UNIVERSITAT ROVIRA i VIRGILI

Virtual screening for the identification of potential antifungal agents targeting *Candida albicans'* lanosterol 14α-demethylase

Treball de Fi de Grau

Autora: Paula Paris Bayarri

Tutor acadèmic: Dr. Aleix Gimeno Vives, Departament de Bioquímica i Biotecnologia, URV (aleix.gimeno@urv.cat)

Ensenyament: Bioquímica i Biologia Molecular

Tarragona, 4 de juny de 2025

The present work has been developed based on the results obtained during the External Internship at the Cheminformatics and Nutrition research group in the Biochemistry and Biotechnology Department of University Rovira I Virgili (URV), under the supervision of Dr. Aleix Gimeno, Dr. Gerard Pujadas and Dr. Santi Garcia-Vallvé.

Table of contents

1. Abstract

Candidiasis, mainly caused by *Candida albicans*, is a serious infection due to its incidence ratios and mortality. Therefore, there is a growing global need to develop novel antifungal agents. The main therapeutic target for *Candida albicans* is lanosterol 14α-demethylase, an enzyme involved in ergosterol biosynthesis. However, resistance to azole-based drugs has been reported. Here, a virtual screening workflow combining ligand- and receptor-based computational approaches has been designed aiming to identify structurally diverse inhibitors of lanosterol 14α-demethylase. As a result, 20 compounds have been identified as potentially novel inhibitors for this target. These compounds should inhibit fungal growth bypassing resistance mechanisms.

2. Introduction

Candidiasis

In the past few years, the human microbiota has become a major focus of scientific research due to its important role in maintaining health and contributing to diseases (El-Sayed et al., 2021). The term microbiota refers to the collection of microorganisms, not only bacteria but also fungi, archaea and viruses, that inhabit in different parts of the human body, such as the gut, skin, oral cavity and mucosal surfaces (such as the vaginal mucosa) (El-Sayed et al., 2021; Lopes & Lionakis, 2022; *Microbiota - MeSH - NCBI*).

Under normal conditions, these microorganisms coexist in a state of dynamic equilibrium known as eubiosis or microbial balance (Iebba et al., 2016). The microbiota helps with key physiological processes, including digestion, regulation of the immune system, and protection against pathogens. However, different factors, such as changes in nutritional habits, environment, lifestyle or the use of antibiotics, among others, can disrupt this balance, leading to a condition known as dysbiosis (Iebba et al., 2016). This condition has been associated with a wide range of disorders such as infections, cancer, autoimmune diseases and metabolic diseases (El-Sayed et al., 2021).

Among the diverse microbial communities inhabiting in the human body, species of the genus *Candida* stand out due to their clinical relevance as opportunistic fungus pathogens. While they typically reside as a harmless commensal under eubiotic conditions, perturbations in the environment can trigger their transition to a pathogenic state(Lopes & Lionakis, 2022). Moreover, *Candida* species are responsible for most of the human fungal infections, with *Candida albicans* being the most frequent cause of opportunistic infections (Lopes & Lionakis, 2022).

In 2022, the World Health Organization (WHO) published the first Fungal Priority Pathogens List (FPPL), aiming to raise global awareness of the growing public health threat by fungal infections and the urgent need for research into developing novel antifungals for these pathogens that are becoming resistant. In this list, fungal pathogens are categorized into critical, high and medium priority. *Candida albicans* is listed in the critical priority group, highlighting its clinical relevance and

the need for developing effective and non-resistant therapies (*WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action*, n.d.). Moreover, *Candida albicans* remains the leading cause of candidiasis worldwide (Katsipoulaki et al., 2024).

The infections caused by the opportunistic fungus *Candida albicans* are termed candidiasis. These infections are classified into two main categories: mucocutaneous and invasive (systemic) candidiasis. On the one hand, mucocutaneous candidiasis affects mucosal surfaces and skin areas, which are colonized by commensal microbiota, primarily manifesting in the genital tract (vulvovaginal candidiasis), oral cavity (oropharyngeal candidiasis) or nails (onychomycosis), among others. On the other hand, invasive candidiasis occurs when *Candida albicans* penetrates deeper tissues, leading to bloodstream infections (candidemia) or disseminated organ involvement (Cornely et al., 2025; Katsipoulaki et al., 2024; Lopes & Lionakis, 2022) (see Figure 1).

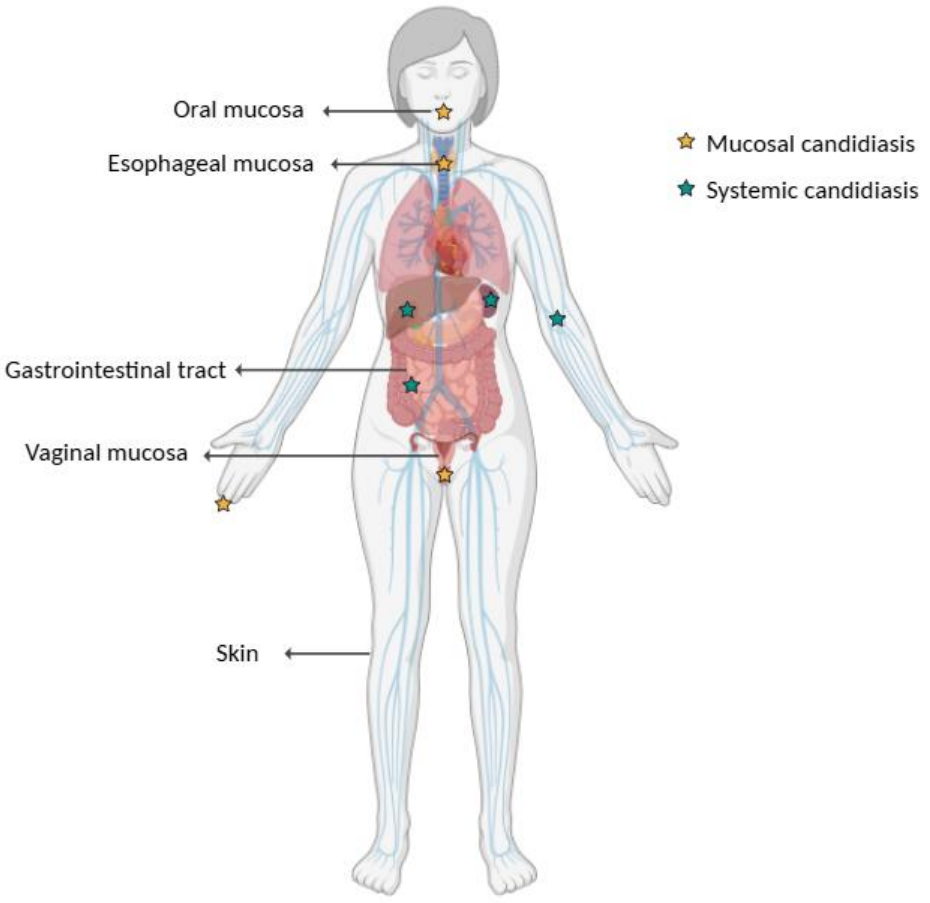


Figure 1. Common infections parts of *Candida albicans* in the human body. Yellow stars indicate regions associated with mucosal candidiasis, while green stars indicate affectations caused by invasive candidiasis. Illustration created with BioRender.

Although mucocutaneous candidiasis is typically non-life-threatening, its high incidence represents a significant clinical and public health concern, especially among immunocompromised patients such as those in the intensive care unit (ICUs), individuals with a positive result for the Human Immunodeficiency Virus (HIV) or post-COVID-19 cases. An example is vulvovaginal candidiasis (VCC), which affects approximately 75% of women of reproductive age at least once in their lifetime, with around 8% suffering recurrent episodes. Epidemiologically, this corresponds to an estimated annual prevalence of 3,871 cases per 100,000 women (David & Solomon, 2023; D'Enfert et al., 2021).

In contrast, when *Candida* species invade the bloodstream (candidemia) and disseminate to internal organs such as the liver, the kidneys or the spleen, the infection becomes life-threatening. The global incidence rate of invasive candidiasis ranges from 2 to 14 cases per 100,000 inhabitants annually, with a mortality rate reaching 40-55%, even when treated with current antifungal therapies (D'Enfert et al., 2021; Soriano et al., 2023).

The high global cases of both systemic and mucocutaneous candidiasis, along with their high rates of mortality, raises concerns about the effectiveness of current antifungal treatments.

Mechanisms of infections of *Candida* species

Candida species employ diverse pathogenic mechanisms to infect the host. Initially, yeast cells adhere to the host surfaces via the expression of adhesins, facilitating the morphological transition to invasive hyphae, which is an important virulence factor. Invasion of the different tissues can happen through two pathways: **(a)** induced endocytosis, mediated by the expression in the fungal surfaces of invasins that mimic host ligands; or **(b)** active penetration, driven by mechanical force from viable hyphae. Moreover, Candida species can form biofilms on abiotic (for example, catheters) and biotic (host cells) surfaces, enhancing resistance to immune defenses and antifungal agents. Hyphae can also secrete hydrolases, which degrade host membranes, and therefore, it promotes tissue invasion. In addition to physical invasion of host cells, fitness traits contribute to the mechanism of infection. Candida can tolerate a wide range of pH values in the human body and even regulate their surrounding pH to

promote optimal growth. During infection, they can adapt their metabolism to different host niches by shifting between glycolysis and glycogenolysis, depending on the nutrient's availability. Another important adaptative mechanism is the rapid response to environmental stress, such as high temperatures, starvation and oxidative stress. This is mediated by the expression of heat shock proteins and antioxidant defense systems. Furthermore, fungi need essential metals such as iron to support growth and pathogenesis, often competing with the host for their acquisition (Mayer et al., 2013).

Described inhibitors for *Candida* species

Antifungal agents used against *Candida* species are classified into four main categories: azoles, polyenes, echinocandins and pyrimidine analogues. However, pyrimidine analogues are not included in first-line clinical treatments (Cornely et al., 2025; David & Solomon, 2023; Hargrove et al., 2017; Katsipoulaki et al., 2024; Lee et al., 2020; McCarty et al., 2021; Pappas et al., 2015; Soriano et al., 2023)

- **Azoles** are the most widely used antifungals, with fluconazole being the most common one. These agents target primarily the biosynthesis of ergosterol, an essential sterol that maintains the fluidity, integrity and permeability of the fungal membrane (David & Solomon, 2023; Hargrove et al., 2017; Katsipoulaki et al., 2024).
- **Echinocandins** are also frequently used. They inhibit (1→3)-β-D-glucan synthase, a key enzyme in fungal cell wall biosynthesis. Their action results in weakened fungal cell wall and ultimately fungal death (David & Solomon, 2023; Hargrove et al., 2017; Katsipoulaki et al., 2024).
- Among **polyenes**, amphotericin B is the most common one. It forms aggregates from ergosterol in the fungal membrane, creating pores that lead to membrane destabilization and cell lysis (Lee et al., 2020). As a result, the membrane lacks ergosterol and therefore fungal growth is impaired (Hargrove et al., 2017).
- **Pyrimidine analogues** are the least used as a treatment for candidiasis due to their toxicity for humans. These compounds mimic the structure of

pyrimidine and interfere in the synthesis of DNA, RNA and protein (David & Solomon, 2023).

These opportunistic fungi possess remarkable adaptability to their environment. Therefore, antifungal resistance has emerged as a major clinical challenge. Resistance to azoles is mainly associated with mutations in genes encoding enzymes involved in the ergosterol biosynthesis, while echinocandins resistance is linked to mutations in FKS1 and FKS2, which encode subunits of the (1→3)-β-D-glucan synthase complex (Lee et al., 2020).

Lanosterol 14α-demethylase

One of the most dangerous opportunistic pathogens that can cause an acute infection is *Candida albicans*, causing the most hospitalizations worldwide (Binjubair et al., 2020). For this reason, this fungus was chosen as the target of this study for the severity of the infections and their frequency.

Fluconazole, an azole antifungal, is one of the first-line agents used to treat candidiasis by targeting lanosterol 14α-demethylase. However, point mutations in the gene ERG11 can lead to fluconazole resistance, limiting its efficacy. (Binjubair et al., 2020; Hargrove et al., 2017). For this reason, recent studies have aimed to develop new inhibitory drugs that are more efficient and potent, hence improving its affinity to the target. The main goal of the design of these new inhibitors is to improve the treatment outcomes and reduce the mortality rate (Monk et al., 2020)

Lanosterol 14α-demethylase is involved in the biosynthesis of ergosterol, a key compound in the fungal membrane to maintain the permeability, fluidity and integrity (Rodrigues, 2018). Its *de novo* biosynthesis involves a series of 30 enzymes that work sequentially, most of them being codified by the ERG gene family (Liu et al., 2019). One of them is lanosterol 14α-demethylase (or Erg11) that is responsible for the removal of the group 14α-methyl from the first cyclized sterol precursor (Hargrove et al., 2017). By inhibiting this enzyme, ergosterol is no longer synthesized, which will affect the fungal cell membrane integrity (Binjubair et al., 2020). Moreover, due to this inhibition, some changes are being done in the cell, such as mitochondrial dysfunction, that will develop into the

generation of toxic reactive oxygen species (ROS). Because of the ROS accumulation in the cell, programmed death will take place, and apoptotic genes will be activated (Schuster et al., 2024).

The target protein lanosterol 14α-demethylase is located in the endoplasmic reticulum (ER) (Hargrove et al., 2017). There is a variable amphipathic transmembrane domain that will anchor to the lipidic bilayer of the ER. However, the catalytic domain is facing the cytosol (see Figure 2) (Monk et al., 2020). Once ergosterol is synthesized, it is transported to the cell membrane through lipid binding/transfer proteins because of its hydrophobicity (Liu et al., 2019).

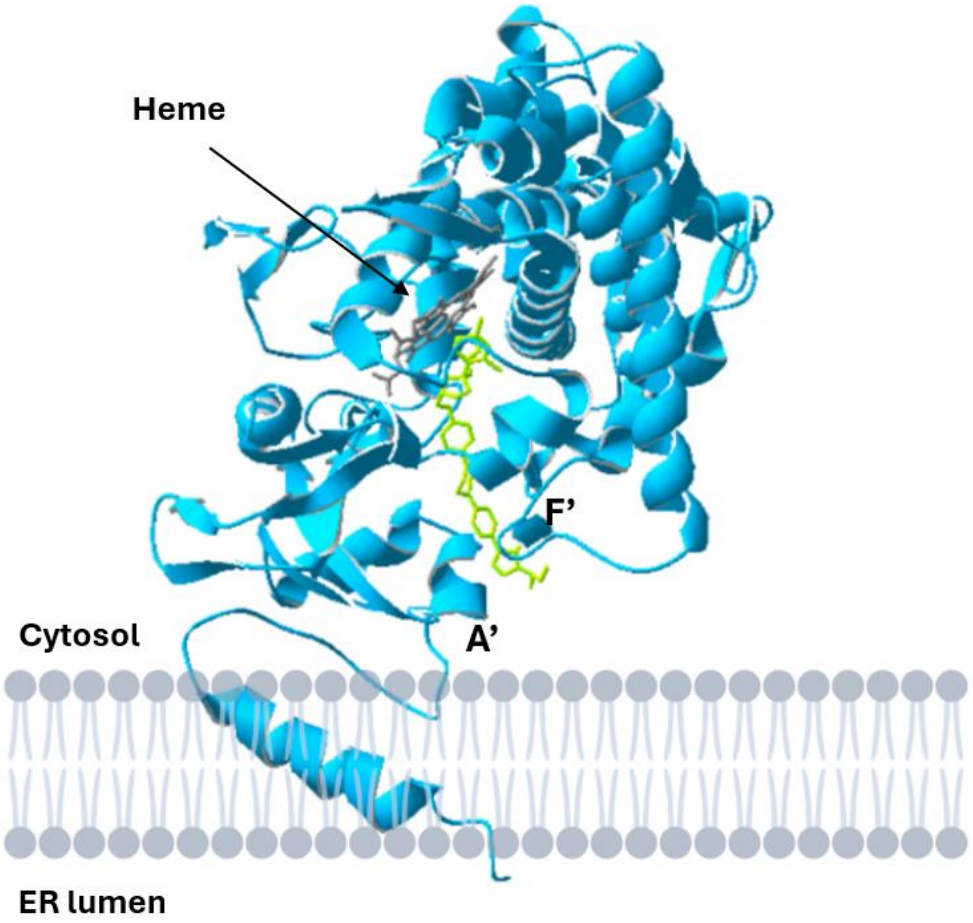


Figure 2. Structural representation of lanosterol 14α-demethylase anchored to the endoplasmic reticulum. Catalytic domain is facing the cytosol, where the synthesis of ergosterol takes place. Moreover, helices A' and F' form the tunnel for the ligand (lanosterol) and other inhibitors facilitating their accommodation to the active site and performing its biological activity. Lanosterol 14α-demethylase, extracted from PDB ID 5V5Z, is represented in blue ribbons. In its binding site, the heme prosthetic group is found in grey sticks format. The crystallized inhibitor is highlighted in green sticks format.

The target enzyme belongs to the cytochrome P450 family, which is a family of hemoproteins. This means that these proteins have a heme prosthetic group in their active site that plays a crucial role in its enzymatic activity (Zhao et al., 2021). One mechanism that azoles have in order to inhibit the activity of this enzyme is by forming an axial coordination bond with the prosthetic heme group that will affect the reduction potential of the iron (Hargrove et al., 2017).

This protein, of over 450 amino acids in sequence length, has a similar structure to all eukaryotes. It is characterized by one opening access tunnel that allows the ligand or inhibitors to reach the active site. This tunnel corresponds to tunnel 2f in Wade's nomenclature, which is formed by helices A' and F', as shown in Figure 2. This tunnel can accommodate the longest part of the lanosterol (substrate) as well as diverse inhibitors (Yu et al., 2015).

Virtual screening

In order to identify potential inhibitors of lanosterol 14α-demethylase, an *in silico* strategy called *Virtual Screening* was implemented. This procedure has become increasingly relevant in recent years due to technological advances in computational drug discovery. Virtual screening consists of a collection of computational techniques designed to identify from a large library of chemical compounds those that with a high likelihood of binding effectively to a specific biological target (Gimeno et al., 2019). These methods can be classified into two main categories: (a) ligand-based methods, which are based on the structural and physicochemical similarities between the small molecules from a library of interest; and (b) receptor-based methods, which focuses on the interactions between the small molecules and the binding site of the target protein. In this study, both approaches were applied in a complementary manner to improve the reliability of the results.

Three crystallized structures of the target protein were identified for this study (PDB IDs: 5TZ1, 5FSA and 5V5Z). All of them exhibit a high degree of structural similarity, as illustrated in Figure 3. Despite this, each structure contains a different co-crystallized ligand. These vary in size, since the ligand of 5TZ1 (VT1 or oteseconazole) is the smallest one, while the largest is the one from 5FSA (X2N or posaconazole), as shown in Figure 4. Among the three, only the ligands

from 5TZ1 and 5FSA have experimental activity values against *Candida albicans*. VT1 is the most active, which has a MIC_{50} value of 0.001 $\mu g \cdot mL^{-1}$ compared to X2N which has a MIC_{50} value of 0.015 $\mu g \cdot mL^{-1}$ (Yates et al., 2017). The MIC_{50} (Minimum Inhibitory Concentration 50) indicates the concentration required to inhibit the 50% of the growth of a microorganism. To facilitate comparisons, these values are often converted into $pMIC_{50}$ values, calculated as:

$$pMIC_{50} = -\log_{10}(MIC_{50})$$

A high $pMIC_{50}$ value implies stronger inhibitory potency. This conversion highlights the superior efficacy of VT1 compared to X2N, since VT1 requires much lower concentration to achieve the same biological effect. Therefore, given its high potency and compact size, the structure 5TZ1 was selected as a reference for the virtual screening process.

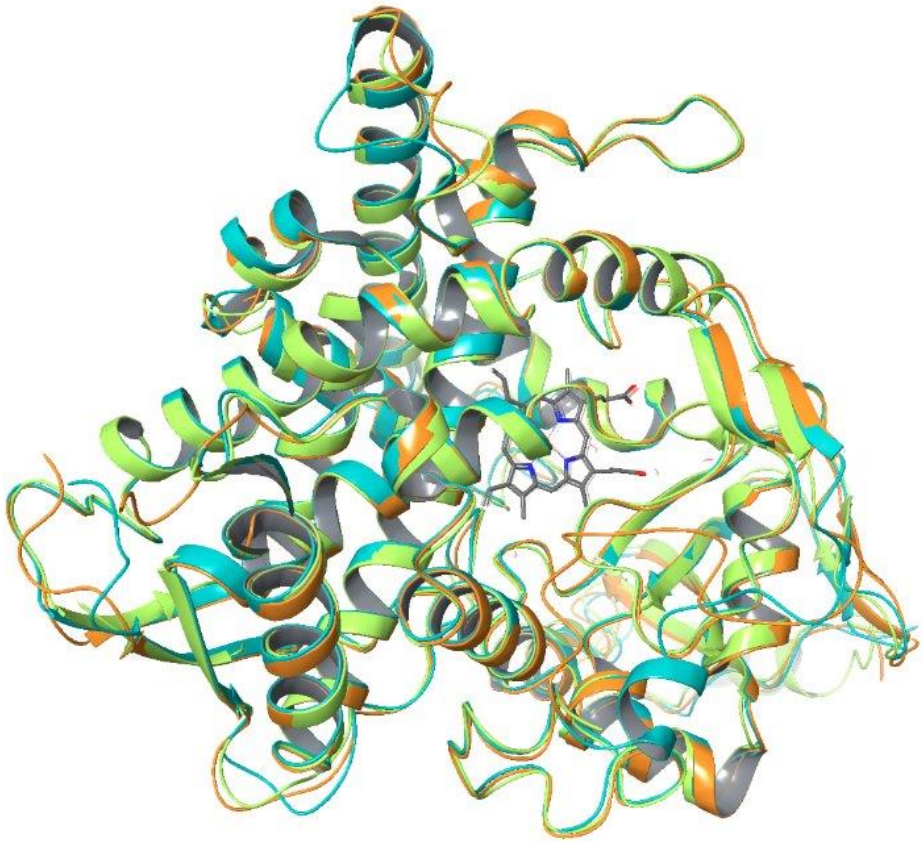


Figure 3. Structural superposition of the three crystallized structures from lanosterol 14α-demethylase in *Candida albicans*. The protein chains, represented in ribbon format, are colored as follows: 5TZ1 in blue, 5V5Z in green and 5FSA in orange. The heme prosthetic group from the three crystallized structures is represented in sticks format and colored by CPK.

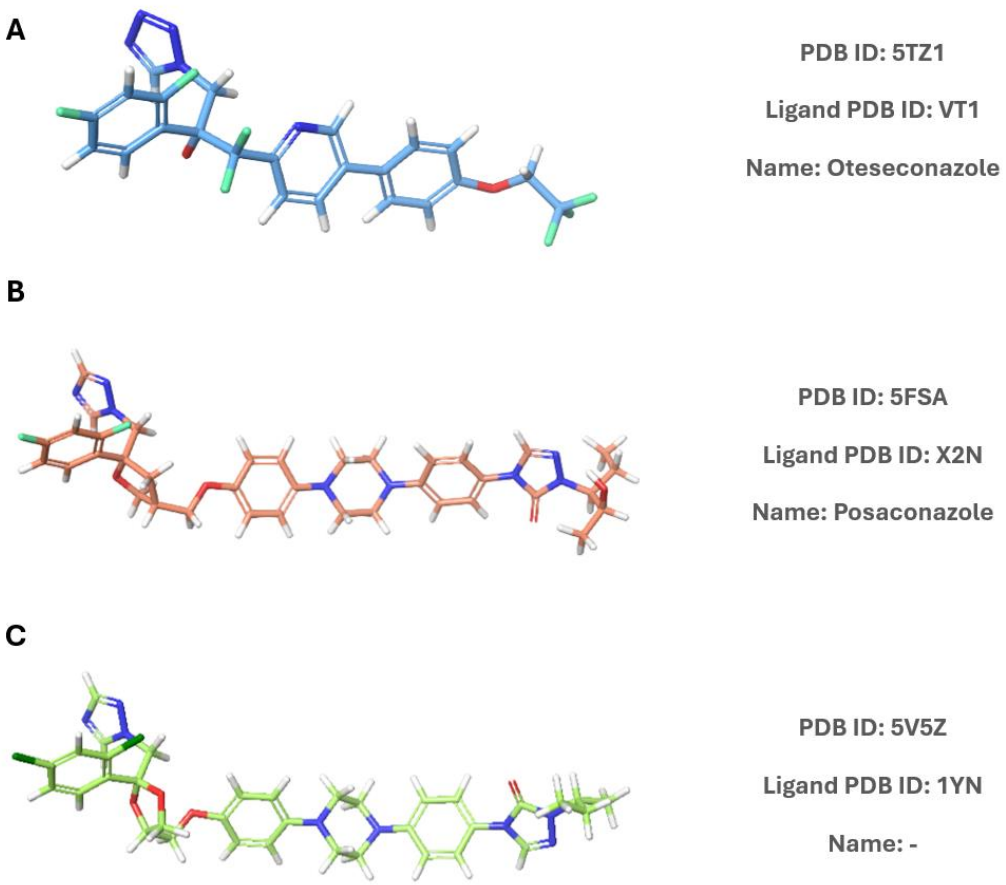


Figure 4. 3D structures of the co-crystalized ligands represented in sticks format and colored using CPK scheme. The backbone of each ligand is distinctly colored: VT1 (oteseconazole, from 5TZ1) in blue (A), X2N (posaconazole, from 5FSA) in light orange (B), and 1YN (from 5V5Z) in light green (C)

To carry out the virtual screening, a customized workflow was developed. The entire process is illustrated in Figure 5, where each step is highlighted in a different color to distinguish the different stages of this study.

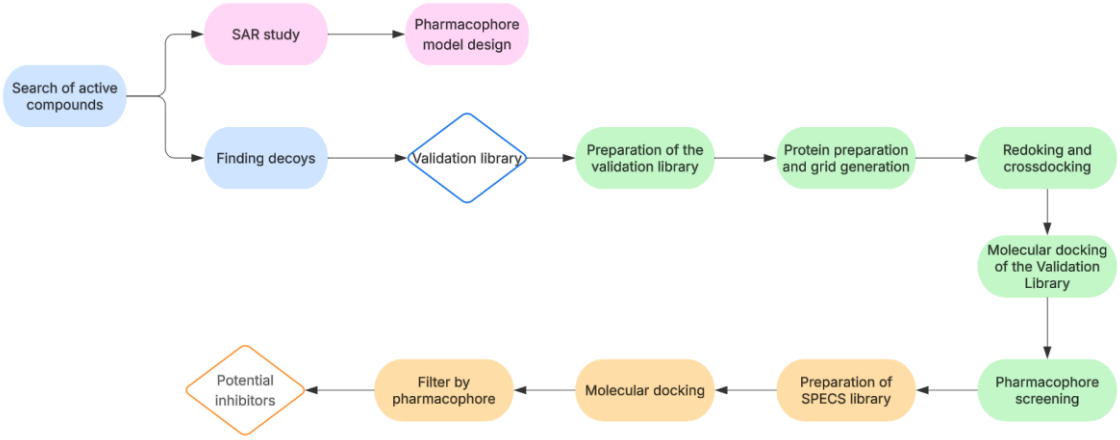


Figure 5. Schematic representation of the complete workflow implemented in this study. Each stage of the ligand-based virtual screening methodology is visually distinguished by using different colors. The process started with the construction of the validation library (blue steps). From the active compounds, a SAR (structure-activity relationship) study was conducted to develop a pharmacophore model that captures the key features contributing to inhibition of the target protein (steps in pink). Next, a receptor-based molecular docking was performed (steps in green) to predict how the compounds from the validation library bind and interact with the target's active site. The resulting poses were screened through the pharmacophore model to assess its enrichment capability. Finally, the same receptor-based molecular docking was also performed in the library of chemical compounds of SPECS (steps in orange) and the resulting hits were screened through the pharmacophore model. This allowed the identification of a final set of potential inhibitors.

Firstly, known active compounds to the target lanosterol 14α-demethylase were identified using a Python script developed in order to automate the retrieval process. For each active compound identified, 50 decoys were generated. These decoys are molecules with similar physicochemical properties to actives but with no reported against the target protein. Together, the actives and decoys formed the validation library. This library was then prepared in order to generate the appropriate ionization and tautomeric states as well as the stereoisomers. Therefore, suitable 3D conformers were generated using OMEGA (Hawkins et al., 2010; OpenEye) and LigPrep (Schrödinger).

On the one hand, OMEGA (Hawkins et al., 2010; OpenEye) is an application developed by OpenEye and is responsible for generating different conformations from 2D structures. This application identifies the rotatable bonds from each molecule and explores the different torsion angles for the generations of possible conformations. Moreover, it uses chemical patterns and structural symmetry in order to reduce the number of redundant structures. The different generated conformations are ordered based on their thermodynamic energy and a final set

is selected based on the structural diversity (calculated by RMSD) and energetic stability.

On the other hand, LigPrep is a tool from Maestro that prepares high quality ligands structures to different procedures that can take place in the same program, such as receptor-based molecular docking. This widget will define the appropriated and relevant tautomeric and ionizations states, ring conformations and stereoisomers to capture the corresponding 3D structure of the ligand (Schrödinger).

Flare (Cheeseright et al., 2006; Cresset®) is an advanced computational drug design platform that helps the visualization and analysis of 3D molecular structures. It supports both ligand-based and structure-based drug design approaches, enabling tasks such as quantitative structure-active relationship (QSAR) models, Activity Miner and Activity Atlas for ligand-based workflows as well as molecular docking and molecular dynamics for structure-based strategies. Mainly, this program was employed in order to perform a QSAR analysis aimed to determine key features that are associated with a greater inhibition activity. The crystal structure obtained from the PDB-REDO 5TZ1 was prepared to ensure the correct protonation and tautomeric states of amino acids residues at pH 7, reflecting a biologically relevant configuration.

An important component of the QSAR analysis was the evaluation of electrostatic complementarity (EC) between the ligand and the protein's active site. EC was computed using the polarizable XED force field. Moreover, EC scores are then calculated, showing scores ranging from 1 to -1, being 1 a perfect complementarity (represented in color green) and -1 electrostatic clashed (represented in color red).

Additionally, the Activity Miner (Cheeseright et al., 2006; Cresset®) tool enables an exploration of complex SAR patterns. This feature compares different molecules based on 3D electrostatic properties, shape and protein interaction patterns. Also, it clusters the different compounds by similarity. A key measure used in this analysis is the disparity, which quantifies the relationship between structural similarity and activity differences with a pair of compounds. To estimate

the disparity between a pair of active compounds, the following formula must be applied:

$$Disparity = \frac{\Delta Activity}{(1 - similarity)}$$

High disparity values indicate that the pair of compounds exhibit high similarity and large differences in activity. In other words, a small change in the structure, makes a big change in activity. These pairs offer the most valuable insight for SAR modelling.

Finally, an Activity Atlas (Cheeseright et al., 2006; Cresset®) model was generated to summarize the electrostatics, hydrophobicity and shape properties of the active compounds. This tool provides a 3D visualization of the regions where these properties contribute more to biological activity. Key features that influence to the potency of the inhibitor were highlighted in the Activity Cliff Summary, offering a guidance for the pharmacophore optimization.

Based on the insights obtained from the SAR study, a pharmacophore model was built using Phase (Dixon, Smondyrev, & Rao, 2006; Dixon, Smondyrev, Knoll, et al., 2006; Schrödinger). Maestro is a software application developed by Schrödinger that provides a comprehensive computational environment for molecular discovery, offering a wide range of tools that help in the development of new drugs. Phase (Dixon, Smondyrev, & Rao, 2006; Dixon, Smondyrev, Knoll, et al., 2006; Schrödinger) applies an algorithm that is useful in the appropriate development of a drug. A pharmacophore consists of a graphical representation summary of different characteristics, both steric and electrostatic, that a drug should have to interact with the binding site of a given target. The selected features embraced the key molecular traits identified in the SAR analysis and therefore are likely to interact with the active site of lanosterol 14α-demethylase, increasing the probability of effective inhibition.

To conclude the SAR study, key molecular interactions between the crystallized ligands and the target's binding site were analyzed using three different platforms: Maestro (Schrödinger), Flare (Cresset®) and a program developed internally by

the research group named CORAL-PIC (Correlation of Residues and Activities for Ligands (with PIC 50)).

Once SAR analysis results were obtained, a molecular docking of the validation library was performed in order to evaluate how many compounds could bind to the protein and how well they could align in the active site. This step was carried with different features from Glide, such as Grid Receptor Generations or Ligand Docking. Glide is a software used in the process of docking in the development of new drugs, as it looks for the favorable interactions between small molecules and the receptor protein (Friesner et al., 2004, 2006; Halgren et al., 2004; Yang et al., 2021; Schrödinger).

Those molecules that were successfully docked, were also filtered through a screening by the chosen pharmacophore (Dixon, Smondyrev, & Rao, 2006; Dixon, Smondyrev, Knoll, et al., 2006; Schrödinger). This pharmacophore model enables to identify only those compounds that have the features previously defined. After applying these filters, the enrichment factor (EF) was calculated. EF is defined as measure of how many times a sample is enriched with active compounds after filtering it (Gimeno et al., 2019). The EF formula used is the following:

$$EF = \frac{\frac{actives\ after\ filter}{total\ compounds\ after\ filter}}{\frac{actives\ before\ filter}{total\ compounds\ before\ filter}}$$

Finally, the same methodology was applied to the Specs library of chemical compounds. Specs is a collection of thousands of compounds available not only to download and perform a virtual screening methodology, but also to purchase those that have stood out in a study. After the correct preparation of the library, compounds were docked to the binding site of lanosterol 14-alpha demethylase. Subsequently, the previously developed pharmacophore model was applied as a filter tool to reduce the number of hits only by selecting those compounds that possess features likely to contribute to inhibitory activity. From this subset, those compounds that presented a most of the pharmacophoric features and showed favorable docking scores were identified as potential inhibitors of the target enzyme.

3. Hypothesis and objectives

The hypothesis of the current research project is that a ligand-based virtual screening approach will be able to enrich a compound library in active compounds against the target lanosterol 14α-demethylase of *Candida albicans* to successfully identify novel hit compounds capable of inhibiting this enzyme.

The main objective of this research project is to develop and implement an efficient virtual screening process to identify potential novel inhibitors of lanosterol 14α-demethylase of *Candida albicans* from a chemical library of compounds.

Specific objectives:

1. Collect active compounds and use them to design a pharmacophore model based on their common structural and chemical features.
2. Apply molecular docking techniques in order to predict the binding poses of compounds and evaluate their binding affinity to the active site of lanosterol 14α-demethylase.
3. Evaluate the performance, reliability and effectiveness of the virtual screening workflow through comprehensive validation analysis.

4. Material and methods

Actives retrieve

In order to automatize the retrieval of active compounds for the target lanosterol 14α-demethylase, a Python script was developed and executed using Google Colab. This script attempts to connect with the ChEMBL database via the *chembl_websource_client* Python package (Chembl-Webresource-Client · PyPI). In addition, RDKit was installed to explore and manipulate molecular structures (RDKit: Open-Source Cheminformatics. https://www.rdkit.org).

To ensure the comprehensive retrieval of all the known active compounds, the script searched for several synonymous names of the target protein. These included: *lanosterol 14α-demethylase*, *sterol 14-alpha demethylase*, *14-alpha sterol demethylase*, *cytochrome P450 51*, *CYP51A1*, *cytochromeP450LI*, *CYPLI*, *cytochrome P450-14DM*, *ergosterol biosynthesis protein 11*, *P450 sterol 14-alpha-demethylase* and *cytochrome P450 14alpha-demethylase*. All of these 11 different names for the same target were found in a preliminary search through ChEMBL database and UNIPROT database, as well as from scientific literature from the documentation of the crystallized structures and other literatures sources.

After retrieving 10,000 initial entries, the dataset was filtered by name and source organism. In order to isolate those entries that belonged to the fungal species, a dataset of 537,509 fungal species was downloaded from Mycobank database (Mycobank). From this subset, only entries from *Candida albicans* were further analyzed. Moreover, compounds only with defined inhibitory values (IC_{50}, MIC_{50}, MIC_{100}, pIC_{50}, K_i and inhibition) were selected for further analysis, as they are measures that indicate some kind of inhibition of the enzyme.

Fingerprint similarity heatmap

In order to evaluate the structural similarity between a set of compounds, a script was used to calculate pairwise similarity based on molecular fingerprints. These fingerprints were generated using Morgan algorithm with a radius of 2, capturing local chemical environments around each atom. Compounds with a Tanimoto

similarity score of 0.6 or higher were gathered in the same cluster, allowing the identification of structurally related structures.

This tool uses different packages such as RDKit (RDKit: Open-Source Cheminformatics. https://www.rdkit.org) and Matplotlib (Hunter, 2007) to represent the results on a heatmap and group the different molecules in clusters based on their similarity.

CORAL-PIC (COrrelation of Residues and Activities for Ligands (with pIC_{50}))

Additionally, another internal script was used in order to estimate the interactions between the crystallized ligand-protein structures. This uses Arpeggio, which is a tool that calculates the contacts in between atoms in a protein (Jubb et al., 2017). These interactions between the different proteins and their corresponding ligands are compilated in a heatmap where it shows the frequency of these contacts.

Decoys generation

The decoys were found by using a platform called LUDe (LIDEB's Useful Decoys) that generates decoys based on a collection of active compounds previously found. For this study, the default LUDe settings were applied, ensuring that up to 50 decoys were generated per active compounds (Gori et al., 2022).

OMEGA

To run OMEGA application (OpenEye, n.d.), several command-line options were applied. *-in* was used to specify the input file containing the set of molecules to be analyze, and *-out* defined the name of the output file, where results were stored. *-maxconfs* parameter was set at 1 in order to generate only one conformer per molecule, based on the number of rotatable bonds. *-strictstereo* was implemented to exclude those compounds with unspecified stereochemistry. *-progress percent* was used in order to monitor the progress in real time, and *-mpi_np* was set at 30 processors to execute the job using 30 processors in MPI mode (Hawkins et al., 2010).

Flare (Cresset)

a) Protein Preparation

The default settings configured for the protein preparation protocol were not changed. These include: allowing all changes to protonation/tautomer state, allowing small side chain movements, removing atoms from residues with incomplete backbone, filling in 1- and 2-residues gaps and allowing asparagine, glutamine and histidine flips.

The calculation method used was “Normal” and the configuration of the cap chains used was “Intelligence Capping”, which will add ACE to the start of the truncated chain and NME to the end. Moreover, waters outside the active site (defined as 6.00 Å from the ligand) were removed and the active site size remained unmodified. The prepared protein structures were then saved as a new entry. Finally, the co-crystallized ligand was automatically extracted and assigned to the Ligands Table. (Bauer & Mackey, 2019a; Cheeseright et al., 2006; Cresset®; Kuhn et al., 2020)

b) Conf Hunt and Align

The different ligands used in this study were aligned using Flare’s Conformation Hunt & Alignment tool. To ensure high accuracy, the “Very accurate but slow” conformation hunt setting was selected, which generates low-energy 3D conformers, including possible stereoisomers if chirality is unspecified. Moreover, the different ligands were “Normal” aligned in those ligands with reported IC_{50} values, and “Aligned by substructure” in the rest of the ligands used in the SAR study. This approach ensures that the conformation of the ligands is reliable (Bauer & Mackey, 2019b; Cheeseright et al., 2006; Cresset®; Kuhn et al., 2020).

c) Activity Atlas

For model type settings, the default option Activity Atlas ™ (AA) was selected, with the model mode set to “Normal” (Cheeseright et al., 2006; Cresset®).

Maestro (Schrödinger)

a) Protein Preparation Workflow

The Protein Preparation Workflow in Maestro was used with mostly default settings (Madhavi Sastry et al., 2013; Schrödinger).

In the Global settings of Specify Protein, the simulation pH was set at 7.4. Moreover, options to process ligands, metals and ions and other as small molecules ("hets") were enabled. The option "Prompt for FASTA file in sequence is missing" was also selected.

Both options of Preprocess were selected (Cap termini, to add the N-acetyl and N-methyl amide groups to the truncated chains of the protein, in case the extremes do not have them; and fill in the missing side chains). Those features that were checked in the foldable *More Options* were: *Assign bond orders: Reassign all (CDD)*, *Replace hydrogens*, *Create Zero-order bonds to metals* and *Disulfide bonds*, *Identify protein features: Antibody* and *Antibody scheme: Kebat*, *Delete waters beyond hets: 8.00 Å*, and *Generate het states (with Epik): pH: 7.4 +/- 2.0, Max states to process automatically 1*.

Settings for Optimize H-Bonds Assignments that were checked were: *Sample water orientations, Consider all ligand states* and *Optimization uses: PROPKA*.

Finally, settings for Minimize and Delete Waters were marked as follows: *Minimize: All atoms - to max RMSD 0.30 Å*, and *Delete waters: Distant from ligands (hets): 5 Å*.

b) LigPrep

Ligands were prepared using the LigPrep function of Maestro. This was set to use *Epik* in a pH of 7.4 ± 1.00 (Madhavi Sastry et al., 2013; Schrödinger).

c) Glide

The Grid Receptor Generation is a feature of Glide used to design a grid based on the ligand from structure 5TZ1. It was located automatically in the active site of the protein, represented with an outer box colored pink. The x, y, z coordinates of the grid used in this study were: 70.82, 66.30, 4.34. The size of the outer box

was: 30.36, 34.36, 29.36. The size of the inner box (represented in color green in Figure 18) was: 11, 15, 10.

Moreover, in order to perform the docking of the different libraries, the Ligand Docking tool was used. Settings for this program were left as the default ones. Only one pose per ligand was generated (Friesner et al., 2004, 2006; Halgren et al., 2004; Schrödinger, n.d.-a; Yang et al., 2021).

d) Phase

One of the tools used in the Phase interface was the “Develop Pharma Hypothesis” function where from a ligand that was selected in the entry list, a pharmacophore was created using a model “Receptor-ligand complex”. Moreover, the method chosen to execute this program was the “Auto (E-Pharmacophore)” where it creates automatically a model of pharmacophore based on the complex ligand-protein. Other settings were left as default.

In order to check whether a collection of compounds has the characteristics of the pharmacophore model, a screening must be done. To perform it, the “Phase Ligand Screening” Phase tool was used. Configuration were set to match a certain number of features designed in the pharmacophore model. Moreover, screening settings options “Use existing conformers” and “Score in place” were checked (Dixon, Smondyrev, & Rao, 2006; Dixon, Smondyrev, Knoll, et al., 2006; Schrödinger).

5. Results and discussion

In order to apply the virtual screening methodology previously described (see Figure 5), a set of steps was established and color-coded for clarity. First, the validation library was created, consisting of 41 known active compounds and 1,776 generated decoys, totaling 1,817 molecules (blue steps shown in Figure 5). After preparing the library a total of 1,358 compounds, the analysis was performed. Subsequently, a molecular docking was carried out using a grid specifically designed for the reference structure (green steps in Figure 5). In parallel, a QSAR study was conducted to identify the key features that contribute most to inhibitory activity. Based on these insights, a pharmacophore model was generated (pink steps in Figure 5). The results of the docking (1,025 molecules) were then screened using the pharmacophore model (green steps in Figure 5), retaining only those small molecules with overlapping characteristics (a total of 62). From this subset, 6 active compounds were successfully recovered, resulting in an EF of 6.61. Following the same validated strategy, the Specs chemical compound library was docked to the target and then screened. A total of 8,052 compounds were successfully docked and filtered through the pharmacophore model (orange steps shown in Figure 5). Hence, 20 molecules were proposed as potential inhibitors of lanosterol 14α-demethylase. The results from each step are shown in Figure 6.

	Validation library			Specs library
	Active	Decoys	EF	
Initially	41	1776	-	202596
Preparation	15	1343	-	144588*
Molecular Docking	15	1010	1.3249	76665
Pharmacophore	6	56	6.6129	8052

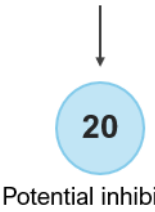


Figure 6. Summary of the number of compounds retained at each step of the virtual screening methodology workflow.

*The Specs library was prepared using OMEGA and LigPrep, using the same protocol as the validation library, but it was additionally filtered by molecular weight with the range of 300 and 730 Da.

5.1. Validation library

5.1.1. Active retrieve

A total of 40 active compounds were found from the ChEMBL database based on their reported inhibitory activity against the target enzyme lanosterol 14α-demethylase from *Candida albicans*. As detailed in the corresponding section in Material and Methods, these compounds presented experimentally validated activity values, specifically IC_{50} or MIC_{50}.

Moreover, the Protein Data Bank (PDB) was checked, and three crystal structures of the target protein bound to different inhibitors were identified. Two of these ligands had already been found during the ChEMBL search, while the third one was not included in the database and lacked associated activity data. Therefore, this compound was excluded from the subsequent structure-activity relationship (SAR) analysis.

Hence, a total of 41 compounds were selected to constitute the set of active molecules. Their structural similarity was evaluated using a dendrogram based on molecular Morgan fingerprints. As shown in Figure 7, a cluster of 17 compounds (highlighted in orange) was identified, indicating a high degree of similarity among them, as they are represented by lighter shades of blue in the heatmap. Structural similarity within this cluster was further explored in the SAR analysis. In contrast, the remaining 24 active molecules displayed greater structural diversity, as they are represented by darker shades of blue in the heatmap. However, this was not the only cluster identified. In total, 6 distinct clusters were observed, each highlighted by different colored lines for visual differentiation. These clusters were formed based on Tanimoto similarity, where compound pairs with a similarity score equal or greater than 0.6, were grouped into the same cluster.

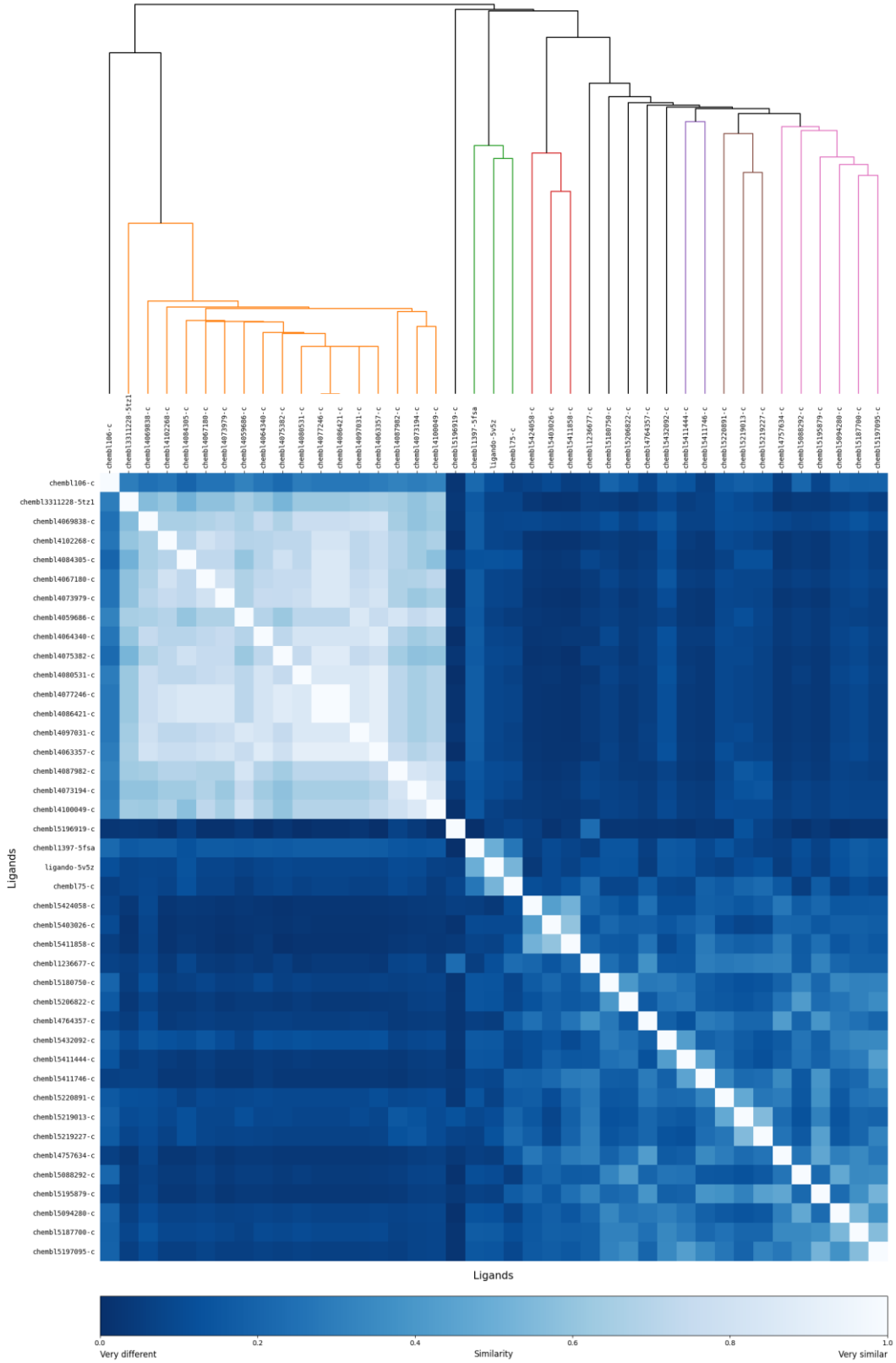


Figure 7. Heatmap representing the structural similarity among the 41 active compounds based on molecular fingerprints. Darker shades of blue indicate high dissimilarity, while lighter shades represent high similarity. Compounds with a Tanimoto similarity ≥ 0.6 are grouped into distinct clusters, each one highlighted with different colors to distinguish visually the groups of structurally similar molecules.

This selected set of actives constituted the basis for the following *in silico* studies, including the pharmacophore model generation, decoy selection and molecular docking.

5.1.2. Development of a pharmacophore model

In order to develop a pharmacophore model, a series of sequential steps were followed. First, a structure-activity relationship (SAR) study was performed to identify the key molecular features responsible for the inhibitory activity. Next, the molecular interactions between the co-crystallized inhibitor and the target protein were identified. Finally, based on these analyses, the pharmacophore model was designed.

SAR study with Flare

In order to execute the SAR study using Flare (Cresset®), the collected active compounds were divided into three groups: crystallized ligands, compounds with reported MIC_{50} values (orange cluster in Figure 7) and compounds with reported IC_{50} values. The same methodology was applied independently to each group, as the experimental activity data types are not directly comparable. Since MIC_{50} and IC_{50} represent different types of inhibition, merging them into a single dataset would compromise the reliability of the analysis. Therefore, they were analyzed separately.

Firstly, the EC of all ligands within each group was analyzed. EC corresponds to the match of electrostatic potential between the surface of the active site of the target protein and the small molecule. The results, summarized in Table 1, showed consistently low EC values across all groups, indicating a poor alignment between the electrostatic surfaces of the ligands and the target protein's active site. This lack of complementarity may lead to unfavorable interactions including steric clashes and electrostatic repulsions. Therefore, this analysis did not provide significant insights into ligand binding or activity trends.

Table 1. Mean and standard deviation of the electrostatic complementarity within the active compounds

	Crystallized	**MIC_{50}**	**IC_{50}**
EC	0.2225 ± 0.118	0.2695 ± 0.04	0.1473 ± 0.0669

To better understand the chemical features that have a significatively impact in the inhibitory activity among the ligands, Activity Miner and Activity Atlas were performed. However, in order for these tools to generate reliable results, a minimum of 20 small molecules is recommended to be used. Among the three groups analyzed, only one met this requirement. Therefore, the results obtained from these two groups should be analyzed carefully, as the outcome may lack of robustness.

The Activity Miner's results for the crystallized ligands (VT1 and X2N) displayed low disparity (2.3) meaning that the compounds exhibit moderate similarity (0.493) and their activities values differ only slightly (Figure 8). This analysis suggested that these compounds, while being structurally different, have a similar effect on the target protein.

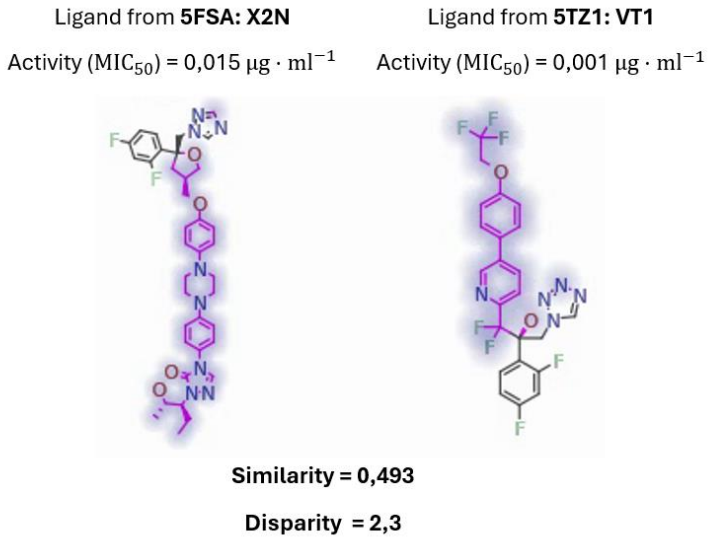


Figure 8. Comparison of the 2D chemical structures of inhibitors X2N (left) and VT1 (right) from crystal structures 5FSA and 5TZ1 respectively. The structural similarity between the two compounds is 0.493 and the disparity value is 2.3. Regions highlighted in purple indicate those areas that are different between the molecules.

Moreover, Activity Atlas was used to identify 3D patterns based on electrostatic and hydrophobicity. However, the obtained results did not provide significant guidance for understanding how an ideal inhibitor to the target should be, as they were inconclusive (Supplementary Figure 1). These uninformative results could be due to the limited dataset as there were only two compounds in the SAR study of the crystallized compounds. As mentioned before, results obtained from a

limited dataset (fewer than 20 compounds) should be inspected carefully as they could fail to capture reliable information about the relationship between activity and structure.

Regarding compounds with reported MIC_{50}, the results of the Activity Miner were more informative. A matrix displaying the disparity scores for each pair of compounds that were analyzed, is shown in Figure 9 and four pairs that have demonstrated the greater disparity were further analyzed. These pairs were compared, and the results were summarized in Figure 10. Higher disparity values meant that they exhibited good structural similarity but have presented significant differences regarding their activity. In other words, a slightly difference in their structure had an impact on the activity of the compound.

	CHEMBL4075382	CHEMBL4084305	CHEMBL4080531	CHEMBL4064340	CHEMBL4063357	CHEMBL4077246	CHEMBL4073979	CHEMBL4086421	CHEMBL4069838	CHEMBL4067180	CHEMBL4097031	CHEMBL4102268	CHEMBL4087982	CHEMBL4100049	CHEMBL4059686	CHEMBL4073194	CHEMBL1397
CHEMBL4075382	0	-3,3	-13,4	-4,4	-3,9	-0,6	-0,4	-1,9	-2,2	0	-5,5	-5,8	3,2	2,6	0	-1,2	0
CHEMBL4084305	3,3	0	-8,4	-1,3	-2,2	2,2	2	0	0	1,6	-3,6	-4,5	4,9	4	1,7	0	0,5
CHEMBL4080531	13,4	8,4	0	14,6	5	7,6	9,8	5,6	8,6	9,2	2,2	1,3	11,6	9,2	5,7	4	2,2
CHEMBL4064340	4,4	1,3	-14,6	0	-0,9	2,2	2,9	1,1	1,3	4	-3,1	-3,2	6,5	5,1	2,2	0,8	0,9
CHEMBL4063357	3,9	2,2	-5	0,9	0	6,8	5,8	1,4	2,2	3,7	-1,6	-2,5	9,7	6,4	3,7	1,9	1,2
CHEMBL4077246	0,6	-2,2	-7,6	-2,2	-6,8	0	0,3	-1,1	-1,3	0,5	-4,2	-4,7	4	3,1	0,7	-0,9	0,2
CHEMBL4073979	0,4	-2	-9,8	-2,9	-5,8	-0,3	0	-1,5	-2	0,2	-5	-5,2	4,3	3,3	0,2	-1,2	0,1
CHEMBL4086421	1,9	0	-5,6	-1,1	-1,4	1,1	1,5	0	0	1,6	-4	-3,5	4,6	4,2	1,6	0	0,5
CHEMBL4069838	2,2	0	-8,6	-1,3	-2,2	1,3	2	0	0	2	-3,8	-4,1	5,2	4	1,2	0	0,6
CHEMBL4067180	0	-1,6	-9,2	-4	-3,7	-0,5	-0,2	-1,6	-2	0	-9,6	-8,7	4,1	3,3	0	-1,4	0
CHEMBL4097031	5,5	3,6	-2,2	3,1	1,6	4,2	5	4	3,8	9,6	0	-1	8,1	7,1	4,3	2,7	1,6
CHEMBL4102268	5,8	4,5	-1,3	3,2	2,5	4,7	5,2	3,5	4,1	8,7	1	0	8,9	7,2	4,9	3,9	1,8
CHEMBL4087982	-3,2	-4,9	-11,6	-6,5	-9,7	-4	-4,3	-4,6	-5,2	-4,1	-8,1	-8,9	0	0	-6,1	-6,6	-1
CHEMBL4100049	-2,6	-4	-9,2	-5,1	-6,4	-3,1	-3,3	-4,2	-4	-3,3	-7,1	-7,2	0	0	-3,9	-5	-1
CHEMBL4059686	0	-1,7	-5,7	-2,2	-3,7	-0,7	-0,2	-1,6	-1,2	0	-4,3	-4,9	6,1	3,9	0	-1,8	0
CHEMBL4073194	1,2	0	-4	-0,8	-1,9	0,9	1,2	0	0	1,4	-2,7	-3,9	6,6	5	1,8	0	0,5
CHEMBL1397	0	-0,5	-2,2	-0,9	-1,2	-0,2	-0,1	-0,5	-0,6	0	-1,6	-1,8	1	1	0	-0,5	0

Figure 9. Disparity matrix of active compounds with reported MIC_{50} values. The matrix, generated in Flare, display the disparity values both in positive and negative. For this reason, values should be considered as absolute values. Color intensity reflects the degree of disparity: red and blue boxes indicate high disparity, while white boxes represent low disparity. The darker the color, the greater disparity between the pair of compounds.

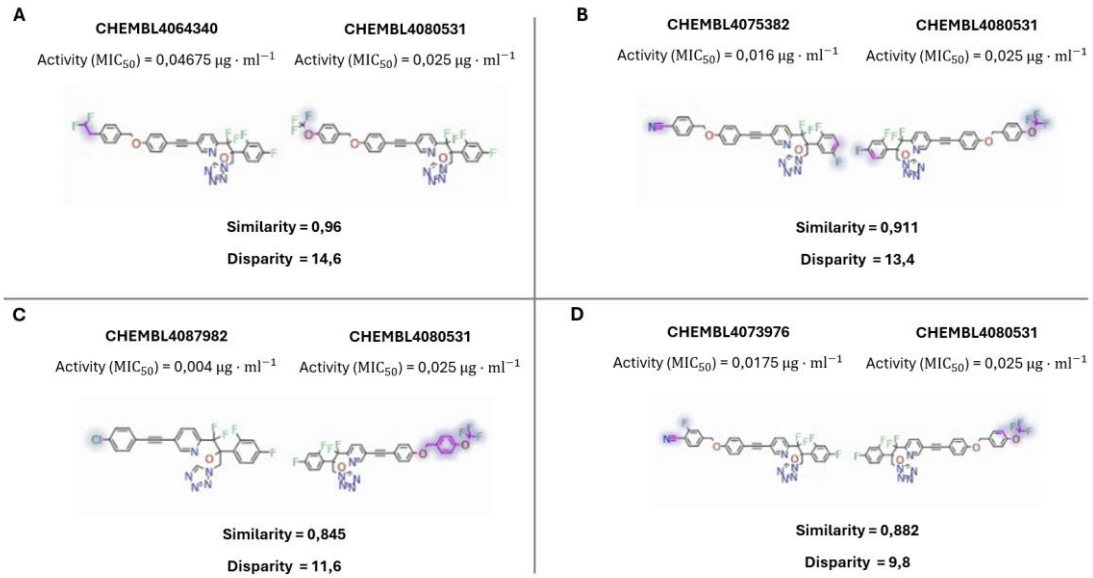


Figure 10. Comparison of the 2D chemical structures of different pairs that presented high disparity in the matrix (Figure 9). Regions highlighted in purple indicate those areas that are different between the pair of molecules.

Furthermore, the Activity Atlas study demonstrated poor results regarding electrostatic charges, likely due to the overall low EC observed among the ligands. However, the Average Hydrophobicity showed that the central region of the inhibitor tends to be hydrophobic (see Figure 11).

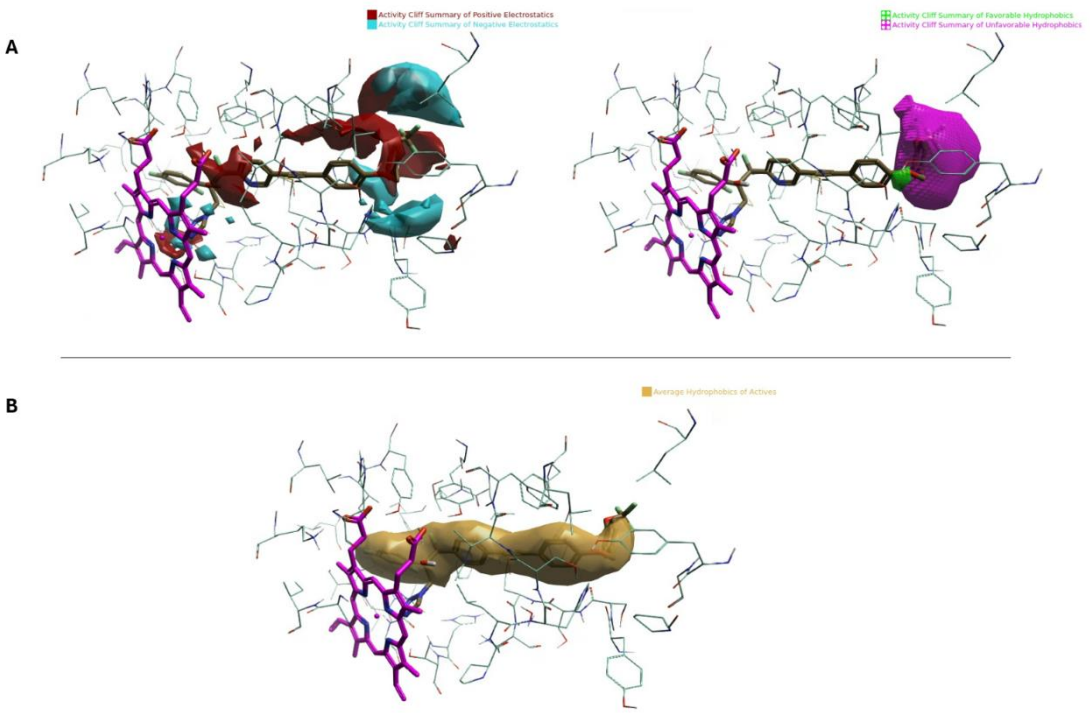


Figure 11. (A) Activity Cliff results comparing all active compounds with reported MIC_{50} values. The molecule CHEMBL4080531 is shown in sticks and colored by CPK. Heme prosthetic group is also shown in sticks colored by CPK and with a pink backbone. The model highlights regions with positive (red surfaces), and negative electrostatics (blue surfaces) (top left), and favorable (green surfaces) and unfavorable (pink surfaces) hydrophobic regions (top right). **(B)** Summary of regions where active compounds commonly exhibited hydrophobic characteristics (yellow surfaces).

Finally, compounds with reported IC_{50} were analyzed by the Activity Miner tool. The overall disparity values between the pairs were low. This is because of the structural differences among the compounds as they are not similar (similarity values in between 0.346 and 0.637). The disparity values are gathered in the disparity matrix shown in Supplementary Figure 2. Due to the low values in disparity obtained from this Activity Miner study, it does not contribute to the better understanding of essential features that should be in the ideal inhibitor.

Finally, Activity Atlas showed that the core of the inhibitors was hydrophobic (Figure 12B), as obtained in those ligands with described MIC_{50} values (Figure 11). Moreover, Activity Atlas reinforced this fact as shown in Figure 12A (top right) with the favorable hydrophobic surfaces. Activity Atlas also showed that near the heme group, compounds should have negative charges in order to form a charge-charge interaction with the iron from the prosthetic group (Figure 12A, top left corner).

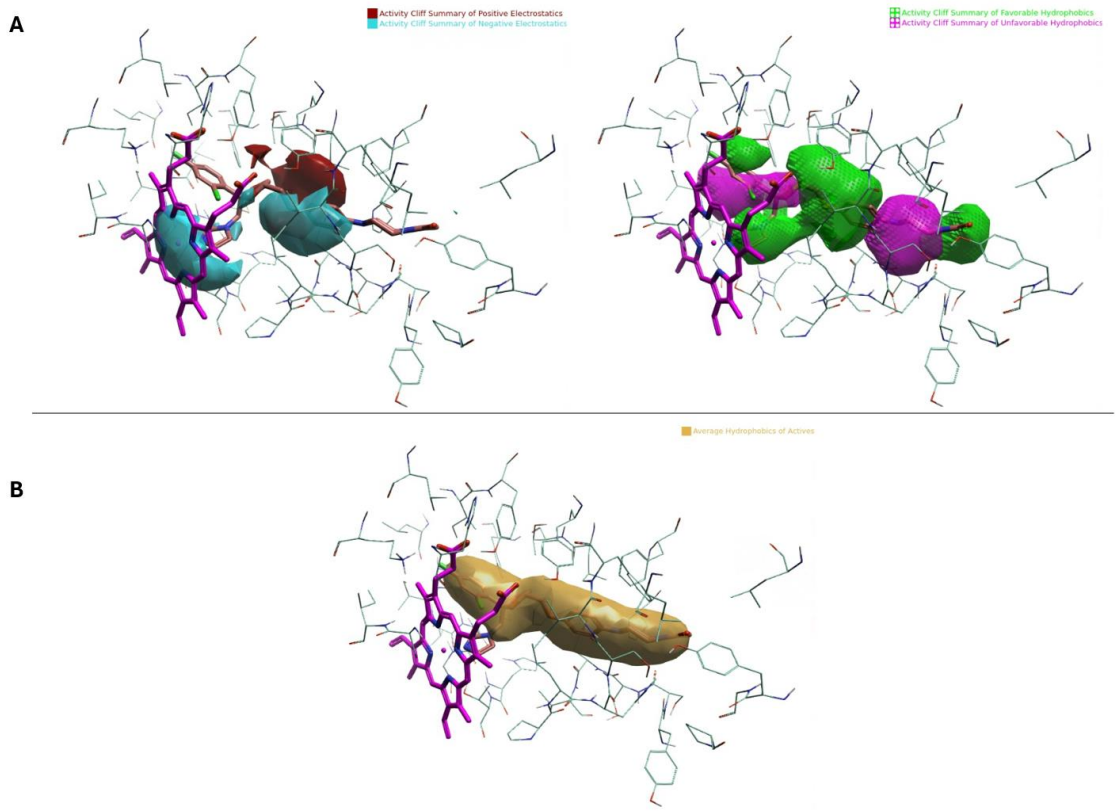


Figure 12. **(A)** Activity Atlas results comparing all active compounds with reported MIC_{50} values. The molecule CHEMBL75 is shown in sticks and colored by CPK. Heme prosthetic group is also shown in sticks colored by CPK and with a pink backbone. The model highlights regions with positive (red surfaces), and negative electrostatics (blue surfaces) (top left), and favorable (green surfaces) and unfavorable (pink surfaces) hydrophobic regions (top right). **(B)** Summary of regions where active compounds commonly exhibited hydrophobic characteristics (yellow surfaces).

Identification of interactions

To identify the key interactions between the ligands and the binding site of the target protein, only the crystallized structures were analyzed by using different tools.

Firstly, a Maestro tool called Ligand Interaction Diagram, was used to determinate the interactions between the ligand and the binding site. As shown in Figure 13, only a few key interactions were observed, being the main ones Pi-Pi stacking and Pi-cation. Additionally, the surrounding environment of the inhibitors, that form a complex with the protein, was mainly hydrophobic as illustrated in Figure 13 with a green line around the ligand.

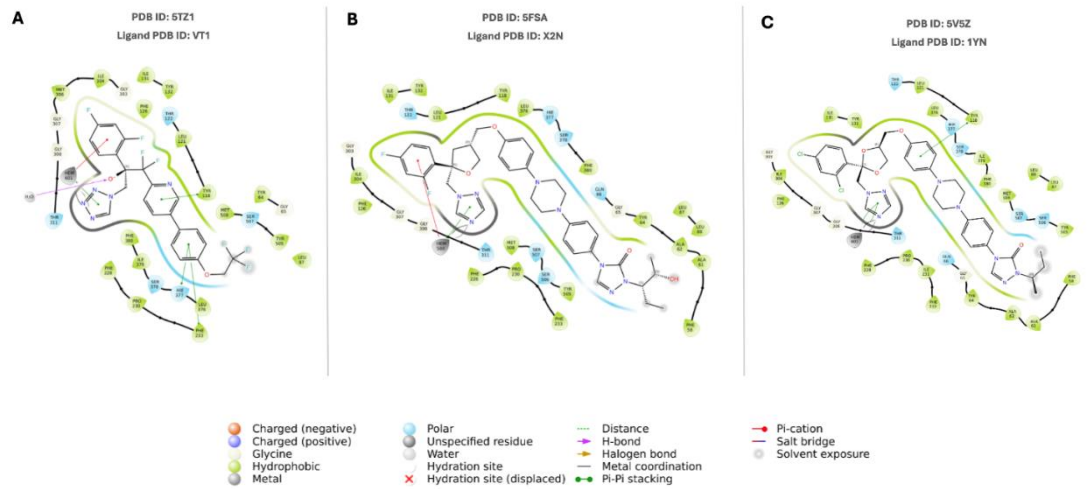


Figure 13. Interactions diagrams of the co-crystallized ligands **(A)** VT1, **(B)** X2N and **(C)**1YN with the active site residues of lanosterol 14α-demethylase obtained with Ligand Interaction Diagram (Maestro, Schrödinger). Green lines represented indicate π-π stacking interactions, red lines show π-cation and pink lines correspond to hydrogen bonds. Green lines around the ligand means a hydrophobic environment, grey lines represent exposure to the iron atom of the heme group, blue regions indicate polar regions, and pale yellow highlight ligand exposure to a glycine residue.

Secondly, similar results were observed using Flare, which identified aromatic-aromatic interactions (pink), salt bridges (also pink) and hydrophobic contacts (grey). Among these interactions, hydrophobic contacts were the most abundant in between the three crystallized structures, as shown in Figure 14.

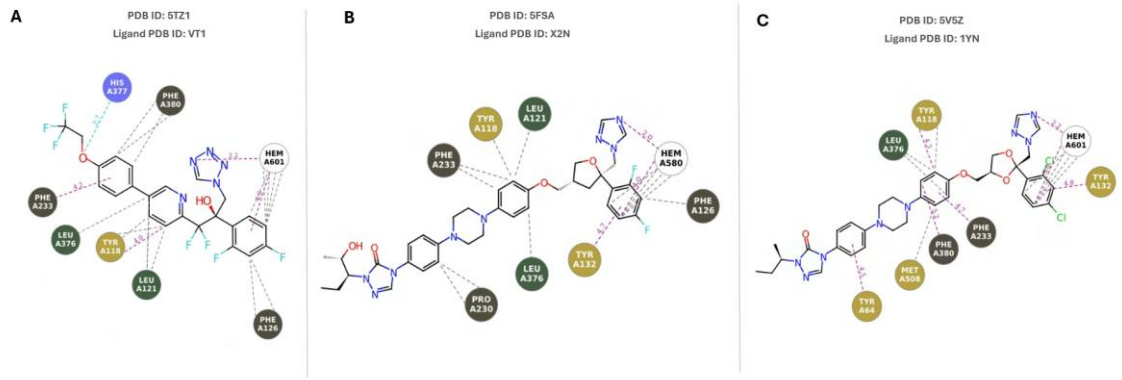


Figure 14. Interactions diagrams of the co-crystallized ligands **(A)** VT1, **(B)** X2N and **(C)** 1YN with the active site residues of lanosterol 14α-demethylase obtained with Flare. Light blue lines represented indicate weak hydrogen bonds, pink lines show aromatic-aromatic and salt bridge/metal interactions, and grey lines indicate hydrophobic contacts.

Finally, the interactions of the different structures were also analyzed by an internal program developed by the research group called CORAL-PIC (COrrelation of Residues and Activities for Ligands (with pIC_{50})). The results,

shown in Figure 15, demonstrated also that the main interactions between the ligands and the protein structure were hydrophobic nature.

Figure 15. Heatmap illustrating the interactions among the three co-crystallized ligands and the active site residues of lanosterol 14α-demethylase. White squares indicate the absence of an interaction. Red squares represent interactions observed in only one of the three structure, yellow squares indicate interactions shared by two structures, and green squares demonstrate common interactions to all three crystal structures.

Overall, these three programs that analyze the interactions between the crystallized ligands and the binding site from the target protein, suggest that hydrophobic interactions are the predominant ones, with aromatic interactions playing a crucial role. Therefore, the pharmacophore designed for this target, should prioritize hydrophobic features and include aromatic rings to support these interactions.

Development of a pharmacophore model

Based on the SAR analysis and the interactions identified, several pharmacophore models were generated using the ligand VT1 from the crystallized structure with PDB ID 5TZ1 as a reference. Multiple pharmacophores were designed to optimize the performance of each model. Some of the generated pharmacophores are listed in the Supplementary Table 1. However, the one that generated better results was the one shown in Figure 16A. Moreover,

Figure 16B illustrate the superposition of the chosen pharmacophoric features with the reference ligand.

This pharmacophore model was selected not only for a good EF, but also the total number of actives compounds that were retrieved. As shown in Supplementary Table 1, although the selected model (first entry) did not present the greatest EF, since the ratio of actives to decoys was lower compared to the second, third and fourth entries, it outperformed the others in terms of number of active compounds screened. Both metrics were equally prioritized for the study.

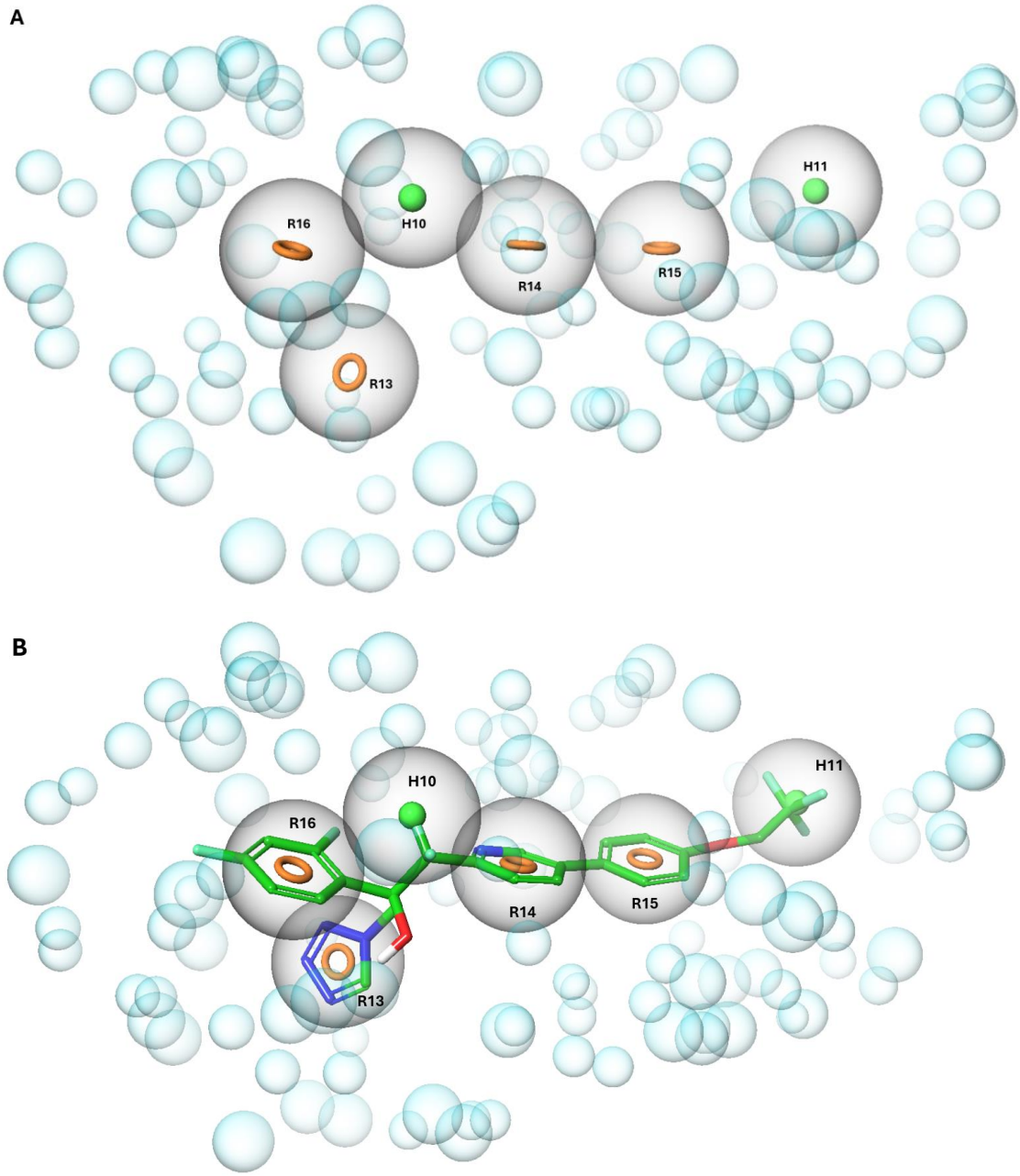


Figure 16. (A) Final pharmacophore model summarizing the six key features. **(B)** Superposition of the reference ligand VT1 with the chosen pharmacophore model. Rings are represented in orange circles and hydrophobic characteristics are represented in green spheres. The grey spheres around each features illustrate the tolerance radius, set at 2.00 Å. Blue spheres represent exclusion volume, that are regions where ligand atoms are not allowed to be placed due to potential steric clashes with active site residues.

The pharmacophore in Figure 16 has 6 features. Two rings (R15 and 16) that are required as they establish key interactions with the heme prosthetic group (R16) and aromatic interactions with the residues from the binding site (R15). Ring R14 was configured as permitted and it corresponds to either a ring or a hydrophobic group. Then, another ring (R13) and two hydrophobic groups (H10 and H11) were

assigned as permitted. These permitted features are not mandatory in the screening process that is performed and described in section 5.1.4. Protein-ligand docking. In contrast, required features must be there for a compound to be selected during the screening. This process, that was performed from a set of different small molecules, retained only those compounds that matched the required pharmacophoric features and the permitted ones.

5.1.3. Decoys generation

In order to generate the decoys, the methodology described in the corresponding section of Materials and Methods was followed. As a result, a total of 1,776 compounds were identified as molecules with physicochemical properties similar to the active compounds, but with a low probability of interacting with the target protein.

The expected number of decoys was 50 per active compound (as established in the settings), with a theoretical total of 2,050 decoys. However, only 1,776 were retrieved. This shortfall of 274 compounds may be due to the high degree of similarity among active compounds that belong to the same cluster. As discussed in the SAR study conducted using Flare, the active compounds with reported MIC_{50} values shared highly similar molecular structures. Such redundancy within the active dataset reduces the chemical space available for decoy selection, as the algorithm must identify compounds that are physicochemical similar but structurally distinct from each active. So, when many actives share almost identical structures, it becomes more difficult to find valid decoys that meet the required criteria. As a result of this, a lower number of decoys were found.

5.1.4. Protein-ligand docking

After combining all the retrieved decoys and active compounds into a single file, the validation library was generated. This library was prepared following the OMEGA and LigPrep protocols. As a result of this step, compounds with undefined chirality were removed. Therefore, the final number of compounds in the validation library was lower, as shown in Figure 6. A total of 1,358 compounds were compiled and used for the subsequent computational analysis.

Before starting the docking process, the reference protein was properly prepared using the Protein Preparation Workflow in Maestro to minimize structural issues. This step was essential to address issues that were identified in the Protein Reliability Report (as shown in Figure 17A), where larger and redder circles indicate more severe or numerous structural problems. The different features that are analyzed are divided into four categories according to their severity. In Figure 17A, Steric Clashes is classified in the second category, indicating that there are important structural issues in the full structure. However, Peptide Planarity belongs to the third category, which is associated to minor structural issues in the full structure. After preparing the protein, these issues changed, as illustrated in Figure 17B, where the same problem regarding Peptide Planarity remains active, but the circle is bigger compared with before preparation. However, the Steric Clashes were solved as the red circle changed into a green one. Nevertheless, Improper Torsion, appeared after preparation. This issue belongs to the third category as well, as it is a minor structural issue in the full structure. Therefore, major structural issues were solved by preparing the protein.

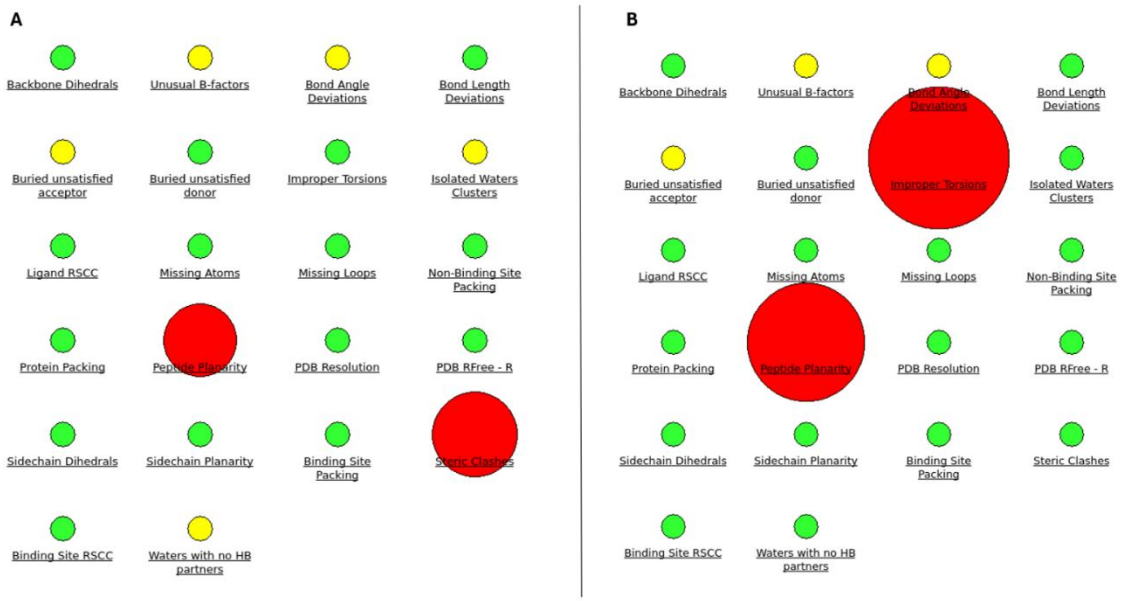


Figure 17. Results of the Protein Reliability Report **(A)** before and **(B)** after preparing the protein with PDB ID 5TZ1 with the Protein Preparation Workflow tool.

The next step in the workflow was the generation of the docking grid, which was based on the crystallized structure of lanosterol 14α-demethylase with PDB ID 5TZ1. Multiple trials were carried out using grids which varied inner box sizes

(green box) to identify the configuration that yielded the best performance. The main objective was to choose the grid that maximized the recovery of actives while minimizing the number of decoys. The grid selected for the virtual screening process is described in detail in the corresponding section in Materials and Methods. The final grid configuration used in this study is shown in Figure 18.

The grid was specifically designed for the ligand VT1 of the reference structure. Instead of a cubic grid, a rectangular inner box was used to include the full extent of the binding site. Several alternative grid configurations with larger inner box sizes were tested. However, the outcome let to unreliable poses of docking (data not shown), as compounds were predicted to bind to the solvent exposed regions of the protein rather than to the binding site. Therefore, the final grid provided the better results.

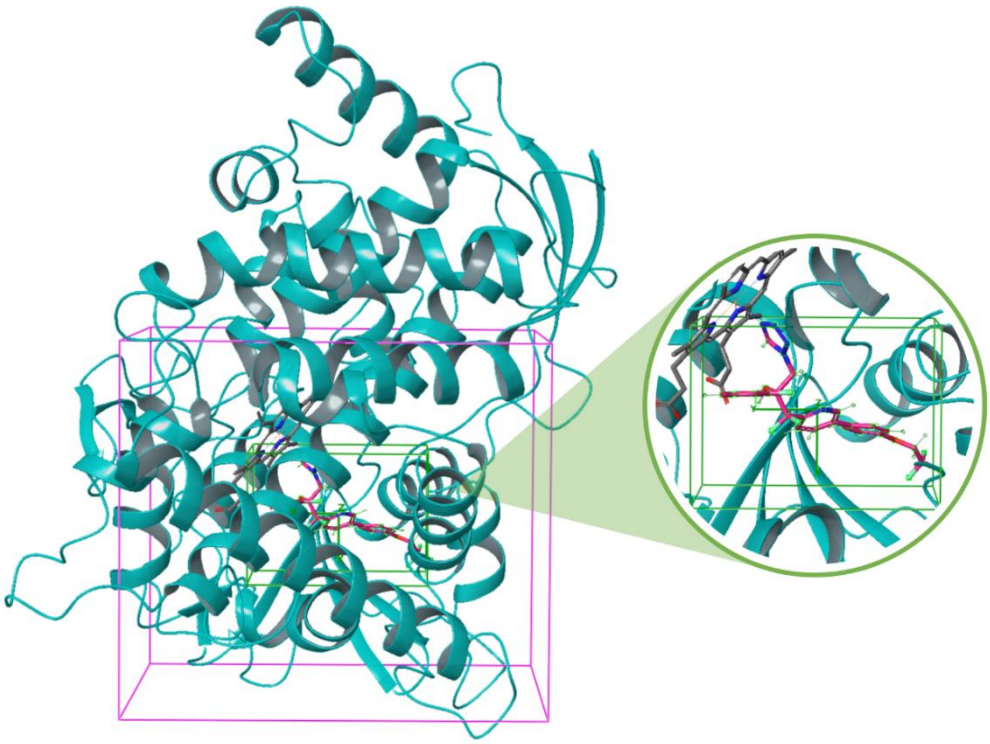


Figure 18. Receptor grid generated for the crystal structure with PDB ID 5TZ1. The pink box represents the outer box, while the green box corresponds to the inner box, which was adjusted to fit the co-crystallized inhibitor VT1. Lanosterol 14α-demethylase is represented in blue ribbon format. In its binding site, heme prosthetic group is found in sticks format and colored in grey. The inhibitor of the crystallized structure is represented in sticks format and colored by CPK, however its backbone is colored in pink

Before conducting the molecular docking of the validation library, a redocking and crossdocking procedures were performed to validate the selected grid. The redocking process consists of performing a docking process between a known

crystallized ligand and its native protein. This method was carried out using the reference ligand with the reference protein (PDB ID: 5TZ1), while the cross-docking was performed by docking the ligands from the other crystallized complexes into the reference protein. The goal of this step was to ensure that crystallized compounds could be properly docked into the protein's binding site. Although all three co-crystallized ligands were docked, the results did not meet expectations, suggesting limitations in the docking accuracy.

The results of the redocking seemed to show an overall good alignment with the original pose, although there are some deviations as illustrated in Figure 19, where the co-crystallized ligand is represented in pink while the redocked is represented in blue. These deviations caused a RMDS of 4.04, which means that these two conformations cannot be considered as equal. In order to consider them as equals, the RMSD value should be less than 2.

However, the crossdocking process is defined as the docking of known crystallized ligands into another protein structure that was not originally crystallized with. This procedure was performed with the ligands from 5FSA and 5V5Z structures and the protein structure of 5TZ1. The corresponding outcome yielded less accurate results, as shown in Figures 20 and 21. Both ligands were docked in an opposite orientation to their crystallized conformations. These results highlight that the docking process has its limitations, as when applied to non-reference small molecules, it may change orientations and some changes in the conformation may occur.

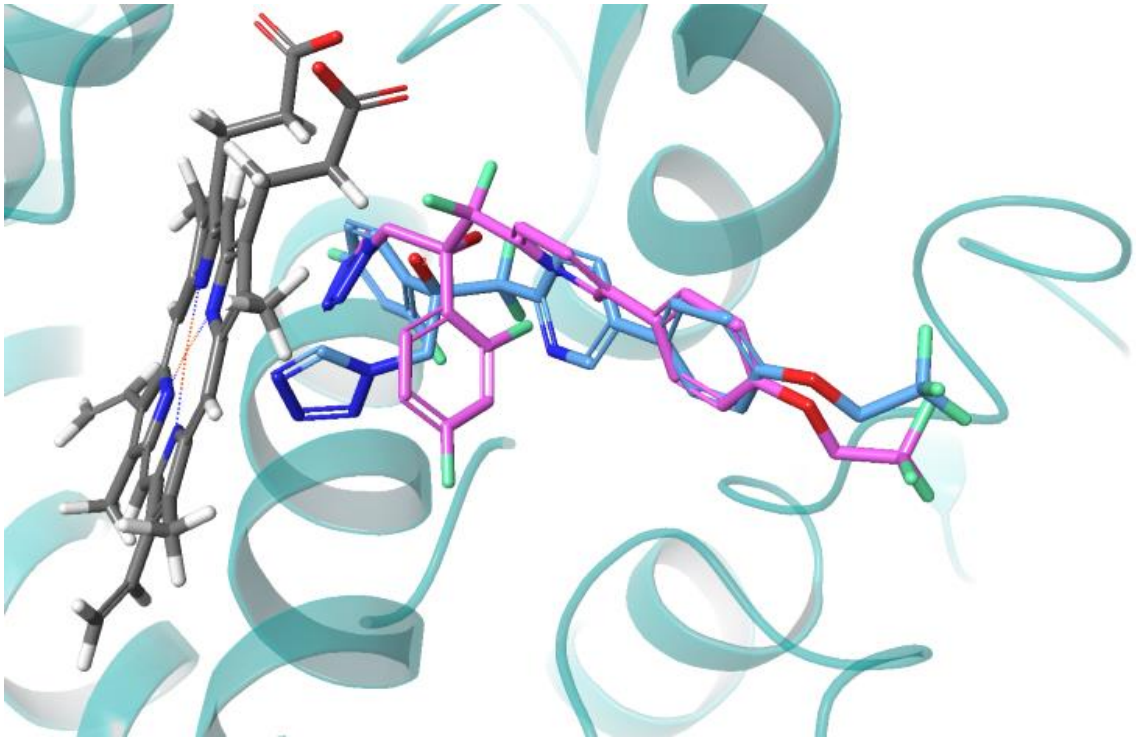


Figure 19. Superposition of the redocked and crystallized VT1 ligand in complex with lanosterol 14α-demethylase with PDB ID 5TZ1. The protein is shown in ribbon format and colored in blue. The heme prosthetic group is displayed in stick representation and colored by CPK. Both ligands are also shown as sticks and colored by CPK. The crystallized VT1 is highlighted with blue carbon backbone, while the redocked VT1 is shown in pink.

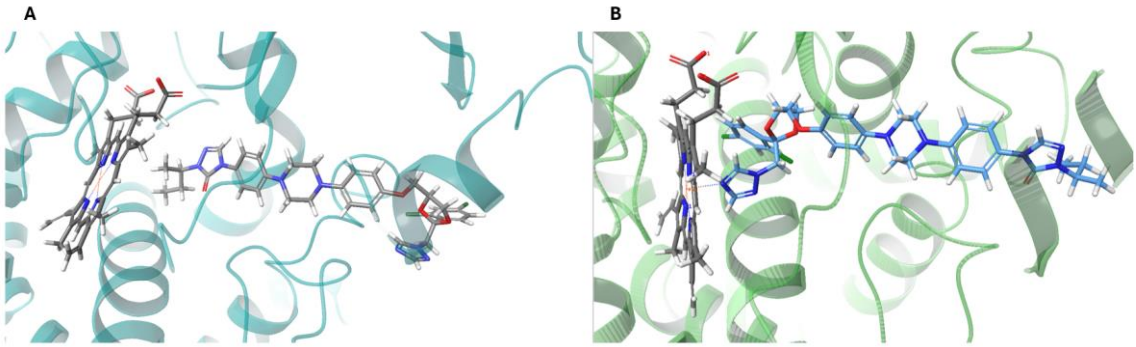


Figure 20. **(A)** Crossdocking of X2N ligand into the 5TZ1 protein structure. **(B)** 5FSA crystallized structure. Protein is shown in ribbon format while heme prosthetic group and ligand are shown in sticks and colored by CPK.

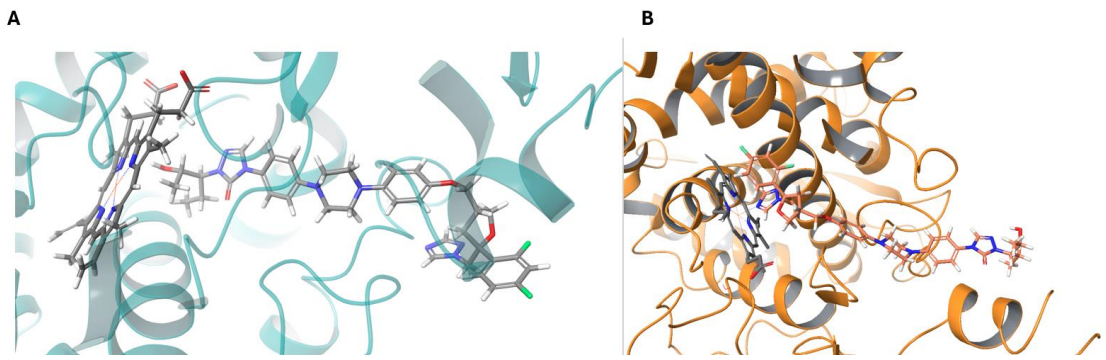


Figure 21. **(A)** Crossdocking of 1YN ligand into the 5TZ1 protein structure. **(B)** 5V5Z crystallized structure. Protein is shown in ribbon format while heme prosthetic group and ligand are shown in sticks and colored by CPK.

Although molecular docking is a widely used computational method for predicting the interactions between a protein and a ligand, its results are not always accurate, as previously discussed. One major limitation is the assumption of the protein rigidity. During a docking process, the protein is often considered to have a static structure, whereas in reality, proteins are dynamic and flexible within the biological environment. This assumption prevents an accurate simulation of protein conformational changes that take place while a ligand is binding. Another factor that could reduce the accuracy is the generation of ligand poses. The algorithm of the docking tries to predict the optimal binding conformation based on the torsion flexibility of the bonds in between atoms, rotations and translations from atoms. However, the predicted pose may have slight differences in comparison with the original one. These factors impact on the reliability of the prediction of docking pose of the ligand (Zhang et al., 2024).

After having the reference protein, the validation library and the grid prepared, molecular ligand-based docking was the next step. In this process, as shown in Figure 6, 1,025 molecules were docked. The EF of this step was 1.3249, meaning that this step performed approximately 1.3 times better than an aleatory selection because all active compounds were retrieved, and some decoys were lost during the process.

To further enrich the docked validation library, several previously designed pharmacophores models were tried out. These trials are summarized in Supplementary Table 1. Among the tested models, one pharmacophore was selected based on its performance, characteristics and the enrichment factor. As

explained earlier, the chosen pharmacophore model consisted of six features (2 rings that were required, 1 permitted hydrophobic or ring, and 1 ring and 2 hydrophobic permitted characteristics). This model was used in order to perform a screening of the docked validation library to filter those compounds that present at least 3 out of the 6 characteristics. As a result of this screening, a total of 62 compounds were retrieved, 6 of which were known actives. The EF that was obtained from this step was 6.61, indicating that the pharmacophore filter increased the likelihood of identifying active compounds in a collection of small molecules by approximately 6.6 times in comparison to a random selection.

5.2. Virtual screening

To identify possible inhibitor candidates for the protein target lanosterol 14α-demethylase, the same methodology was applied. The same protein with PDB ID 5TZ1 was prepared and the same grid was used to carry out the molecular docking process using the SPECS chemical compounds library. This dataset was previously prepared using OMEGA and LigPrep. Additionally, all compounds were filtered using ADME criteria to ensure their suitability for oral administration in humans, meaning that they would be able to overcome administration, absorption, distribution, metabolism and excretion barriers. Moreover, a molecular weight filter was also applied, retaining compounds between 300 and 730 Da, according to the molecular weight of known active compounds.

A total of 144,588 compounds were initially selected for the molecular docking, and 76,664 compounds were successfully docked. However, during this process, some unexpected issues with license and power outages were encountered. These incidents limited the number of compounds that were docked. Therefore, fewer compounds were docked than originally planned.

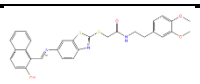
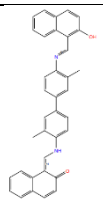
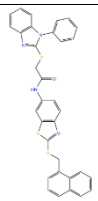
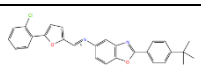
These docked results were filtered by the pharmacophore previously designed. Thus, a total of 8,051 compounds exhibited at least three of the six characteristics that were defined as pharmacophoric features. To refine the selection further, only those compounds matching five or more pharmacophoric features were retained for downstream analysis. A greater number of matching features would increase the likelihood of true inhibitory activity against lanosterol 14α-demethylase.

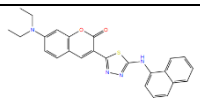
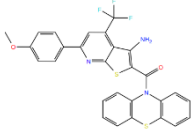
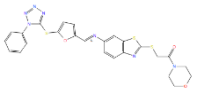
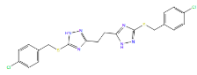
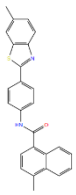
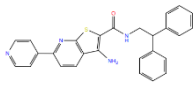
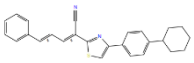
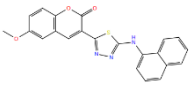
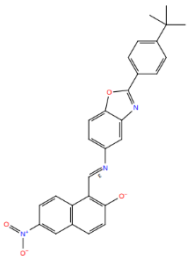
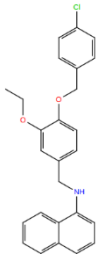
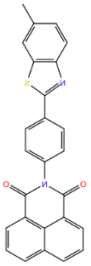
Therefore, a total of 106 compounds were selected as potential inhibitors of the target.

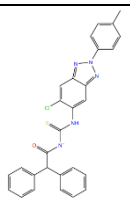
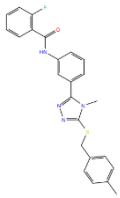
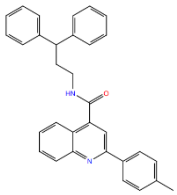
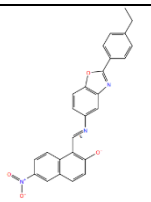
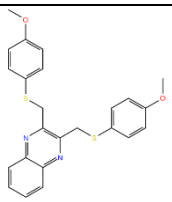
These 106 compounds were sorted by docking score, a score that represents the predicted binding affinity. Therefore, those with more negative scores suggest a higher likelihood of strong binding to the active site of lanosterol 14α-demethylase.

Based on the ranking, the selected candidates as potential inhibitors were the first 20 compounds, collected in Table 2. As shown in Figure 22, these molecules exhibited differences from each other, because dark blue boxes predominate in the heatmap of similarity. However, only one of them matched all six pharmacophoric features and displayed a docking score of -10.035 kcal/mol, suggesting that with a high likelihood, this compound will bind strongly to the active site. This molecule stands out as a promising inhibitor candidate and was further analyzed.

Table 2. Overview of the 20 selected molecules as potential inhibitors of lanosterol 14α-demethylase that matched 5 or 6 of the six pharmacophoric features and showed the most favorable docking score. Pharmacophoric features are ordered based on the order described on the pharmacophore previously designed in Figure 16 (H10, H11, R13, R14, R15 and R16). Features marked with (+) mean that they are fulfilled, while those indicated with (-) are not.

ID	2D structure	Docking score (kcal/mol)	Matches	Pharmacophoric features
78315		-11,547	5	H(+) H(-) R(+) R(+) R(+) R(+)
7885		-11.524	5	H(+) H(-) R(+) R(+) R(+) R(+)
48415		-11.445	5	H(+) H(-) R(+) R(+) R(+) R(+)
9019		-10.970	5	H(+) H(-) R(+) R(+) R(+) R(+)

40673		-10.836	5	H(+) H(-) R(+) R(+) R(+) R(+)
48535		-10.564	5	H(+) H(-) R(+) R(+) R(+) R(+)
76205		-10.554	5	H(+) H(-) R(+) R(+) R(+) R(+)
49680		-10.453	5	H(+) H(+) R(-) R(+) R(+) R(+)
81335		-10.451	5	H(+) H(-) R(+) R(+) R(+) R(+)
41475		-10.439	5	H(-) H(+) R(+) R(+) R(+) R(+)
81806		-10.379	5	H(+) H(+) R(-) H(+) R(+) R(+)
40669		-10.353	5	H(+) H(-) R(+) R(+) R(+) R(+)
85712		-10.338	5	H(+) H(-) R(+) R(+) R(+) R(+)
49295		-10.212	5	H(+) H(-) R(+) R(+) R(+) R(+)
6982		-10.158	5	H(+) H(-) R(+) R(+) R(+) R(+)

106312		-10.140	5	H(+) H(-) R(+) R(+) R(+) R(+)
52374		-10.107	5	H(+) H(+) R(-) R(+) R(+) R(+)
23419		-10.035	6	H(+) H(+) R(+) R(+) R(+) R(+)
72593		-10.014	5	H(+) H(-) R(+) R(+) R(+) R(+)
65155		-9.835	5	H(-) H(+) R(+) R(+) R(+) R(+)

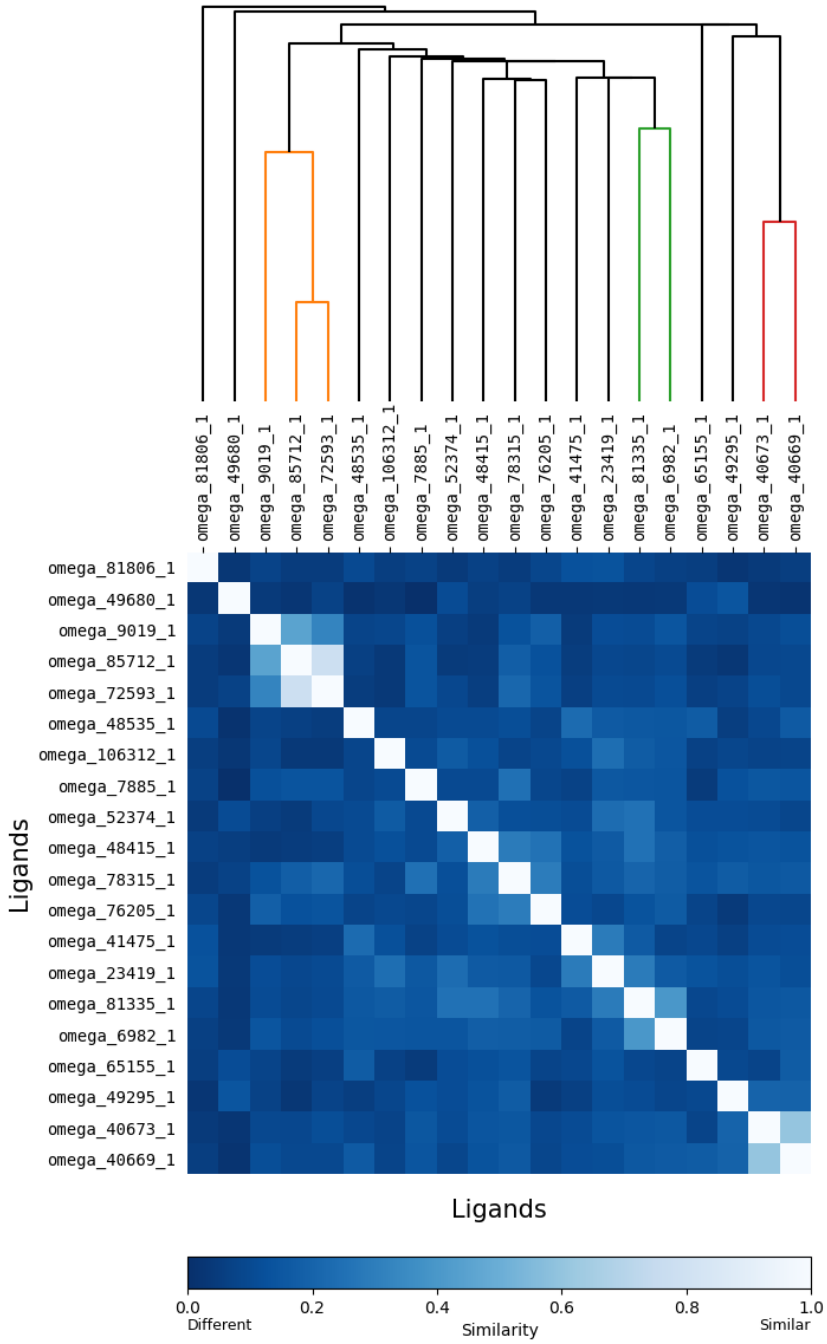


Figure 22. Similarity heatmap of potential candidates to inhibit lanosterol 14α-demethylase of *Candida albicans*. The darker the blue is, the more different the two compounds are, while the whiter the square is, the more similar they are. Compounds with a Tanimoto similarity ≥ 0.6 are grouped into distinct clusters, each one highlighted with different colors to visually distinguish the groups of structurally similar molecules.

When the selected compound was superposed to the pharmacophore model (Figure 23), it matched successfully all six characteristics. This alignment confirms that the small molecule met the structural requirements in order to

become biologically active to the target. Therefore, this compound could be a strong candidate to inhibit lanosterol 14α-demethylase.

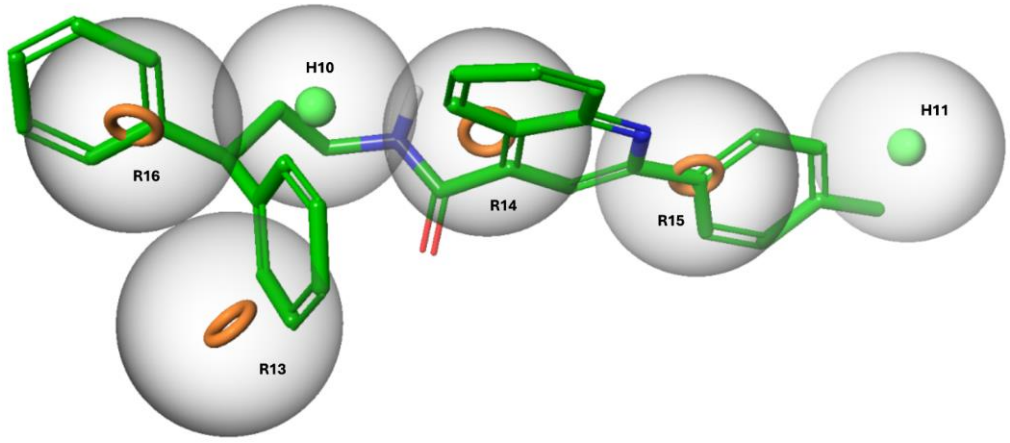


Figure 23. Superposition of the pharmacophore compound with ID 23419, which is one of the 20 compounds proposed as potential inhibitors to the target lanosterol 14α-demethylase that matches all six pharmacophoric features. Rings are represented in orange circles and hydrophobic characteristics are shown in green spheres. The grey spheres around each features illustrate the tolerance radius, set at 2.00 Å.

Moreover, when comparing the potential inhibitor with the reference ligand within the binding site of the target protein, interesting interactions were observed, as shown in Figure 24. Notice that these formed interactions are π-π (highlighted in blue), and they occur between aromatic rings from the ligand and the surrounding residues of the active-site. There are three types of π-π interactions: (a) edge-to-face stacked or T-shaped, (b) offset stacked, and (c) face to surface stacked (Rutledge et al., 2010; Zhuang et al., 2019). T-shaped and offset-stacked were observed in this study. Aromatic rings exhibit non-uniform charge distribution, with a positive electrostatic potential at the center and negative charge density around the edges. This is due to the delocalized π-electrons, which concentrate electron density on the periphery.

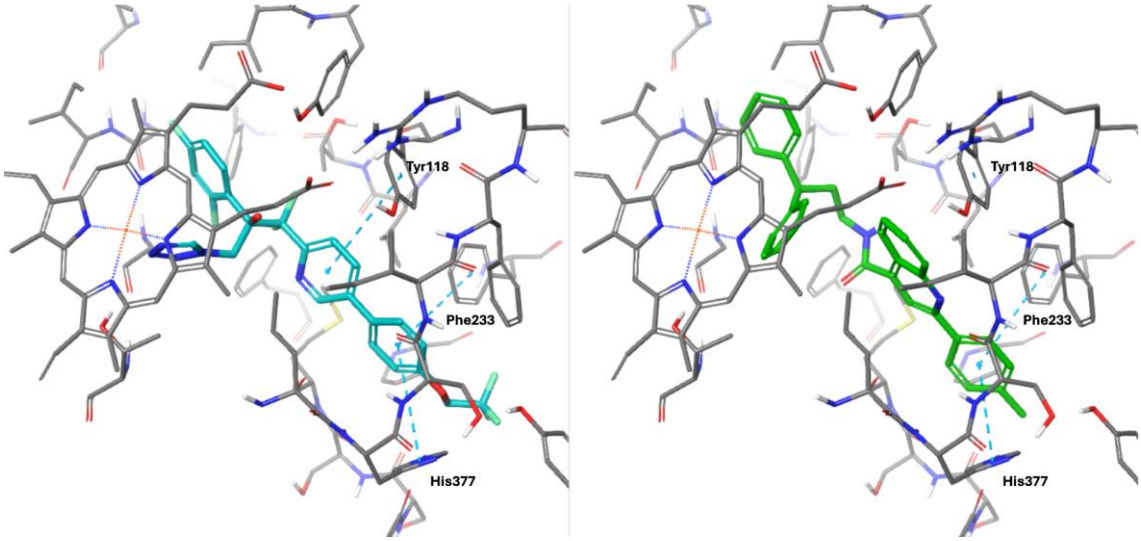


Figure 24. Schematic representations of the interactions between VT1 (left) and potential inhibitor with ID 23419 (right). The complex ligand-protein is represented in sticks and colored by CPK. Interactions π-π are represented with blue dotted lines among the residues Tyr118, Phe233 and His377

In both ligands, interactions between His377 and Phe233 appear to be offset stacked π-π interactions, where the negative edge of an aromatic ring aligns to a positive charged center of another aromatic ring, therefore enabling the electrostatic complementarity. These two rings are facing each other. However, only reference inhibitor (VT1 from the 5TZ1 crystal structure) exhibits a T-shape interaction, based on the position of the different aromatic rings of the different structures. In this type of interaction, the rings are not facing each other rather than they are in T-shape form, where the positive electrostatic surface from one ring aligns with the negative electrostatic surface from the edge of another ring.

However, it is important to note that protein-ligand complexes are dynamic rather than static structures. Therefore, although this T-shape π-π interaction is not observed that it is being formed in the docking pose of the potential candidate inhibitor, this does not mean its formation under physiological conditions. The two aromatic rings that could establish this interaction are spatially close, suggesting that conformational flexibility of the ligand or protein could enable a favorable π-π interaction to occur.

6. Conclusions

Due to the high incidence rate associated with candidiasis, there is a growing need to develop novel and more effective inhibitors. Therefore, in this study, a complete virtual screening was successfully implemented in order to identify potential inhibitors of lanosterol 14α-demethylase, using the Specs chemical compound library. Through the automatization of retrieving active compounds, a pharmacophore model was designed to capture essential inhibitory features. Molecular docking from the validation library (active and decoys) was performed and then filtered by the pharmacophore model to evaluate the enrichment factor of the step, achieving an EF above 6.6 in the validation set. Finally, the same strategy was applied to the Specs chemical compound library and after the docking process and the pharmacophoric screening, a set of 20 molecules with high predicted affinity for the binding site and good pharmacophoric alignment were identified as promising candidates to inhibit fungal growth and to treat the infection of candidiasis. The structurally diverse nature of the identified compounds should avoid the fungal resistance mechanisms reported for azole-based compounds. Therefore, the defined strategy demonstrated its potential to identify inhibitors of lanosterol 14α-demethylase in *Candida albicans*, as evidenced by the favorable EF obtained during the screening of the validation library.

Despite these promising results, this performed study has its limitations. The main one is that all gathered results were generated exclusively through computational predictions. Therefore, future work should validate these hit compounds trough *in vitro* assays to confirm their efficacy as inhibitors of lanosterol 14α-demethylase and their mechanism of action.

7. Acknowledgements

I would like first to express my sincere gratitude to Dr. Aleix Gimeno, not only for his supervision and guidance throughout this project, but also for all the unconditional support during all the process. Moreover, I would like to thank Dr. Gerard Pujadas and Dr. Santi Garcia-Vallvé for giving me the opportunity to join the Cheminformatics and Nutrition research group for the past few months.

Also, my heartfelt thanks to Ariadna Llop, the PhD student of the research group, for the constant support and assistance, particularly with the dendrogram script. I would like to thank Said, as well, for allowing me to use his CORAL-PIC script, which was useful to understand the properties of the protein binding site.

Last but not least, I am deeply grateful to the rest of the members of the group for their unconditional support, help and encouragement. I could not have done this without you.

8. References

Bauer, M. R., & Mackey, M. D. (2019a). Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein-Ligand Complexes. *Journal of Medicinal Chemistry*, *62*(6), 3036–3050. https://doi.org/10.1021/ACS.JMEDCHEM.8B01925,

Bauer, M. R., & Mackey, M. D. (2019b). Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein-Ligand Complexes. *Journal of Medicinal Chemistry*, *62*(6), 3036–3050. https://doi.org/10.1021/ACS.JMEDCHEM.8B01925/SUPPL_FILE/JM8B01925_SI_002.ZIP

Binjubair, F. A., Parker, J. E., Warrilow, A. G., Puri, K., Braidley, P. J., Tatar, E., Kelly, S. L., Kelly, D. E., & Simons, C. (2020). Small-Molecule Inhibitors Targeting Sterol 14α-Demethylase (CYP51): Synthesis, Molecular Modelling and Evaluation Against Candida albicans. *ChemMedChem*, *15*(14), 1294–1309. https://doi.org/10.1002/CMDC.202000250,

Cheeseright, T., Mackey, M., Rose, S., & Vinter, A. (2006). 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,

chembl-webresource-client · PyPI. (n.d.). Retrieved June 4, 2025, from https://pypi.org/project/chembl-webresource-client/

Cornely, O. A., Sprute, R., Grothe, J. H., Koehler, P., Meis, J. F., Reinhold, I., Seidel, D., Mellinghoff, S. C., Rahimli, L., Rahn, S., Salmanton-García, J., Stemler, J., Toebben, C., Cornely, O. A., Sprute, R., Bassetti, M., C-A Chen, S., Groll, A. H., Kurzai, O., … Zhu, L.-P. (2025). Global guideline for the diagnosis and management of candidiasis: an initiative of the ECMM in cooperation with ISHAM and ASM. *The Lancet Infectious Diseases*, *25*(5), e280–e293. https://doi.org/10.1016/S1473-3099(24)00749-7

Cresset®. (n.d.-a). *Activity Atlas™* (Flare, V10.0).

Cresset®. (n.d.-b). *Activity Miner™* (Flare, V10.0).

Cresset®. (n.d.-c). *Flare™* (Flare, V10.0).

David, H., & Solomon, A. P. (2023). Molecular association of Candida albicans and vulvovaginal candidiasis: focusing on a solution. *Frontiers in Cellular and Infection Microbiology*, *13*, 1245808. https://doi.org/10.3389/FCIMB.2023.1245808

D'Enfert, C., Kaune, A. K., Alaban, L. R., Chakraborty, S., Cole, N., Delavy, M., Kosmala, D., Marsaux, B., Fróis-Martins, R., Morelli, M., Rosati, D., Valentine, M., Xie, Z., Emritloll, Y., Warn, P. A., Bequet, F., Bougnoux, M. E., Bornes, S., Gresnigt, M. S., … Brown, A. J. P. (2021). The impact of the fungus-host-microbiota interplay upon Candida albicans

infections: Current knowledge and new perspectives. *FEMS Microbiology Reviews*, *45*(3). https://doi.org/10.1093/FEMSRE/FUAA060,

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,

Dixon, S. L., Smondyrev, A. M., & Rao, S. N. (2006). PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chemical Biology and Drug Design*, *67*(5), 370–372. https://doi.org/10.1111/J.1747-0285.2006.00384.X,

El-Sayed, A., Aleya, L., & Kamel, M. (2021). Microbiota's role in health and diseases. *Environmental Science and Pollution Research International*, *28*(28), 36967. https://doi.org/10.1007/S11356-021-14593-Z

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,

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/JM051256O,

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? *International Journal of Molecular Sciences*, *20*(6). https://doi.org/10.3390/IJMS20061375,

Gori, D. N. P., Alberca, L. N., Rodriguez, S., Alice, J. I., Llanos, M. A., Bellera, C. L., & Talevi, A. (2022). LIDeB Tools: A Latin American resource of freely available, open-source cheminformatics apps. *Artificial Intelligence in the Life Sciences*, *2*, 100049. https://doi.org/10.1016/J.AILSCI.2022.100049

Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., & Banks, J. L. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*, *47*(7), 1750–1759. https://doi.org/10.1021/JM030644S,

Hargrove, T. Y., Friggeri, L., Wawrzak, Z., Qi, A., Hoekstra, W. J., Schotzinger, R. J., York, J. D., Peter Guengerich, F., & Lepesheva, G. I. (2017). Structural analyses of Candida albicans

sterol 14α-demethylase complexed with azole drugs address the molecular basis of azole-mediated inhibition of fungal sterol biosynthesis. *Journal of Biological Chemistry*, *292*(16), 6728–6743. https://doi.org/10.1074/JBC.M117.778308/ASSET/AF77DE73-940D-4CB9-AEDC-D935434E9C69/MAIN.ASSETS/GR14.JPG

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)*, *9*, 90–95.

Iebba, V., Totino, V., Gagliardi, A., Santangelo, F., Cacciotti, F., Trancassini, M., Mancini, C., Cicerone, C., Corazziari, E., Pantanella, F., & Schippa, S. (2016). Eubiosis and dysbiosis: the two sides of the microbiota. *The New Microbiologica*, *39*(1), 1–12.

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,

Katsipoulaki, M., Stappers, M. H. T., Malavia-Jones, D., Brunke, S., Hube, B., & Gow, N. A. R. (2024). Candida albicans and Candida glabrata : global priority pathogens . *Microbiology and Molecular Biology Reviews*, *88*(2). https://doi.org/10.1128/MMBR.00021-23/ASSET/D2C45409-5CC8-49D3-8090-870E8A52B91E/ASSETS/IMAGES/LARGE/MMBR.00021-23.F004.JPG

Kuhn, M., Firth-Clark, S., Tosco, P., Mey, A. S. J. S., MacKey, M., & Michel, J. (2020). Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *Journal of Chemical Information and Modeling*, *60*(6), 3120–3130. https://doi.org/10.1021/ACS.JCIM.0C00165,

Lee, Y., Puumala, E., Robbins, N., & Cowen, L. E. (2020). Antifungal Drug Resistance: Molecular Mechanisms in Candida albicans and Beyond. *Chemical Reviews*, *121*(6), 3390. https://doi.org/10.1021/ACS.CHEMREV.0C00199

Liu, J.-F., Xia, J.-J., Nie, K.-L., Wang, F., & Deng, L. (2019). Outline of the biosynthesis and regulation of ergosterol in yeast. *World Journal of Microbiology and Biotechnology*, *35*(7), 98. https://doi.org/10.1007/s11274-019-2673-2

Lopes, J. P., & Lionakis, M. S. (2022). Pathogenesis and virulence of Candida albicans. *Virulence*, *13*(1), 89–121. https://doi.org/10.1080/21505594.2021.2019950,

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,

Mayer, F. L., Wilson, D., & Hube, B. (2013). Candida albicans pathogenicity mechanisms. *Virulence*, *4*(2), 119–128. https://doi.org/10.4161/viru.22913

McCarty, T. P., White, C. M., & Pappas, P. G. (2021). Candidemia and Invasive Candidiasis. *Infectious Disease Clinics of North America*, *35*(2), 389–413. https://doi.org/10.1016/J.IDC.2021.03.007

Microbiota - MeSH - NCBI. (n.d.). Retrieved June 4, 2025, from https://www.ncbi.nlm.nih.gov/mesh/68064307

Monk, B. C., Sagatova, A. A., Hosseini, P., Ruma, Y. N., Wilson, R. K., & Keniya, M. V. (2020). Fungal Lanosterol 14α-demethylase: A target for next-generation antifungal design. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1868*(3), 140206. https://doi.org/10.1016/J.BBAPAP.2019.02.008

Mycobank. (n.d.). Retrieved June 4, 2025, from https://www.mycobank.org/

OpenEye, C. M. Sciences. http://www. eyesopen. com. (n.d.). *OMEGA 6.0.0.1.* .

Pappas, P. G., Kauffman, C. A., Andes, D. R., Clancy, C. J., Marr, K. A., Ostrosky-Zeichner, L., Reboli, A. C., Schuster, M. G., Vazquez, J. A., Walsh, T. J., Zaoutis, T. E., & Sobel, J. D. (2015). Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *62*(4), e1. https://doi.org/10.1093/CID/CIV933

RDKit: Open-source cheminformatics. https://www.rdkit.org. (n.d.).

Rodrigues, M. L. (2018). The multifunctional fungal ergosterol. *MBio*, *9*(5). https://doi.org/10.1128/MBIO.01755-18,

Rutledge, L. R., Churchill, C. D. M., & Wetmore, S. D. (2010). A preliminary investigation of the additivity of π-π or π+-π stacking and T-Shaped interactions between natural or damaged DNA nucleobases and histidine. *Journal of Physical Chemistry B*, *114*(9), 3355–3367. https://doi.org/10.1021/JP911990G,

Schrödinger, L. (n.d.-a). *Schrödinger Release 2025-2: Glide* (2025–2).

Schrödinger, L. (n.d.-b). *Schrödinger Release 2025-2: LigPrep* (New York, NY).

Schrödinger, L. (n.d.-c). *Schrödinger Release 2025-2: Maestro*.

Schrödinger, L. (n.d.-d). *Schrödinger Release 2025-2: Phase*.

Schuster, M., Kilaru, S., & Steinberg, G. (2024). Azoles activate type I and type II programmed cell death pathways in crop pathogenic fungi. *Nature Communications 2024 15:1*, *15*(1), 1–21. https://doi.org/10.1038/s41467-024-48157-9

Soriano, A., Honore, P. M., Puerta-Alcalde, P., Garcia-Vidal, C., Pagotto, A., Gonçalves-Bradley, D. C., & Verweij, P. E. (2023). Invasive candidiasis: current clinical challenges and unmet needs in adult populations. *Journal of Antimicrobial Chemotherapy*, *78*(7), 1569. https://doi.org/10.1093/JAC/DKAD139

WHO fungal priority pathogens list to guide research, development and public health action. (n.d.). Retrieved May 24, 2025, from https://www.who.int/publications/i/item/9789240060241

Yang, Y., Yao, K., Repasky, M. P., Leswing, K., Abel, R., Shoichet, B. K., & Jerome, S. V. (2021). Efficient Exploration of Chemical Space with Docking and Deep Learning. *Journal of Chemical Theory and Computation*, *17*(11), 7106–7119. https://doi.org/10.1021/ACS.JCTC.1C00810,

Yates, C. M., Garvey, E. P., Shaver, S. R., Schotzinger, R. J., & Hoekstra, W. J. (2017). Design and optimization of highly-selective, broad spectrum fungal CYP51 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, *27*(15), 3243–3248. https://doi.org/10.1016/J.BMCL.2017.06.037

Yu, X., Cojocaru, V., Mustafa, G., Salo-Ahen, O. M. H., Lepesheva, G. I., & Wade, R. C. (2015). Dynamics of CYP51: implications for function and inhibitor design. *Journal of Molecular Recognition*, *28*(2), 59–73. https://doi.org/10.1002/jmr.2412

Zhang, X., Shen, C., Zhang, H., Kang, Y., Hsieh, C. Y., & Hou, T. (2024). Advancing Ligand Docking through Deep Learning: Challenges and Prospects in Virtual Screening. *Accounts of Chemical Research*, *57*(10), 1500–1509. https://doi.org/10.1021/ACS.ACCOUNTS.4C00093,

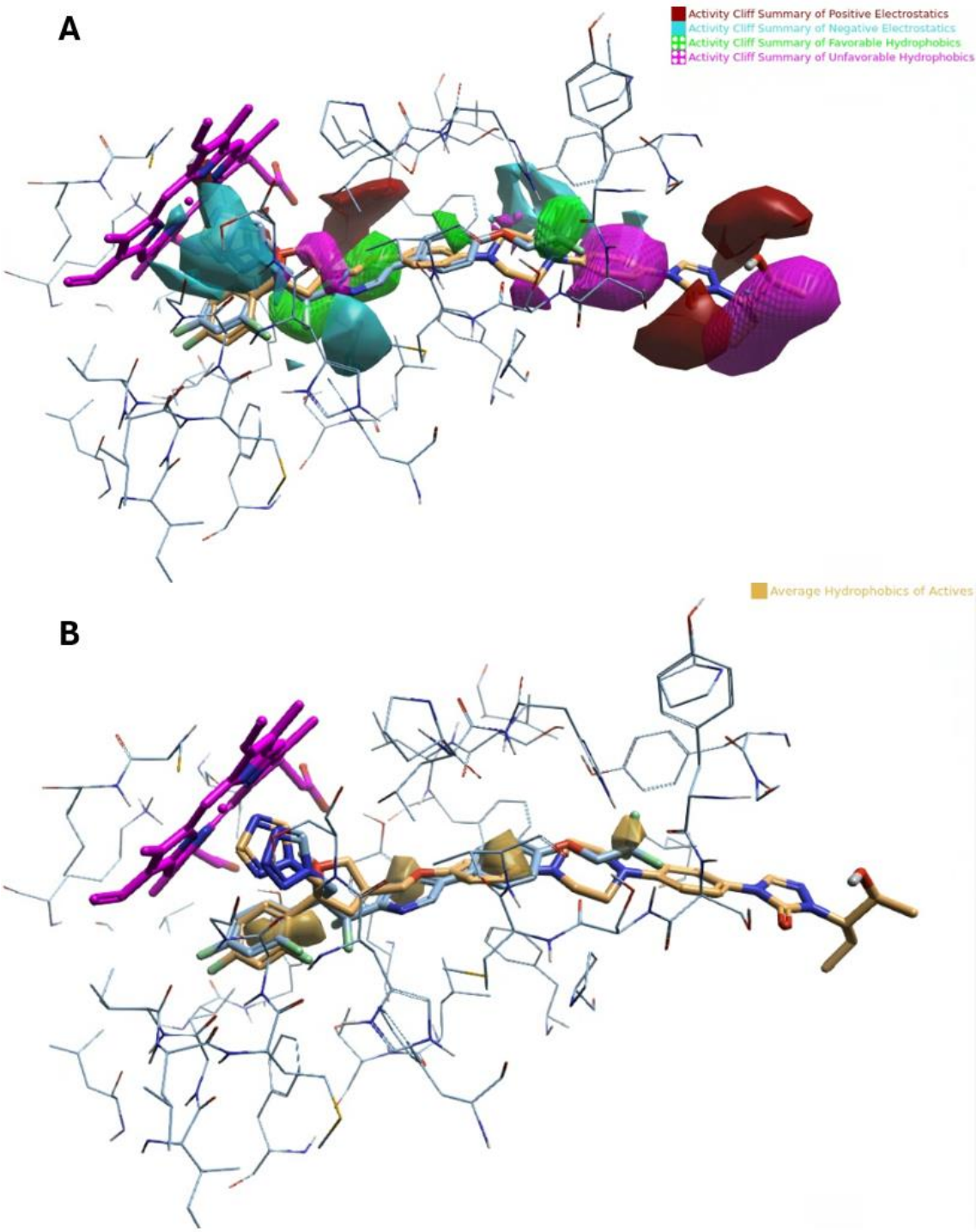
Zhao, M., Ma, J., Li, M., Zhang, Y., Jiang, B., Zhao, X., Huai, C., Shen, L., Zhang, N., He, L., & Qin, S. (2021). Cytochrome p450 enzymes and drug metabolism in humans. *International Journal of Molecular Sciences*, *22*(23). https://doi.org/10.3390/IJMS222312808,

Zhuang, W. R., Wang, Y., Cui, P. F., Xing, L., Lee, J., Kim, D., Jiang, H. L., & Oh, Y. K. (2019). Applications of π-π stacking interactions in the design of drug-delivery systems. *Journal of Controlled Release*, *294*, 311–326. https://doi.org/10.1016/j.jconrel.2018.12.014

OpenAI. (2025). ChatGPT (GPT-4 model). https://chat.openai.com/

Google Gemini. (2023). *Gemini* (Large language model). Google.

9. Supplementary information



Supplementary Figure 1. **(A)** Activity Cliff results from comparing ligands VT1 (shown in sticks, colored by CPK and a blue backbone) and X2N (shown in sticks, colored by CPK and a light orange backbone). Heme prosthetic group is also shown in sticks colored by CPK and with a pink backbone. The model highlights regions the positive (red surfaces), negative electrostatics (blue surfaces), and favorable (green surfaces) and unfavorable (pink surfaces) hydrophobic regions. **(B)** Summary of regions where active compounds commonly exhibited hydrophobic characteristics (yellow surfaces).

	CHEMBL5180750	CHEMBL5094280	CHEMBL4757634	CHEMBL4764357	CHEMBL5424058	CHEMBL5196919	CHEMBL5219013	CHEMBL5219227	CHEMBL5206822	CHEMBL5088292	CHEMBL5220891	CHEMBL75	CHEMBL1236677	CHEMBL5432092	CHEMBL106	CHEMBL5411746	CHEMBL5411444	CHEMBL5403026	CHEMBL5411858	CHEMBL5187700	CHEMBL5195879	CHEMBL5197095
CHEMBL5180750	0	0,10	-0,50	-1,20	-0,80	-2,00	0,60	0,70	0,40	0,50	0,00	0,90	0,30	0,60	0,90	-0,10	0,10	0,20	0,50	1,20	0,70	0,40
CHEMBL5094280	-0,10	0	-0,80	-1,50	-1,20	-2,40	0,50	0,80	0,30	0,40	-0,10	0,80	0,20	0,60	0,70	-0,20	0,00	0,20	0,40	1,20	0,70	0,40
CHEMBL4757634	0,50	0,80	0	-0,80	-0,50	-1,50	1,00	1,20	0,70	0,90	0,40	1,20	0,70	1,00	1,20	0,20	0,50	0,60	0,80	1,60	1,00	0,80
CHEMBL4764357	1,20	1,50	0,80	0	0,20	-0,80	1,50	1,60	1,30	1,70	1,20	1,80	1,20	1,40	1,80	0,80	1,00	1,20	1,40	2,10	1,70	1,40
CHEMBL5424058	0,80	1,20	0,50	-0,20	0	-1,00	1,20	1,50	1,20	1,10	0,80	1,60	1,20	1,40	1,30	0,70	0,80	1,00	1,20	1,90	1,50	1,10
CHEMBL5196919	2,00	2,40	1,50	0,80	1,00	0	2,20	2,40	2,10	2,20	2,00	2,50	1,80	2,70	2,40	1,50	1,80	1,80	2,10	2,60	2,20	1,90
CHEMBL5219013	-0,60	-0,50	-1,00	-1,50	-1,20	-2,20	0	0,20	-0,10	-0,10	-0,50	0,30	-0,20	0,10	0,30	-0,60	-0,40	-0,30	-0,10	0,70	0,20	0,00
CHEMBL5219227	-0,70	-0,80	-1,20	-1,60	-1,50	-2,40	-0,20	0	-0,30	-0,20	-0,70	0,20	-0,50	-0,10	0,10	-0,80	-0,60	-0,40	-0,20	0,50	0,00	-0,20
CHEMBL5206822	-0,40	-0,30	-0,70	-1,30	-1,20	-2,10	0,10	0,30	0	0,10	-0,40	0,40	-0,10	0,20	0,40	-0,40	-0,30	-0,10	0,00	0,80	0,30	0,10
CHEMBL5088292	-0,50	-0,40	-0,90	-1,70	-1,10	-2,20	0,10	0,20	-0,10	0	-0,50	0,40	-0,10	0,20	0,40	-0,50	-0,30	-0,20	0,00	0,80	0,30	0,00
CHEMBL5220891	0,00	0,10	-0,40	-1,20	-0,80	-2,00	0,50	0,70	0,40	0,50	0	1,10	0,20	0,60	0,80	-0,10	0,10	0,20	0,40	1,30	0,70	0,50
CHEMBL75	-0,90	-0,80	-1,20	-1,80	-1,60	-2,50	-0,30	-0,20	-0,40	-0,40	-1,10	0	-0,50	-0,20	0,00	-1,00	-0,70	-0,60	-0,40	0,40	-0,10	-0,30
CHEMBL1236677	-0,30	-0,20	-0,70	-1,20	-1,20	-1,80	0,20	0,50	0,10	0,10	-0,20	0,50	0	0,30	0,50	-0,40	-0,20	0,00	0,20	0,70	0,30	0,20
CHEMBL5432092	-0,60	-0,60	-1,00	-1,40	-1,40	-2,70	-0,10	0,10	-0,20	-0,20	-0,60	0,20	-0,30	0	0,10	-0,70	-0,50	-0,40	-0,10	0,60	0,10	-0,10
CHEMBL106	-0,90	-0,70	-1,20	-1,80	-1,30	-2,40	-0,30	-0,10	-0,40	-0,40	-0,80	0,00	-0,50	-0,10	0	-0,80	-0,60	-0,40	-0,30	0,40	-0,10	-0,30
CHEMBL5411746	0,10	0,20	-0,20	-0,80	-0,70	-1,50	0,60	0,80	0,40	0,50	0,10	1,00	0,40	0,70	0,80	0	0,30	0,40	0,60	1,20	0,70	0,60
CHEMBL5411444	-0,10	0,00	-0,50	-1,00	-0,80	-1,80	0,40	0,60	0,30	0,30	-0,10	0,70	0,20	0,50	0,60	-0,30	0	0,20	0,40	1,20	0,50	0,40
CHEMBL5403026	-0,20	-0,20	-0,60	-1,20	-1,00	-1,80	0,30	0,40	0,10	0,20	-0,20	0,60	0,00	0,40	0,40	-0,40	-0,20	0	0,20	0,80	0,40	0,20
CHEMBL5411858	-0,50	-0,40	-0,80	-1,40	-1,20	-2,10	0,10	0,20	0,00	0,00	-0,40	0,40	-0,20	0,10	0,30	-0,60	-0,40	-0,20	0	0,70	0,20	0,00
CHEMBL5187700	-1,20	-1,20	-1,60	-2,10	-1,90	-2,60	-0,70	-0,50	-0,80	-0,80	-1,30	-0,40	-0,70	-0,60	-0,40	-1,20	-1,20	-0,80	-0,70	0	-0,60	-0,80
CHEMBL5195879	-0,70	-0,70	-1,00	-1,70	-1,50	-2,20	-0,20	0,00	-0,30	-0,30	-0,70	0,10	-0,30	-0,10	0,10	-0,70	-0,50	-0,40	-0,20	0,60	0	-0,20
CHEMBL5197095	-0,40	-0,40	-0,80	-1,40	-1,10	-1,90	0,00	0,20	-0,10	0,00	-0,50	0,30	-0,20	0,10	0,30	-0,60	-0,40	-0,20	0,00	0,80	0,20	0

Supplementary Figure 2. Disparity matrix of active compounds with reported IC_{50} values. The matrix, generated in Flare, and display the disparity values both in positive and negative. For this reason, values should be considered as absolute values. Color intensity reflects the degree of disparity: red and blue boxes indicate high disparity, while white boxes represent low disparity. The darker the color, the greater disparity between the pair of compounds.

Supplementary Table 1. Set of different pharmacophoric models that performed better results. Capital letters in the pharmacophoric model correspond to required characteristics, while lower case letters correspond to permitted features. Note: first entry corresponds to the chosen model for this study, and the other 10 entries correspond to the models of pharmacophores that presented a good outcome.

Pharmacophoric model	Tolerance ratio	Number of matches	Retrieved actives	Total retrieved	EF
h_{10} h_{11} r_{13} r_{14} R_{15} R_{16}	2.00 Å	3 out of 6	6	62	6.61
h_{10} R_{13} r_{14} R_{15} R_{16}	2.00 Å.	4 out of 5	1	5	13.67
h_{11} r_{14} R_{15} R_{16}	2.00 Å	3 out of 4	2	15	9.11
h_{10} h_{11} r_{14} R_{15} R_{16}	h_{10} and h_{11} = 3.0 Å r_{14} R_{15} R_{16}= 2.00 Å	5 out of 5	1	9	7.59
h_{10} r_{14} R_{15} R_{16}	h_{10} = 2.50 Å r_{14} R_{15} R_{16} = 2.00 Å	3 out of 4	2	22	6.21
h_{10} h_{11} r_{14} R_{15} R_{16}	h_{10} and h_{11} = 3.50 Å r_{14} R_{15} R_{16} = 2.00 Å	5 out of 5	1	12	5.69
H_{10} h_{11} r_{14} R_{15} R_{16}	H_{10} and h_{11} = 3.0 Å r_{14} R_{15} R_{16} = 2.00 Å	4 out of 5	2	26	5.26
r_{14} R_{15} R_{16}	2.00 Å	2 out of 3	6	91	4.51
H_{10} r_{14} R_{15} R_{16}	H_{10} = 3.00 Å r_{14} R_{15} R_{16}= 2.00 Å	3 out of 4	2	34	4.02
r_{14} R_{15} R_{16}	2.50 Å	2 out of 3	7	136	3.52
r_{14} R_{15} R_{16}	4.00 Å	3 out of 3	10	209	3.27