



# Structural Bioinformatics: Methods development and applications to biomedical problems

Ph.D. course in Biochemistry  
XXXVI Cycle (2020-2023)

Candidate

Gianmarco Pascarella

Supervisor

Dr. Veronica Morea

Coordinator

Prof. Maria Luisa Mangoni

## Index.

Aim of the thesis.	5
1. Introduction.	7
1.1 Bioinformatics background.	7
1.2 Structural Bioinformatics.	8
1.3 Virtual Screening.	9
1.4 Homology Modelling.	10
1.5 Alphafold.	11
1.6 Protein Structural Alignment.	13
2. The $\sigma_1$ receptor as a target against neurodegenerative disease	14
2.1 Background	14
2.2 Aim of the work	17
2.3 Methods	18
2.3.1 Identification of known drugs able to bind $\sigma_1$ receptor and increase growth of Huntington disease patient-derived cells	18
2.3.2 Investigation of the entry pathway and molecular nature of $\sigma_1$ receptor ligands	20
2.4 Results	23
2.4.1 Identification of known drugs able to bind $\sigma_1$ receptor and increase growth of Huntington disease patient-derived cells	23
2.4.2 Investigation of the entry pathway and molecular nature of $\sigma_1$ receptor ligands	38
2.5 Conclusions	54

3. Computational study of the mode of COX binding of thiocanthal and thiocanthol, two novel biologically active compounds that can be obtained from extra virgin olive oil by eco-sustainable procedures.	61
3.1 Background	61
3.2 Aim of the work	64
3.3 Methods	64
3.4 Results	67
3.5 Conclusions	71
4. Identification of the molecular basis of the antiproliferative activity of <i>Arabidopsis thaliana</i> sirtuins	74
4.1 Background	74
4.2 Aim of the work	76
4.3 Methods	77
4.4 Results	78
4.5 Conclusions	86
5. Identification of the structural determinants underling the substrate specificity of the Arabidopsis N-acetyltransferase activity 2 protein	88
5.1 Background	88
5.2 Aim of the work	90
5.3 Methods	90
5.4 Results	93
5.5 Conclusions	104
6. Implementation of a pipeline to identify consensus results from protein structure alignment programs.	107
6.1 Background	107
6.2 Aim of the work	108
6.3 Methods	109

6.4 Results	110
6.5 Conclusions	118
7. General conclusions	121
8. Publications resulting from this thesis	125
References	126
Glossary	140
Appendix A	142
Appendix B	143

## Aim of the thesis.

In this thesis, I have tackled several biological and biomedical problems by structural bioinformatics methods, in collaboration with experimental groups:

- 1) Implementation a virtual screening (VS)-aided drug-repositioning procedure to investigate whether known drugs may be endowed with therapeutic activity against Huntington's disease (HD), a very severe neurodegenerative disease for which no therapy is currently available (Chapter 2 and Battista et al., 2021)
- 2) Identification of the potential physiological ligands of the  $\sigma_1$  receptor, an endoplasmic reticulum resident protein of great pharmacological interest since its agonists and antagonists have neuroprotective and analgesic activity, respectively (Chapter 2 and Pascarella et al., 2023).
- 3) Inference of the potential affinity and mode of binding towards COX-2 and COX-1 enzymes of thiocanthol and thiocanthal, two novel anti-inflammatory compounds that our collaborators obtained from olive oil precursors through eco-sustainable procedures (Chapter 3 and Di Risola et al, submitted for publication to Green Chemistry).
- 4) Identification of the sequence and structure determinants underlying the antiproliferative activity of *Arabidopsis thaliana* sirtuins, which has been demonstrated by our collaborators (Chapter 4 and Bruscalupi et al, 2023).
- 5) Identification of the structural determinants underlying the substrate specificity of the *Arabidopsis thaliana* N-acetyltransferase 2, whose kinetic and thermodynamic parameters have been determined by our experimental collaborators (Chapter 5 and Mattioli et al, 2022).

- 6) The development of an automated method to identify the consensus regions that are structurally aligned by different protein structure alignment methods, which usually give largely different results in the case of divergent protein structures (Chapter 6 and Pascarella et al, in preparation).

# 1. Introduction.

## 1.1 Bioinformatics background.

Bioinformatics is an interdisciplinary computational science involving informatics, biochemistry, physics, mathematics, and biology principles. The term “bioinformatics” appeared for the first time in the 1970s, thanks to Ben Hesper and Paulien Hogeweg who used it to define their research as “the study of informatic processes in biotic systems”. They started from the consideration that the defining properties of life was information processing in its various forms, like information accumulation during evolution, information transmission from DNA to intra- and intercellular processes, and the interpretation of such information at multiple levels and assumed that information processing could be a “useful metaphor for understanding living systems.” (Hogeweg 2011).

Despite two of the major bioinformatics databases were already established, namely GenBank, founded in 1982, and the Protein Data Bank (PDB), founded in 1973, only in the 1990s bioinformatics started to emerge as a science, both in Europe, with the “Bioinformatics in the 90s” conference, held in Maastricht in 1991, and in the USA, with the establishment of the National Center for Biotechnology Information (NCBI) (Attwood et al. 2011). This field of science began a rapidly and steady expansion from the 2000s onwards, due to the increasing amount of computational power available. According to Altman and Dugan, bioinformatics can be divided into two main branches of molecular biology. The first branch is based on the central dogma of molecular biology: DNA sequences are transcribed into mRNA sequences, and mRNA sequences

are translated into protein sequences. In this case, bioinformatics applications are used to transfer information at any stage of the central dogma, such as the organization and control of the genes in the DNA sequence, the identification of transcriptional units in DNA, the prediction of protein structure from sequence, and the analysis of molecular function. The second branch is based on the scientific method: formulation of the hypothesis, design of the experiments to test it, evaluation of the results and extension or modification of the hypothesis itself according to data. In this case bioinformatics applications are used to create systems to generate hypotheses, design experiments, store and organise information in databases, test the compatibility between data and models and remodulate hypotheses (Altman & Dugan 2005)

## 1.2 Structural Bioinformatics.

Structural bioinformatics is the branch of bioinformatics involved in the investigation, analysis, prediction, representation, and maintenance of information about biomolecular structures, such as protein structures and DNA complexes. For this reason, structural bioinformatics has two main objectives that are interrelated: the development of methods for manipulating information about biomolecular structures and the application of methods to solve biological problems (Altman & Dugan 2005). The large amount of data produced by genome-wide sequencing projects, high-throughput methods for expression analysis and the increase of structures stored in the PDB, is the new challenge of structural bioinformatics, which is focusing on emerging and increasingly promising techniques based on artificial intelligence (AI).



In this thesis we have pursued both of the objectives of structural bioinformatics: on the one hand, we have applied three of the most frequently used methods in structural bioinformatics, namely protein structure comparison, protein structure prediction and virtual screening, to solve biological problems; on the other hand, we have developed a method to improve the manipulation of protein structures.

### 1.3 Virtual Screening.

Virtual Screening (VS) is a computational approach to screen databases of small molecules (such as metabolites, drugs, etc.) to identify compounds with potentially desired biological activity against a given target. This technique can be used in drug discovery to screen a large number of molecules either already existing or even created from scratch with the aim of assembling a small dataset of potential candidates for further *in vitro* experiments (Lavecchia & Di Giovanni 2013). For example, with sufficient computational power, it is possible to screen all the conceivable molecules, as well as the existing ones, against a given target, to extract only the ones above a reasonable threshold, to reduce the number of molecules to be experimentally tested. This approach is a crucial alternative in pharmaceutical research, which usually utilises expensive and time-consuming processes.

VS is divided into two main categories: ligand-based (LBVS) and structure-based (SBVS). The first one utilises structure-activity data from a set of known active molecules, such as similarity and substructure searching, pharmacophore and three-dimensional shape matching, to create the dataset

useful in experimental methods. The second one, instead, uses the 3D structure of the target, experimentally or computationally determined, to perform ligands-target docking and rank them based on binding affinity predictions and/or complementarity with the binding site(Lavecchia & Di Giovanni 2013). In general, in a SBVS simulation there are three fundamental parameters: the identification of a valid target; the construction of a valid and robust dataset of ligands; and the correct definition of the binding site. Indeed, if one or more of these three parameters is not accurate the results might not be representative of the real situation. Typical errors include target structures determined with low resolution (i.e., structure not well determined experimentally); a small set of ligands; or ligands with unfavourable properties related to the binding site, for example too big or too small with respect to the binding site.

#### 1.4 Homology Modelling.

Homology modelling is a technique utilised to build the structure of a protein in the absence of experimental information except the amino acid sequence. This computational approach is based on previous knowledge, derived from both experimental work and database analyses, that i) protein 3D structures are determined by their amino acid sequences and that ii) proteins with similar sequences have similar structures. Currently, in the protein structure database (PDB) (Berman et al. 2000) more than 200,000 structures are registered, while the number of protein sequences is about 220 millions, according to Blastp (Camacho et al. 2009) info box. Due to this discrepancy, computational methods capable of transferring structural information between proteins are

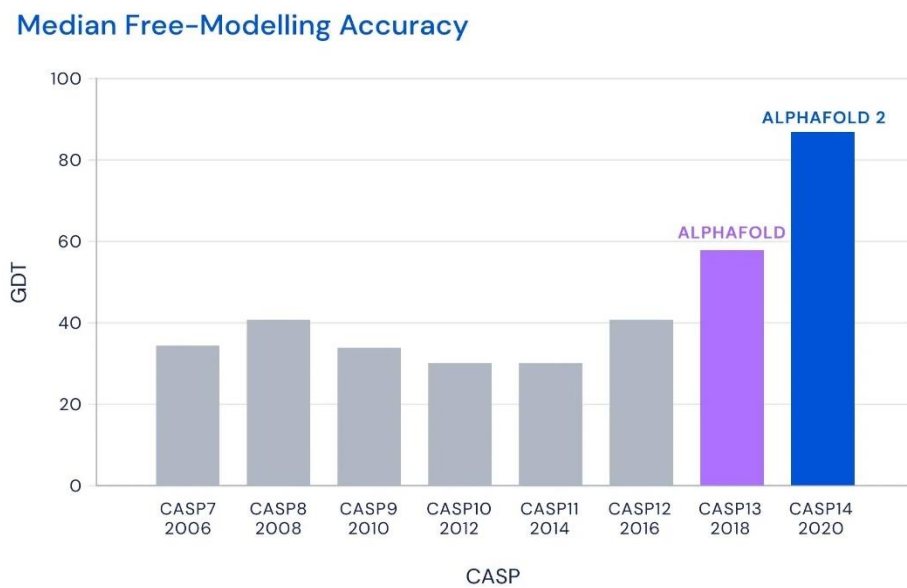
essential tools. Among all methods to predict protein structures, homology modelling has long been considered to be the most accurate, as well as time- and cost-effective, and it has been widely used to generate 3D structures of proteins employed in virtual screening and docking simulations (Muhammed & Aki-Yalcin 2019).

This type of prediction is based on multiple steps: I) research of a structure (template) in the PDB database that is expected to be similar to that of the protein to model (target); II) Initial target-template alignment and subsequential refinement; III) Transfer of template coordinates to the target to generate the backbone; IV) Loop and side-chain modelling; V) Model optimization and VI) Model validation (Krieger et al. 2005).

## 1.5 Alphafold.

In the last few years, Artificial Intelligence (AI) and machine learning have emerged as powerful resources and have been implemented in almost all fields of human knowledge, ranging from internet browsers to vehicles, and in scientific research. The prediction of 3D protein structures gained a particular benefited from the developments in this field. Indeed, in 2018, DeepMind, a Google subsidiary, developed the first version of Alphafold (AF) (Senior et al. 2020), a method based on Artificial Intelligence. AF participated in the 13<sup>th</sup> edition of the Critical Assessment of Structure Prediction (CASP) experiment (Kryshtafovych et al. 2019), which is a community-wide experiment to assess the results of protein structure prediction methods in blind tests, and ranked first among all programmes. The improvement of AF compared to other

methods was significant, in that not only it performed way better than the best methods of CASP12, but there was an average of 15% better accuracy over other groups of CASP13. In the following CASP (CASP14, 2020) DeepMind introduced the second version of AF, Alphafold2, which further improved performance, by about 70%, compared to the first version (Figure 1.1), as measured by Global Distance Test (GDT) values (GDT indicates the percentage of atoms of two structures whose distance is lower than a specified threshold, see Appendix A).



*Fig.1.1. The graph shows the average GDT value of the best programme for each CASP edition.*

## 1.6 Protein Structural Alignment.

This approach aims at finding structural similarities between two or more proteins. This is useful for example for evolutionary inferences, since structure is more conserved than sequence during evolution, therefore proteins with similar structures can be evolutionarily related even if their sequences have diverged beyond the ability of current sequence comparison methods to recognize these relationships. Additionally, structural alignments are useful to obtain more accurate sequence alignment than those provided by sequence comparison methods and, therefore, to recognize homology at a residue level, as well as at a global protein level. Finally, structure alignments are useful for functional inferences, since protein function, in addition to structure, is often conserved during evolution. Currently available protein structure alignment methods are widely different from one another in the following features. i) The protein region that they try to align: for example, some programmes only consider C $\alpha$ -C $\alpha$  atoms, others the whole main-chain atoms. ii) The type of output that they provide: the oldest and the simplest programs return only a global structure similarity value, while others also provide a structure alignment in terms of coordinate superposition and/or sequence alignments corresponding to the structure superposition. iii) The type of scoring function that they utilise: the most frequently used parameter to evaluate structure similarity is root-mean square deviation (RMSD, see Appendix A), while other methods use other scoring functions, such as GDT-TS (see Appendix A), in addition, or as an alternative, to RMSD.

## 2. The $\sigma$ 1 receptor as a target against neurodegenerative disease

### 2.1 Background.

The sigma-1 receptor ( $\sigma$ 1R) is an intracellular receptor expressed in central nervous system regions and known to regulate calcium signalling and cell protection. The  $\sigma$ 1R gene encodes a 24 kDa protein of 223 amino acids, which is anchored to the endoplasmic reticulum and plasma membranes. The human  $\sigma$ 1-R is implicated in a variety of diseases affecting the CNS, such as depression, drug addiction and neuropathic pain (Schmidt et al. 2016). Interestingly,  $\sigma$ 1-R is an evolutionary isolate with no discernible similarity to other proteins. The only known  $\sigma$ 1-R homologue is the yeast C-8 sterol isomerase ERG2. ERG2 is one of the proteins involved in the biosynthesis of ergosterol, which is essential to modulate fungal cell membrane fluidity, like cholesterol does in animal cells. In particular, ERG2 catalyses the reaction that shifts the delta-8 double bond to delta-7 position in the B ring of sterols, thereby converting fecosterol to episterol (Jordá & Puig 2020). The endogenous ligand of this receptor is still unknown, although N,N-dimethyl tryptamine, sphingosine, and myristic acid have been proposed to act as  $\sigma$ 1R endogenous modulators (Chu & Ruoho 2016). Many molecules have been shown to have agonist or antagonist activity, based on their ability to recapitulate the phenotype of receptor overexpression or knockdown, respectively (Nguyen et al. 2015). Agonists have been associated with

cytoprotective activity (Mancuso et al. 2012; Maher et al. 2018). Antagonists have analgesic effect in both animals and humans (Hayashi et al. 2011; Castany et al. 2018). Additionally,  $\sigma$ 1R antagonists potentiate signalling by G-protein coupled receptors (Kim et al. 2010; Navarro et al. 2010) whereas agonists increase IP3-dependent calcium flux, at least with IP3 receptors type 3 (Wu & Bowen 2008), and inhibit sodium and potassium channel current (Aydar et al. 2002). While the detailed mechanism of  $\sigma$ 1R activity has not been unveiled yet, modulation of  $\sigma$ 1R activity by agonists has been shown to significantly attenuate oxidative stress, neuroinflammation, hypoxia, apoptotic pathways, and other processes caused by neurodegenerative diseases such as Huntington disease (HD) (Geva et al. 2016), Alzheimer's disease (AD) (Ryskamp et al. 2019), Parkinson's Disease (PD) (Francardo et al. 2019) and amyotrophic lateral sclerosis (ALS) (Ionescu et al. 2019). The 3D structures of the human  $\sigma$ 1-R receptor (Hs $\sigma$ 1-R) have been experimentally determined by X-ray crystallography in two different works conducted by the same group. In the first work, the structure of  $\sigma$ 1-R was determined in complex with PD144418 (PDB ID: 5HK1), which has antagonist activity, and 4-IBP (PDB ID: 5HK2), whose activity has not been clearly classified (Schmidt et al. 2016). In the second work, the X-ray crystal structure of  $\sigma$ 1R in complex with two additional antagonists, namely haloperidol and NE-100, and the (+)-pentazocine agonist, has also been determined. Intriguingly, the overall structure of  $\sigma$ 1R did not change significantly upon agonist vs. antagonist binding. In addition, the mechanism of access to the binding site was investigated by molecular dynamics (MD) simulations (Schmidt et al. 2018).

Accelerated MD simulations (aMD) of the Hs $\sigma$ 1-R monomer inserted in a hydrated lipid bilayer were performed in four different conditions (Schmidt et

al. 2018): (i) on the (+)-pentazocine-bound structure (PDB ID: 6DK1); (ii) on the haloperidol-bound structure (PDB ID: 6DJZ); (iii) in condition (i) except for the absence of the ligand; and (iv) in condition (i) with the ligand placed in water at a distance  $> 10 \text{ \AA}$  from the protein. As a result of these simulations, it was proposed that major conformational rearrangements of the protein should take place to make the binding site accessible to the ligand. The first of these rearrangements is the opening of the 'lid' of the  $\beta$ -barrel, following disruption of the backbone hydrogen bonds between W136 in  $\beta$ -strand 6 and A161 in  $\beta$ -strand 9. This lid, comprising  $\beta$ -strands 6 and 7 and the loop that connects them, points toward the aqueous medium and is the region of the  $\beta$ -barrel farthest from the ligand. Then, the backbone hydrogen bonds between E123 and R175 are broken and the  $\beta$ -strands 5 and 10, where these residues are located, respectively, separate from each other, thus exposing the binding pocket. Finally, the ligand enters the binding site and assumes a position similar (i.e.,  $\text{RMSD} < 3.0 \text{ \AA}$ ) to that observed in the crystal structure (Schmidt et al. 2016; Schmidt et al. 2018). Two pathways were investigated: (i) pathway 1, directed towards the aqueous solvent, through a polar region occluded by Q135, H154 and E158 in  $\beta$ -strands 6 and 8 and in the loop between  $\beta$ -strands 8 and 9, respectively; and (ii) pathway 2, directed towards the membrane, through the  $\alpha$ -helices 4 and 5 that are in contact with the membrane. Based on the magnitude of the force and the time required for the ligand to be completely dissociated from the protein, the pathway connecting the ligand binding site with the aqueous milieu was proposed to be the most likely ligand access route, in agreement with the results of the previously performed MD studies (Schmidt et al. 2016; Schmidt et al. 2018). This hypothesis implies that, to reach the binding site, the ligand would initially interact with the polar residues Q135,



H154 and E158 even if, based on lipid/water partition coefficients, the three compounds studied in this work would preferentially be associated with lipid environments (Rossino et al. 2020). In contrast with the results of MD simulations, the results of a recent work performed on the  $\sigma$ 1-R homologue from *Xenopus laevis* (Xl $\sigma$ 1-R) indicate that the ligand is more likely to enter the binding site from the membrane side (pathway 2), thanks to conformational changes determining an opening between the  $\alpha$ 4 and  $\alpha$ 5 helices, rather than from the aqueous medium (pathway 1), following major structural rearrangements of the cupin-fold domain. Interestingly, although the structures in coordinate files 7W2B (“closed” conformation) and 7W2E (“open-like” conformation) were solved in the putative apo-form, an electron density peak was identified in proximity to the Xl $\sigma$ 1-R binding site in the Fo-Fc electron density map of both structures (Meng et al. 2022).

## 2.2 Aim of the work.

In these works, we established several objectives to achieve. In the first work, we performed a virtual screening (VS) of a library of compounds approved by the FDA for clinical use. In order to identify one or more already known drugs with the ability to potentially bind  $\sigma$ 1R. The selected compounds were analysed, by our collaborators, to verify in vitro whether the predicted data were supported by experimental data. In the second work, we tried to contribute to the elucidation of the entrance mechanism of the ligand in the binding site. We, also, tried to identify the physiological ligand with the auxilium of virtual screening on a large number of compounds (comprising

human metabolites, sterols, etc.). Thanks to our collaborator, we also analysed hypotheses about the possible entry modes of ligands in the binding site with molecular dynamics (MD) starting from the data produced by our VS.

## 2.3 Methods.

### 2.3.1 Identification of known drugs able to bind $\sigma_1$ receptor and increase growth of Huntington disease patient-derived cells.

The atomic coordinates of all the  $\sigma_1$ R 3D structures were downloaded from the PDB (<https://www.rcsb.org/>) (Berman et al. 2000). The structures were visual inspected with Chimera (Pettersen et al. 2004). The structure in coordinate files 5HK1, which has been solved with the highest resolution (2.51 Å) (Schmidt et al. 2016), was used to perform virtual screening. Crystallographic waters and ligand molecules were removed using Chimera. AutoDock Tools (ADT) v. 1.5.6 was used to add hydrogen atoms, merge non-polar hydrogen atoms and automatically assign Gasteiger charges.

The ligand dataset was composed by: I) 1576 compounds FDA approved and available for sale from the ZINC15 database (<http://zinc15.docking.org/>) (Sterling & Irwin 2015); II) ligands coming from literature, which are proved to bind  $\sigma_1$ R; III) ligands present in the available  $\sigma_1$ R 3D structures. The FDA compounds from Zinc were converted to PDB format in two steps: first, the SMILE format was converted to mol2 with MarvinSketch 18.26 (<https://chemaxon.com>); second, the mol2 format was converted to PDB using Open Babel v. 2.3.1 (O'Boyle et al. 2011). All ligands were then converted in

pdbqt format using a script from ADT 1.5.6 (Morris et al. 2009) using the additional parameters: “-A ‘hydrogen\_bonds’” to add hydrogens and build bonds among non-bonded atoms; and “-U ‘nphs\_lps’” to merge both non-polar hydrogens and lone pairs.

Virtual screening was performed with the program Autodock Vina (Trott & Olson 2009) and using the following parameters for the binding site: spacing value at 0.375 Å; center on coordinates 12.168, 36.423 and -34.778; and  $30 \times 24 \times 34$  grid points. The Vina parameters were been: “-num\_modes 100”, which represents the maximum number of binding modes to generate; and “-energy\_range 9”, in order to maximize the energy difference between the best binding mode and the worst one and keeping all other values as default, as Vina parameters. After virtual screening, the 20 hits with the best Vina score were extracted using the “vina\_screen\_get\_top.py” script from Autodock Vina tools (Trott & Olson 2009).

The 20 lowest energy results from virtual screening against 5HK1\_A were docked again into the same monomer using ADT. The same space searching parameters reported for virtual screening were used, plus: 100 genetic algorithm runs; population size of 150; RMS cluster tolerance of 2 Å. All other parameters were left as default.

Python scripts were developed to extract the best 20 from the virtual screening results, ranked by energy, and to perform a preliminary analysis on these compounds (Pettersen et al. 2004). The analysis comprised the reconstruction of the complexes receptor-ligand, the extraction from the VS results of the energy values, and a calculation of protein-ligand interactions such as hydrogen bonds, number of contacts and number of unfavourable interactions (clashes). The script utilized to extract this information was ran with the

Chimera bash interface. Scripts interfacing with Chimera were written in Python version 2.7, required by Chimera. All other scripts were written in Python version 3.6. The 20 complexes protein-ligand were also visually inspected with the program PyMol (<https://www.pymol.org/>).

### 2.3.2 Investigation of the entry pathway and molecular nature of $\sigma 1$ receptor ligands.

The 3D structures of human and *X. laevis*  $\sigma 1$ -R used in this work were downloaded from the Protein Data Bank (PDB: <https://www.rcsb.org/>; accessed on 18 March 2022) (Berman et al. 2000).

To perform the VSs, different structures were selected: I) 5HK1, determined in complex with the PD144418 antagonist,

which has been solved with the highest resolution (2.51 Å); II) 6DK1, which is the only available Hs $\sigma 1$ -R structure that has been determined in complex with a classic agonist, namely (+)-pentazocine (Schmidt et al. 2018); III) 7W2E and 7W2B from *Xenopus laevis*  $\sigma 1$ -R structures, that both shown a not yet identified ligand in the binding site (Meng et al. 2022), while the 7W1B structure was determined in a “closed” conformation, like that observed in Hs $\sigma 1$ -R structures, the 7W1E structure displayed an “open-like” conformation, where the  $\alpha 4$  helix rotates slightly away from  $\alpha 5$ , thus enlarging an opening that may allow ligand entry; iv) the coordinates of ERG2 model were downloaded from AlphaFold protein database (Jumper et al. 2021).

Structure visualization and analysis were performed using the programs InsightII (Dayringer et al. 1986), Swiss-PDBViewer (Guex & Peitsch 1997),

Chimera (Pettersen et al. 2004) and PyMol (<https://www.pymol.org>). For each of these structures, we selected chain A, we removed the water molecules and all others molecule present in the coordinates files. AutoDock Tools (ADT) v.1.5.6 was used to add hydrogen atoms, merge non-polar hydrogen atoms and automatically assign Gasteiger charges.

The dataset utilised in the VS against the Hs $\sigma$ 1-R structure in coordinate files 5HK1\_A or 6DK1\_A or against yeast ERG2 model was constructed with the following compounds: I) molecules from ZINC15 database tagged as “metabolites” (15,871 compounds); II) molecules tagged as “FDA approved” or “World not-FDA” (1538 and 3192 compounds, respectively), which have been approved as drugs by the FDA or other regulatory agencies, excluding redundant molecules; III) Due to the demonstrated ability of neurosteroids to act as  $\sigma$ 1-R agonists or antagonists (Maurice & Su 2009), and the fact that the homologous ERG2 protein present in fungi binds steroid-based compounds, we selected all the compounds belonging to the sterol\_lipids, sterols, steroids, androgens, estrogens, cholesterol or ergosterol biosynthetic pathway categories and available for sale from the LIPID MAPS database (<https://www.lipidmaps.org/>; accessed on 1 July 2021) (Sud et al. 2007), comprising 3761, 1593, 362, 101, 63, 442 and 347 compounds, respectively; IV) molecules derived from literature that showed binding affinity for  $\sigma$ 1-R and/or  $\sigma$ 2-R, to use as positive controls. The final comprehensive dataset comprised 21,359 non-redundant compounds (several compounds are present in more than one of the listed categories). The structures of these compounds were downloaded from ZINC, whenever available, or from the ChEbi (Hastings et al. 2016) or PubChem (Bolton et al. 2011) databases.

For VS against  $Xl\sigma 1$ -R A chain monomers in coordinate files 7W2E or 7W2B (i.e., 7W2E\_A and 7W2B\_A, respectively), or against the yeast ERG2 model, we used a 1332 molecule dataset obtained from the yeast metabolome database (YMDB)(Ramirez-Gaona et al. 2017) by eliminating all compounds whose molecular weight was higher than 400 Da. This was done to speed up the VS procedure, since a preliminary visual inspection of the density map in the ligand binding site of 7W2E\_A and 7W2B\_A had revealed that the unknown  $Xl\sigma 1$ -R ligand in these two apo structures was not larger than a cholesterol molecule, which has a molecular weight of about 387 Da.

All compounds present in the selected datasets were converted: (i) from the .smi or .sdf to the PDB format, using an ad hoc developed bash script that included the “molconvert” command from MarvinSketch v18.26 (<https://chemaxon.com>, accessed on 1 July 2021); (ii) from the PDB to the pdbqt format, using a script from AutoDock Tools v1.5.6 (ADT) (Morris et al. 2009) where the following parameters were added: “-A ‘hydrogen\_bonds’” to both add hydrogens and build bonds among non-bonded atoms; and “-U ‘nphs\_lps’” to merge both non-polar hydrogens and lone pairs.

For VS against  $Hs\sigma 1$ -R structures in coordinate files 5HK1\_A and 6DK1\_A, the following space searching parameters were adopted: spacing value at 0.375 Å; centre on coordinates 12.168, 36.423 and -34.778; and  $30 \times 24 \times 34$  grid points. For virtual screening against  $Xl\sigma 1$ -R monomers 7W2E\_A and 7W2B\_A, only coordinates and dimension of the binding pocket were changed: centre on coordinates -31.000, -26.000 and 34.000; and  $12 \times 12 \times 12$  grid points. VS was performed using the program VINA (Trott & Olson 2009) and the same parameters for all four  $\sigma 1$ -R structures and the ERG2 model, namely: “-num\_modes 100”, which represents the maximum number

of binding modes to generate; and “-energy\_range 9”, in order to maximize the energy difference between the best binding mode and the worst one. For each VS, was used the “vina\_screen\_get\_top.py” script from AutoDock Vina tools (Trott & Olson 2009) to extract the best 20 ligands from the VS results. Previously developed Python scripts were used to parse VINA output files and perform a preliminary analysis of the selected ligand–receptor complexes. In particular, the pose energy of each ligand was extracted from the VINA pdbqt file; features of protein–ligand interactions such as hydrogen bonds, number of contacts and number of unfavourable interactions (clashes) were calculated by the structure visualization and analysis program Chimera, following rebuilding of receptor–ligand complexes. Information on the clinical indication and mechanism of action of each compound was manually obtained from KEGG (Kanehisa & Goto 2000) and DrugBank (Wishart et al. 2018).

## 2.4 Results.

### 2.4.1 Identification of known drugs able to bind $\sigma_1$ receptor and increase growth of Huntington disease patient-derived cells.

All the 3D structures of  $\sigma_1$ R experimentally determined by X-ray crystallography and available from the Protein Data Bank (PDB: <https://www.rcsb.org/>) (Berman et al. 2000) contain quaternary assemblies of three monomers, each comprising an N-terminal transmembrane helix and C-terminal ligand binding domain.

Structure comparisons indicate that the ligand binding domain, which comprises the whole ligand binding site, is highly conserved in all of the monomers present in the different structures, as previously reported by the authors (Schmidt et al. 2016; Schmidt et al. 2018) and shown by the low root mean square deviation (RMSD), values reported in Table 2.1, which are a measure of structural difference based on the distance between equivalent pairs of atoms, in Å. Therefore, the monomer chosen for virtual screening, i.e., chain A in coordinate files 5HK1 (5HK1\_A), which is the 3D structure determined with the highest resolution, adequately represents the structures of other monomers.

*Table 2.1. All-against-all structure comparison of  $\sigma 1R$  ligand binding domain in the available 3D structures determined by X-ray crystallography. The ligand binding domain comprises residues 29-212 in 5HK2 chain B and 35-218 in all other chains. RMSD values calculated after optimal all-against-all pairwise superposition of either Ca or all atoms are reported in the lower and upper part of the matrix, respectively.*

		5HK1			5HK2			6DJZ			6DK0			6DK1		
		A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
5HK1	A	-	0.34	0.45	0.37	0.44	0.28	0.14	0.36	0.46	0.17	0.37	0.47	0.38	0.58	0.53
	B	0.66	-	0.31	0.31	0.33	0.36	0.35	0.14	0.34	0.37	0.18	0.35	0.49	0.44	0.44
	C	0.77	0.60	-	0.48	0.19	0.49	0.46	0.34	0.14	0.49	0.37	0.19	0.60	0.51	0.42
5HK2	A	0.73	1.23	0.79	-	0.46	0.33	0.37	0.33	0.51	0.39	0.34	0.52	0.50	0.55	0.54
	B	0.76	0.61	0.49	0.79	-	0.47	0.44	0.35	0.20	0.47	0.37	0.20	0.55	0.50	0.40
	C	0.59	0.69	0.66	0.70	0.81	-	0.28	0.37	0.51	0.28	0.38	0.52	0.44	0.58	0.56
6DJZ	A	0.18	0.67	0.78	0.70	0.77	0.60	-	0.35	0.46	0.11	0.36	0.46	0.35	0.57	0.52
	B	0.66	0.18	0.62	0.65	0.62	0.68	0.66	-	0.34	0.37	0.11	0.35	0.48	0.41	0.43
	C	0.72	0.62	0.19	0.60	0.50	0.78	0.72	0.62	-	0.49	0.36	0.14	0.58	0.50	0.39
6DK0	A	0.26	0.69	0.26	0.75	0.80	0.61	0.22	0.67	0.75	-	0.37	0.49	0.33	0.56	0.52
	B	0.66	0.22	0.64	0.66	0.65	0.70	0.67	0.16	0.64	0.68	-	0.37	0.48	0.40	0.43
	C	0.73	0.62	0.24	0.83	0.50	0.80	0.73	0.63	0.17	0.75	0.64	-	0.56	0.46	0.36
6DK1	A	0.55	0.82	0.92	0.83	0.88	0.75	0.53	0.80	0.87	0.52	0.80	0.86	-	0.53	0.47
	B	0.93	0.67	0.83	0.84	0.80	0.98	0.92	0.61	0.79	0.93	0.63	0.77	0.83	-	0.40
	C	0.89	0.79	0.61	0.90	0.71	0.91	0.89	0.78	0.57	0.89	0.78	0.55	0.81	0.79	-



The chemical formulas, clinical indication and information on the mechanism of action of the 20 virtual screening hits having the best predicted interaction energy with 5HK1\_A are reported in Figure 2.1. The interaction of these 20 FDA drugs with  $\sigma$ 1R was further investigated by computational docking to 5HK1\_A binding site. This step was performed because the free binding energy predicted by docking methods has an accuracy of  $\sim$ 2–3 kcal/mol standard deviation (Huey et al. 2007), therefore ranking of poses based on this parameter alone is not reliable. Conversely, highly populated clusters have been shown to be enriched in compounds that show strong binding in experimental tests (Cosconati et al. 2010). The poses of the 20 FDA drugs having the best energy among those in the largest clusters produced by molecular docking, were visually analysed using the Chimera program (Pettersen et al. 2004). To evaluate the likelihood of ligand-receptor binding, we took into account: i. Parameters calculated by Chimera, such as the number of hydrogen bonds and non-polar interactions, and number of unfavourable Van der Waals contacts, if present; and ii. Whether ligand moieties comprised in previously reported pharmacophore models (i.e., the positively charged group and hydrophobic regions) were involved among the aforementioned interactions.

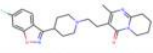

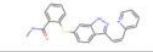
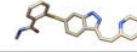
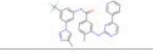
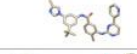
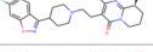
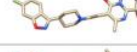
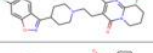

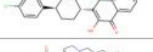



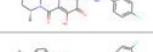

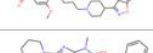

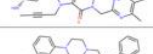
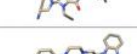
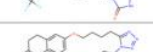




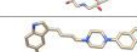



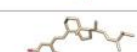
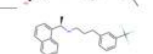
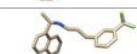
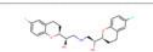
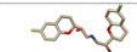
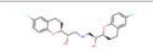
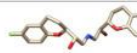
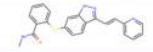

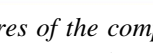

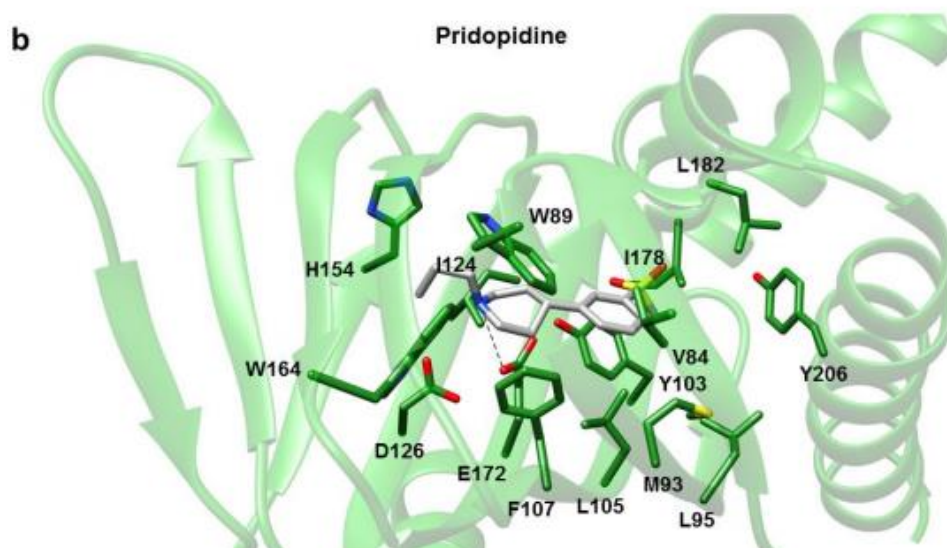
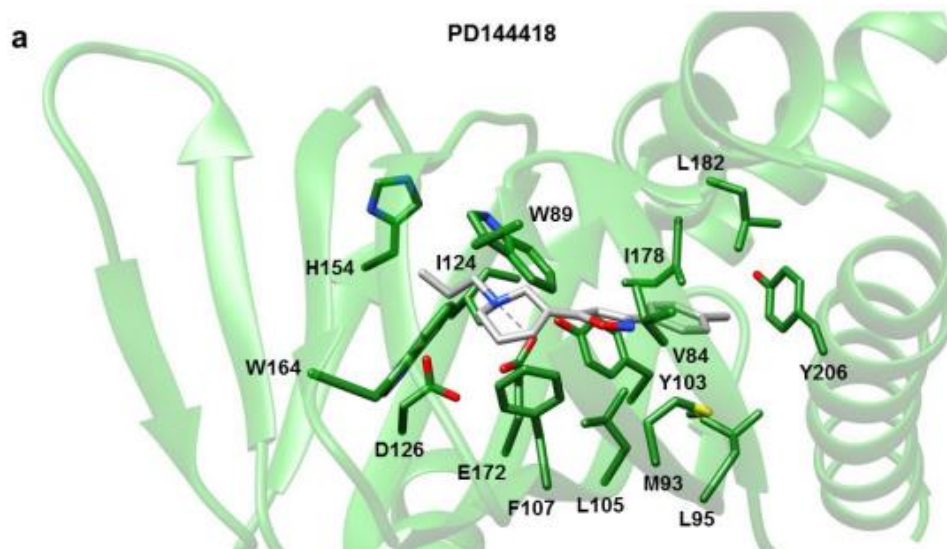
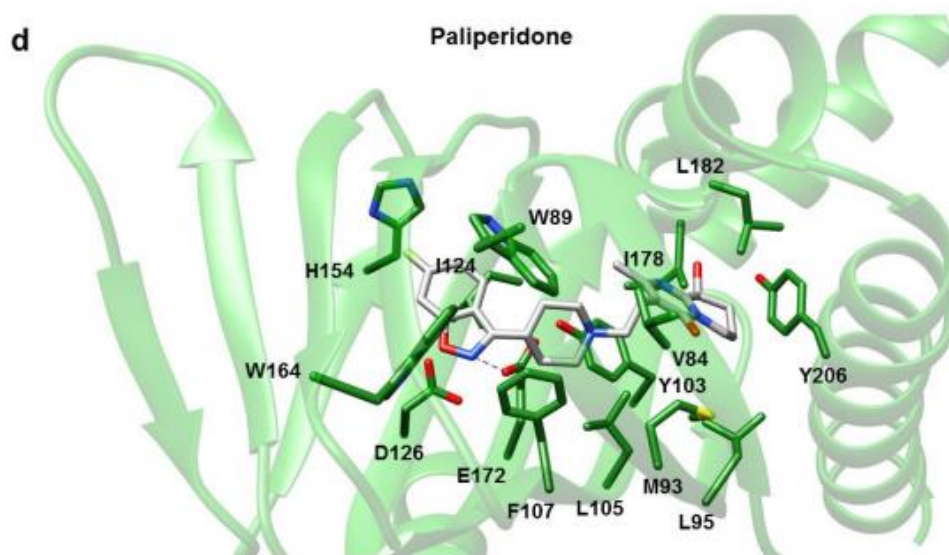
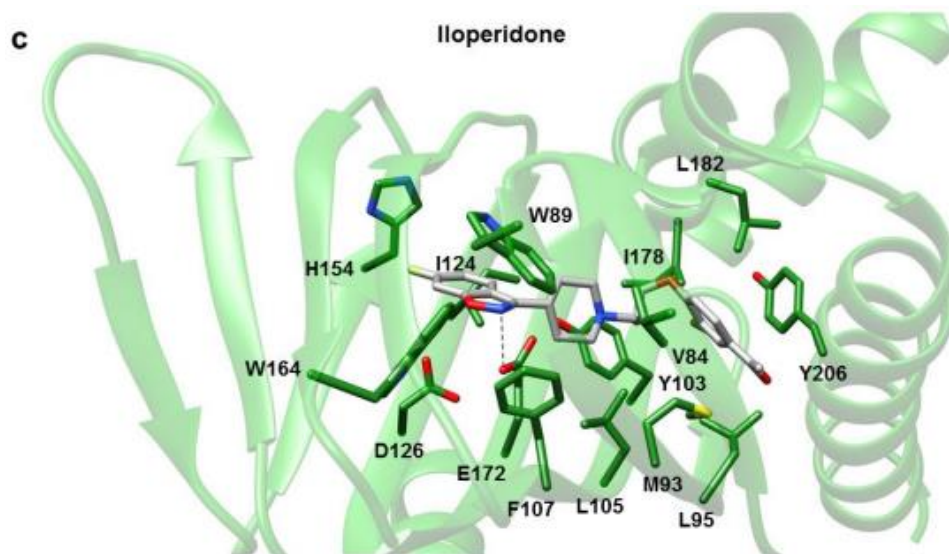
Drug name ZINC ID Chemical formula	IUPAC	Clinical indication - mechanism of action	2D structure	3D conformation
Risperidone ZINC: 538312 C23H27FN4O2	3-[2-[4-(6-fluoro-1,2-benzoxazol-3-yl)piperidin-1-yl]ethyl]-2-methyl-6,7,8,9-tetrahydropyrido[1,2-a]pyrimidin-4-one	Psychosis - Adrenergic receptor and dopamine D2-receptor antagonist		
Axitinib (cis) ZINC: 11616882 C22H18N4OS	N-methyl-2-((3-[(E)-2-(pyridin-2-yl)ethenyl]-1H-indazol-6-yl)sulfonyl)benzamide	Antineoplastic - VEGFR inhibitor		
Nilotinib (*) ZINC: 6716957 C28H22F3N7O	4-methyl-N-[3-(4-methyl-1H-imidazol-1-yl)-5-(trifluoromethyl)phenyl]-3-[[4-(pyridin-3-yl)pyrimidin-2-yl]amino]benzamide	Chronic myeloid leukemia (CML) - Tyrosine kinase inhibitor		
Paliperidone (*) ZINC: 4214700 C23H27FN4O3	3-[2-[4-(6-fluoro-1,2-benzoxazol-3-yl)piperidin-1-yl]ethyl]-9-hydroxy-2-methyl-4H,6H,7H,8H,9H-pyrido[1,2-a]pyrimidin-4-one	Schizophrenia - 5HT2A and D2 antagonist (risperidone metabolite)		
Paliperidone (*) ZINC: 1481956 C23H27FN4O3	3-[2-[4-(6-fluoro-1,2-benzoxazol-3-yl)piperidin-1-yl]ethyl]-9-hydroxy-2-methyl-4H,6H,7H,8H,9H-pyrido[1,2-a]pyrimidin-4-one	Schizophrenia - 5HT2A and D2 antagonist (risperidone metabolite)		
Atovaquone ZINC: 100017856 C22H19ClO3	2-hydroxy-3-[[1-(1,4)-4-(4-chlorophenyl)cyclohexyl]-1,4-dihydronaphthalene-1,4-dione	Malaria - Cytochrome b inhibitor		
Droperidol ZINC: 19796080 C22H22FN3O2	1-[1-[4-(4-fluorophenyl)-4-oxobutyl]-1,2,3,6-tetrahydropyridin-4-yl]-2,3-dihydro-1H-1,3-benzodiazol-2-one	Neuroleptanalgesia - Dopamine receptor antagonist		
Dolutegravir ZINC: 58581064 C20H19F2N3O5	(3S,7R)-N-[(2,4-difluorophenyl)methyl]-11-hydroxy-7-methyl-9,12-dioxo-4-oxa-1,8-diazatricyclo[8.4.0.0.0']{3,8}tetradeca-10,13-diene-13-carboxamide	HIV - HIV-1 integrase inhibitor		
Iloperidone (*) ZINC: 1548097 (C24H27FN2O4)	1-[4-[3-[4-(6-fluoro-1,2-benzoxazol-3-yl)piperidin-1-yl]propoxy]-3-methoxyphenyl]ethan-1-one	Schizophrenia - Dopamine D2 and 5-HT2A receptor antagonist		
Linagliptin (*) ZINC: 3820029 C25H28N8O2	8-[(3R)-3-aminopiperidin-1-yl]-7-(but-2-yn-1-yl)-3-methyl-1-[(4-methylquinazolin-2-yl)methyl]-2,3,6,7-tetrahydro-1H-purine-2,6-dione	Diabetes mellitus type 2 - Muscarinic antagonist		
Flibanserin (*) ZINC: 52716421 C20H21F3N4O	1-(2-[4-(3-(trifluoromethyl)phenyl)piperazin-1-yl]ethyl)-2,3-dihydro-1H-1,3-benzodiazol-2-one	Hypoactive sexual desire disorder (HSDD) - 5-HT1A and 5-HT2A receptor ligand		
Clostrazol ZINC: 1552174 C20H27N5O2	6-[4-(1-cyclohexyl-1H-1,2,3,4-tetrazol-5-yl)butoxy]-1,2,3,4-tetrahydroquinolin-2-one	Vasodilator, Antiplatelet - cAMP phosphodiesterase III inhibitor		
Nebivolol ZINC: 11681534 C22H25F2NO4	N-methyl-2-((3-[(E)-2-(pyridin-2-yl)ethenyl]-1H-indazol-6-yl)sulfonyl)benzamide	Hypertension - $\beta_1$ receptor blocker		
Vilazodone (*) ZINC: 1542113 C26H27N5O2	5-[4-[4-(5-cyano-1H-indol-3-yl)butyl]piperazin-1-yl]-1-benzofuran-2-carboxamide	Depression - serotonin reuptake inhibitor		
Azilsartan medoxomil ZINC: 14210642 C30H24N4O8	(5-methyl-2-oxo-2H-1,3-dioxol-4-yl)methyl 2-ethoxy-1-[(4-[2-(5-oxo-4,5-dihydro-1,2,4-oxadiazol-3-yl)phenyl]phenyl)methyl]-1H-1,3-benzodiazole-7-carboxylate	Hypertension - Angiotensin II receptor antagonist		
Doxercalciferol ZINC: 4641374 C28H44O2	(1R,3S,5Z)-5-[2-[[1R,3aS,4E,7aR)-1-[[2R,3E,5R)-5,6-dimethylhept-3-en-2-yl]-7a-methyl-octahydro-1H-inden-4-ylidene]ethylidene]-4-methylidene-cyclohexane-1,3-diol	Hyperparathyroidism - Parathyroid hormone synthesis, calcium homeostasis		
Cinacalcet ZINC: 1550499 C22H22F3N	[(1R)-1-(naphthalen-1-yl)ethyl]((3-[3-(trifluoromethyl)phenyl]propyl)amino)amine	Hyperparathyroidism - Calcimimetic		
Nebivolol ZINC: 4213946 C22H25F2NO4	1-(6-fluoro-3,4-dihydro-2H-1-benzopyran-2-yl)-2-[(2-(6-fluoro-3,4-dihydro-2H-1-benzopyran-2-yl)-2-hydroxyethyl)amino]ethan-1-ol	Hypertension - $\beta_1$ receptor blocker		
Nebivolol ZINC: 5844792 C22H25F2NO4	1-(6-fluoro-3,4-dihydro-2H-1-benzopyran-2-yl)-2-[(2-(6-fluoro-3,4-dihydro-2H-1-benzopyran-2-yl)-2-hydroxyethyl)amino]ethan-1-ol	Hypertension - $\beta_1$ receptor blocker		
Axitinib (trans) ZINC: 11616882 C22H18N4OS	N-methyl-2-((3-[(E)-2-(pyridin-2-yl)ethenyl]-1H-indazol-6-yl)sulfonyl)benzamide	Antineoplastic - Tyrosine kinase inhibitor		

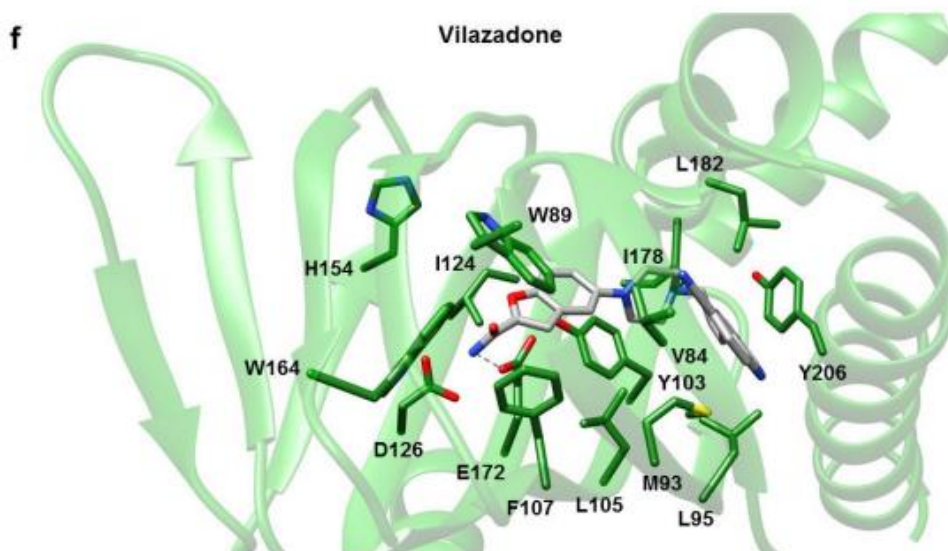
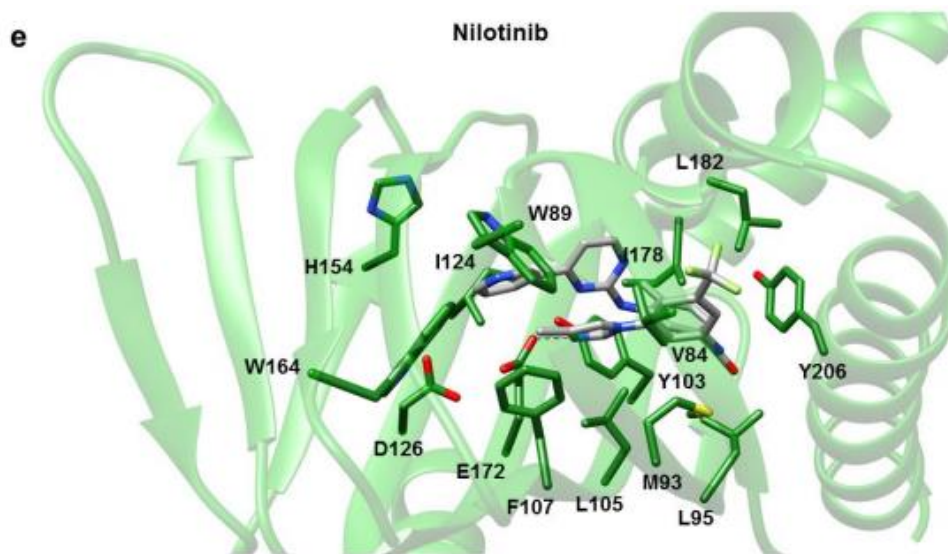
Figure 2.1. Clinical indication and chemical structures of the compounds listed in Table 2. Compounds selected for experimental validation by SPR are indicated by an asterisk (\*).

Additionally, we visually inspected each of the poses to assess whether additional interactions might occur, in case small conformational adjustments

with respect to the poses predicted by docking programs were allowed. At the end of this analysis, we selected six compounds to be experimentally evaluated for  $\sigma_1$ R binding by SPR: nilotinib, paliperidone, iloperidone, linagliptin, flibanserin and vilazodone. The complexes of these ligands and of pridopidine, for comparison purposes, with 5HK1\_A predicted by molecular docking are shown in Figure 2.2a–h.







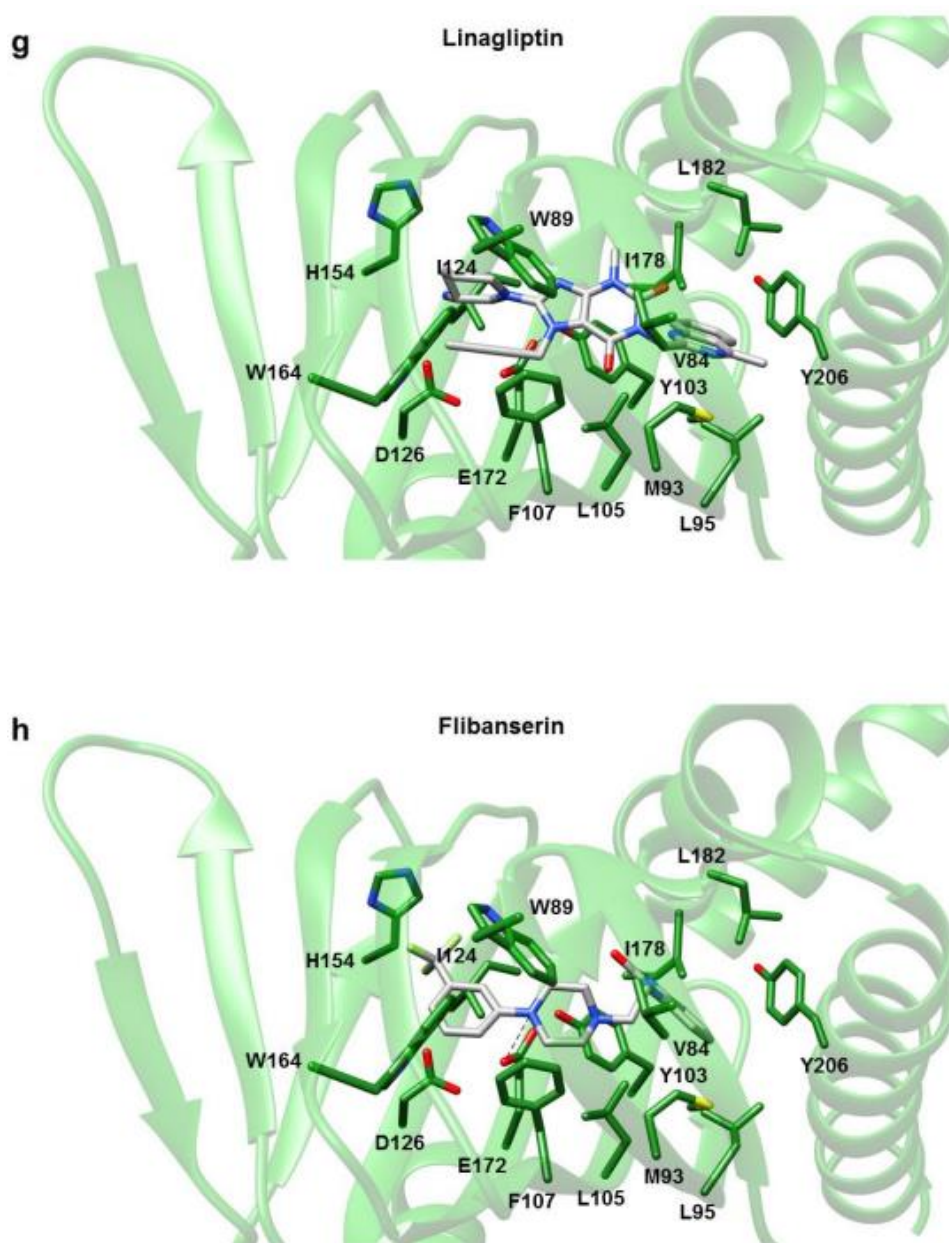


Figure 2.2. Ligand binding to  $\sigma 1R$ . The protein backbone in co-ordinate file 5HK1 is shown as transparent green ribbon.  $\sigma 1R$  residues involved in ligand binding in the 3D structures listed in Table 1 are labelled, shown as sticks and coloured by atom type: C, green; N, blue;

*O*, red. Ligands are shown as sticks and coloured by atom-type: C, white; N, blue; O, red; S, yellow; F, light green. Glu172 side-chain oxygen and the closest ligand atom able to establish a polar interaction with it are connected by a dashed line. The co-ordinates of the complex with PD144418 (a) have been experimentally determined by X-ray crystallography (PDB ID: 5HK1). The conformation of Pridopidine (b) was predicted by virtual screening. The conformation of all the other ligands corresponds to the best energy pose of the most populated cluster predicted by docking simulations (see Materials and Methods): (c) Iloperidone; (d) Paliperidone; (e) Nilotinib; (f) Vilazodone; (g) Linagliptin; (h) Flibanserin.

Interestingly, known  $\sigma$ 1R ligands were found at much lower virtual screening ranking positions. Pridopidine was at position 458 and haloperidol and N,N-dimethyltryptamine, the highest and lowest ranking among known ligands, were at positions 51 and 1002, respectively. However, the differences in predicted binding energies between the highest-ranking known ligands and the 20 highest ranking drugs are relatively small (Table 2a, b), especially when the lack of accuracy of binding energy differences  $<3$  kcal/mol is taken into account (Cosconati et al. 2010). As an example, the binding energy difference between haloperidol and the 1st ranked hit (risperidone) is 1.7 kcal/mol, between haloperidol and pridopidine is 2.2 kcal/mol, and between haloperidol and N,N-dimethyltryptamine is 1.6 kcal/mol. Examination of the poses of the six selected drugs in comparison with that of pridopidine shows that the latter is somewhat smaller and, therefore, establishes a lower number of hydrophobic interactions, while the polar interactions with Glu172 is maintained. The lower number of interactions might account for the worse interaction energy and, therefore, ranking, predicted by Vina.

*Table 2.2. Results of virtual screening and computational docking procedures. (a) Twenty FDA approved compounds having the best interaction energy with the 1R structure in coordinate file 5HK1\_A. Vina Best E: energy of the ligand-receptor pose calculated by Vina. Autodock largest cluster and lowest energy cluster: Best E, Mean E and #pos represent the energy of the best pose in the cluster, the mean energy of the poses in the cluster and the*



number of poses in the cluster, respectively. #hb, #cla and #con: number of hydrogen bonds, clashes and contacts calculated by Chimera. Vina/Autodock RMSD (Å): root-mean square deviation (in angstroms) calculated after optimal superposition of the pose selected by Vina and the best pose of the largest cluster calculated by Autodock Compounds selected for experimental validation by SPR are indicated by an asterisk (\*). (b) Ranking and Best energy calculated by Vina for known  $\sigma$ IR ligands in the compound library used for virtual screening.

(a)

Drug Name	ZINC ID Number	Vina Best E (kcal/mol)	Chimera			Vina/ATD Comparison
			#hb	#cla	#con	RMSD (Å)
Risperidone	538312	-12.6	0	0	70	2.9
Axitinib (cis)	11616882	-12.6	3	0	59	0.5
Nilotinib (*)	6716957	-12.3	4	1	110	6.4
Paliperidone (*)	4214700	-12.2	0	0	72	3.1
Paliperidone (*)	1481956	-12.2	0	0	80	10.7
Atovaquone	100017856	-11.9	0	0	51	2.9
Droperidolo	19796080	-11.9	0	0	58	3.3
Dolutegravir	58581064	-11.9	3	0	58	9.4
loperidone (*)	1548097	-11.7	1	0	70	10.7
Linagliptin (*)	3820029	-11.7	0	1	97	0.6
Flibanserin (*)	52716421	-11.6	0	0	48	1.0
Cilostazol	1552174	-11.6	0	0	71	9.4
Nebivolol	11681534	-11.6	2	0	62	10.1
Vilazodone (*)	1542113	-11.6	2	2	81	6.1
Azilsartan	14210642	-11.5	2	1	123	9.0
Medoxomil						
Doxercalciferol	4641374	-11.5	0	2	96	3.4
Cinacalcet	1550499	-11.5	0	0	52	8.7
Nebivolol	4213946	-11.5	5	0	62	2.4
Nebivolol	5844792	-11.5	4	0	61	2.5
Axitinib (trans)	11616882	-11	0	0	63	2.3

(b)

Ligand Name (PDB)	ZINC ID
Haloperidol (GMJ)	537822
4-IBP (61V)	1642602
PDI44418 (61W)	5862
NE-100 (GKY)	598622
Pridopidine	22063703
(+)-Pentazocine (GM4)	596
Phencyclidine	968311
Remoxipride	2021799
Captodiamine	2040210
N,N-Dimethyltryptamine	897457

ATD Largest Cluster			Chimera				ATD Lowest Energy Cluster		
Best E (kcal/mol)	Mean E (kcal/mol)	#pos	#hb	#cla	#con	Best E (kcal/mol)	Mean E (kcal/mol)	#pos	
-11.5	-11.1	37	0	1	99	-11.5	-11.1	37	
-10.9	-10.8	41	1	0	70	-10.9	-10.8	41	
-7.8	-5.5	16	2	11	148	-9.5	-3.8	6	
-11.5	-10.7	46	1	2	111	-11.5	-10.9	4	
-11.1	-10.6	14	1	1	93	-11.9	-11.2	4	
-11.1	-10.7	73	0	1	68	-11.3	-11.0	27	
-9.8	-9.4	27	0	1	83	-10.2	-9.9	13	
-9.4	-9.2	52	0	0	71	-10.3	-9.9	43	
-10.2	-9.7	39	0	2	84	-10.4	-9.6	7	
-12.4	-10.0	33	1	6	112	-12.4	-10.0	33	
-9.4	-9.2	71	0	0	59	-10.0	-9.4	11	
-9.2	-8.7	25	1	0	68	-9.7	-9.2	19	
-10.0	-8.7	73	1	1	81	-10.0	-8.7	73	
-9.2	-8.2	38	2	7	126	-9.4	-7.6	7	
-7.9	-4.4	18	0	12	143	-9.0	-4.4	8	
-12.2	-10.1	55	1	2	112	-12.2	-10.1	55	
-10.4	-9.8	23	1	0	85	-10.4	-9.8	23	
-10.9	-9.5	58	2	0	78	-10.9	-9.5	58	
-11.1	-9.3	62	4	0	84	-11.1	-9.3	62	
-7.6	-7.6	75	0	0	61	-7.7	-7.6	15	
Vina Best E (kcal/mol)			Ranking						
-10.9							51		
-10.7							71		
-10.1							139		
-8.9							406		
-8.7							458		
-8.6							516		
-8.5							529		
-7.9							723		
-7.9							733		
-7.1							1002		

In vitro SPR experiments show that all the compounds predicted by virtual screening to interact with the receptor are indeed able to bind to  $\sigma$ 1R, with dissociation constant ( $K_D$ ) values in the micromolar range (Table 2.3). Interestingly, flibanserin, iloperidone and linagliptin ( $K_D < 10 \mu\text{M}$ ) showed higher affinity values for  $\sigma$ 1R than pridopidine ( $K_D$  about  $15 \mu\text{M}$ ). This value falls in a different concentration range with respect to the previously reported inhibition constant value ( $K_i = 81.7 \text{ nM}$ ) (Sahlholm et al. 2013), which, however, was determined in significantly different experimental conditions. In the previous assay  $\sigma$ 1R was inserted in cell membranes and saturated with  $[^3\text{H}](+)\text{-pentazocine}$ ; the pridopidine  $K_i$  value was then measured based on its

ability to displace the radioactive ligand. Conversely, the  $K_D$  value measured in SPR experiments indicates the binding affinity between pridopidine and the purified  $\sigma 1R$  in conditions where ligand and receptor are allowed to interact directly.

*Table 2.3  $\sigma 1R$  affinity for selected compounds and pridopidine, as a control, measured by SPR experiments (\*) Pridopidine  $K_i = 81.7$  nM in a displacement assay (Sahlholm et al. 2013).*

FDA Name	ZINC ID	$K_D$ ( $\mu$ M)
Pridopidine	ZINC000022063703	$14.8 \pm 1.0$ (*)
Flibanserin	ZINC000052716421	$4.9 \pm 1.1$
Iloperidone	ZINC000001548097	$5.1 \pm 0.6$
Linagliptin	ZINC000003820029	$9.6 \pm 1.0$
Nilotinib	ZINC000006716957	$22.0 \pm 3.0$
Paliperidone	ZINC000004214700	$46.0 \pm 21$
Vilazodone	ZINC000001542113	$52.0 \pm 9.0$

To investigate whether cell proliferation rate was a suitable parameter to indicate the ability of selected drugs to interact with  $\sigma 1R$  in a cellular context and exert an agonist or antagonist function, we cultured fibroblasts in the presence of the known  $\sigma 1R$  agonist pridopidine. In the presence of pridopidine, the cell number of both HD and healthy lines was significantly increased with respect to the DMSO control after 72 h and, in the case of HD2 cells, at 48 h as well (Figure 2.3a–d). This increase was not paralleled by a reduced number of dead cells, which was only reduced to a small, not significant extent (Figure 2.3e–h).

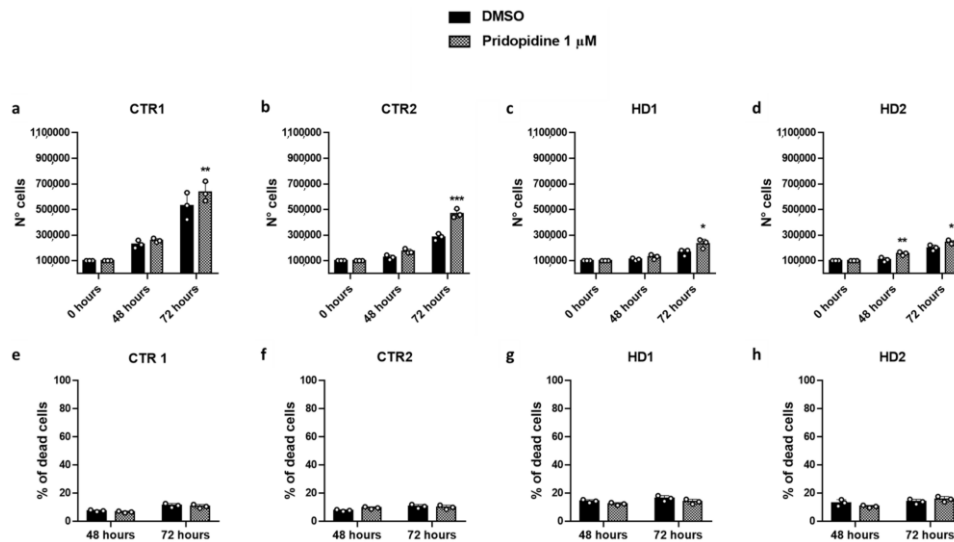


Figure 2.3. Pridopidine effect on CTR1 (a), CTR2 (b), HD1 (c) and HD2 (d) cell growth, and CTR1 (e), CTR2 (f), HD1 (g) and HD2 (h) percentage of dead cells. Fibroblast numbers were determined by direct counting of stained and unstained cells after 48 and 72 h and plotted against culture time. Data are shown as mean  $\pm$  SD. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.0001$ , compared with DMSO treatment (Two-way ANOVA, followed by uncorrected Fisher's LSD for multiple comparisons).

The same assay was conducted with all the six drugs, that were tested at the same concentration used for pridopidine (i.e., 1  $\mu$ M) and all data were normalized with respect to the same cell lines treated only with the DMSO vehicle.

All of the selected drugs significantly increased fibroblast number at 72 h, in either HD1 or HD2 cell lines, and some of them at 48 h as well. Iloperidone, paliperidone and nilotinib had a consistent beneficial effect, since they significantly increased both HD1 and HD2 fibroblast number at 72 h, and paliperidone at 48 h as well (Figure 2.4c, d). However, while iloperidone did not have any effect on healthy fibroblasts at either 72 or 48 h, paliperidone significantly decreased CTR1 and increased CTR2 cell number at 72 h and

nilotinib significantly decreased CTR2 cell number at 72 h (Figure 2.4a, b). The effect of the other drugs on the different cell lines is less consistent. The number of HD1 fibroblasts is significantly increased by linagliptin, flibanserin and vilazodone at 72 h, and by vilazodone at 48 h as well, but none of these drugs increases the number of HD2 fibroblasts, flibanserin even determining a significant reduction in cell number at 72 h (Figure 2.4c, d). As far as control cell lines are concerned, all of these compounds determined significant increases in CTR2 and/or CTR1 cell number: CTR2 cells were increased by linagliptin and vilazodone at 72 h and by linagliptin and flibanserin at 48 h; CTR1 cells were increased by vilazodone at 72 h (Figure 2.4a, b).

Importantly, the effect of the tested drugs on HD fibroblasts was comparable to and, in several cases, better than that of pridopidine. In particular, at 72 h, the number of HD1 fibroblasts in the presence of linagliptin, paliperidone or vilazodone, and the number of HD2 fibroblasts in the presence of iloperidone or paliperidone, was significantly higher than the number of the respective fibroblasts in the presence of pridopidine.

In agreement with the above results, treatment with the selected drugs significantly increased the growth rates of either HD1 or HD2 fibroblasts to an extent comparable to, and, in several cases, better than that of pridopidine.

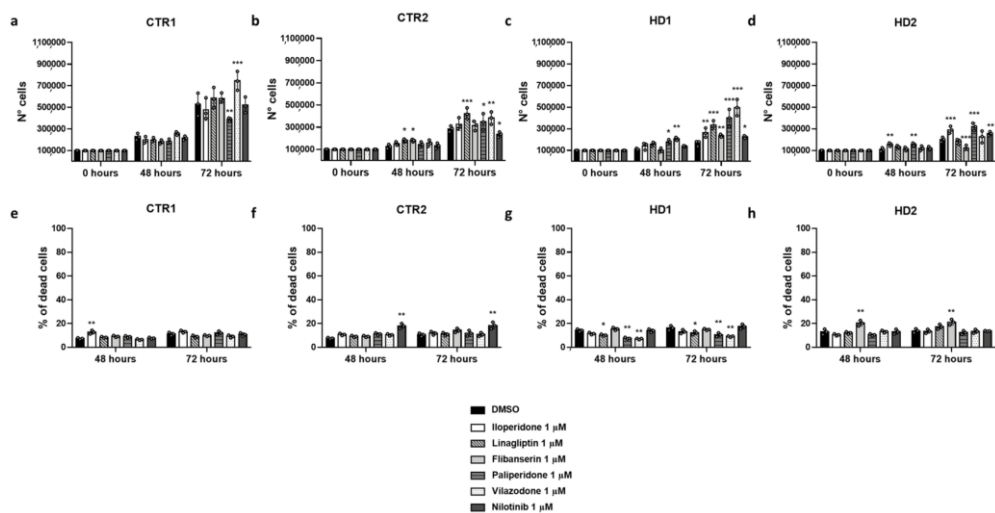


Figure 2.4 Effect of iloperidone, linagliptin, flibanserin, paliperidone, vilazodone and nilotinib on CTR1 (a), CTR2 (b), HD1 (c) and HD2 (d) cell growth, and percentage of dead cells in CTR1 (e), CTR2 (f), HD1 (g) and HD2 (h) cell lines. Fibroblast numbers were determined by direct counting of stained and unstained cells after 48 and 72 h and plotted against culture time. Data are shown as mean  $\pm$  SD. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.0001$ , compared with DMSO treatment (Two-way ANOVA, followed by uncorrected Fisher's LSD for multiple comparisons).

## 2.4.2 Investigation of the Entry Pathway and Molecular Nature of $\sigma_1$ Receptor Ligands.

To try and identify common structures among  $H\sigma_1$ -R ligands, we performed VS on a 21,359-compounds dataset. This dataset comprises: all human metabolites; several categories of steroid-based compounds (i.e., sterol lipids, sterols, steroids, androgens, estrogens and compounds belonging to the cholesterol or ergosterol biosynthetic pathway); known ligands of the  $\sigma_1$ -R receptor and/or of the  $\sigma_2$ -R receptor, which is known to share several ligands with  $\sigma_1$ -R in spite of their different overall structure, as positive controls; and

compounds approved for clinical use by the FDA or other regulatory agencies, such as EMA (European Medicines Agency). This dataset comprises several compounds that are reported to bind human  $\sigma$ 1R with very low affinity (i.e.,  $K_i > 10,000$  nM) by the Psychoactive Drug Screening Program (PDSP) Ki database (Roth et al. 2016) which we used as negative controls.

The VS was performed against two of the five available H $\sigma$ 1-R structures, namely the structures in coordinate files 5HK1, since it was solved with the highest resolution, and 6DK1\_A, since it is the only one determined in complex with an agonist, rather than antagonist, compound, and against the molecular model of yeast ERG2 built by the AlphaFold2 program.

The results of these VS experiments are summarized in Table 2.4 for two subsets of hits, defined on the basis of the values of their receptor binding energy calculated by the program used for VS ( $E_{\text{calc}}$ ), namely: (i) the 20 hits with the lowest  $E_{\text{calc}}$ ; and (ii) all the hits whose  $E_{\text{calc}}$  does not differ more than 3 kcal/mol from the lowest  $E_{\text{calc}}$ . The rationale for choosing the first set of hits is that it is commonly reported to be selected for detailed analyses in the literature, due to the fact that 20 is a small enough number of compounds for visual inspection. However, it has been reported that  $E_{\text{calc}}$  values have a standard deviation of 2–3 Kcal/mol (Huey et al. 2007). It follows that hits whose  $E_{\text{calc}}$  differs by less than 3.0 Kcal/mol from the  $E_{\text{calc}}$  of the best hit may have an actual binding energy to the receptor similar to, or even better than, that of the best hit, and may, therefore, be all considered as “best hits”. For this reason, we chose to analyse in greater detail this second set of hits, which, from now on, will be referred to as “best-E3”. We found that the “best-E3” subsets for the VS against H $\sigma$ 1-R in coordinate files 5HK1 and 6DK1 comprise 1666 and 2987 hits, respectively, with  $E_{\text{calc}}$  between  $-13.10$  and  $-10.10$  kcal/mol and

between  $-13.20$  and  $-10.20$  kcal/mol, respectively. Comparison of these results (data not shown) shows that 1271 compounds among the “best-E3” for both structures, whereas 395 and 1717 compounds are among the “best-E3” hits only for 5HK1 and 6DK1, respectively.

Given the high similarity between the Hs $\sigma$ 1-R structures in coordinate files 5HK1 and 6DK1, these results indicate that VS results are significantly affected even by the very small side-chain variations induced upon Hs $\sigma$ 1-R binding by different ligands.

The “best-E3” results of the VS against ERG2 molecular model fall into an  $E_{\text{calc}}$  range between  $-11.7$  and  $-8.7$  Kcal/mol, which is higher than those of the “best-E3” resulting from VS against the Hs $\sigma$ 1-R structure in coordinate sets 5HK1 and 6DK1, although the difference is not significant when the expected 2–3 kcal/mol standard deviation on  $E_{\text{calc}}$  values is taken into account (Huey et al. 2007). Due to this expected standard deviation, the  $E_{\text{calc}}$  values for fecosterol and episterol, which are the substrate and product of the reaction catalysed by the ERG2 protein, respectively, are higher (i.e.,  $-10.3$  and  $-9.4$  kcal/mol, respectively) than the best hit (phaseolinisoflavan,  $E_{\text{calc}} = -11.7$  kcal/mol), which does not contain a steroid nucleus; additionally, other compounds belonging to the ergosterol synthesis pathway and, therefore, likely to have structures able to bind ERG2, have an  $E_{\text{calc}}$  similar to, or higher than, that of unrelated compounds.

*Table 2.4. Summary of the results of VS experiments against the 3D structures of Hs $\sigma$ 1-R in coordinate files 5HK1 and 6DK1 and the molecular model of yeast ERG2. 5HK1\_A and 6DK1\_A: monomer with chain ID “A” in coordinate files 5HK1 and 6DK1, respectively. ERG2: molecular model of yeast ERG2 protein built by AlphaFold2. B\_20: 20 hits with lowest calculated interaction energy ( $E_{\text{calc}}$ ) with the target structure. E\_3.0: “best-E3”, namely hits whose  $E_{\text{calc}}$  with the target structure is  $\leq 3.0$  kcal/mol higher than that of the best hit. Nb. Hits:*



number of hits in each results subset (i.e., B\_20 and E\_3.0).  $E_{calc}$  (kcal/mol): range of interaction energy with the target protein in each results subset. Hb, Contacts and Clashes: range of hydrogen bonds, overall contacts and unfavourable van der Waals contacts between ligand and target protein in each subject. Ago-Ant: known  $\sigma$ 1-R and/or  $\sigma$ 2-R binders. Metab, FDA and World: molecules tagged as “metabolites + for sale”, “FDA approved + for sale” and “World-not FDA + for sale” in the ZINC15 database (Sterling & Irwin 2015). Ste\_Lip, Sterols, Steroids, Androg, Estrog, Chol\_P and Ergo\_P: molecules belonging to the sterol\_lipids, sterols, steroids, androgens, estrogens, cholesterol or ergosterol biosynthetic pathway categories, respectively, and available for sale in the LIPID MAPS database (Sud et al. 2007).

	5HK1_A		6DK1_A		ERG2	
	B_20	E_3.0	B_20	E_3.0	B_20	E_3.0
Subset	B_20	E_3.0	B_20	E_3.0	B_20	E_3.0
Nb. Hits	20	1679	20	3005	20	2574
$E_{calc}$ (kcal/mol)	-13.1(-12.3)	-13.1(-10.1)	-13.2(-12.5)	-13.2(-10.5)	-11.7(-11.0)	-11.7(-8.7)
Hb	0-2	0-3	0-1	0-6	0-2	0-10
Cont	40-100	29-152	41-100	28-175	44-95	22-143
Clashes	0-4	0-7	0-3	0-7	0-3	0-8
Ago-Ant	0	8	1	5	0	10
Metab	2	431	3	681	10	1320
Ste_Lip	0	436	4	1192	4	627
Sterols	7	412	11	706	2	299
Steroids	0	109	0	138	1	218
Androg	0	52	0	68	1	79
Estrog	0	42	0	6	0	45
Chol_P	0	135	4	254	1	140
Ergo_P	1	124	1	169	0	85
FDA	3	140	0	144	0	312
World	8	248	1	261	4	43

To verify whether the “best-E3” subsets were enriched with specific structures with respect to the whole 21,359 compounds dataset used for VS, we compared the number of compounds belonging to each category (e.g., metabolites, agonists/antagonists, etc.) comprised in this initial dataset with the number of hits of the same category comprised in the “best-E3” subsets, resulting from VS towards the Hs $\sigma$ 1-R structures in coordinate files 5HK1 and 6DK1, and towards the ERG2 molecular model (Table 2.5). Examination of these values shows that the “best-E3” subsets resulting from VS against the 5HK1 structure and ERG2 model are significantly enriched in compounds belonging to the agonists/antagonists category, which comprises experimentally validated  $\sigma$ 1-R ligands, the ratio between the percentage of this category among the “best-

E3” hits and among the starting set of compounds (R) being 2.8 for the 5HK1 structure and 2.4 for the ERG2 model. Conversely, no variation in the percentage of this category is observed in the results of VS against 6DK1. The percentage of compounds belonging to the “metabolites” category is significantly reduced among the “best-E3” subsets for both the Hs $\sigma$ 1-R structures and ERG2 model, with R values of 0.3, 0.3 and 0.7, respectively, whereas compounds approved by the FDA and other regulatory agencies do not show a regular trend (Table 2.5). Interestingly, compounds having a steroid-based structure (i.e., sterol lipids, sterols, steroids, androgens, estrogens and compounds in the cholesterol or ergosterol pathway) are significantly enriched in the “best E3” subsets of all three proteins, i.e., both the Hs $\sigma$ 1-R structures and the ERG2 model, with respect to categories comprising compounds with very diverse chemical structures, such as metabolites and compounds approved for clinical use by the FDA or other regulatory agencies. In detail, steroid-based compounds are less than 25% of the total number of compounds in the 21,359 compounds dataset used for VS, and 61%, 69% and 47% of compounds among the “best-E3” subsets of results against the Hs $\sigma$ 1-R structure in coordinate files 5HK1 and 6DK1 and the ERG2 molecular model, respectively, which corresponds to an enrichment in steroid-based compounds of 2.5, 2.8 and 1.9 folds, respectively.

*Table 2.5. Comparison between the percentage of compounds in each category used for VS against the 3D structures of Hs $\sigma$ 1-R in coordinate files 5HK1 and 6DK1, or the molecular model of yeast ERG2, and the percentage of compounds in the same categories found in the “best-E3” subset obtained from VS against each structure.*

Category	Compounds Used for VS		"Best-E3"								
			5HK1			6DK1			ERG2		
	Nb	%	Nb	%	R	Nb	%	R	Nb	%	R
Ago-Ant	36	0.1	8	0.4	2.8	5	0.1	1.0	10	0.3	2.4
Metab	15,871	58.1	433	20.2	0.3	694	19.0	0.3	1384	43.3	0.7
Ste_Lip	3761	13.8	436	20.4	1.5	1192	32.7	2.4	627	19.6	1.4
Sterols	1593	5.8	412	19.2	3.3	706	19.3	3.3	299	9.3	1.6
Steroids	362	1.3	109	5.1	3.8	138	3.8	2.9	218	6.8	5.1
Androg	101	0.4	52	2.4	6.6	68	1.9	5.0	79	2.5	6.7
Estrog	63	0.2	42	2.0	8.5	6	0.2	0.7	45	1.4	6.1
Chol_P	442	1.6	135	6.3	3.9	254	7.0	4.3	140	4.4	2.7
Ergo_P	347	1.3	124	5.8	4.6	169	4.6	3.6	85	2.7	2.1
FDA	1538	5.6	142	6.6	1.2	157	4.3	0.8	269	8.4	1.5
World	3192	11.7	248	11.6	1.0	261	7.2	0.6	43	1.3	0.1
All Ste	6669	24	1310	61	2.5	2533	69	2.8	1493	47	1.9
All Div	20,601	75	823	38	0.5	1112	30	0.4	1696	53	0.7
All Cat	27,306	100	2141	100	1	3650	100	1	3199	100	1

Taken together, these results indicate that the program used for VS has a good ability to recognize actual Hs $\sigma$ 1-R ligands, which are enriched among the "best-E3" hits of both the highest resolution coordinate file 5HK1 and the homologous ERG2 protein model (although not among the "best-E3" hits of the lower resolution coordinate file 6DK1), as well as to identify low-affinity Hs $\sigma$ 1-R binders, and that steroid-based compounds are among the Hs $\sigma$ 1-R preferred ligands.

To obtain further information about the nature of physiological binders of  $\sigma$ 1-R proteins, we tried to identify the compound(s) giving rise to the electron density peak near the binding site of the Xl $\sigma$ 1-R structure in coordinate files 7W2B and 7W2E. We first performed a VS of the 1332 yeast metabolites dataset against the two Xl $\sigma$ 1-R apo structures in coordinate files 7W2B\_A and 7W2E\_A. The results of this VS are summarized in Table 2.6.

*Table 2.6. Summary of the results of VS experiments against the 3D structures of Xl $\sigma$ 1-R in coordinate files 7W2B and 7W2E. 7W2E\_A and 7W2B\_A: monomer with chain ID "A" in coordinate files 7W2E and 7W2B, respectively.*

Subset	7W2E_A		7W2B_A	
	B_20	E_3.0	B_20	E_3.0
Nb. Hits	20	143	20	90
E <sub>calc</sub> (kcal/mol)	-10.8-(-9.2)	-10.8-(-7.8)	-11.1-(-9.3)	-11.1-(-8.1)
Hb	0-5	0-8	0-6	0-8
Contacts	31-93	18-100	37-110	22-110
Clashes	0-3	0-6	0-5	0-5

To select compounds likely to fit in the electron density maps in the ligand binding site of Xlσ1-R structures, we visually inspected the 2D structures of the 88 “best-E3” hits that are common between the results of VS against the Xlσ1-R structures in coordinate files 7W2B and 7W2E (data not shown). We selected five compounds (Table 2.7) based on the following criteria: (i) the compatibility of their molecular shape with the electron density peaks observed in the apo 7W2B and 7W2E structures; and (ii) the fact that their molecular structures were quite different from one another and, at the same time, each of them was similar to other compounds satisfying the first criteria.

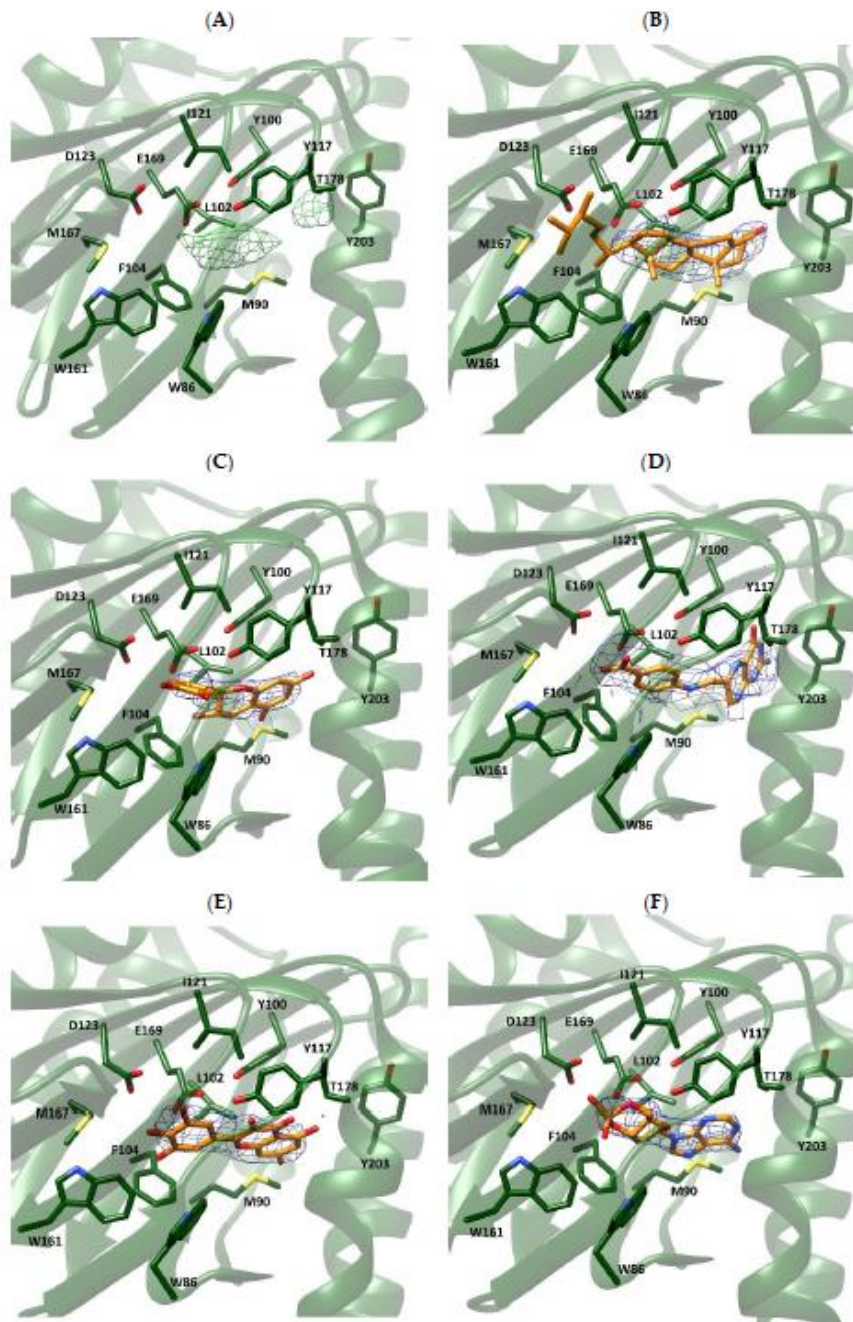
*Table 2.7 Average B-factor values of the five selected molecules after fitting in the electron density map of the Xlσ1-R structure in chain C of coordinate files 7W2E (7W2E\_C) and 7W2B (7W2B\_C). For each compound in each structure, occupancy = 1.00.*

YMDB ID	Ligand Name	Average B-Factor with 7W2E_C	Average B-Factor with 7W2B_C
YMDB00543	Ergosterol	86.4	114.7
YMDB01653	Catechin	107.2	131.8
YMDB00293	7,8-Dihydropteroic acid	126.2	121.1
YMDB01754	Myricetin	116.1	108.3
YMDB00452	3',5'-Cyclic dAMP	163.8	139.0

For both Xlσ1-R structures, we selected the chain where the electron density map peak found in the proximity of the Xlσ1-R binding site in the Fo-Fc map is most intense, namely chain C for both coordinate files 7W2B (7W2B\_C)

and 7W2E (7W2E\_C). Then, we tested the selected compounds for their ability to fit the electron density peak in 7W2B\_C and 7W2E\_C, and calculated the average B-factor values of the resulting complexes (Table 2.7).

Additionally, visual inspection of the generated complexes (Figure 2.5) indicated that the five selected compounds fit very well in the electron density map peak of both 7W2B\_C and 7W2E\_C. In line with the enrichment in steroid-based compounds in the “best-E3” results of VS experiments, ergosterol was the compound giving rise to the lowest B-factor value in the complex with coordinate file 7W2E\_C and the second lowest B-factor value in complex with 7W2B\_C. As shown in Figure 2.5 (panels B and H), the four A-D rings making the steroid nucleus are within the electron density peak, with only part of the ergosterol long chain substituent at position 17 falling outside the electron density.



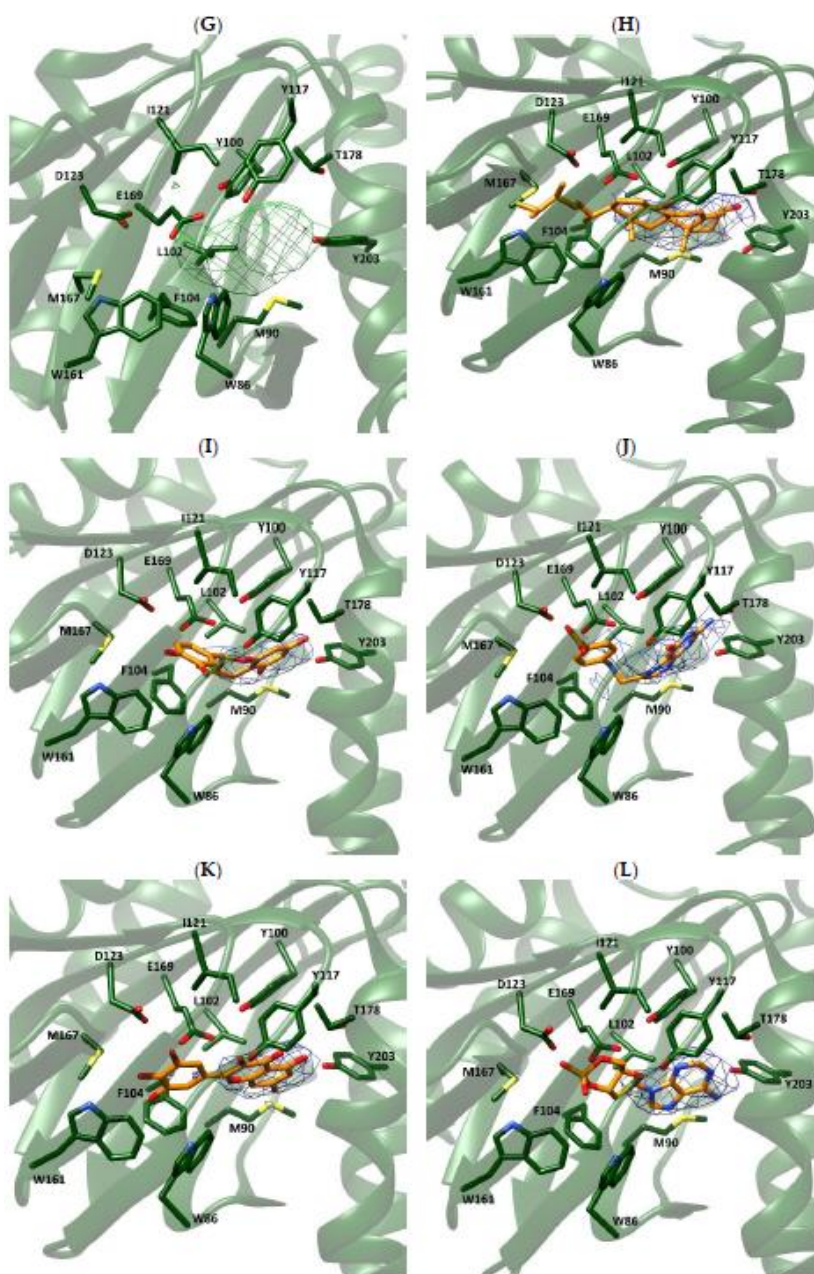


Figure 2.5. Fitting of selected compounds into the electron density within the ligand binding site of XI\_1-R. The protein is shown as ribbon and colored green. The side-chains of residues surrounding the ligand binding site are shown as sticks and coloured by atom-type: C, N, O and S atoms are green, blue, red and yellow, respectively. The structures in coordinate files

*7W2B and 7W2E are shown in panels (A–F) and (G–L), respectively. Ligands in panels (B–F) and (H–L) are shown as sticks and coloured by atom-type in the same way as protein side-chains, except that C atoms are orange. Ligands are: ergosterol, panels (B, H); catechin, panels (C, I); 7,8-dihydropteroic acid, panels (D, J); myricetin, panels (E, K); and 30,50-cyclic dAMP, panels (F, L).*

Based on the results of VS experiments and electron density map fitting, indicating that compounds comprising a steroid nucleus are likely to be among the preferred  $\sigma$ 1-R ligands, we inspected the results of VS against Hs $\sigma$ 1-R in coordinate sets 5HK1 and 6DK1 (data not shown), to identify a steroid-based compound suitable for experimental assessment of Hs $\sigma$ 1-R binding ability. We selected 16,17-didehydroprogesterone (LIPID MAPS ID: LMST02030163), because: (i) it is the compound with the lowest  $E_{\text{calc}}$  among the “best-E3” hits of VS against coordinate file 5HK1 comprising a steroid nucleus; (ii) it is a human endogenous compound; and (iii) it has a very short chain substituent at position 17. The molecular model of the complex between the Hs $\sigma$ 1-R in coordinate file 5HK1 and 16,17-didehydroprogesterone is shown in Figure 2.6. Examination of the  $\sigma$ 1-R residues at a distance  $\leq 4.0$  Å from the ligand reveals that the carbonyl oxygen in position 3 of 16,17-didehydroprogesterone may establish a polar interaction with the side-chain carboxylic group of E172, in the protonated state, thus replacing the basic amino group shared by classic pharmacophoric models. The non-polar remaining regions of 16,17-didehydroprogesterone establish hydrophobic interactions with hydrophobic residues lining the ligand binding site (i.e., V84, W89, M93, Y103, L105, F107, W164, I178, L182, A185 and Y206), most of which are the same residues that interact with ligands present in experimentally determined structures. Additionally, the carbonyl oxygen in position 20 of 16,17-didehydroprogesterone may establish a polar interaction with the side-chain hydroxylic group of T181.



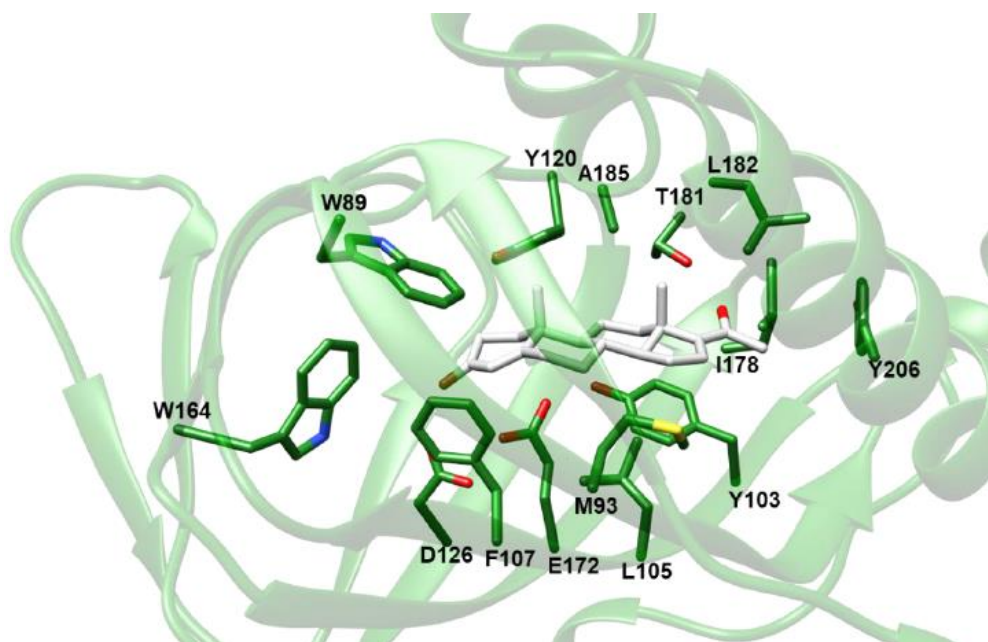


Figure 2.6 Molecular model of the complex between Hs $\sigma$ 1-R and 16,17-didehydroprogesterone built by VINA. The protein is shown as ribbon and coloured green. The ligand and the side-chains of residues at a distance  $\leq 4.0$  Å from the ligand are shown as sticks and coloured by atom-type: N, O and S atoms are blue, red and yellow, respectively; C is green for the protein and white for the ligand. The only exception is V84, which was removed from the picture for clarity.

Thanks to the fact that two of the tryptophan residues (i.e., Trp89 and Trp164) are part of the previously identified Hs $\sigma$ 1-R binding site, we were able to perform fluorescence titration to measure the affinity of selected molecules for this receptor. To validate the method, we performed fluorescence titration using pridopidine and iloperidone in addition to 16,17-didehydroprogesterone (Figure 2.7), since both molecules have been previously demonstrated to bind Hs $\sigma$ 1-R with high affinity using different techniques (Sahlholm et al. 2013). Pridopidine was initially shown, by  $^3\text{H}$ [(+)-pentazocine displacement experiments (Sahlholm et al. 2013), to have a  $K_i$  value for Hs $\sigma$ 1-R in the nanomolar

range (81.7 nM). Subsequently, we reported that both pridoipidine and iloperidone have  $K_D$  values towards  $Hs\sigma 1$ -R in the micromolar range, as measured by surface plasmon resonance (SPR) experiments.

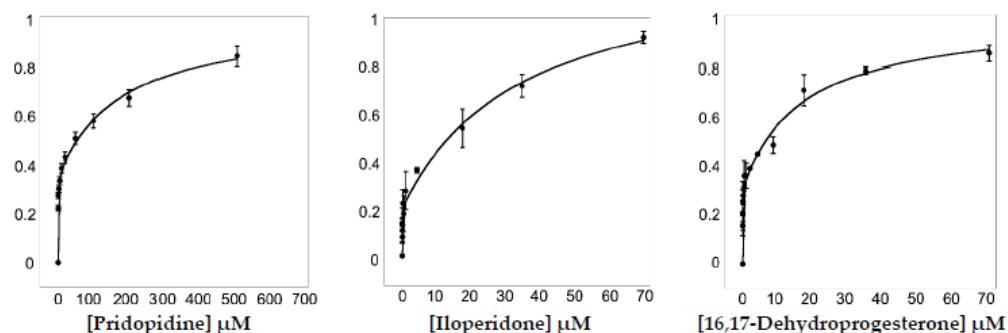


Figure 2.7. Titration of  $Hs\sigma 1$ -R with pridoipidine (left panel), iloperidone (central panel) and 16,17-didehydroprogesterone (right panel).

Data analysis suggested that  $Hs\sigma 1$ -R has two binding sites for the examined ligands, i.e., one high-affinity site and one low-affinity site. The  $K_D$  values of pridoipidine and iloperidone for the  $Hs\sigma 1$ -R high-affinity site are 254 and 19 nM, and therefore they are both higher than those previously measured by SPR (i.e., 15 and 5  $\mu$ M, respectively), although in both cases iloperidone resulted to have a higher affinity for  $Hs\sigma 1$ -R (Battista et al. 2021). The  $K_D$  value of 16,17-didehydroprogesterone for the high affinity site was 10 nM, very similar to that of iloperidone and 25-fold better than that of pridoipidine, indicating that 16,17-didehydroprogesterone is a very high affinity ligand for  $Hs\sigma 1$ -R.

Table 2.8. Dissociation constant ( $K_D$ ) values determined by fluorescence titration.

Ligands	$K_{D1}$ ( $\mu$ M)	$K_{D2}$ ( $\mu$ M)
Pridoipidine	$0.254 \pm 0.122$	$177 \pm 67.1$
Iloperidone	$0.019 \pm 0.015$	$33.10 \pm 14.41$
16,17-didehydroprogesterone	$0.010 \pm 0.004$	$15.81 \pm 4.03$

The dynamics of the apo form of the trimeric Hs $\sigma$ 1-R, embedded into a bilayer resembling the membrane composition of the ER, as shown in Figure 2.8, was investigated by means of all-atoms MD simulations for 1.5  $\mu$ s. Along the simulated trajectory, the RMSD of the backbone atoms of the trimer calculated with respect to the initial atomic positions increased to about 0.6 nm, while the RMSD of the single monomers in the last 500 ns fluctuated between 0.3 and 0.5 nm (data not shown). Overall, the secondary structures of the three monomers were found to remain rather stable along the simulation time, as confirmed by the time evolution of the number of H-bonds calculated within the single monomers (data not shown). Notably, the root mean square fluctuation of protein residues averaged over the last 500 ns of simulation showed different values for the three monomers, with the average fluctuations of monomer B larger than monomer C, and fluctuations of monomer C larger than monomer A (data not shown). Apart from the different fluctuations involving cytosolic protein loops, possibly due to a limited sampling time, significant differences between monomer B and the other monomers were found for residues 115–128 (in strands 4 and 5), and residues 172–188 (in strand 10 and helix 4). Looking at the simulated structures, we found that these two regions underwent a significant conformational change in monomer B. In Figure 2, the distance between the 4-helix (residues 180–188) and the coil between strands 4 and 5 (residues 118–121, shown in yellow and grey in panels B–C, respectively), both of which rest on the lipid bilayer, was monitored along the simulated trajectory, and a significant spacing between these two regions was observed in the last 150 ns of simulation. This conformational change also involved the 5-strand (residues 123–125, partially unfolded at the end of the simulation) and the 10-strand (residues 172–175), resulting into the

opening of the substrate cavity (Figure 2.9). Intriguingly, we found a possible correlation between such a conformational change and the breaking/formation of salt bridges between three residues, namely R175, E102 and E123, located in strand 10, 3 and 5, respectively (Figure 2.10). By monitoring the distances between R175 and the two residues E102 and E123 (Figure 2.10), we found three different behaviours for the three monomers. In monomer A, R175 formed an almost permanent salt bridge with E123. The same occurred in monomer B up to 1  $\mu$ s of simulation; subsequently, the salt bridge between R175 and E123 was lost and a new salt bridge between R175 and E102 was formed. In monomer C, for the first 600 ns the behaviour was similar to the behaviour found in monomer A, and then both the R175/E123 and R175/E102 salt bridges were lost. The breakage of the salt bridge between R175 and R123 occurring in monomer B after 1  $\mu$ s is likely to affect the subsequent spacing between the 5 and 10 strands, to which E123 and R175 belong, with the consequent opening of the cavity. In this context, E102 in the 3 strand may play a crucial role in triggering the opening of the substrate cavity, by forming a salt bridge with Arg175 and, therefore, inducing the breaking of the salt bridge between R175 and E123.

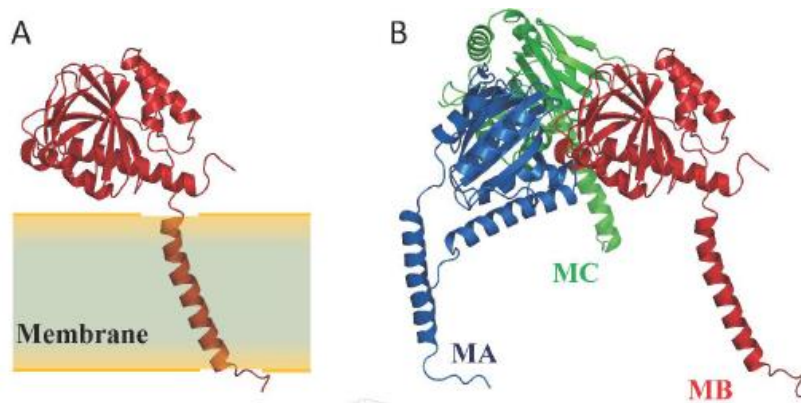


Figure 2.8. Structure of  $Hs\sigma 1-R$  protein. (A) A single monomer is shown as ribbon. The position of the monomer with respect to the membrane region of the protein is highlighted. (B) The homotrimer is shown as ribbon. The three monomers (MA, MB and MC) are coloured blue, red and green, respectively.

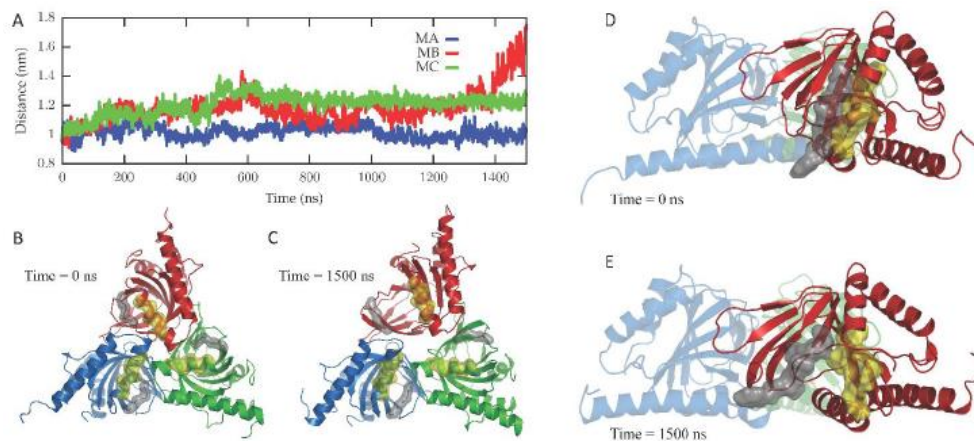


Figure 2.9. Conformational changes occurring in the  $Hs\sigma 1-R$  protein along the simulated trajectory. (A) The distances between the mass centres of the backbone atoms of residues 118–121 and 180–188 for the three monomers (namely, MA, MB and MC) are reported as a function of time. (B, C) Cartoon representation of the  $Hs\sigma 1-R$  protein viewed from the membrane side at 0 ns (B) and 1500 ns (C) of the MD simulation. The three monomers are coloured blue (M1), red (M2) and green (M3). The surface of residues 118–121 and 180–188 of the three monomers is also shown and coloured grey and yellow, respectively. (D, E) Cartoon representation of the  $Hs\sigma 1-R$  protein C-terminal domains, external to the membrane, at 0 ns (D) and 1500 ns (E) of the MD simulation. The surface of residues 122–126 and 171–176 of M2 (red) is also shown and coloured grey and yellow, respectively, highlighting the opening of the  $Hs\sigma 1-R$  binding site that occurs along the simulated trajectory.

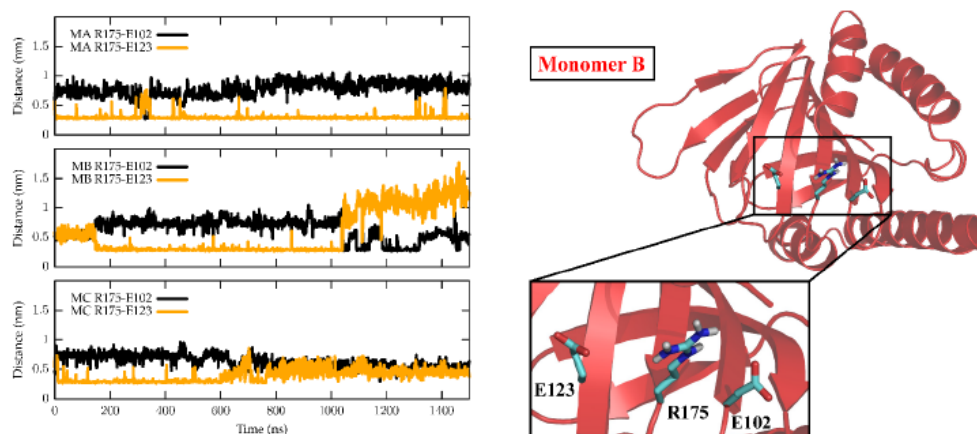


Figure 2.10. Time evolution of salt bridges between residues R175, E102 and E123 along the MD simulation. In the left panel, the minimum distances calculated between R175 and E102 (black) and between R175 and E123 (orange) are reported as a function of the simulated time for the three monomers, namely M1 (top panel), M2 (middle panel) and M3 (bottom panel). Right panel: cartoon representation of M2 in the starting conformation of the MD simulation, which is virtually identical to the crystallographic conformation. Zoomed-in inset: residues R175, E102 and E123 are shown as sticks and coloured by atom type: C, cyan; N, blue; O, red; H, white.

## 2.5 Conclusions.

The ER-resident  $\sigma 1$ -R is being intensively studied because of its involvement in several CNS disorders and because of the neuroprotective activity of its agonists. Many experimental and computational studies have provided valuable information on putative entrance and exit pathways to and from the ligand binding site, and on a number of compounds able to bind  $\sigma 1$ -R and elicit or inhibit specific activities. However, two essential receptor features, such as the route of ligand access to the binding site and the nature of the physiological

ligand(s), have not yet been unequivocally determined. On the other hand, both pieces of information would be required to both understand the physio-pathological role of  $\sigma$ 1-R and to design novel, higher affinity and higher specificity ligands endowed with specific agonistic or antagonistic activities. Among these uncertainties,  $\sigma$ 1-R is drawing attention as a therapeutic target because its agonists have been shown to be able to counteract neurodegenerative disease processes. As an example, one of these agonists, pridopidine, is safe (Squitieri et al. 2013; Reilmann et al. 2019) and has beneficial effects on functional capacity (McGarry et al. 2020) and brain connectivity (Sánchez-Castañeda et al. 2017) of HD patients, as well as beneficial effects for HD cell and mouse models (Geva et al. 2016; Ryskamp et al. 2017; Eddings et al. 2019; Squitieri et al. 2015).

For these reasons, in these two works, we focalised our attention on three objectives: I) Try to identify an already known drug able to act on  $\sigma$ 1-R; II) According to data present in literature, speculate on the possible physiological ligand and III) suggest a way in which the ligand could enter in the binding site.

To achieve the first goal, we performed a virtual screening (VS) of a library comprising all FDA-approved drugs available on Zinc15 database. The selected top20, coming from the first VS, were redocked and analysed with a combination of own and known tools. Finally, we visually inspected these 20 compounds and selected six drugs FDA-approved: iloperidone, paliperidone, flibanserin, linagliptin, vilazodone and nilotinib. In vitro SPR experiments demonstrated that the computational strategy was successful, since each of the selected compounds was proved to be able to bind  $\sigma$ 1R at concentrations in the micromolar range. Importantly, the value of ligand  $\sigma$ 1R complexes KD is in

the same range as that measured for pridopidine, and even lower for three of these compounds (flibanserin, iloperidone, linagliptin). Importantly, experiments performed on HD fibroblasts showed that the six selected compounds were able to increase the growth of HD patient fibroblasts, albeit to a different extent from one another, indicating that, like pridopidine, they also exert agonistic activity upon  $\sigma$ 1R binding. Among the six tested drugs, iloperidone appears to have the most favourable effects. It had 3-fold higher affinity than pridopidine for  $\sigma$ 1R in vitro and significantly improved HD fibroblast growth. Iloperidone activity was specific for HD cells, since it did not affect healthy cell growth to a significant extent. Iloperidone is an anti-psychotic drug used for schizophrenia treatment, which is able to bind several dopamine and serotonin receptor isotypes (Mauri et al. 2014). In particular, it binds dopamine D2 and serotonin 5-HT2A receptors in the caudate nucleus and putamen of the brain (Kongsamut et al. 1996). The second and the third goals were more theoretical, but they gave us some interesting data to speculate on. With the VS on H $\sigma$ 1R and its homologous ERG2, we were able to try to understand the possibly nature of the endogenous ligand of H $\sigma$ 1R. We performed a VS on a large dataset comprising approved drugs (FDA and from other World agency), human metabolites (endogenous or imported with the diet), known ligands of H $\sigma$ 1R and H $\sigma$ 2R and a large amount of lipids (such as sterols, steroids) according to the role of the only known homologous (fungal ERG2). VS results were sorted based on their predicted interaction energy ( $E_{\text{calc}}$ ) with H $\sigma$ 1-R. Since individual  $E_{\text{calc}}$  values are expected to have a 2–3 kcal/mol standard deviation from the actual protein–ligand interaction energy, we focused our analysis of the results of each VS on the set of hits whose  $E_{\text{calc}}$  value was within one standard deviation from that of the best hit (“best-E3”



subsets). The VS results shown that, for all the three structures tested, statistically the percentage of lipidic compounds present in the best-E3 subsets were higher than other type of molecules. Especially if we consider the metabolites which were, in total, an order of magnitude higher than the lipid compounds themselves. These data suggest that the physiological agonist(s) could be a sterol/steroid. Next, to get additional clues about putative physiological  $\sigma$ 1-R ligands, we took advantage of the two recently determined apo structures of Xl $\sigma$ 1-R that showed an electron density peak in the ligand binding site. We performed a VS against both structures using a library of ligands comprising all yeast metabolites with MW  $\leq$ 400 Da, because the Xl $\sigma$ 1-R protein used for X-ray studies was expressed in and purified from yeast, and because inspection of the electron density map indicated that the yeast metabolite giving rise to the unfitted electron density was not larger than a cholesterol molecule. Visual inspection of the 88 compounds that resulted in the “best-E3” subsets obtained from VS against both structures led us to select five structurally diverse molecules that were likely to best fit the experimentally determined electron density on the basis of both their size and shape. According to both visual inspection and measurement of average B-factor values, and in agreement with these selection criteria, all five molecules were shown to fit well in the electron density map, within the ligand binding site of both the closed and the open-like form of apo Xl $\sigma$ 1-R, ergosterol being the best fitting compound. Since both VS against Hs $\sigma$ 1-R structures and fitting into the electron density map of Xl $\sigma$ 1-R structures indicated steroid-based molecules as preferred  $\sigma$ 1-R ligands, we decided to measure experimentally the affinity of one steroid-based compound against human Hs $\sigma$ 1R. We selected 16,17-didehydroprogesterone because it is a human endogenous compound, it

does not contain long substituents that may affect  $\sigma$ 1-R binding and it is within the group of hits predicted by VS procedures to bind H $\sigma$ 1-R with highest affinity. To this end, we implemented a fluorescence titration procedure and used both pridopidine and iloperidone as positive controls. The experimental data denote that 16,17-didehydroprogesterone is an even better H $\sigma$ 1-R ligand than iloperidone and pridopidine, the KD1 value for the high-affinity site being 10 nM, which is two orders of magnitude lower than that of pridopidine for the same site in the same assay. Analysis of the complex between H $\sigma$ 1-R and 16,17-didehydroprogesterone built by the VINA program indicates that the interaction mode of the ligand with the receptor was very similar to that observed for the ligands present in experimentally determined structures of complexes with H $\sigma$ 1-R, and to the shared features of pharmacophoric models. The main difference is in the replacement of the basic amino group shared by those ligands and pharmacophoric models with the carbonyl oxygen at position 3 of 16,17-didehydroprogesterone in the polar interaction with the conserved E172. In this model, the carbonyl oxygen of the ligand is expected to act as the electron donor and the side-chain carboxylic group of E172 is expected to be protonated and act as an electron acceptor. Given the results of this work and the well-known ability of steroid-based molecules to act as  $\sigma$ 1-R agonists or antagonists, we suggest that pharmacophoric models for H $\sigma$ 1-R ligands should be expanded to include an oxygen-atom-containing group, with the aim to establish a polar interaction with E172, as an alternative to a basic nitrogen. Finally, we used molecular dynamics (MD) to evaluate how physiological ligand could enter or exit in the binding site. First, we chose the trimeric structure of H $\sigma$ 1-R, as opposed to the monomer used in previous simulations, because the trimer is the minimal quaternary assembly that is present in all

Hs $\sigma$ 1-R and Xl $\sigma$ 1-R structures. In our system, while the overall secondary structures of the three monomers were substantially stable during the simulation time, significant conformational changes occurred in two regions of monomer B flanking the ligand binding site, and led to the opening of a cavity between the ligand binding site and the lipid bilayer. The first of these regions comprises residues 115–128, including part of the  $\beta$ 4 and  $\beta$ 5 strands and the loop comprised between them, and the second region comprises residues 172–188, including part of the  $\beta$ 10 strand and the 4 helix. These results are in partial agreement with the results of previous MD simulations performed on Hs $\sigma$ 1-R, which highlighted structure alterations affecting residues E123 ( $\beta$ 5) and R175 ( $\beta$ 10) (Schmidt et al. 2018) each of which is comprised in one of the regions that unfolds in our studies. However, our results indicate that ligand entrance and exit occur via the protein side leaning on the membrane (pathway 2), whereas both previous MD studies point to an opening towards the aqueous medium (pathway 1) (Figure 6.11) (Schmidt et al. 2018; Rossino et al. 2020).

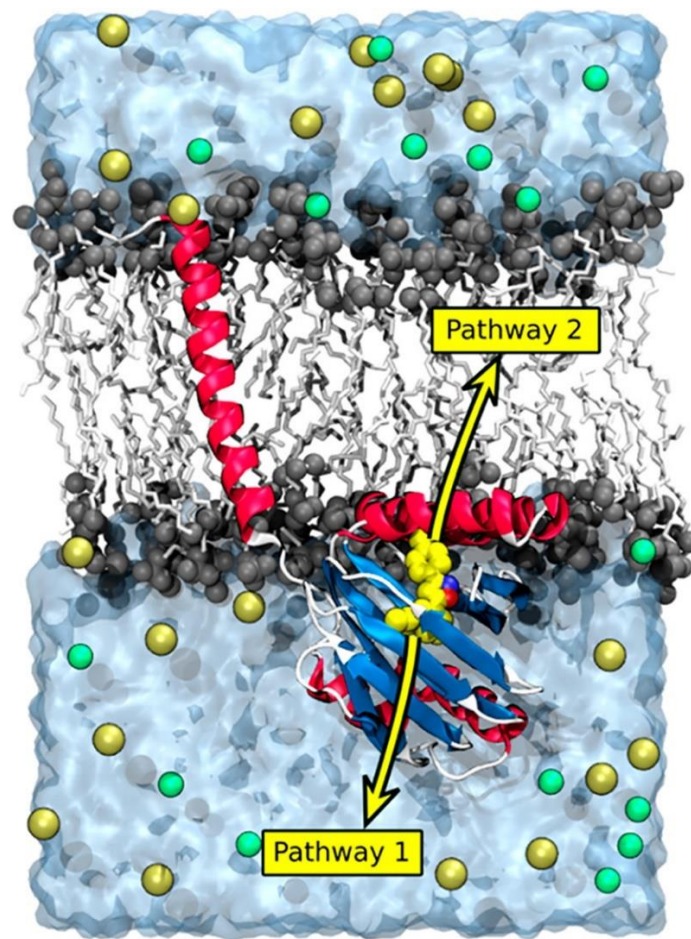


Figure 2.11. Ligand access pathways to the binding pocket. The presumed ligand-accessible pathways are illustrated using yellow arrows. Hs $\sigma$ 1-R bound to PD144418 is represented. Hs $\sigma$ 1-R is shown in a cartoon representation and coloured by secondary structure. PD144418 is displayed as spheres coloured by atom type: carbon in yellow, oxygen in red, and nitrogen in blue. Lipid head groups and tails are shown as grey spheres and white sticks, respectively. Water molecules are represented by a blue transparent surface. Na<sup>+</sup> and Cl<sup>-</sup> ions are shown in light yellow and green spheres, respectively. Figure adapted from Rossino et al. 2020.

3. Computational study of the mode of COX binding of thiocanthal and thiocanthol, two novel biologically active compounds that can be obtained from extra virgin olive oil by eco-sustainable procedures.

### 3.1 Background.

Extra virgin olive oil (EVOO) is a good source of antioxidants and bioactive compounds, and its content of phenolic compounds plays a central role in the beneficial effects of the Mediterranean diet (Servili et al. 2014). The two most abundant phenols found in EVOO are the secoiridoids oleocanthal and oleacein, which are derivatives of the simple phenols tyrosol and hydroxytyrosol, respectively (Figure 3.1) (Lozano-Castellón et al. 2020).

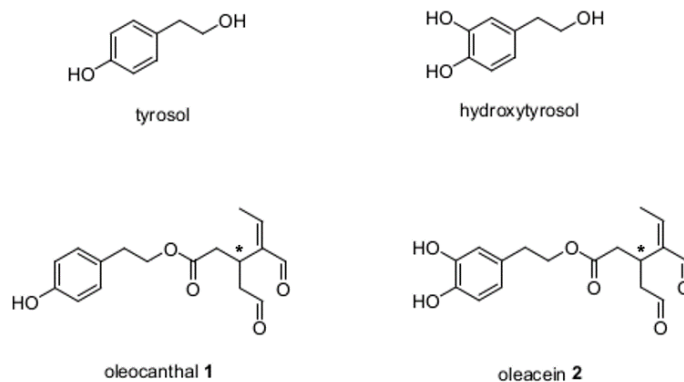


Figure 3.1. Simple phenols (tyrosol and hydroxytyrosol) and secoiridoids (oleocanthal 1 and oleacein 2) found in EVOO. The asterisk indicates the position of the chiral center (3S carbon).

Oleocanthal and oleacein are synthesized, as 3S enantiomers, by the combination of two processes, the first occurring in *Olea europaea* and the second during EVOO production by olive crushing and malaxation. Both oleocanthal and oleacein are endowed with several important biological properties and pharmacological activities, including anticancer, antioxidant, anti-inflammatory, and neuro- and cardiovascular protective effects. Notably, the anti-inflammatory effects have been ascribed to the inhibition of cyclooxygenase (COX) activity, both at translational and post-translational levels (Parkinson & Keast 2014). Because of these activities, in the last decade oleocanthal and oleacein have been proposed to be both ideal lead compounds and multifunctional drugs themselves.

Given the lability of the aldehyde functional groups of oleocanthal and oleacein, and the presence of these compounds at highly variable concentrations in olive oil only, the strategies developed to obtain them by extraction and purification from EVOO or chemical synthesis until now are time-consuming and expensive, and yield very low product amounts. For these reasons, oleocanthal and oleacein are unsuitable for large scale production (Francioso et al. 2020; Grewal et al. 2020).

As mentioned above, the presence of highly reactive aldehydic functions makes oleocanthal and oleacein susceptible to oxidative/degradative processes and, therefore, it is an obstacle for the obtainment of these compounds from EVOO. However, the chemical reactivity of aldehydic groups can also be exploited to transform oleocanthal and oleacein into more hydrophilic molecules. To this end, we introduced a sulfonate group in the iridoid skeleton of the oil-soluble oleocanthal and oleacein *via* metabisulfite reduction. The novel hydro-soluble derivatives thus obtained, which we named thiocanthal

and thiocanthol (Figure 3.2), were subsequently isolated using a two-step and eco-sustainable strategy. Due to the introduction of another chiral carbon, both of them are present as a mixture of diastereomers: 3*S*,8*R*-thiocanthol and 3*S*,8*S*-thiocanthol; and 3*S*,8*R*-thiocanthal and 3*S*,8*S*-thiocanthal.

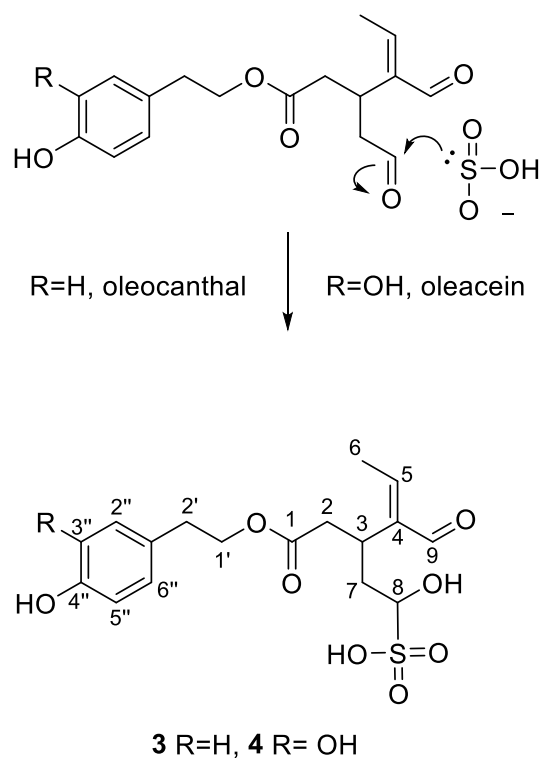


Figure 3.2. Reaction mechanism of regioselective sulfonate attack to oleocanthal (*R*= H) or oleacein (*R*= OH) and generation of thiocanthal (**3**) and thiocanthol (**4**).

After the purification, the sulfur-containing derivatives were characterized, and their biological activity as COX inhibitors was tested and compared with that of the respective oleocanthal and oleacein precursors. We found that the COX-inhibitory activity of the sulfonate derivatives is substantially unaltered compared to their precursors, indicating that thiocanthal and thiocanthol may

be exploited as more convenient lead compounds for drug development than the respective oleocanthal and oleacein precursors.

### 3.2 Aim of the work.

The overall aim of this work was to try to find a “green” method to extract oleocanthal or oleacein, two compounds endowed with anti-inflammatory activity, from olive oil.

Our collaborators developed an environmentally compatible procedure to extract and purify two chemical derivatives of these compounds, named thiocanthal and thiocanthol, and determined their biological activity as COX inhibitors.

Our aim in this work was to investigate putative modes of COX-binding by compounds thiocanthal, thiocanthol, oleocanthal and oleacein, and compare them with those of ibuprofen and other well-known COX ligands. To this end, we have employed computational techniques (VS) to calculate the predicted binding energy of the interaction between the ligands under investigation and the COX-2 and COX-1 enzymes, and build and analyse models of the ligand-receptor complexes.

### 3.3 Methods.

*Receptor preparation.* The atomic co-ordinates of human COX-1 and COX-2 3D structures that had been experimentally determined by X-ray



crystallography and are publicly available through the PDB (<https://www.rcsb.org/>) (Berman et al. 2000) were downloaded. Structures in co-ordinate files with PDB IDs 6Y3C (Resolution = 3.36 Å) (Miciaccia et al. 2021) and 5F19 (Resolution = 2.06 Å) (Lucido et al. 2016), which have been determined with best resolution among those available for human COX-1 and COX-2, respectively, were selected for computational VS. Each of these structures comprises two chains, one for each of the monomers that takes part in the biologically active homodimer. For both structures, chain A was selected. Crystallographic waters and ligand molecules were removed using Chimera. AutoDock Tools (ADT) (Morris et al. 2009) v.1.5.6 was used to add hydrogen atoms, merge non-polar hydrogen atoms and automatically assign Gasteiger charges.

*Ligand preparation.* The small molecule compound dataset used for computational VS comprised:

- I) The four compounds analysed in this work, namely oleocanthal, oleacein, thiocanthal and thiocanthol, whose structures were manually drawn using MarvinSketch 19.12 (<https://chemaxon.com>).
- II) All ligands in complex with human COX-1 and COX-2 in the experimentally determined structures available from the PDB.
- III) Ibuprofen, a known COX-binding drug, whose co-ordinates were downloaded from the PDB.
- IV) Etoricoxib, celecoxib, indomethacin and aspirin, which are known to be selective COX-2 inhibitors, whose co-ordinates were downloaded from the ZINC database (<http://zinc15.docking.org/>) (Sterling & Irwin 2015).

All compounds present in these datasets were converted, first, from the .smi to the PDB format, using an *ad hoc* developed bash script that includes the

“molconvert” command from MarvinSketch 19.12; then, from the PDB to the pdbqt format, using a script from AutoDock Tools 1.5.6 (ADT) (Morris et al. 2009) where the following parameters were added: “-A ‘hydrogen\_bonds’” to both add hydrogens and build bonds among non-bonded atoms; and “-U ‘nphs\_lps’” to merge both non-polar hydrogens and lone pairs.

*Virtual Screening.* Computational VS was performed using the following space searching parameters: For 5F19: spacing value at 0.375 Å; center on coordinates 15, 43 and 63; and 16X16X16 grid points; For 6Y3C: spacing value at 0.375 Å; center on coordinates -33, -44.4; and 16X16X16 grid points. These parameters were chosen following visual inspection of COX-1 and COX-2 binding pockets, after optimal superposition of the two structures, using the Autodock Vina tool included in the USCF Chimera suite. VS was performed using the program VINA (Trott & Olson 2009) with the following parameters: “--num\_modes 100”, which represents the maximum number of binding modes to generate, and “--energy\_range 9”, to maximize the energy difference between the best binding mode and the worst one. Additionally, all conformations (poses) were kept, rather than only those with a VINA score better than a given threshold. All other parameters had default values. The ordered list of the VS results was extracted with “vina\_screen\_get\_top.py” script from Autodock Vina tools (Trott & Olson 2009).

*Results analysis.* Python scripts were developed to parse Vina output files and perform a preliminary analysis of COX-ligand complexes. In particular, the pose energy of each ligand was extracted from the Vina pdbqt file; features of protein-ligand interactions, such as hydrogen bonds, number of contacts and number of unfavourable interactions (clashes), were calculated using the structure visualization and analysis program Chimera (Pettersen et al. 2004),

following re-building of receptor-ligand complexes. Structure visualization and analysis were performed using the programs CHIMERA, InsightII (Accelrys Inc.) and Swiss-PDBViewer (Guex & Peitsch 1997). Protein residues are defined to be in contact with the ligand if they have at least one atom at a distance 4.0 from a ligand atom.

### 3.4 Results.

The numerical results of our VS are shown in Table 3.1. Taking into account that the values of  $E_{\text{calc}}$  produced by the program used for VS have been reported to have a standard deviation of 2–3 Kcal/mol (Huey et al. 2007), examination of this table indicates that both stereoisomers of thiocanthal and thiocanthol are predicted to bind to COX-2 with affinity not statistically different from that of oleocanthal, oleacein and ibuprofen, and of the compounds with the highest (celecoxib) and lowest (salicylic acid) COX-2 binding  $E_{\text{calc}}$ .

*Table 3.1. Results of VS towards COX-2 (top) and COX-1 (bottom). Only representative compounds are shown, namely, the four compounds discussed in this work (i.e., oleocanthal, oleacein, thiocanthal, thiocanthol); well-known anti-inflammatory drugs (e.g., ibuprofen, indomethacin, aspirin); selective COX-2 inhibitors; and other compounds whose structure has been determined in complex with COX proteins available from the PDB. Source: Compounds were downloaded from the PDB, whenever available, or ZINC database; t.w.: compounds studied in this work. ID (PDB name): identifier of the ZINC database, PDB file, or number assigned in this work. In the case of PDB files, the PDB three-letter code used to indicate compound co-ordinates is reported in parenthesis. Common Name: Name of compound in the ZINC "molecule name" or PDB "synonyms" fields or used in the text.  $E_{\text{calc}}$  (kcal/mol): Energy of compound-COX interaction calculated by VINA. #Hb, #CO and #CL: total number of hydrogen bonds, contacts, and unfavourable interactions (clashes) between compound and COX calculated by Chimera, respectively. Activity: Compound activity from DrugBank (i.e.,*

NSAID: nonsteroidal anti-inflammatory drug; SC2I: selective COX-2 inhibitor), PDB file (i.e., Cofactor) or this work (i.e., C.I.: COX inhibitor).

Source	ID (PDB name)	Common Name	E <sub>calc</sub>	Hb	CO	CL	Activity
PDB	5IKV (FLF)	Flufenamic acid	-7,8	0	58	0	NSAID
PDB	5IKR (ID8)	Mefenamic acid	-7,7	0	55	0	NSAID
PDB	4PH9 (IBP)	Ibuprofen	-7	1	44	0	NSAID
PDB	5IKT (TLF)	Tolfenamic acid	-7	1	52	0	NSAID
PDB	5IKQ (JMS)	Meclofenamic acid	-6,8	1	55	2	NSAID
t.w.	1	Oleocanthal	-6,1	0	91	0	C.I.
t.w.	2	Oleacein	-6	5	69	3	C.I.
PDB	5F1A (SAL)	Salicylic acid	-5,7	0	22	0	NSAID
ZINC	53	Aspirin	-5,6	0	35	0	NSAID
t.w.	3	Thiocanthal (8R)	-5,6	0	94	2	C.I.
t.w.	4	Thiocanthol (8R)	-5,3	4	113	1	C.I.
	3	Thiocanthal (8S)	-4,8	0	72	1	C.I.
PDB	5KIR (RCX)	Rofecoxib	-4,5	1	74	4	NSAID
ZINC	601283	Indomethacin	-4,4	0	95	3	NSAID
	4	Thiocanthal (8S)	-3,6	1	81	5	C.I.
ZINC	2570895	Celecoxib	-3	1	107	4	SC2I
ZINC	579472	Etoricoxib	-1,1	2	104	8	SC2I
PDB	5F1A (COH)	Protoporphyrin	15,8	2	194	34	Cofactor

Source	ID (PDB name)	Common Name	E <sub>calc</sub>	Hb	CO	CL	Activity
ZINC	2570895	Celecoxib	-10.2	3	71	0	SC2I
ZINC	601283	Indomethacin	-9.6	0	45	0	NSAID
PDB	5IKV (FLF)	Flufenamic acid	-9.3	0	30	0	NSAID
ZINC	579472	Etoricoxib	-9.2	3	66	0	SC2I
PDB	5KIR (RCX)	Rofecoxib	-9.1	1	50	0	NSAID
PDB	5IKQ (JMS)	Meclofenamic acid	-8.4	0	40	0	NSAID
PDB	5IKR (ID8)	Mefenamic acid	-8.3	0	33	0	NSAID
PDB	5IKT (TLF)	Tolfenamic acid	-8.0	0	43	0	NSAID
t.w.	3	Thiocanthal (8S)	-7.9	1	67	0	C.I.
t.w.	4	Thiocanthal (8R)	-7.8	4	85	0	C.I.
t.w.	4	Thiocanthal (8S)	-7.8	4	81	0	C.I.
t.w.	3	Thiocanthal (8R)	-7.6	1	71	0	C.I.
PDB	4PH9 (IBP)	Ibuprofen	-7.3	0	28	0	NSAID
t.w.	1	Oleocanthal	-7.1	1	56	0	C.I.
t.w.	2	Oleacein	-7.0	1	59	0	C.I.
ZINC	53	Aspirin	-6.1	0	29	0	NSAID
PDB	5F1A (SAL)	Salicylic acid	-5.5	0	23	0	NSAID
PDB	5F1A (COH)	Protoporphyrin	0.9	3	139	16	Cofactor

Additionally, the E<sub>calc</sub> values of thiocanthol, thiocanthal, oleocanthal, oleacein and ibuprofen towards COX-2 are lower than those towards COX-1 and, in the

case of both 3S,8S-thiocanthal and 3S,8S-thiocanthal the difference in  $E_{\text{calc}}$  values is higher than 3 Kcal/mol, as in the case of the selective COX-2 inhibitors celecoxib and etoricoxib.

Structure analysis of COX-2 complexes produced by VS and those downloaded from the PDB, indicates that most COX-2 residues predicted to be involved in interactions with thiocanthal, thiocanthol, oleocanthal and oleacein are also involved in ligand binding in experimentally determined 3D structures. A model of the COX-2 complexes with 3S,8R-thiocanthal and 3S,8R-thiocanthal built by VINA is shown in Figure 3.3 (COX-2 complexes with the 3S,8S-thiocanthal and 3S,8S-thiocanthal diastereoisomers are very similar, suggesting that the different configuration at the level of carbon atom (does not significantly affect ligand binding)).

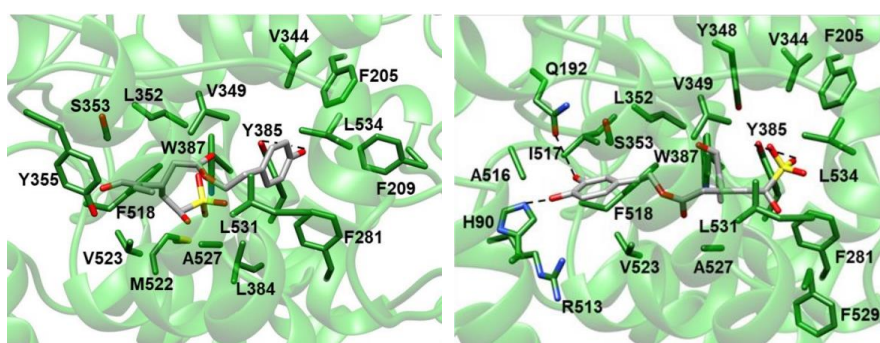


Figure 3.3. Model of human COX-2 in complex with 3S,8R-thiocanthal (left panel) and 3S,8R-thiocanthal (right panel) built by the VINA program. COX-2 Ca atoms are shown as a green-coloured ribbon. Ca atoms and side-chains of residues interacting with the ligand are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, green. Ligands are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, white. Hydrogen bonds are shown as dashed lines.

For comparison purposes, models of COX-2 complexes with oleocanthal, oleacein and ibuprofen are shown in Figure 3.4 and the model of COX-2

complex with celecoxib and the experimental structures of COX-2 complexes with rofecoxib and salicylic acid are shown in Figure 3.5.

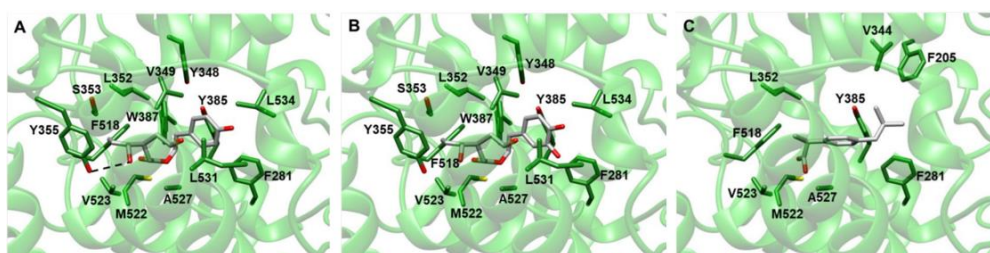


Figure 3.4. Models of human COX-2 in complex with oleocanthal 1 (A), oleacein 2 (B) and ibuprofen (C) built by the VINA program. COX-2 Ca atoms are shown as green-coloured ribbon. Ca atoms and side-chains of residues interacting with the ligand are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, green. Ligands are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, white. Hydrogen bonds are shown as dashed lines.

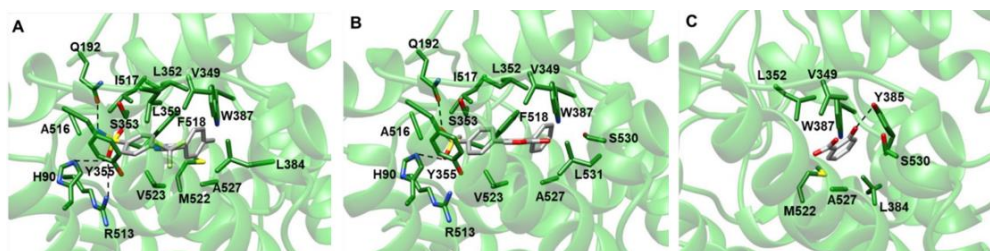


Figure 3.5. Models of human COX-2 in complex with celecoxib (A) built by the VINA program and experimental structures of human COX-2 in complex with rofecoxib (B) and salicylic acid (C) downloaded from the PDB. COX-2 Ca atoms are shown as green-coloured ribbon. Ca atoms and side-chains of residues interacting with the ligand are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, green. Ligands are shown as sticks and coloured by atom-type: N, blue; O, red; S, yellow; C, white. Hydrogen bonds are shown as dashed lines.

### 3.5 Conclusions.

EVOO is a good source of antioxidants and bioactive compounds. Tyrosol, hydroxytyrosol, oleacein and oleocanthal show important roles in antioxidant and anti-inflammatory processes and as cardio- and neuroprotective agents. However, oleacein and oleocanthal purification is an expensive and cumbersome procedure, with low yields and involving large consumption of organic solvents and chemicals. In this work a new green procedure to obtain two purified derivatives of oleacein and oleocanthal from EVOO was developed. The reactivity of the aldehydic moieties of these compounds was exploited to introduce a sulfonate group via metabisulfite reduction of the iridoid skeleton. The novel water-soluble derivatives were isolated using a two-step non-organic chromatographic strategy taking the advantage of the newly introduced ionizable group and the pre-existing phenolic and catecholic moieties. The novel purified sulfur-containing derivatives, named thiocanthal and thiocanthol, showed anti-inflammatory properties *in vitro*, since they inhibit the COX activity to a similar extent to their bioactive precursors oleacein and oleocanthal. The results of our *in-silico* analysis support and strengthen these results, since thiocanthal and thiocanthol are predicted to bind to COX-2 with affinity in the same range as well-known COX-2 binders and higher than their binding affinity towards COX-1. Additionally, analysis of the models of COX-2 and COX-1 complexes with the compounds studied in this work (i.e., thiocanthol and thiocanthal), their precursors (i.e., oleacein and oleocanthal) and other well-known COX inhibitors, obtained by VS indicate that COX residues involved in the interaction with our new compounds and with known COX inhibitors are largely the same (table 3.2).

Table 3.2. Interactions between COX2 and known ligands observed in PDB, Protein Data Bank, structures (left) or predicted by VS, Virtual Screening, (right). Residue names are coloured red, blue, and black if the atoms involved in the interaction with the ligand are negative or partially negative, positive or partially positive, and hydrophobic, respectively. The label ".BB" indicates that the interactions occur with the backbone. TLF, Tolfenamic acid. SAL, Salicylic acid. RCX, Rofecoxib. JMS, Meclofenamic acid. IDB, Iodipamide. FLF, Flufenamic acid. BOG, Beta-Octylglucoside. Oleac, oleacein. Oleoc, oleocanthal. Thiol, thiocanthol.

	PDB							VS							
	TLF (5IKT)	SAL (5F1A)	RCX (5KIR)	JMS (5IKQ)	ID8 (5IKR)	FLF (5IKV)	BOG (5F1A)	Oleac	Oleoc	Thiol	Thial	FLF	IBP	RCX	Celec
S 53			H90								H90			H90	H90
H 90															
M 113											M113.BB				
V 116						V116									
R 120						R120									
E 179							E179								
K 180							K180								
R 185							R185								
Q 192			Q192								Q192			Q192	Q192
F 205										F205	F205		F205		
T 206										T206					
F 209										F209		F209			
M 223															
V 344										V344	V344		V344		
Y 348	Y348			Y348				Y348	Y348		Y348	Y348			
V 349	V349	V349	V349	V349	V349	V349		V349	V349	V349	V349	V349		V349	V349
L 352	L352	L352	L352	L352	L352	L352		L352	L352	L352	L352		L352	L352	L352
S 353	S353		S353	S353	S353	S353		S353	S353	S353	S353		S353	S353	S353
Y 355	Y355		Y355	Y355	Y385	Y355		Y355	Y355	Y355				Y355	Y355
L 359						L359									L359
F 381								F381	F381	F381	F381	F381	F381		
L 384		L384								L384		L384			
Y 385		Y385		Y385		Y385		Y385	Y385	Y385	Y385	Y385	Y385		
W 387	W387	W387	W387	W387	W387	W387		W387	W387	W387	W387	W387			W387
R 438							R438								
E 486							E486								
R 513			R513								R513			R513	R513
A 516			A516								A516			A516	A516
I 517			I517								I517			I517	I517
F 518			F518						F518	F518	F518		F518	F518	F518
M 522	M522	M522		M522	M522	M522		M522	M522	M522			M522	M522	M522
V 523	V523		V523		V523	V523		V523	V523	V523	V523		V523	V523	V523
G 526	G526	G526		G526	G526	G526		G526	G526	G526	G526	G526	G526	G526	G526
A 527	A527	A527	A527	A527	A527	A527		A527	A527	A527	A527	A527	A527	A527	A527
F 529								F529.BB			F529.BB				
S 530	S530	S530	S530	S530	S530	S530									
L 531	L531		L531	L531	L531	L531		L531	L531	L531	L531	L531.BB	L531.BB	L531	L531
G 533												G533			
L 534								L534	L534	L534	L534	L534			

In conclusion, this work presents a novel water-compatible synthesis and a rapid and eco-friendly purification method to produce the two novel COX



inhibitors thiocanthal and thiocanthol from EVOO. These compounds are attractive candidates for future *in vivo* studies and applications as potential selective COX-2 inhibitors.

## 4. Identification of the molecular basis of the antiproliferative activity of *Arabidopsis thaliana* sirtuins.

### 4.1 Background.

The sirtuin family comprises highly conserved NAD-dependent enzymes with broad cellular functions, including life span regulation, DNA repair, metabolism, stress resistance, proliferation, and energy production (*Chang & Guarente 2014; Carafa et al. 2016*). These functions result from sirtuin ability to remove a large array of acyl modifications from cell proteins. Most sirtuins act as deacetylases of histone and non-histone proteins, and some of them show additional enzymatic activities (*Bheda et al. 2016*). Since these enzymes are regulated by NAD availability, and hence by the nutritional state of the cell, they play important roles in growth and development as well. Protein substrates have been identified for many sirtuins, especially those involved in energy metabolism or stress defence.

Seven members of the sirtuin family have been found in mammals. All of them share a conserved NAD<sup>+</sup> binding catalytic domain but differ in subcellular localization, enzymatic activity, and function. SIRT1, SIRT6 and SIRT7 are predominantly located in the nucleus; SIRT2 is both in the nucleus and in the cytoplasm; SIRT3, SIRT4 and SIRT5 are primarily present in mitochondria (*Carafa et al. 2016*). Sirtuins have been identified also in plants and shown to have different enzymatic activities and localization, but the actual function has been demonstrated for some plant sirtuins only (*Zheng 2020*). Plant sirtuins

have been studied mainly in *Arabidopsis thaliana* and *Oryza sativa*, in both of which only two genes have been detected. Based on the analysis of conserved domain regions, AtSRT1 and AtSRT2 gene products have been reported to belong to class II and class IV within the sirtuin family, and to have highest sequence similarity with human SIRT6 (HsSIRT6) and SIRT4 (HsSIRT4), respectively (Pandey et al. 2002). AtSRT1 and AtSRT2 exert deacetylase activity (Hollender & Liu 2008) but no other enzymatic activity has been described so far. Deacetylase activity has been detected for OsSRT1 and OsSRT2 (Huang et al. 2007; Zhong et al. 2013) as well. Moreover, histone deacetylation activity has been reported for OsSRT1 (Lu et al. 2018). In both *Arabidopsis* and rice, AtSRT1 and OsSRT1 are mainly localized in the nucleus (Huang et al. 2007; Liu et al. 2017), and AtSRT2 and OsSRT2 in mitochondria (König et al. 2014), although nuclear localization of AtSRT2 has also been reported (Wang et al. 2010). The functions of plant sirtuins are far from being completely clarified. AtSRT1 deacetylates and represses several genes involved in plant stress response (Liu et al. 2017). Further, AtSRT1 regulates primary metabolism by modulating the transcriptional factor AtMBP-1 by direct lysine-deacetylation of the protein (Liu et al. 2017). AtSRT2 is likely to play a role in energy metabolism, since it directly deacetylates mitochondrial proteins such as ATP synthase and ADP/ATP carriers (König et al. 2014). Moreover, a negative role in plant basal defence has been proposed, since AtSRT2 expression is downregulated in response to infection with the tomato pathogen, *Pseudomonas syringae* pv. tomato DC3000 (Wang et al. 2010). The two plant sirtuins are usually involved in different and independent processes, but in at least one study they have been reported to work together in the same metabolic pathway, by interacting with the ENAP1 factor and mediating

ethylene-induced transcriptional repression by H3K9 deacetylation (Zhang et al. 2018). Therefore, data available so far only suggest that plant sirtuins, like animal homologs, play a role in metabolic pathways related to energy production.

Based on sequence similarity between AtSRT1 and SIRT6, and between AtSRT2 and SIRT4, we hypothesized that Arabidopsis sirtuins play a role similar to that of the human enzymes, which are engaged in proliferation control. To evaluate this hypothesis, *srt1* and *srt2* mutants of *Arabidopsis* were produced, which lack the SRT1 and SRT2 gene, respectively, and evaluated, in comparison with wild-type (WT) plants, for several parameters related to cell proliferation including plant weight, root length and variation in DNA synthesis and GDH activity, which we hypothesized to mediate sirtuin engagement in cell proliferation control, as previously reported for mammalian SIRT4 (Jeong et al., 2013). Our role in this work was to perform bioinformatics analyses to identify the molecular determinants of AtSRT1 and AtSRT2 biological activity.

#### 4.2 Aim of the work.

Our aim in this work was to identify the sequence and structure determinants underlying the antiproliferative activity of Arabidopsis SRT1 and SRT2 demonstrated by our experimental collaborators. To this end, we obtained the molecular models of Arabidopsis SRT1 and SRT2 and human SIRT4 and compared them with one another and with the experimentally determined

structures of human SIRT6, both at the global and at the substrate binding site level.

### 4.3 Methods.

The amino acid sequences of AtSRT1 and AtSRT2 proteins and of their human homologs HsSIRT6 and HsSIRT4 were downloaded from the UniProt web site (<https://www.uniprot.org/>). The BLAST program (Altschul et al. 1997) was used for pair-wise sequence comparisons, and to search NCBI sequence databases for sequences homologous to those given as input. ClustalO (Sievers et al. 2011) was used to generate multiple sequence alignments (Table 4.1). The experimentally determined 3D structures of HsSIRT6 were downloaded from the Protein Data Bank (PDB; [rcsb.org](https://www.rcsb.org/)) (Berman et al. 2000). The 3D atomic models of AtSRT1, AtSRT2, HsSIRT4 and HsSIRT6 were downloaded from the AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>) (Jumper et al. 2021). Structure visualization, superposition and analysis were carried out using the Chimera program (Pettersen et al. 2004). Structure comparisons were performed with Chimera and structural searches in the PDB were performed using the Protein structure comparison service PDBe-Fold at the European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm>). Structurally conserved regions were defined as the largest regions comprising residue pairs whose C $\alpha$ -C $\alpha$  distance is  $\leq 2.5$  Å. Two residues are defined to be in contact if they comprise at least one atom at a distance  $\leq 4.0$  Å from one of the atoms of the other residue.

#### 4.4 Results.

AtSRT1 (UniProt ID: Q9FE17) and AtSRT2 (UniProt ID: Q94AQ6) protein sequences comprise 473 and 376 amino acids (a.a.), respectively. Pairwise BLAST alignment of the two sequences encompasses 293 residues with 28% sequence identity and 20% insertions or deletions (Table 4.1), suggesting that they are distantly related to each other, in agreement with (Pandey et al. 2002). The closest human homologs of AtSRT1 and AtSRT2, namely HsSIRT6 (UniProt ID: Q8N6T7) and HsSIRT4 (UniProt ID: Q9Y6E7), respectively, comprise 355 a.a. and 314 a.a., respectively.

*Table 4.1. Results of BLAST pairwise sequence comparisons among AtSRT1, AtSRT2, HsSirt6 and HsSirt4. The E-values associated with each alignment, the percentage of sequence identity and, in parenthesis, the number of identical residues vs. the number of aligned residues are reported*

	<b>AtSRT1</b>	<b>AtSRT2</b>	<b>HsSirt6</b>	<b>HsSirt4</b>
<b>AtSRT1</b>	-	1e <sup>-29</sup>	4e <sup>-95</sup>	7e <sup>-26</sup>
<b>AtSRT2</b>	28% (81/293)	-	8e <sup>-30</sup>	1e <sup>-76</sup>
<b>HsSirt6</b>	51% (141/277)	28% (87/308)	-	2e <sup>-24</sup>
<b>HsSirt4</b>	30% (78/256)	43% (125/293)	28% (84/296)	-

In the ClustalO generated multiple sequence alignment of the four proteins, about 250 residues are aligned (Figure 4.1). Most of the aligned residues are located within the catalytic domain region, and 109 residues (i.e., 43% of these regions) are identical in all four proteins. In addition to the aligned regions, AtSRT2 contains a longer N-terminal sequence, and AtSRT1 and HsSirt6 have longer C-terminal tails, respectively.

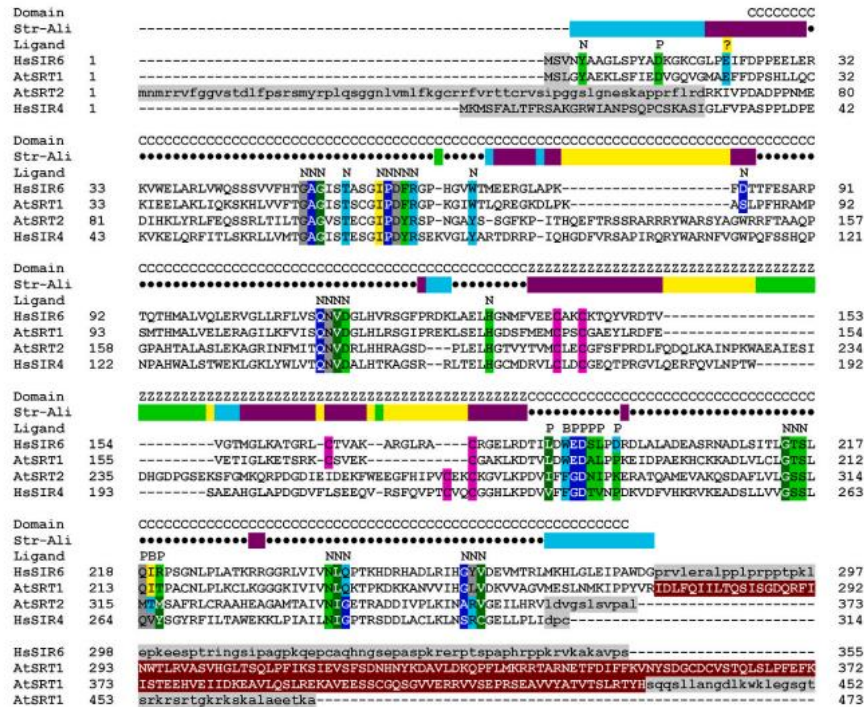
AtSRT1	-----MSLGYAEKLSFIED-----VG	16
HsSirt6	-----MSVNYAAGLSFYAD-----KG	16
AtSRT2	mnmrrvfvgvstldlfpssmyrplqsggnLVMLFKGCRFRVTRTCRVSIPIGGSLGNESKA	60
HsSirt4	-----MKMSFALT-FR-----SAKGRWIANP-SQ	22
	: : :	
AtSRT1	QVGM--AE-FFDPShLLQCKIEELAKLIQKSKHLVFTGAGISTSCGIPDFRGPkGIWT	72
HsSirt6	KCGL---PE-IFDPPEELERKRWELARLVWQSSVVFTGAGISTASGIPDFRGPkGVWT	72
AtSRT2	PPRFLRDRKIVPDADPPNMEDIHKLYRLFQSSRLTILTGAGVSTECGIPDYRSPNGAY-	119
HsSirt4	PCSKASIGLFFVSPPLDPEKVKELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGly	82
	. : * : : . * . : . * * * : * * * * : * . : :	
AtSRT1	LQREGKDLPK-----ASLFFHRAMPSMTHMALVELERAGILKfVIS	113
HsSirt6	MEERGL-APK-----FDTFESARPTQTHMALVQLERVGLLRfLVS	112
AtSRT2	-SSGfKPIThQEFTRSSRRRRYWARsYAGWRRFTAAPGPAHTALASLEKAGRINfMIT	178
HsSirt4	ARTDRRPIQHGDfVRSAPIRQRYWARNFVgWPQfSSHQPNAHWALSTWEKLGKlyWLVt	142
	: * * * * * : * * * * * : : : :	
AtSRT1	QNVdGLHRSgIPREKLSELHGDSFMEMCPSCGAeyLRDFEVETIGLKETSrk-----	166
HsSirt6	QNVdGLHVRSGfPRDKLAELHGnMFVEECAKCKTQYVRDTVVGTMGLKATGRl-----	165
AtSRT2	QNVDRlHHRAGS-DPL-ELHGTVYTMVMECCGfSfPRDLFQDQLKAINPKWAEAIESID	235
HsSirt4	QNVdALHTKAGS-RRLELHGCMdRVLCLDCGEGTPrGVlQERfQVlNPTWAEAHGL-	199
	* * * * * : * * * * * * * * * * : * * * * * :	
AtSRT1	-----CS-----VE-----KCGAKLKDVTLDWEDALPPKEID	193
HsSirt6	-----CT-----VAKARGLRACRGELRDTILDWEDSLPDRDLA	198
AtSRT2	HGDPGSEKsFGMKQRPDGDIEIDeKfWEEGfHPVCEKCKGVlKPDVIFFGDNIPKERAT	295
HsSirt4	-----APDGDVFLSEB-QVRSfQVPTCVCGHGLKPDVVFfGDTVNPDKVD	244
	* . * : : : * :	
AtSRT1	PAEKHCKKADLVLCGLTSLQITPACNLPLKCLKGGGKIVIVNLQKTPKDKKANVVIHGLV	253
HsSirt6	LADEASRNADLSITLGTSLQIRPSGNLPLATKRGGRLVIVNLQPTKHDRHADLRlHGyV	258
AtSRT2	QAMEVAKQSDAFVLGSSLMTMSAFRLCRAAHEAGAMTAIVNIGETRADDIVPLKINARV	355
HsSirt4	FVHKRVKEADSLLVGSSLQVYSGYRfILTAWEKLPfIALLNIGfTRSDDLACLKLNsrc	304
	. : : * : : * * . : . . * : * * * : : :	
AtSRT1	DKVvAGVMESlnmkippyvridlfqiiltqsisgdqrfinwtlrvasvghltsqlpfiks	313
HsSirt6	DEVMTRLMKHlgleipawdgrvrleralpplprpptkplekpeespringsipagpkqe	316
AtSRT2	GEILHRVLDVgslsvpal-----	373
HsSirt4	GELLPLIDPC-----	314
	. : : : :	
AtSRT1	ievsvfSDnhnykdavldkqpfmlkrtrarnetfdiffkvnysdgcDcvstqlslpfefki	362
AtSRT1	steehveiidkeavlgslrekaveesscgqsgvverrvvseprseavvyatvtslrtyhs	422
AtSRT1	qqslLangdlkwlklegsgtsrkrsrtrtgrkrkskalaeeetka-----	473
HsSirt6	pcaqhngsepaspkrerptspaphrppkrvkvakavps-----	355

Figure 2. Multiple sequence alignment among AtSRT1, AtSRT2, HsSirt6 and HsSirt4 protein sequences generated by Clustal O (1.2.4). Identical, highly similar and similar residues according to Clustal O default parameters are indicated by “\*”, “:” and “.” Symbols, respectively.

Several 3D structures of HsSIRT6 fragments, determined by X-ray crystallography, are available from the PDB (data not shown). These fragments encompass 285-298 HsSIRT6 residues within the 1–298 region, out of 355 total HsSIRT6 residues. Conversely, the 3D structures of HsSIRT4, AtSRT1, AtSRT2 and of the 57-residues in the C-terminal HsSIRT6 region have not been experimentally determined. To perform structure analyses, we used the atomic models of all the four proteins built by AlphaFold2 (Figure 4.2), which

has been recently demonstrated to largely outperform all other protein structure prediction methods and produce models of accuracy comparable to known structures in blind tests (Jumper et al. 2021). In the case of HsSIRT6 the AlphaFold2 model is essentially identical to the experimentally determined structures in the common regions.

**A**





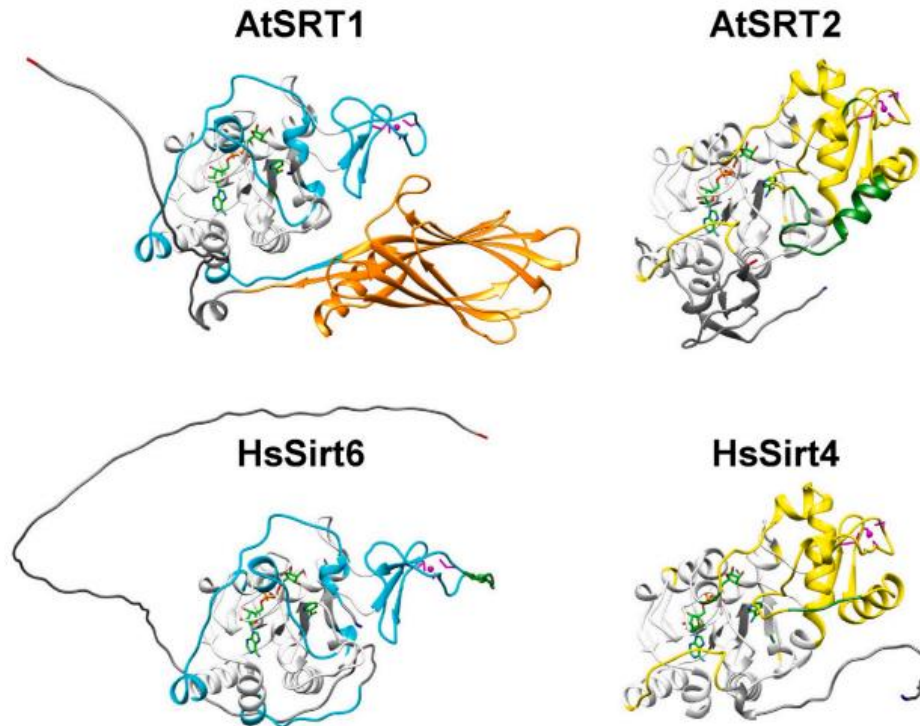
**B**

Figure 4.1. A. Structure-based sequence alignment between AtSRT1, AtSRT2, HsSirt6 and HsSirt4. Amino acids are indicated by one-letter code, and sequence numbering is reported at the beginning and end of each sequence block. HsSirt6 residues are colour-coded based on the type of interaction that they establish with ADP-ribose, NCA and/or peptide substrate in one or more of the experimentally determined HsSirt6 structures, as follows: cyan and blue, polar interactions (i.e., salt-bridges or hydrogen bonds) involving side-chain and main-chain atoms, respectively; yellow, van der Waals interactions between hydrophobic atoms; light and dark green, both hydrophobic and polar interactions, involving side-chain and main-chain atoms, respectively; grey, van der Waals interactions between polar and hydrophobic atoms. HsSirt4, AtSRT1 and AtSRT2 residues are colour-coded like HsSirt6 if they are predicted to have conserved main-chain conformation and comprise chemical groups able to establish polar or van der Waals interactions with ADP-ribose, NCA and/or peptide ligands. Additionally, in all sequences, zinc-binding cysteine residues are coloured magenta; residues belonging to AtSRT1 unique Ig-like domain are dark red; unstructured regions external to the structurally conserved ones, namely, N-terminal regions of AtSRT2 and HsSirt4 and C-terminal regions of AtSRT1 and HsSirt6, are grey. In the “Domain” rows, “C” and “Z” upper-case letters indicate residues belonging to the catalytic and Zn-binding domain, respectively. In the “Str-Ali” (i.e., Structurally Aligned) rows, the “•” symbol and cyan, yellow and purple

*background indicate regions that are structurally conserved among all four proteins, between AtSRT1 and HsSirt6, between AtSRT2 and HsSirt4, both between AtSRT1 and HsSirt6 and between AtSRT2 and HsSirt4 (but not between any member of the first pair and any member of the second pair), respectively; green indicates residues are not structurally conserved but are comprised between conserved regions. In the “Ligand” rows, “N”, “P” and “B” uppercase letters below the horizontal bars indicate residues involved in interactions with ligands deriving from NAD hydrolysis (i.e., ADP-ribose and NCA), peptide substrates or both, respectively. B. Molecular models built for AtSRT1, AtSRT2, HsSirt6 and HsSirt4. The Ca carbon atoms are shown as ribbons and coloured white, cyan, yellow, green and grey according to structural conservation, as in the horizontal bars above the sequence alignment in the top panel; additionally, residues belonging to the AtSRT1 unique Ig-like domain are orange, and the unstructured AtSRT2 and HsSirt4 N-terminal regions and AtSRT1 and HsSirt6 C-terminal regions are grey. N-terminal and C-terminal residues are blue and red, respectively. The side-chains of zinc-binding cysteine residues are shown as sticks and the zinc atoms as a magenta-coloured sphere. The ADP-ribose and NCA moieties deriving from the hydrolysis of the NAD cofactor are shown as stick and coloured by atom-type: C, green; O, red; N, blue; P, orange.*

Structural analysis of these models (Figure 4.1) revealed that the conformation of the catalytic domain, which is involved in both NAD cofactor and peptide substrate binding, is largely conserved in all four proteins. Indeed, 169 out of 226 residues (i.e., 75%, indicated by “●” symbol in Figure 4.2A, assume similar conformations in all four proteins. This is highlighted by the root-mean square deviation (RMSD) values calculated after optimal superposition of Ca atoms of the 169 structurally aligned residues and the percentage of sequence identity in the same regions (Table 4.2). At variance with the catalytic domain, the zinc-binding domain assumes two different conformations in the four proteins: one of the two conformations is shared by AtSRT1 and HsSIRT6, and the other conformation is shared by AtSRT2 and HsSIRT4 (see the “Domain” row in Figure 4.1A).

Table 4.2. Pairwise comparisons of AtSRT1, AtSRT2, HsSirt6 and HsSirt4 3D models and sequences over the 169 residues that are structurally conserved in all four proteins. The RMSD values (in Å) calculated after optimal pair-wise superpositions of Ca atoms, the percentage of sequence identity and number of identical residues, in parentheses, are reported.

	<b>AtSRT1</b>	<b>AtSRT2</b>	<b>HsSirt6</b>	<b>HsSirt4</b>
<b>AtSRT1</b>	-	1.10 Å	0.59 Å	1.18 Å
<b>AtSRT2</b>	39% (65)	-	1.10 Å	0.79 Å
<b>HsSirt6</b>	58% (98)	37% (63)	-	1.22 Å
<b>HsSirt4</b>	36% (61)	47% (80)	36% (60)	-

For this reason, the structurally conserved regions between AtSRT1 and HsSIRT6 (i.e., 269 structurally aligned residues) and between AtSRT2 and HsSIRT4 (i.e., 275 structurally aligned residues) are significantly more extended than the structurally conserved regions among all four proteins (Table 4.3).

Table 4.3 Pairwise comparisons of AtSRT1, AtSRT2, HsSIRT6 and HsSIRT4 3D models and sequences over the longest structurally conserved regions in each pair of proteins. Root-mean square deviation (RMSD) values (in Å), calculated after optimal pairwise superpositions of Ca atoms, are followed by the length of the structurally conserved regions, in parenthesis. Percentages of sequence identity are followed by the number of identical residues vs. the number of structurally aligned residues, in parenthesis.

	<b>AtSRT1</b>	<b>AtSRT2</b>	<b>HsSirt6</b>	<b>HsSirt4</b>
<b>AtSRT1</b>	-	1.12 Å (143)	0.81 Å (269)	1.17 Å (166)
<b>AtSRT2</b>	42% (60/143)	-	1.16 Å (171)	0.92 Å (275)
<b>HsSirt6</b>	51% (138/269)	37% (63/171)	-	1.23 Å (178)
<b>HsSirt4</b>	37% (61/166)	44% (122/275)	34% (60/178)	-

In addition to the catalytic and zinc-binding domains, all four proteins comprise unique regions: i) AtSRT1 contains one additional immunoglobulin (Ig)-like domain (see below) between the catalytic domain and the ~40-residue

C-terminal tail, which is predicted to be unstructured; ii) AtSRT2 comprises a ~65-residue N-terminal region predicted to form a three-strand  $\beta$ -sheet with the 10-residue C-terminal tail; iii) HsSIRT4 has a ~30 residues N-terminal region and HsSIRT6 has a ~75 residues C-terminal region, both of which are predicted to be unstructured. Interestingly, the values reported in Table 4.3 and Table 4.2 indicate that structure and sequence similarity between AtSRT1 and AtSRT2 is higher than that between HsSIRT6 and HsSIRT4, suggesting that the Arabidopsis proteins are more closely related to each other than their human counterparts.

HsSIRT6 residues in contact with ADP-ribose and/or nicotinamide (NCA), which are the molecules deriving from the hydrolysis of the NAD cofactor, and/or with the peptide substrate, comprising either N(6)-acetyl-L-lysine or N(6)-tetradecanoyl-L-lysine, in the experimentally determined HsSIRT6 structures analysed in this work, are highlighted in Fig. 4.1, together with the type of interaction (i.e., polar or non-polar) that they establish with the ligand. In the same Figure, AtSRT1, AtSRT2 and HsSirt4 residues occurring at structurally equivalent positions with respect to those involved in ligand binding in HsSirt6 structures, are coloured based on the type of interaction that they are predicted to establish with the same ligands. A summary of the expected conservation of interactions between each pair of proteins is reported in Table 4.4. In agreement with results of sequence and structure analysis, AtSRT1 has highest conservation of functional residues with HsSirt6, and AtSRT2 with HsSirt4.

*Table 4.4 Expected conservation of interactions with ligand molecules (i.e., ADP-ribose, NCA and peptide substrate) between each pair of proteins analysed in this work. Str-Ali, Ide, Sim,*

*Diff*: Number of residues that are structurally aligned, identical, expected to establish similar interactions and expected to establish different interactions, respectively.

		Cofactor				Peptide Substrate			
		Str-Ali	Ide	Sim	Diff	Str-Ali	Ide	Sim	Diff
AtSRT1	HsSirt6	29	27	2	–	11	8	2	1
AtSRT2	HsSirt4	26	24	2	3	10	3	6	2
AtSRT1	AtSRT2	26	17	6	6	10	1	6	4
HsSirt6	HsSirt4	26	16	7	6	10	2	5	4
AtSRT1	HsSirt4	26	16	7	6	10	3	5	3
AtSRT2	HsSirt6	26	17	6	6	10	1	6	4

As far as regions external to the conserved core of the catalytic domain are concerned, sequence database searches did not provide indications about their possible function. The unstructured C-terminal tails of either AtSRT1 or HsSirt6 (comprising a.a. 433–473 and 279–355, in the respective sequences) have BLAST-detectable sequence similarity only with sirtuin homologs from Brassicaceae and animal species, respectively. The unstructured N-terminal regions of AtSRT2 and HsSIRT4 (comprising a.a. 1–69 and 1–31, respectively) do not match any *A. thaliana* or human sequence with significant E-values (AtSRT2 matches two human proteins below threshold: oestrogen-induced tag 6 and egl nine homolog 2 with 36% sequence identity over 47 residues). Additionally, no experimentally determined 3D structure present in the PDB was detected by the PDBeFold server to have similar conformation to any of these N- or C-terminal sirtuin regions. Conversely, a structural search in the PDB archive performed by the PDBeFold server using the unique all- $\beta$  Ig-like domain (a.a. 274–432) of AtSRT1 as query, revealed that this domain is structurally similar to human GAS41 (RMSD value: 1.84 Å over 107 structurally aligned residues between AtSRT1 molecular model and GAS41 3D structure in PDB entry: 5vna, chain D). GAS41 is a chromatin-associated protein belonging to the YEATS family, which is involved in the recognition of acetyl-lysine in histone proteins, with a preference for H3K18 and H3K27 peptides (Cho et al. 2018). GAS41 has been suggested to be a reader of

diacetylated histones, since it contains a C-terminal coiled-coil domain that is responsible for protein dimerization and binds to diacetylated H3 peptides with higher affinity than monoacetylated peptides.

*In vitro* and *in vivo* experiments, performed by our collaborators, demonstrated that AtSRT1 and AtSRT2 exert an inhibitory role on DNA replication control. Indeed, both *srt1* and *srt2* mutants of *Arabidopsis*, where the SRT1 and SRT2 gene is not present, respectively, showed: i) at a biochemical level, an increase in both DNA synthesis and activity of GDH, the enzyme that catalyses the deamination of glutamate to  $\alpha$ -ketoglutarate ( $\alpha$ KG), which fuels the TCA cycle, thereby providing both energy and metabolites that are necessary for proliferation (Jeong et al., 2013); and ii) at the phenotypic level, an increase in plant weight and root length.

#### 4.5 Conclusions.

In this work, both the *Arabidopsis* sirtuins were demonstrated, for the first time, to play a role in cell proliferation control, which is mediated by the inhibition of DNA duplication rate and GDH activity.

To rationalize this activity of AtSRT1 and AtSRT2 on a molecular basis, we took advantage of their close evolutionary relationships with the two human sirtuins HsSIRT6 and HsSIRT4, whose function and, in the case of HsSIRT6, 3D-structure, have been extensively characterized. Bioinformatics analyses indicated that the catalytic domain regions involved in interactions with NAD cofactor and peptide substrates are highly conserved in the four proteins, both in terms of main-chain conformation and of the chemical-physical properties

of functional residues, which are only slightly less conserved at positions involved in peptide binding with respect to those interacting with NAD. These results indicate that the four proteins are likely able to bind the same, or similar peptides, and catalyse the same de-acylation and ADP-ribosylation reactions demonstrated for HsSIRT6 and HsSIRT4, respectively.

Moreover, we found an evolutionary relationship between the unique Ig-like domain comprised in AtSRT1 and the human GAS41 protein, which suggests that the function of this domain is to facilitate AtSRT1 interaction with specific substrates, possibly with similar features to those recognized by GAS41. Since protein domains that are fused in a single gene in a species are generally interaction and functional partners in species where they are encoded by different genes, the observation that the AtSRT1 sirtuin domain is fused to a GAS41-like domain in Arabidopsis indicates that HsSirt6 and GAS41, although encoded by different genes, may be interaction and functional partners in human as well.

Interestingly, SRT1 and SRT2 have higher 3D structure conservation and sequence identity with each other than SIRT6 and SIRT4, respectively, both within the catalytic domain and over the whole sequences, suggesting that the two Arabidopsis proteins are closer relatives than their human counterparts.

## 5. Identification of the structural determinants underlying the substrate specificity of the Arabidopsis *N*-acetyltransferase activity 2 protein.

### 5.1 Background.

Polyamine acetylation has an important regulatory role in several organisms, since it affects both polyamine function and homeostasis (Pegg 2008; Tavladoraki et al. 2012). Acetylation reduces the number of positive charges on polyamines, thus altering their capacity to interact with several macromolecules. In mammals, acetylated spermine (Spm) and spermidine (Spd) are substrates of peroxisomal polyamine oxidases (PAOs) and are readily secreted into the extracellular space, thus playing a key role in the control of intracellular polyamine levels and of cellular processes related to polyamine content (Pegg 2008). Acetylated polyamines have been also detected in several plant species, among which *Nicotiana plumbaginifolia*, *Helianthus tuberosus* and Arabidopsis (Del Duca et al. 1995; Mesnard et al. 2000; Tassoni et al. 2000; Lie Fliniaux et al. 2004; Hennion et al. 2012; Toumi et al. 2019), but only limited information exists on their physiological roles. However, a role of acetylated polyamines in plant development and responses to environmental and biotic stress has been suggested (Tassoni et al. 2000; Hennion et al. 2006; Adio et al. 2011; Jammes et al. 2014)(*Neuwald & Landsman 1997*). Spermidine/Spermine N<sup>1</sup>-acetyltransferases (SSATs) are GCN5-related N-acetyltransferases that catalyse the transfer of acetyl groups



to the aminopropyl end(s) of Spd and Spm, using acetyl-coenzyme A (acetyl-CoA) as a cofactor. Human SSAT1 (HsSSAT1) is a highly regulated enzyme with broad substrate specificity, being able to acetylate substrates with the general structure  $H_2N(CH_2)_3NHR$ , including  $N^1$ -acetylSpm, sym-norspermine, and sym-norspermidine, but not putrescine (Put),  $N^1$ -acetylSpd, and sym-homospermidine, which have terminal aminobutyl groups (*Della Ragione & Pegg 1983; Hegde et al. 2007; Pegg 2008*). In contrast to HsSSAT1, *E. coli* Spd acetyltransferase (SAT), encoded by the *speG* gene, transfers the acetyl group from acetyl-CoA to either end of Spd (aminopropyl or aminobutyl end) (Fukuchi et al. 1994; Sugiyama et al. 2016). Interestingly, the acetyltransferase encoded by the human gene locus BC011751, annotated as HsSSAT2 (Coleman et al. 2004; Han et al. 2006) because of its sequence similarity with HsSSAT1, does not acetylate Spm or Spd, but rather thialysine, a structural analogue of L-lysine, which is considered to be a source of metabolites with antioxidant properties and can act as an antimetabolite by competing with L-lysine (Jun et al. 2003; Coleman et al. 2004; Proietti et al. 2020).

In *Arabidopsis* genome, two SSAT-like genes have been identified in adjacent positions: AtNATA1 (N-acetyltransferase activity 1; At2g39030) and AtNATA2 (At2g39020). AtNATA1 is highly regulated (Adio et al. 2011; Jammes et al. 2014) whereas AtNATA2 is constitutively expressed (Adio et al. 2011; Lou et al. 2016). The AtNATA1 gene was shown to encode for a protein with N-acetyltransferase activity, though discrepancy exists among published data regarding AtNATA1 substrate specificity. Some data indicate that 1,3-diaminopropane (Dap) is the best substrate, followed by thialysine (Jammes et al. 2014), while other data indicate that ornithine (Orn) (Adio et al. 2011) or Put (Lou et al. 2016) are best substrates. Furthermore, it was

suggested that AtNATA1, together with arginine decarboxylase 1, which was shown to have also N<sup>δ</sup>-acetylOrn decarboxylase activity, provides a pathway for the synthesis of acetylated Put from N<sup>δ</sup>-acetylOrn (Lou et al. 2020). Conversely, no information on AtNATA2 substrate specificity is available so far.

## 5.2 Aim of the work.

The overall aim of this work was to try and identify the physiological function and substrate specificity of AtNATA2.

To this end, our collaborators expressed AtNATA2 in a heterologous bacterial system and determined the catalytic properties of the recombinant protein towards several substrates (including Spm, Nor-Spm, Spd, N-monoacetyl-Put, Orn, Dap, and thialysine).

In parallel, we built molecular models of AtNATA1 and AtNATA2, and performed a comparative analysis of these models and of the experimentally determined structures of the HsSSAT1 and HsSSAT2 human homologues to identify the structural determinants of substrate specificity of each of these enzymes.

## 5.3 Methods.

The sequences of AtNATA1 and AtNATA2 were used to search the NCBI database for homologous proteins of known 3D structure using the BLASTp

algorithm (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)(Altschul et al. 1997). This search retrieved the proteins listed in Table 5.1, the coordinate files of which were downloaded from the Protein Data Bank (PDB; <http://www.rcsb.org/>)(Berman et al. 2000). Based on the E-values calculated by BLASTp, both AtNATA1 and AtNATA2 match most closely *Pseudomonas aeruginosa* probable N-acetyltransferase (E-values  $1e^{-24}$  and  $4e^{-21}$ , respectively), then HsSSAT2 (E-values  $3e^{-17}$  and  $2e^{-15}$ , respectively), followed by mouse SSAT1 (MmSSAT1; E-values  $2e^{-16}$  and  $3e^{-12}$ , respectively) and HsSSAT1 (E-values  $2e^{-15}$  and  $4e^{-11}$ , respectively).

Table 5.1. Proteins of known 3D structure homologous to AtNATA1 and AtNATA2 analysed in this work. Ligand: PDB ID of the bound ligand, namely: ACO, acetyl-co-enzyme A; CoA, coenzyme A; NHQ, *N*<sup>1</sup>-spermine-acetyl-coenzyme A bi-substrate analogue; SPM, spermine.

Protein name	Organism	UniProt ID	PDB ID	Res. (Å)	PDB Ligand ID	Reference
HsSSAT2	<i>Homo sapiens</i>	Q96F10	2BEI	1.84	ACO	Han et al., 2006
HsSSAT1	<i>Homo sapiens</i>	P21673	2JEV	2.30	NHQ	Hegde et al., 2007
MmSSAT1	<i>Mus musculus</i>	P48026	3BJ8	2.3	SPM, CoA	Montemayor et al., 2008
Probable <i>N</i> -acetyltransferase	<i>Pseudomonas aeruginosa</i>	Q9I640	2FE7	2.00	-	Unpublished

Considering that AtNATA1 and AtNATA2 E-values and percentage sequence identity with the bacterial enzyme were only marginally better than those with HsSSAT2, HsSSAT2 was chosen as a template for AtNATA1 and AtNATA2 model building, because the alignments with both Arabidopsis proteins presented a lower number of inserted residues that were grouped in a smaller

number of regions, which is the most commonly observed situation in protein evolution. Additionally, the structure determination of the *P. aeruginosa* protein is not associated with a peer-reviewed publication describing the details of the structure.

The 3D models of AtNATA1 and AtNATA2 were built using Swiss-Model (<https://swissmodel.expasy.org/>) (Guex & Peitsch 1997). The UCSF Chimera package (Pettersen et al. 2004) was used for 3D structure visualization, analysis, and comparisons, including pairwise structure superimpositions. To model AtNATA1 and AtNATA2 complexes with acetyl-CoA and substrates, each 3D model was optimally superimposed to the HsSSAT1 structure, and the co-ordinates of the CoA cofactor covalently bound to Spm were imported into each model. The Spm moiety was computationally modified to obtain Put, Orn and other ligands.

The structure-based multiple sequence alignment (MSA) shown in Fig. 5.1 was obtained as follows:

- ClustalO (<https://www.ebi.ac.uk/Tools/msa/clustalo>) (Sievers et al. 2011) was used to produce an initial MSA comprising AtNATA1, AtNATA2, their homologues of known structure listed in Table 5.1, and additional homologues from other species, namely HsSSAT2-like from *Mus musculus* (MmSSAT2), and SSAT from *L. major* (LmSSAT), *C. elegans* (CeSSAT) and *S. pombe* (SpSSAT);
- Structure superimpositions between each pair of structures present in the initial MSA were obtained as follows: An initial structure alignment was generated automatically by Chimera. A structural superimposition was then performed including only structurally aligned residues the C $\alpha$  atoms of which were at a distance not higher than 2.2 Å. This procedure was

repeated until all and only residue pairs satisfying this distance threshold were superimposed.

- The initial MSA produced by ClustalO was thus manually adjusted based on the structure superposition between HsSSAT2, HsSSAT1 and the 3D models built for AtNATA1 and AtNATA2 (see below) to produce the structure based MSA shown in Fig. 5.1. Of the two monomers present in the functional homodimer, HsSSAT2 chain B (comprising residues 2–59 and 70–170) was used in all comparisons, because it is more complete than chain A (comprising residues 3–30, 35–60 and 68–170). Chain B was also chosen as a reference for AtNATA2 and AtNATA1 models since they were built using HsSSAT2 as a template. Chain A of HsSSAT1 was selected because it provides the best structure superposition with HsSSAT2 chain B, whereas chain C of MmSSAT1 was chosen because it is the only one in complex with Spm. Throughout the text, residues belonging to the monomer chosen as reference are indicated by their one-letter code name and number along the sequence, and residues from the other monomer by the one-letter code, number, and chain name. As an example, in HsSSAT1, E92 is from the reference monomer (chain A) and W154.B from the other one (chain B).

## 5.4 Results.

AtNATA1 and AtNATA2 genes, and their protein products, AtNATA1 and AtNATA2, are closely related to each other. In fact, they have similar gene structure (being both intron-less), 76% nucleotide sequence identity and 79%

amino acid sequence identity. AtNATA1 and AtNATA2 also have 31–32% identical amino acids with the thialysine-specific HsSSAT2 over 152 aligned residues and 27–28% identical amino acids with the Spm/Spd-specific HsSSAT1 over 119 aligned residues (Fig. 5.1). With respect to the human SSATs, AtNATA1 and AtNATA2 have an additional N-terminal region of 22 and 29 amino acids, respectively (Fig. 5.1). Moreover, these N-terminal regions are predicted to comprise a potential PEST motif (AtNATA1 with high probability, and AtNATA2 with low probability) and are found upstream of a methionine residue which is well aligned to the first methionine residue of HsSSAT1 and HsSSAT2, and thus may have a regulatory role at a post-transcriptional level. AtNATA1 and AtNATA2 also present a long insertion (amino acid regions 45–105 and 75–135, respectively) with respect to homologous proteins of known structure (Fig. 5.1), which bears a putative PEST motive. BLASTp searches indicated that these insertions are only present in plant AtNATA-like sequences, and that neither of them is homologous to proteins of known structure.

a

SA	.....●.....	
AtNATA2	maaaappppptaapepnmvaplispi <del>gh</del> PMFSRIRLATPSDVVPIHKLIHQMAVGERLTH	60
AtNATA1	-----mapptaapepntvpetsptgh <del>rm</del> FSRIRLATPTDVVPIHKLIHQMAVGERLTH	53
HsSSAT2	-----mASVRIREAKEGDCGDILRLIRELAEEBEKLS	32
MmSSAT2	-----mastrireaesdcgdimrmirelaefeklsh	32
LaSSAT	-----msaplvtvrraeredtqcmydlimelaiyerape	34
ImSSAT	-----msaagvtvrraeredtqrmydlimelaiyerape	34
SpSSAT	-----mgsvrirsvikedlptvyqfikelaefekced	32
CeSSAT	-----mknfeivtvtpdhaeqliismihelaefekmks	32
HsSSAT1	-----mAKFVIRPATAADCSDILRLIKELAKyeymee	32
MmSSAT1	-----makfkipatasdcsdilrlikelakyeymed	32

C

SA	.....●.....	
AtNATA2	LFSATESGLASTLFTSrfQSF <del>TV</del> FLLEVsrsfpfpatitsspspdf <del>tpffk</del> thnldpid	120
AtNATA1	LFVATESGLASTLFNSrfQAV <del>TV</del> FLLEIspspfpt <del>thd-assp</del> dft <del>pfle</del> thkvd <del>lpie</del>	112
HsSSAT2	QVKISEEALRADGF <del>GDN--PFYHCLVAE</del> lpagk-----	65
MmSSAT2	qvkiseealradgfgen--pffhclvaeiipagge-----	65
LaSSAT	cvvvskaemeeeegfger--plwsafvae <del>lqetd</del> -----	65
ImSSAT	cvvvskaemeeeegfger--plwsafvae <del>lqetn</del> -----	65
SpSSAT	qveatipnlevafgfgf--idevtpvaygvfi-----	60
CeSSAT	svvntaeklrkdienka--vhgfia-----	55
HsSSAT1	QVILTEKDLLEDGFGEH--PFYHCLVAEVPkehwt-----	65
MmSSAT1	qviltekdlqedgfgeh--pfyhclvaevpkehwt-----	65

D

SA	.....●.....	
AtNATA2	dpesynfspdm <del>lndv</del> VVAGFV <del>LF</del> FPNYSFLskPGFYIEDIFVREPYRRKGF <del>GSM</del> LLTAV	180
AtNATA1	dpdre <del>kfl</del> pd <del>klndv</del> VVAGFV <del>LF</del> FPNYP <del>SFL</del> akQGFYIEDIFMREPYRRKGF <del>GK</del> LLTAV	172
HsSSAT2	-----l <del>l</del> gpCVVGYGIYFYI <del>Ys</del> TWK-GRTIYLEDIYVMPEYR <del>QCG</del> ISKIIKKV	113
MmSSAT2	-----sqgslvvyg <del>lyy</del> fiystwt-grnvylediyvmpqyr <del>gqg</del> lgtkiikkv	113
LaSSAT	-----dvkprvigmalyyyrystwr-grmlyledf <del>vvt</del> esyrgigtgkmlferv	113
ImSSAT	-----dvkprvigmalyyyrystwr-grmlyledf <del>ivt</del> eshrgigagkmlferv	113
SpSSAT	-----eendqpagmaiyf <del>lnfst</del> wt <del>svrg</del> iyledlyv <del>rpqf</del> gkgygsyllsyl	109
CeSSAT	-----figeepagmnlfyayst <del>twv</del> -gqylhmedlyirp <del>qfr</del> mg <del>lart</del> lwkk1	103
HsSSAT1	-----peghsIVGFAMYYFTYdPwI-GKLLYLEDFFVMSDYR <del>GFG</del> ISEILKNL	113
MmSSAT1	-----peghsivgfamyyf <del>tydpwi</del> -gkllyledffvmsdyr <del>gfg</del> gseilknl	113

A

SA	.....●.....	
AtNATA2	AKQAVKMGYGRV <del>EWV</del> LDWNVNAIN <del>KF</del> YEQMGAQILQ---EWRVCRLTGDAL <del>EAFDQV</del> ni---	236 (100%)
AtNATA1	AKQAVKLGVRV <del>EW</del> VIDWNVNAIN <del>F</del> YEQMGAQVFK---EWRVCRLTGDAL <del>QAIDK</del> ni---	228 (79%)
HsSSAT2	AEVALDKGCSQFRLAVLD <del>WNR</del> AMDLYKALGAQDLTeaEGWHFFCFQGEATR <del>KL</del> AGK----	170 (31%)
MmSSAT2	aevalnk <del>gcs</del> qf <del>rlav</del> ld <del>wn</del> kkavnlykflgaqdlteseg <del>wsf</del> fegeamrelagr---	170 (31%)
LaSSAT	lqqakeegchgm <del>vqald</del> wn <del>tpain</del> fykkydaeida---g <del>wv</del> nctlkf-----	158 (30%)
ImSSAT	lqqakeegchgm <del>vqald</del> wn <del>tpain</del> fykkydaeidp---g <del>wv</del> nclmkf-----	158 (30%)
SpSSAT	areslrig <del>rldw</del> vld <del>wn</del> grai <del>evy</del> ekag <del>qk</del> vg---g <del>ws</del> m <del>rv</del> tgenlkalad <del>kl</del> pgnf	168 (35%)
CeSSAT	aelardk <del>givr</del> l <del>ewav</del> ld <del>wn</del> knaialydtvdvynltkseg <del>wft</del> fm <del>dga</del> a <del>ink</del> fade----	160 (28%)
HsSSAT1	SQVAMRCRCsSMHFLV <del>aew</del> nepsin <del>fy</del> kr <del>rg</del> ASDLSseEGWRL <del>Fk</del> idkeyllk <del>matee</del> ----	171 (27%)
MmSSAT1	sqvam <del>kr</del> cc <del>ssm</del> hfl <del>v</del> aew <del>n</del> epsin <del>fy</del> kr <del>rg</del> as <del>d</del> ls <del>s</del> ee <del>g</del> w <del>r</del> l <del>f</del> kidkeyllk <del>maae</del> ----	171 (28%)

B

Figure 5.1. Structure-based multiple sequence alignment comprising AtNATA1, AtNATA2 and homologous proteins from different species. Insertions and N-terminal extensions in AtNATA1 and AtNATA2 are underlined. “■”, “□” and “●” symbols indicate the 115, 37 and 4 residues, respectively, which are structurally conserved in: i) AtNATA1 and AtNATA2 models, and HsSSAT2 and HsSSAT1 structures (HsSSAT2 residues: 3–26, 33–48, 50–59, 70–81, 83, 85, 87, 89–122, 124–129, 145–149, 153–157); ii) AtNATA1 and AtNATA2 models, and HsSSAT2 structure, but not HsSSAT1 (HsSSAT2 residues: 27–32, 82, 84, 88, 123, 130–144, 158–169); iii) HsSSAT2 and HsSSAT1 structures, but not AtNATA1 and AtNATA2 models (HsSSAT2 residues: 2, 49, 86, 152), respectively. Residues that are structurally aligned to each other are uppercase. Residues not observed in the 3D structure of HsSSAT2 chain B and residues that are not present in AtNATA1 or AtNATA2 models are shown in italic. Conserved structural domains in GCN5-related N-acetyl transferases (Neuwald and Landsman, 1997) are indicated by labelled horizontal bars (A, B, C, D). Positions involved in acetyl-CoA or polyamine binding are labelled with letters ‘a’ or ‘p’, respectively. Residues the identity of which is conserved with respect to AtNATA2, are black-shaded if they are found at positions labelled with ‘a’ or ‘p’ and grey-shaded otherwise. Numbers in parentheses indicate the percentage of amino acid sequence identity to AtNATA2. At, *Arabidopsis thaliana*; Hs, *Homo sapiens*; Mm, *Mus musculus*; Lm, *Leishmania major*; Sp, *Schizosaccharomyces pombe*. Ce, *Caenorhabditis elegans*.

To elucidate the structural determinants of substrate specificity, 3D models of AtNATA1 and AtNATA2 homodimers have been obtained (Figure 5.2). Inspection of these models and of structure-based sequence alignments (Figure 5.1) shows that the long insertions are located on the external sides of both proteins and are, therefore, not expected to affect the rest of the fold (Figure 5.2).



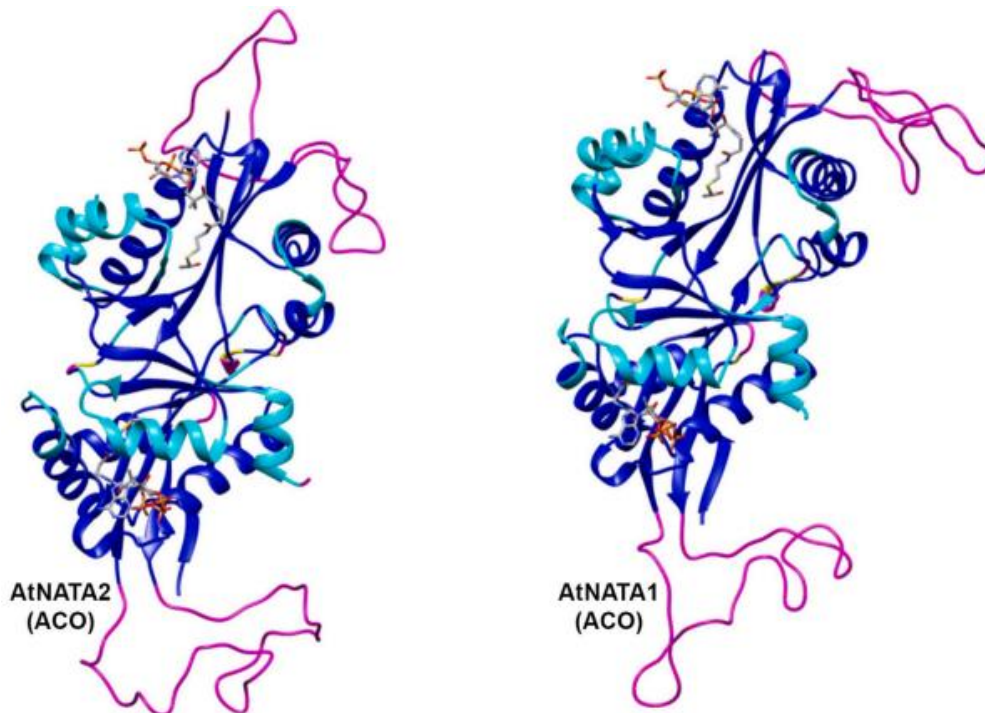


Figure 5.2. 3D atomic models of AtNATA1 and AtNATA2 homodimers in complex with acetyl-CoA. In the models, main-chain atoms are represented as ribbon and colour-coded as follows: Blue, residues that are structurally conserved with respect to both HsSSAT2 and HsSSAT1 structures (indicated by “■” in Fig. 5.1); Cyan, residues that are structurally conserved with respect to HsSSAT2 structure, but not HsSSAT1 (indicated by “□” in Fig. 5.1); Yellow, residues that are structurally conserved between HsSSAT2 and HsSSAT1 structures, but not in AtNATA1 and AtNATA2 models (indicated by “●” in Fig. 5.1); Magenta, residues that are not structurally conserved in either of the aforementioned groups of structures. Acetyl-CoA (ACO) is shown as sticks and coloured by atom type (C, green; N, blue; O, red; P, orange).

The core of the fold is predicted to be conserved with respect to HsSSAT2, to which both AtNATA1 and AtNATA2 monomers are aligned for 152 out of 170 total HsSSAT2 residues. Indeed, residues involved in acetyl-CoA binding are structurally conserved in AtNATA1 and AtNATA2 with respect to HsSSAT2, most of them being either identical or substituted by residues with similar physical-chemical properties (Figure 5.1). However, differences in

residue size are observed at a few positions, namely HsSSAT2 residues G102, Q103, I105, L127 and L139, which in AtNATA1 and AtNATA2 are replaced by the bulkier R, K, F, W and F, respectively. With respect to HsSSAT1, AtNATA1 and AtNATA2 conserve the GCN5-related N-acetyltransferase fold in the A and C regions, but they are expected to diverge from HsSSAT1 in the B and D regions. Indeed, among HsSSAT1 residues involved in acetyl-CoA binding (Lu et al. 1996; Bewley et al. 2006; Zhu et al. 2006), residues Y100, R101, G104 and G106 are conserved in AtNATA1 and AtNATA2, and residues Y27 and L91 are conservatively substituted by F and I residues, respectively (Figure 5.1). On the other hand, the short C terminal MATEE sequence, which was shown to be important in HsSSAT1 rapid *in vitro* degradation and for the stabilizing effect of polyamine analogues on HsSSAT1 (Coleman & Pegg 1997), is not present in the two Arabidopsis proteins or in HsSSAT2 (Figure 5.1).

Comparative analysis of the ligand binding pockets of AtNATA2 and AtNATA1 with those of HsSSAT2, HsSSAT1 and MmSSAT1 revealed that they present significant differences in shape, dimension, and chemical-physical nature of surrounding residues (Figure 5.1, 5.3 and 5.4). These differences result from the fact that only a small fraction of HsSSAT1 acidic, hydrophobic, and aromatic residues lining the Spm binding site are conserved in terms of both main-chain position and identity in AtNATA2, AtNATA1 and HsSSAT2 (Figure 5.1; Figure 5.3). These are: i) D93 and E92, which are within salt-bridge distance from the first (N1) and second (N2) Spm amino groups and participate in HsSSAT1 catalytic mechanism (Hegde et al. 2007; Bewley et al. 2006); and ii) W154 from the monomer not taken as reference (W154.B; see Methods), which makes hydrophobic contacts with the methylene groups

between the second (N2) and third (N3) Spm amino group. Furthermore, L128 residue of HsSSAT1 catalytic site is conserved in main chain position, but its side chain is replaced by different aliphatic side chains, such as those of isoleucine in AtNATA1 (I187), valine in AtNATA2 (V195) and alanine in HsSSAT2 (A128).

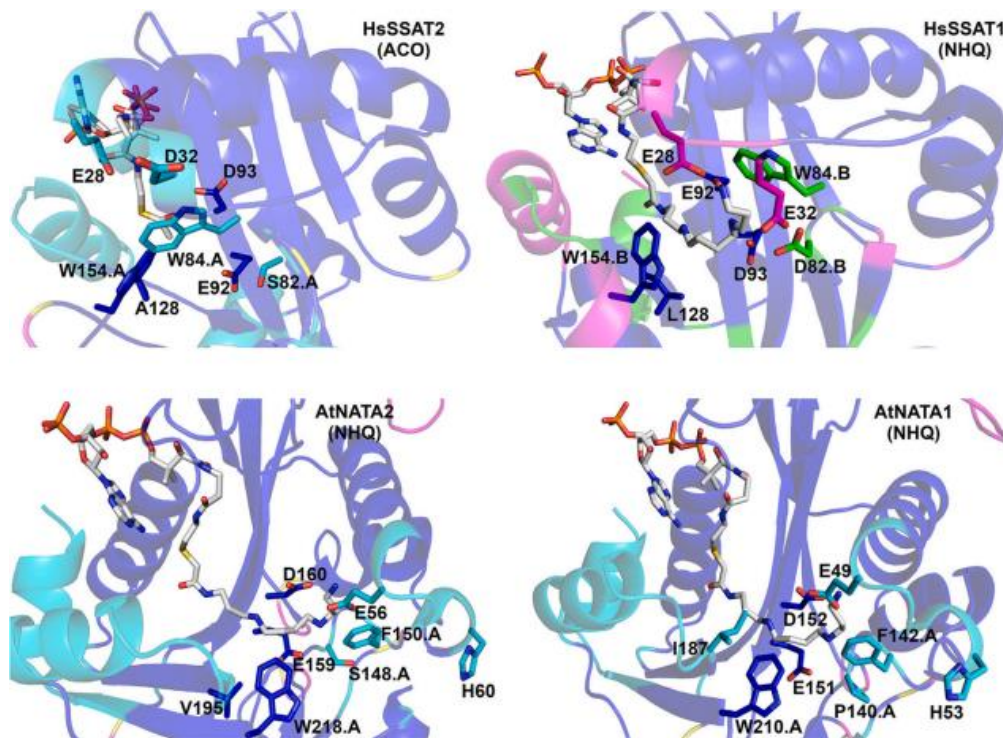


Figure 5.3. Comparison among substrate binding-site residues of human SSAT structures and Arabidopsis NATA models. Ribbon representation of the 3D structures of HsSSAT2 and HsSSAT1 homodimers, which have been experimentally determined by Xray crystallography (Han et al. 2006; Hegde et al. 2007), and of the atomic models of AtNATA2 and AtNATA1 homodimers, built in this work. Colour-coding is as follows: Blue, regions that are structurally conserved among HsSSAT2, HsSSAT1, AtNATA2 and AtNATA1 (indicated by “■” in Fig. 5.1); Cyan, regions that are structurally conserved among HsSSAT2, AtNATA2 and AtNATA1, but not HsSSAT1 (indicated by “□” in Fig. 5.1); Green, regions that are structurally conserved among HsSSAT1, AtNATA2 and AtNATA1, but not HsSSAT2; Magenta, other regions. Protein ligands, i.e., acetyl-CoA (ACO; HsSSAT2) and N1-Spm-acetyl-coenzyme A bi-substrate

*analogue (NHQ) comprising covalently linked acetyl-CoA and Spm (HsSSAT1, AtNATA2 and AtNATA2) are shown as sticks and coloured by atom type: N, O, C, S and P atoms are blue, red, white, yellow, and orange, respectively. Residues interacting with Spm in HsSSAT1, and structurally equivalent residues in HsSSAT2, AtNATA2 and AtNATA1, are shown as sticks and coloured by atom type: N, O and S atoms are blue, red and yellow, respectively; C atoms are blue, cyan, green or magenta depending on the colouring of the ribbon.*

All of the other HsSSAT1 residues involved in interactions with Spm present significant differences in HsSSAT2 and the Arabidopsis homologues: i) HsSSAT1 E28, which interacts with the third (N3) and fourth (N4) Spm amino groups, is conserved in type in the other three enzymes (E28 in HsSSAT2, E49 in AtNATA1 and E56 in AtNATA2), but its main-chain and side-chain are placed in a different position with respect to the other ligand-binding site residues; ii) E32, which interacts with Spm N4, is replaced by histidine in AtNATA2 (H60) and AtNATA1 (H53), and conservatively substituted by aspartate in HsSSAT2 (D32); iii) D2.B, which is salt-bridged to Spm N3, is replaced by the hydrophobic proline in AtNATA1 (P140.A) and by the polar serine in AtNATA2 (S148.A) and HsSSAT2 (S82.A); iv) W84.B, which packs with the Spm methylene groups between N3 and N4, is conserved in HsSSAT2 (W84) and replaced by phenylalanine in AtNATA2 (F150.A) and AtNATA1 (F142.A). As shown in Fig. 5.4, W84.B of HsSSAT1 packs with the methylene groups of Spm comprised between N3 and N4, whereas in HsSSAT2 structure the loop comprising W84.B assumes a different conformation that places the side-chain of HsSSAT2 in a position that would clash with Spm distal region. In the 3D models of AtNATA2 and AtNATA1, the homologous loops are predicted to assume a conformation similar to that of HsSSAT2, and the side chains of F150.A and F142.A are predicted to assume a position similar to that

of W84.A in HsSSAT2 and, therefore, overlap with the Spm region comprised between N3 and N4 as well (Fig. 5.4).

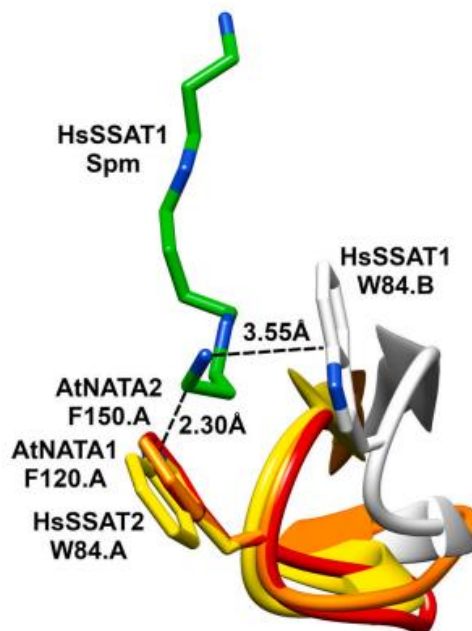


Figure 5.4. Comparison among the substrate binding-site pockets of human SSAT structures and Arabidopsis NATA models. The loop regions comprising residue W84.B of HsSSAT1 and HsSSAT2, F150.A of AtNATA2 and F142.A of AtNATA1 are shown as ribbon and coloured white, yellow, orange and red, respectively. The side chains of these residues and the Spm ligand of HsSSAT1 are shown as sticks and coloured by atom type (N, blue; O, red; C, white, yellow, orange, red and green in HsSSAT1, HsSSAT2, AtNATA2, AtNATA1 and Spm, respectively).

Our collaborators expressed recombinant AtNATA2 in *E. coli*, purified it by affinity chromatography to electrophoretic homogeneity, and analysed its catalytic properties by an *in vitro* assay based on the quantification of CoA-SH released during the reaction (Bode et al. 1993; Coleman et al. 2004; Lin et al. 2010; Jammes et al. 2014). The results of these analyses indicate that

recombinant AtNATA2 is significantly more active towards Dap and thialysine than other substrates (Figure 5.5 and Table 5.2).

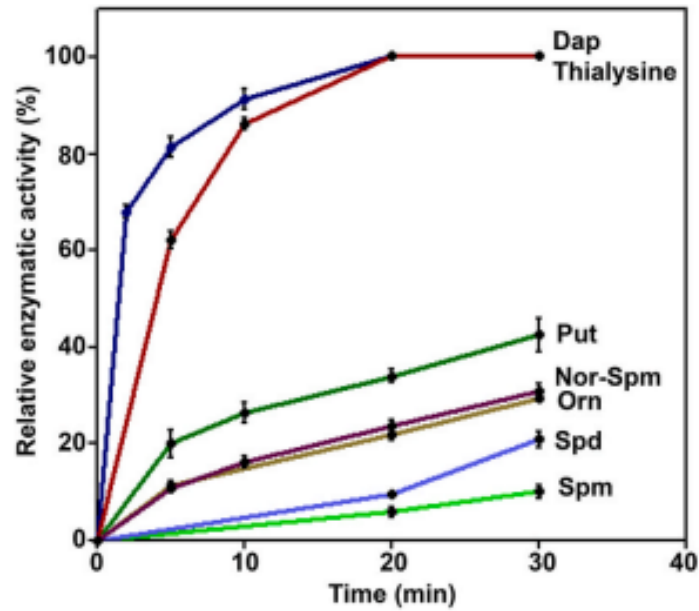


Figure 5.5. Time course of acetyltransferase activity of recombinant AtNATA2 *in vitro* with different acetyl acceptors. Catalytic activity is reported as percentage relative to the maximum activity. Each point represents the mean value from three independent experiments. Bars indicate standard error (SE). Dap: 1,3-diaminopropane; Nor-Spm: norspermine; Orn: ornithine; Put: putrescine; Spd: spermidine; Spm: spermine.

Table 5.2 Catalytic parameters of recombinant AtNATA2 with the indicated substrates. Values indicate means of three replicates  $\pm$  standard deviation (SD). Acetyl-CoA: acetyl-coenzyme A; Dap: 1,3-diaminopropane; Nor-Spm: norspermine; Orn: ornithine; Put: putrescine; Spd: spermidine; Spm: spermine.

	$k_{\text{cat}}$ ( $\text{s}^{-1}$ )	$K_{\text{m}}$ (mM)	$k_{\text{cat}}/K_{\text{m}}$ ( $\text{mM}^{-1}\text{s}^{-1}$ )
Dap	$1.02 \pm 0.20$	$0.77 \pm 0.18$	1.33
Thialysine	$0.78 \pm 0.20$	$0.90 \pm 0.36$	0.87
Put	$0.07 \pm 0.023$	$1.29 \pm 0.37$	0.05
Orn	$0.02 \pm 0.005$	$0.29 \pm 0.07$	0.07
Nor-Spm	$0.03 \pm 0.010$	$2.73 \pm 0.83$	0.011
Spd	$0.02 \pm 0.004$	$1.70 \pm 0.80$	0.012
Spm	$0.01 \pm 0.006$	n.d.*	n.d.
N-monoacetyl-Put	$0.01 \pm 0.005$	n.d.	n.d.
Acetyl-CoA	n.d.	$0.42 \pm 0.06$	n.d.

\*n.d. = not determined.

AtNATA2 catalytic activity towards Dap was further shown to be higher at 30°C than 23°C, and at basic than acidic pH, reaching a maximum at pH 8.5 (Figure 5.6). Furthermore, AtNATA2 was shown to be catalytically active towards N-monoacetyl-Put, though with a 7-fold lower  $k_{\text{cat}}$  value than towards Put (Table 5.2). In addition, stoichiometric analysis evidenced a molar ratio of around 2 between the released CoA-SH and consumed Dap or Put, indicating that Dap and Put are di-acetylated during the AtNATA2 catalysed reaction.

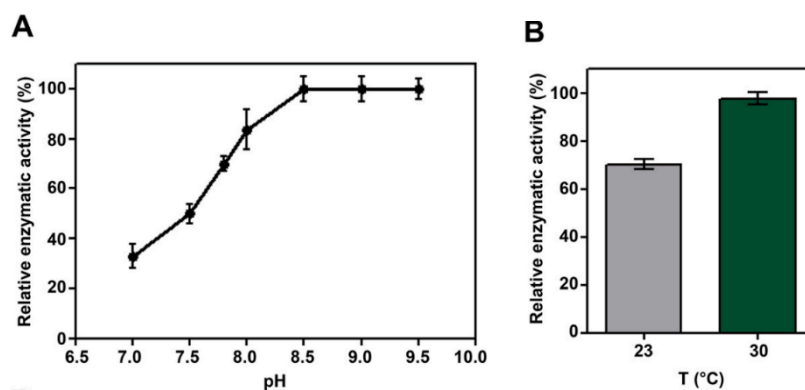


Figure 5.6. Catalytic activity of recombinant AtNATA2 as a function of pH (A) and temperature (B). Catalytic activity was measured using Dap as a substrate. Data are expressed as percent of maximum activity. Each point represents the mean value from three independent experiments and bars indicate SE.

Our structure analyses indicate that the preference of AtNATA2, AtNATA1 and HsSSAT2 for short substrates, comprising only two amino groups (such as Dap, Put and Orn), over substrates comprising three or four amino groups, is likely contributed by the fact that although HsSSAT1 residues interacting with Spm regions proximal to the acetyl-CoA cofactor (from N1 to N2) are mostly conserved in both Arabidopsis protein and HsSSAT2, residues interacting with Spm distal regions (from N3 to N4) assume different conformation (like E28 and W84.B) and/or are replaced by residues differing in size (like W84.B), chemical nature (e. g., E32 replacement by histidine in AtNATA2 and AtNATA1) or both (D82.B replacement by serine and proline). Based on these observations, the lack of activity of HsSSAT1 towards short diamines, like Put and Dap, may be ascribed to the interaction of HsSSAT1 acidic residues, such as E28, E32 and D82.B, which are in contact with Spm N3 or N4 in the crystal structure, with one of the external amino groups of the diamines. This would result in the positioning of the other amino group in a location too distant from the acetyl-CoA cofactor for a reaction to occur.

## 5.5 Conclusions.

In the present study, the activity of recombinant AtNATA2 were studied by *in vitro* measurements of catalytic activity, and bioinformatics analyses of both



AtNATA2 and its AtNATA1, HsSSAT1 and HsSSAT2 homologues were performed to provide a structural explanation for the observed catalytic behaviour.

The best polyamine substrate of AtNATA2 resulted to be Dap, whereas the catalytic efficiency towards Put, Spd and Spm and Orn is low. These data are consistent with data previously obtained for the closely related AtNATA1, which was reported to have high catalytic activity with Dap, and low with Spd and Spm (Jammes et al. 2014). However, differently from AtNATA1, which has 2.5-fold lower activity with thialysine (a metabolite which has not been detected in plants yet) than with Dap, recombinant AtNATA2 has high catalytic activity with thialysine as well (65% of the catalytic activity towards Dap). This difference between AtNATA1 and AtNATA2 suggests that small differences in the amino acid sequence of SSATs can influence substrate specificity. This was also observed for HsSSAT1 and HsSSAT2 that, despite high sequence identity (46%), exhibit different substrate specificity, HsSSAT1 being Spm- and Spd-specific, and HsSSAT2 thialysine-specific (Coleman et al. 2004). Data presented herein appear to be in contrast with previous data showing that recombinant AtNATA1 is mainly active with Put and Orn, the acetylated forms of which have been associated with defense responses (Adio et al. 2011; Lou et al. 2016). These contradictory data regarding recombinant AtNATA1 and AtNATA2 proteins may be reconciled by the possibility that the native enzymes acetylate all three metabolites under conditions in which intracellular Put or Orn levels exceed Dap levels, as was shown for Arabidopsis plants (Sánchez-López et al. 2009). Previous structural and biochemical studies suggested that HsSSAT1 and MmSSAT1 catalyse the acetyl transfer reaction through an acid/base-based catalytic mechanism, during which a

catalytic base deprotonates the N1-amine of the polyamine, and a catalytic acid protonates the sulphur of acetyl-CoA (*Hegde et al. 2007; Montemayor & Hoffman 2008*). Indeed, Y140 and E92 residues are positioned well inside the catalytic site to fulfil the role of a catalytic acid and base, respectively (*Hegde et al. 2007; Montemayor & Hoffman 2008*). The present structural studies showed that these amino acids are conserved in AtNATA1 and AtNATA2, suggesting that the general catalytic mechanism is conserved in these plant enzymes. The structural analyses performed in this work also suggested that the preference of AtNATA1, AtNATA2 and HsSSAT2 for short amine substrates can be ascribed to the different shape and chemical-physical nature of their ligand binding pockets with respect to those of HsSSAT1 and MmSSAT1, in particular in the regions interacting with Spm distal portions with respect to the acetyl-CoA cofactor. Overall, this study gives an insight into polyamine acetylation in plants and contributes to understanding polyamine metabolism and physiological roles. Moreover, it provides a comparison among the structural determinants of substrate specificity of different plant and human SSAT homologues.

## 6. Structural alignment programmes evaluation and creation of a pipeline to compare them.

### 6.1 Background.

Protein folding is a hierarchical process constituted by four subsequential stages that start with the amino acid sequence of the protein (primary structure). The folding process starting from the amino acid sequence organises the protein in ever more complex structures. The first level of rearrangement is constituted by motifs such as secondary structures (i.e.,  $\alpha$  helices and  $\beta$  sheets) and loops, which are combined into increasingly complex arrangements in the tertiary (3D) structure. This type of organisation in general represents the minimum functional unit of a protein, but often is associated with either identical or different 3D structures to create the final level of protein organisation (quaternary structure).

The three-dimensional protein structures are determined with different types of experiments such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and electron microscopy. All the experimentally determined structures of biological macromolecules are stored in the Protein Data Bank (PDB: <https://www.rcsb.org/>) (Berman et al. 2000), which is the central repository of protein structures. Typically, there are two ways to compare proteins, i.e., based on protein sequences or protein structures. The 3D structures of homologous proteins are more conserved than their respective sequences, which may diverge to the extent that sequence comparison methods fail to recognize their evolutionary relationships. For this reason, structure

alignments provide more accurate comparisons of proteins than sequence alignments.

As a result, many programs have been developed to compare 3D structures that differ in multiple aspects, including: sets of atoms that are used in the comparison (i.e., alpha carbons, backbone atoms, all atoms), type of scoring (RMSD, GDT, etc.), working environment (server or local), and output types (score, structure-based sequence alignment, superposed coordinates, etc.).

Despite their usefulness, current protein structure alignment programs (PSAPs) have several limitations. First of all, most of them provide reliable structural alignments in the case of highly conserved structures, but commit errors with divergent structures, the entity of the error increasing with the extent of structural divergence. Additionally, it is not generally known which programs provide the best results for specific protein families or in general.

## 6.2 Aim of the work.

The long-term aim of this work is to perform an assessment of the ability of currently available PSAPs to produce correct protein structural alignments (PSA). As first step to reach this goal, we performed the following tasks:

- I) We performed an extensive search of the available PSAPs and catalogued them.
- II) We identified a dataset of protein pairs (PPs) such that: i) The structural similarity between the two proteins of each PP is high enough to make it possible to structurally align them using a PSAP. ii) The coordinate files of the protein structures among all different PPs are diverse enough (in terms of

structure divergence, structure completeness, presence of alternate residue conformations and other peculiarities that might be present in PDB files) to make a pipeline able to process them all reasonably robust.

III) We developed a pipeline that compares the pairwise PSAs provided by the different PSAPs, and identifies consensus regions, i.e., regions of the two input proteins that are aligned in the same way by the different PSAPs.

The following step will be:

IV) The implementation in the pipeline of the PSAs produced by PSAPs that are available only as servers.

V) The search in the literature for highly reliable PSAs and the production of novel manually curated PSAs starting from the consensus regions identified by our procedure, to obtain a set PSAs representative of different protein folds.

VI) The evaluation of the accuracy of each PSAP by comparing the PSAs produced by them with manually curated PSAs.

### 6.3 Methods.

An extensive research for the available PSAPs was performed using both commonly used search-engines, with either general (such as Google <https://www.google.com>) or research specific (e.g., PubMed <https://pubmed.ncbi.nlm.nih.gov>; Google scholar <https://scholar.google.com>) scope, and information aggregators such as Wikipedia (<https://en.wikipedia.org>).

The dataset of PPs chosen to test the method was the ensemble of Target-Template pairs investigated in the 14<sup>th</sup> edition of the Critical Assessment of

Structure Prediction (CASP) website (<https://predictioncenter.org/casp14>). CASP is an international experiment to assess the accuracy of protein structure prediction methods, which is run every two years since 1994 (Moult et al. 1995). The proteins of unknown structure that have to be predicted are called Targets, whereas proteins present in the PDB that have high structure similarity with the Targets and often used a starting point to build the models are called Templates. This dataset was chosen to test our method because it comprises a relatively high number of PPs such that: i) the two proteins of each PP have detectable structure similarity with each other; and ii) the proteins in all PPs have variable degree of structure similarity. As far as point i) is concerned, in CASP the extent of structure similarity and divergence of each PPs is qualitatively assigned. The “TBM-Easy” and “TBM-Hard” categories indicate Targets-Template PPs that have high and low structure similarity with each other, respectively; conversely, the “FM” category indicates Targets that have little to no significant similarity with other structures present in the PDB. Additionally, the chosen dataset can be easily extended to incorporate the sets of Target-Template PPs investigated in previous or subsequent CASP editions. All the scripts incorporated in the pipeline to compare the PSAs coming from the collected PSAPs and identify consensus regions were written in python, version 3 (Van Rossum & Drake 2009).

## 6.4 Results.

By utilising the search-engines cited in the Methods section, 115 programmes performing protein structural alignments (PSA) were identified (Appendix B).

However, only 80 of these are currently available. These 80 software were divided into two categories: server-based only or downloadable. As a first step, we focussed on this second category, which comprises 59 PSASs. These programs were downloaded and installation was attempted. This step was successful for 30 programmes only, due to problems such as missing libraries or incompatibility of old versions of programming languages with current operating systems. The 30 PSAPs that could be installed software were tested to investigate the type of output. This consisted in: i) overall similarity measures, expressed most commonly by RMSD or GDT values; ii) structure-based sequence alignments; iii) coordinates of the superimposed structures or rotation-translation matrices from which the coordinates of the superimposed structures could be obtained; iv) any combination of the above. For our comparison purposes, we selected all the methods whose output was either directly a structure-based sequence alignment (SBSA) or was such that a SBSA could be derived from it (e.g., coordinates of the superimposed structures). The 15 programmes that satisfy this last requirement are listed in Table 6.1.

*Table 6.1. List of PSAPs that provide an output from which a SBSA can be obtained, selected to setup the pipeline.*

<b>NAME</b>	<b>Last Update</b>	<b>Language</b>	<b>Author</b>
CAB-Align	20-03-2015	C++	(Terashi & Takeda-Shitaka 2015)
DeepAlign	15-08-2018	C++	(Wang et al. 2011)
Fr-TM-align	11-01-2009	Fortran	(Pandit & Skolnick 2008)
Kpax	27-04-2019	NA	(Ritchie et al. 2012)
LGA	02-09-2019	C	(Zemla 2003)
LOVOALIGN	16-11-2018	Fortran	(Martínez et al. 2007)
MAMMOTH	14-08-2006	Fortran	(Ortiz et al. 2009)
Mapsci	14-08-2009	C++	(Ilinkin et al. 2010)

Matt	12-02-2208	C	(Menke et al. 2008)
mTM-align	16-09-2018	C++	(Dong et al. 2018)
Mustang	04-01-2017	C++	(Konagurthu et al. 2006)
parMATT	14-04-2019	C	(Shegay et al. 2019)
SHEBA	09-05-2007	C	(Jung & Lee 2000)
SPalign	18-07-2012	C++	(Yang et al. 2012)
TM-align	22-08-2019	NA	(Zhang & Skolnick 2005)

The pipeline used to compare the output of all the programmes was written in python and comprises different scripts, whose tasks are schematized in the flowchart in Figure 6.1.

- 5) The first script, “PSAP-launcher”, runs each of the 15 selected PSAPs on a given pair of input structures and collects their output, i.e., 15 different PSAs, written in a different format for each PSAP.
- ii) The second script, “PSA-parser” translates the specific output of each PSAP into a new format that we called Fasta2Excel. This is a variant of the Fasta format where each amino acid-encoding letter is separated from the other amino acid-encoding letters by a space and, therefore, ready to be imported in the Microsoft Excel software for visual examination (Figure 6.1).



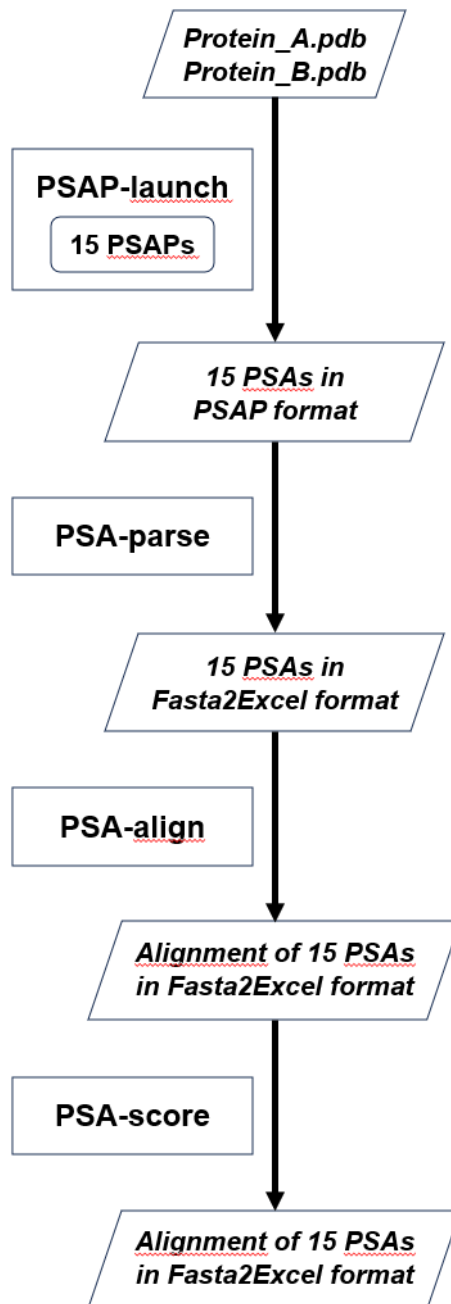


Figure 6.1. An example of the schematisation of the process produced by the scripts.

An example of output of this script after uploading to Excel is shown in Figure 6.2. In spite of the fact that the two structures given as input to the different PSAPs, i.e., PDB files 6FRH and 6TMM, are quite similar to each other (the 6FRH target was assigned to the TBM-easy CASP category) it can be observed that the PSAs produced by the different PSAPs differ in several regions.

CAB	>6frh	-	-	-	M	L				V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K										
CAB	>6tmm	L	I	L	N	L				V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I										
Deep	>6frh	M	L	R	E	-				L	T	V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K								
Deep	>6tmm	-	-	-	-	L				I	Q	V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	M	-	Q	V	S	I								
ErTMalign	>6frh	M	-	-	L	R				P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K											
ErTMalign	>6tmm	-	L	I	L	N				E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I											
Kpax	>6frh	-	-	-	M	L				T	V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K									
Kpax	>6tmm	L	I	L	-	-				Q	V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I									
LGA	>6frh	M	L	R	E	S				G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K												
LGA	>6tmm	L	I	L	N	L				N	N	H	N	F	F	A	N	G	V	L	A	H	A	M	-	Q	V	S	I												
Loyalign	>6frh	-	A	I	S	G		■	■	N	D	I	I	V	H	A	S	I	E	Q	D	K																			
Loyalign	>6tmm	G	A	F	V	S		■	■	N	G	V	L	A	H	A	-	M	Q	V	S	I																			
Matt	>6frh	M	L	R	-	-				E	E	V	F	D	L	T	V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K	L	G	G
Matt	>6tmm	-	-	-	L	I				V	F	I	Y	H	I	Q	V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I	-	-	-
Mammoth	>6frh	-	-	-	M	L	R			I	I	V	H	A	S	I	E	Q	D																						
Mammoth	>6tmm	L	I	L	N	L				V	L	A	H	A	M	Q	V	S	-																						
Mapsci	>6frh	M	L	R	E	-				G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K												
Mapsci	>6tmm	L	I	L	N	L				N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I												
MtAlign	>6frh	M	-	-	L	R		■	■	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K											
MtAlign	>6tmm	-	L	I	L	N		■	■	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I											
Mustang	>6frh	M	L	R	-	-				P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K	-										
Mustang	>6tmm	L	I	L	N	L				E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	M	Q	V	S	-	-	I										
parMatt	>6frh	M	L	R	-	-				E	E	V	F	D	L	T	V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K	L	G	G
parMatt	>6tmm	-	-	-	L	I				V	F	I	Y	H	I	Q	V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I	-	-	-
Sheba	>6frh	M	L	R	E	S				V	P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K										
Sheba	>6tmm	L	I	L	N	L				V	E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I										
SPalign	>6frh	M	L	R	E	S				L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K													
SPalign	>6tmm	L	I	L	N	L				N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I													
tmAlign	>6frh	M	-	-	L	R				P	G	L	R	N	F	V	A	N	D	I	I	V	H	A	S	I	E	Q	D	K											
tmAlign	>6tmm	-	L	I	L	N				E	N	N	H	N	F	F	A	N	G	V	L	A	H	A	-	M	Q	V	S	I											

Figure 6.2. Output of the “PSAP-parser” script run on the Target-Template pair 6FRH-6TMM imported in Excel.

iii) The third script, “PSA-align”, aligns the 15 PSAs produced by the 15 PSAPs to each other, to facilitate their comparison. To illustrate how the script works, we will refer to the two proteins present in each PSA as “Protein\_A” and “Protein\_B”. To obtain an alignment of the different PSAs, the script aligns the Protein\_A sequences of all PSAs to each other by inserting gaps at appropriate positions both in the Protein\_A sequences and in the Protein\_B sequences, in such a way that Protein\_A-Protein\_B alignment present in each PSA is maintained. Operatively, the script works as follows. First, it selects as

a reference the PSA produces by LGA (Zemla 2003), which is the PSAP used by CASP14 organizers. Then it compares the Protein\_A sequences of each of the 14 PSAs produced by the other PSAPs (Protein\_A-PSAP) to the Target sequence of the PSA produced by LGA (Protein\_A-LGA). At positions where Protein\_A-LGA and the Protein\_A-PSAP differ, a gap is introduced either in the protein\_A-LGA or in the Protein\_A-PSAP sequence, until these sequences are aligned. Importantly, every time a gap is inserted into any Protein\_A sequence, a gap is also inserted at the corresponding position into the Protein\_B sequence of the same PSA, so that the alignment between Protein\_A and Protein\_B sequence of each PSA is preserved.

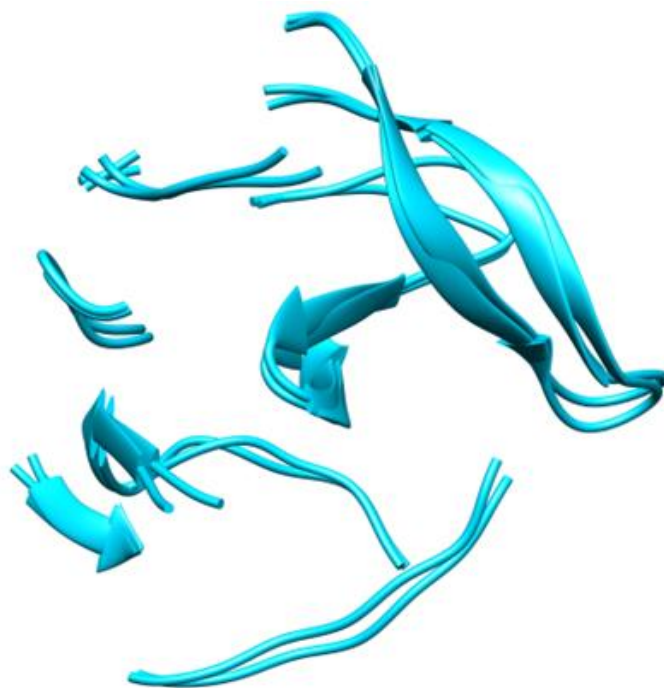
iv) The fourth script, “PSA-entropy”, calculates the variability at each position of the PSA alignment produced by the previous script by using Shannon’s entropy. This script assigns a value of “0” to all positions where the PSAs produced by the 15 PSAPs are identical and positive values to all positions where the PSAs produced by the 15 PSAPs differ, which are proportional to the extent of variation, until 3.67 which is the value if all the 15 PSAPs differ from each other.

The final output of the pipeline incorporating these scripts is a file in the Fasta2Excel format that comprises: all the 15 PSAs aligned to one another; the numbering of the Protein\_A and Protein\_B sequences; and the position-specific score based on Shannon’s entropy that measures the variation between the different PSAs at each position. Figure 6.3 shown an example of the pipeline output imported in the Microsoft Excel software.

>6Tmm	Lga	-	-	-	-	-	0	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	-	13	-	14	15	16	17	18	19	-	
>6Tmm	Lga	-	-	-	-	-	G	A	F	V	S	N	T	-	Q	I	T	M	-	A	-	-	D	-	K	-	Q	K	K	F	I	N	-	
>6Frh	Lga	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	-	13	14	15	16	17	18	19	20	-	
>6Frh	Lga	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	-	T	G	K	R	V	P	I	K	-	
>6Frh	Shoba	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	-	13	14	15	16	17	18	19	20	-	
>6Frh	Shoba	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	-	T	G	K	R	V	P	I	K	-	
>6Frh	parMatt	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	11	12	-	-	13	-	14	-	15	16	17	18	19	20	-	
>6Frh	parMatt	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	A	S	-	-	T	-	G	-	K	R	V	P	I	K	-	
>6Frh	DeepAlign	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	DeepAlign	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	SPalign	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	SPalign	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Mapsci	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	Mapsci	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Cab	-	-	-	-	-	-2	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	12	-	13	-	14	-	15	16	17	18	19	20	-
>6Frh	Cab	-	-	-	-	-	S	G	A	I	S	G	D	S	-	L	I	S	L	-	A	S	-	T	-	G	-	K	R	V	P	I	K	-
>6Frh	tmAlign	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	tmAlign	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Manmoth	-	-	-	-	-	-1	1	2	3	4	5	6	7	8	9	10	11	-	12	-	-	13	-	14	-	15	16	17	18	19	20	21	
>6Frh	Manmoth	-	-	-	-	-	G	A	I	S	G	D	S	L	I	S	L	A	-	S	-	-	T	-	G	-	K	R	V	P	I	K	D	
>6Frh	Matt	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	11	12	-	-	13	-	14	-	15	16	17	18	19	20	-	
>6Frh	Matt	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	A	S	-	-	T	-	G	-	K	R	V	P	I	K	-	
>6Frh	Frimalign	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	Frimalign	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	mTMalign	-6	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	mTMalign	M	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Kpax	-6	-5	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	Kpax	M	L	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Lovodalign	-	-	-	-	-	-	1	2	3	4	5	6	-	7	8	9	10	-	11	-	-	12	13	14	-	15	16	17	18	19	20	-	
>6Frh	Lovodalign	-	-	-	-	-	-	A	I	S	G	D	S	-	L	I	S	L	-	A	-	-	S	T	G	-	K	R	V	P	I	K	-	
>6Frh	Mustang	-	-	-	-	-	-1	1	2	3	4	5	6	-	7	8	9	10	-	11	12	13	14	-	-	-	15	16	17	18	19	20	-	
>6Frh	Mustang	-	-	-	-	-	G	A	I	S	G	D	S	-	L	I	S	L	-	A	S	T	G	-	-	-	-	K	R	V	P	I	K	-
seq	prog	0.567	0.353	0.0	0.0	0.0	0.353	0.353	0.0	0.0	0.0	0.0	0.0	0.353	0.353	0.353	0.353	0.353	0.567	0.722	0.567	0.353	1.199	0.997	0.906	0.567	0.0	0.0	0.0	0.0	0.0	0.353		

Figure 6.3. Structure alignment between the two proteins in coordinate files 6TMM and 6FRH produced by the 15 different PSAPs analysed in this work. The sequences and sequence numbering of both proteins are reported for each PSAP. The bottom line of the alignment indicates the values of Shannon's entropy. There are 207 consensus positions that are structurally aligned in the output of all the PSAPs. This regions have score "0" and are highlighted by a cyan background. Shannon's entropy is 3.37 over 217 alignment positions (1.6%).

Analysis of the results of the pipeline described above for different CASP14 Target-Template pairs shows that, in the case of Targets assigned to the TBM-easy category, the regions that are structurally aligned in the output of all the 15 programmes are relatively large, as shown in the example in Figure 6.3. Visual examination and calculation of RMSD values after optimal least-square superposition of these structurally aligned regions indicates that they actually comprise highly conserved structures (Figure 6.4).



*Figure 6.4. Regions of the 6TMM-6FRH Target-Template pair that are structurally aligned in the output of all 15 examined PSAP, and that are highlighted by a cyan background in Figure 6.3. The RMSD value calculated after optimal least-square fit superimposition of these regions is 0.83Å.*

Conversely, in the case of Targets assigned to the “TBM-hard” category, the consensus structurally aligned regions are significantly reduced, as shown in the example in Figure 6.5.

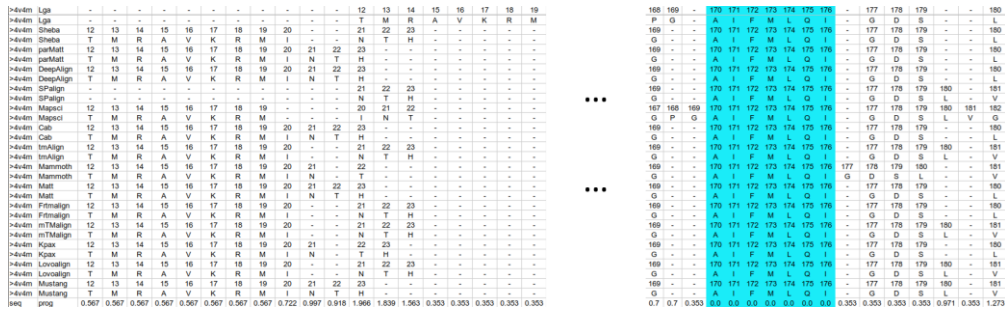


Figure 6.5. Structure alignment between the two proteins in coordinate files 4V4M and 6S44 produced by the 15 different PSAPs analysed in this work. The sequences and sequence numbering of both proteins are reported for each PSAP. The bottom line of the alignment indicates the values of Shannon's entropy. There are 8 consensus positions that are structurally aligned in the output of all the PSAPs. These regions have score "0" in the bottom line and are highlighted by a cyan background. Shannon's entropy is 219.73 over 314 alignment positions (70%).

A statistical analysis of the results of the pipeline in terms of extent of consensus structurally aligned regions, and RMSD values of these regions after optimal least-square fit superposition, as a function of % age sequence identity both within the consensus structurally aligned regions and between the whole sequences of each protein pair is currently ongoing.

## 6.5 Conclusions.

We developed an automated pipeline, comprising four main Python scripts, that takes two protein structures as input, launches the 15 PSAPs that are currently available for download, still working and able to provide a PSA for the two proteins, and generates an alignment of the 15 PSAs provided by the PSAPs, where the consensus regions that are structurally aligned by all the 15 PSAPs are aligned and a score for the other regions is provided, based on the

degree of variability among the PSAs provided by the 15 PSAPs measured by Shannon's entropy.

This pipeline is useful for both inexperienced users and researchers that routinely compare protein structures, since programmes described in the literature or reported in databases to perform protein structure alignments are large in number and highly heterogeneous in terms of output, regions that they take into account for the alignment and, most important, accuracy. Additionally, many of them are not working anymore. The pipeline provides users with the results of the 15 currently available, still working and downloadable PSAPs, and with information on the degree of consensus among the different methods by entering a single command, without the need to search, download and install any of the PSAPs. Starting from the regions with the lowest values of Shannon's entropy, which indicate the highest consensus among the PSAs generated by the 15 PSAPs and, therefore, are expected to be the regions where the PSA is most accurate, the user can decide whether to extend the PSA to other regions with increasing values of Shannon's entropy, based on visual examination of the PSA obtained including these regions and calculation of parameters like RMSD or GDT.

In the future, we are going to:

- i) Run the pipeline on a larger dataset, including Target-Template pairs from all other CASP experiments, to test the pipeline robustness on a much larger number of protein coordinate files.
- ii) Make the pipeline available to the scientific community through the development of a web-based server.

ii) Implement the pipeline by incorporating also the results of server-only PSAPs, some which are popularly used, to have the complete set of PSAs produced by all currently available and working PSAPs.

A long-term goal is the assessment of the performance of the different PSAPs incorporated in the pipeline by comparison of the PSAs generated by them with manually curated and highly reliable PSAs reported in the literature or generated by ourselves, for protein pairs with diverse folds, level of structure similarity, length, and other relevant parameters.



## 7. General conclusions

In conclusion, in this thesis we have applied well-established structural bioinformatics methods to address biotechnological and biomedical problems, and we have developed a method to identify the most similar regions among two protein structures based on the consensus among the results of different structure superposition programs.

- 1) I have implemented a VS-based drug repositioning procedure that allowed six FDA-approved drugs to be proposed for a novel use in HD therapy, which are directly amenable to off-label clinical use. This procedure uses the  $\sigma 1$  receptor as a target and a library of FDA-approved compounds as potential ligands, and the accurate visual inspection of the twenty VS-generated complexes with highest predicted binding affinity to critically evaluate the results and select six compounds for experimental testing. All the six selected compounds were demonstrated to directly bind Hs $\sigma 1$ R *in vitro*, in SPR experiments, and increase the growth of HD patient fibroblasts, indicating that they exert agonistic activity on Hs $\sigma 1$ R, like the prototypic Hs $\sigma 1$ R ligand pridopidine. The most active of these six compounds was the anti-psychotic drug iloperidone, which is also devoid of effects on the growth of normal fibroblasts (Chapter 2 and Battista et al., 2021).
- 2) I have implemented an extensive VS procedure that allowed steroid-based compounds to be proposed to be among the potential physiological ligands of the  $\sigma 1$  receptor. As a result of this procedure, the set of compounds with statistically significant highest affinity

towards the human  $\sigma$ 1 receptor presented an enrichment in compounds comprising a steroidal/sterolic nucleus, as well as in known  $\sigma$ 1R ligands, but not in known non-binders or in categories of compounds with diverse scaffolds, like human metabolites and approved drugs. The hypothesis of steroid-based ligands was supported by results produced by our collaborators: i) the good fitting of ergosterol in the electron density in the binding site of the  $\sigma$ 1 receptor from *Xenopus laevis*, which was determined in the absence of ligands; ii) the ability of the endogenous steroid 16,17-didehydroprogesterone to bind human  $\sigma$ 1 receptor with affinity comparable to that of iloperidone and higher than that of a prototypic  $\sigma$ 1 receptor ligand in fluorescence titration assays; and iii) the results of molecular dynamics simulations that indicate that the ligands access the binding site from the membrane side of the  $\sigma$ 1 receptor (Chapter 2 and Pascarella et al., 2023).

- 3) The results of computational docking towards the structures of COX-2 and COX-1 enzymes indicate that thiocanthal and thiocanthol, the new compounds obtained from olive oil through eco-sustainable procedures, interact with COX-2 involving the same residues that interact with other ligands in known structures, with predicted affinity in the same range as that of COX-2 inhibitors clinically used as anti-inflammatory drugs, and higher for COX-2 than COX-1 (Chapter 3 and Di Risola et al, submitted for publication to Green Chemistry).
- 4) To identify the sequence and structure determinants underlying the antiproliferative activity of Arabidopsis sirtuins demonstrated by our collaborators, I produced the molecular models of Arabidopsis SRT1 and SRT2 and human SIRT4 and compared them with one another and

with the experimentally determined structures of human SIRT6, both at the global and at the substrate binding site level. Human SIRT6 and SIRT4 were chosen because they are the closest homologues of the *Arabidopsis* sirtuins and are functionally characterised. Sequence and structure comparisons highlighted that the residues responsible for the de-acylation and ADP-ribosylation reactions demonstrated for HsSIRT6 and HsSIRT4 are conserved. Additionally, we found an evolutionary relationship between the unique Ig-like domain comprised in AtSRT1 and the human GAS41 protein, which suggests that the function of this domain is to facilitate AtSRT1 interaction with specific substrates, possibly with similar features to those recognized by GAS41 (Chapter 4 and Bruscalupi et al, 2023).

- 5) To explain the catalytic activity of the *Arabidopsis thaliana* AtNATA2 protein towards several substrates, which had been determined by our experimental collaborators, I built a molecular model of the *Arabidopsis thaliana* AtNATA1 and AtNATA2 proteins and performed a comparative analysis of these models and of the experimentally determined structures of the human SSAT1 and SSAT2 homologues. These analyses allowed me to identify specific features in the binding sites of the *Arabidopsis* proteins and of human SSAT2, in terms of both shape and identity of residues lining the binding site, that differ from the binding site of human SSAT1 and might explain their different substrate specificity, in particular the preference of the former three enzymes for shorter amines with respect to human SSAT1 substrates (Chapter 5 and Mattioli et al, 2022).

6) I have implemented an automated pipeline to identify the consensus results among the 15 protein structure alignment programs that are available for download and are functional after installation. Indeed, many of these programs have developed since the first protein structures were determined and the necessity to compare them arose. However, these programs present several problems: many of them are not functional anymore; their ability to superimpose protein structures decreases with the decrease in structural similarity; and it is not clear which program performs better on specific protein structures or in general. To address these problems, we have performed an extensive recognition of the programs reported in the literature and/or in databases, have tested all of them to identify the programs that are still functional, and have developed a pipeline to compare the results of these programmes and identify the consensus regions in their output, based on an easily interpretable scoring function. Since consensus regions among the output of different programs are more likely to correspond to regions of actual structure similarity than be found by all programs by mistake, this pipeline represents a valuable aid to all users who need to perform structure comparisons as part of their research activity (Chapter 6 and Pascarella et al, in preparation).

## 8. Publications resulting from this thesis

\*: equal contribution

Battista, T.\*, **G. Pascarella\***, D.S. Staid\*, G. Colotti, J. Rosati, A. Fiorillo, A. Casamassa, et al. 2021. ‘Known Drugs Identified by Structure-Based Virtual Screening Are Able to Bind Sigma-1 Receptor and Increase Growth of Huntington Disease Patient-Derived Cells’. *International Journal of Molecular Sciences* 22 (3).  
<https://doi.org/10.3390/ijms22031293>.

Bruscalupi, G., P. Di Micco, C.M. Failla, **G. Pascarella**, V. Morea, M. Saliola, A. De Paolis, S. Venditti, and M.L. Mauro. 2023. ‘Arabidopsis Thaliana Sirtuins Control Proliferation and Glutamate Dehydrogenase Activity’. *Plant Physiology and Biochemistry* 194.  
<https://doi.org/10.1016/j.plaphy.2022.11.007>.

Di Risola, D., Mattioli, R., Federico, R., **Pascarella, G.**, Fontana, M., Dainese, E., Dufrusine, B., Ciogli, A., Gasparrini, F., Morea, V., Villani, C., Mosca, L., and Francioso, A. Green synthesis and two-step chromatographic separation of thiocanthal and thiocanthol: Two novel biologically active sulfur derivatives of oleocanthal and oleacein from extra virgin olive oil. Submitted for publication to *Green Chemistry*.

Mattioli, R\*., **G. Pascarella\***, R. D’Incà, A. Cona, R. Angelini, V. Morea, and P. Tavladoraki. 2022. ‘Arabidopsis N-Acetyltransferase Activity 2 Preferentially Acetylates 1,3-Diaminopropane and Thialysine’. *Plant Physiology and Biochemistry* 170.  
<https://doi.org/10.1016/j.plaphy.2021.11.034>.

**Pascarella, G.\***, L. Antonelli\*, D. Narzi, T. Battista, A. Fiorillo, G. Colotti, L. Guidoni, V. Morea, and A. Ilari. 2023. ‘Investigation of the Entry Pathway and Molecular Nature of  $\Sigma 1$  Receptor Ligands’. *International Journal of Molecular Sciences* 24 (7).  
<https://doi.org/10.3390/ijms24076367>.

**Pascarella, G.**, Via, A., and Morea, V. Implementation of a pipeline to identify consensus results from protein structure alignment programs. In preparation.

## References

- Adio AM, Casteel CL, de Vos M, Kim JH, Joshi V, Li B, Juéry C, Daron J, Kliebenstein DJ & Jandera G (2011) Biosynthesis and defensive function of N $\delta$ -Acetylornithine, a jasmonate-induced arabidopsis metabolite. *Plant Cell* 23, 3303–3318.
- Altman RB & Dugan JM (2005) Defining Bioinformatics and Structural Bioinformatics. In *Structural Bioinformatics*. John Wiley and Sons Inc., pp.1–14.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402.
- Attwood TK, Gisel A, Eriksson N-E & Bongcam-Rudloff E (2011) Concepts, Historical Milestones and the Central Place of Bioinformatics in Modern Biology: A European Perspective. In *Bioinformatics - Trends and Methodologies*. InTech.
- Aydar E, Palmer CP, Klyachko VA & Jackson MB (2002) The sigma receptor as a ligand-regulated auxiliary potassium channel subunit. *Neuron* 34, 399–410. Available at: <http://www.cell.com/article/S0896627302006773/fulltext>.
- Battista T, Pascarella G, Staid DS, Colotti G, Rosati J, Fiorillo A, Casamassa A, Vescovi AL, Giabbai B, Semrau MS, Fanelli S, Storici P, Squitieri F, Morea V & Ilari A (2021) Known drugs identified by structure-based virtual screening are able to bind sigma-1 receptor and increase growth of huntington disease patient-derived cells. *Int J Mol Sci* 22.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28, 235–242. Available at: <https://pubmed.ncbi.nlm.nih.gov/10592235/>.
- Bewley MC, Graziano V, Jiang J, Matz E, Studier FW, Pegg AE, Coleman CS & Flanagan JM (2006) Structures of wild-type and mutant human spermidine/spermine N 1-acetyltransferase, a potential therapeutic drug target. *Proc Natl Acad Sci U S A* 103, 2063–2068.
- Bheda P, Jing H, Wolberger C & Lin H (2016) The Substrate Specificity of Sirtuins. *Annu Rev Biochem* 85, 405–429.
- Bode R, Thureau A-M & Schmidt H (1993) Characterization of acetyl-CoA: L-lysine N $\epsilon$ -acetyltransferase, which catalyses the first step of carbon

- catabolism from lysine in *Saccharomyces cerevisiae*. *Arch Microbiol*, 400.
- Bolton EE, Chen J, Kim S, Han L, He S, Shi W, Simonyan V, Sun Y, Thiessen PA, Wang J, Yu B, Zhang J & Bryant SH (2011) PubChem3D: a new resource for scientists. *J Cheminform* 3, 32. Available at: [/pmc/articles/PMC3269824/](https://pubmed.ncbi.nlm.nih.gov/29497125/).
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K & Madden TL (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10.
- Carafa V, Rotili D, Forgione M, Cuomo F, Serrettiello E, Hailu GS, Jarho E, Lahtela-Kakkonen M, Mai A & Altucci L (2016) Sirtuin functions and modulation: from chemistry to the clinic. *Clin Epigenetics* 8.
- Castany S, Gris G, Vela JM, Verdú E & Boadas-Vaello P (2018) Critical role of sigma-1 receptors in central neuropathic pain-related behaviours after mild spinal cord injury in mice. *Sci Rep* 8. Available at: <https://pubmed.ncbi.nlm.nih.gov/29497125/>.
- Chang HC & Guarente L (2014) SIRT1 and other sirtuins in metabolism. *Trends in Endocrinology and Metabolism* 25, 138–145.
- Cho HJ, Li H, Linhares BM, Kim E, Ndoj J, Miao H, Grembecka J & Cierpicki T (2018) GAS41 Recognizes Diacetylated Histone H3 through a Bivalent Binding Mode. *ACS Chem Biol* 13, 2739–2746.
- Chu UB & Ruoho AE (2016) Biochemical pharmacology of the sigma-1 receptor. *Mol Pharmacol* 89, 142–153. Available at: <http://dx.doi.org/10.1124/mol>.
- Coleman CS, Stanley BA, Jones AD & Pegg AE (2004) Spermidine/spermine-N1-acetyltransferase-2 (SSAT2) acetylates thialysine and is not involved in polyamine metabolism. *Biochemical Journal* 384, 139–148. Available at: [/biochemj/article/384/1/139/41451/Spermidine-spermine-N1-acetyltransferase-2-SSAT2](https://pubmed.ncbi.nlm.nih.gov/15111111/) [Accessed October 13, 2023].
- Coleman CS & Pegg AE (1997) Proteasomal degradation of spermidine/spermine N1-acetyltransferase requires the carboxyl-terminal glutamic acid residues. *Journal of Biological Chemistry* 272, 12164–12169.
- Cosconati S, Forli S, Perryman AL, Harris R, Goodsell DS & Olson AJ (2010) Virtual screening with AutoDock: Theory and practice. *Expert Opin Drug Discov* 5, 597–607. Available at: [/pmc/articles/PMC3083070/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/20111111/).

- Dayringer HE, Tramontano A, Sprang SR & Fletterick RJ (1986) Interactive program for visualization and modelling of proteins, nucleic acids and small molecules. *J Mol Graph* 4, 82–87.
- Dong R, Peng Z, Zhang Y & Yang J (2018) MTM-align: An algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* 34, 1719–1725.
- Del Duca S, Beninati S & Serafini-Fracassini D (1995) Polyamines in chloroplasts: Identification of their glutamyl and acetyl derivatives. *Biochemical Journal* 305, 233–237.
- Eddings CR, Arbez N, Akimov S, Geva M, Hayden MR & Ross CA (2019) Pridopidine protects neurons from mutant-huntingtin toxicity via the sigma-1 receptor. *Neurobiol Dis* 129, 118–129. Available at: <https://pubmed.ncbi.nlm.nih.gov/31108174/>.
- Francardo V, Geva M, Bez F, Denis Q, Steiner L, Hayden MR & Cenci MA (2019) Pridopidine Induces Functional Neurorestoration Via the Sigma-1 Receptor in a Mouse Model of Parkinson’s Disease. *Neurotherapeutics* 16, 465–479. Available at: <https://pubmed.ncbi.nlm.nih.gov/30756361/>.
- Francioso A, Federico R, Maggiore A, Fontana M, Boffi A, D’Erme M & Mosca L (2020) Green Route for the Isolation and Purification of Hydroxytyrosol, Tyrosol, Oleacein and Oleocanthal from Extra Virgin Olive Oil. *Molecules* 25. Available at: </pmc/articles/PMC7464626/> [Accessed October 12, 2023].
- Fukuchi JI, Kashiwagi K, Takio K & Igarashi K (1994) Properties and structure of spermidine acetyltransferase in Escherichia coli. *Journal of Biological Chemistry* 269, 22581–22585.
- Geva M, Kusko R, Soares H, Fowler KD, Birnberg T, Barash S, Merenlender-Wagner A, Fine T, Lysaght A, Weiner B, Cha Y, Kolitz S, Towfic F, Orbach A, Laufer R, Zeskind B, Grossman I & Hayden MR (2016) Pridopidine activates neuroprotective pathways impaired in Huntington Disease. *Hum Mol Genet* 25, 3975–3987. Available at: <https://pubmed.ncbi.nlm.nih.gov/27466197/>.
- Grewal R, Reutzel M, Dilberger B, Hein H, Zotzel J, Marx S, Tretzel J, Sarafeddin A, Fuchs C & Eckert GP (2020) Purified oleocanthal and ligstroside protect against mitochondrial dysfunction in models of early Alzheimer’s disease and brain ageing. *Exp Neurol* 328, 113248.
- Guex N & Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. Available at: <https://pubmed.ncbi.nlm.nih.gov/9504803/>.



- Han BW, Bingman CA, Wesenberg GE & Phillips GN (2006) Crystal structure of Homo sapiens thialysine N  $\epsilon$ -acetyltransferase (HsSSAT2) in complex with acetyl coenzyme A. *Proteins: Structure, Function, and Bioinformatics* 64, 288–293.
- Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P & Steinbeck C (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* 44, D1214–D1219. Available at: <https://pubmed.ncbi.nlm.nih.gov/26467479/>.
- Hayashi T, Tsai SY, Mori T, Fujimoto M & Su TP (2011) Targeting ligand-operated chaperone sigma-1 receptors in the treatment of neuropsychiatric disorders. *Expert Opin Ther Targets* 15, 557–577. Available at: <https://pubmed.ncbi.nlm.nih.gov/21375464/>.
- Hegde SS, Chandler J, Vetting MW, Yu M & Blanchard JS (2007) Mechanistic and structural analysis of human spermidine/spermine N 1-acetyltransferase. *Biochemistry* 46, 7187–7195. Available at: <https://pubs.acs.org/doi/abs/10.1021/bi700256z> [Accessed October 13, 2023].
- Hennion F, Bouchereau A, Gauthier C, Hermant M, Vernon P & Prinzing A (2012) Variation in amine composition in plant species: How it integrates macroevolutionary and environmental signals. *Am J Bot* 99, 36–45.
- Hennion F, Frenot Y & Martin-Tanguy J (2006) High flexibility in growth and polyamine composition of the crucifer *Pringlea antiscorbutica* in relation to environmental conditions. *Physiol Plant* 127, 212–224.
- Hogeweg P (2011) The roots of bioinformatics in theoretical biology. *PLoS Comput Biol* 7.
- Hollender C & Liu Z (2008) Histone deacetylase genes in arabidopsis development. *J Integr Plant Biol* 50, 875–885.
- Huang L, Sun Q, Qin F, Li C, Zhao Y & Zhou DX (2007) Down-regulation of a Silent Information Regulator2-related histone deacetylase gene, OsSRT1, induces DNA fragmentation and cell death in rice. *Plant Physiol* 144, 1508–1519.
- Huey R, Morris GM, Olson AJ & Goodsell DS (2007) Software news and update a semiempirical free energy force field with charge-based desolvation. *J Comput Chem* 28, 1145–1152. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.20634>.

- Ilinkin I, Ye J & Janardan R (2010) Multiple structure alignment and consensus identification for proteins. Available at: <http://www.biomedcentral.com/1471-2105/11/71>.
- Ionescu A, Gradus T, Altman T, Maimon R, Saraf Avraham N, Geva M, Hayden M & Perlson E (2019) Targeting the Sigma-1 Receptor via Pridopidine Ameliorates Central Features of ALS Pathology in a SOD1 G93A Model. *Cell Death Dis* 10. Available at: <https://pubmed.ncbi.nlm.nih.gov/30824685/>.
- Jammes F, Leonhardt N, Tran D, Bousserouel H, Véry AA, Renou JP, Vavasseur A, Kwak JM, Sentenac H, Bouteau F & Leung J (2014) Acetylated 1,3-diaminopropane antagonizes abscisic acid-mediated stomatal closing in Arabidopsis. *Plant Journal* 79, 322–333.
- Jordá T & Puig S (2020) Regulation of Ergosterol Biosynthesis in *Saccharomyces cerevisiae*. *Genes (Basel)* 11, 1–18. Available at: <https://pubmed.ncbi.nlm.nih.gov/32679672/>.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P & Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583. Available at: </pmc/articles/PMC8371605/>.
- Jun DY, Rue SW, Han KH, Taub D, Lee YS, Bae YS & Kim YH (2003) Mechanism underlying cytotoxicity of thialysine, lysine analog, toward human acute leukemia Jurkat T cells. *Biochem Pharmacol* 66, 2291–2300.
- Jung J & Lee B (2000) Protein structure alignment using environmental profiles. *Protein Eng* 13, 535–543. Available at: [http://ftp.embl-heidelberg.de/pub/databases/pdb\\_select](http://ftp.embl-heidelberg.de/pub/databases/pdb_select).
- Kanehisa M & Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28, 27–30. Available at: <https://pubmed.ncbi.nlm.nih.gov/10592173/>.
- Kim FJ, Kovalyshyn I, Burgman M, Neilan C, Chien CC & Pasternak GW (2010)  $\sigma_1$  Receptor modulation of G-protein-coupled receptor signaling: Potentiation of opioid transduction independent from receptor binding. *Mol Pharmacol* 77, 695–703. Available at: <https://pubmed.ncbi.nlm.nih.gov/20089882/>.

- Konagurthu AS, Whisstock JC, Stuckey PJ & Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins: Structure, Function and Genetics* 64, 559–574.
- Kongsamut S, Roehr JE, Cai J, Hartman HB, Weissensee P, Kerman LL, Tang L & Sandrasagra A (1996) Iloperidone binding to human and rat dopamine and 5-HT receptors. *Eur J Pharmacol* 317, 417–423. Available at: <https://pubmed.ncbi.nlm.nih.gov/8997630/>.
- König AC, Hartl M, Pham PA, Laxa M, Boersema PJ, Orwat A, Kalitventseva I, Plöckinger M, Braun HP, Leister D, Mann M, Wachter A, Fernie AR & Finkemeier I (2014) The Arabidopsis class II sirtuin is a lysine deacetylase and interacts with mitochondrial energy metabolism. *Plant Physiol* 164, 1401–1414.
- Krieger E, Nabuurs SB & Vriend G (2005) Homology Modeling. In *Structural Bioinformatics*. John Wiley and Sons Inc., pp.509–523.
- Kryshchak A, Schwede T, Topf M, Fidelis K & Moult J (2019) Critical assessment of methods of protein structure prediction (CASP)–Round XIII. *Proteins* 87, 1011–1020. Available at: <https://pubmed.ncbi.nlm.nih.gov/31589781/> [Accessed October 12, 2023].
- Lavecchia A & Di Giovanni C (2013) *Send Orders of Reprints at reprints@benthamscience.net Current Medicinal Chemistry*, Available at: <http://www.enamine.net/>.
- Lie Fliniaux O, Raynaud-Le Grandic S, Baltora-Rosset S, Bienaime C, Robins RJ & Fliniaux M (2004) Altered nitrogen metabolism associated with de-differentiated suspension cultures derived from root cultures of *Datura stramonium* studied by heteronuclear multiple bond coherence (HMBC) NMR spectroscopy. Available at: <https://academic.oup.com/jxb/article/55/399/1053/524940> [Accessed October 13, 2023].
- Lin HJ, Lien YC & Hsu CH (2010) A high-throughput colorimetric assay to characterize the enzyme kinetic and cellular activity of spermidine/spermine N1-acetyltransferase 1. *Anal Biochem* 407, 226–232.
- Liu X, Wei W, Zhu W, Su L, Xiong Z, Zhou M, Zheng Y & Zhou DX (2017) Histone Deacetylase AtSRT1 Links Metabolic Flux and Stress Response in Arabidopsis. *Mol Plant* 10, 1510–1522.
- Lou YR, Ahmed S, Yan J, Adio AM, Powell HM, Morris PF & Jander G (2020) Arabidopsis ADC1 functions as an N $\delta$ -acetylornithine decarboxylase. *J Integr Plant Biol* 62, 601–613.

- Lou YR, Bor M, Yan J, Preuss AS & Jander G (2016) Arabidopsis NATA1 acetylates putrescine and decreases defense-related hydrogen peroxide accumulation. *Plant Physiol* 171, 1443–1455.
- Lozano-Castellón J, López-Yerena A, Rinaldi de Alvarenga JF, Romero del Castillo-Alba J, Vallverdú-Queralt A, Escribano-Ferrer E & Lamuela-Raventós RM (2020) Health-promoting properties of oleocanthal and oleacein: Two secoiridoids from extra-virgin olive oil. *Crit Rev Food Sci Nutr* 60, 2532–2548. Available at: <https://www.tandfonline.com/doi/abs/10.1080/10408398.2019.1650715> [Accessed October 12, 2023].
- Lu L, Berkey KA & Casero RA (1996) RGFSGS is an amino acid sequence required for acetyl coenzyme A binding and activity of human spermidine/spermine N1 acetyltransferase. *Journal of Biological Chemistry* 271, 18920–18924.
- Lu Y, Xu Q, Liu Y, Yu Y, Cheng ZY, Zhao Y & Zhou DX (2018) Dynamics and functional interplay of histone lysine butyrylation, crotonylation, and acetylation in rice under starvation and submergence. *Genome Biol* 19.
- Lucido MJ, Orlando BJ, Vecchio AJ & Malkowski MG (2016) THE CRYSTAL STRUCTURE OF ASPIRIN ACETYLATED HUMAN CYCLOOXYGENASE-2: INSIGHT INTO THE FORMATION OF PRODUCTS WITH REVERSED STEREOCHEMISTRY. *Biochemistry* 55, 1226. Available at: </pmc/articles/PMC4775376/> [Accessed October 12, 2023].
- Maher CM, Thomas JD, Haas DA, Longen CG, Oyer HM, Tong JY & Kim FJ (2018) Small-Molecule Sigma1 Modulator Induces Autophagic Degradation of PD-L1. *Molecular Cancer Research* 16, 243–255. Available at: <https://pubmed.ncbi.nlm.nih.gov/29117944/>.
- Mancuso R, Oliván S, Rando A, Casas C, Osta R & Navarro X (2012) Sigma-1R Agonist Improves Motor Function and Motoneuron Survival in ALS Mice. *Neurotherapeutics* 9, 814–826. Available at: <https://pubmed.ncbi.nlm.nih.gov/22935988/>.
- Martínez L, Andreani R & Martínez JM (2007) Convergent algorithms for protein structural alignment. *BMC Bioinformatics* 8.
- Mauri MC, Paletta S, Maffini M, Colasanti A, Dragogna F, Di Pace C & Altamura AC (2014) Clinical pharmacology of atypical antipsychotics: An update. *EXCLI J* 13, 1163–1191. Available at: <https://pubmed.ncbi.nlm.nih.gov/26417330/>.

- Maurice T & Su TP (2009) The pharmacology of sigma-1 receptors. *Pharmacol Ther* 124, 195–206. Available at: <https://pubmed.ncbi.nlm.nih.gov/19619582/>.
- McGarry A, Leinonen M, Kieburtz K, Geva M, Olanow CW & Hayden M (2020) Effects of Pridopidine on Functional Capacity in Early-Stage Participants from the PRIDE-HD Study. *J Huntingtons Dis*, 1–10. Available at: <https://pubmed.ncbi.nlm.nih.gov/33164941/>.
- Meng F, Xiao Y, Ji Y, Sun Z & Zhou X (2022) An open-like conformation of the sigma-1 receptor reveals its ligand entry pathway. *Nature Communications* 2022 13:1 13, 1–11. Available at: <https://www.nature.com/articles/s41467-022-28946-w>.
- Menke M, Berger B & Cowen L (2008) Matt: Local flexibility aids protein multiple structure alignment. *PLoS Comput Biol* 4, 88–99.
- Mesnard F, Azaroual N, Marty D, Fliniaux MA, Robins RJ, Vermeersch G & Monti JP (2000) Use of <sup>15</sup>N reverse gradient two-dimensional nuclear magnetic resonance spectroscopy to follow metabolic activity in *Nicotiana plumbaginifolia* cell-suspension cultures. *Planta* 210, 446–453.
- Miciaccia M, Belviso BD, Iaselli M, Cingolani G, Ferorelli S, Cappellari M, Loguercio Polosa P, Perrone MG, Caliandro R & Scilimati A (2021) Three-dimensional structure of human cyclooxygenase (hCOX)-1. *Sci Rep* 11, 4312. Available at: </pmc/articles/PMC7900114/> [Accessed October 12, 2023].
- Montemayor EJ & Hoffman DW (2008) The crystal structure of spermidine/spermine N1- acetyltransferase in complex with spermine provides insights into substrate binding and catalysis. *Biochemistry* 47, 9145–9153.
- Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ, Ruth H, Lindstrom W, Sanner MF, Belew RK, Goodsell DS & Olson AJ (2009) Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30, 2785–2791. Available at: </pmc/articles/PMC2760638/?report=abstract>.
- Moult J, Pedersen JT, Judson R & Fidelis K (1995) A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics* 23, ii–iv. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1002/prot.340230303> [Accessed October 12, 2023].

- Muhammed MT & Aki-Yalcin E (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des* 93, 12–20.
- Navarro G, Moreno E, Aymerich M, Marcellino D, McCormick PJ, Mallol J, Cortés A, Casadó V, Canela EI, Ortiz J, Fuxe K, Lluís C, Ferré S & Franco R (2010) Direct involvement of  $\sigma$ -1 receptors in the dopamine D1 receptor-mediated effects of cocaine. *Proc Natl Acad Sci U S A* 107, 18676–18681. Available at: <https://pubmed.ncbi.nlm.nih.gov/20956312/>.
- Neuwald AF & Landsman D (1997) GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem Sci* 22, 154–155.
- Nguyen L, Lucke-Wold BP, Mookerjee SA, Cavendish JZ, Robson MJ, Scandinaro AL & Matsumoto RR (2015) Role of sigma-1 receptors in neurodegenerative diseases. *J Pharmacol Sci* 127, 17–29.
- O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T & Hutchison GR (2011) Open Babel: An Open chemical toolbox. *J Cheminform* 3, 33. Available at: <http://jcheminf.springeropen.com/articles/10.1186/1758-2946-3-33>.
- Ortiz AR, Strauss CEM & Olmea O (2009) MAMMOTH (Matching molecular models obtained from theory): An automated method for model comparison. *Protein Science* 11, 2606–2621.
- Pandey R, Müller A, Napoli CA, Selinger DA, Pikaard CS, Richards EJ, Bender J, Mount DW & Jorgensen RA (2002) Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Res* 30, 5036–5055.
- Pandit SB & Skolnick J (2008) Fr-TM-align: A new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinformatics* 9.
- Parkinson L & Keast R (2014) Oleocanthal, a Phenolic Derived from Virgin Olive Oil: A Review of the Beneficial Effects on Inflammatory Disease. *Int J Mol Sci* 15, 12323. Available at: </pmc/articles/PMC4139846/> [Accessed October 12, 2023].
- Pegg AE (2008) Spermidine/spermine-N1-acetyltransferase: A key metabolic regulator. *Am J Physiol Endocrinol Metab* 294, 995–1010. Available at: <https://journals.physiology.org/doi/10.1152/ajpendo.90217.2008> [Accessed October 13, 2023].

- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605–1612. Available at: <https://pubmed.ncbi.nlm.nih.gov/15264254/>.
- Proietti G, Rainone G, Hintzen JCJ & Mecinović J (2020) Exploring the Histone Acylome through Incorporation of  $\gamma$ -Thialysine on Histone Tails. *Bioconjug Chem* 31, 844–851.
- Della Ragione F & Pegg AE (1983) Studies of the specificity and kinetics of rat liver spermidine/spermine N1-acetyltransferase. *Biochem J* 213, 701–706.
- Ramirez-Gaona M, Marcu A, Pon A, Guo AC, Sajed T, Wishart NA, Karu N, Feunang YD, Arndt D & Wishart DS (2017) YMDB 2.0: a significantly expanded version of the yeast metabolome database. *Nucleic Acids Res* 45, D440. Available at: [/pmc/articles/PMC5210545/](https://pubmed.ncbi.nlm.nih.gov/27111111/).
- Reilmann R, McGarry A, Grachev ID, Savola JM, Borowsky B, Eyal E, Gross N, Langbehn D, Schubert R, Wickenberg AT, Papapetropoulos S, Hayden M, Squitieri F, Kieburtz K, Landwehrmeyer GB, Agarwal P, Anderson KE, Aziz NA, Azulay JP, Bachoud-Levi AC, Barker R, Bebak A, Beuth M, Biglan K, Blin S, Bohlen S, Bonelli R, Caldwell S, Calvas F, Carlos J, Castagliuolo S, Chong T, Chua P, Coleman A, Corey-Bloom J, Cousins R, Craufurd D, Davison J, Decorte E, De Michele G, Dornhege L, Feigin A, Gallehawk S, Gauteul P, Gonzales C, Griffith J, Gustov A, Guttman M, Heim B, Heller H, Hjerminde L, Illarioshkin S, Ivanko L, Jaynes J, Jenckes M, Kaminski B, Kampstra A, Konkel A, Kopishinskaya S, Krystkowiak P, Komati SK, Kwako A, Lakoning S, Latipova G, Leavitt B, Loy C, MacFarlane C, Madsen L, Marder K, Mason S, Mendis N, Mendis T, Nemeth A, Nevitt L, Norris V, O'Neill C, Olivier A, Orth M, Owens A, Panegyres P, Perlman S, Preston J, Priller J, Puch A, Quarrell O, Ragosta D, Rialland A, Rickards H, Romoli AM, Ross C, Rosser A, Rudzinska M, Russo C V, Saft C, Segro V, Seppi K, Shannon B, Shprecher D, Simonin C, Skitt Z, Slawek J, Soliveri P, Sorbi S, Suski V, Stepniak I, Sungmee P, Temirbaeva S, Testa C, Torvin-Moller A, Uhl S, Vangsted-Hansen C, Verny C, Wall P, Walker F, Wasserman P, Witkowski G, Wright J, Zalyalova Z & Zielonka D (2019) Safety and efficacy of pridopidine in patients with Huntington's disease (PRIDE-HD): a phase 2, randomised, placebo-controlled, multicentre, dose-ranging study. *Lancet Neurol* 18, 165–176. Available at: <https://pubmed.ncbi.nlm.nih.gov/30563778/>.

- Ritchie DW, Ghoorah AW, Mavridis L & Venkatraman V (2012) Fast protein structure alignment using Gaussian overlap scoring of backbone peptide fragment similarity. *Bioinformatics* 28, 3274–3281.
- Rossino G, Orellana I, Caballero J, Schepmann D, Wünsch B, Rui M, Rossi D, González-Avendaño M, Collina S & Vergara-Jaque A (2020) New Insights into the Opening of the Occluded Ligand-Binding Pocket of Sigma1 Receptor: Binding of a Novel Bivalent RC-33 Derivative. *J Chem Inf Model* 60, 756–765. Available at: <https://pubs.acs.org/doi/abs/10.1021/acs.jcim.9b00649> [Accessed October 12, 2023].
- Van Rossum G & Drake FL (2009) Python 3 Reference Manual; CreateSpace. *Scotts Valley, CA*, 242. Available at: <https://www.python.org/> [Accessed October 12, 2023].
- Roth BL, Lopez E, Patel S & Kroeze WK (2016) The Multiplicity of Serotonin Receptors: Uselessly Diverse Molecules or an Embarrassment of Riches? <http://dx.doi.org/10.1177/107385840000600408> 6, 252–262. Available at: <https://journals.sagepub.com/doi/10.1177/107385840000600408>.
- Ryskamp D, Wu J, Geva M, Kusko R, Grossman I, Hayden M & Bezprozvanny I (2017) The sigma-1 receptor mediates the beneficial effects of pridopidine in a mouse model of Huntington disease. *Neurobiol Dis* 97, 46–59. Available at: <https://pubmed.ncbi.nlm.nih.gov/27818324/>.
- Ryskamp D, Wu L, Wu J, Kim D, Rammes G, Geva M, Hayden M & Bezprozvanny I (2019) Pridopidine stabilizes mushroom spines in mouse models of Alzheimer’s disease by acting on the sigma-1 receptor. *Neurobiol Dis* 124, 489–504. Available at: <https://pubmed.ncbi.nlm.nih.gov/30594810/>.
- Sahlholm K, Århem P, Fuxe K & Marcellino D (2013) The dopamine stabilizers ACR16 and (-)-OSU6162 display nanomolar affinities at the  $\sigma$ -1 receptor. *Mol Psychiatry* 18, 12–14. Available at: <http://www.nature.com/mp>.
- Sánchez-Castañeda C, de Pasquale F, Caravasso CF, Marano M, Maffi S, Migliore S, Sabatini U & Squitieri F (2017) Resting-state connectivity and modulated somatomotor and default-mode networks in Huntington disease. *CNS Neurosci Ther* 23, 488–497. Available at: <https://pubmed.ncbi.nlm.nih.gov/28464463/>.
- Sánchez-López J, Camañes G, Flors V, Vicent C, Pastor V, Vicedo B, Cerezo M & García-Agustín P (2009) Underivatized polyamine analysis in



- plant samples by ion pair LC coupled with electrospray tandem mass spectrometry. *Plant Physiology and Biochemistry* 47, 592–598.
- Schmidt HR, Betz RM, Dror RO & Kruse AC (2018) Structural basis for  $\sigma 1$  receptor ligand recognition. *Nat Struct Mol Biol* 25, 981–987. Available at: <https://pubmed.ncbi.nlm.nih.gov/30291362/>.
- Schmidt HR, Zheng S, Gurpinar E, Koehl A, Manglik A & Kruse AC (2016) Crystal structure of the human  $\sigma 1$  receptor. *Nature* 532, 527–530. Available at: <http://www.nature.com/articles/nature17391>.
- Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Židek A, Nelson AWR, Bridgland A, Penedones H, Petersen S, Simonyan K, Crossan S, Kohli P, Jones DT, Silver D, Kavukcuoglu K & Hassabis D (2020) Improved protein structure prediction using potentials from deep learning. *Nature* 2020 577:7792 577, 706–710. Available at: <https://www.nature.com/articles/s41586-019-1923-7> [Accessed October 12, 2023].
- Servili M, Sordini B, Esposto S, Urbani S, Veneziani G, Maio I Di, Selvaggini R & Taticchi A (2014) Biological Activities of Phenolic Compounds of Extra Virgin Olive Oil. *Antioxidants* 3, 1–23. Available at: [www.mdpi.com/journal/antioxidants](http://www.mdpi.com/journal/antioxidants) [Accessed October 12, 2023].
- Shegay M V, Suplatov DA, Popova NN, Švedas VK & Voevodin V V (2019) parMATT: parallel multiple alignment of protein 3D-structures with translations and twists for distributed-memory systems. *Bioinformatics*.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7.
- Squitieri F, Landwehrmeyer B, Reilmann R, Rosser A, De Yebenes JG, Prang A, Ivkovic J, Bright J & Rembratt Å (2013) One-year safety and tolerability profile of pridopidine in patients with Huntington disease. *Neurology* 80, 1086–1094. Available at: <https://pubmed.ncbi.nlm.nih.gov/23446684/>.
- Squitieri F, Di Pardo A, Favellato M, Amico E, Maglione V & Frati L (2015) Pridopidine, a dopamine stabilizer, improves motor performance and shows neuroprotective effects in Huntington disease R6/2 mouse model. *J Cell Mol Med* 19, 2540–2548. Available at: <https://pubmed.ncbi.nlm.nih.gov/26094900/>.
- Sterling T & Irwin JJ (2015) ZINC 15 - Ligand Discovery for Everyone. *J Chem Inf Model* 55, 2324–2337. Available at: <https://pubmed.ncbi.nlm.nih.gov/26479676/>.

- Sud M, Fahy E, Cotter D, Brown A, Dennis EA, Glass CK, Merrill AH, Murphy RC, Raetz CRH, Russell DW & Subramaniam S (2007) LMSD: LIPID MAPS structure database. *Nucleic Acids Res* 35. Available at: <https://pubmed.ncbi.nlm.nih.gov/17098933/>.
- Sugiyama S, Ishikawa S, Tomitori H, Niiyama M, Hirose M, Miyazaki Y, Higashi K, Murata M, Adachi H, Takano K, Murakami S, Inoue T, Mori Y, Kashiwagi K, Igarashi K & Matsumura H (2016) Molecular mechanism underlying promiscuous polyamine recognition by spermidine acetyltransferase. *International Journal of Biochemistry and Cell Biology* 76, 87–97.
- Tassoni A, Van Buuren M, Franceschetti M, Fornalè S & Bagni N (2000) Polyamine content and metabolism in *Arabidopsis thaliana* and effect of spermidine on plant development. *Plant Physiology and Biochemistry* 38, 383–393.
- Tavladoraki P, Cona A, Federico R, Tempera G, Viceconte N, Saccoccio S, Battaglia V, Toninello A & Agostinelli E (2012) Polyamine catabolism: Target for antiproliferative therapies in animals and stress tolerance strategies in plants. *Amino Acids* 42, 411–426.
- Terashi G & Takeda-Shitaka M (2015) CAB-align: A flexible protein structure alignment method based on the residue-residue contact area. *PLoS One* 10.
- Toumi I, Pagoulatou MG, Margaritopoulou T, Milioni D & Roubelakis-Angelakis KA (2019) Genetically modified heat shock protein90s and polyamine oxidases in *Arabidopsis* reveal their interaction under heat stress affecting polyamine acetylation, oxidation and homeostasis of reactive oxygen species. *Plants* 8.
- Trott O & Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31, NA–NA. Available at: <http://doi.wiley.com/10.1002/jcc.21334>.
- Wang C, Gao F, Wu J, Dai J, Wei C & Li Y (2010) *Arabidopsis* putative deacetylase AtSRT2 regulates basal defense by suppressing PAD4, EDS5 and SID2 expression. *Plant Cell Physiol* 51, 1291–1299.
- Wang S, Peng J & Xu J (2011) Alignment of distantly related protein structures: Algorithm, bound and implications to homology modeling. *Bioinformatics* 27, 2537–2545.
- Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, MacIejewski A, Gale N, Wilson A, Chin L, Cummings R, Le Di, Pon A,

- Knox C & Wilson M (2018) DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 46, D1074–D1082. Available at: <https://pubmed.ncbi.nlm.nih.gov/29126136/>.
- Wu Z & Bowen WD (2008) Role of sigma-1 receptor C-terminal segment in inositol 1,4,5-trisphosphate receptor activation: Constitutive enhancement of calcium signaling in MCF-7 tumor cells. *Journal of Biological Chemistry* 283, 28198–28215. Available at: </pmc/articles/PMC2661391/?report=abstract>.
- Yang Y, Zhan J, Zhao H & Zhou Y (2012) A new size-independent score for pairwise protein structure alignment and its application to structure classification and nucleic-acid binding prediction. *Proteins: Structure, Function and Bioinformatics* 80, 2080–2088.
- Zemla A (2003) LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 31, 3370–3374.
- Zhang F, Wang L, Ko EE, Shao K & Qiao H (2018) Histone deacetylases SRT1 and SRT2 interact with ENAP1 to mediate ethylene-induced transcriptional repression. *Plant Cell* 30, 153–166.
- Zhang Y & Skolnick J (2005) TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33, 2302–2309.
- Zheng W (2020) Review: The plant sirtuins. *Plant Science* 293.
- Zhong X, Zhang H, Zhao Y, Sun Q, Hu Y, Peng H & Zhou DX (2013) The Rice NAD<sup>+</sup>-Dependent Histone Deacetylase OsSRT1 Targets Preferentially to Stress- and Metabolism-Related Genes and Transposable Elements. *PLoS One* 8.
- Zhu YQ, Zhu DY, Yin L, Zhang Y, Vornrhein C & Wang DC (2006) Crystal structure of human spermidine/spermine N1- acetyltransferase (hSSAT): The first structure of a new sequence family of transferase homologous superfamily. *Proteins: Structure, Function and Genetics* 63, 1127–1131.

## Glossary.

Abbreviation	Definition
$\sigma$ 1R	Sigma1-receptor
AD	Alzheimer's disease
ADT	AutoDock Tools
AF	AlphaFold
AI	Artificial intelligence
ALS	Amyotrophic lateral sclerosis
CASP	Critical Assessment of Structure Prediction
CNS	Central nervous system
COA	Coenzyme-A
DAP	1,3-diaminopropane
EMA	European Medicines Agency
FDA	Food and Drug Administration
GDT	Global Distance Test
HD	Huntington's disease
LBVS	Ligand-based Virtual Screening
MD	Molecular dynamics
NCBI	National Center for Biotechnology Information
ORN	Ornithine
PAO	Polyamine oxidase
PD	Parkinson's Disease
PDB	Protein Data Bank
PUT	Putrescine

RMSD	Root-mean square deviation
SAT	Spd-acetyltransferase
SBVS	Structure-based Virtual Screening
SPD	Spermidine
SPM	Spermine
SPR	Surface Plasmon Resonance
SSAT	Spermidine/Spermine N1-acetyltransferases
VS	Virtual Screening
YMDB	Yeast metabolome database

## Appendix A.

### **GDT\_TS (Global Distance Test Total Score).**

This parameter indicates the percentage of atoms of two structures whose distance is lower than a specified threshold.

$GDT\_TS = (GDT\_P1 + GDT\_P2 + GDT\_P4 + GDT\_P8)/4$ , where  $GDT\_Pn$  denotes percent of residues under distance cutoff  $\leq n\text{\AA}$  (CASP14 definition)

### **RMSD (Root Mean Square Deviation)**

It is a parameter calculated using the following formula:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n d_i^2}$$

Where  $d_i$  is the distance between a given atom and its counterpart in the reference structure.

### **Shannon's Entropy**

It is a parameter calculated using the following formula:

$$H(X) = - \sum_x p(x) \log_2(p(x))$$

Where  $H(X)$  is the entropy to a variable  $x$  which can assume  $n$  values and  $p(x)$  is the probability that a given event occurs.

## Appendix B.

<b>Name</b>	<b>Download Server</b>	<b>Set</b>	<b>Author</b>
CAB-Align	D	F	Terashi et Takeda-Shitaka
DeepAlign	D	F	S. Wang and J. Xu
Fr-TM-align	D	F	S.B. Pandit and J. Skolnick
Kpax	D	F	D.W. Ritchie et. al.
LGA	D/S	F	A. Zemla
LOVOALIGN	D	F	Andreani et al.
MAMMOTH	D/S	F	CEM Strauss and AR Ortiz
MAMMOTH-MULT	D	F	D. Lupyan
Mapsci	D/S	F	Ilinkin et al R. Dong, Z. Peng, Y. Zhang and J. Yang
mTM-align	D/S	F	Yang
MUSTANG	D	F	A.S. Konagurthu et al.
parMATT	D	F	Shegay et al.
SHEBA	D	F	J Jung and B Lee
SPalign	D/S	F	Y. Yang et.al.
TM-align	D/S	F	Y. Zhang and J. Skolnick
FAST	D	FNW	J. Zhu
MIC/MICAN	D	FNW	S.Minami et. al.
BLOMAPS	D	NF	W-M. Zheng and S. Wang
Chimera	D	NF	E. Meng et al.
DaliLite	D/S	NF	L. Holm
GANGSTA+	D/S	NF	A. Guerler and E.W. Knapp
MASS	D/S	NF	O. Dror and H. Wolfson
MatAlign	D	NF	Z. Aung and K.L. Tan
MMLigner	D/S	NF	J. Collier et al.
ProBiS	D/S	NF	J. Konc and D. Janezic
pyMCPSC	D	NF	Sharma A, Manolakos ES
PyMOL	D	NF	W. L. DeLano
SSAP	D	NF	C. Orengo and W. Taylor
STAMP	D	NF	R. Russell and G. Barton
THESEUS	D	NF	D.L. Theobald and D.S. Wuttke

3D-Blast	D/S	NW	Yang and Tung
CBA	D	NW	J. Ebert
CLEPAPS	D	NW	W-M. Zheng and S. Wang
ComSubstruct	D	NW	N. Morikawa
CSR	D/S	NW	M. Petitjean
CTSS	D	NW	T. Can
EXPRESSO	D/S	NW	C. Notredame et al.
Fit3D	D/S	NW	F. Kaiser et al.
Flexsnap	D	NW	Salem et al.
FoldMiner	D	NW	Shapiro and Brutlag
LOCK 2	D	NW	J. Shapiro
LSQRMS	D	NW	Alexandrov and Gerstein
Matt	D/S	NW	M. Menke
MSVNS for MaxCMO	D	NW	D. Pelta et al.
ProCKSI	D	NW	D. Barthel et al.
ProSMoS	D/S	NW	S. Shi et al.
QP Tableau Search	D	NW	A.Stivala et al.
RCSB PCT	D/S	NW	A. Prlic et al.
SA Tableau Search	D	NW	A.Stivala et al.
SARST	D/S	NW	W-C. Lo et al.
SAS-Pro	D	NW	Shah and Sahinidis
SCALI	D/S	NW	X. Yuan and C. Bystroff
Staccato	D	NW	M. Shatsky and H. Wolfson
Structal	D/S	NW	Gerstein and Levitt
TopMatch	D/S	NW	M. Sippl and M. Wiederstein
URMS	D	NW	K. Kedem
Vorolign	D	NW	F. Birzele et al.
YAKUSA	D/S	NW	M. Carpentier et al.
MOE	D	P	Chemical Computing Group
CLICK	S	-	M. Nguyen
CE	S	-	I. Shindyalov
Esript	S	-	Robert et Gouet
FATCAT	S	-	Y. Ye and A. Godzik
FlexProt	S	-	M. Shatsky and H. Wolfson
KENOBI/K2/K2SA	S	-	Z. Weng
MISTRAL	S	-	C. Micheletti and H. Orland



msTALI	S	-	P. Shealy and H. Valafar
mulPBA	S	-	A.P. Joseph et. al.
MultiProt	S	-	M. Shatsky and H. Wolfson
POSA	S	-	Y. Ye and A. Godzik
PrISM	S	-	B. Honig
ProFit	S	-	ACR. Martin
Protein3Dfit	S	-	D. Schomburg
RAPIDO	S	-	R. Mosca and T.R. Schneider
SALIGN	S	-	M.S. Madhusudhan et al.
SSM (PDBeFOLD)	S	-	E. Krissinel
STRAP	S	-	C. Gille
SuperPose	S	-	Malti et al.
TOPS++FATCAT	S	-	M. Veeramalai et al.
VAST	S	-	S. Bryant
3D-Blast	NA	-	L. Mavridis et. al.
3DCOMB	NA	-	S. Wang and J. Xu
ALADYN	NA	-	Potestio et al.
CAALIGN	NA	-	T.J. Oldfield
C-BOP	NA	-	E. Sandelin
CE-MC	NA	-	C. Guda
CURVE	NA	-	D. Zhi
deconSTRUCT	NA	-	ZH. Zhang et al.
DEDAL	NA	-	P. Daniluk and B. Lesyng
DEJAVU	NA	-	GJ. Kleywegt
EpitopeMatch	NA	-	S. Jakushev
FASE	NA	-	J. Vesterstrom and W. R. Taylor
FLASH	NA	-	E.S.C. Shih and M-J Hwang
GANGSTA	NA	-	B. Kolbeck
LOCK	NA	-	AP. Singh
MALECON	NA	-	S. Wodak
Matchprot	NA	-	S. Bhattacharya et al.
Matras	NA	-	K. Nishikawa
MAX-PAIRS	NA	-	A. Poleksic
MIRAGE-align	NA	-	K. Hung et. al.
MolCom	NA	-	S.D. O'Hearn
MolLoc	NA	-	M.E. Bock et al.

PRIDE	NA	-	S. Pongor
SABERTOOTH	NA	-	F. Teichert et al.
SARF2	NA	-	N. Alexandrov
Smolign	NA	-	H. Sun
SSGS	NA	-	G. Wainreb et al.
STON	NA	-	C. Eslahchi et al.
SWAPSC	NA	-	Mario A. Fares
TABLEAUsearch	NA	-	A.S. Konagurthu et al.
TALI F	NA	-	X. Mioa
TetraDA	NA	-	J. Roach
TOPOFIT	NA	-	VA. Ilyin
TOPS+ COMPARISON	NA	-	M. Veeramalai and D. Gilbert
TS-AMIR	NA	-	J. Razmara et. al.

Legend: D (download), S (Server-based), D/S (both download/server-based), F (produces Fasta output file, FNW (produces fasta file output but with problems in the output), NF (works but does not produce Fasta output file), NW (does not works – installation or other problems), NA (download or server links no more actives)

License: “all rights reserved”