al., 2006), based on the predicted protein-ligand complexes of reference (Figure 7). These structures were used to manually pick pharmacophoric features based on the SAR analysis results. Moreover, a receptor-based excluded volume was created to emulate the binding site occupancy.

The poses generated in the docking (validation library and Specs library) were screened against the structure-based pharmacophores, using the "Use existing conformers and Score in place" settings. The compounds that did not match the features in the pharmacophore were discarded. To validate the pharmacophore models, the results from the screening of the validation library were used to calculate the EF. Multiple pharmacophore models were tested for both protein-ligand complexes, and the two that produced the best EF are presented in this work (see Results 6.5).

The formula for calculating the EF was the following:

$$EF = \frac{\frac{a2}{a2+d2}}{\frac{a1}{a1+d1}} \quad (1)$$

In which: actives before the virtual screening step (a1), decoys before the virtual screening step (b1), actives after the virtual screening step (a2), and decoys before the virtual screening step (b2).

5.8. Shape and Charge Density Similarity Protocol

Shape and charge density similarity were evaluated using EON (OpenEye, 2025). Similarly to the structure-based pharmacophores, the predicted poses from the two clusters of actives of the validation library (Figure 7) served as the reference, also called a query. After calculation, EON reports the Shape Tanimoto Similarity and the Charge Tanimoto Similarity on a 0-1 scale, where values closer to 1 indicate higher similarity. Additionally, EON also calculates the Shape +

Charge Combo, which represents the sum of the Shape and Charge Tanimoto Similarity. To filter the Specs library, the validation library was applied to establish a threshold, or cut-off, based on the Shape + Charge Combo results.

5.9. Post-Screening Clustering, ADME Profiling and Protein-Ligand Interactions

After completing the full virtual screening workflow up to the shape and charge similarity filter, the remaining compounds were subjected to a clustering analysis. Molecules were clustered using Morgan fingerprints and a Tanimoto Similarity cutoff of 0.60. The results were then represented in a dendrogram. If clusters were formed, only the compound with the highest Shape + Charge Combo was retained.

The remaining compounds were then evaluated using SwissADME (Daina et al., 2017) to profile their physicochemical, pharmacokinetic, and medicinal chemistry properties. Three factors were considered: gastrointestinal (GI) absorption, blood-brain barrier (BBB) permeation, and the Brenk filter to exclude compounds with highly reactive or undesirable groups.

The subset that passed this ADME profiling was subjected to a second clustering step using the same parameters. However, this time the objective was to combine the compounds that passed the two independent virtual screening workflows, ensuring that there were no highly similar molecules between the two sets.

Finally, the protein-ligand interactions of the selected hits were analyzed with Arpeggio (Jubb et al., 2017) and CORAL-PIC. Arpeggio was utilized to calculate the interactions between the protein (7M1V) and the ligands. Afterwards, the interactions calculated were represented with CORAL-PIC, a program developed by Said Trujillo de León (Cheminformatics and Nutrition Research Group).

6. Results

6.1. Known Allosteric Inhibitors Clustering

A total of 173 allosteric inhibitors with reported IC50 values for ZIKV NS2B/NS3 protease were found in the literature. Among them, the crystallized ligand from the 7M1V is present (PubChem CID: 257767). All of them were subjected to similarity analysis and subsequent clustering. The results can be visualized in the dendrogram representation (Figure 6).

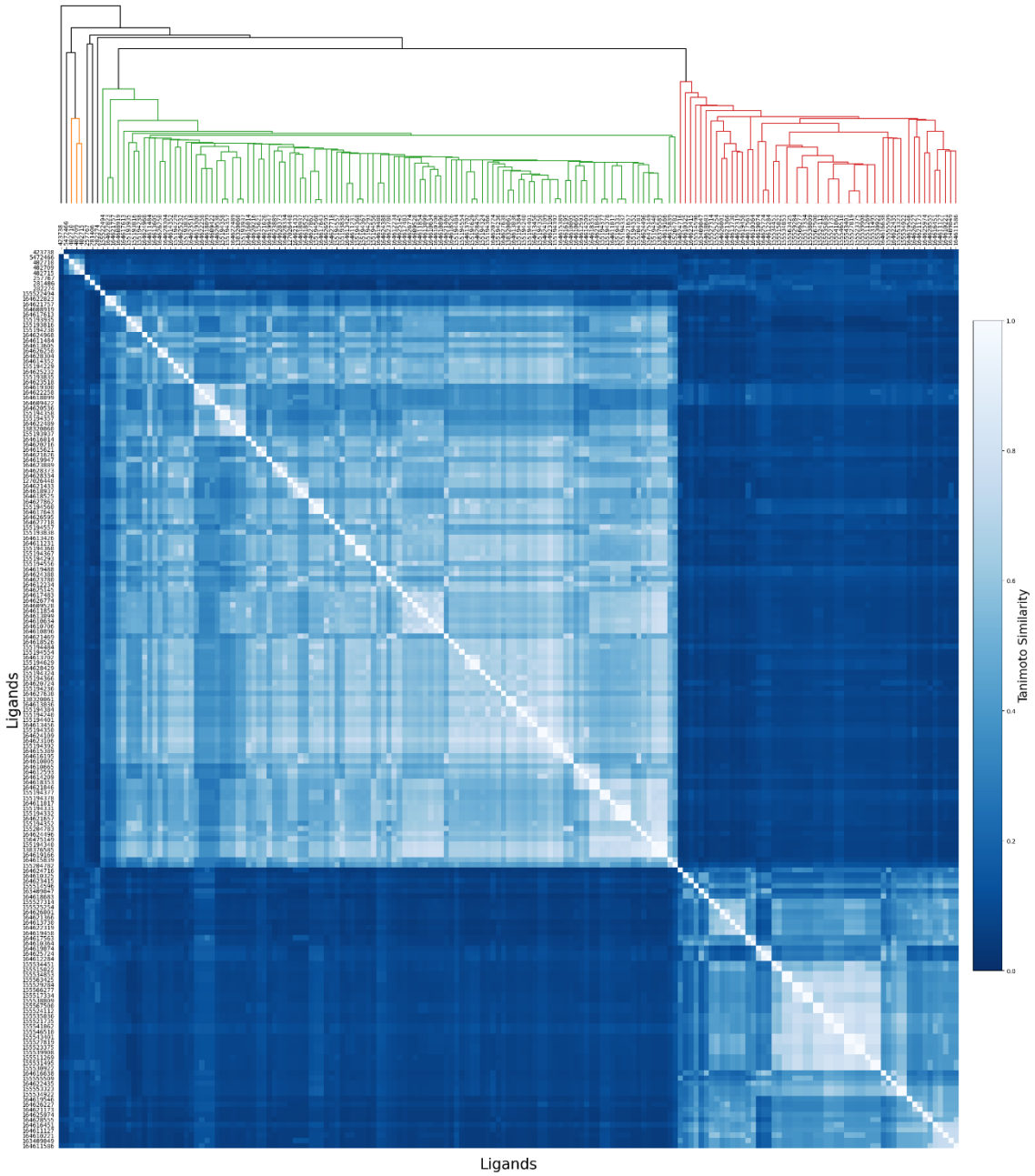


Figure 6. *Dendrogram representation of the Tanimoto similarity between the 173 known allosteric inhibitors against the ZIKV NS2B/NS3 protease.*

As can be seen, three clusters were identified. Each cluster was named in order of the number of ligands, cluster 1 being the one with the most ligands, and cluster 3 the one with the fewest ligands. Notedly, the crystallized ligand from 7M1V did not form part of any cluster.

In cluster 1 (green cluster), a total of 111 compounds were found. They are a group formed by three variable substituents (R2, R5, and R6) and a pyrazine-based core, with IC50 values ranging from 0.13 μM to >50 μM. SAR studies were conducted with these compounds (Nie et al., 2021), making them valuable for gaining knowledge about relevant and irrelevant physicochemical features for NS2B/NS3 protease allosteric inhibition. In cluster 2 (red cluster), a total of 54 compounds were found. In this case, they are a group of combined proline-based and benzo[d]thiazole-based compounds, with IC50 values from 0.32 μM to 21.9 μM. As compounds from cluster 1, SAR studies were developed with them (Maus et al., 2021; Millies et al., 2019). Additionally, docking studies were performed, complementing SAR studies and providing insights into protein-ligand interactions. Lastly, in cluster 3 (yellow cluster), only three compounds were found, with IC50 values ranging from 0.97 μM to >10 μM.

Both cluster 1 and cluster 2 were selected for further *in silico* SAR interpretation and to construct a validation library for virtual screening, based on the number of compounds available and the amount of prior research. Although the bound allosteric ligand in the 7M1V structure was the only experimental pose available, it was not considered for developing the virtual screening strategy as it was not related to the compounds from clusters 1 and 2.

6.2. Validation of Binding Poses by Superposition to Ligand-Based Pharmacophore References

The aligned active compounds used to produce the ligand-based pharmacophores served as a reference for superposing the corresponding predicted docking poses, as described in Section 5.3. The docking pose reporting the best superposition for cluster 1 was the first pose of the most active compound (PubChem CID: 155194384; IC50: 0.13 μM), with a RMSD value of 0.32 Å (Figure 7A). For cluster 2, the best superposition for cluster 2 was the fifth pose of the third most active compound (PubChem CID: 155539908; IC50: 0.58

μM), with a RMSD value of 0.65 Å (Figure 7B). The specific atoms selected for the superposition are shown in Figure 7.

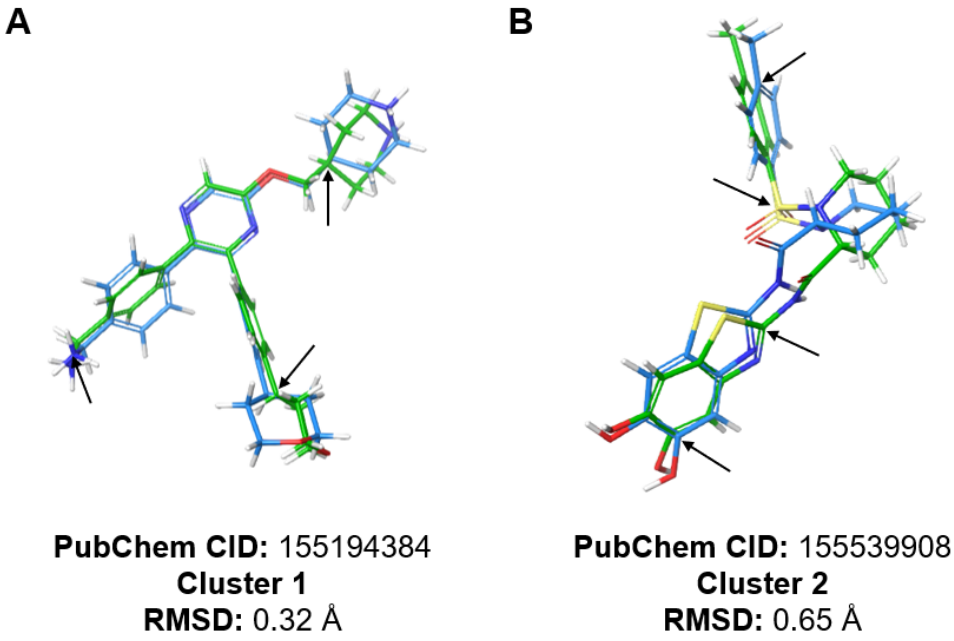


***Figure 7**. Superposition of the docking poses with the best RMSD from cluster 1 (A) and cluster 2 (B). The conformation reference from the ligand-based pharmacophore is represented in blue, whereas the docking poses are represented in green. The atoms selected to superpose are marked with black arrows. Obtained with Maestro (Schrödinger, 2025c).*

By comparing and finding a consensus between poses obtained using two different methods, this methodology enabled the indirect validation of binding poses in the context of a lack of experimental poses and high conformational flexibility. These binding poses and their respective protein-ligand complex with 7M1V were retained as references for subsequent SAR interpretation, to develop the docking grid (cluster 1, PubChem CID: 155194384), and for the construction of the structure-based pharmacophore models.

6.3. Interpretation of Structure-Activity Relationships: Physicochemical Patterns and Binding Mechanism Hypothesis

The SAR analysis was carried out independently for cluster 1 and cluster 2. For each series, conformations and alignments were obtained using their respective validated reference conformations (Figure 7). Then, the 3D landscape of their SAR was generated with Activity Atlas (Cheeseright et al., 2006a) and was inspected within the binding site context to identify general patterns or features likely related to allosteric inhibition.

Although performed separately, a common picture emerged from both series. The compounds showed a predominantly hydrophobic character, with an activity cliff highlighting favorable hydrophobic volumes in contact with a hydrophobic pocket in the allosteric site (Figure 8A, 8B, and 8C). This hydrophobic pocket is precisely the one that the NS2B protein renders accessible when it dissociates from the NS2B/NS3 protease, which potentially underlines a mechanism of allosteric inhibition. Furthermore, both the benzothiazole core (cluster 1) and the R6 substituent (cluster 2) were found to bind to the same subpocket of the allosteric site (Figure 8C). These groups have been reported to enable allosteric inhibitors, in the case of the benzothiazole group, and to exhibit higher inhibitory activity, observed with the R6 substituent (Millies et al., 2019; Nie et al., 2021). For that reason, physicochemical features derived from them are found to be key for developing an effective pharmacophoric filter.

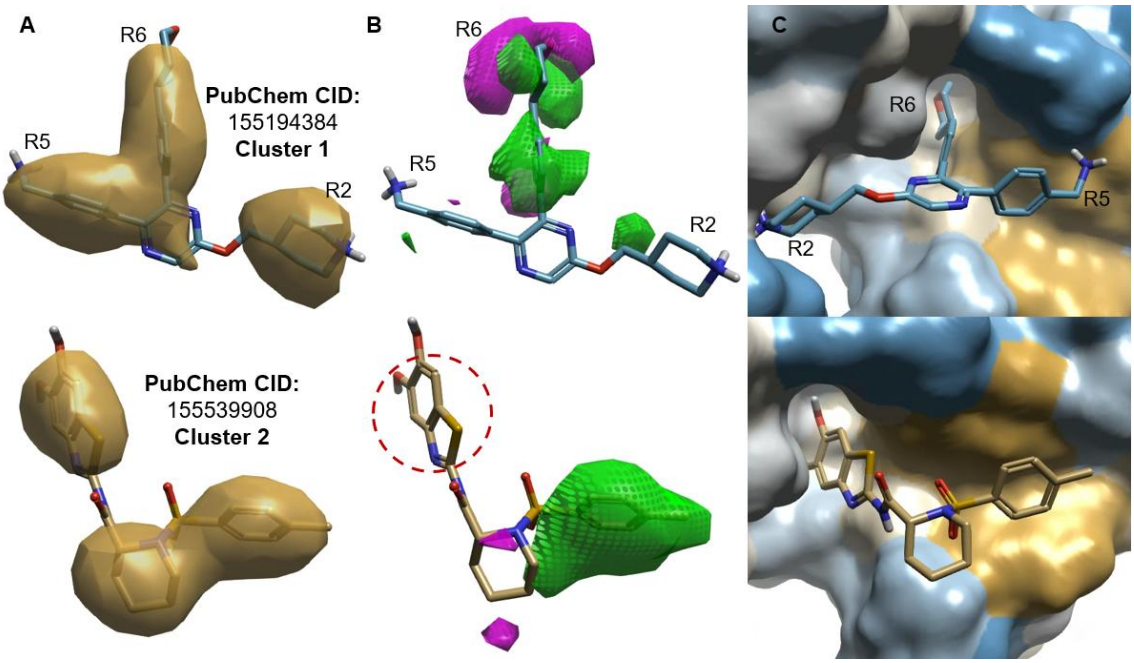


Figure 8. *Summary of the insights gained from Activity Atlas (Cheeseright et al., 2006a) using the reference ligands from both clusters. (A) Average hydrophobic field of ligands. (B) Activity cliffs for unfavorable hydrophobics (magenta) and favorable hydrophobics (green). The benzothiazole group is marked with a red circle. (C) Binding poses of the reference ligands within the protein allosteric pocket. The hydrophobic surface of the protein is illustrated (yellow=hydrophobic, blue=hydrophilic, white=neutral). Obtained with Flare (Cresset, 2025).*

To complement these findings, the predicted interactions of the most active and representative active ligands of each cluster were analyzed (Figure 9). As expected, all of them presented multiple hydrophobic/aromatic-aromatic interactions with the protein. Specifically, interactions with VAL1155 and

TRP1083 were conserved in all of them. The ligands also formed hydrogen bonds with residues such as LEU1149, LEU1076, THR1118, and ASP1122. More importantly, ligands from cluster 2 made hydrogen bonds with ASN1152, which is the molecular switch between the open and closed conformation. This interaction was made by one of the hydroxyl groups added to the benzothiazole core. Thus, this hydroxyl group represents a crucial structural feature to prevent the activity of the NS2B/NS3 protease.

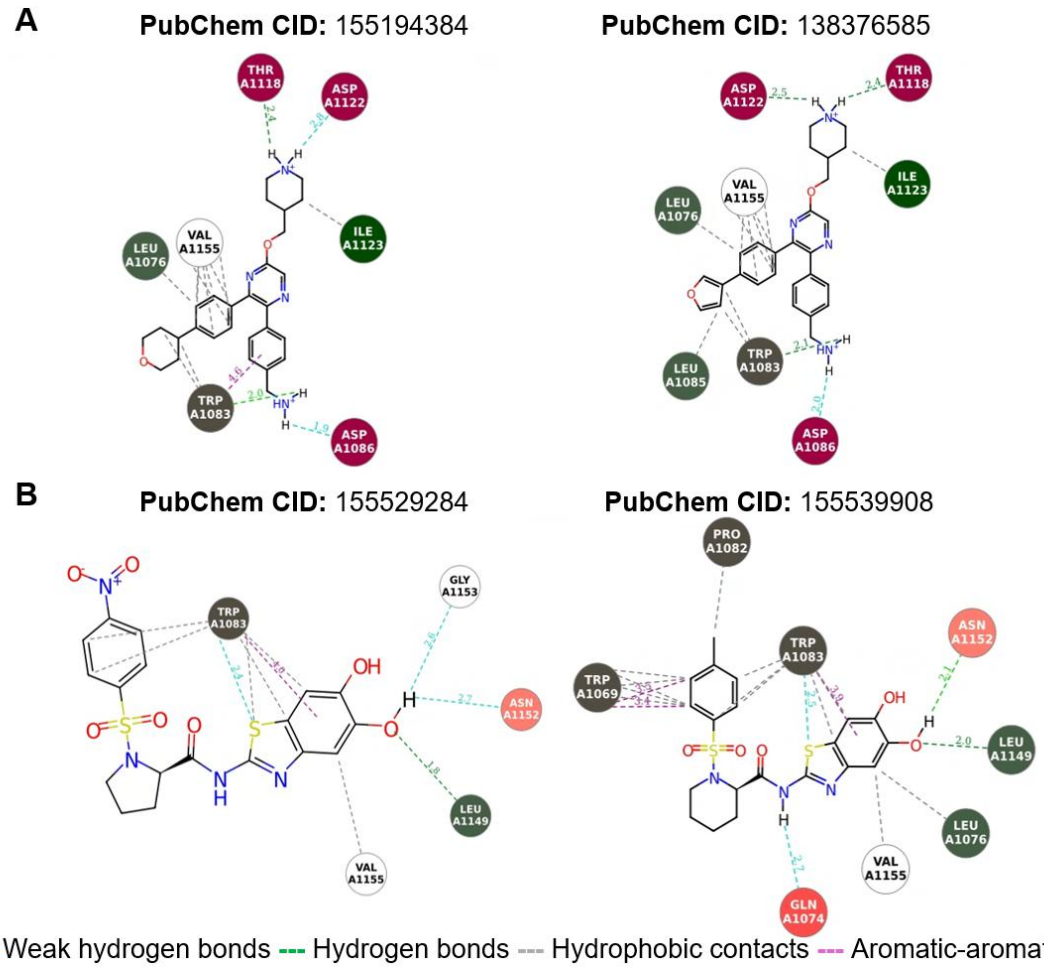


Figure 9. *2D diagrams of interactions from the most relevant and active compounds from cluster 1 (A) and cluster 2 (B).*

6.4. Protein-Ligand Docking Output for Validation Library and Specs Library

Docking was performed to generate possible poses and discard compounds that could not be accommodated in the binding site (see Section 5.6). For that reason, docking scores were not considered.

- Validation library: 38,288 poses for actives and decoys were generated. A total of 33 compounds returned no acceptable pose and were removed from the output.
- Specs library: 2,128,350 poses were generated. A total of 2395 compounds returned no acceptable pose and were removed from the output.

The docking poses generated were then used to screen against the structure-based pharmacophores.

6.5. Structure-Based Pharmacophores Validation and Screening

Despite the consensus in the SAR interpretation of cluster 1 and cluster 2, structure-based pharmacophores were built separately based on each cluster (Figure 10). This was decided to capture more cluster-dependent features and to maximize the identification of scaffolds in the screening.

- Model based on cluster 1 (Pharmacophore 1): this model included three required pharmacophoric features. As aromatic rings can engage in both π-π stacking and hydrophobic contacts, two aromatic rings were incorporated in the regions that created most interactions with the hydrophobic region of the allosteric site. One located in the R6 region (x=3.45, y=1.26, z=-10.27), and the other in the R5 region (x=5.87, y=3.57, z=-7.07). Hydrophobic contacts were also allowed to match these features. Lastly, a positive ionic feature was used in place of a donor in the R5 region that formed hydrogen bonds with the allosteric site (x=3.12, y=6.76, z=-5.67). This was decided because hydrogen bonds are heavily dependent on direction, and small changes in conformation could prevent potential hits from matching.
- Model based on cluster 2 (Pharmacophore 2): the second model included four required pharmacophoric features. Three aromatic rings were incorporated in the same fashion as in the first model. Two of them were in the benzothiazole region (x=2.04, y=1.98, z=-12.59; x=3.69, y=2.33, z=-11.16), and the other was in a region that also created a great number of hydrophobic/aromatic-aromatic interactions with the hydrophobic region (x=6.51, y=8.10, z=-6.57). The last feature was a positive ionic, used

similarly as in the model based on cluster 1. It was in the hydroxyl group region that formed a hydrogen bond with ASN152 (x=1.97, y=1.48, z=-15.79).

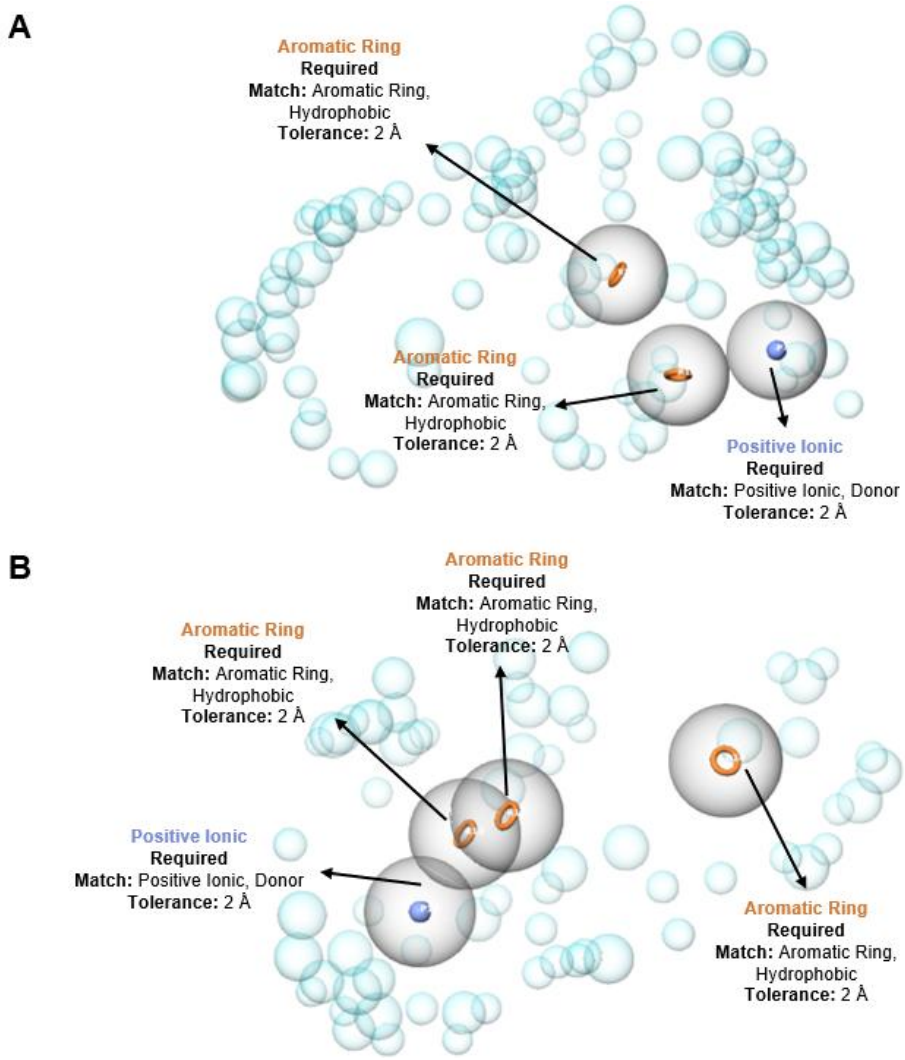


Figure 10. *Structure-based pharmacophore models based on cluster 1 (A) and cluster 2 (B). Obtained with Maestro (Schrödinger, 2025c).*

The docking poses previously obtained for the validation library were screened against the two pharmacophore models. A total of 288 poses passed pharmacophore 1, and 20 passed pharmacophore 2. To calculate the EF, only one pose per compound was counted. The EF of pharmacophore 1 was 10.11, and the EF of pharmacophore 2 was 12.84. After screening the docking poses generated for the Specs library, the number of poses was reduced to 3,735 for pharmacophore 1 and to 548 for pharmacophore 2.

6.6. Shape and Charge Density Similarity Screening

The final virtual screening step involved performing a similarity screening to compare the shape and charge of the compounds with the conformational references (Figure 7). First, it was necessary to set up a threshold with the validation library to then discard the Specs compounds that did not satisfy the cutoff. The distribution of shape + charge Tanimoto similarity (Shape + Charge Tanimoto Combo) for actives and decoys is illustrated in Figure 11. For cluster 1, actives and decoys were drastically separated at a 0.8 shape + charge Tanimoto combo value, thus easily establishing it as the threshold value. Nonetheless, the limited number of remaining compounds for cluster 2 complicated the decision. In this case, a stricter approach was preferred, as it reduces the risk of getting false positives even at the cost of excluding some active compounds. Therefore, the threshold was set to 0.8 as well, resulting in the retention of four actives and one decoy.

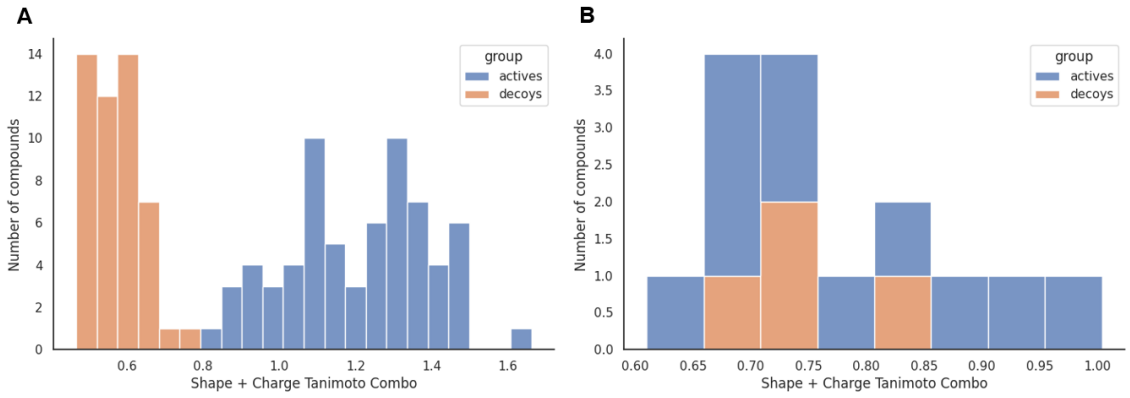


Figure 11. *Histogram distribution of actives and decoys by shape + charge Tanimoto similarity (Shape + Charge Tanimoto Combo). The validation library compounds from cluster 1 were distributed in 22 bins (A), whereas the validation library compounds from cluster 2 were distributed in 8 bins (B). Obtained with Matplotlib (Hunter, 2007).*

After screening the Specs library, some compounds were represented multiple times, as different poses or protonation states satisfied the applied filter. In those cases, only the conformer with the highest shape + charge Tanimoto combo was conserved. Following this refinement, a total of 29 compounds were retained for screening based on cluster 1, and 32 compounds for screening based on cluster 2.

6.7. Post-Screening Hit Compounds Selection and Analysis

As previously explained in Section 5.9, the diversity of the compounds retained from the virtual screening was assessed. For each cluster formed, only the compound with the highest shape + charge Tanimoto was left. This resulted in 8 hit compounds being selected for the screening based on cluster 1, and 22 for the other based on cluster 2. Afterwards, SwissADME (Daina et al., 2017) was used to review their oral bioavailability and general safety. Some of these properties are summarized in the BOILED-Egg diagrams (Figure 12).

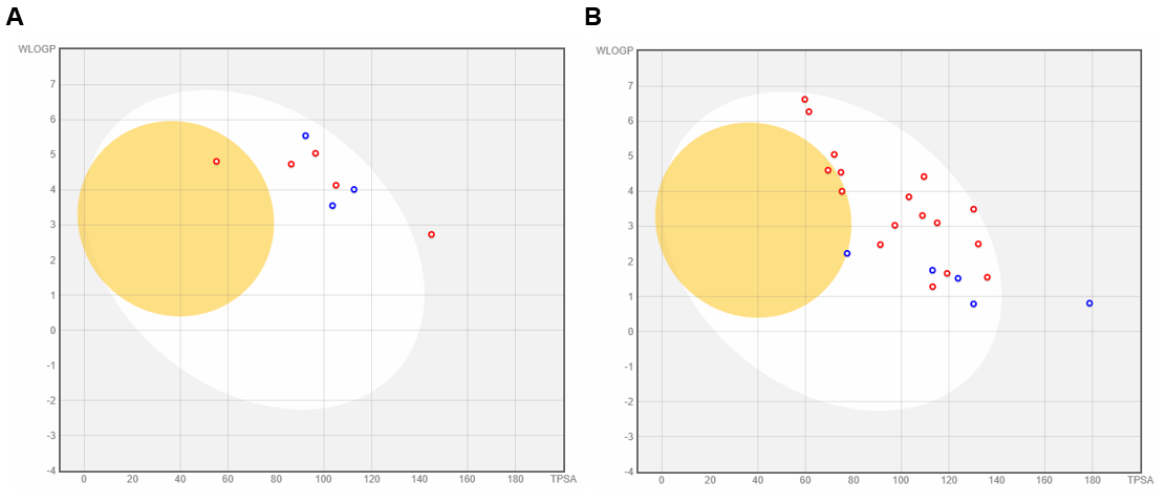


Figure 12. *Boiled-Egg diagrams of hit compounds obtained from screening with cluster 1 (A) and cluster 2 (B) references. Adapted from SwissADME (Daina et al., 2017).*

The BOILED-Egg diagram is a graphical model that predicts passive gastrointestinal absorption and blood-brain barrier permeation based on lipophilicity (WLOGP) and polarity (TPSA). In this representation, the white region represents compounds that are expected to be efficiently absorbed through the gastrointestinal tract, and the yellow region indicates compounds that are likely to pass the blood-brain barrier. Each dot (compound) is also color-coded according to its interactions with the P-glycoprotein (P-gp) transporter, which regulates absorption, distribution, and excretion of compounds such as drugs (Elbahnsi et al., 2024). Blue dots indicate compounds that are expected to be exported from the central nervous system by P-gp, whereas red dots indicate compounds that are not likely to be transported by P-gp.

The two compounds that were out of the white region were removed, since they could not act as efficient oral drugs. Also, because one of the main reasons for concern about ZIKV infection is its ability to infect the nervous system,

compounds predicted to pass the blood-brain barrier were kept. Notably, these molecules tested negative for transport by the P-gp, which ensures they would not be eliminated from the nervous system. The last SwissADME (Daina et al., 2017) property considered was the Brenk rule. All compounds that had chemical groups considered highly reactive by the Brenk rule were eliminated. At the end of the evaluation, 20 hit compounds remained.

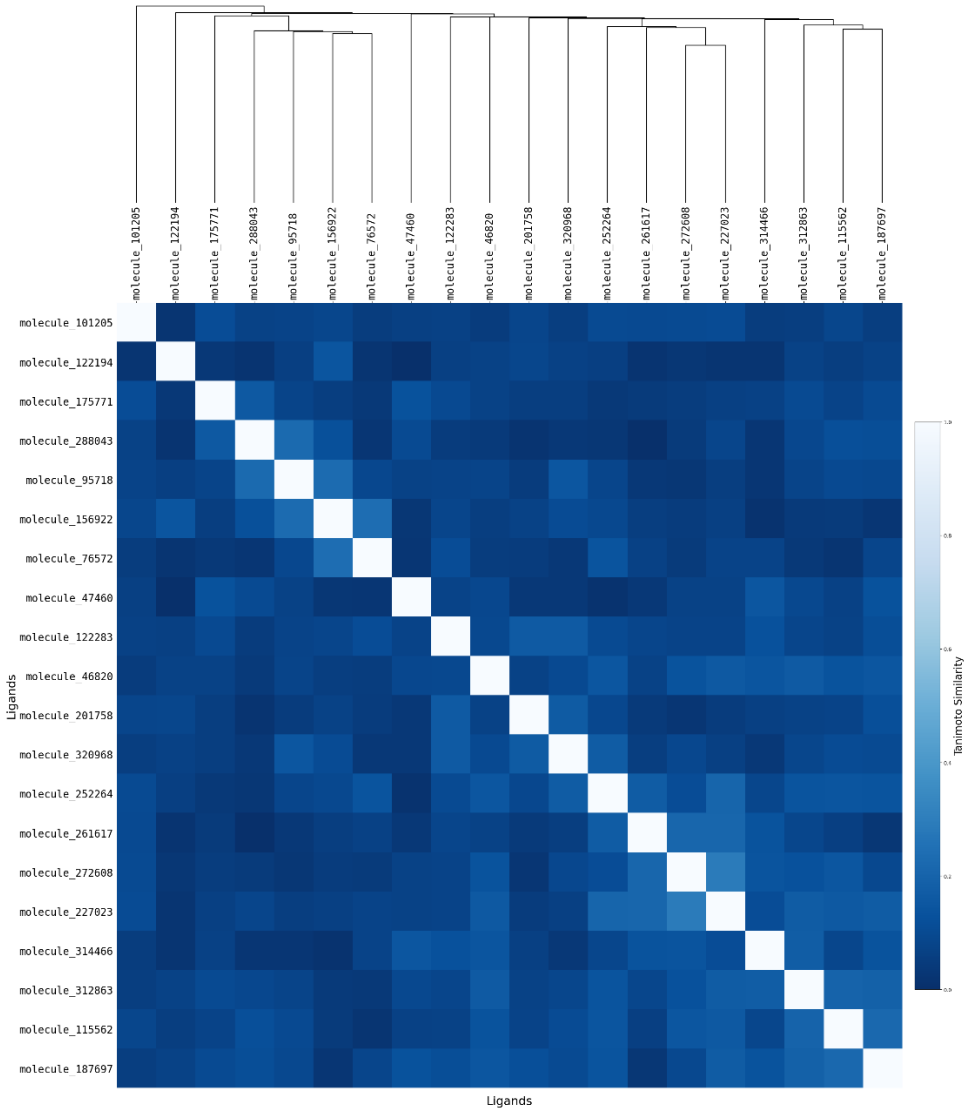


Figure 13. *Dendrogram representation of the structural similarity of the 20 final hit compounds.*

The 20 hit compounds were then unified to evaluate the similarity between them, and the results were represented in a dendrogram (Figure 13). As can be seen, neither of them formed a cluster, which confirms chemical diversity among the selected hits.

Lastly, the interactions formed between the NS2B/NS3 protease and the 20 hit compounds were studied (Figure 14). As expected, the hit compounds formed predominantly hydrophobic interactions, primarily involving TRP1069, LEU1076, TRP1083, and VAL1155. In the case of aromatic residues such as TRP1069 and TRP1083, aromatic-aromatic interactions were also formed, while VAL1155 participated in carbon-π interactions. Beyond these apolar contacts, a particularly relevant observation was that 14 hit compounds established polar interactions with ASN1152, the residue known as the molecular switch.

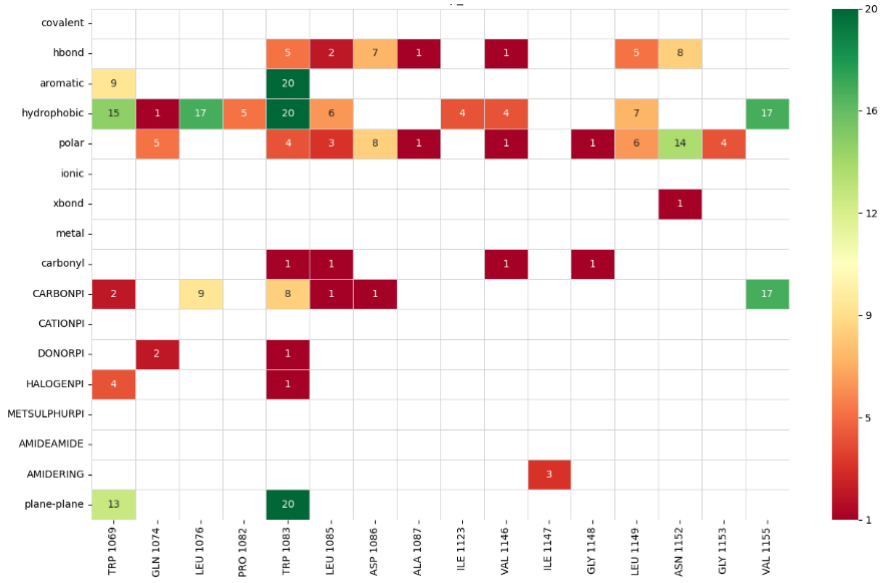


Figure 14. *Heatmap indicating the predicted interactions between the NS2B/NS3 allosteric pocket and the 20 hit compounds. The color pattern indicates the number of hit compounds that form each interaction. Obtained with CORAL-PIC.*

A table (Table S1) collecting the 20 hit compounds can be found in the Supplementary Information.

7. Discussion

Altogether, this virtual screening strategy identified 20 novel allosteric inhibitors against the ZIKV NS2B/NS3 protease that converge on a shared binding mechanism while presenting distinct scaffolds. The predicted inhibition mechanism should be similar to that of previously identified pyrazine-, proline-, and benzothiazole-based allosteric inhibitors, as their SAR informed the design of the virtual screening filters (Maus et al., 2021; Millies et al., 2019; Nie et al., 2021). Those studies showed that allosteric inhibitors from these series interacted with a hydrophobic region of the protein that becomes accessible when NS2B shifts away from NS3pro (Shiryaev et al., 2023). Moreover, proline and benzothiazole-based inhibitors formed interactions with ASN1152, which previous research has identified as a key residue in switching from closed to open conformations of the NS2B/NS3 (Brecher et al., 2017). These findings align with previous docking studies carried out with the proline and benzothiazole-based compounds, which also identified interactions within the hydrophobic region and ASN1152 residue (Millies et al., 2019; Nie et al., 2021).

The 20 hit compounds were confirmed to interact mainly with the same residues as the references, hence confirming the importance of incorporating SAR knowledge into the development of a virtual screening strategy. Based on this knowledge, allosteric inhibition of ZIKV appears to work through a mechanism that combines interaction with the specific hydrophobic region and with ASN1152, thereby preventing NS2B from associating with them and promoting the closed conformation.

The structure-based and ligand-based rational approaches developed were key to the effectiveness of virtual screening. The docking permitted overcoming the conformational flexibility of the validation library, although not used to predict affinity. The structure-based pharmacophores permitted the incorporation of the observed hydrophobic and polar features from the SAR interpretation, thus prioritizing compounds that could potentially act by the exact inhibition mechanism as references. Finally, the shape + charge similarity filtering helped minimize the high specificity of pharmacophoric features by preserving general shape and electrostatic field while allowing more scaffold diversity. Then, the assessment of ADME properties ensured that hit compounds had the desired

physicochemical and pharmacokinetic attributes to act as effective drugs against ZIKV. This is especially highlighted for the two hit compounds predicted to permeate through the blood-brain barrier, as they could act against ZIKV replication during neurological complications.

Nonetheless, the applied strategy had several limitations. On the one hand, the protein structure used as the receptor for structure-based steps is a mutant built to stabilize the open conformation. While structural analyses showed that the structure had no major conformational changes, subtle local differences may affect the screening. On the other hand, no experimental binding poses were available for the series used to interpret the SAR and build the validation library. For this reason, pose validation was indirect, based on agreement between docking and ligand-based pharmacophore models. However, the conformations could differ from the biological ones. Related to that, proline and benzothiazole-based compounds (cluster 2) presented high flexibility in the allosteric site, which further complicated validation.

8. Conclusions

Although cases have declined, Zika virus remains a concern due to the risk of resurgence or first-time emergence in regions with competent vectors, and the possibility of complications affecting pregnancy and the nervous system. For this reason, the current project has centered on developing a virtual screening targeting an allosteric site of the NS2B/NS3 protease to contribute to the research of specific antivirals for Zika virus.

Overall, this work demonstrates the potential of integrating in-depth SAR analysis into the development of a virtual screening strategy. By using SAR to define essential recognition features and translate them into docking, pharmacophore, and shape/electrostatic similarity criteria, the approach allows the identification of novel inhibitors that share inhibition mechanisms but preserve structural diversity. In this case, 20 structurally diverse compounds were identified with drug-like properties. Notably, two compounds are predicted to cross the blood-brain barrier, making them potential candidates for studying ZIKV inhibition in the nervous system.

Next steps should focus on converting these *in silico* candidates into lead compounds. To do this, their binding mode should be confirmed by in vitro assays and, if possible, by obtaining crystallized protein-ligand complexes. Moreover, enzyme and cell-based inhibition assays should be conducted to evaluate their inhibitory potency.

9. Bibliography

Attiq, N., Arshad, U., Brogi, S., Shafiq, N., Imtiaz, F., Parveen, S., Rashid, M., & Noor, N. (2022). Exploring the anti-SARS-CoV-2 main protease potential of FDA approved marine drugs using integrated machine learning templates as predictive tools. *International Journal of Biological Macromolecules*, *220*, 1415–1428. https://doi.org/10.1016/j.ijbiomac.2022.09.086

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. In *Nucleic Acids Research* (Vol. 28, Issue 1). http://www.rcsb.org/pdb/status.html

Brecher, M., Li, Z., Liu, B., Zhang, J., Koetzner, C. A., Alifarag, A., Jones, S. A., Lin, Q., Kramer, L. D., & Li, H. (2017). A conformational switch high-throughput screening assay and allosteric inhibition of the flavivirus NS2B-NS3 protease. *PLoS Pathogens*, *13*(5). https://doi.org/10.1371/journal.ppat.1006411

Cereto Massague, A., Garcia-Vallvé, S., & Pujadas, G. (2012). DecoyFinder, a tool for finding decoy molecules. *Journal of Cheminformatics*, *4*(S1), P2. https://doi.org/10.1186/1758-2946-4-S1-P2

Cheeseright, T., Mackey, M., Rose, S., & Vinter, A. (2006a). Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *Journal of Chemical Information and Modeling*, *46*(2), 665–676. https://doi.org/10.1021/ci050357s

Cheeseright, T., Mackey, M., Rose, S., & Vinter, A. (2006b). Molecular field extreme as descriptors of biological activity: Definition and validation. *Journal of Chemical Information and Modeling*, *46*(2), 665–676. https://doi.org/10.1021/ci050357s

Chen, X., Yang, K., Wu, C., Chen, C., Hu, C., Buzovetsky, O., Wang, Z., Ji, X., Xiong, Y., & Yang, H. (2016). Mechanisms of activation and inhibition of Zika virus NS2B-NS3 protease. In *Cell Research* (Vol. 26, Issue 11, pp. 1260–1263). Nature Publishing Group. https://doi.org/10.1038/cr.2016.116

Cresset. (2025). *Flare* (10.0.1). Cresset.

Dablander, M., Hanser, T., Lambiotte, R., & Morris, G. M. (2023). Exploring QSAR models for activity-cliff prediction. *Journal of Cheminformatics*, *15*(1). https://doi.org/10.1186/s13321-023-00708-w

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), 42717. https://doi.org/10.1038/srep42717

Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and

preliminary results. *Journal of Computer-Aided Molecular Design*, *20*(10–11), 647–671. https://doi.org/10.1007/s10822-006-9087-6

Elbahnsi, A., Dudas, B., Cisternino, S., Declèves, X., & Miteva, M. A. (2024). Mechanistic insights into P-glycoprotein ligand transport and inhibition revealed by enhanced molecular dynamics simulations. *Computational and Structural Biotechnology Journal*, *23*, 2548–2564. https://doi.org/10.1016/j.csbj.2024.06.010

Feng, Y. (2024). Recent advances in the study of zika virus structure, drug targets, and inhibitors. In *Frontiers in Pharmacology* (Vol. 15). Frontiers Media SA. https://doi.org/10.3389/fphar.2024.1418516

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*, *47*(7), 1739–1749. https://doi.org/10.1021/jm0306430

Gimeno, A., Ojeda-Montes, M. J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., Pujadas, G., & Garcia-Vallvé, S. (2019). The light and dark sides of virtual screening: What is there to know? In *International Journal of Molecular Sciences* (Vol. 20, Issue 6). MDPI AG. https://doi.org/10.3390/ijms20061375

Giordano, D., Biancaniello, C., Argenio, M. A., & Facchiano, A. (2022). Drug Design by Pharmacophore and Virtual Screening Approach. In *Pharmaceuticals* (Vol. 15, Issue 5). MDPI. https://doi.org/10.3390/ph15050646

Gratton, R., Agrelli, A., Tricarico, P. M., Brandão, L., & Crovella, S. (2019). Autophagy in zika virus infection: A possible therapeutic target to counteract viral replication. *International Journal of Molecular Sciences*, *20*(5). https://doi.org/10.3390/ijms20051048

Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A., & Stahl, M. T. (2010). Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, *50*(4), 572–584. https://doi.org/10.1021/ci100031x

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science \& Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling*, *60*(12), 6065–6073. https://doi.org/10.1021/acs.jcim.0c00675

Johnston, R. C., Yao, K., Kaplan, Z., Chelliah, M., Leswing, K., Seekins, S., Watts, S., Calkins, D., Chief Elk, J., Jerome, S. V., Repasky, M. P., & Shelley, J. C. (2023). Epik: pKa and Protonation State Prediction through Machine Learning. *Journal of Chemical Theory and Computation*, *19*(8), 2380–2388. https://doi.org/10.1021/acs.jctc.3c00044

Joosten, R. P., Long, F., Murshudov, G. N., & Perrakis, A. (2014). The PDB_REDO server for macromolecular structure model optimization. *IUCrJ*, *1*(4), 213–220. https://doi.org/10.1107/S2052252514009324

Jubb, H. C., Higueruelo, A. P., Ochoa-Montaño, B., Pitt, W. R., Ascher, D. B., & Blundell, T. L. (2017). Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *Journal of Molecular Biology*, *429*(3), 365–371. https://doi.org/10.1016/j.jmb.2016.12.004

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2025). PubChem 2025 update. *Nucleic Acids Research*, *53*(D1), D1516–D1525. https://doi.org/10.1093/nar/gkae1059

Kralj, S., Jukič, M., & Bren, U. (2023). Molecular Filters in Medicinal Chemistry. *Encyclopedia*, *3*(2), 501–511. https://doi.org/10.3390/encyclopedia3020035

Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, *27*(3), 221–234. https://doi.org/10.1007/s10822-013-9644-8

Maus, H., Barthels, F., Hammerschmidt, S. J., Kopp, K., Millies, B., Gellert, A., Ruggieri, A., & Schirmeister, T. (2021). SAR of novel benzothiazoles targeting an allosteric pocket of DENV and ZIKV NS2B/NS3 proteases. *Bioorganic and Medicinal Chemistry*, *47*. https://doi.org/10.1016/j.bmc.2021.116392

Maus, H., Hammerschmidt, S. J., Hinze, G., Barthels, F., Pérez Carrillo, V. H., Hellmich, U. A., Basché, T., & Schirmeister, T. (2023). The effects of allosteric and competitive inhibitors on ZIKV protease conformational dynamics explored through smFRET, nanoDSF, DSF, and 19F NMR. *European Journal of Medicinal Chemistry*, *258*. https://doi.org/10.1016/j.ejmech.2023.115573

Meewan, I., Shiryaev, S. A., Kattoula, J., Huang, C. T., Lin, V., Chuang, C. H., Terskikh, A. V., & Abagyan, R. (2023). Allosteric Inhibitors of Zika Virus NS2B-NS3 Protease Targeting Protease in “Super-Open” Conformation. *Viruses*, *15*(5). https://doi.org/10.3390/v15051106

Metzler, A. D., & Tang, H. (2024). Zika Virus Neuropathogenesis—Research and Understanding. In *Pathogens* (Vol. 13, Issue 7). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/pathogens13070555

Millies, B., Von Hammerstein, F., Gellert, A., Hammerschmidt, S., Barthels, F., Göppel, U., Immerheiser, M., Elgner, F., Jung, N., Basic, M., Kersten, C.,

Kiefer, W., Bodem, J., Hildt, E., Windbergs, M., Hellmich, U. A., & Schirmeister, T. (2019). Proline-Based Allosteric Inhibitors of Zika and Dengue Virus NS2B/NS3 Proteases. *Journal of Medicinal Chemistry*, *62*(24), 11359–11382. https://doi.org/10.1021/acs.jmedchem.9b01697

Mwaliko, C., Nyaruaba, R., Zhao, L., Atoni, E., Karungu, S., Mwau, M., Lavillette, D., Xia, H., & Yuan, Z. (2021). Zika virus pathogenesis and current therapeutic advances. In *Pathogens and Global Health* (Vol. 115, Issue 1, pp. 21–39). Taylor and Francis Ltd. https://doi.org/10.1080/20477724.2020.1845005

Nie, S., Yao, Y., Wu, F., Wu, X., Zhao, J., Hua, Y., Wu, J., Huo, T., Lin, Y. L., Kneubehl, A. R., Vogt, M. B., Ferreon, J., Rico-Hesse, R., & Song, Y. (2021). Synthesis, Structure-Activity Relationships, and Antiviral Activity of Allosteric Inhibitors of Flavivirus NS2B-NS3 Protease. *Journal of Medicinal Chemistry*, *64*(5), 2777–2800. https://doi.org/10.1021/acs.jmedchem.0c02070

OpenEye, C. M. S. (2025). *EON 3.1.0*.

Paggi, J. M., Pandit, A., & Dror, R. O. (2025). *The Art and Science of Molecular Docking*. https://doi.org/10.1146/annurev-biochem-030222

Rabe, I. B., Hills, S. L., Haussig, J. M., Walker, A. T., dos Santos, T., San Martin, J. L., Gutierrez, G., Mendez-Rico, J., Rodriguez, J. C., Elizondo-Lopez, D., Gonzalez-Escobar, G., Chanda, E., Al Eryani, S. M., Kodama, C., Yajima, A., Kakkar, M., Kato, M., Wijesinghe, P. R., Samaraweera, S., … Rojas, D. P. (2025). A Review of the Recent Epidemiology of Zika Virus Infection. In *American Journal of Tropical Medicine and Hygiene* (Vol. 112, Issue 5, pp. 1026–1035). American Society of Tropical Medicine and Hygiene. https://doi.org/10.4269/ajtmh.24-0420

Schrödinger, L. (2025a). *Canvas* (Schrödinger Release 2025-02).

Schrödinger, L. (2025b). *LigPrep* (Schrödinger Release 2025-02).

Schrödinger, L. (2025c). *Maestro* (Schrödinger Release 2025-02).

Schrödinger, L. (2025d). *QikProp* (Schrödinger Release 2025-02).

Schrödinger, L. (2025e). *The Pymol Molecular Graphics System* (Version 3.0).

Seidel, T., Wieder, O., Garon, A., & Langer, T. (2020). Applications of the Pharmacophore Concept in Natural Product inspired Drug Design. In *Molecular Informatics* (Vol. 39, Issue 11). Wiley-VCH Verlag. https://doi.org/10.1002/minf.202000059

Shiryaev, S. A., Cieplak, P., Cheltsov, A., Liddington, R. C., & Terskikh, A. V. (2023). Dual function of Zika virus NS2B-NS3 protease. *PLoS Pathogens*, *19*(11). https://doi.org/10.1371/journal.ppat.1011795

Shiryaev, S. A., Farhy, C., Pinto, A., Huang, C. T., Simonetti, N., Ngono, A. E., Dewing, A., Shresta, S., Pinkerton, A. B., Cieplak, P., Strongin, A. Y., & Terskikh, A. V. (2017). Characterization of the Zika virus two-component

NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. *Antiviral Research*, *143*, 218–229. https://doi.org/10.1016/j.antiviral.2017.04.015

Sirohi, D., & Kuhn, R. J. (2017). Zika Virus Structure, Maturation, and Receptors. In *Journal of Infectious Diseases* (Vol. 216, pp. S935–S944). Oxford University Press. https://doi.org/10.1093/infdis/jix515

Tajik, S., Farahani, A. V., Ardekani, O. S., Seyedi, S., Tayebi, Z., Kami, M., Aghaei, F., Hosseini, T. M., Nia, M. M. K., Soheili, R., & Letafati, A. (2024). Zika virus tropism and pathogenesis: understanding clinical impacts and transmission dynamics. In *Virology Journal* (Vol. 21, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12985-024-02547-z

Vázquez, J., López, M., Gibert, E., Herrero, E., & Javier Luque, F. (2020). Merging ligand-based and structure-based methods in drug discovery: an overview of combined virtual screening approaches. In *Molecules* (Vol. 25, Issue 20). MDPI AG. https://doi.org/10.3390/molecules25204723

Vittorio, S., Lunghini, F., Morerio, P., Gadioli, D., Orlandini, S., Silva, P., Jan Martinovic, Pedretti, A., Bonanni, D., Del Bue, A., Palermo, G., Vistoli, G., & Beccari, A. R. (2024). Addressing docking pose selection with structure-based deep learning: Recent advances, challenges and opportunities. In *Computational and Structural Biotechnology Journal* (Vol. 23, pp. 2141–2151). Elsevier B.V. https://doi.org/10.1016/j.csbj.2024.05.024

Wu, M.-R., Nissim, L., Stupp, D., Pery, E., Binder-Nissim, A., Weisinger, K., Enghuus, C., Palacios, S. R., Humphrey, M., Zhang, Z., Maria Novoa, E., Kellis, M., Weiss, R., Rabkin, S. D., Tabach, Y., & Lu, T. K. (2019). A high-throughput screening and computation platform for identifying synthetic promoters with enhanced cell-state specificity (SPECS). *Nature Communications*, *10*(1), 2880. https://doi.org/10.1038/s41467-019-10912-8

Yao, Y., Huo, T., Lin, Y. L., Nie, S., Wu, F., Hua, Y., Wu, J., Kneubehl, A. R., Vogt, M. B., Rico-Hesse, R., & Song, Y. (2019). Discovery, X-ray Crystallography and Antiviral Activity of Allosteric Inhibitors of Flavivirus NS2B-NS3 Protease. *Journal of the American Chemical Society*, *141*(17), 6832–6836. https://doi.org/10.1021/jacs.9b02505

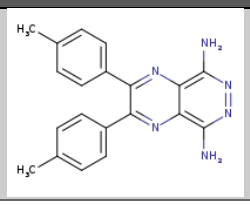
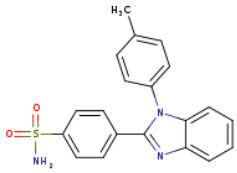
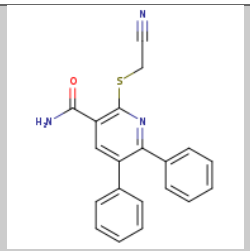
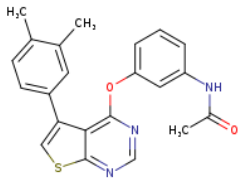
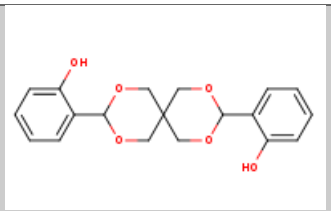
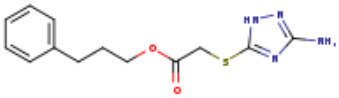
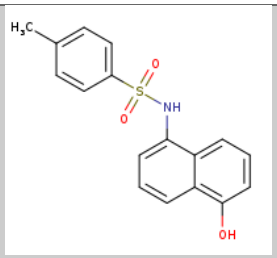
10. Self-assessment

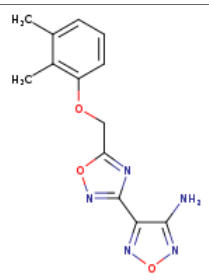
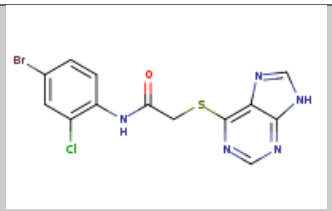
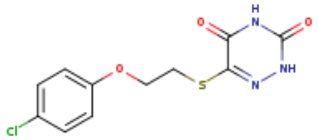
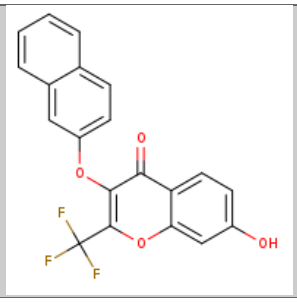
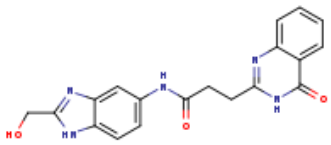
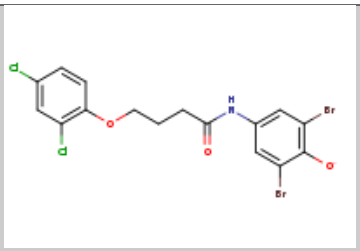
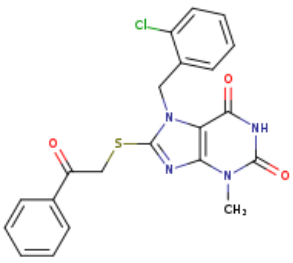
I began this project with a genuine interest in learning more about bioinformatics and cheminformatics, but also with worry and insecurity about whether I could manage to develop a real research project. Despite the knowledge acquired in *Bioinformàtica, Enginyeria de Proteïnes,* and *Estructura i Funció de Biomolècules*, it was my first contact with computational techniques for drug discovery, such as virtual screening. For that reason, I am sincerely grateful for the guidance, patience, and trust of all my supervisors, and for my internship teammates, with whom I created an environment of mutual support, exchange of ideas, and understanding of each other's working rhythm.

The project was quite a challenge because of its complexity and methodological limitations. It required several changes in strategy and even redoing parts of the work. Although challenging and nerve-wracking, I appreciate the way it allowed me to develop my adaptability, critical thinking, and problem-solving skills. Moreover, I gained familiarity with the typical steps of a virtual screening workflow and learned to work with software commonly employed in this field.

Altogether, it was an experience that helped me become a more complete professional and offered an initial sense of the challenges I can expect from here on. I now finish with more confidence, and I am better prepared for the next steps in my career path.

11. Supplementary Information

Compound ID	Structure	Molecular Weight (g/mol)	WLogP	TPSA (Å^2)
molecule_288043		342.40	3.55	103.60
molecule_95718		363.43	4.73	86.36
molecule_115562		345.42	4.13	105.07
molecule_101205		389.47	5.54	92.35
molecule_122194		344.36	2.23	77.38
molecule_312863		292.36	1.66	119.19
molecule_156922		313.37	4.54	74.78

molecule_175771		287.27	1.75	113.09
molecule_272608		398.67	3.31	108.86
molecule_314466		299.73	1.28	113.14
molecule_201758		372.29	6.62	59.67
molecule_252264		363.37	1.52	123.76
molecule_261617		496.99	6.27	61.39
molecule_46820		440.90	3.10	115.05

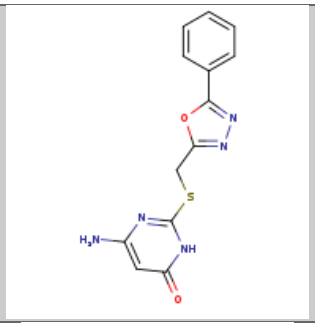
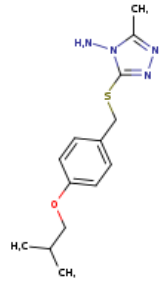
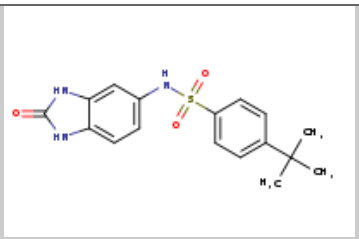
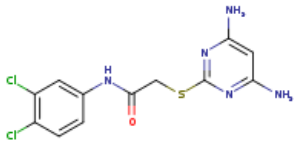
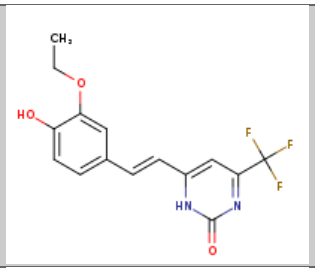
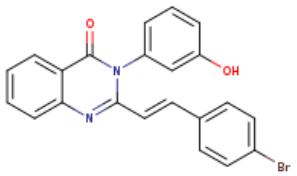
molecule_187697		301.32	1.55	135.99
molecule_47460		292.40	2.48	91.26
molecule_76572		345.42	3.84	103.20
molecule_227023		344.22	2.50	132.22
molecule_122283		326.27	4.00	75.21
molecule_320968		419.27	4.81	55.12

Table S1. *Summary of the compound ID, structure, and relevant physicochemical properties of the 20 hit compounds.*