UNIVERSITAT ROVIRA i VIRGILI

VIRTUAL SCREENING FOR NOVEL ANTICANCER AGENTS TARGETING THE PDZ2 DOMAIN OF SYNTENIN-1

Guillem Arasa Estivill

FINAL DEGREE PROJECT IN BIOTECHNOLOGY

Academic supervisor/tutor:	Dr. Gerard Pujadas Anguiano Bachelor's degree in Biotechnology Biochemistry and Biotechnology Department gerard.pujadas@urv.cat
In cooperation with:	Cheminformatics and Nutrition research group Universitat Rovira i Virgili
Supervisor/s:	Dr. Santiago Garcia-Vallvé Bachelor's degree in Biotechnology Biochemistry and Biotechnology Department santi.garcia-vallve@urv.cat

June 2023

Jo, Guillem Arasa Estivill, amb DNI 39926939-C, soc coneixedor de la guia de prevenció del plagi a la URV Prevenció, detecció i tractament del plagi en la docència: guia per a estudiants (aprovada el juliol 2017) (http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/) i afirmo que aquet TFG no constitueixen cap de les conductes considerades com a plagi per la URV.

Tarragona, 7 de juny de 2023.

ARASA ESTIVILL GUILLEMN - 39926939C

Firmado digitalmente por ARASA ESTIVILL GUILLEMN - 39926939C
Fecha: 2023.06.07 11:41:20 +02'00'

Index

1. Center information

This study was developed in the Cheminformatics and Nutrition research group at Universitat Rovira i Virgili. This group is led by the lecturers Dr. Gerard Pujadas and Dr. Santiago Garcia from the Biotechnology Department and its location is at C/ Marcel·lí Domingo,1 in Campus Sescelades.

The Cheminformatics and Nutrition research group focuses on the application of cheminformatics for the discovery of new drugs (using virtual screening with protein-ligand docking, pharmacophores or electrostatic/shape comparisons). The compounds are sourced from chemical compounds databases (either natural or synthetic) and may be incorporated as bioactive ingredients or drugs (depending on their origin).

The most recent studies of the research group have been focused on the discovery of novel inhibitors for SARS-CoV-2 with great interest on the main protease (M-pro) and the papain-like protease (PL-pro) of this virus. Although this study is aimed to the novel anticancer drug discovery, the workflow relies on the previous methodologies developed during the COVID-19 studies.

Guillem Arasa Estivill
Universitat Rovira i Virgili

2. Abstract

Syntenin-1, also known as mda-9 or syndecan binding protein, is an overexpressed protein found in advanced and aggressive forms of certain cancers with metastatic activity. It is a 298 amino acids length protein and its two main domains, PDZ1 and PDZ2, are responsible for the protein-protein interactions with peptides such as syndecan, neurexin or c-Src. In this study, a virtual screening workflow was used in order to find potential inhibitory molecules that could bind to the PDZ2 binding site. Initially, the 3D structure of different PDZ2-drug complexes had been analyzed and the main intermolecular interactions at the binding site were determined, considering Val209, Gly210 and Phe211 as the minimal hydrogen bonds required for the ligand activity. After this step, the PDB syntenin-1 structures were analyzed and ranked after a crossdocking and a redocking of the ligands, considering 8AAO structure the most appropriate for protein-ligand docking. After, this structure was used to dock the Specs library containing more than 170.000 synthetic compounds with proper ADMET properties and no PAINS features. Finally, all the ligands that could perform hydrogen bonds with Val209, Gly210 and Phe211 had been filtered with a pharmacophore model based on the 8AAP ligand, considered the most potent inhibitor to date of syntenin-1. This experiment resulted in a subset of 6 molecules that may be potential inhibitors of this target and, therefore, new anticancer drugs.

Keywords: syntenin-1, mda-9, cancer, PDZ2, virtual screening, protein-ligand docking.

3. Abbreviations

ADMET: Administration, Distribution, Metabolism, Excretion, Toxicity

LBVS: Ligand-Based Virtual Screening

MD: Molecular Docking

mda-9: melanoma differentiation associated gene-9.

PDB: Protein Data Bank

PDZ: Post-synaptic density protein PSD95/SAP90, Drosophila tumor suppressor DLGA, tight junction protein ZO-1 (Zonula Occludens 1).

PPI: Protein-Protein Interactions

RMSD: Root Mean Square Deviation

SBVS: Structure-Based Virtual Screening

SDCBP: syndecan binding protein

VS: Virtual Screening

4. Introduction

Cancer, as the second leading cause of natural mortality in Spain between 2021 and 2022 (Instituto Nacional de Estadística, 2022), is currently one of the main focuses of biomedical research alongside COVID-19. Among the population it is very close to mortality caused by diseases of the circulatory system; in men it is the first cause (295,5 deaths per 100.000 inhabitants) and in women, the second (189,9 deaths per 100.000 inhabitants). Cancer is a very generic term that encompasses a wide group of multifactorial and multigenic diseases in which a group of cells grow in an uncontrolled manner and have the capacity to develop metastatic activity both in nearby and distant organs or tissues. Key features include uncontrolled cell growth, evasion of pro-apoptotic signals and evasion of the immune system, among others (*What Is Cancer?*, 2021). Malignant cells accumulate multiple changes in the genome, which may include mutations, deletions, or chromosomal duplications. One of the most identifiable characteristics of these cells lies in the loss of the proteomic regulation in comparison with normal cells, as both gene overexpression and silencing can be found. This dysregulation can be further applied to the development of therapeutic techniques or identification of biomarkers (Kwon et al., 2021).

4.1 Syntenin-1, from gene to protein

Among all the overexpressed molecules in cancer cells syntenin-1 can be found in many versions of aggressive and invasive tumors such as melanomas, gliomas and head and neck squamous cells cancer with chemoresistance, stemness and metastatic activity (Cui et al., 2016; Kegelman et al. 2015; Mir et al., 2021; Das et al., 2012).

Syntenin-1, also known as melanoma differentiation associated gene-9 (*mda-9*), syndecan binding protein (SDCBP) or pro-TGF-alpha cytoplasmatic domain-interacting protein, is a homodimeric protein encoded by the *mda-9* gene located at 8q12.1. It is composed of 29.965 pb in 9 exons (Philley et al., 2016) that transcribe a 2,1kb length cDNA. The encoded protein consists of 298 amino acids with 33 kDA weight and is a member of the PDZ domain protein family. These domains – named after the post-synaptic density protein PSD95/SAP90, *Drosophila* tumor suppressor DLGA, tight junction protein ZO-1 (zonula

occludens 1)) – are globular regions of approximately 80 to 100 amino acids and are the most abundant domain class with protein-protein interactions (also referred as PPI) in the human proteome (Christensen et al., 2019). This characteristic is due to the PDZ ability to assemble with C-terminal regions of certain proteins or complexes, which is related to their high conservation among vertebrate species. They have been classically divided into three groups depending on the C terminal region of the peptides to which they are attached: I (-S/T-X-Φ), II (Φ-X-Φ) and III (D/E-X-Φ) – (where T is threonine, S is serine, D is aspartic acid, E is glutamic acid, Φ is an hydrophobic residue, X is any amino acid) (Kegelman et al., 2015; Christensen et al., 2019). Despite that this simple classification is currently being on revision due to the discovery of alternative and non-canonical bindings with their target, syntenin-1 is still considered interacting with class I and II peptides and other minor groups of lower affinity (Christensen et al., 2019). These PDZ domains are composed of six beta sheets and two alpha helices that form the groove where the peptides are bound which are located antiparallel to the β2 sheet (Shimada et al., 2019).

Syntenin-1 is composed of two PDZ domains (referred as PDZ1 and PDZ2) connected by a short flexible linker of 4 amino acids (Arg193-Pro194-Phe195-Glu196) and are flanked with an N-terminal region (consisting of 112 amino acids) and a C-terminal region (of 24 amino acids). The first encoded region, which mainly interacts with Ub to give rise to a transmembrane cargo Ub-dependent and can recruit the transcription factor Sox4, it is autoinhibited by phosphorylation; on the contrary the second one is essential for the correct assembly of the dimer and its localization (Choi et al., 2016; Kegelman et al., 2015) (Figure 1).

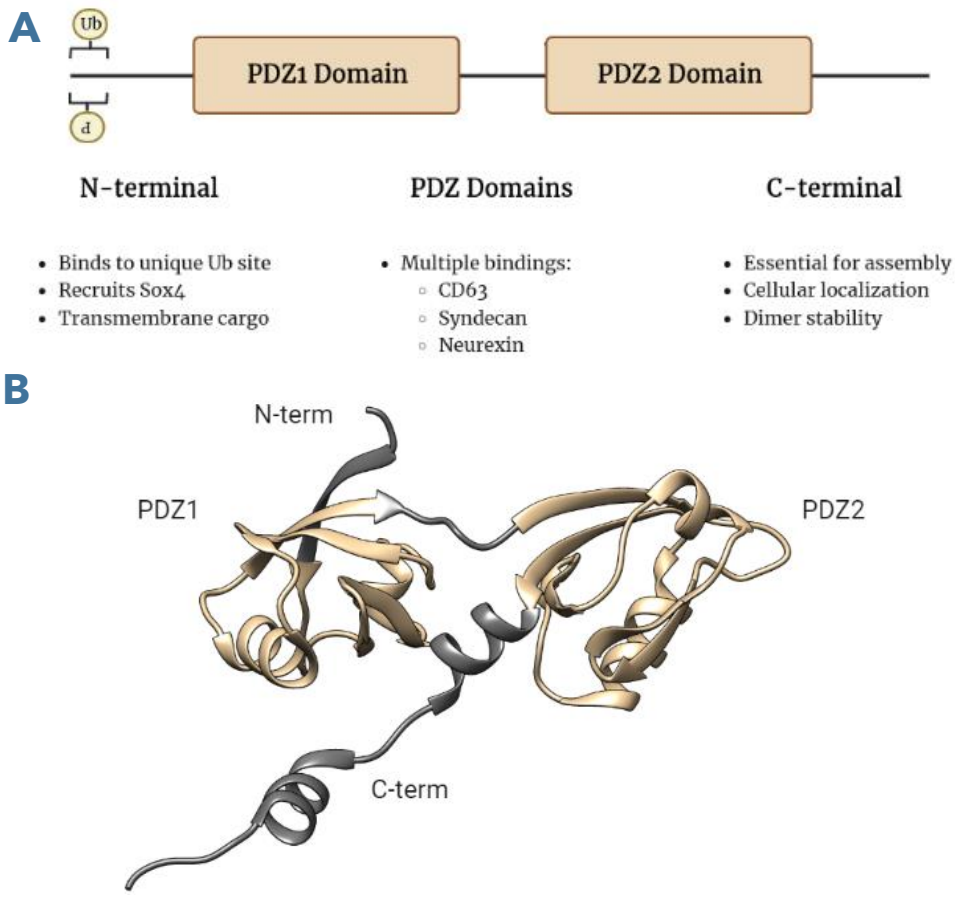

Figure 1 - Syntenin-1 aminoacidic structure of 7FT7.
A) Syntenin-1 sequence
B) Syntenin-1 3D structure. The N-terminal region is not complete (100 amino acids missing). There isn't any available structure in the PDB with the complete sequence.
Photos made with biorender.com

In turn, PDZ domains themselves are capable of interacting with multiple different proteins among them. In the case of PDZ1 it preferentially has IL-R5α (Geijsen et al., 2001), CD63 (Latysheva et al., 2006) or neurexin (Koroll et al., 2001) as binding targets; on the other hand, PDZ2 interacts not only with neurexin but also with neurofascin (Koroll et al., 2001), c-Src (Boukerche et al., 2010) and syndecan-4 (Yu et al., 2019). It is this latter protein that gives syntenin-1 one of its main names, syndecan-binding protein (Shimada et al., 2019).

This broad set of PPIs allows syntenin-1 to participate in many cell functions including cell adhesion, neuronal synapse development, immune regulation, exosome biogenesis, apoptosis modulation and angiogenesis (Yu et al., 2019; Shimada et al., 2019; Philley et al., 2016; Kegelman et al., 2015; Kegelman et al., 2017). Curiously, this protein is able to interact with the E (envelope) protein of the worldwide known coronavirus family, an integral membrane protein which is

accumulated during viral infection. This interaction could play a crucial role in virulence and severity of the acute damaging process (Jimenez-Guardeño et al., 2014).

The different interactions between the domains may be due to a few structural factors. Firstly, although both crystallized domains have high structural similarity, they share only 26% sequence identity (Figure 2). Furthermore, the binding groove of PDZ1 is narrower than that of PDZ2 (Figure 3). Finally, there are electrostatic differences in the binding site; the first domain has positive charge clusters that are absent in the second domain (Kang et al., 2003; Shimada et al., 2019) (Figures S1 and S2). Notwithstanding the PDZ1 interactions are usually weaker and the targets of both domains are different, it is currently being studied that a cooperation between both syntenin-1 domains may be required for an optimal activity with peptides such as syndecan (Grembecka et al., 2006; Martinez et al., 2023).

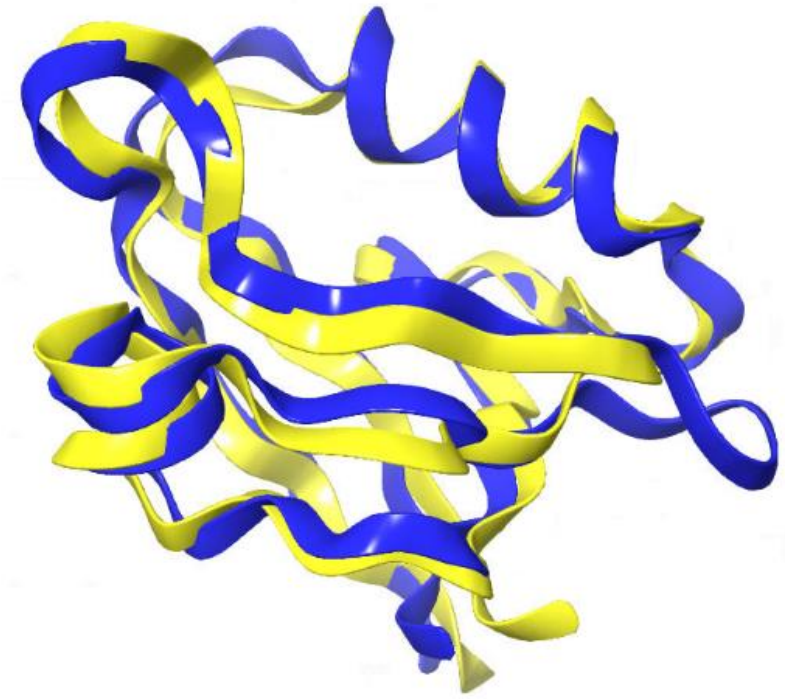

Figure 2 - PDZ domains superposition.
This figure shows a secondary structure superposition of the PDZ1 and PDZ2 domains from 8AAP structure.

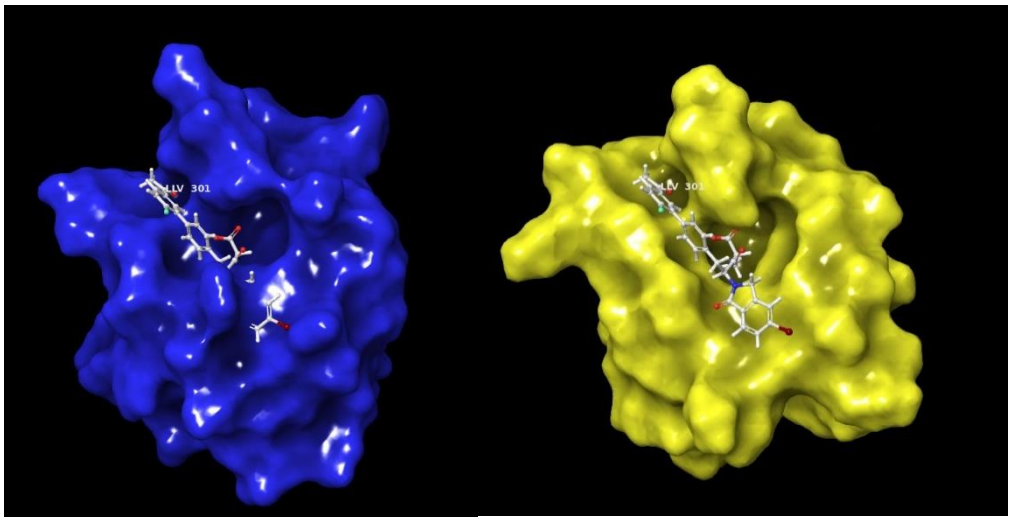

Figure 3 - Syntenin-1 surface comparison.
PDZ1 (blue) and PDZ2 (yellow) domains of 8AAP structural comparison. The ligand shown in the picture is the ligand of 8AAP, which is only capable to bind to PDZ2, as the binding site of the first domain is smaller.

4.2 Syntenin-1 in cancer

Since its identification in 1997 as a syndecan-4 interacting protein (Grootjans et al., 1997) this structure has been widely studied due to its multifunctionality as it plays important roles in cell-cell adhesion, cellular signaling or vesicular transport. While these functions may be associated with a normal cellular behavior, nevertheless, it is known that its overexpression is related to lung, colon, head and neck squamous cell carcinoma or melanoma cancers and metastasis. In addition, syntenin-1 is also identified in advanced tumor stages, survivability, and cancer recurrence (Shimada et al., 2019; Lin et al., 1998).

Syntenin-1 can be bound to Src cellular kinase which activates the nuclear factor kappa-B (NF-κB) signaling route. This pathway activates cellular motility, invasion, and metastasis genes such as matrix metalloprotease MMP-2. This MMP-2 is also related to metastasis development so, mda-9 overexpression may lead to MMP-2 increase. Syntenin-1-related metastasis studies are extended to exosome release. It is shown that syntenin-1 decrease in tumor cells can reduce exosome biogenesis, nanometric particles directly related to metastatic processes (Imjeti et al., 2017).

On the other hand, syntenin-1 can also activate Smad, which forms complexes with VEGFR2. Those complexes may lead to *de novo* blood vessels formations through angiogenic proteins overexpression (Das et al., 2020) (Figure 4).

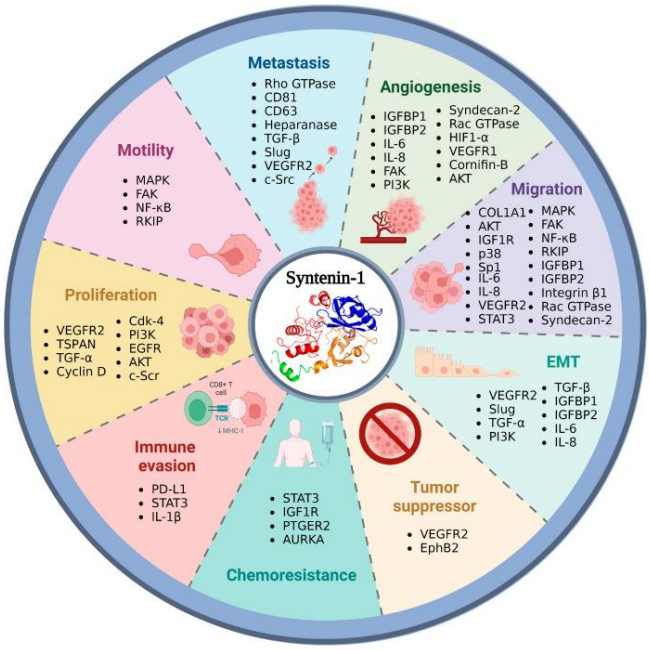

Figure 4 – Syntenin-1 and its interactions as a cancer-related protein. Original photo from Pintor-Romero et al., 2023.

4.3 Virtual Screening

Drug development involves millions to billions of dollars and may take further than 10 years each (Schlander et al., 2021; Gurung et al., 2021). The annual exponential increase in computational power has led to the optimization and application of *in silico* techniques that can overcome other drug development methods in the initial steps. The use of computer-aided drug development (also known as CADD) with cheminformatics approximations has allowed the creation of the virtual screening (VS) methods (Tripathi et al., 2022) which can be used to identify certain molecules from an initial library of drug-like compounds that may be capable to bind to a target structure (Gimeno et al., 2019). Despite high-throughput screening is the best method to identify lead drugs to study their absorption, distribution, metabolism, excretion, and toxicity (ADMET) it requires more money and time (Arul Murugan et al., 2022). Moreover, VS can allow the user to study multiple structurally different compounds to gain a broad knowledge of biochemical interactions (Gimeno et al., 2019).

VS methods can be divided into two main families depending on the structural information used by the researcher: ligand-based VS (LBVS) and structure-based VS (SBVS) (Vázquez et al., 2020). LBVS uses structural and physical chemistry information from previously studied molecules, typically found in public databases such as PubChem or ZINC. They rely on molecular similarity properties including geometric, thermodynamic, or electronic, and can also use 1D, 2D and 3D information (Banegas-Luna et al., 2018). These inputs are typically used in pharmacophore design, which is defined as "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biologic target and to trigger (or block) its biological response" (Wermuth et al., 1998). LBVS is often used along the initial steps of drug discovery as it consumes few resources to process lots of information (Gimeno et al., 2019). Despite of the usefulness of LBVS, these methods can be considered not potent enough to achieve further and more complete analysis. SBVS includes pathways based on the tridimensional structure of the target protein and the interactions resulting from the intermolecular contacts with the ligands, which are displayed in affinity rankings (Gimeno et al., 2019; Vázquez et al., 2020). Pharmacophores can be also obtained using either co-crystallized ligands on the PDB, the binding site residues, or flexible clusters of ligands that are not yet docked (Pinto et al., 2020). Despite the techniques already commented, the most common method among computational chemists could be the molecular docking between the target protein and a library of ligands; molecular docking (also called protein-ligand docking) can analyze and hypothesize the most optimal conformation and orientation of all compounds in the binding site, commonly known as *pose* (Torres et al., 2019; Gimeno et al., 2019). It consists of two main steps: sampling and scoring. Throughout sampling step, protein-ligand docking programs generate all possible poses of the ligand (considered flexible) in the binding site (considered as rigid in order to save resources). After this first stage, scoring gives a rating to all poses generated and discriminates between the most probable poses (Ballante et al., 2021; Salmaso et al., 2018; Pantsar et al., 2018).

5. Objectives

- Understand the differences between the syntenin-1 PDZ1 and PDZ2 domains and to target one of them.
- Provide further information on the tridimensional structure of the syntenin-1 and its intermolecular interactions.
- Identify hit molecules from a large compound library that could inhibit syntenin-1 activity by using VS.
- Provide future recommendations for syntenin-1 inhibitor discovery.

6. Hypothesis

The hypotheses of the current research project are that:

a) VS may be a good methodology to discover potential hit molecules targeting PDZ2 in syntenin-1.
b) The initial library of compounds may include a subset of compounds with potential inhibitory effects on the activity of syntenin-1 through its binding to the PDZ2 domain, thereby demonstrating their potential as anti-metastatic agents.

7. Methods and Materials

As a starting point for this project, **PubMed** was used to download the articles necessary to understand the insights of the syntenin-1 protein and the inhibitors in development (so-called as *actives* molecules). PubMed is a free access database that contains more than 35 million citations and abstracts of biomedical literature coordinated by the National Institutes for Health or NIH (*About – PubMed,* 2022).

After a brief knowledge of the syntenin-1 context, **PubChem** was used to collect information of possible actives (in parallel with the PubMed compound research). PubChem is an open chemistry database at the NIH that provides information on both small and large molecules and collects information on structures, identifiers, or activity (Kim et al., 2023). The PubChem and PubMed compounds were used

to validate the VS results and, in the case of the active ligands, were used for the decoy selection through the **DecoyFinder** tool. This user-free graphical interface helps finding a set of decoy molecules for a given input of active molecules (Cereto-Massagué et al., 2012) by using many chemical properties such as molecular weight or number of rotational bonds. This program was used to develop a set of 50 decoys for each active found in the bibliography. For DecoyFinder a ChEMBL database subset of compounds was used as input.

All structural information and diffraction data used in this study were obtained from **PDB,** the Protein Data Bank. It is a freely accessible database of tridimensional structural information for large molecules (proteins, DNA and RNA). It contains up to 200.000 crystallized compounds (Burley et al., 2019). **JMol** is a free, open-source viewer of molecular structures in 3D (Hanson & Lu, 2017). This program was used to identify the binding sites and the crystallographic contacts between chains A and B of each protein at the beginning of the study.

All the structures used in this project were analyzed using **PoseView,** an online server developed by the University of Hamburg that generates two-dimensional diagrams of complexes with known 3D structures and their interactions (hydrogen bond and hydrophobic contacts). This program (Stierand & Rarey, 2007) was used to determine the more important interactions between the target protein and its ligands and to select the proteins that could be used for the following protein preparations.

The **Maestro** suite was used for all structure analyses, protein and ligand preparations and molecular docking. It is an advanced graphical user interface used for computational drug discovery. It is developed by Schrodinger company and the Maestro versions used in this project were the 2022-04 and 2023-01 releases. Maestro consists of a group of applications depending on the purpose of the investigations:

- Maestro interface: It was used for the visual inspection of the structures and ligands and can manage all the following utilities.
- PrimeX: Performs refinements to the 3D structures based on the X-ray and diffraction data. This task was used to develop electron density maps that

would be used to select default or alternative positions for the amino acid residues. In this project these maps were analyzed:

  - Fo-Fc: Shows the differences between the experimental data and the atomic models.
  - 2Fo-Fc: Calculates the electron density by using the observed diffraction data and the calculated by the atomic model.
  - Since no alternate side chains were found near the binding site that could affect the ligand interaction, so the alternates positions were chosen only in obvious cases. Nevertheless, these positions would not have affected the molecular docking results.

- Protein Reliability Report: (the Isolated Waters Clusters and Waters with no HB partners were not considered to be as important by the Maestro tutorials).
- Protein Preparation Workflow: The deposited PDB structures often contain multiple errors such as missing hydrogen atoms, incomplete side chains, incorrect valences, flipped residues or imprecise protonation states. Those errors may be fixed manually by the crystallographers and cheminformatic experts, but the time and the difficulty of these processes can be drastically reduced by the Protein Preparation Workflow of the Maestro interface.
  - Preparation Workflow: This is an automated process that includes Preprocess, Optimize H-bond Assignments and Minimize and Delete Waters. The first step corrects structural defects and adds missing information (such as the lack of side chains). The Optimize H-bond Assignment task helps to correctly adjust hydrogen bonds and is capable to flip Asn, Gln and His side chains. Finally, Minimize and Delete Waters optimizes possible steric clashes due to hydrogen addition.
    - The default parameters were not modified since Maestro does not recommend this for inexperienced users.
  - Diagnostics: This non-automatic process reports possible steric clashes, non-corresponding valences, missing atoms or alternative residue positions. The positions suggested by Maestro are the result of duplicated information in the input file and are both considered as correct but there can be only one position at the

same time. To select the most proper conformation it is helpful to use the electron density map generator tool, PrimeX.

  - Substructure: It is used to delete atoms, residues, solvents or ligands. In this project only the proteins and ligands were used but the solvents were not retained because no water molecules gave rise to the intermolecular interactions necessary for inhibition.

- LigPrep: If the ligand is part of the crystallized structure, it must be extracted and have its tridimensional structure deleted. To obtain the native conformation of the ligand (the one it should have when is not bound to the target), the 3D molecule is converted to a SMILES (unidimensional chemical representation) format and then imported into the Maestro interface. The LigPrep task makes it easy to obtain the lowest energy conformation of the ligand. LigPrep is used to obtain multiple outputs depending on the conformation, protonation, stereochemistry, or tautomers of an unprepared ligand. The parameters chosen in this step were also the predefined values, except for the pH range, which was set to 7.0 ± 1.0 to match the pH set for the Protein Preparation.
- Receptor Grid Generator: For the proper ligand docking Glide (the molecular modelling software responsible for the ligand docking calculations (Friesner et al., 2004) requires an input of information representing the active or binding site of the receptor protein called grid. These grids are represented by purple boxes covering the region where the ligand should bind and can be identified using the Receptor Grid Generation panel on the Maestro interface. The Receptor Grid Generator holds three main tabs that had been modified for the different structures:
  - Receptor tab: It allows the user to identify the ligand (and thus define the grid position). Although the parameters were not changed, each ligand had to be selected. The grid generator process was useful to discard the 7FT7 structure as it could not include the main binding residues in the grid.
  - Site tab: This tab is used to define the volume and position of the ligand grid generated.
  - Constraints tab: This tab is used to include possible constraints in the molecular docking calculations. A constraint is a receptor-ligand

interaction that is considered important for the binding mode and that can be useful to enrich and speed up the docking calculations, since Glide (this program is included in the Maestro interface) is able to discard the outputs that do not satisfy the interactions.

  - Positional constraint: it defines a spherical region that should be occupied by a specific atom of the ligand.
  - Hydrogen bond constraint: Requires a specific H-bond between the receptor and the ligand.

- Ligand Docking: This panel is used to set up and run docking jobs using previously calculated receptor grids with Glide calculations. In this project the Glide SP precision was used, and the default options were chosen, except for:
  - Ligands: Each ligand was selected after their preparation, either for cross and redockings or for the molecular docking of the Specs library.
  - Output: 5 poses per ligand (with cross and redockings) or 1 pose per ligand (Specs library docking).
  - Constraints: Depending on the goal of each experiment the constraints were used or not.

The **ChimeraX USCF** and **sPDBv** were also used in this project. ChimeraX USCF is a free-of-charge program for interactive visualization and analysis of molecular structures and related data developed by at University of California (Pettersen et al., 2021). ChimeraX was used to visualize the main residues of the binding site. In the other hand, sPDBv is a free use application developed by Guex and Peitsch (1997) that allows the simultaneous analysis of several proteins. This program was used to superimpose the 3D structures of the PDZ1 and PDZ2 domains and to compare their binding sites.

Specs is the world's largest provider of research compounds with +350.000 screening molecules (specs.net). The Specs library provided all compounds that would be used for the ligand docking after being filtered by the Cheminformatics and Nutrition research group. The filtering of the compounds was based on the elimination of compounds that did not meet suitable ADMET characteristics (by using QikProp, an ADMET characteristics prediction tool of Maestro) or of those

PAINs (pan-assay interference) compounds. These structures are characterized by a very high false positive rate in VS studies (Bolz et al., 2021).

Pharmit is a free web server that facilitates VS by searching for small molecules based on their structural and chemical similarity (Koes, 2016). This server was used to develop the pharmacophore model after the molecular docking of the Specs library.

8. Results and discussion

Human syntenin-1 (UniProt code: O00560) is a protein deposited in the Protein Data Bank (PDB) database. For this current job, the main objective will be to discover potential hit molecules that may inhibit syntenin-1 activity by binding to the PDZ2 domain. Those potential drugs will be found through the VS of a library of compounds derived from Specs. Along this section, the methodology of the preparation of the target proteins and ligands and the results of the molecular docking will be explained.

8.1 Protein preparation and selection

In order to use the best structure for the virtual screening, the desired structure must be already selected, filtered and prepared so the results are closer to the *in vivo* conditions. Along this section all parameters and steps will be explained to choose the most proper structure for the protein-ligand docking.

Firstly, the PDB database was used to download all human syntenin-1 structures. The 54 structures found are different of each other and may have sequence variations. With a script developed in the Quimioinformàtica i Nutrició research group the sequences were separated depending on sequence length or presence of mutations. 6R9H was chosen as the reference structure because during the first steps it was the most recent structure in the PDB (along with 6RLC) and had a potential inhibitor drug co-crystallized (Table 1).

Table 1 - PDB structure analysis
List of all structures deposited in the PDB classified depending on sequence length and mutations. The reference structure was 6R9H.

1obz, 1v1t, 1w9e, 1w9o, 1w9q, 1ybo, 6r9h, 6rlc, 8aai, 8aao, 8aak, 8aap	- PDZ1 and PDZ2 - 166 amino acids - No mutations
4z33	- PDZ1 and PDZ2 - 166 amino acids - 1 mutation: ○ P5I
1n99	- PDZ1 and PDZ2 - 163 amino acids - 1 mutation: ○ M3MSE (selenomethionine)
7fta, 7ftb, 7ftc, 7ftd, 7fsg, 7fsh, 7fsi, 7fsj, 7fsk, 7fsl, 7fsm, 7fsn, 7fso, 7fsp, 7fsq, 7fsr, 7fss, 7fst, 7fsu, 7fsv, 7fsw, 7fsx, 7fsy, 7fsz, 7ft0, 7ft1, 7ft2, 7ft3, 7ft4, 7ft5, 7ft6, 7ft7, 7ft8, 7ft9, 8blu, 8blv	- PDZ1, PDZ2, N-term, C-term - 195 amino acids - 4 mutations: ○ G1M, M3Q, D4G, P5I
1oby, 1obx	- PDZ2 - 76 amino acids - No mutations
1r6j	- PDZ2 - 82 amino acids - No mutations

For the current project, the PDZ2 domain was the one and only domain needed as the main goal was to dock a library in the PDZ2 binding site. Although the presence of mutations in a bunch of structures, all mismatches were far from the binding site (localized in the N terminal sequence, unnecessary for the molecular docking as it is autoinhibited during the cellular activity), so no structure was neither discarded by mutation nor length. In the other hand, the 195 amino acid length structures contain the C-terminal sequence, far from the binding site.

8.2 PDZ Domain Interactions

Syntenin-1 contains two PDZ domains (PDZ1 and PDZ2) that can be bound to peptides such as syndecan-4. This interaction has led to the study of potential inhibitor drugs in both domains to avoid the PPI. Those PPI are slightly different between the two domains and could be due to the binding site size or the positive charge cluster in the first domains. Thus, it was important to define the domain that would be targeted in order to be more specific in the drug discovery.

The structural superposition of PDZ domains was useful to identify the two binding sites and to compare their structure. As it has been mentioned along this thesis, PDZ1 is narrower than the PDZ2 groove and has a Arg128 and His175 in the middle of the binding site (Figure 5). This suggested that it could be more difficult to find a potential or active compound for PDZ1 in comparison to PDZ2. Despite the possibility to find inhibitors that could bind to both domains, it has been proposed to focus only on one domain (*i.e.,* PDZ2).

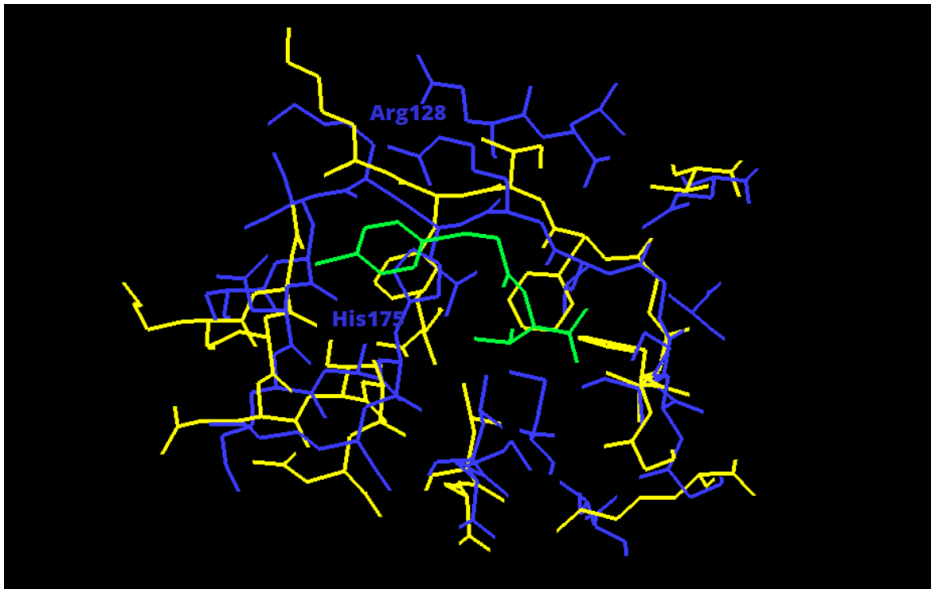

Figure 5 - Syntenin-1 PDZ domains comparison. The PDZ1 (blue) and the PDZ2 (yellow) domains of 8AAP were superposed. The Arg128 and His175 are labeled for a clearer visualization.

Currently there are a few studies focused on the synthesis or discovery of inhibitors of syntenin-1, targeting either PDZ1 or PDZ2 domains. The PDZ1i lead compound from Kegelman et al. (2017) is the most tested molecule *in vitro* and has given good safety and inhibition results (Pradhan et al., 2021; Das et al., 2019; Bhoopathi et al., 2019) but does not appreciably bind to PDZ2. Despite the

potential patient safety, this molecule shows two main disadvantages: PDZ1i exceeds 500 Da (so that cannot be considered as small drug) and there is no bibliography reported showing the intermolecular interactions between this ligand and the syntenin-1. The compound PI1a has been recently synthesized by Tang et al. (2023) and has shown an *in vitro* inhibitory potential by having a K_d value of 110 µM but the research has not been continued yet.

On the other hand, it is worth to mention the PDZ2 results, which could be more promising as it had been further studied and seemed to be more accessible to drugs. The most active compounds currently under study have shown IC_{50} values of about 400 nM (Hoffer et al., 2023; Leblanc et al., 2020; Garcia et al., 2021). Those low values PDZ2 are due to the structure-activity relationship approaches. Parallelly, this domain has been more structurally analyzed than PDZ1 (it could be evident just by the difference on the number of isolated PDZ2 structures deposited on the PDB database and the PubMed bibliography) (Table 2 and Table S1).

Table 2 - Most active PDZ2 inhibitors

Compound	pIC_{50}	IC_{50} µM
118/SYNTi	6,398	0,4
117	6,086	0,8
116	5,824	1,5
94	5,398	4,0
95	5,377	4,2
122	5,284	5,2
91	5,276	5,3

After this analysis, the PDZ2 was chosen as the target of the current investigation, as it is shown to have more druggability. Nevertheless, it is worth to mention that if some peptides could bind to both PDZ domains of syntenin-1 it could happen the same with the small inhibitory drugs. Although the earlier studied compounds show no appreciable activity in PDZ1, Leblanc et al. (2020) highlight that the PPI mechanisms remain unclear.

The next step was to study the contacts between the syntenin-1 chains. As a homodimer, both chains are in contact in the crystal structure and interact with each other. With the JMol interface the PDZ2 binding site was identified (Figure 6) and showed that the crystallographic contacts between chains would probably have no effect on the binding activity when the drug is present. This information let to work only with one chain for every structure and could ease further analysis.

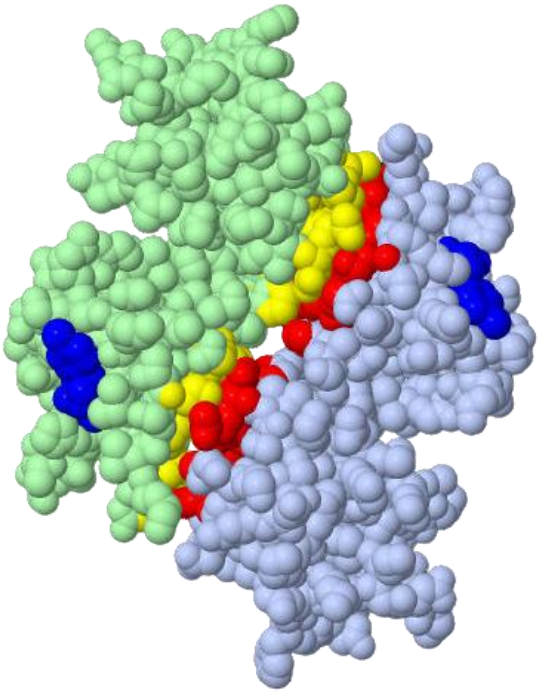

Figure 6 – 6R9H structure representation in JMol
Ligands are shown in dark blue, contact zone are yellow and red colored. The light green and blue represent both syntenin-1 chains.

8.3 Molecular interaction

Among all crystallized structures obtained from the PDB, only the ones with a co-crystallized ligand for PDZ2 were selected, as these structures would give more information of the protein-ligand interactions. By using PoseView, a computational tool capable of generating bidimensional interaction diagrams of non-covalent contacts, and the Ligand Interaction Diagram of the Maestro software the main intermolecular interactions were identified (Figure 7, Table 3).

Table 3 - Main interactions with PDZ2 domain
In the present table there are shown only structures with co-crystallized drug-like compounds.
- The green squares indicate hydrophobic interactions.
- N and O indicate Nitrogen and Oxygen atoms
- PA and PD indicate Proton Acceptor and Donor
- Pi-Pi indicates Pi stacking contacts

PDZ2	6r9h	6rlc	8aai	8aak	8aao	8aap
His208			Nδ - PD			
Val209	N - PD	N - PD	N - PD	N - PD	N - PD	N - PD
Gly210	N - PD	N - PD	N - PD	N - PD	N - PD	N - PD
Phe211	N - PD O - PA	N - PD O - PA		O - PA	N - PD O - PA	N - PD O - PA
Ile212						
Phe213	Pi - Pi	Pi - Pi	N - PD Pi - Pi	N - PD Pi - Pi	N - PD Pi - Pi	N - PD Pi - Pi
Lys214						
Thr219						
Val222						
Asp251						
Ala255						
Leu258						
Thr260						

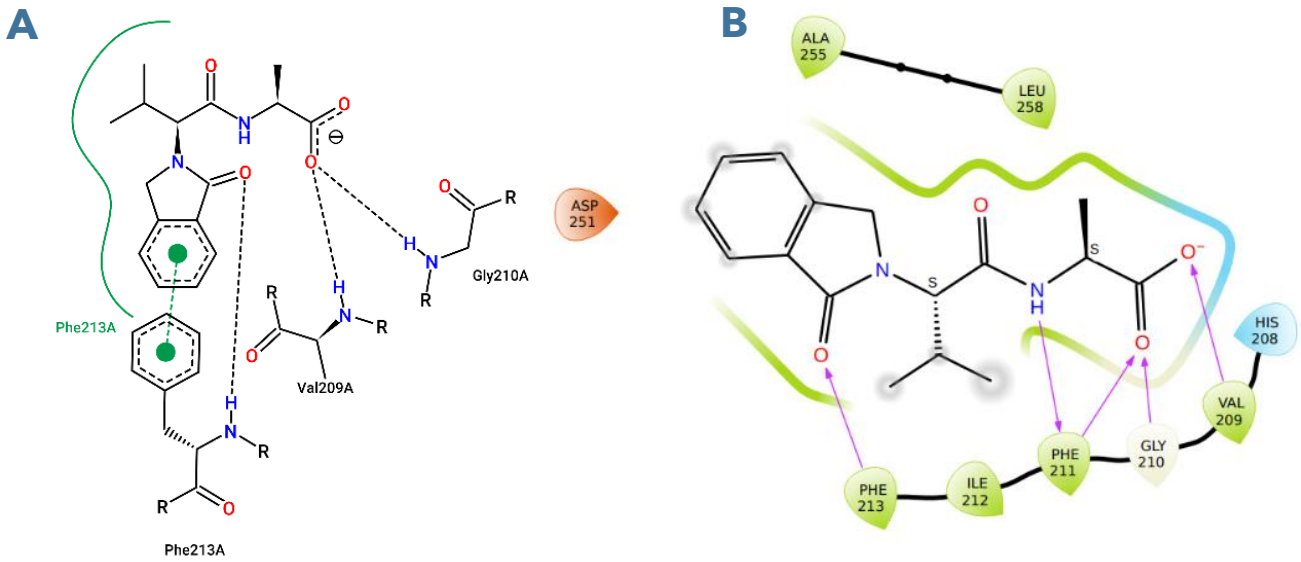

Figure 7 - PDZ2 interactions with ligands; 8AAI_A structure

A) With PoseView: Pi-Pi interactions are represented as a dot green line. Hydrophobic contacts are represented as a solid green line. H-bonds are represented as a dot black line.

B) With Maestro: H-bond contacts are represented in a purple line. Hydrophobic contacts are represented in green; polar regions are represented in light blue.

As seen in Table 3, intermolecular interactions comprise pi-pi stackings among aromatic rings, hydrogen bonds with both backbone or sidechains and hydrophobic interactions. Although, the most conserved interactions are:

- **Val209:**
    - Hydrogen bond with the backbone nitrogen that acts as a proton donor.
- **Gly210:**
    - Hydrogen bond with the backbone nitrogen that acts as a proton donor.
- **Phe211:**
    - Hydrogen bond with the backbone nitrogen that acts as a proton donor.
    - Hydrogen bond with the backbone oxygen that acts as a proton acceptor.
- **Ile212:**
    - Hydrophobic region
- **Phe213:**
    - Hydrogen bond with the backbone nitrogen that acts as a proton donor.
- **Ala255:**
    - Hydrophobic region

These interactions have been also mentioned by the authors who had already developed inhibitor molecules and studied tridimensional structure (Hoffer et al., 2023; Leblanc et al., 2020; Garcia et al., 2021). From this list, the most important ones for drug stability are the H-bond ones (Val209, Gly210, Phe211, Phe213) (Figure 8).

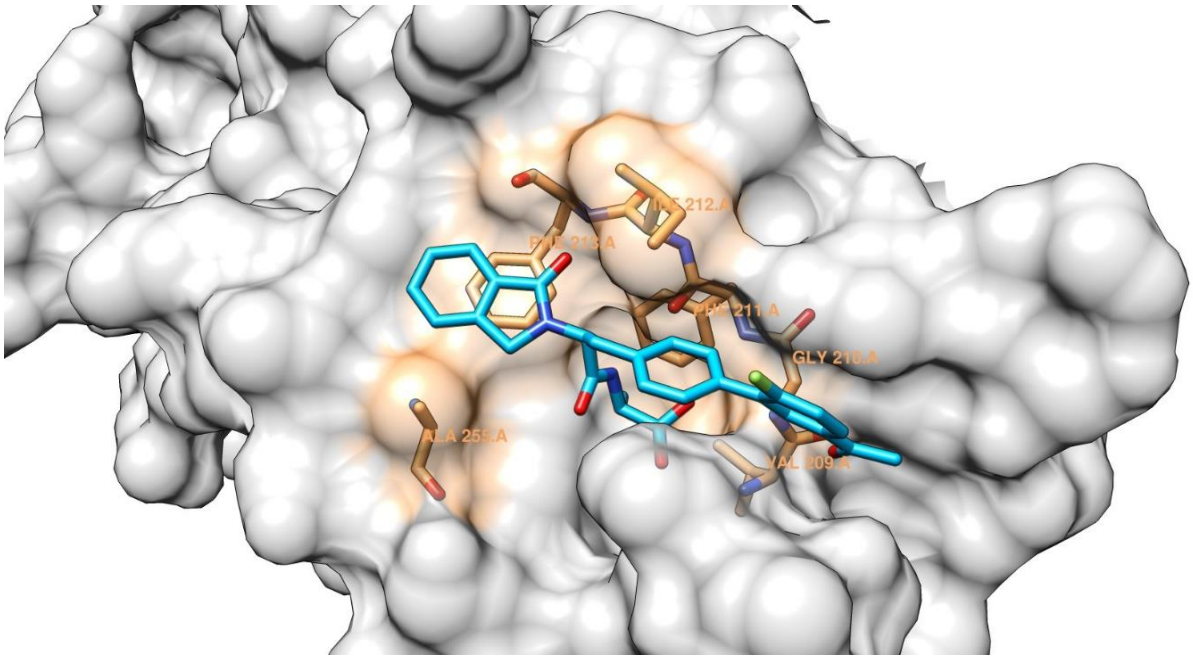

Figure 8 - Most conserved residues with 8AAO

8.4 Protein Selection

Although the high number of crystal structures available for the study, there were only 14 ones with a drug-like molecule in the binding site (Table 4). The discarded proteins were due to distinct reasons but mainly because of the lack of intermolecular interactions with the ligand. There were 10 structures with a peptide ligand binding site that were useful for identifying the main hydrogen bonds but, as the objective was the development of drug-like compounds, the structures used were only those bound to small drugs. Thus, the binding sites we worked with would be adapted to small molecules.

The PDB structure 7FT7 was discarded for two reasons:

- Although PoseView detects H-bond between the co-crystallized ligand and the Phe213 (one of the most shared interactions with the inhibitors) it did not interact with any of the desired residues (Val209, Gly210 and Phe211).
- During the Protein Preparation (one of the following steps) it was not possible to generate a grid that could cover all desired residues (Val209 was excluded).

Table 4 - Protein selection by ligand

No co-crystallized ligand	Allosteric ligand		Ligand in the binding site but no interaction	
1n99	7ftc	7ft4	7fsi	8blu
1r6j	7fsv	7fsg	7fsj	7fsk
	7ft7	7fss		
	7fsy	7ft6		
	7fsr	7fsp		
	7ftd	7fsm		
	7ft8	7fsl		
	7ft3	7fsw		
	7fsn	7fst		
	7ft0	7ftb		
	7ft2	7fsq		
	7fsu			
Others	**Drug-like ligand in binding site**		**Peptidic ligand in binding site**	
7ft7	8aao	7ft9	1w9o	1v1t
	6rlc	7fsh	1w9q	4z33
	8aap	7fsx	1ybo	8blv
	8aak	7ft1	1w9e	1oby
	6r9h	7ft5	1obz	1obx
	8aai	7fsz		
	7fso	7fta		

8.5 Docking Preparation

8.5.1 Protein Preparation Workflow

All 17 syntenin-1 structures were analyzed with the Protein Reliability Report pre and post processing. Thus, the tool allows the user to do a visual inspection of the errors in any PDB structure and see the improvements of the Protein Preparation Workflow calculations. As an example, the 8AAI crystal structure has been studied. The asymmetric unit (.pdb file) held 4 chains (so 2 proteins) and, in order to choose the best chain (as only one chain was necessary) the global quality was evaluated (Figure 9).

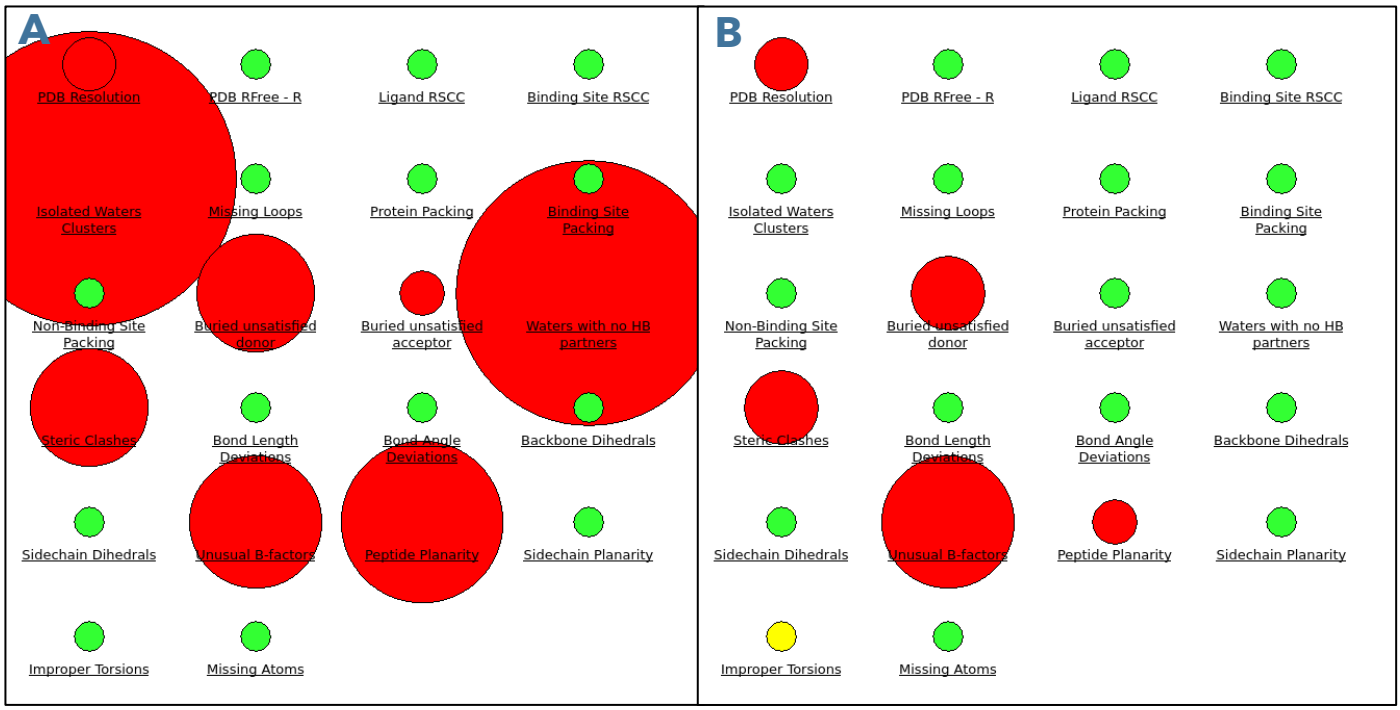

Figure 9 - Protein Reliability Report results of 8AAI structure

A) Protein Reliability Report before Protein Preparation
B) Protein Reliability Report after Protein Preparation

It was evident the number of problems the original structure had, especially in Unusual B-factors, Steric Clashes or Peptide Planarity analysis. After the Protein Preparation Workflow is evident the quality improvement in the crystal structure; although the Unusual B-factor remain the same, it is worth mentioning that errors shown in the image are result of the sum of the individual chain errors (Figure S3).

The Diagnostics tab of the Protein Preparation Workflow was used after a PrimeX analysis but only a few cases showed alternate positions, and those residues were far from the binding site, so the default positions were preferently chosen, in exception of cases like Figure S4.

8.5.2 Ligand Preparation

When the ligand is part of the crystallized structure it has to be extracted and have its tridimensional structure deleted. In order to obtain the experimental pose of the ligand (the one that should have when is not bound to the target), the 3D molecule is changed to a SMILES format (unidimensional chemical representation) and then imported to the Maestro interface. With the LigPrep task, it becomes easy to obtain the lowest energy conformation of the ligand. LigPrep

is used to obtain multiple outputs depending on the conformation, protonation, stereochemistry or tautomers of a non-prepared ligand. The parameters chosen in this step were also the predefined values except the pH range, which was set to 7.0 ± 1.0, so it was the same as the expected for the Protein Preparation (Figure 10).

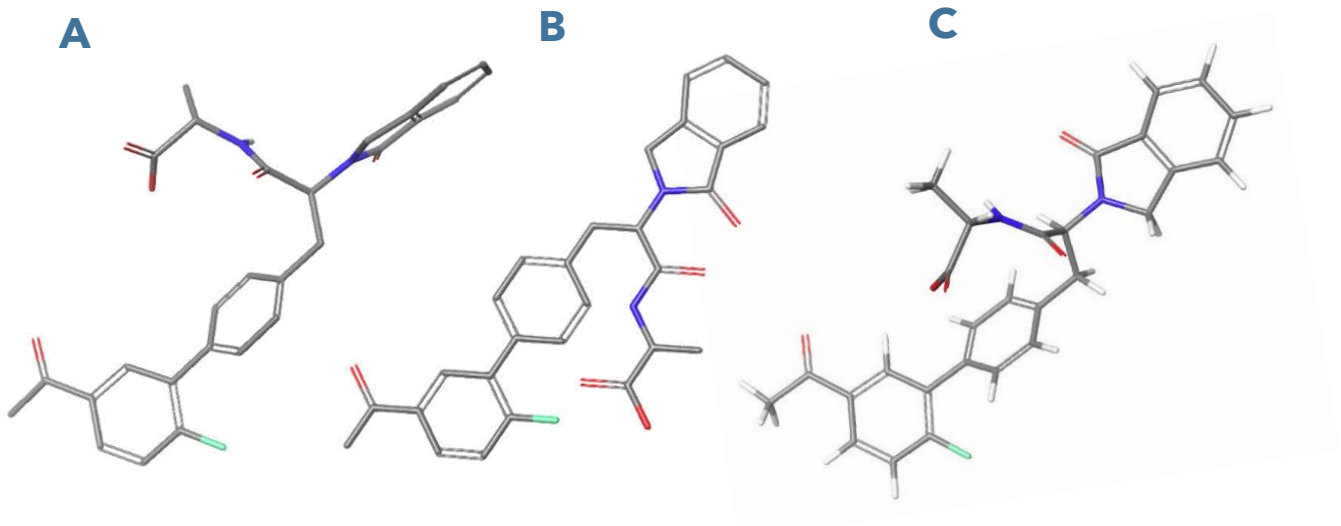

Figure 10 - 8AAO ligand
A) Cocristallized ligand before LigPrep
B) Bidimensional representation of the ligand
C) Ligand after the LigPrep processing

8.5.3 Receptor Grid Generation

Before the Ligand Docking is done, every structure had to have its grid defined with the ligand fit in the green box so, only when was required, the green box volume was set bigger than the default (Figure S5). The grid files are necessary for Glide (Grid-based Ligand Docking with Energetics) calculations, as they represent the position and the size of the binding site.

8.5.4 Constraints

Along with the ligand-receptor interaction study it was concluded that the main atoms in the groove may be the Val209, Gly210, Phe211 (Phe213 was also shared in many structures but not as present as the other three). Val209 and Gly210 share the N proton donor hydrogen bond and Phe211 has either N as donor or O as an acceptor. Despite this, the oxygen of the backbone was more relevant for Phe211 rather than the N (when referred to constraints) as it was

considered that the nitrogen H-bond could be achieved if the oxygen H-bond was present (Figure 11).

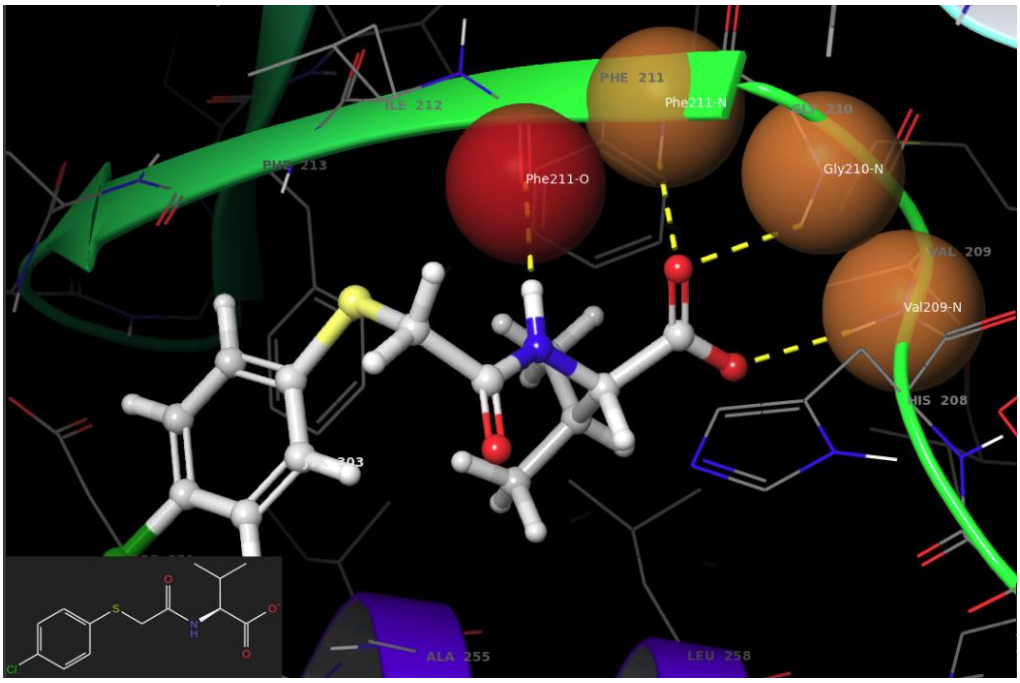

Figure 11 - 6R9H H-bond constraints. The nitrogen H-bond are constraints are set in color orange, the oxygen H-bond constraint is set in red.

8.6 Crossdockings and redockings

As mentioned previously, it is important to set the best structure to have the most reliable interactions between protein and ligand. To evaluate the capability to bind to a ligand, every protein had its own ligand docked after the LigPrep (redocking) and the other ligands from the rest of structures (crossdockings) (Figure 12).

The Ligand Docking task was used to both crossdocking and redocking. The default energy thresholds, dielectric constants or molecular sizes were not modified at all. Despite, in this point the constraints were modified depending on the results obtained.

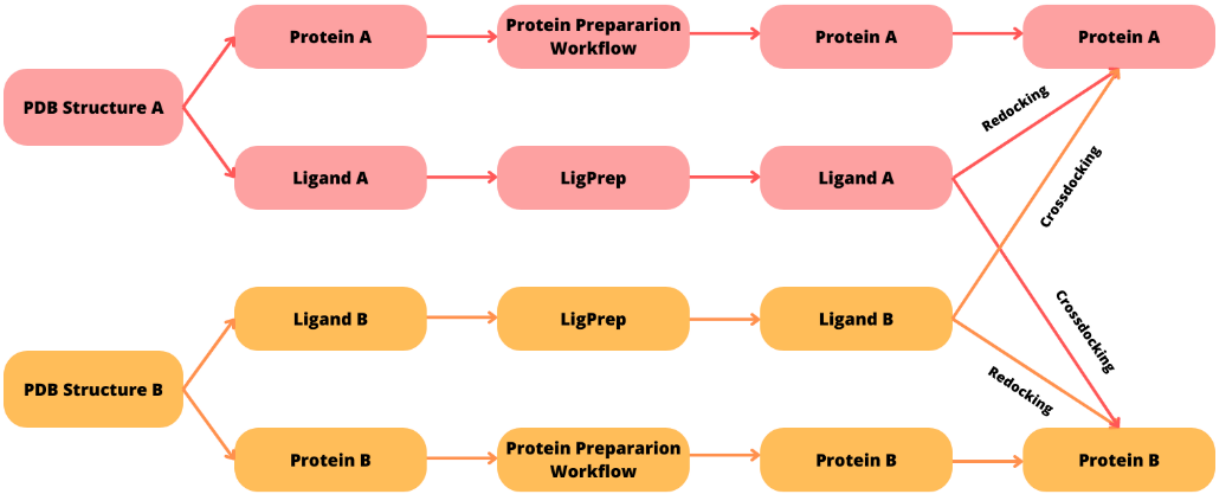

Figure 12 - Crossdockings and redockings workflow

8.6.1 Constraints

The main goal in this step was to identify the PDB structure that could rearrange its own and the other ligands with the lowest RMSD (root mean square deviation) value. This number is a comparison of the predicted and the experimental pose for the ligand and gives information of the deviation between the corresponding atoms of two (or more) compounds; the smaller the value, the more similar are the two poses (Reva et al., 1998). Three value ranges for the RMSD calculations were established: 0.0 - 1.0 Å as Great results, 1.0 - 2.0 Å as Good results, > 2.0 Å as Bad results.

As a first subset, the structures 6R9H, 6RLC, 8AAI, 8AAK, 8AAO and 8AAP had their RMSD calculated (Table 5). No constraints were chosen as these calculations were also focused on the comprehension of the most important intermolecular interactions. After defining the most important residues in the binding site, another grid was created with 4 hydrogen bond constraints (Val209 – N, Gly210 – N, Phe211 – O and Phe213 – N) (Table 6). With this input, the results suggested that, with a high number of restrictions, the ligand would not acquire proper poses because Maestro had not been capable of generating an output for many ligands. Another docking was executed but instead of forcing the usage of the 4 mentioned constraints the option "at least 2 of them" was chosen (Table 7). In that case, although the 6R9H and 6RLC ligands (the smallest ones) showed some poses, the rest of results showed no improvement.

Table 5 - RMSD value calculations. The ligand docking was performed with no constraints

Without Constraints						
RMSD value (SMARTS) – Å						
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**
6r9h	0,9714	0,7411	0,9404	3,0583	10,2537	6,2481
6rlc	1,755	0,2963	1,2921	2,2449	1,3762	1,0388
8aai	1,2661	1,5613	0,8546	1,7215	0,8757	0,5036
8aak	2,0294	2,4402	1,3413	1,2323	1,1935	1,2986
8aao	1,0388	2,2093	0,8179	1,7004	0,5202	0,4671
8aap	0,8182	2,1581	0,8636	1,7147	0,6139	0,5707

Table 6 - RMSD value calculations. The ligand docking was performed with 4 H-bond constraints, using all 4.
The constraints were: Val209 – N, Gly210 – N, Phe211 – O, Phe213 – N.

4 Constraints obligatory						
RMSD value (SMARTS) – Å						
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**
6r9h			0,9477	2,1721		0,5084
6rlc			0,9222	1,5022	1,1405	
8aai			0,7541	2,3875		0,7596
8aak			1,6096	2,0874	1,1357	1,4103
8aao			0,8324	1,9546	0,3389	0,4737
8aap			0,8272	2,0379	0,614	0,6192

Table 7 - RMSD value calculations. The ligand docking was performed with 4 H-bond constraints, using at least 2.
The constraints were: Val209 – N, Gly210 – N, Phe211 – O, Phe213 – N.

4 Constraints – at least 2						
RMSD value (SMARTS) – Å						
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**
6r9h	2,1356	2,1425	0,9163	4,57		7,5133
6rlc	1,2239	1,3021	1,3898	2,0137	6,9393	6,8406
8aai			0,7549	2,4337	4,5205	0,5863
8aak	4,7843					
8aao	1,3396	1,5253	0,7197		0,386	2,313
8aap	6,0167		0,9523	5,1666	2,5246	1,0335

Then positional constraints were used for the ligand docking. To be more specific on the interactions that should be targeted, the proton acceptors of the ligands that could interact with the Val209 or Gly210 were used to generate the constraints (Figure 13, Tables 8 and 9). The usage of these constraints resulted in pretty good results, especially on structures 8AAI, 8AAP and 8AAO.

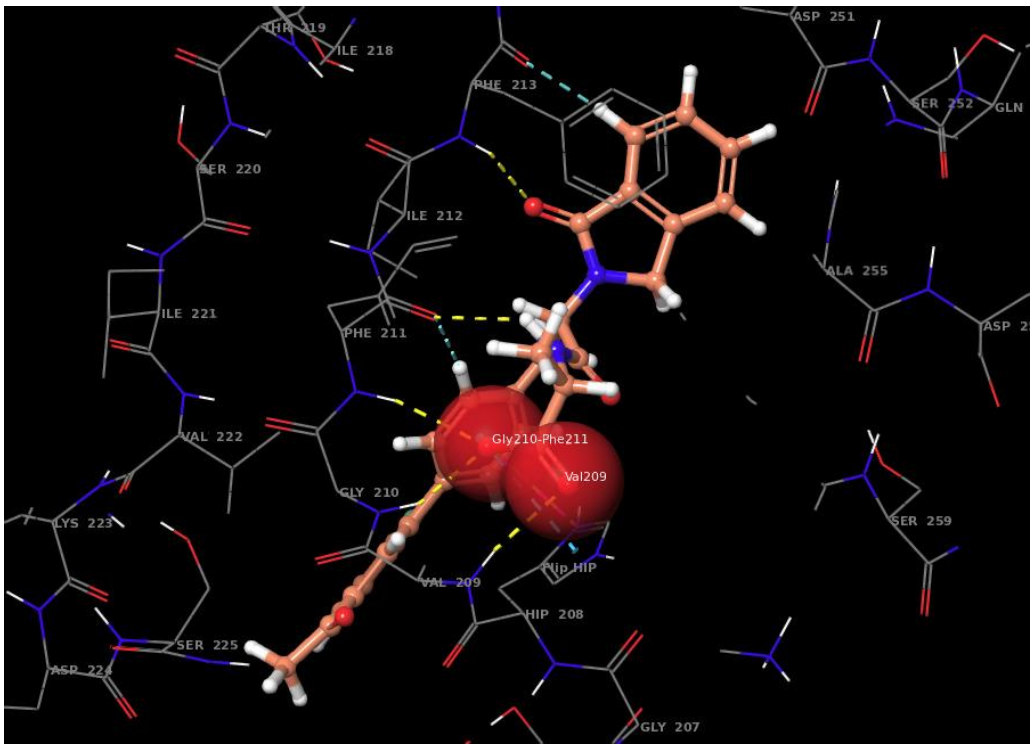

Figure 13 - Positional Constraints on 8AAO
The selected constraints are the oxygen atoms that create H-bonds with Val209 – N, Gly210 – N, Phe211 – N.

Table 8 - RMSD value calculations. The ligand docking was performed with 2 Positional Constraints, using all 2.
The constraints targets were: Val209 – N, Gly210 – N, Phe211 – N.

2 Positional Constraints obligatory						
RMSD value (SMARTS) – Å						
	6r9h_lig	6rlc_lig	8aai_lig	8aak_lig	8aao_lig	8aap_lig
6r9h	0,8093	1,7032	0,9332	2,5616	9,8432	6,1056
6rlc	1,1798	0,7573	1,4845	1,2218	1,1862	6,3437
8aai	0,9941	1,8514	0,7773	1,1909	1,4954	2,2551
8aak	1,7605	1,5645	1,4939	1,7994	3,832	1,3623
8aao	0,8971	1,1142	0,8013	2,0677	0,7545	0,4893
8aap	1,1437	1,4678	0,7395	1,3653	0,4302	0,381

Table 9 - RMSD value calculations. The ligand docking was performed with 2 Positional Constraints, using at least 1.
The constraints targets were: Val209 – N, Gly210 – N, Phe211 – N.

2 Positional Constraints – At least 1						
RMSD value (SMARTS) – Å						
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**
6r9h	0,8008	1,003	0,9735	1,1539	9,4925	5,7259
6rlc	1,1798	0,7573	1,4845	1,2218	1,1862	6,3437
8aai	1,2179	1,8644	0,7507	1,7073	0,6909	2,3399
8aak	1,3439	1,5486	1,4067	1,8186	7,7749	1,4437
8aao	0,8973	1,3278	0,7544	1,3864	1,7133	1,2637
8aap	0,7275	1,4864	0,699	1,2324	0,3754	1,6811

The hydrogen bond constraints were revaluated after the positional constraint, but this time using only Val209 – N, Gly210 – N and Phe21 – O. At this point, all the structures were tried to be docked as the main interactions were already confirmed. By using the mentioned H-bond constraints two dockings were analyzed, forcing 3 interactions or at least 2 of them (Tables 10 and 11). These results indicate various things:

- Val209 – N, Gly210 – N, Phe211 – O may be good constraints to apply on the compound library docking.
- The structures 8AAO and 8AAP are both adequate to use as model for the ligand docking.
- The best dockings were for the 6R9H, 6RLC, 8AAI, 8AAK, 8AAO and 8AAP structures, especially on those ligands.

Table 10 - RMSD value calculations. The ligand docking was performed with 3 H-bond constraints, using all 3. The constraints targets were: Val209 – N, Gly210 – N, Phe211 – O.

3 Contraints Val-Gly-Phe // RMSD (SMARTS) – Å														
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**	**7fso_lig**	**7ft9_lig**	**7fsh_lig**	**7fsx_lig**	**7ft1_lig**	**7fsz_lig**	**7fta_lig**	**7ft5_lig**
6r9h	0,8209	1,0195	0,8291	2,7540	8,2455	5,1012	5,2794	4,2592	3,1924		2,8366	5,317	1,9439	2,7826
6rlc	1,4238	1,0256	1,362	2,1225	10,6324	6,235	5,0992	5,6292	5,3133		5,4774	5,3704	1,5246	2,6656
8aai	1,1507	1,3354	0,8705	1,6038	8,3097	0,4668	5,3586	3,5851	4,6094			5,1984	0,9877	
8aak	1,7584	1,8600	1,6581	2,0059	1,2262	1,455	6,822	3,5294	4,7116			5,6638	1,4738	
8aao	1,231	1,0746	0,7947	1,8392	0,3206	0,799	4,5599	3,331	4,6239			5,6387	1,4057	
8aap	1,0197	1,5937	0,6718	1,8068	0,8234	0,5735	6,8622	4,0718	4,8145			5,4638	1,5716	
7fso										1,0000				
7ft9	2,3120	2,2304	1,7399	2,5757			6,4325	2,5234	4,9651			6,0622	7,0043	
7fsh	3,6189	2,8626	2,8763	2,8682	4,9431	5,9814	7,008	2,6612	4,7106			6,5352	7,4078	
7fsx	2,5615	2,3539	1,6745	4,769	1,5815	4,0499	6,9887	2,9843	4,3239			6,1781	0,6263	
7ft1	3,8742	3,1758	3,5883	2,8012	3,7661	4,1078	6,9398	3,0535	5,0299			6,9038	7,3074	
7fsz	2,3024	1,7617	1,6571	2,012	8,076		7,0676	3,1304	4,8436			5,4963	0,668	
7fta	2,4844	2,0649	2,0167	2,7303	9,0527		6,8611	6,3754					0,8126	
7ft5	2,4742	3,3151	1,8153	2,4668	2,7691	2,8596	6,3238	3,5878	4,9855			6,0268	7,0214	

Guillem Arasa Estivill
Universitat Rovira i Virgili

Table 11 - RMSD value calculations. The ligand docking was performed with 3 H-bond constraints, using qt least 2. The constraints targets were: Val209 – N, Gly210 – N, Phe211 – O.

3 Contraints Val-Gly-Phe, at least 2 // RMSD (SMARTS) – Å														
	6r9h_lig	**6rlc_lig**	**8aai_lig**	**8aak_lig**	**8aao_lig**	**8aap_lig**	**7fso_lig**	**7ft9_lig**	**7fsh_lig**	**7fsx_lig**	**7ft1_lig**	**7fsz_lig**	**7fta_lig**	**7ft5_lig**
6r9h	0,8209	1,0195	0,8291	2,754	8,2455	5,1012	5,2794	4,2592	3,1824		2,8366	5,4001	1,9439	2,6989
6rlc	1,7543	1,0256	1,362	2,1225	10,6324	6,235	5,0992	5,7651	5,3133		5,4774	5,4111	1,5246	2,6656
8aai	1,842	1,8078	0,861	2,4289	1,2992	0,5725	5,8419	5,7442	2,8786	3,63	6,1266	5,0036	0,99	2,7437
8aak	1,7422	1,1742	1,4324		3,8207	1,5965	5,7296	5,2009	2,9657		5,6853	5,6199	1,5042	2,5651
8aao	1,4385	1,1742	0,8958	1,329	0,567	1,5175	5,5987	5,4635	4,9315		5,8255	5,7803	1,3779	2,4841
8aap	1,0446	1,5541	0,8833	1,7571	0,8142	2,6266	4,5918	5,9534	4,9185		4,6179	5,5108	2,08	2,7412
7fso	2,2329	3,6364	1,8915	4,6572		8,1863	6,3652	5,9151	4,9265	1,2267	5,349	5,23	0,7314	1,955
7ft9														
7fsh	3,6193	3,3335	3,024	2,9927	5,2303	5,7953	6,1825	5,9186	5,1153	2,4372	5,5042	6,373	1,648	2,0706
7fsx	2,5375	2,3539	1,6745	4,769	1,5815	4,0499	6,9887	2,9843	4,3239			6,1781	0,6263	
7ft1	3,8435	3,0529	3,0354	2,7911	7,1549	7,731	5,7977	5,8594	5,6124	2,2685	5,424	6,8495	2,1247	2,417
7fsz	2,3	1,8768	1,6527	2,2876	5,4651	8,6374	5,9002	4,3424	2,9476	1,4912	2,8165	5,4111	0,3975	2,3744
7fta	2,5205	2,4758	1,9523	6,173	5,8723	7,3822	6,1583	5,7124	2,2724	5,3004	3,3386	5,5827	0,4135	1,3429
7ft5	2,8914	2,1538	1,7904	4,1836	1,9467	5,9003	6,2665	5,8417	3,9441	1,0591	5,3495	5,4011	0,7001	0,8385

These results may be consistent with the ligand similarity based on Tanimoto similarity of the co-crystallized molecules of the structures (Figure 14). 8AAI, 8AAK, 8AAO and 8AAP structures are the most related structures and the ligands are derivatives of the same compound, so the chemical properties would be similar. The 6R9H and 6RLC show also quite similarity and the RMSD values are lower than the rest of structures that show no correlation. Despite of this, the 6RLC and 6R9H have higher RMSD values with 8AAI, 8AAK, 8AAO and 8AAP that could be explained by the size of the ligands, as the first two have about 310 Da of MW and the other 4 are about 500 Da. The rest of the ligands, as they are smaller (lower or about 200 Da and show no similarity), may be the reason for the higher RMDS. Otherwise, it should be considered the origin of the crystal structures. 8AAI, 8AAK, 8AAO and 8AAP were crystallized by Hoffer et al. (2023) group, 6RLC and 6R9H after Leblanc et al. (2020) experiments. The rest of structures shown in figure 14 were crystallized by one research group, but at the moment of the publication of this project, the original article was not available and could not be inspected.

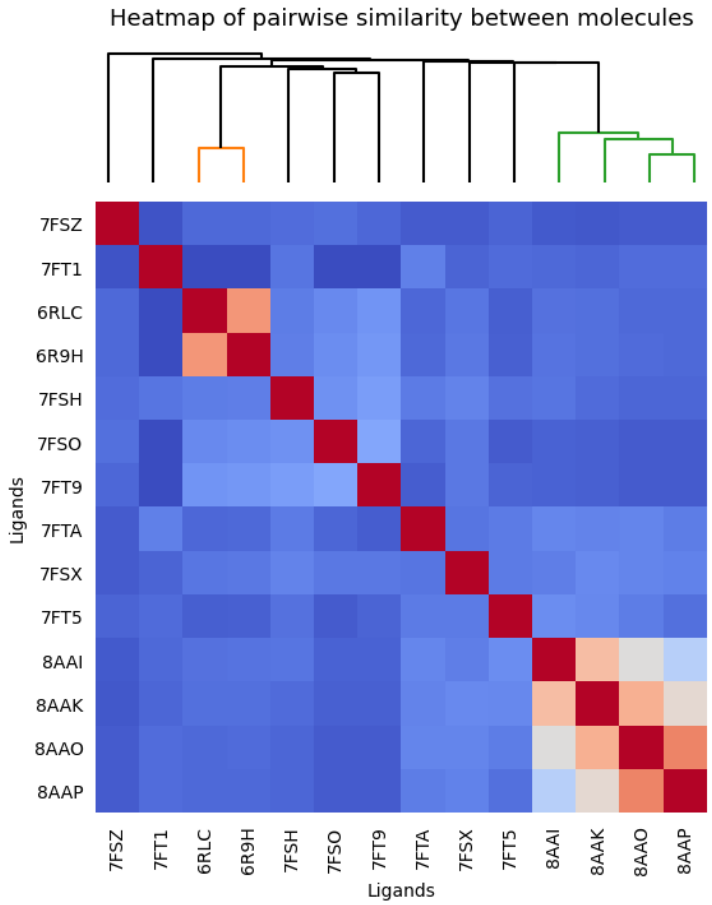

Figure 14 - Tanimoto similarity of the cocrystallized ligands. The blue colors represt no correlation; the red colors represent more similarity.

8.7 Virtual Screening

8.7.1 Molecular Docking of Specs library

Having the constraint validated, the set of 174.925 compounds of the Specs library was docked with the 8AAO grid as input. This library is a subset of compounds that had been already filtered by the Cheminformatics and nutrition group according to whether they satisfied ADMET properties and were not PAINS molecules (a class of compounds that may interact unspecifically and generate a higher false positive ratio (Magalhães et al., 2021)). The goal of this molecular docking was to discard all those molecules that could not fit in the binding site of PDZ2 domain and orient the rest of molecules at the 8AAO binding site. The docking resulted in 110.792 poses for the Specs compounds, meaning that 63.133 poses could not satisfy the binding interactions. A set of decoys was generated with the DecoyFinder with a ChEMBL library input, which gave 650 decoys (50 for each of the 13 active compound). The decoys were also filtered with those constraints and passed from 1.377 poses (obtained after the LigPrep of the 650 decoys) to 885 poses. Finally, the 13 active compounds were also docked in the 8AAO grid, but all active molecules were maintained.

8.7.2 Pharmacophore development

Not only the capability of the Specs molecules to fit on the binding site was important but also to present some interactions present on the main structures. A pharmacophore model was developed with the Pharmit interface (as mentioned in the introduction, a pharmacophore enables to filter molecules based on structural and chemical similarities). For the pharmacophore development only the 8AAP structure was used since it was co-crystallized with the most potent inhibitor at the moment (Compound SYNTi, Table S1). Although 8AAO was also bound to an active molecule (Compound 95, Table S1), it was considered not active enough for the goal of this step. The features chosen for this step were three hydrogen donors, one hydrogen acceptor and two aromatic regions (Figure 15). This combination was the one with the highest EF value.

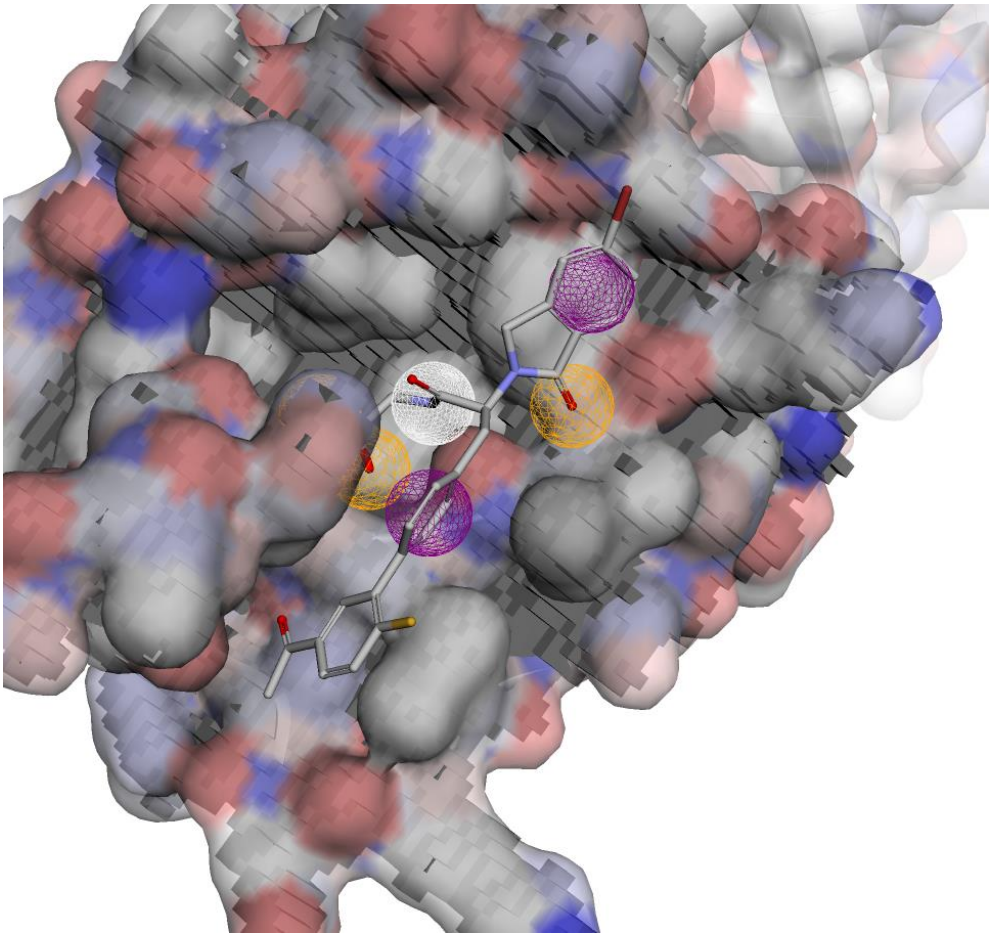

Figure 15 - Features chosen for the pharmacophore development. The white spheres represent H-donor, yellow spheres represent H-acceptor, purple spheres represent aromatic regions. Each sphere has a 1.3 Å radius.

Although 110.792 poses were obtained from the Ligand Docking of Specs library, the modification of the output files (the change from .mae to .sdf extensions) reduced the number of conformers by 5.167. These eliminated components were repeated structures with different protonation, which the program detected as duplicates.

Of the 13 initial active components, only 7 were retained after filtering by the pharmacophore; in turn, of the 885 decoy poses, none passed the filter. This filter also considerably reduced the number of poses in the Specs library, passing from 105.625 to only 6 final molecules (Figure 16).

Workflow steps	Validation			Virtual Screening
	Actives	Decoys	EF	Specs library
Starting database	13	1.377	-	174.925
Ligand Docking	13	885	1,54	110.792
Pharmacophore	7	0	69,08	6

Figure 16 - Results of the Specs library filtering

9. Conclusions

In this study, a virtual screening workflow of a library drug-like molecules from the company Specs has been carried out, with the aim of finding potential inhibitors that could bind to the PDZ2 domain of syntenin-1. This protein is associated with tumor development, especially in more aggressive cancers with metastatic activity. The Maestro interface has been mainly used for the protein-ligand docking and Pharmit was used for the screening of the library of components by the pharmacophore development.

First, the intermolecular interactions between the drug-like ligands and the structures that were available in the PDB were studied, with the aim of finding out which ones could be the most important for the inhibition of syntenin-1. PDZ2 domain of syntenin-1 was proposed as a target since, according to the reported literature, this seemed to be the one with the highest inhibitory potential due to its structural and chemical characteristics. In addition, in the PDZ2 domain there were reported a greater number of potential drugs, much more active, that would be useful for the pharmacophore filtering. Subsequently, it was considered which was the most suitable structure for Ligand Docking by cross-docking and re-docking and it was concluded that the 8AAP and 8AAO structures were the ones with the lowest RMSD values, the second one being chosen for the subsequent procedures. However, in future studies it might be useful to use more than one structure (when these have low RMSD values), especially if more are

deposited on the PDB in the future. As discussed above, structures with more similar ligands (according to Tanimoto similarity) gave better results than in the case of those structures with non-similar ligands.

The constraints selected by protein-ligand docking seem to give good results with the validation of the 8AAO structure, as those selected hydrogen bonds (Val209-N, Gly210-N, Phe211-O) seem to be the minimum necessary to have good inhibitory components. These constraints were validated using a set of 650 decoys extracted from a ChEMBL library. Ligand Docking allowed to discard those structures that could not satisfy the requested constraints, eliminating more than 60,000 poses. However, during the library processing process (the change from .mae to .sdf file) approximately 5,000 components were lost which, despite being duplicate compounds but with different protonation, it would have been desirable to work with them as well.

Finally, a pharmacophore was developed with the aim of filtering out all those compounds that could not meet certain pharmacological characteristics. The ligand SYNTi, co-crystallized with the 8AAP structure, was used for the development of this pharmacophore, as it is currently considered to be the most potent PDZ2 inhibitor (IC_{50} = 400 nM). The constraints selected by the pharmacophore were those that gave the highest EF (EF = 69,08), and this was reflected in the filtering result of the Specs library, which went from 105.625 poses to only 6. In further studies, these compounds should be tested in order to see whether they are good inhibitors.

As future insights in the study of syntenin-1, it would be of great interest that there could be an increase in the literature with respect to intermolecular protein-ligand interactions by both the PDZ1 and PDZ2 domains. In the first case, it would be important to study possible inhibitors (since only the 1W9E structure had a co-crystallized ligand), especially focused on drug-like molecules; for PDZ2, it would be of interest to explore other methods for inhibition, such as covalent ligands.

Finally, VS methods may be very useful for drug discovery and development due to the increasing computational power and the growing information stored in drug-like databases. It would be useful in future studies to be able to use multiple

docking interfaces (such as AutoDock Vina or DOCK) or pharmacophore development programs to choose those structures that are shared between the results of different applications, as due to the duration of this project, it was not possible to carry out this methodology.

Guillem Arasa Estivill
Universitat Rovira i Virgili

10. Self-assessment

At first, I would like to express my gratitude to Dr. Santi Garcia and Dr. Gerard Pujadas for allowing me to do my project in the Cheminformatics and nutrition research group. I contacted Gerard because the lessons I did with him were pretty much enjoyable and I think that it has been the best decision I have made. The working environment has been especially positive, for example during the day to day with Ariadna, Riccardo and Camilla, and the treatment received from my two supervisors has been excellent too. The weekly meetings were good checkpoints to update the knowledge acquired and I think that they had been so useful.

This work has helped me to understand what is involved in the study of an extensive bibliography, because the project has required understanding what the structure was or how it interacted, among others, and this understanding of the academic literature took me more time than expected. I am very satisfied with the knowledge I have acquired so far, and I have learned a lot more about chemistry and intramolecular interactions. Also, this group has been a great challenge for me as I had not had that much knowledge about this field.

If it would have been possible, I consider it would had been useful to have extensive programming and Ubuntu knowledge to work more comfortably and not to depend on the other colleagues on the lab. If my academic and professional future allows me, I will have to invest more time in learning more on bioinformatics.

11. Bibliography

About - PubMed. (2022, October 5). PubMed. Last access on June 1, 2023, from https://pubmed.ncbi.nlm.nih.gov/about/

Arul Murugan, N., Ruba Priya, G., Narahari Sastry, G., & Markidis, S. (2022). Artificial intelligence in virtual screening: Models versus experiments. *Drug discovery today*, *27*(7), 1913–1923. https://doi.org/10.1016/j.drudis.2022.05.013

Ballante, F., Kooistra, A. J., Kampen, S., de Graaf, C., & Carlsson, J. (2021). Structure-Based Virtual Screening for Ligands of G Protein-Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacological reviews*, *73*(4), 527–565. https://doi.org/10.1124/pharmrev.120.000246

Banegas-Luna, A. J., Cerón-Carrasco, J. P., & Pérez-Sánchez, H. (2018). A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future medicinal chemistry*, *10*(22), 2641–2658. https://doi.org/10.4155/fmc-2018-0076

Bhoopathi, P., Pradhan, A. K., Bacolod, M. D., Emdad, L., Sarkar, D., Das, S. K., & Fisher, P. B. (2019). Regulation of neuroblastoma migration, invasion, and in vivo metastasis by genetic and pharmacological manipulation of MDA-9/Syntenin. *Oncogene*, *38*(41), 6781–6793. https://doi.org/10.1038/s41388-019-0920-5

Bolz, S. N., Adasme, M. F., & Schroeder, M. (2021). Toward an Understanding of Pan-Assay Interference Compounds and Promiscuity: A Structural Perspective on Binding Modes. *Journal of chemical information and modeling*, 61(5), 2248–2262. https://doi.org/10.1021/acs.jcim.0c01227

Boukerche, H., Aissaoui, H., Prévost, C., Hirbec, H., Das, S. K., Su, Z. Z., Sarkar, D., & Fisher, P. B. (2010). Src kinase activation is mandatory for MDA-9/syntenin-mediated activation of nuclear factor-kappaB. *Oncogene*, *29*(21), 3054–3066. https://doi.org/10.1038/onc.2010.65

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Caruso, U., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R.,

Guranovic, V., Guzenko, D., Hudson, B., Liang, Y., Lowe, R., Peisach, E., . . . Ioannidis, Y. (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1), D520–D528. https://doi.org/10.1093/nar/gky949

Cereto-Massagué, A., Guasch, L., Valls, C., Mulero, M., Pujadas, G., & Garcia-Vallvé, S. (2012). DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics (Oxford, England),* 28(12), 1661–1662. https://doi.org/10.1093/bioinformatics/bts249

Choi, Y., Yun, J. H., Yoo, J., Lee, I., Kim, H., Son, H. N., Kim, I. S., Yoon, H. S., Zimmermann, P., Couchman, J. R., Cho, H. S., Oh, E. S., & Lee, W. (2016). New structural insight of C-terminal region of Syntenin-1, enhancing the molecular dimerization and inhibitory function related on Syndecan-4 signaling. *Scientific reports*, *6*, 36818. https://doi.org/10.1038/srep36818

Cui, L., Cheng, S., Liu, X., Messadi, D. V., Yang, Y., & Hu, S. (2016). Syntenin-1 is a promoter and prognostic marker of head and neck squamous cell carcinoma invasion and metastasis. *Oncotarget*, *7*(50), 82634–82647. https://doi.org/10.18632/oncotarget.13020

Christensen, N. R., Čalyševa, J., Fernandes, E. F. A., Lüchow, S., Clemmensen, L. S., Haugaard-Kedström, L. M., & Strømgaard, K. (2019). PDZ Domains as Drug Targets. *Advanced therapeutics*, *2*(7), 1800143. https://doi.org/10.1002/adtp.201800143

Das, S. K., Bhutia, S. K., Kegelman, T. P., Peachy, L., Oyesanya, R. A., Dasgupta, S., Sokhi, U. K., Azab, B., Dash, R., Quinn, B. A., Kim, K., Barral, P. M., Su, Z. Z., Boukerche, H., Sarkar, D., & Fisher, P. B. (2012). MDA-9/syntenin: a positive gatekeeper of melanoma metastasis. *Frontiers in bioscience (Landmark edition)*, *17*(1), 1–15. https://doi.org/10.2741/3911

Das, S. K., Kegelman, T. P., Pradhan, A. K., Shen, X. N., Bhoopathi, P., Talukdar, S., Maji, S., Sarkar, D., Emdad, L., & Fisher, P. B. (2019). Suppression of Prostate Cancer Pathogenesis Using an MDA-9/Syntenin (SDCBP) PDZ1 Small-Molecule

Inhibitor. *Molecular cancer therapeutics*, *18*(11), 1997–2007. https://doi.org/10.1158/1535-7163.MCT-18-1019

Das, S. K., Maji, S., Wechman, S. L., Bhoopathi, P., Pradhan, A. K., Talukdar, S., Sarkar, D., Landry, J., Guo, C., Wang, X. Y., Cavenee, W. K., Emdad, L., & Fisher, P. B. (2020). MDA-9/Syntenin (SDCBP): Novel gene and therapeutic target for cancer metastasis. *Pharmacological research*, *155*, 104695. https://doi.org/10.1016/j.phrs.2020.104695

Friesner, R. A., Banks, J. L., Murphy, R. C., Halgren, T. A., Klicic, 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

Garcia, M., Hoffer, L., Leblanc, R., Benmansour, F., Feracci, M., Derviaux, C., Egea-Jimenez, A. L., Roche, P., Zimmermann, P., Morelli, X., & Barral, K. (2021). Fragment-based drug design targeting syntenin PDZ2 domain involved in exosomal release and tumour spread. *European journal of medicinal chemistry*, *223*, 113601. https://doi.org/10.1016/j.ejmech.2021.113601

Gangemi, R., Mirisola, V., Barisione, G., Fabbi, M., Brizzolara, A., Lanza, F., Mosci, C., Salvi, S., Gualco, M., Truini, M., Angelini, G., Boccardo, S., Cilli, M., Airoldi, I., Queirolo, P., Jager, M. J., Daga, A., Pfeffer, U., & Ferrini, S. (2012). Mda-9/syntenin is expressed in uveal melanoma and correlates with metastatic progression. *PloS one*, *7*(1), e29989. https://doi.org/10.1371/journal.pone.0029989

Geijsen, N., Uings, I. J., Pals, C., Armstrong, J., McKinnon, M., Raaijmakers, J. A., Lammers, J. W., Koenderman, L., & Coffer, P. J. (2001). Cytokine-specific transcriptional regulation through an IL-5Ralpha interacting protein. *Science (New York, N.Y.)*, *293*(5532), 1136–1138. https://doi.org/10.1126/science.1059157

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), 1375. https://doi.org/10.3390/ijms20061375

Grembecka, J., Cierpicki, T., Devedjiev, Y., Derewenda, U., Kang, B. S., Bushweller, J. H., & Derewenda, Z. S. (2006). The binding of the PDZ tandem of syntenin to target proteins. *Biochemistry*, *45*(11), 3674–3683. https://doi.org/10.1021/bi052225y

Grootjans, J. J., Zimmermann, P., Reekmans, G., Smets, A., Degeest, G., Dürr, J., & David, G. (1997). Syntenin, a PDZ protein that binds syndecan cytoplasmic domains. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(25), 13683–13688. https://doi.org/10.1073/pnas.94.25.13683

Guex, N., & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15), 2714–2723. https://doi.org/10.1002/elps.1150181505

Gurung, A. B., Ali, M. A., Lee, J., Farah, M. A., & Al-Anazi, K. M. (2021). An Updated Review of Computer-Aided Drug Design and Its Application to COVID-19. *BioMed research international*, *2021*, 8853056. https://doi.org/10.1155/2021/8853056

Hanson, R. M., & Lu, X. J. (2017). DSSR-enhanced visualization of nucleic acid structures in Jmol. *Nucleic acids research*, 45(W1), W528–W533. https://doi.org/10.1093/nar/gkx365

Hoffer, L., Garcia, M., Leblanc, R., Feracci, M., Betzi, S., Ben Yaala, K., Daulat, A. M., Zimmermann, P., Roche, P., Barral, K., & Morelli, X. (2023). Discovery of a PDZ Domain Inhibitor Targeting the Syndecan/Syntenin Protein-Protein Interaction: A Semi-Automated "Hit Identification-to-Optimization" Approach. *Journal of medicinal chemistry*, *66*(7), 4633–4658. https://doi.org/10.1021/acs.jmedchem.2c01569

Hughes, J. P., Rees, S., Kalindjian, S. B., & Philpott, K. L. (2011). Principles of early drug discovery. British journal of *pharmacology*, 162(6), 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x

Imjeti, N. S., Menck, K., Egea-Jimenez, A. L., Lecointre, C., Lembo, F., Bouguenina, H., Badache, A., Ghossoub, R., David, G., Roche, S., & Zimmermann, P. (2017). Syntenin mediates SRC function in exosomal cell-to-cell communication. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(47), 12495–12500. https://doi.org/10.1073/pnas.1713433114

Instituto Nacional de Estadística (2022, December 19). *Defunciones según causa de muerte: Año 2021 (datos definitives) y primer semestre de 2022 (datos provisionales)*. [Press release]. https://www.ine.es/prensa/edcm_2021.pdf

Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C., & Enjuanes, L. (2014). The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS pathogens*, *10*(8), e1004320. https://doi.org/10.1371/journal.ppat.1004320

Kang, B. S., Cooper, D. R., Jelen, F., Devedjiev, Y., Derewenda, U., Dauter, Z., Otlewski, J., & Derewenda, Z. S. (2003). PDZ tandem of human syntenin: crystal structure and functional properties. *Structure (London, England: 1993)*, *11*(4), 459–468. https://doi.org/10.1016/s0969-2126(03)00052-2

Kashyap, R., Roucourt, B., Lembo, F., Fares, J., Carcavilla, A. M., Restouin, A., Zimmermann, P., & Ghossoub, R. (2015). Syntenin controls migration, growth, proliferation, and cell cycle progression in cancer cells. *Frontiers in pharmacology*, *6*, 241. https://doi.org/10.3389/fphar.2015.00241

Kegelman, T. P., Das, S. K., Emdad, L., Hu, B., Menezes, M. E., Bhoopathi, P., Xu, J., Pellecchia, M., Sarkar, D., & Fisher, P. B. (2015). Targeting tumor invasion: the roles of MDA-9/Syntenin. *Expert Opinion on Therapeutic Targets*, *19*(1), 97–112. https://doi.org/10.1517/14728222.2014.959495

Kegelman, T. P., Wu, B., Das, S. K., Talukdar, S., Beckta, J. M., Hu, B., Emdad, L., Valerie, K., Sarkar, D., Furnari, F. B., Cavenee, W. K., Wei, J., Purves, A., De, S. K., Pellecchia, M., & Fisher, P. B. (2017). Inhibition of radiation-induced glioblastoma invasion by genetic and pharmacological targeting of MDA-

9/Syntenin. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 370–375. https://doi.org/10.1073/pnas.1616100114

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, 51(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956

Koes, D. R. (2016). Pharmit: interactive exploration of chemical space. *Nucleic Acids Research*, 44(W1), W442–W448. https://doi.org/10.1093/nar/gkw287

Koroll, M., Rathjen, F. G., & Volkmer, H. (2001). The neural cell recognition molecule neurofascin interacts with syntenin-1 but not with syntenin-2, both of which reveal self-associating activity. *The Journal of biological chemistry*, *276*(14), 10646–10654. https://doi.org/10.1074/jbc.M010647200

Kwon, Y. W., Jo, H., Bae, S., Seo, Y., Song, P., Song, M., & Yoon, J. H. (2021). Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery. *Frontiers in Medicine*, *8*. https://doi.org/10.3389/fmed.2021.747333

Latysheva, N., Muratov, G., Rajesh, S., Padgett, M., Hotchin, N. A., Overduin, M., & Berditchevski, F. (2006). Syntenin-1 is a new component of tetraspanin-enriched microdomains: mechanisms and consequences of the interaction of syntenin-1 with CD63. *Molecular and cellular biology*, *26*(20), 7707–7718. https://doi.org/10.1128/MCB.00849-06

Leblanc, R., Kashyap, R., Barral, K., Egea-Jimenez, A. L., Kovalskyy, D., Feracci, M., Garcia, M., Derviaux, C., Betzi, S., Ghossoub, R., Platonov, M., Roche, P., Morelli, X., Hoffer, L., & Zimmermann, P. (2020). Pharmacological inhibition of syntenin PDZ2 domain impairs breast cancer cell activities and exosome loading with syndecan and EpCAM cargo. *Journal of extracellular vesicles*, *10*(2), e12039. https://doi.org/10.1002/jev2.12039

Lin, J. J., Jiang, H., & Fisher, P. B. (1998). Melanoma differentiation associated gene-9, mda-9, is a human gamma interferon responsive gene. *Gene*, *207*(2), 105–110. https://doi.org/10.1016/s0378-1119(97)00562-3

Magalhães, P. R., Reis, P. B. P. S., Vila-Viçosa, D., Machuqueiro, M., & Victor, B. L. (2021). Identification of Pan-Assay INterference compoundS (PAINS) Using an MD-Based Protocol. *Methods in molecular biology (Clifton, N.J.)*, 2315, 263–271. https://doi.org/10.1007/978-1-0716-1468-6_15

Martinez, J. C., Ruiz-Sanz, J., Resina, M. J., Montero, F., Camara-Artigas, A., & Luque, I. (2023). A calorimetric and structural analysis of cooperativity in the thermal unfolding of the PDZ tandem of human Syntenin-1. *International journal of biological macromolecules*, *242*(Pt 1), 124662. Advance online publication. https://doi.org/10.1016/j.ijbiomac.2023.124662

Mir, C., Garcia-Mayea, Y., Garcia, L., Herrero, P., Canela, N., Tabernero, R., Lorente, J., Castellví, J., Allonca, E., García-Pedrero, J. M., Rodrigo, J. P., Carracedo, A., & LLeonart, M. E. (2021). SDCBP Modulates Stemness and Chemoresistance in Head and Neck Squamous Cell Carcinoma through Src Activation. *Cancers*, *13*(19), 4952. https://doi.org/10.3390/cancers13194952

Pantsar, T., & Poso, A. (2018). Binding Affinity via Docking: Fact and Fiction. *Molecules (Basel, Switzerland)*, *23*(8), 1899. https://doi.org/10.3390/molecules23081899

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein science: a publication of the Protein Society*, 30(1), 70–82. https://doi.org/10.1002/pro.3943

Philley, J. V., Kannan, A., & Dasgupta, S. (2016). MDA-9/Syntenin Control. *Journal of cellular physiology*, *231*(3), 545–550. https://doi.org/10.1002/jcp.25136

Pinto, G. P., Hendrikse, N. M., Stourac, J., Damborsky, J., & Bednar, D. (2022). Virtual screening of potential anticancer drugs based on microbial products. *Seminars in cancer biology*, *86*(Pt 2), 1207–1217. https://doi.org/10.1016/j.semcancer.2021.07.012

Pintor-Romero, V. G., Hurtado-Ortega, E., Nicolás-Morales, M. L., Gutiérrez-Torres, M., Vences-Velázquez, A., Ortuño-Pineda, C., Espinoza-Rojo, M., Navarro-Tito, N., & Cortés-Sarabia, K. (2023). Biological Role and Aberrant Overexpression of

Syntenin-1 in Cancer: Potential Role as a Biomarker and Therapeutic Target. *Biomedicines*, 11(4), 1034. https://doi.org/10.3390/biomedicines11041034

Pradhan, A. K., Maji, S., Bhoopathi, P., Talukdar, S., Mannangatti, P., Guo, C., Wang, X. Y., Cartagena, L. C., Idowu, M., Landry, J. W., Sarkar, D., Emdad, L., Cavenee, W. K., Das, S. K., & Fisher, P. B. (2021). Pharmacological inhibition of MDA-9/Syntenin blocks breast cancer metastasis through suppression of IL-1β. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(21), e2103180118. https://doi.org/10.1073/pnas.2103180118

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Folding & design*, 3(2), 141–147. https://doi.org/10.1016/s1359-0278(98)00019-4

Salmaso, V., & Moro, S. (2018). Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Frontiers in pharmacology*, *9*, 923. https://doi.org/10.3389/fphar.2018.00923

Schlander, M., Hernandez-Villafuerte, K., Cheng, C. Y., Mestre-Ferrandiz, J., & Baumann, M. (2021). How Much Does It Cost to Research and Develop a New Drug? A Systematic Review and Assessment. *PharmacoEconomics*, *39*(11), 1243–1269. https://doi.org/10.1007/s40273-021-01065-y

Shimada, T., Yasuda, S., Sugiura, H., & Yamagata, K. (2019). Syntenin: PDZ Protein Regulating Signaling Pathways and Cellular Functions. *International journal of molecular sciences*, *20*(17), 4171. https://doi.org/10.3390/ijms20174171

Stierand, K., & Rarey, M. (2007). From Modeling to Medicinal Chemistry: Automatic Generation of Two-Dimensional Complex Diagrams. *ChemMedChem*, 2(6), 853–860. https://doi.org/10.1002/cmdc.200700010

Tang, H., Wang, L., Li, S., Wei, X., Lv, M., Zhong, F., Liu, Y., Liu, J., Fu, B., Zhu, Q., Wang, D., Liu, J., Ruan, K., Gao, J., & Xu, W. (2023). Inhibitors against Two PDZ Domains of MDA-9 Suppressed Migration of Breast Cancer Cells. *International journal of molecular sciences*, *24*(4), 3431. https://doi.org/10.3390/ijms24043431

Tripathi, N. M., & Bandyopadhyay, A. (2022). High throughput virtual screening (HTVS) of peptide library: Technological advancement in ligand discovery. *European journal of medicinal chemistry*, *243*, 114766. https://doi.org/10.1016/j.ejmech.2022.114766

Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key Topics in Molecular Docking for Drug Design. *International journal of molecular sciences*, *20*(18), 4574. https://doi.org/10.3390/ijms20184574

Vázquez, J., López, M., Gibert, E., Herrero, E., & Luque, F. J. (2020). Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules (Basel, Switzerland)*, *25*(20), 4723. https://doi.org/10.3390/molecules25204723

What Is Cancer? (2021, October 11). National Cancer Institute. https://www.cancer.gov/about-cancer/understanding/what-is-cancer

Wermuth, C., Ganellin, C., Lindberg, P. & Mitscher, L. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry*, *70*(5), 1129-1143. https://doi.org/10.1351/pac199870051129

Yu, Y., Li, S., Wang, K., & Wan, X. (2019). A PDZ Protein MDA-9/Syntenin: As a Target for Cancer Therapy. *Computational and structural biotechnology journal*, *17*, 136–141. https://doi.org/10.1016/j.csbj.2019.01.002

Guillem Arasa Estivill
Universitat Rovira i Virgili

12. Supplementary Information

Table S 1 - Compunds reported in the bibliography

Origen	Compound	pIC_{50}	IC_{50} µM	K_d (µM)
Hoffer et al., 2023	118/SYNTi	6,398	0,4	
Hoffer et al., 2023	117	6,086	0,8	
Hoffer et al., 2023	116	5,824	1,5	
Hoffer et al., 2023	94	5,398	4,0	
Hoffer et al., 2023	95	5,377	4,2	
Hoffer et al., 2023	122	5,284	5,2	
Hoffer et al., 2023	91	5,276	5,3	
Hoffer et al., 2023	96	5,201	6,3	
Garcia et al., 2021	62	5,120	7,6	
Hoffer et al., 2023	98	5,114	7,7	
Hoffer et al., 2023	92	5,056	8,8	
Hoffer et al., 2023	97	5,004	9,9	
Hoffer et al., 2023	93	5,000	10,0	
Hoffer et al., 2023	85	4,886	13,0	
Hoffer et al., 2023	51	4,854	14,0	
Hoffer et al., 2023	52	4,699	20,0	
Hoffer et al., 2023	90	4,481	33,0	
Garcia et al., 2021	45	4,480	33,1	
Leblanc et al., 2020	SyntOff	4,456	35,0	
Hoffer et al., 2023	84	4,456	35,0	
Hoffer et al., 2023	89	4,357	44,0	
Garcia et al., 2021	57	4,330	46,8	
Hoffer et al., 2023	86	4,201	63,0	
Hoffer et al., 2023	29	4,174	67,0	
Garcia et al., 2021	61	4,160	69,2	
Hoffer et al., 2023	19/E5	4,119	76,0	
Garcia et al., 2021	58	4,110	77,6	
Hoffer et al., 2023	87	4,108	78,0	
Garcia et al., 2021	64	4,020	95,5	
Garcia et al., 2021	65	3,950	112,2	
Hoffer et al., 2023	82	3,893	128,0	
Garcia et al., 2021	60	3,750	177,8	
Hoffer et al., 2023	88	3,745	180,0	
Garcia et al., 2021	59	3,700	199,5	
Garcia et al., 2021	63	3,700	199,5	
Garcia et al., 2021	66	3,700	199,5	
PubChem	CID 164628214	3,699	200,0	
Hoffer et al., 2023	24	3,672	213,0	
Tang et al., 2023	PI2a			500
Tang et al., 2023	PI2b			610

Guillem Arasa Estivill
Universitat Rovira i Virgili

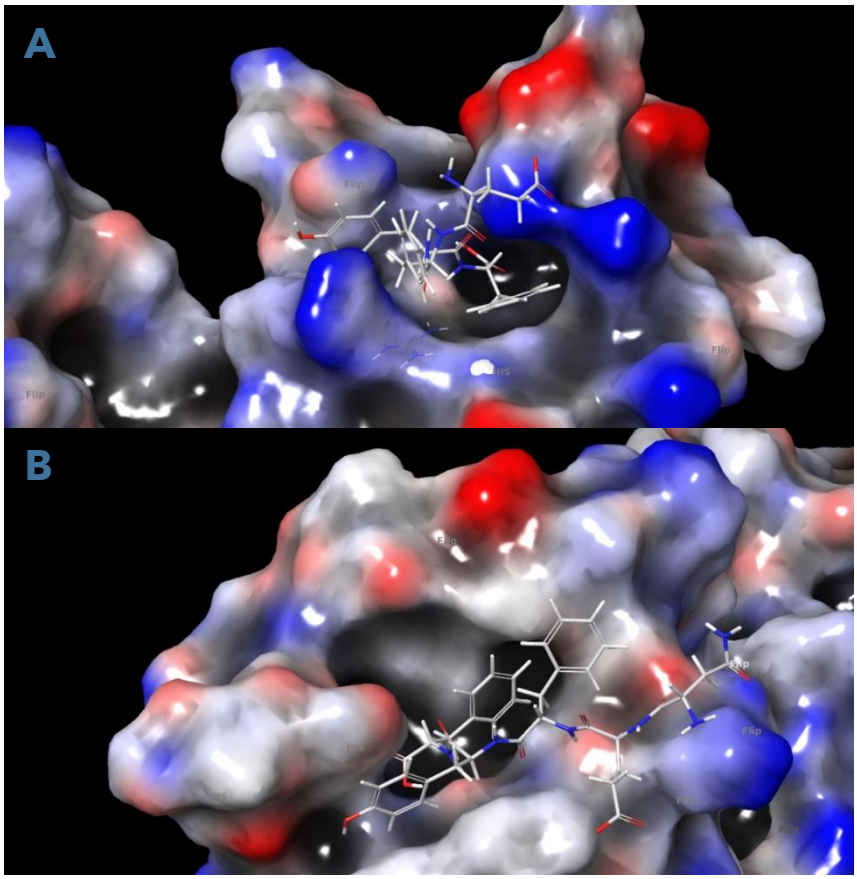

Figure S 1 - Comparison of the electrostatic potentials of PDZ1 (A) and PDZ2 (B)
Blue represents positive charges, red represent negative charges, grey represent neutral.

```
6r9h_2dom         194   PFERTITMHK DSTGHVGFIF K---NGKITS IVK-DSSAAR NGLLTEHNIC
6r9h_1dom         111    DPREVILCK DQDGKIGLRL KSIDNGIFVQ LVQANSPASL VGLRFGDQVL
                   *  . . * *  * .*. . *   ** .   .*  .*.*.   **     .

6r9h_2dom         240   EINGQNVIGL KDSQIADILS TS-GTVVTIT IMPA
6r9h_1dom         160   QINGENCAGW SSDKAHKVLK QAFGEKITMT IRD□R-
                 .***.*  *         .*   . *  .*.* *
```

Figure S 2 - Sequence alignment of PDZ domains

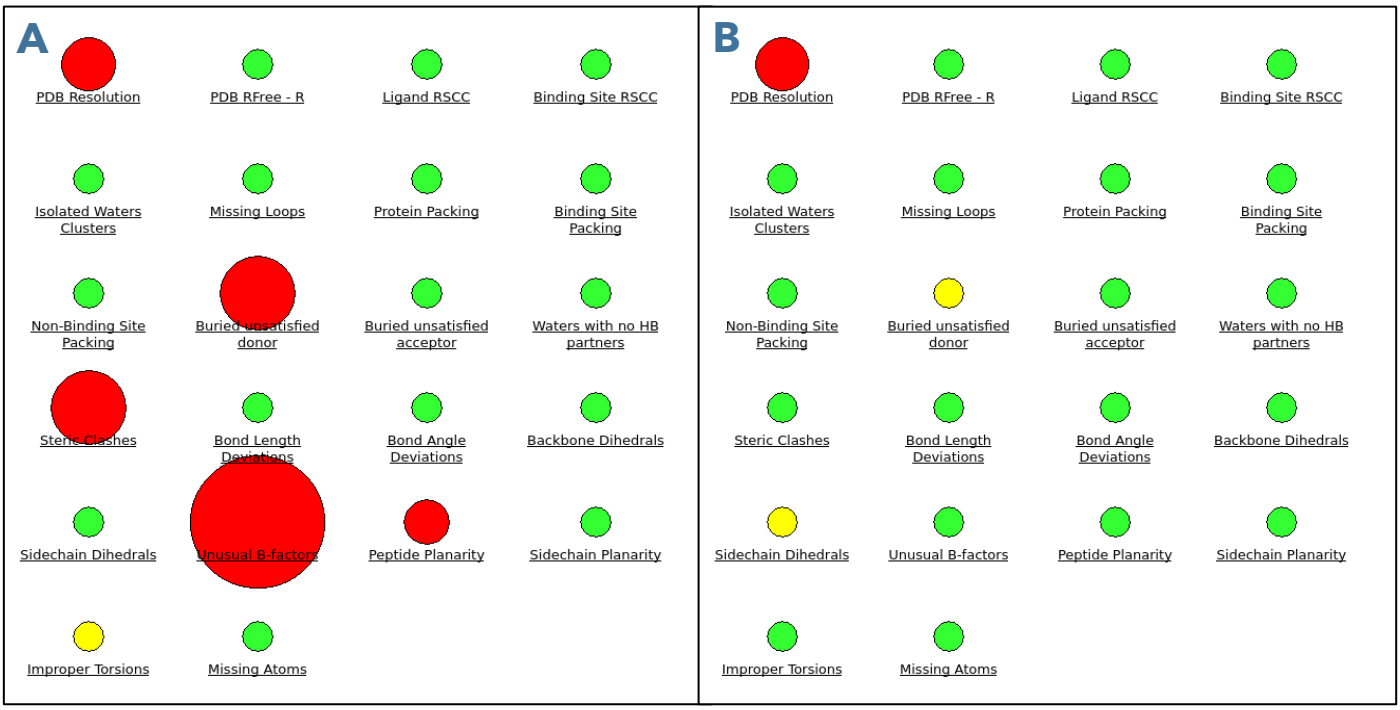

Figure S 3 - Protein Reliability Report of the 8AAI structure

A) Protein Reliability Report of the 8AAI structure prepared (4 chains)
B) Protein Reliability Report of the 8AAI structure prepared chain A

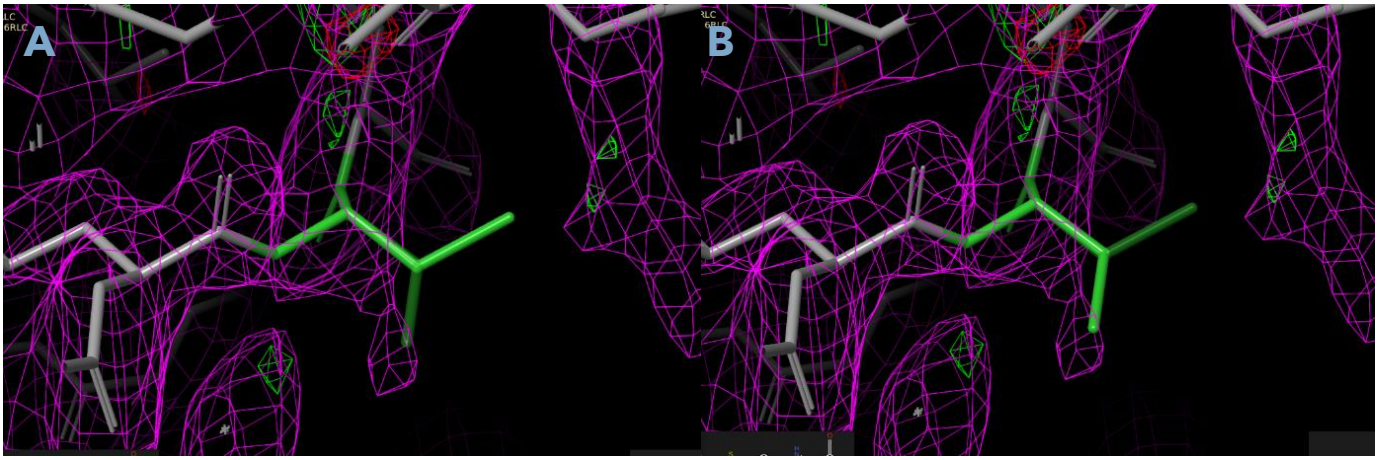

Figure S 4 - Possible residues of Cys118 of 6RLC_C:
A) Default position
B) Alternate position
Fo-Fc map is displayed on colours red and green; 2Fo-Fc is displayed on color purple

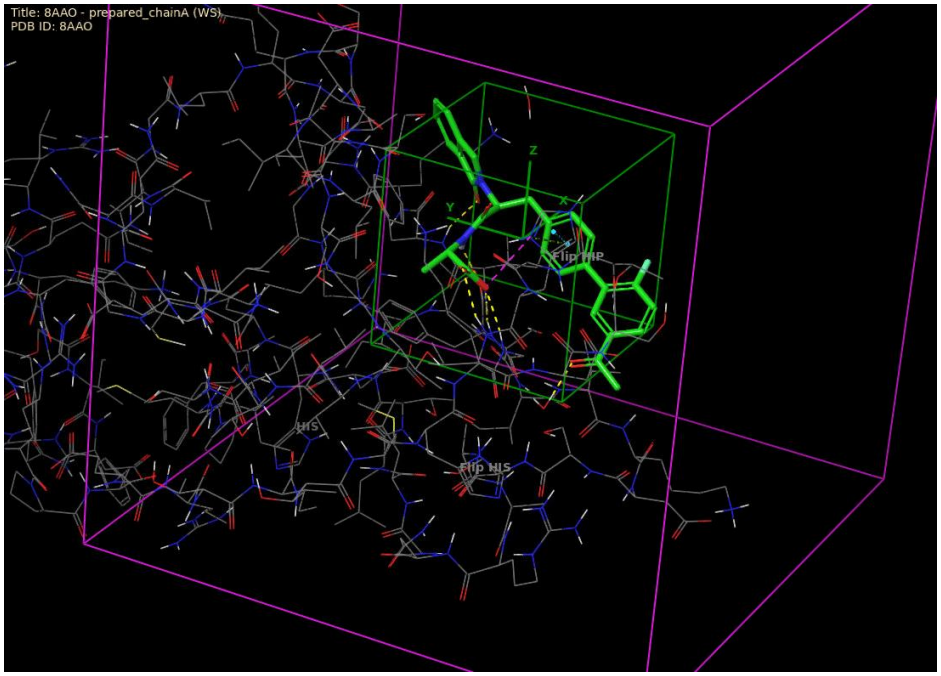

Figure S 5 - 8AAO grid generator
The purple box represents the grid; the green box represents the ligand position and approximate volume