
PhD Programme in Life Sciences



SAPIENZA
Università di Roma

PhD Programme in Life Sciences

XXXV Cycle

Application of Computer Aided Drug Design techniques for the identification
of bioactive compounds

Candidate
Marianna Bufano

Supervisor
Prof. Romano Silvestri

Tutor
Prof. Antonio Coluccia

Coordinator
Prof. Francesca Cutruzzolà

To all my best “docked” people

Abstract

Molecular modeling is used in drug discovery in order to estimate molecular properties, to plan experiments and predict their outcomes, thus, to make decisions about them. Indeed, computational approaches are now key components in the drug discovery process and can help to speed up the release of new and improved active molecules.

The application of molecular modeling techniques in the early stage of drug discovery is one of the topics of this thesis. Here I present some works concerning virtual screening studies to identify compounds able to modulate some proteins involved in certain diseases.

AN-465-J137-985 shows an inhibiting binding capability and reduce the affinity of the C-SH3 domain for Gab2 (in s A549 and H1299 lung cancer cell line.)

(S)-RS4690 proved promising as new therapeutic agent against WNT-dependent colon cancer selectively binding to DVL1 PDZ (with an EC_{50} of $0.49 \pm 0.11 \mu\text{M}$ and the growth of HCT116 cells that did not present the APC mutation with an EC_{50} value $7.1 \pm 0.6 \mu\text{M}$.)

RS6212 was found to have potent anticancer activity inhibiting lactate dehydrogenase in the micromolar range in several cancer cell lines, (such as the aerobic glycolysis-dependent Med1-MB cell line, the CRC HCT116 and SW620, the lung cancer A549, and the pancreatic PANC-1 cancer cells.)

All these inhibitors deserve to be further investigated as a starting point for the development of novel anticancer strategies.

Moreover, another computational technique used in drug discovery is molecular dynamics. It helps the knowledge about the stability of a ligand-protein complex and allowed us to simulate the conformational change after binding.

Co-solvent molecular dynamics was useful to identify a suitable binding site for AM-001, an allosteric inhibitor of EPAC1.

Accelerated molecular dynamics let us the recognition of the putative hot-spot residues involved in CCRL2-Chemerin binding.

Introduction	1
1. Ligand-protein binding	1
2. Computational approaches to predict ligand–protein binding	3
2.1. Molecular docking	4
2.1.1. Sampling	5
2.1.2. Scoring functions	8
2.2. Virtual Screening	11
2.2.1. Pharmacophore concept and types	17
3. Molecular dynamics simulations	20
3.1. Enhanced sampling technique	25
3.1.1. Enhanced sampling that does not use CVs	25
3.1.2. Enhanced sampling that uses CVs	28
4. References	31
5. Aim of the work	44
Identification of new inhibitors using Virtual Screening techniques	46
1. Grb2 protein	50
1.1. Materials and methods	53
1.1.1. Molecular modeling studies	53
1.2. Identify binding inhibitors by virtual screening	53
1.3. In vitro test of the inhibitors	57
1.4. In cellula test of AN-465-J137-985	62

1.5.	Conclusions	64
1.6.	References	66
2.	DVL1	71
2.1.	Materials and methods	73
2.2.	Virtual screening	75
2.3.	HPLC separations of the enantiomers	78
2.4.	Binding mode and molecular dynamics simulation	78
2.5.	Binding assays	80
2.6.	In vitro assays	81
2.7.	ROS production	84
2.8.	DVL1 binding features	85
2.9.	Conclusions	86
2.10.	References	88
3.	LDHA	95
3.1.	Materials and methods	101
3.1.1.	Molecular modeling	101
3.1.2.	Molecular dynamics	101
3.2.	Virtual screening studies	102
3.3.	In vitro antitumor activity of the identified LDHA inhibitors	105
3.4.	Enzymatic assay	108
3.5.	Additional cancer cells	110
3.6.	Effect on HCT116 CRC cells lacking LDHA	110
3.7.	Effect on apoptosis and autophagy	110

3.8.	Combined effect of compound 18 and complex I inhibition	113
3.9.	Computational studies	114
3.9.1.	Binding mode of compound 18	114
3.9.2.	Molecular dynamics simulations	117
3.9.3.	Drug-like properties	119
3.10.	Conclusions	120
3.11.	References	121
Application of molecular dynamics technique to the study of ligand-protein binding and protein-protein binding		132
1. Exploring Epac1 interactions with the allosteric inhibitor AM-001		132
1.1.	Materials and methods	139
1.1.1.	Epac structures	139
1.1.2.	Molecular dynamics	139
1.1.3.	PCA	140
1.1.4.	Molecular docking	141
1.2.	Analyses of the active conformation	141
1.3.	Occupancy maps analyses of the inactive conformation	145
1.4.	Conclusions	159
1.5.	References	160

2. Exploring CCRL2-Chemerin binding	171
2.1. Materials and methods	173
2.1.1. Structural comparison of modeled proteins	173
2.1.2. Protein-protein docking	174
2.1.3. Accelerated molecular dynamics	174
2.1.4. Trajectories analyses	177
2.2. Modeling CCRL2 and Chemerin	177
2.3. CCRL2-Chemerin protein-protein docking	180
2.4. Selection of CCRL2-chemerin binding models	182
2.5. Proposed interaction models for CCRL2-chemerin binding	187
2.6. Conclusions	194
2.7. References	195
3. Laboratory equipment	202
4. Conclusions and future perspectives	203
5. Figures index	206
6. Tables index	209

Introduction

1. Ligand-protein binding

The effects produced by a molecule arise from its specific and reversible interaction with a receptor. Such interaction can modify the biochemical processes in which the receptor participates, stimulating or inhibiting them.

The driving forces of the association between protein and ligands are the result of various interactions and energy exchanges among the protein, ligand, water, and ions. Gibbs free energy is the most important thermodynamic quantity for the characterization of the driving forces; the protein-ligand binding occurs spontaneously only when the system free energy is negative. Moreover, the amount of the difference in free energy between the complex state and the unbound free state (*i.e.*, the magnitude of the negative free energy change upon binding) determines the stability of the complex. The decrease in system free energy can be considered the force that drives the protein–ligand binding⁽¹⁾.

The concept of receptors was introduced in 1900 by Paul Ehrlich, an immunologist who speculated that side chains on cells of the immune system combine with bacterial toxins in order to turn into antibodies. Later, these side chains were renamed receptors⁽²⁾. Since those times, this term was used to indicate any macromolecule targeted by a ligand. The specificity of the ligand-receptor link at the basis of the efficacy and manageability of most drugs was soon compared to the substrate-enzyme specificity, which is described through the key-lock model by Fisher. Fisher describes the interaction between enzymes and substrates as a recognition between a lock that receives its key: this can happen only when exact geometric complementarity between the substrate (key) and enzyme (lock) occurs⁽³⁾. However, this theory didn't explain the allosteric modulation or why different molecules can fit to the same receptor.

Introduction

Later, Koshland presented the induced fit theory in which the interaction also required the specific orientation of catalytic groups but, in order for that to happen, the substrate needs to provoke a change in the three-dimensional relationship of the active site residues which is beneficial for their binding⁽⁴⁾.

On the contrary, Monod, Wytman and Changeux proposed the conformational selection model where proteins can adopt multiple conformational states (active, inactive, intermediate) and the protein subunit adopts the required conformation only when the ligand is bound to it⁽⁵⁾.

It is important to stress that all three mechanisms may exist all at once or in a consequential manner, covering a wide spectrum of binding events. Indeed, Csermely et al.⁽⁶⁾ proposed a new theory that unified the induced fit and selection models, leading to an evolution of the selection model. These models were helpful for the understanding of biological mechanism. Thus, they paved the way for the beginning of new insight in drug discovery, for the development of new molecules that bind to a specific energy state of the protein (or to a site that influence the activation of the protein) and could work as inhibitors, activators or allosteric modulators.

In my thesis I will discuss several computational techniques applied in drug discovery which are useful for understanding the behavior of some proteins and the discovery of proteins' new inhibitors involved in specific diseases.

2. Computational approaches to predict ligand–protein binding

Over the time, several methods have been used to study the ligand-protein binding. Among these methods, computational methodologies have become crucial components in drug discovery process.

Computer Aided Drug Design (CADD) techniques can be used for three reasons: virtual screening, *hit/lead* optimization and design of novel compounds.

A huge database of compounds is examined via virtual screening in order to search for binding capacity to a target and a subset of compounds is selected and suggested for *in vitro* testing. The purpose is to increase the *hit* rate of new drugs by reducing the number of compounds to test experimentally. Another application of CADD is the optimization of *hit/lead* compounds driven by the rationalization of the structure-activity relationship, by merging together fragments into new chemotypes or by improving drug metabolism and pharmacokinetics properties⁽⁷⁾.

We can classify CADD methods in structure-based and ligand-based. The choice to use one or another is usually based on the availability of information about the target. Structure-based method is generally used when crystal structures are available, while ligand-based are used when few information are known about the structure.

After the individuation of key elements to bind a target, design of new compounds may be attempted.

In the following paragraphs I will convey about the main computational techniques in drug discovery and the application of those in our research.

2.1 Molecular Docking

Molecular docking consists in the prediction of ligand conformation and orientation within a targeted binding site. The molecular docking requires that the 3D structure of the macromolecule is identified. It may be a X-ray crystal structure (or NMR) or a structure obtained by computational techniques from a known structure with high identity (such as homology modeling).

Docking is featured by a sampling method and a scoring function^(8,9,10). Sampling refers to both the ligand and the protein: ligand conformational sampling is an essential step that generates a ligand multiconformer database to be used in ligand sampling; protein sampling refers to the flexibility given to the binding site during the simulation.

Secondly, scoring functions are used to estimate the binding affinities of ligand poses approximating the ligand-receptor free energy using parameters such as the number of hydrogen bonds, lipophilicity, ionic interactions or entropy penalties⁽¹⁰⁹⁾.

Through docking we aim to predict the stable drug interactions by inspecting and modelling drug molecular interactions between drug and target receptor molecules.

According to the classification of docking techniques⁽¹¹⁾, we can classify the method in:

a. Rigid docking

In this method, many possible ligand binding orientations can take place giving six degrees of freedom (three translational and three rotational) to the ligand.

This approximation can conceptually approach the lock and key theory and is mainly used for protein-protein docking, where the number of conformational degrees of freedom is too high to be sampled⁽¹¹⁾.

Generally, in these methods the binding site and the ligand are approximated by “hot” points and the superposition of matching points is evaluated⁽¹²⁾.

b. Semi-flexible docking

Another approach is to consider only the conformational space of the ligand, setting the protein as rigid. The ligand has more freedom degrees, but the semi-flexible docking assumes that the protein has the right conformation to accommodate the ligand.

c. Flexible docking

In this method the protein and the ligand are treated both as flexible. One of the simplest attempts in this case is to treat only some side chains of the protein as flexible, while backbones are kept fixed. In one of the earliest studies, Leach included side-chain flexibility using a rotamer library⁽¹³⁾. From this on, researchers have tried to develop techniques that integrate an increase in the protein flexibility including the backbone atoms. However, compared to the side-chain flexibility methods, these techniques are more challenging for the scoring function and time-consuming.

This classification can be systematically extended by describing sampling algorithms.

2.1.1 Sampling

Sampling consists of algorithms that generate ligand poses that match the space chosen for binding pocket. This step is complicated by the number of freedom degrees of the small molecule, increasing the conformational space to sample. There are three types of ligand sampling: shape matching, systematic search, and stochastic algorithms.

a. Shape matching

Ligand and receptor interaction results from the complementary between the two surfaces. The conformational searching was made to find the best steric overlap between them by minimizing the angle between the surfaces of opposing molecules. The goal of the shape algorithm is to find the best shape in the shortest time possible, so the benefit of shape matching is its computational efficiency. However, a rigid body approximation of the ligand is given, ignoring any kind of conformational movement.

b. Systematic search

All freedom degrees are given to the ligand to allow binding conformations. Each degree of freedom is associated to a set of finite values and all the values of each coordinate are explored in a combinatorial way⁽¹⁴⁾. There are three types of systematic search methods: exhaustive search, fragmentation and conformational ensemble.

In the exhaustive search, flexible-ligand docking is carried out by methodically rotating all possible rotatable bonds of the ligand in a given interval, allowing the number of combinations to expand as the number of rotatable bonds increases. Thus, to work around this problem and make it easy, geometric/chemical constrains are applied to the initial ligand poses, and then the ligand conformations are further subject to more accurate refinement/optimization procedures. Glide⁽¹⁵⁾ is an example of software that involves a stage of exhaustive search.

Fragmentation is a method in which ligand binding conformation is increased using multiple rigid fragments docked in the binding cavity. Fragments will be rotated

and linked. Examples of softwares utilizing fragmentation are FlexX⁽¹⁶⁾ and Hammerhead⁽¹⁷⁾.

The conformational ensemble usually performs a group of pre-generated ligand conformations docked into the protein and followed by a ranking of the docked poses from different docking runs according to their energy scores⁽¹⁸⁾. Example of this docking methods are offered by EUDOC⁽¹⁹⁾ or FLOG⁽²⁰⁾.

c. Stochastic search

Stochastic algorithm search ligand conformation and orientation (translation/rotation) randomly, according to probabilistic criteria.

There are four types of stochastic algorithms⁽⁸⁾: Monte Carlo (MC) methods, evolutionary algorithms (EA), Tabu search methods, and swarm optimization (SO) methods. Monte Carlo method is based on Metropolis Monte Carlo algorithm based on Boltzmann probability function:

$$P \sim \exp[-(E_l - E_0)/k_B T]$$

where E_l and E_0 are the energy score before and after the modification, k_B the Boltzmann constant, and T the temperature of the system. In this method random changing of the ligand is accepted only if there's an improving of the energy score.

This algorithm was implemented by docking software such as Autodock⁽²¹⁾, Autodock vina⁽²²⁾ or ROSETTALIGAND⁽²³⁾.

Genetic algorithms are inspired by Darwin's theory of evolution. The GA operates on a population of "individuals," each of them offers a potential resolution to the ligand-protein docking problem. Each "individual" is scored using a fitness function that takes into account both the intramolecular ligand energy and the total interaction energy between the protein and the ligand molecule⁽²⁴⁾.

The most popular type of EAs is the genetic algorithms (GAs). GOLD^(25,26), Autodock 3 & 4⁽²⁷⁾, MolDock⁽²⁸⁾, rDOCK⁽²⁹⁾ are the examples that have implemented evolution algorithms.

In Tabu algorithm the poses of the ligand are generated randomly, but we receive an output in which the ligand conformation will not be equal to others in order not to explore the same conformational space. This will be possible by measuring the rmsd of each new pose generated with respect to the previous one.

Swarm optimization (SO) took inspiration from swarm pattern in which conformations of a ligand are influenced by the best position of its group's components. For example, PLANTS⁽³⁰⁾ (Protein-Ligand ANTSsystem) is based on ant colony optimization mimicking the behavior of ants who exchange information on where to find food by leaving pheromones. Each degree of ligand freedom is metaphorically associated to a pheromone vector.

2.1.2 Scoring functions

The scoring function is correlated to each conformation. This predicts the associated energy with a physical or empirical function. Generally, this can be related with the accuracy of the algorithm⁽³¹⁾. Binding mode could be predicted by the conformation with the lowest energy score. We can assemble the scoring function into three group:

a. *Force field scoring function*

Force field (FF) scoring functions^(31,32,33) usually derives from both experimental data and *ab initio* quantum mechanical calculations according to the principles of physics.

The energy of the interaction between a macromolecule and its ligand as well as the internal energy of the ligand (such as steric strain induced by binding) are often quantified by molecular mechanics force fields.

This scoring function is composed by two energy components, the VDW term express by Lennard-Jones function and the electrostatic term express by Coulomb function.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

where r_{ij} stands for the distance between protein atom i and ligand atom j , A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic charges. The distance-dependent dielectric constant $\varepsilon(r_{ij})$ in the Coulombic term is introduced to consider the effect of the solvent. Examples of force field based scoring functions are: GoldScore⁽³⁴⁾, AutoDock⁽²⁷⁾, GBVI/WSA⁽³⁵⁾.

b. Empirical scoring functions

Empirical scoring functions consider several weighted energy terms such as van der Waals, electrostatics, hydrogen bond, desolvation, entropy, hydrophobicity in order to estimate the affinity of the complex.

$$\Delta G = \sum_i W_i \cdot \Delta G_i$$

where ΔG_i represent individual empirical energy terms, and the corresponding coefficients W_i are obtained from regression analysis using experimentally determined binding energies and X-ray structural information⁽³⁶⁻⁴⁰⁾.

The LUDI⁽⁴¹⁾ scoring function was the first example of an empirical one. Other empirical scoring functions are: GlideScore^(15,42), ChemScore⁽⁴³⁾, PLANTS_{CHEMPLP}⁽³⁰⁾.

c. Knowledge-based scoring function

This model starts from known ligand-protein interactions and it converts experimental frequencies of atom-atom interactions into energies component. For protein–ligand studies, the potentials are calculated by

$$w(r) = -k_b T \ln[\rho(r)/\rho^*(r)]$$

where k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom pair at distance r in the training set, and $\rho^*(r)$ is the pair density in a reference state where the interatomic interactions are zero.

After the potential parameters $w(r)$ are derived, the energy of ligand binding for a given complex is simply the sum of the interaction terms for all the protein-ligand atom pairs in the complex⁽⁴⁴⁾.

Example of knowledge-based scoring functions are DrugScore^(45,46) and GOLD/ASP⁽⁴⁷⁾.

d. Consensus scoring functions

This method is not really a function derived from equations, but it combines multiple scoring functions in order to minimize errors from different scoring functions⁽⁴⁸⁾.

Molecular docking can be well applied at multiple stages of the drug design and discovery process for three main purposes: (1) predicting the binding mode of a known active ligand, (2) predicting the binding affinities of analogous compounds from a known active series, (3) and identifying new ligands using virtual screening (VS).

2.2. Virtual Screening

The goal of VS is to find new bioactive compounds from huge chemical libraries by using information of the protein target (structure-based VS) or existing bioactive ligands (ligand-based VS)⁽⁴⁹⁾.

Ligand-based approach use SAR (structure activity relationship) derived from known active ligands in order to identify new candidate molecules for experimental evaluation. Instead, structure-based VS largely use the docking approach, where the target protein or receptor is used to dock the candidate molecules and rank them based on their predicted binding affinity (docking score). Ligand- and Structure-based approaches can be used alone, in a consequential way, or with both approaches simultaneously, based on the necessities of the researchers.

VS can be considered as the *mining* of chemical spaces with the aim of identifying molecules that possess a desired property⁽⁵⁰⁾.

VS is highly dependent on the quantity and the quality of the resources and there's no "official" workflow to follow; however, researchers have to apply their computational knowledge and experience to find the active drug candidate from drug databases and chemical libraries, applying the best possible source of tools. Generally, a VS workflow is featured by:

1. Selection of chemical libraries

Database libraries must fit the purpose of the experiment before its selection for screening. Many databases are publicly available, and researchers can choose to cover a particular class of compounds (structural or pharmacological) or diverse classes of molecules.

2. Preprocessing of chemical libraries

This step consists of selecting compounds from chemical libraries by removing the duplicate structures, tautomers,

charged molecules or select them in a specific range of molecular weights.

3. Filtering of drug like properties

One can also decide to select molecules for the screen relatively to drug like properties such as:

a. ADMET filter

This filter is useful for getting information about absorption, distribution, metabolism and excretion or toxicity.

It's important to consider, in the physical properties of the compound, the right compromise between its solubility and hydrophobicity. Generally, a drug needs to be soluble enough for reaching the blood stream, but also lipophilic enough for penetrating the cell membrane lipidic bilayer. These properties can be predicted *in silico* with the help of mathematical algorithms.

For example, Lipinski's rule of 5 is the most popular method for predicting the bioavailability of a compound ⁽⁵¹⁾. The rule is based on the ADME properties of known drugs, assuming that most of the orally active drugs for humans fulfill three of the following four criteria.

b. Lipinsky rule of five

It is a criteria for oral bioavailability. Generally, an orally bioavailable drug can violate at most one of these rules: (1) molecular weight

(MW) is over 500, (2) calculated octanol/water partition coefficient (logP) is over 5, (3) presence of more than 5 HBDs, and (4) presence of more than 10 HBAs.

However, Lipinski's rule of five fails to distinguish between drugs and nondrugs; it rather serves as a

method to predict compounds with poor absorption or permeability (except for antibiotics).

Therefore, compounds in the screening library that fulfill Lipinski's rule of five are more likely to be orally active and can be filtered either at early stages or at the end of the VS.

4. Screening

a. Ligand based approach

Molecules with physical and chemical properties such as those of the known ligands are identified using, for example, QSAR models or pharmacophore-based search. The ligand-based approach is possible without protein information and can be employed for scaffold hopping. This method has its roots in the principle introduced by Johnson and Maggiora⁽⁵²⁾, which states that similar compounds have similar properties. Thus, compounds with high similarity to reference compounds are likely to behave in a similar manner. So, similarity can be computed with different methods such as:

1. *Molecular fingerprint methods*

Fingerprints are sequences of bits. Each bit involves certain features of a molecule. There are various types of fingerprints: sub-structure keys-based fingerprints, topological fingerprints, circular fingerprints, or pharmacophore fingerprints. Bits allow us to compare two molecules and then quantify a similarity, using, for example, similarity metrics (tanimoto coefficient) or supervised machine learning.

2. *3D shape similarity*

The basis of 3D-shape similarity lies in the fact that two molecules with a similar shape are likely to fit in the same binding pocket and exhibit similar biological activity⁽⁵³⁾. In this approach, the

3D shape of the compounds in the library is compared to those of known active compounds, that represent the reference structure. This method does not consider the specific structure or properties of the reference ligands and only relies on the shape of the molecules.

3. *Electrostatic potential similarity*

Electrostatic interactions are important factors that need to be considered, because the ligand has to match the target electrostatic environment for the binding to occur. Thus, electrostatic potential of different molecules could be compared to that of the reference ligand and could potentially match the electrostatic potential of the target to have activity on it. Also compounds with different chemical structure can have similar electrostatic distribution and can be chosen in order to obtain structural diversity.

4. *Pharmacophore based similarity*

A set of ligands are used to design a pharmacophore which includes common features between them that are required for the binding with the target protein. The creation of the pharmacophore can be also based on a single ligand. The compounds that match the pharmacophore are predicted to achieve the same interactions with the biological target and their binding is expected to result in the same biological response.

b. Receptor-based approach:

In this approach the protein structure of interest is available, and a compound library of small molecules (available via purchase or synthesis) is explored using techniques like protein ligand docking that is the most famous one in VS⁽⁵⁴⁾. Active-site-directed SBPs

for the molecular recognition between a ligand and a target protein can be utilized to choose compounds that bind to the active sites of physiologically significant targets with known 3D structures,

Another approach used in this case is the generation of a pharmacophores structure-base that takes into account the distribution of features of the receptor, obtaining them in a way that differs from the ligand-based method. The advantage is the discrimination of conformations of compounds that don't fit the binding site using for example excluded volumes.

- c. A combination of ligand and receptor-based approach can also be used considering information from both. The two techniques are combined in a sequential or a parallel way. The sequential combination consists, in general, in the application of the fastest and less refined approaches in the first part, and the most computational expensive techniques in the second step. In the parallel combination method, selected compounds of both methods are compared either as a complementary selection (top ranked compounds from each method) or a consensual selection (compounds selected by both approaches) ⁽⁵⁵⁾.
- d. Machine learning techniques: machine learning techniques like support vector machine (SVM) and binary kernel discrimination (BKD) can be applied in a few cases of VS.

Many other theoretical approaches are in continuous elaboration and their applications are interesting for VS experiments. One of these approaches is the Relaxed Complex Scheme (RCS). The RCS uses Molecular dynamics simulations to explore an ensemble of low energy structures of the protein-ligand complex that will be used as starting point to dock small molecules libraries^(56,57). It combines the advantages of docking algorithms

with dynamic structural information provided by MD simulations, explicitly accounting for the flexibility of both the receptor and docked ligands⁽⁵⁵⁾.

5. Hit selection

This is the most difficult and subjective step. It is crucial to select compounds that are different from one another: this can be achieved by grouping the hit compounds according to their structure. To do so, one of the methods is clustering structure using, for example, hierarchical clustering, k-means clustering or HDBSCAN.

Compounds that showed the presence of a toxic group (e.g. nitro, aniline, hydantoin, alkyl halide peroxide, and carbazide), which are related with metabolism-mediated toxicity, are deleted from the result of the filter. Otherwise, compounds containing aldehydes and epoxides may be considered inappropriately electrophilic, whereas others such as thiols are redox active. Other compounds may be autofluorescent and others may aggregate⁽⁵⁹⁾ at certain concentrations and produce false positives in some assays. These molecules that include groups to be avoided, chemically reactive compounds, and assay-interfering compounds, are often referred to as PAINS (pan-assay interfering substances) or frequent hitters⁽⁶⁰⁾.

Once hits are selected from the final screening process, the researcher has to synthesize or buy the hits for further study. The selected hits have to go through different *in vitro/in vivo* bioassays for final confirmation of their pharmacological actions. Compounds showing hopeful pharmacological activity are considered as the leading ones for further preclinical and clinical studies to establish them as the final drug candidates.

2.2.1 Pharmacophore concept and types

In this section I will discuss about the concept of pharmacophore, which is a crucial part of the virtual screenings carried out during my years of PhD.

Introducing the term pharmacophore in the year 1909, Ehrlich⁽⁶¹⁾, named the “father of drug discovery,” defined it as “*a molecular framework that carries (phoros) the essential features responsible for a drug’s (pharmacon) biological activity.*” Although the first definition of the term was credited to Ehrlich, it was Kier who introduced the physical chemical concept in the late 1960s and early 1970s when describing common molecular features of ligands of important central nervous system receptors. This was labeled as “muscarinic pharmacophore” by Kier^(62,63). In the past, pharmacophore models were mainly worked out manually, assisted using simple interactive molecular graphics visualization programs. Later, the growing complexities of molecular structures required refined computer programs for the determination and use of pharmacophore models. In the evolution of computational chemistry, the initial definition of a pharmacophore model, conceived as an essential geometric representation of the key molecular interactions, remains unchanged. With the advances in computational chemistry, various automated tools for pharmacophore modeling and applications were developed⁽⁶⁴⁾.

There are two ways to generate a pharmacophore model. The first method is ligand-based modeling, where a set of active molecules are superimposed and common chemical features which are necessary for their bioactivity are extracted; the second is structure-based modeling performed by searching potential interaction points between the macromolecular target and ligands.

- A. Ligand-based pharmacophore model: if macromolecule structure is not available this model can be one of the solutions to screen compounds. The ligand-based pharmacophore model usually starts by

the extraction of common chemical features from the structures of a known set of ligands, that are representative of fundamental interactions between the ligands and a specific macromolecular target.

Firstly, a set of active ligands will represent a training set; they are docked to generate multiple conformation of ligands and exemplify their flexibility. This conformational analysis can also be done in the pharmacophore modeling process (generating conformers during the screening). Multiple ligands of the test set will be aligned to the training set; then, essential common chemical features will be created to build the pharmacophore model. In the generation of the pharmacophore models, it is important to consider the type of ligand molecules, the size of the data set and its chemical diversity⁽⁶⁵⁾. Features are chosen based on interaction with the target (H-bond acceptor or donor, negative or positive charge, hydrophobic and aromatic features) and information from the SAR can be included. The pharmacophore must be validated to understand if it can differentiate active compounds from inactive compounds. A screening of a set of actives and a set of decoys (compounds with similar structure but inactive against the target) is useful in this step.

- B. Structure-based pharmacophore model: Structure-based pharmacophore modeling is directly dependent on the 3D structures of macromolecular targets or macromolecule ligand complexes. A general protocol can involve the selection of complementary chemical features of the active site, their spatial relationships, and then the developing of a pharmacophore model assembled with chosen features⁽⁶⁶⁾.

Structure-based pharmacophore modeling can be further classified into two subclasses: macromolecule-ligand complex based and macromolecule (without ligand) based. The first is based on the identification of key interaction between ligand and target protein (for example, LigandScout⁽⁶⁷⁾ or Phase in Maestro Schrodinger Suites^(68,69) are the most famous programs that incorporate this scheme).

The main limitation of this process is the requirement for the 3D structure of the macromolecule-ligand complex. Consequently, it cannot be applied when no ligands targeting the binding site of interest are known. This can be solved by the macromolecule-based approach.

The identification of too many chemical features, or, on the contrary, the selection of few features could inconveniently occur when using structure-based pharmacophore modeling. A pharmacophore model consisting of too many chemical features (for example, more than eight) is not suitable for practical applications. Instead, the identification of few features could lead to the screening of too many compounds and therefore to insufficient structural selectivity. Thus, I think it is always beneficial to pick usually from five to eight features to create a reliable pharmacophore hypothesis.

3. Molecular dynamics simulations

Molecular dynamics (MD) simulations, first developed in the late 1970s⁽⁷⁰⁾, is a computational technique that reproduce the movement of macromolecule and ligands over the time. Simulations can show biological processes such us protein folding, conformational change or ligand binding using physic model that rule interatomic interactions.

The first application study on a simple system was carried out to study collisions among hard spheres⁽⁷¹⁾; later, McCammon et. al⁽⁷⁰⁾ were the firsts that published MD of a biomolecule. In particular, they reproduce the motion of a 58-residues Bovine Pancreatic Trypsin Inhibitor (BPTI) in 9.2 ps simulation carried out in vacuum with a crude molecular mechanics potential.

The movement of the system and his trajectory is generated by the integration of Newton's law motion ($F = ma$):

$$\frac{dx_i^2}{dt^2} = \frac{F_{xi}}{m_i}$$

This equation describes the acceleration of a mass m_i under a force F_{xi} in a specific direction xi .

The integration is divided into many small stages, each separated in time by fixed δt . The total force on each particle in the configuration at a time t is calculated as the vector sum of its interactions with other particles. Accelerations can be determined from the force and then they are combined with velocities at a time $t + \delta t$. The force is assumed to be constant during the time step^(72,73).

Several integration algorithms, which derive Newton's equations by a discrete time numerical approximation, are available.

The velocity-Verlet integrator is reported in the following equations as an example to compute position and velocity of an atom i at the time step $t+\delta t$, starting from the previous step, $r(t-dt)$.

$$r_i(t + \delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2$$

$$v_i(t + \delta t) = v_i(t) + \frac{1}{2}[a_i(t) + a_i(t + \delta t)]\delta t$$

where $r_i(t)$, $v_i(t)$ and $a_i(t)$ are respectively position, velocity and acceleration of atom i at time t , and $r_i(t + \delta t)$, $v_i(t + \delta t)$ and $a_i(t + \delta t)$ are respectively position, velocity and acceleration of atom i at time $t + \delta t$.

The most demanding part of molecular dynamics simulations is the calculation of the force on each particle of the system.

Acceleration is calculated from the forces acting on atom i according to Newton's second law, and forces are computed from the force field, according to the following equation:

$$a_i(t) = \frac{d^2r_i(t)}{dt^2} = \frac{F_i(t)}{m_i} = - \frac{dV(r(t))}{dr_i(t)}$$

where $V(r(t))$ is the potential energy function retrieved by the force field.

The force-field, which is the sum of the bonded (intramolecular) and nonbonded energy components, is a straightforward function that approximates the potential energy. The basic form of the function comprises, in the bonded portion, bond stretching and bending described by harmonic potential, and torsional potential described by a trigonometric function. Nonbonded terms consist of van der Waals and Coulomb electrostatic interactions between couples of atoms. A successful force field in drug design should work well both for biological molecules and the organic molecules that interact with them⁽⁷³⁾.

The most used force field in molecular dynamics are CHARMM⁽⁷⁴⁾, AMBER^(75,76), OPLS⁽⁷⁷⁾ and GROMOS⁽⁷⁸⁾.

As an example, the components of AMBER force field are reported⁽⁷⁶⁾:

$$\begin{aligned}
E_{total} = & \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \\
& + \sum_{dihedrals} \frac{V_n}{2} [+ \cos(n\phi - \gamma)] \\
& + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]
\end{aligned}$$

Where the first term represents the energy between covalently bonded atoms; the second term represents the energy due to the geometry of electron orbitals involved in covalent bonding; the third term represents the energy for twisting a bond due to bond order (e.g., double bonds) and neighboring bonds or lone pairs of electrons; and the fourth term represents the non-bonded energy between all atom pairs.

MD simulate an isolated system by solving the Newton's equation in a system featured by constant temperature and pressure, thus energy, and number of particles. Constant temperature is conserved by thermostat⁽⁷⁹⁾ which allows for changes in kinetic energy and also pressure is maintained with barostat algorithms through the regulations of the volume⁽⁸⁰⁾.

Periodic boundary condition is used to simulate a finite system and create a cell of a certain size, where coordinates and speeds are repeated in space directions⁽⁸⁰⁾ in a periodic way. Thus, a spherical cutoff with a radius of at least 10 Å can be used and it is employed in the calculation of the short-ranged van der Waals terms. Instead, the calculation of the long-range electrostatic interactions is extended to all the periodic cells; usually, the Ewald sum methods is used for this calculation, but to reduce computational cost Particle Mesh Ewald (PME) could be used⁽⁸¹⁾ to take advantage of the Fast Fourier transform⁽⁸²⁾.

Molecular dynamics simulations are useful when we have to consider protein motion. Indeed, in static models (like X-RAY crystallography or NMR) we can obtain much information about

the macromolecule structure, but, when a small molecule binds to its target in solution, it faces to a macromolecule in constant motion. Upon binding, the ligand may induce conformational changes that are not always sampled when the ligand is absent⁽⁸³⁾. Information on the binding mode of a protein and ligand can be obtained by molecular docking; however, the predicted position and orientation of ligands, and the observation of some interactions also depends on additional optimization of the complementarity between protein and ligand. The use of MD simulations can not only optimize the local steric clashes between protein and ligand but also correct and optimize the initial binding mode of the ligand during molecular docking. So, MD trajectories can be considered as sampling engines that reproduce protein conformations usable for Multiple Protein Conformations docking applications. For example, McCammon et al. developed the so called Relaxed-Complex Scheme (RCS): mini-libraries of compounds are docked by AutoDock⁽²⁷⁾ against a large ensemble of snapshots derived from unliganded protein MD trajectories. This approach is founded on the conformational selection binding model^(84,85,57).

Alternative conformers obtained through MD can also give insights into cryptic or allosteric binding sites⁽⁸⁶⁾. For example, Schame et al. identified an alternative binding site named “trench”, close to the active site of the HIV-1 integrase⁽⁸⁷⁾.

MD has further applications as a docking-coupled technique. In particular, it can be used for the assessment of the stability and for the refinement of docking poses: incorrect poses are likely to be unstable and dissociate from the complex, while realistic ones will be stable⁽⁸⁸⁾.

Simulations in explicit solvent may allow to estimate the contribution of water during binding.

Additionally, by using MD, we can examine the entire binding process, which may shed light on the metastable states that the ligand reached during the simulation, on alternative binding sites, on the function of water during binding, and on conformational rearrangements that come before, during or after binding.

However, the observation of a binding event during a classical MD simulation is very rare and a high number of MD steps is required to simulate slow processes, such as large domain motions and binding (μs - ms). These processes, in particular, demand the overcoming of a high energy barrier, which corresponds to sparsely populated states in the conformational energy landscape; in this scenario, the simulated system becomes trapped in a local minimum, making classical MD insufficient to extensively explore the conformational space⁽⁸⁹⁾.

3.1. Enhanced sampling technique

These methods add a bias force/potential to the system to let it overcome limitations of classical MD and escape from local minima, implying an acceleration of conformational sampling. These methods have been conceived primarily to study folding, binding or unbinding processes, sharing the principal idea of sampling enhancement and overcoming high energy barriers. Enhanced sampling techniques can be divided into methods that make use of collective variables to introduce the bias and methods that do not. The employment of a collective variable (CV) is based on the idea that a complex system can be decomposed into one or a combination of reaction coordinates describing the process of interest⁽⁹⁰⁾.

3.1.1. Enhanced sampling that does not use CVs

a. Replica exchange molecular dynamics

Replica exchange is also known as Parallel tempering (PT) and its first application was in 1999⁽⁹¹⁾. This method is featured by the replication of parallel simulations that starts with the same configuration but carried out with different temperatures. At defined time intervals multiple copies with similar T are exchanged in temperature with a probability determined by the energy (E) and temperature (T) of the system. The exchanges are calculated based on Metropolis criterion:

$$P(T_1 \rightarrow T_2) = \begin{cases} 1 & \text{for } [\beta_2 - \beta_1](E_1 - E_2) \leq 0 \\ e^{-[\beta_2 - \beta_1](E_1 - E_2)} & \text{for } [\beta_2 - \beta_1](E_1 - E_2) > 0 \end{cases}$$

where $\beta = 1/k_B T$ (with k_B the Boltzmann constant).

Temperatures are updated by rescaling the velocities of the parent simulations ($v1$ and $v2$ to $v1'$ and $v2'$) according to the following equation:

$$\begin{cases} v'_1 = \sqrt{\frac{T_2}{T_1}} v_1 \\ v'_2 = \sqrt{\frac{T_1}{T_2}} v_2 \end{cases}$$

It's important to select the range of temperatures and various strategies have been proposed to guide the choice⁽⁹²⁾.

The exchange of temperature or Hamiltonian⁽⁹³⁾ can lead the system to the convergence of MD simulation, thus REMD has been extensively adopted in various ensemble simulations, both to study protein folding and ligand binding⁽⁹⁴⁾.

b. Accelerated molecular dynamics

Accelerated MD (aMD) is a simulation method that doesn't require the selection of reaction coordinate or CVs; thus, aMD can be useful for investigating the biomolecular conformational space without a priori knowledge or restraints. It was proposed by McCammon⁽⁹⁵⁾ and it is applied to the system to let it exit from the local minimum. This is possible by adding a bias energy to the potential energy function according to the set of variables (a threshold energy (E) and a boosting constant (α)) to reduce the potential energy barrier and thus accelerate transitions between different low-energy states. This allows aMD to sample distinctive biomolecular conformations and uncommon barrier-crossing events that are not obtainable in cMD simulations.

In particular, when the potential energy ($V(r)$) is lower than a certain cutoff (E), the bias is added giving a modified potential ($V^*(r)=V(r)+ \Delta V(r)$); otherwise, the simulation continues in the true-unbiased potential ($V^*(r)=V(r)$).

The bias function is reported as follows:

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))}$$

where E is the potential energy cutoff and α is a tuning parameter determining the depth of the modified potential energy basin.

The choice of E and α is crucial since they control how quickly the molecular dynamics will accelerate. One guideline for selecting E is that it should be higher than the minimum of $V(r)$, V_{\min} , close to the starting structure. If E is smaller than V_{\min} , the simulation will always be conducted using the true potential and will therefore be a classic MD simulation. At higher values of E α has to be set higher than 0, in order to retain the basic shape of the potential energy surface, indeed when $\alpha=0$ the system goes through a random walk as most of the changed potential becomes isoenergetic. When $\alpha = E - V_{\min}$, this will allow to maintain the primary shape of the landscape⁽⁹⁵⁾.

There are two type of boosts, the “dihedral-boost”⁽⁹⁵⁾ and “dual-boost”⁽⁹⁶⁾.

In dihedral-boost aMD, boost potential is applied to all dihedrals in the system with input parameters (E_{dihed} , α_{dihed}). In dual-boost aMD, a total boost potential is applied to all atoms in the system in addition to the dihedral boost, *i.e.*, (E_{dihed} , α_{dihed} ; E_{total} , α_{total}):

$$\begin{aligned} E_{\text{dihed}} &= V_{\text{dihed_avg}} + \alpha_1 \times N_{\text{res}}, & \alpha_1 &= \alpha_2 \times \frac{N_{\text{res}}}{5} \\ E_{\text{total}} &= V_{\text{total_avg}} + b_1 \times N_{\text{atoms}}, & \alpha_1 &= b_2 \times N_{\text{atoms}} \end{aligned}$$

where N_{res} is the number of protein residues, N_{atoms} is the total number of atoms, and $V_{\text{dihed_avg}}$ and $V_{\text{total_avg}}$ are the average dihedral and total potential energies calculated from short cMD simulations, respectively.

Originally, aMD was only applied to torsional angles⁽⁸³⁾ but was subsequently extended to all force field terms including explicit solvent⁽⁹⁷⁾.

3.1.2. Enhanced sampling that uses CVs

a. Steered molecular dynamics (SMD)

The SMD method was introduced in 1997 by Klaus Schulten⁽⁹⁸⁾. In this technique an external force is applied to drive out the ligand from the binding site to investigate how the molecule reacts to mechanical stress and to clarify structure–function relationship of a macromolecular complex involving either protein–ligand or protein–protein interactions. The added force is assumed to be related to the binding strength and it is correlated with conformation changes of proteins or molecules in the system under investigation⁽⁹⁹⁾. Usually, this force is applied to only one term of the molecule, instead the other terminus is free to move; it can be fixed or it can change during the simulation. Other possibilities involve constant forces or application of forces on different CVs, such as nonlinear coordinates that can help to explore conformational rearrangement of protein domains⁽⁹⁸⁾. It is important to choose the direction of the force because the ligand might crash with “obstacle” during the way out from the protein. Moreover, integration with the targeted molecular dynamics (TMD) are reported: in TMD a bias force is applied to conduct the system from an initial to a desired final configuration⁽¹⁰⁰⁾, leading to the individuation of a path that can be used as set of directions for a SMD simulation⁽⁹⁸⁾.

b. Random acceleration molecular dynamics (RAMD)

It is an extension of SMD and it was developed to study the exit of a ligand from its target binding site. The difference between them consists in the application of a randomly oriented force to drive out the ligand; in this case, we don't choose the direction of the force and, when the ligand meets an “obstacle” during the exit pathway, the escape direction is changed automatically. In particular, the direction of the force is chosen stochastically and maintained for a set number of MD steps. If during this time interval the average velocity of the ligand is lower than a specified cut-off (or, in other terms, if the distance covered by the ligand is

lower than a cutoff distance, meaning that probably a rigid “obstacle” has been met), a new force direction is assigned to allow the ligand to search for alternative exit pathways. This process is repeated until the ligand displacement exceeds a predefined distance from its starting position, then it is considered to have been dissociated from the macromolecule⁽¹⁰¹⁾.

c. Umbrella sampling

In umbrella sampling⁽¹⁰²⁾, the energy potential is modified to allow for an easier transition over the energy barrier. The additional energy term is applied along one or a combination of CVs. This is carried out in separate windows that overlap with a CV reference value (ζ_{ref}). Bias potentials should be chosen to ensure that sampling along the reaction coordinate is as uniform as possible, and this is usually a harmonic biasing potential. The strength of the bias must be high enough to allow the energy barriers to be crossed, but sufficiently low to let system distributions of different windows overlap. The sampling in each individual window can be performed using conventional molecular dynamics (cMD) or employing enhanced sampling techniques such as Hamiltonian replica exchange⁽¹⁰³⁾. The post-processing analyses is carried out using techniques such as the weighted histogram analysis method WHAM⁽¹⁰⁴⁾.

d. Metadynamics

Metadynamics is another technique designed to accelerate rare events and reconstruct the free energy profile. Here, the bias potential is introduced in the form of a Gaussian function to one or more collective variable in order to avoid that the system revisit configurations already sampled⁽¹⁰⁵⁾. Indeed, bias-Gaussians are deposited in the CV space with a given frequency, and at each iteration the bias is the sum of the already deposited Gaussians. During the simulation, the system, instead of being trapped in the bottom of a “well”, is moved away from the hill of deposited Gaussians, and enters a new minimum. The process continues until all the minima are counterweighed by the bias potential⁽¹⁰⁶⁾.

Metadynamics can be used to explore binding/unbinding processes⁽¹⁰⁷⁾ and, with the application of funnel metadynamics⁽¹⁰⁸⁾, can also be used for the estimation of binding free energy.

4. References

1. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct.* 2007;36:21-42. doi: 10.1146/annurev.biophys.36.040306.132550. PMID: 17201676.
2. Prüll CR. Part of a scientific master plan? Paul Ehrlich and the origins of his receptor concept. *Med Hist.* 2003 Jul;47(3):332-56. PMID: 12905918; PMCID: PMC1044632.
3. Fischer E. Einfluss der configuration auf die wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* 1894;27:2984–2993
4. Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A.* 1958 Feb;44(2):98-104. doi: 10.1073/pnas.44.2.98. PMID: 16590179; PMCID: PMC335371.
5. Monod J, Wyman J, Changeux JP. On the nature of allosteric transitions: a plausible model. *J Mol Biol.* 1965 May;12:88-118. doi: 10.1016/s0022-2836(65)80285-6. PMID: 14343300.
6. Csermely P, Palotai R, Nussinov R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci.* 2010 Oct;35(10):539-46. doi: 10.1016/j.tibs.2010.04.009. Epub 2010 Jun 11. PMID: 20541943; PMCID: PMC3018770.
7. Sliwoski G, Kothiwale S, Meiler J, Lowe EW Jr. Computational methods in drug discovery. *Pharmacol Rev.* 2013 Dec 31;66(1):334-95. doi: 10.1124/pr.112.007336. PMID: 24381236; PMCID: PMC3880464.
8. Huang SY, Zou X. Advances and challenges in protein-ligand docking. *Int J Mol Sci.* 2010 Aug 18;11(8):3016-34. doi: 10.3390/ijms11083016. PMID: 21152288; PMCID: PMC2996748.
9. Abagyan R, Totrov M. High-throughput docking for lead generation. *Curr Opin Chem Biol.* 2001 Aug;5(4):375-82. doi: 10.1016/s1367-5931(00)00217-9. PMID: 11470599.
10. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov.* 2004 Nov;3(11):935-49. doi: 10.1038/nrd1549. PMID: 15520816.

11. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins*. 2002 Jun 1;47(4):409-43. doi: 10.1002/prot.10115. PMID: 12001221.
12. Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. *J Comput Aided Mol Des*. 2002 Mar;16(3):151-66. doi: 10.1023/a:1020155510718. PMID: 12363215.
13. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol*. 1994 Jan 7;235(1):345-56. doi: 10.1016/s0022-2836(05)80038-5. PMID: 8289255.
14. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct*. 2003;32:335-73. doi: 10.1146/annurev.biophys.32.110601.142532. Epub 2003 Jan 28. PMID: 12574069.
15. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem*. 2004 Mar 25;47(7):1750-9. doi: 10.1021/jm030644s. PMID: 15027866.
16. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol*. 1996 Aug 23;261(3):470-89. doi: 10.1006/jmbi.1996.0477. PMID: 8780787.
17. Welch W, Ruppert J, Jain AN. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol*. 1996 Jun;3(6):449-62. doi: 10.1016/s1074-5521(96)90093-9. PMID: 8807875.
18. Lorber DM, Shoichet BK. Hierarchical docking of databases of multiple ligand conformations. *Curr Top Med Chem*. 2005;5(8):739-49. doi: 10.2174/1568026054637683. PMID: 16101414; PMCID: PMC1364474.
19. Pang YP, Perola E, Xu K, Prendergast FG. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J Comput*

Chem. 2001 Nov 30;22(15):1750-1771. doi: 10.1002/jcc.1129. PMID: 12116409.

20. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des.* 1994 Apr;8(2):153-74. doi: 10.1007/BF00119865. PMID: 8064332.

21. Morris GM, Goodsell DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J Comput Aided Mol Des.* 1996 Aug;10(4):293-304. doi: 10.1007/BF00124499. PMID: 8877701.

22. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010 Jan 30;31(2):455-61. doi: 10.1002/jcc.21334. PMID: 19499576; PMCID: PMC3041641.

23. Meiler J, Baker D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins.* 2006 Nov 15;65(3):538-48. doi: 10.1002/prot.21086. PMID: 16972285.

24. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct.* 2007;36:21-42. doi: 10.1146/annurev.biophys.36.040306.132550. PMID: 17201676.

25. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* 1997 Apr 4;267(3):727-48. doi: 10.1006/jmbi.1996.0897. PMID: 9126849.

26. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J Mol Biol.* 1995 Jan 6;245(1):43-53. doi: 10.1016/s0022-2836(95)80037-9. PMID: 7823319.

27. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662

28. Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. *J Med Chem*. 2006 Jun 1;49(11):3315-21. doi: 10.1021/jm051197e. PMID: 16722650.
29. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol*. 2014 Apr 10;10(4):e1003571. doi: 10.1371/journal.pcbi.1003571. PMID: 24722481; PMCID: PMC3983074.
30. Korb O, Stützle T, Exner TE (2006) PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In: Dorigo M, Gambardella LM, Birattari M, Martinoli A, Poli R, Stützle T (eds) *Ant Colony Optimization and Swarm Intelligence*. Springer Berlin Heidelberg, pp 247–258
31. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007;36:21-42. doi: 10.1146/annurev.biophys.36.040306.132550. PMID: 17201676.
32. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP. Molecular mechanics methods for predicting protein-ligand binding. *Phys Chem Chem Phys*. 2006 Nov 28;8(44):5166-77. doi: 10.1039/b608269f. Epub 2006 Sep 1. PMID: 17203140.
33. Meng, EC; Shoichet, BK; Kuntz, ID. Automated docking with grid-based energy approach to macromolecule-ligand interactions. *J. Comput. Chem* 1992, 13, 505–524
34. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins*. 2003 Sep 1;52(4):609-23. doi: 10.1002/prot.10465. PMID: 12910460.
35. Corbeil CR, Williams CI, Labute P. Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des*. 2012 Jun;26(6):775-86. doi: 10.1007/s10822-012-9570-1. Epub 2012 May 8. PMID: 22566074; PMCID: PMC3397132.
36. Jain AN. Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des*. 1996 Oct;10(5):427-40. doi: 10.1007/BF00124474. PMID: 8951652.

37. Head, RD; Smythe, ML; Oprea, TI; Waller, CL; Green, SM; Marshall, GR. Validate a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc* 1996; 118: 3959–3969
38. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des.* 1997 Sep;11(5):425-45. doi: 10.1023/a:1007996124545. PMID: 9385547.
39. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM, Waldman M. LigScore: a novel scoring function for predicting binding affinities. *J Mol Graph Model.* 2005 Apr;23(5):395-407. doi: 10.1016/j.jmgm.2004.11.007. Epub 2004 Dec 25. PMID: 15781182.
40. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des.* 2002 Jan;16(1):11-26. doi: 10.1023/a:1016357811882. PMID: 12197663.
41. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des.* 1994 Jun;8(3):243-56. doi: 10.1007/BF00126743. PMID: 7964925.
42. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem.* 2006 Oct 19;49(21):6177-96. doi: 10.1021/jm051256o. PMID: 17034125.
43. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des.* 1997 Sep;11(5):425-45. doi: 10.1023/a:1007996124545. PMID: 9385547.

44. Huang, S-Y; Zou, X. Mean-force scoring functions for protein-ligand binding. *Annu. Rep. Comput. Chem* 2010; 6: 281–296.
45. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol.* 2000 Jan 14;295(2):337-56. doi: 10.1006/jmbi.1999.3371. PMID: 10623530.
46. Velec HF, Gohlke H, Klebe G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem.* 2005 Oct 6;48(20):6296-303. doi: 10.1021/jm050436v. PMID: 16190756.
47. Mooij WT, Verdonk ML. General and targeted statistical potentials for protein-ligand interactions. *Proteins.* 2005 Nov 1;61(2):272-87. doi: 10.1002/prot.20588. PMID: 16106379.
48. Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem.* 1999 Dec 16;42(25):5100-9. doi: 10.1021/jm990352k. PMID: 10602695.
49. Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr Med Chem.* 2004 Jan;11(1):71-90. doi: 10.2174/0929867043456287. PMID: 14754427.
50. Fox S, Farr-Jones S, Yund MA. High Throughput Screening for Drug Discovery: Continually Transitioning into New Technology. *J Biomol Screen.* 1999;4(4):183-186. doi: 10.1177/108705719900400405. PMID: 10838437.
51. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2001 Mar 1;46(1-3):3-26. doi: 10.1016/s0169-409x(00)00129-0. PMID: 11259830.
52. Johnson, M.A.; Maggiora, G.M. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons: New York, NY, USA, 1990

53. Kumar A, Zhang KYJ. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Front Chem*. 2018 Jul 25;6:315. doi: 10.3389/fchem.2018.00315. PMID: 30090808; PMCID: PMC6068280.
54. Ripphausen P, Nisius B, Peltason L, Bajorath J. Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J Med Chem*. 2010 Dec 23;53(24):8461-7. doi: 10.1021/jm101020z. Epub 2010 Oct 7. PMID: 20929257.
55. Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem*. 2013;20(23):2839-60. doi: 10.2174/09298673113209990001. PMID: 23651302.
56. Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*. 2003 Jan;68(1):47-62. doi: 10.1002/bip.10218. PMID: 12579579.
57. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des*. 2008 Sep;22(9):693-705. doi: 10.1007/s10822-007-9159-2. Epub 2008 Jan 15. PMID: 18196463; PMCID: PMC2516539.
58. Campello, R.J.G.B.; Moulavi, D.; Sander, J. *Density-Based Clustering Based on Hierarchical Density Estimates*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 160–172.
59. Seidler J, McGovern SL, Doman TN, Shoichet BK. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J Med Chem*. 2003 Oct 9;46(21):4477-86. doi: 10.1021/jm030191r. PMID: 14521410.
60. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem*. 2010 Apr 8;53(7):2719-40. doi: 10.1021/jm901137j. PMID: 20131845.
61. Ehrlich P. Ueber den jetzigen Stand der Chemotherapie. *Ber Dtsch Chem Ges* 1909;42:17 47

62. Kier LB. Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. *Mol Pharmacol.* 1967 Sep;3(5):487-94. PMID: 6052710.
63. Kier LB. *MO Theory in drug research.* New York, NY: Academic Press; 1971
64. Leach AR, Gillet VJ, Lewis RA, Taylor R. Three-dimensional pharmacophore methods in drug discovery. *J Med Chem.* 2010 Jan 28;53(2):539-58. doi: 10.1021/jm900817u. PMID: 19831387.
65. Poptodorov K, Luu T, Hoffmann RD. In: Langer T, Hoffmann RD, editors. *Methods and principles in medicinal chemistry, pharmacophores and pharmacophores searches, vol. 2.* Weinheim, Germany: Wiley-VCH; 2006.
66. Langer T. Pharmacophores in Drug Research. *Mol Inform.* 2010 Jul 12;29(6-7):470-5. doi: 10.1002/minf.201000022. Epub 2010 Jun 28. PMID: 27463325.
67. Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model.* 2005 Jan-Feb;45(1):160-9. doi: 10.1021/ci049885e. PMID: 15667141.
68. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des.* 2006 Oct-Nov;20(10-11):647-71. doi: 10.1007/s10822-006-9087-6. Epub 2006 Nov 24. PMID: 17124629.
69. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des.* 2006 May;67(5):370-2. doi: 10.1111/j.1747-0285.2006.00384.x. PMID: 16784462.
70. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature.* 1977 Jun 16;267(5612):585-90. doi: 10.1038/267585a0. PMID: 301613.
71. Alder BJ, Wainwright TE (1957) Phase transition for a hard sphere system. *J Chem Phys* 27:1208–1209

72. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev.* 2006 May;106(5):1589-615. doi: 10.1021/cr040426m. PMID: 16683746; PMCID: PMC2547409.
73. Leach AR (2001) *Molecular Modelling: Principles and Applications*, illustrated. Pearson Education
74. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* 1998 Apr 30;102(18):3586-616. doi: 10.1021/jp973084f. PMID: 24889800.
75. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004 Jul 15;25(9):1157-74. doi: 10.1002/jcc.20035. Erratum in: *J Comput Chem.* 2005 Jan 15;26(1):114. PMID: 15116359.
76. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". *J. Am. Chem. Soc.* 117 (19): 5179–5197
77. Jorgensen WL, Tirado-Rives J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc.* 1988 Mar 1;110(6):1657-66. doi: 10.1021/ja00214a001. PMID: 27557051.
78. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem.* 2004 Oct;25(13):1656-76. doi: 10.1002/jcc.20090. PMID: 15264259.
79. Hünenberger, P. Thermostat algorithms for molecular dynamics simulations. In *Advanced Computer Simulation*; Holm,

- C.; Kremer, K., Eds.; Springer: Berlin, Germany, 2005; 173:105–149, doi: 10.1007/b99427
80. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press, Inc.: San Diego, CA, 2001; p 638
81. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for ewald sums in large systems *J. Chem. Phys.* 1993, 98:10089–10092 DOI: 10.1063/1.464397
82. Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., Pedersen L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* 1995;103:8577–8593. doi: 10.1063/1.470117
83. Koshland DE Jr. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci U S A* 1958; (44):98 - 104; <http://dx.doi.org/10.1073/pnas.44.2.98>; PMID: 16590179
84. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc.* 2002 May 22;124(20):5632-3. doi: 10.1021/ja0260162. PMID: 12010024.
85. Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers.* 2003 Jan;68(1):47-62. doi: 10.1002/bip.10218. PMID: 12579579.
86. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol.* 2011 Oct 28;9:71. doi: 10.1186/1741-7007-9-71. PMID: 22035460; PMCID: PMC3203851.
87. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. Discovery of a novel binding trench in HIV integrase. *J Med Chem.* 2004 Apr 8;47(8):1879-81. doi: 10.1021/jm0341913. PMID: 15055986.
88. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev.* 2006 Sep;26(5):531-68. doi: 10.1002/med.20067. PMID: 16758486.
89. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature.* 2007 Dec 13;450(7172):964-72. doi: 10.1038/nature06522. PMID: 18075575.

90. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J Med Chem.* 2016 May 12;59(9):4035-61. doi: 10.1021/acs.jmedchem.5b01684. Epub 2016 Feb 8. PMID: 26807648.
91. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151. doi:10.1016/S0009-2614(99)01123-9
92. Patriksson A, van der Spoel D. A temperature predictor for parallel tempering simulations. *Phys Chem Chem Phys.* 2008 Apr 21;10(15):2073-7. doi: 10.1039/b716554d. Epub 2008 Feb 25. PMID: 18688361.
93. Fukunishi H, Watanabe O, Takada S On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys* 2002 116:9058
94. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J Comput Aided Mol Des.* 2013 Dec;27(12):989-1007. doi: 10.1007/s10822-013-9689-8. Epub 2013 Dec 3. PMID: 24297454; PMCID: PMC4154199.
95. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys.* 2004 Jun 22;120(24):11919-29. doi: 10.1063/1.1755656. PMID: 15268227.
96. Hamelberg D, de Oliveira CA, McCammon JA. Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. *J Chem Phys.* 2007 Oct 21;127(15):155102. doi: 10.1063/1.2789432. PMID: 17949218.
97. Voter AF. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events *Phys. Rev. Lett.* 1997 May; 78(20): 3908-391. doi: 10.1103/PhysRevLett.78.3908
98. Izrailev, S., S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers, and K. Schulten. Steered molecular dynamics. In *Computational Molecular Dynamics: Challenges, Methods and Ideas*, Lecture Notes in Computational Science and

- Engineering, Vol. 4. P. Denflhard, J. Hermans, B. Leimkuhler, A. Mark, R. Skeel, and S. Reich, editors. Springer-Verlag, 1998, Berlin. 39–65.
99. Lu H, Schulten K. Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins*. 1999 Jun 1;35(4):453-63. PMID: 10382673.
100. Schlitter J, Engels M, Krüger P, Jacoby E, Wollmer A Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T <-> R Transition in Insulin. *Mol Simul* 1993; 10:291-308. Doi: 10.1080/08927029308022170
101. Kokh DB, Doser B, Richter S, Ormersbach F, Cheng X, Wade RC. A workflow for exploring ligand dissociation from a macromolecule: Efficient random acceleration molecular dynamics simulation and interaction fingerprint analysis of ligand trajectories. *J Chem Phys*. 2020 Sep 28;153(12):125102. doi: 10.1063/5.0019088. PMID: 33003755.
102. Torrie GM, Valleau JP Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J Comput Phys* 1977; 23:187–199
103. Sugita Y., Kitao A., Okamoto Y. Multidimensional replica-exchange method for free-energy calculations. *J Chem Phys* 2000;113:6042–6051
104. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 1992; 13:1011–1021 doi: 10.1002/jcc.540130812
105. Laio A, Gervasio FL Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* 2008; 71(12):6601. Doi: 10.1088/0034-4885/71/12/126601
106. Barducci A, Bonomi M, Parrinello M Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2011; 1:826–843. Doi: 10.1002/wcms.31
107. Gervasio FL, Laio A, Parrinello M. Flexible docking in solution using metadynamics. *J Am Chem Soc*. 2005 Mar 2;127(8):2600-7. doi: 10.1021/ja0445950. PMID: 15725015.

108. Limongelli V, Bonomi M, Parrinello M. Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci U S A*. 2013 Apr 16;110(16):6358-63. doi: 10.1073/pnas.1303186110. Epub 2013 Apr 3. PMID: 23553839; PMCID: PMC3631651.

109. Zhang S, Golbraikh A, Tropsha A. Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces. *J Med Chem*. 2006 May 4;49(9):2713-24. doi: 10.1021/jm050260x. PMID: 16640331; PMCID: PMC2773514.

Aim of the work

A drug development project usually begins with a pharmacological hypothesis, which states that the modulation of a certain molecular biological mechanism might be helpful in treating diseases in which it is involved. In a project like this the next step is to identify hits, for example small molecules that can affect this modulation. These hit compounds are then developed into hit series using medicinal chemistry. Series could be also optimized having regard to pharmacological and pharmacokinetic qualities. At the end, the drug candidate could be turned into new medicine through clinical development. The application of structure-based virtual screening in the hit identification stage of drug discovery is one of the subjects of this thesis.

A protocol developed in our research group that consist of 5 steps (identification of the target, validation of docking software, docking of compounds libraries, pharmacophore generation and filter, visual inspection) was applied on several targets, among which I focused my work on the identification of small molecules that block, for example, a protein-protein interaction or the overexpression of proteins overexpressed in some type of cancers. In particular, the protein-protein interaction between Gab2 and Grb2, the PDZ domain of DVL1 protein and the lactate dehydrogenase enzyme.

Another important tool used for the theoretical investigation of biological processes is molecular dynamics simulations that study a system's evolution through time and provides insight into atomic-scale processes that take place in biological systems, such as the function of protein flexibility in ligand binding.

It is a useful method to have an ideal way to obtain multiple conformations of macromolecular targets. Indeed, molecular dynamics simulations were used to probe the stability of complexes between some proteins under study and new compounds selected from Virtual Screening. Also, the behavior of some scaffold proteins after ligand or protein binding could be

Aim of the work

study using implemented approach of classical molecular dynamics simulations, which let us to investigate about their possible conformational landscape.

Specifically, co-solvent molecular dynamics was useful in the identification of a suitable binding site for a known allosteric inhibitor of the EPAC1 (exchange protein activated by cAMP) protein. Furthermore, accelerated molecular dynamics was used to explore a larger conformational sampling for understanding the interaction between CCRL2 and one its ligand, chemerin, as a starting point in the discovery of new therapeutic agents.

Identification of new inhibitors using Virtual Screening techniques

In this chapter I will present the application of a validated virtual screening protocol aimed to identify new modulators of protein target, validated as crucial effectors, in the development and progression of cancer diseases.

The developed VS protocol was featured by 5 steps:

1. Identification of the target

An extensive study on biological function and essential pockets for biological activity of the protein is carried out. First, we check that the crystal structure of the protein is available, for example on protein data bank. Alternatively, we proceed with homology modeling starting from the amino acid sequence of the protein. If the target is part of a protein family, the protein showing the highest percentage of identity can be taken as reference. Even if the crystal structure is available, its quality needs to be carefully evaluated. If the target is not already studied the presence of known ligands of the same chosen pocket will help in the identification of new ones. This will be useful especially for the validation of the chosen binding site. When no ligands are found we can proceed with the recognition of essential amino acid for the activity (for example those observed from mutagenesis studies or involved in the interaction with another domain of the target or another protein).

2. Validation of the docking software

The presence of known ligands in the crystal structure will be also useful for the validation of the docking program. These will be docked in the binding site and the software that gives a binding pose with the lowest Root Mean Square Deviation (RMSD) relative to the crystal structure ligand could be used.

RMSD is the measure of the average distance between the observed heavy atom positions of the ligand and those predicted

Identification of new inhibitors using Virtual Screening technique

by the program (more specifically, the top-ranked solution from the program). This is calculated with the following formula, and it is expressed in Å:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

Where δ_i is the distance between atom i and either a reference structure or the mean position of the N equivalent atom. Generally, an RMSD lower than 3 Å is defined as an acceptable value^{1,2}.

We used many different programs like PLANTS, GOLD, Glide, Autodock 4 or Smina.

It could be that information about co-crystallized inhibitors are not available, and it was not possible to run RMSD computation. Then, in our approach, the visualization and the analysis of the main interactions between the ligand and the receptor, and the comparison with literature was used to validate the docking procedure.

3. Docking of in-house and commercial compounds libraries

At this point compounds from our in-house database (around 7000 molecules) and compounds from commercial libraries are downloaded and docked. Before this step databases are prepared deleting compounds with molecular weight lower than 200 and higher than 550 and filtered based on Lipinski rule of five. The

¹ Vieth M, Hirst JD, Kolinski A, Brooks CL. Assessing energy functions for flexible docking. *J Comput Chem.* 1998;19:1612–1622

² Bursulaya BD, Totrov M, Abagyan R, Brooks CL 3rd. Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des.* 2003 Nov;17(11):755-63

Identification of new inhibitors using Virtual Screening technique

pre-processing could also involve removal of duplicates, elimination of counter ions, filtering chemically reactive groups (e.g., electrophiles, metal chelators, Michael acceptors), undesirable atoms (e.g., organometallic complexes); and other such groups, dissociation/protonation equilibria of acids and bases, protomeric equilibria, tautomeric equilibria (or predominant tautomers). Overall, libraries are prepared and revised relative to the studied target and adapted on the electrostatic nature of the pocket.

Pharmacophore generation and filter

Based on the availability of the information about the target and/or ligands, a pharmacophore is created using a ligand-based or a structure-based method previously described in the introduction.

All docked compounds are filtered with the generated pharmacophore.

Features are chosen and a tolerance is attributed; some features are selected to be mandatory or permitted based on the importance in the studied target.

4. Visual inspection

The resulted compounds from the filter are checked over. During examination, hydrogen bonding networks that take place with binding site residues were taken into account, frequently in relation to co-crystallized ligands or other ligands with a similar chemotype. According to general principles, the environment can have a substantial impact on the strength, stability, and consequently final contribution of hydrogen bonds to binding affinity. Hydrogen bonds established with side chain atoms are typically weaker and frequently more susceptible to side chain mutations than hydrogen bonds made to the protein backbone. Also salt bridge, p-p cation and hydrophobic interaction are

Identification of new inhibitors using Virtual Screening technique

considered. In particular, molecules with solvent exposed hydrophobic groups, twisted amide or ester groups, or unfavorable cis-trans isomery are not considered. Chemical diversity, novelty, and commercial availability are usually criteria that drive the selection too.

The reported procedure was applied mainly to protein target involved in cancer diseases such as:

- 1) Grb2
- 2) DVL
- 3) LDH5A

Identification of new inhibitors using Virtual Screening technique

1) Grb2 protein

GAB2 (GRB2-associated binding protein2) is a scaffold protein featured by a variety of structural domains and docking sites that act as a basis for the assembly of signaling networks.

The association between GAB proteins and their binding partners is mediated by several structural motifs that are highly conserved, such as an N-terminal Pleckstrin homology (PH) domain, a central proline-rich region, and multiple phospho-tyrosine residues. The PH domain participates in the membrane localization of GAB2 by binding to phospholipids in cell membranes, particularly phosphatidylinositol phosphates^(1,2).

Numerous PXXP motifs found in the proline-rich region act as docking sites for proteins bearing the Src homology 3 (SH3) domain.

In lung⁽³⁾, stomach⁽⁴⁾ and breast⁽⁵⁾ malignancies, *gab2* was discovered to be overexpressed. Its significance in hematological tumors, including juvenile myelomonocytic leukemia⁽⁶⁾, chronic myelogenous leukemia (CML)⁽⁷⁾, acute leukemia⁽⁸⁾, and acute lymphoblastic leukemia⁽⁹⁾, is also undeniably established. BCR-ABL binds *Gab2*, which subsequently activates several pathways like RAS-RAF-MEK-ERK, PI3K-AKT, and JAK-STAT that are implicated in increasing leukemogenesis⁽¹⁰⁾. Therefore, *Gab2* is necessary for BCR-ABL-induced cell transformation.

GAB2 binds to the Grb2 adaptor protein through the SH3 domain using a "canonical" (PXXPXR) and "atypical" (PXXXXRXXKP) binding motif⁽¹¹⁾. GRB2 is the main upstream regulator of GAB2 and indirectly recruits it to the activated plasma membrane receptors. These include RTKs (EGFR, KIT), cytokine receptors (IL-1, IL-3, IL-15, TPO, EPO, KITL, M-CSF, Flt310, gp130), Fc receptors (FcεR1, FcγR1), T- and B-cell antigen receptors, and G-protein-coupled receptors⁽¹²⁾. GRB2 binds GAB2 via its C-terminal SH3 domain, and the complex binds to phosphorylated tyrosine residues on membrane receptors' intracellular domains⁽¹³⁾. Some receptors, such as the γ -chain of IL-2 and IL-3 receptors, lack GRB2 binding sites. In this case, the GRB2-GAB2 complex

Identification of new inhibitors using Virtual Screening technique

and tyrosine-phosphorylated receptors must be connected by a SHC protein in order to function for these signaling cascades^(14,15).

Grb2 is an adaptor protein widely expressed and required for several fundamental cellular processes, it also serves as a crucial downstream intermediary in a number of oncogenic signaling pathways. One SH2 domain is flanked by two SH3 domains in the modular structure of the mature 25 kDa Grb2 protein⁽¹⁶⁾. Grb2 is featured by three domains: SH2 domain, which is a conserved sequence of 100 amino acids, two Src homology 3 (SH3) domains, which have a conserved sequence of around 50 amino acids. Through its SH2 domain, Grb2 can interact directly with receptor tyrosine kinases (e.g., hepatocyte growth factor receptor, platelet derived growth factor receptor etc.) and non-receptor tyrosine kinases, such as focal adhesion kinase (FAK) and Bcr/Abl⁽¹⁷⁾. It can interact also to substrates of tyrosine kinases, via preferential binding to the phosphopeptide motif pYXNX (where N is asparagine and X any residue). The carboxyl and amino-terminal SH3 domains bind proline-rich regions within interacting proteins.

Grb2 is constitutively associated with Sos, a guanine-nucleotide exchange factor that promotes GDP–GTP exchange on Ras. Ras is activated along with the downstream MAPK cascade when the growth factor receptor is activated and tyrosyl phosphorylated by Grb2, which brings Sos1 close to membrane-bound Ras.⁽¹⁸⁾

Grb2 signaling has been directly linked to the etiology of a number of distinct human cancers. Chromosome 17(q22), where the human *grb2* gene is located, is known to be duplicated in leukemias and solid tumors⁽¹⁹⁾.

Disrupting the interaction between Gab2 and Grb2 that is implicated in the onset and progression of different types of cancers^(20,21,22,23) could be valuable for potential anticancer drug target.

It is tempting to hypothesize that blocking the interaction between the C-SH3 domain of Grb2 and Gab2 would be an effective chemotherapeutic approach given the role it plays in cell

Identification of new inhibitors using Virtual Screening technique

proliferation and transformation. By keeping to these assumptions, we can successfully identify and characterize a lead compound that inhibits this important protein-protein interaction. The molecule was first chosen using virtual screening and then created using chemical synthesis. As outlined below, both in vitro and in cellular investigations support the lead compound's capacity to successfully block Grb2's C-SH3 domain's ability to bind to a peptide that mimics the 503 to 524 sequences of Gab2, and they also demonstrate how well it can stop the proliferation of cancer cell lines.

Identification of new inhibitors using Virtual Screening technique

1.1) Materials and methods

Compounds AN-153-I158560 and AN-465-J137-985 were purchased from Life Chemicals Europe GmbH (Unterhaching, Germany) (Figure S1, Supporting Information). Compounds F0526-1467, F2096-1321, F5030-1061, F5139-0164 and F6599-2263 were purchased from Specs (Zoetermeer, The Netherlands).

1.1.1) Molecular modeling studies

The Grb2/Gab2 structures were downloaded from the PDB, pdb code 2W0Z and 2VWF⁽¹³⁾. Hydrogen atoms were added to the protein, using Maestro protein preparation wizard.⁽²⁴⁾ Ligand structures were built with Maestro and minimized using the MMFF94x force field until a rmsd gradient of 0.05 kcal/(mol·Å) was reached. The docking simulations were performed using PLANTS.⁽²⁵⁾ We set a binding lattice of 12 Å radius using all default settings used. The pharmacophore model was obtained by Phase⁽²⁶⁾. The polar features had a tolerance of 2 Å while the hydrophobic features had a tolerance of 2.5 Å. The commercially available compounds library (Maybridge, [www.Maybridge.com] Specs [www.SPECS.net] and Lifechemical [www.lifechemicals.com] about 1,800,000 derivatives) was firstly filtered out by the rule of five⁽²⁷⁾, then the obtained training set (about 1,400,000) was docked at the Grb2 C-SH3 domain. All the docking proposed binding conformation (10 per molecule) were filtered out by the pharmacophore model and the 50 first ranked derivatives were visual inspected. The images in the manuscript were created with PyMOL⁽²⁷⁾.

1.2) Identifying binding inhibitors by Virtual Screening

We evaluated commercially available databases (Maybridge, Specs and LifeChemicals) using docking and pharmacophore models to find useful drugs that could disrupt the interaction between C-SH3 and Gab2. To guarantee that the chosen compounds have druglike properties, the entire training set (about 2,000,000 variants) was first filtered by Lipinski Rule of Five⁽²⁸⁾

Identification of new inhibitors using Virtual Screening technique

compliance. Plants was used to dock the obtained training set to the Grb2 C-SH3 domain.

A pharmacophore model was created using the Gab2 core binding areas (Gab2a PRxxK pdb code 2W0Z and Gab2b VNRxxK pdb code 2VWF)⁽¹³⁾ to collect the pharmacophoric interactions of Gab2 with its target. In Figure 1. The fitting to the pharmacophore model was used to rank all docking poses. The best compounds were visually inspected and seven derivatives (AN-153-I158560, F0526-1467, F2096-1321, F5030-1061, F5139-0164, F6599-2263 and AN-465-J137-985) were chosen for the additional analyses.

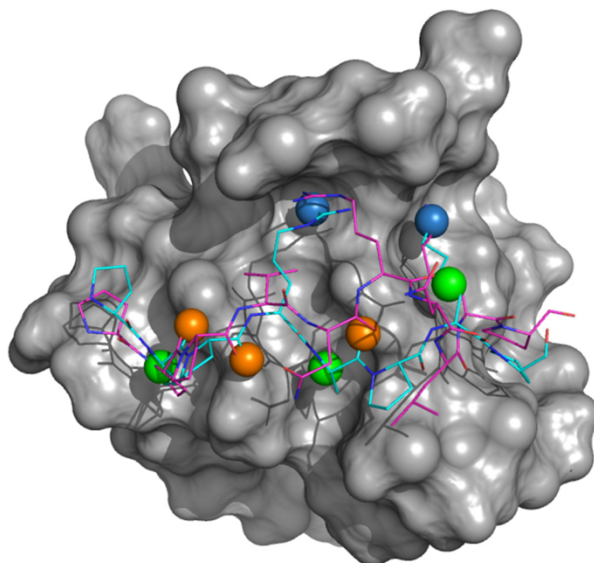
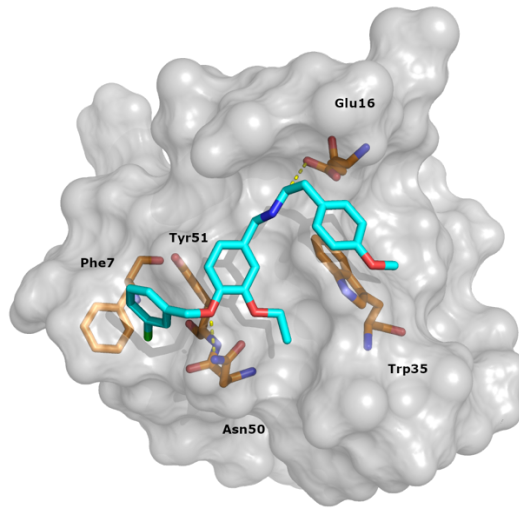


Figure 1. Pharmacophore model for Grb2 C-SH3:Gab2 complex. The Grb2 C-SH3 domain is reported as gray surface. The Gab2 core binding regions are reported as lines, Gab2a PPPRPPKP cyan and Gab2b PPVNRNLKP magenta. The pharmacophore queries are depicted as sphere: green for hydrophobic, orange for H-bond acceptor and light blue for positive charge and H-bond donor.

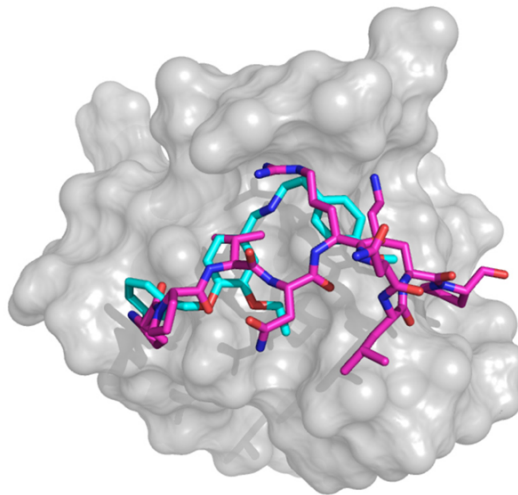
Identification of new inhibitors using Virtual Screening technique

As detailed below, out of these seven compounds, experiments confirmed an inhibiting binding capability only in the case of AN-465-J137-985. Hence, to provide additional insights and to elucidate its properties, we resorted to characterize computationally the structural featured of its binding to C-SH3. In particular, the analyses of the docking proposed AN-465-J137-985 binding mode led us to identify a series of crucial contacts with its cognate target, reported in Figure 2. We observed T-shaped aromatic interactions between the chlorobenzyl ring with Phe7 and between the central phenyl ring with Tyr51. The Trp35 and the methoxyphenyl moiety are engaged in another aromatic interaction. We also highlighted two polar contacts: one included the Glu16 carboxylic moiety and the secondary amine function, and the other was an H-bond between the C-SH3 Asn50 amide moiety and the ether oxygen atom that connected the two aromatic rings (Figure 2A). The superimposition of the AN-465-J137-985 proposed binding and Gab2a core binding showed that the secondary amine moiety was in the same region of the Arg515 side chain and was able to make similar contacts, and the three aromatic rings were superposable with Pro512 and Val513 side chains and the Arg515 backbone simulating the hydrophobic contact observed for the substrate. Lastly the ether oxygen atom between the aromatic rings is superimposable with the Pro512 amide oxygen atom doing the same polar contact (Figure 2B).

Identification of new inhibitors using Virtual Screening technique



A



B

Figure 2. AN-465-J137-985 proposed binding mode. (A) Proposed binding mode for derivative AN-465-J137-985. The Grb2 C-SH3 domain is reported as gray surface. AN-465-J137-985 is reported as cyan stick; residues involved in contacts are reported as orange. H-bond is depicted as yellow dot lines. (B) Superimposition of AN-465-J137-985 proposed binding and Gab2 core binding.

Identification of new inhibitors using Virtual Screening technique

1.3) In Vitro test of the inhibitors

They used equilibrium binding studies to determine experimentally whether the seven derivatives AN-153-I158560, F0526-1467, F2096-1321, F5030-1061, F5139-0164, F6599-2263, and AN-465-J137-985 could effectively target the interaction between the C-SH3 domain of Grb2 and Gab2. Similar to their earlier research^(29,30), they devised the procedures to test the C-SH3 domain using a peptide that mimicked a particular section of Gab2 between residues 503 and 524 (Gab2*). By taking advantage of the two naturally occurring tryptophan residues of the C-SH3 domain in positions 35 and 36, the binding reaction was monitored by following the change of the intrinsic tryptophan fluorescence emission at increasing concentration of ligand. Experiments were conducted in buffer HEPES 50 mM, 0.5 M NaCl, pH 7.0 at 25 °C in the presence of 20% v/v DMSO, to ensure the solubility of the potential inhibitory molecules. To calculate the affinity between C-SH3 Grb2 and Gab2* in these experimental conditions we conducted a control experiment by challenging a constant concentration of C-SH3 (2 μ M) versus increasing concentrations of Gab2* (ranging from 1 to 38 μ M). The dependence of the normalized fluorescence signal at 330 nm at different concentration of Gab2* is reported in Figure 3.

Identification of new inhibitors using Virtual Screening technique

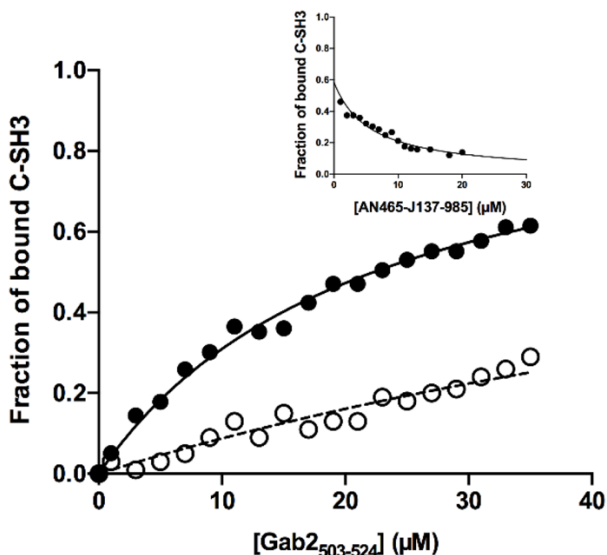


Figure 3. Equilibrium binding titration monitored by change in intrinsic fluorescence emission of C-SH3 at different concentrations of Gab2* in absence (full circles) and in presence of 5 μM AN-465-J137-985 (empty circles). Continuous and broken lines are the best fit to a hyperbolic function. Inset panel. Titration of the complex between C-SH3 and Gab2*, a constant concentration of 1 μM and 25 μM respectively, with varying concentrations of AN-465-J137-985.

Data were satisfactorily fitted with a hyperbolic equation, the calculated K_D being $22 \pm 2 \mu\text{M}$ ($X^2 = 0.9916$). then, they turned to performing equilibrium binding studies in the presence of the various compounds in solution at constant concentrations of 5 M to test the inhibitory effect of the proposed molecules. While experiments with the molecules AN-153-I158560, F0526-1467, F2096-1321, F5030-1061, F5139-0164, and F6599-2263 did not show any discernible change in the C-SH3 domain's affinity for Gab2*, an equilibrium binding titration with the molecule AN-465-J137-985 showed a definite effect on the binding reaction. The dependence of the normalized fluorescence signal at 330 nm at different concentration of Gab2* compared to the control experiment is reported in Figure 3. The calculated K_D in the presence of AN-465-J137-985 at the concentration of 5 μM was

Identification of new inhibitors using Virtual Screening technique

$100 \pm 15 \mu\text{M}$ ($X^2 = 0.9916$), which corresponds to a decrease of affinity of C-SH3 for Gab2 of ~ 5 fold as compared to the data in the absence of inhibitor. This result clearly shows an inhibitory effect of the AN-465-J137-985 molecule on the binding between C-SH3 and Gab2. It should be noted that although it is obvious that the presence of AN-465-J137-985 causes a decrease in affinity between the C-SH3 domain and Gab2*, due to the low affinity between the interacting partners, interpretation of quantitative inhibition from equilibrium experiments alone is very challenging. To estimate the inhibition constant of AN-465-J137-985, they resorted to titrate a constant concentration of a pre-incubated complex involving C-SH3 and Gab2* with varying concentrations of AN-465-J137-985 (Figure 3, inset panel). As expected, the observed fluorescence decreases with increasing concentrations of AN-465-J137-985, indicating a decrease in affinity with an apparent inhibition constant of about $5 \pm 1 \mu\text{M}$. To deduce the mechanism whereby AN-465-J137-985 affects the C-SH3:Gab2 binding reaction, they performed pseudo-first order kinetic binding experiments. By taking advantage of a stopped-flow apparatus, a solution containing a constant concentration of the C-SH3 domain at the concentration of $2 \mu\text{M}$ was rapidly mixed with an excess of Gab2* at different concentrations ranging from 0 to $12 \mu\text{M}$. All the traces recorded at different concentrations of Gab2* were satisfactorily fitted with a single exponential equation and the calculated observed rate constants k_{obs} were plotted as a function of the concentration of Gab2* (Figure 4). The analysis of the linear dependence of the k_{on} returned the value of the microscopic association rate constant $k_{on} = 40.3 \pm 0.5 \mu\text{M}^{-1} \text{ s}^{-1}$, corresponding to the slope of the fitting line.

Identification of new inhibitors using Virtual Screening technique

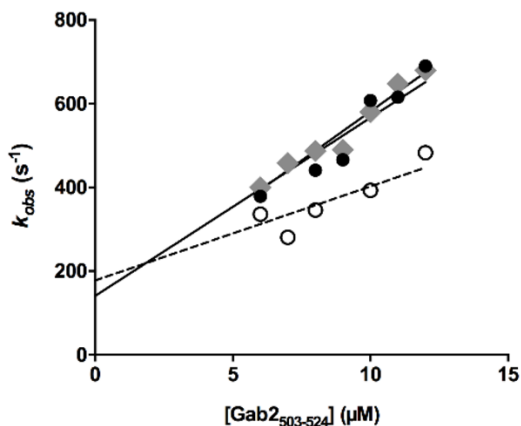


Figure 4. Observed rate constants measured through stopped-flow kinetic binding experiments by rapidly mixing a constant concentration of C-SH3 versus different concentrations of Gab2* in absence (full circles) and in presence of 5 μM AN-465-J137-985 (empty circles). The data recorded in the absence of DMSO and AN-465-J137-985 are reported in gray diamonds. Lines are the best fit to a linear equation.

The analysis of the data reported in Figure 4 allows in theory to extrapolate the microscopic dissociation rate constant by estimating the intercept on the y-axis. However, this process is typically prone to significant errors, which could result in an incorrect computation of k_{off} . Thus, they turned to measure directly k_{off} through displacement experiments⁽³¹⁾. They specifically tested a preincubated complex of C-SH3 and a dansylated variant of Gab2* at molar ratio 1:1 and at fixed concentration of 2 μM versus a high excess of Gab2* (80 μM). In agreement with theory, the fluorescence change upon displacement displayed a single exponential behavior and was found insensitive to displacer concentration, with a calculated k_{off} of $170 \pm 10 \text{ s}^{-1}$.

To test the effect of the presence of the AN-465-J137-985 on binding kinetics they repeated stopped-flow binding and displacement experiments by adding the AN-465-J137-985 molecule at the fixed concentration of 5 μM to the solution

Identification of new inhibitors using Virtual Screening technique

containing C-SH3, for binding experiments, and to the solution containing the preincubated C-SH3: Gab2* complex for displacement experiments.

The calculated observed rate constants were plotted as function of the concentrations of Gab2* (Figure 2) and fitted with a linear equation, the k_{on} being $20.6 \pm 1.5 \mu\text{M}^{-1} \text{s}^{-1}$. Microscopic dissociation rate constant k_{off} was directly measure by displacement experiment, being $190 \pm 20 \text{s}^{-1}$. Importantly, these values are consistent with what can be extrapolated from the pseudo-first order data reported in Figure 4. It is of interest to notice that while the k_{off} values in absence and in presence of AN-465-J137-985 are comparable, a decrease of k_{on} by a factor of 2 is clearly appreciable. Taken together our results clearly demonstrate that while the dissociation rate constant of Gab2 by the C-SH3 domain is mostly unaffected by the presence of the AN-465-J137-985 molecule, there is a clear effect on the recognition event between the two molecules that leads to a decrease in the binding affinity.

Identification of new inhibitors using Virtual Screening technique

1.4) In Cellula tests of AN-465-J137-985

Two lung cancer cell lines were employed to examine the effects of AN-465-J137-985 and determine whether it may have any anti-cancer properties. They treated A549 and H1299 lung cancer cell lines with increasing doses of AN-465-J137-985 and evaluated cell survival using the trypan blue exclusion assay. The data in Figure 5A indicate that AN-465-J137-985 significantly inhibits the growth of both cancer cell lines indicating a lethal dose 50 (LD50) value of about 5 and 7 μM for H1299 and A549, respectively.

Identification of new inhibitors using Virtual Screening technique

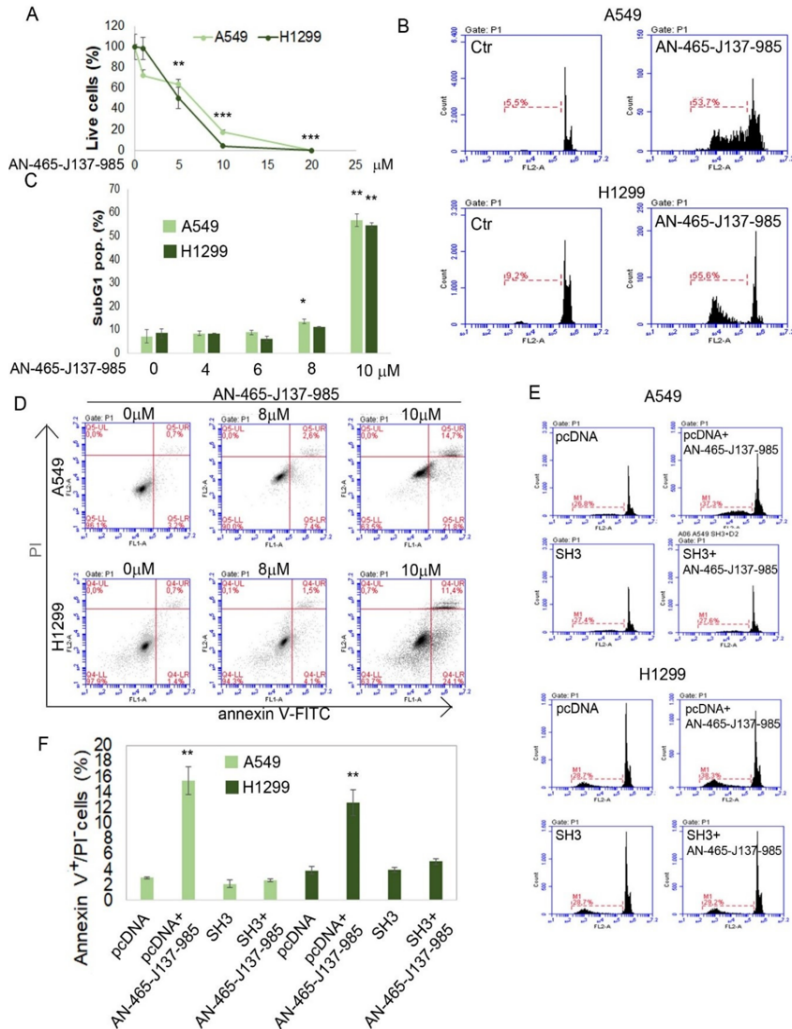


Figure 5. Effect of AN-465-J137-985 in A549 and H1299 lung cancer cells. (A) Evaluation of cell numbers after 48 h treatment with increasing concentrations of AN-465-J137-985 by trypan blue exclusion assay. (B-C) PI staining of cells after 48 h incubation with increasing concentrations of AN-465-J137-985. (B) Representative experiments of A549 and H1299 treated or not with 10 μM AN-465-J137-985. The percentage of sub-G1 population is reported in red. (C) Average \pm standard deviation of the percentage of the Sub-G1 population. (D) Annexin-V/PI staining of cells 24 h after treatment with 8 and 10 μM AN-465-J137-985. (E) PI staining of cells 72h after transfection with the plasmid encoding for C-SH3 of GRb2 or a control plasmid (PCDNA) incubated for 48 h in presence or absence of AN-465-J137-985 8 μM. (F) Annexin-V/PI staining percentage of cells 48h

Identification of new inhibitors using Virtual Screening technique

after transfection with the plasmid encoding for C-SH3 of Grb2 or a control plasmid (PCDNA) incubated for 24 h in the presence or absence of AN-465-J137-985 8 μM . * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Each experiment have been repeated in triplicate.

To better understand the mechanism that decreases cell survival they used propidium iodide (PI) and annexin-V/PI staining, in order to evaluate the possible activation of the apoptotic machinery. The cell cycle analysis reported in Figure 5B,C shows that AN-465-J137-985 at 10 μM concentration induces a strong accumulation of the cells in the sub-G1 population (around 50% for both cell lines) at 48 h after the treatment with respect to DMSO treated cells (Ctr), indicating the induction of a strong apoptotic effect.

Figure 5D shows annexin-V/PI staining of A549 and H1299 cells after the treatment with DMSO as control or with AN-465-J137-985 at 8 and 10 μM for 24 h. The drug-induced accumulation of the cells in the lower right quadrant (annexin-V positive/PI negative) confirm the hypothesis of the activation of the apoptotic mechanism.

To validate the selectivity of AN-465-J137-985, we overexpressed the C-SH3 fragment of Grb2 in both cell lines; the overexpression in the cells transfected with the plasmid was confirmed by RT-PCR (data not shown). The effect of 8 μM AN-465-J137-985 on cells transfected with the plasmid encoding for C-SH3 or a control plasmid (PCDNA) was evaluated 48 h after the transfection using PI (Figure 5E) and annexin-V/PI staining (Figure 5F) after 24 and 48 h treatment, respectively. Data in Figure 5E,F show that SH3 overexpression is able to rescue almost completely (~95%) the death effect induced in both cell lines by the treatment with AN-465-J137-985, indicating that C-SH3 fragment of Grb2 is a direct target of AN-465-J137-985.

1.5) Conclusions

The effects of the interaction between the C-SH3 domains of Grb2 and Gab2 on cell physiology, as well as the start and

Identification of new inhibitors using Virtual Screening technique

progression of cancer, are well documented^(4,5). On the basis of these findings, it is crucial to create an anticancer pharmacological strategy that aims to prevent the binding of C-SH3 and Gab2. A potent tool that can be used to find prospective compounds capable of interacting with macromolecular targets is structure-based drug design and screening. We describe here the successful identification of a compound with a notable effect on reducing the affinity of the C-SH3 for Gab2 by utilizing a synergy between computational work and in vitro and in cellula studies. We established a chemical synthesis methodology that provides the path for upcoming advancements and adjustments of the main molecule. The cellular tests on A549 and H1299, two different cancer cell lines, successfully validated the inhibitory impact discovered by the equilibrium and kinetic binding assay, confirming the molecule's anti-tumorigenic properties. While it is not yet feasible to rule out cross-reactivity with additional SH3 domains, the in cellula tests carried out using cell lines that had been transfected with SH3 seem to validate the selectivity of AN-465-J137-985. This hypothesis will be furthered by animal model-based studies in the future and by enhancing the observed KD through chemical modifications.

Coluccia A, Bufano M, La Regina G, Puxeddu M, Toto A, Paone A, Bouzidi A, Musto G, Badolati N, Orlando V, Biagioni S, Masci D, Cantatore C, Cirilli R, Cutruzzolà F, Gianni S, Stornaiuolo M, Silvestri R. Anticancer Activity of (*S*)-5-Chloro-3-((3,5-dimethylphenyl)sulfonyl)-*N*-(1-oxo-1-((pyridin-4-ylmethyl)amino)propan-2-yl)-1*H*-indole-2-carboxamide (RS4690), a New Dishevelled 1 Inhibitor. *Cancers* (Basel). 2022 Mar 7;14(5):1358. doi: 10.3390/cancers14051358.

Identification of new inhibitors using Virtual Screening technique

1.6. References

1. Maroun CR, Moscatello DK, Naujokas MA, Holgado-Madruga M, Wong AJ, Park M. A conserved inositol phospholipid binding site within the pleckstrin homology domain of the Gab1 docking protein is required for epithelial morphogenesis. *J Biol Chem.* 1999 Oct 29;274(44):31719-26. doi: 10.1074/jbc.274.44.31719. PMID: 10531383.
2. Isakoff SJ, Cardozo T, Andreev J, Li Z, Ferguson KM, Abagyan R, Lemmon MA, Aronheim A, Skolnik EY. Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast. *EMBO J.* 1998 Sep 15;17(18):5374-87. doi: 10.1093/emboj/17.18.5374. PMID: 9736615; PMCID: PMC1170863.
3. Xu XL, Wang X, Chen ZL, Jin M, Yang W, Zhao GF, Li JW. Overexpression of Grb2-associated binder 2 in human lung cancer. *Int J Biol Sci.* 2011 Apr 18;7(5):496-504. doi: 10.7150/ijbs.7.496. PMID: 21552417; PMCID: PMC3088873.
4. Lee SH, Jeong EG, Nam SW, Lee JY, Yoo NJ, Lee SH. Increased expression of Gab2, a scaffolding adaptor of the tyrosine kinase signalling, in gastric carcinomas. *Pathology.* 2007 Jun;39(3):326-9. doi: 10.1080/00313020701329773. PMID: 17558859.
5. Bentires-Alj M, Gil SG, Chan R, Wang ZC, Wang Y, Imanaka N, Harris LN, Richardson A, Neel BG, Gu H. A role for the scaffolding adapter GAB2 in breast cancer. *Nat Med.* 2006 Jan;12(1):114-21. doi: 10.1038/nm1341. Epub 2005 Dec 20. PMID: 16369543.
6. Mohi MG, Williams IR, Dearolf CR, Chan G, Kutok JL, Cohen S, Morgan K, Boulton C, Shigematsu H, Keilhack H, Akashi K, Gilliland DG, Neel BG. Prognostic, therapeutic, and mechanistic implications of a mouse model of leukemia evoked by Shp2 (PTPN11) mutations.

Identification of new inhibitors using Virtual Screening technique

- Cancer Cell. 2005 Feb;7(2):179-91. doi: 10.1016/j.ccr.2005.01.010. PMID: 15710330.
7. Gu S, Chan WW, Mohi G, Rosenbaum J, Sayad A, Lu Z, Virtanen C, Li S, Neel BG, Van Etten RA. Distinct GAB2 signaling pathways are essential for myeloid and lymphoid transformation and leukemogenesis by BCR-ABL1. *Blood*. 2016 Apr 7;127(14):1803-13. doi: 10.1182/blood-2015-06-653006. Epub 2016 Jan 15. PMID: 26773044; PMCID: PMC4825414.
 8. Zatkova A, Schoch C, Speleman F, Poppe B, Mannhalter C, Fonatsch C, Wimmer K. GAB2 is a novel target of 11q amplification in AML/MDS. *Genes Chromosomes Cancer*; 2006 Sep; 45(9):798-807. PubMed ID: 16736498
 9. Wöhrle FU, Daly RJ, Brummer T. Function, regulation and pathological roles of the Gab/DOS docking proteins. *Cell Commun Signal*. 2009 Sep 8;7:22. doi: 10.1186/1478-811X-7-22. PMID: 19737390; PMCID: PMC2747914.
 10. Chu S, Li L, Singh H, Bhatia R. BCR-tyrosine 177 plays an essential role in Ras and Akt activation and in human hematopoietic progenitor transformation in chronic myelogenous leukemia. *Cancer Res*. 2007 Jul 15;67(14):7045-53. doi: 10.1158/0008-5472.CAN-06-4312. PMID: 17638918.
 11. Lock LS, Royal I, Naujokas MA, Park M. Identification of an atypical Grb2 carboxyl-terminal SH3 domain binding site in Gab docking proteins reveals Grb2-dependent and -independent recruitment of Gab1 to receptor tyrosine kinases. *J Biol Chem*. 2000 Oct 6;275(40):31536-45.
 12. Nishida K, Hirano T. The role of Gab family scaffolding adapter proteins in the signal transduction of cytokine and growth factor receptors. *Cancer Sci*. 2003 Dec;94(12):1029-33. doi: 10.1111/j.1349-7006.2003.tb01396.x. PMID: 14662016.

Identification of new inhibitors using Virtual Screening technique

13. Harkiolaki M, Tsirka T, Lewitzky M, Simister PC, Joshi D, Bird LE, Jones EY, O'Reilly N, Feller SM. Distinct binding modes of two epitopes in Gab2 that interact with the SH3C domain of Grb2. *Structure*. 2009 Jun 10;17(6):809-22. doi: 10.1016/j.str.2009.03.017. PMID: 19523899.
14. Edmead CE, Fox BC, Stace C, Ktistakis N, Welham MJ. The pleckstrin homology domain of Gab-2 is required for optimal interleukin-3 signalsome-mediated responses. *Cell Signal*. 2006 Aug;18(8):1147-55. doi: 10.1016/j.cellsig.2005.09.002. Epub 2005 Nov 4. PMID: 16275030.
15. Leahy M, Lyons A, Krause D, O'Connor R. Impaired Shc, Ras, and MAPK activation but normal Akt activation in FL5.12 cells expressing an insulin-like growth factor I receptor mutated at tyrosines 1250 and 1251. *J Biol Chem*. 2004 Apr 30;279(18):18306-13. doi: 10.1074/jbc.M309234200. Epub 2004 Feb 12. PMID: 14963047.
16. Lowenstein EJ, Daly RJ, Batzer AG, Li W, Margolis B, Lammers R, Ullrich A, Skolnik EY, Bar-Sagi D, Schlessinger J. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell*. 1992 Aug 7;70(3):431-42. doi: 10.1016/0092-8674(92)90167-b. PMID: 1322798.
17. Dharmawardana PG, Peruzzi B, Giubellino A, Burke TR Jr, Bottaro DP. Molecular targeting of growth factor receptor-bound 2 (Grb2) as an anti-cancer strategy. *Anticancer Drugs*. 2006 Jan;17(1):13-20. doi: 10.1097/01.cad.0000185180.72604.ac. PMID: 16317285.
18. Tari AM, Lopez-Berestein G. GRB2: a pivotal protein in signal transduction. *Semin Oncol*. 2001 Oct;28(5 Suppl 16):142-7. doi: 10.1016/s0093-7754(01)90291-x. PMID: 11706405.
19. Huebner K, Kastury K, Druck T, Salcini AE, Lanfrancone L, Pelicci G, Lowenstein E, Li W, Park SH, Cannizzaro

Identification of new inhibitors using Virtual Screening technique

- L, et al. Chromosome locations of genes encoding human signal transduction adapter proteins, Nck (NCK), Shc (SHC1), and Grb2 (GRB2). *Genomics*. 1994 Jul 15;22(2):281-7. doi: 10.1006/geno.1994.1385. PMID: 7806213.
20. Simister PC, Feller SM. Order and disorder in large multi-site docking proteins of the Gab family--implications for signalling complex formation and inhibitor design strategies. *Mol Biosyst*. 2012 Jan;8(1):33-46. doi: 10.1039/c1mb05272a. Epub 2011 Sep 20. PMID: 21935523.
 21. Zhao C, Yu DH, Shen R, Feng GS. Gab2, a new pleckstrin homology domain-containing adapter protein, acts to uncouple signaling from ERK kinase to Elk-1. *J Biol Chem*. 1999 Jul 9;274(28):19649-54. doi: 10.1074/jbc.274.28.19649. PMID: 10391903.
 22. Nishida K, Wang L, Morii E, Park SJ, Narimatsu M, Itoh S, Yamasaki S, Fujishima M, Ishihara K, Hibi M, Kitamura Y, Hirano T. Requirement of Gab2 for mast cell development and KitL/c-Kit signaling. *Blood*. 2002 Mar 1;99(5):1866-9. doi: 10.1182/blood.v99.5.1866. PMID: 11861309.
 23. Adams SJ, Aydin IT, Celebi JT. GAB2--a scaffolding protein in cancer. *Mol Cancer Res*. 2012 Oct;10(10):1265-70. doi: 10.1158/1541-7786.MCR-12-0352. Epub 2012 Aug 7. PMID: 22871571; PMCID: PMC3810274.
 24. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013 Mar;27(3):221-34. doi: 10.1007/s10822-013-9644-8. Epub 2013 Apr 12. PMID: 23579614.
 25. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J*

Identification of new inhibitors using Virtual Screening technique

- Chem Inf Model. 2009 Jan;49(1):84-96. doi: 10.1021/ci800298z. PMID: 19125657.
26. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. Chem Biol Drug Des. 2006 May;67(5):370-2. doi: 10.1111/j.1747-0285.2006.00384.x. PMID: 16784462.
 27. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
 28. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001 Mar 1;46(1-3):3-26. doi: 10.1016/s0169-409x(00)00129-0. PMID: 11259830.
 29. Toto A, Bonetti D, De Simone A, Gianni S. Understanding the mechanism of binding between Gab2 and the C terminal SH3 domain from Grb2. Oncotarget. 2017 Jul 18;8(47):82344-82351. doi: 10.18632/oncotarget.19323. PMID: 29137268; PMCID: PMC5669894.
 30. Malagrino F, Troilo F, Bonetti D, Toto A, Gianni S. Mapping the allosteric network within a SH3 domain. Sci Rep. 2019 Jun 4;9(1):8279. doi: 10.1038/s41598-019-44656-8. PMID: 31164678; PMCID: PMC6547694.
 31. Antonini E, Brunori M. *Hemoglobin and Myoglobin in Their Reactions with Ligands*; North-Holland: Amsterdam, The Netherlands, 1971

Identification of new inhibitors using Virtual Screening technique

2) DVL1

The Wnt pathway has an important role in development and oncogenic processes. The first level of regulation starts with the binding of a Wnt ligand which prefers specific types of Fzd and co-receptors⁽¹⁾.

A ubiquitin-dependent proteasome degradation complex keeps β -catenin levels low in the cell in the absence of Wnts.⁽²⁾

In the "canonical" WNT β -catenin pathway, recruitment of the PDZ family member Dishevelled (DVL) at the plasma membrane occurs when morphogens WNTs bind to Frizzled Receptors (FZDs). DVL recruits the Axin–GSK3 complex to the membrane by interacting with Axin. Once stabilized and accumulated in the cell, β -catenin moves to the nucleus where it interacts with and controls the activity of the T-cell factor (TCF) and lymphoid enhancer-binding factor (LEF) DNA-binding proteins to operate as a transcriptional co-activator.^(3,4) When TCF/LEF proteins link to β -catenin, transcription of Wnt responsive genes is enabled. This promotes cell differentiation and proliferation throughout both developmental and neoplastic processes.^(5,6)

Non-canonical Wnt pathways or β -catenin independent pathways are those that do not depend on β -catenin/Tcf or β -catenin/Lef binding for modulating downstream signaling. Examples of these are Wnt/PCP pathway, Wnt/Ca²⁺ pathway or Wnt–FYN–STAT3 pathway⁽⁷⁾.

The inactivation of the APC destruction complex is possible thanks to the binding of DVL to Frizzled. DVL is a 736 amino acid long scaffold protein made of three consecutive domains, namely DIX, PDZ and DEP. The DIX domain is constituted by 80-85 amino acid and contributes to DVL oligomerization. The 80-90 amino acid long PDZ domain binds to: i) a PDZ binding site (a conserved KTXXXW signaling sequence) allocated at the C-terminal tail of the receptors; ii) a discontinuous domains spanning through the intracellular loops of the receptors. The 90-100 amino acid long DEP domain binds to the C-terminal tail of

Identification of new inhibitors using Virtual Screening technique

FZD and to its intracellular loops and it has been involved in the intracellular transduction of the WNT- β -catenin signaling⁽⁸⁾.

Thus, the WNT signal cascade, which is activated by DVL, leads to the transcription of oncogenes in pathological circumstances. The findings also indicates that RNA interference and DVL knockdown stop the growth of tumors^(9,10).

The three orthologs that belong to DVL proteins, DVL-1, DVL-2, and DVL-3, are very similar to one another in terms of sequence and folding. Despite DVL-2 isoform is the most prevalent (representing 95% of the pool) variations in its expression have little impact on WNT signaling; however, canonical signaling was more responsive to variations in DVL-1 or DVL-3 abundance⁽¹¹⁾. At the same time DVL-1 is one of the least abundant of the total DVL protein pool and play a crucial role in stimulating phosphorylation and activation of the WNT transcriptional activity of the LDL receptor-related protein 6 (LRP6), an ortholog member of the LRP family^(12,13). Additionally, the DVL-1 and DVL-3 proteins are crucial for cancer chemoresistance. Indeed, multidrug-resistant colorectal cancer cells have higher levels of both DVL-1 and DVL-3 expression. Moreover, DVL-1 and DVL-3 increase the levels of multidrug resistance protein 1 (P-gp/MDR1), multidrug resistance-associated protein 2 (MRP2), breast cancer resistance protein (BCRP), Survivin and Bcl-2, which are correlated with multidrug resistance^(14,15).

PDZ domain of DVL proteins is able to bind the Frizzled receptor in order to allow the WNT signaling propagation⁽¹⁶⁾. So, WNT pathway may be downregulated by molecules that target the PDZ protein domain of DVL, which would decrease cancer growth. FJ9 was the first non-peptide antagonist of DVL protein–protein interaction reported to suppress β -catenin–dependent tumor cell growth⁽¹⁷⁾. Representative examples of small molecules include sulindac⁽¹⁸⁾, CalBioChem-322338⁽¹⁹⁾, N-benzoyl-2-aminobenzoic acid⁽²⁰⁾, phenoxyacetic acid⁽²¹⁾ and indole-3-carboxamide⁽²²⁾ derivatives. Significant work is being done to create DVL PDZ domain-specific small molecules that could be

Identification of new inhibitors using Virtual Screening technique

used as anticancer drugs.⁽²³⁾ Given the significant role DVL1 plays in tumor development and chemoresistance, our research has focused on finding novel small compounds that can disrupt DVL1's ability to bind to FZD. Our team is heavily active in the search for WNT pathway modulators. We recently reported FzM1⁽²⁴⁾, which inhibits DVL binding to FZD4 by interacting to an allosteric site found in the intracellular loop 3 (ICL3). Furthermore, by specifically targeting the PDZ1 domain, we recently discovered that 3-benzyl-5-chloro-N-(4-(hydroxymethyl)phenyl)-1H-indole-2-carboxamide (2) as an inhibitor of Na⁺/H⁺ exchanger 3 regulating factor 1 (NHERF1)⁽²⁵⁾. The same C-terminal region that FZDs employ to make connections with DVL via the PDZ domains is involved in NHERF1's interactions with FZDs. Our aim was the research of DVL1 inhibitors that exhibit a high level of selectivity toward NHERF1. A specific DVL1 inhibitor might have increased activity since it was found that NHERF1 knock-out favors the DVL binding to FZDs⁽²⁶⁾. On the other hand, DVL inhibition may be reduced when NHERF1 and DVL are both inhibited. In this research, we carried out structure-based virtual screening studies to find DVL1 inhibitors that are selective toward NHERF1 inhibitors. Through these investigations, the molecule RS4690 (1), a powerful and specific inhibitor of DVL1, was successfully discovered. The (S)-1 pure enantiomer inhibited DVL1 more than the (R)-enantiomer. The WNT pathway block and the anticancer efficacy were confirmed by binding competition assays and biological testing in cells.

2.1) Materials and methods

The PDZ structure of DVL1 was obtained by homology modelling. The template structures, PDZ of DVL2, were available at the PDB (pdb code 3CBZ, 3CBY and 3CC0)⁽¹⁶⁾, while the PDZ of DVL1 sequence was retrieved from UniProt database (Available online: <https://www.uniprot.org/> accessed on 9 February 2022) under the code O14640. The sequence

Identification of new inhibitors using Virtual Screening technique

alignment labels showed 75% sequence identity and 85% sequence similarity (positives). The hydrogen atoms were added to the protein using Maestro protein preparation wizard⁽²⁷⁾ and minimized, keeping all the heavy atoms fixed until an rmsd gradient of 0.05 kcal/(mol·Å) was reached. Ligand structures were built with Maestro and minimized using the MMFF94x force field until a rmsd gradient of 0.05 kcal/(mol·Å) was reached. The docking simulations were performed using PLANTS⁽²⁸⁾. We set a binding lattice of 12 Å radius using all default settings. The pharmacophore model was obtained by Phase.⁽²⁹⁾ The pharmacophore features were fixed at the interaction points identified by compound 1 binding mode analyses. A hydrophobic query was also added in a sub-pocket lined by Leu12, Val75 and Ile81. The fitting to this query addressed the desired selectivity between DVL1 and NHERF1 PDZ domain. The polar features had a tolerance of 2 Å while the hydrophobic features had a tolerance of 2.5 Å. The training set obtained by docking computations was scored by the fitting to the pharmacophore model; the selected compounds must match at least five features, and hydrophobic and aromatic queries were equivalent. Molecular dynamics was performed with the AMBER 20 suite.⁽³⁰⁾ The inhibitors were parametrized by Leap of Amber. The structures of complexes were solvated in a periodic octahedron simulation box using TIP3P water molecules, providing a minimum of 10 Å of water between the protein surface and any periodic box edge. Ions were added to neutralize the charge of the total system. The water molecules and ions were energy-minimized, keeping the coordinates of the protein–ligand complex fixed (1000 cycles), and then the whole system was minimized (10,000 cycles). Following minimization, the entire system was heated to 298 K (20 ps). The production (10 ns) simulation was conducted at 298 K with constant pressure and periodic boundary conditions. Shake bond length condition was used (ntc = 2). Production was carried out on a Cineca Marconi 100 supercomputer. Trajectory analysis was carried out by CPPTRAJ program.⁽³¹⁾ The ADME evaluations were carried out

Identification of new inhibitors using Virtual Screening technique

by swissADME server, a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules⁽³²⁾. The pictures reported in the manuscript were generated with PyMOL.⁽³³⁾

2.2) Virtual screening

We used a virtual screening method with a pharmacophoric model set as the primary filter to find novel DVL1 inhibitors following the procedure mentioned before. We focused on the PDZ domain of DVL1 because PDZ domains are appealing targets for drug discovery due to their critical involvement in the transmission of cellular signals and their abnormal activations in many human illnesses⁽³⁴⁾. Instead of focusing on complete signaling cascades, as is often accomplished by receptor antagonists, finding small molecules that can disrupt the protein-protein interactions in which PDZs are implicated demonstrated to be quite advantageous⁽³⁵⁾. Our group developed an inhibitor of NHERF1 PDZ1 domain, compound 2. It was tested in HEK293 cells transiently expressing both HA-tagged-FZD4 (HA-FZD4) and the chimeric construct DVL-GFP. DVL-GFP is localized in signalosomes, which are punctate intracellular structures, in the absence of HA-FZD4⁽³⁶⁾. When FZD-4 is expressed, DVL-GFP is attracted on the cell's PM, dramatically altering its location. Drugs that block FZD4/DVL binding decrease the quantity of DVL-GFP recruited on the PM, as was previously described⁽²⁴⁾. Compound 2 reduced DVL-GFP binding to HA-FZD4 in dose-response experiments with an EC₅₀ of $5.10 \pm 0.09 \mu\text{M}$.

Starting from this information we thought that this compound could bind also to DVL PDZ domain, and we developed a pharmacophore model to aid in the discovery of new DVL1 PDZ binders using the knowledge we gathered from binding mechanism of compound 2. We compared the two PDZ structures and we found some differences between them. For NHERF1, the carboxylic binding loop's upper portion was defined by Tyr24,

Identification of new inhibitors using Virtual Screening technique

Ile79, and Val86, whereas for DVL1, we observed Leu12, Val75, and Ile81. The DVL1 PDZ's smaller residues created a gap that might be able to hold a hydrophobic group.

For this reason, the pharmacophore model also received a new hydrophobic query, set as required during filtering of poses, in order to clearly discriminate between the inhibitory effects of DVL1 and NHERF1 (Figure 6). At the end, the pharmacophore model is featured by seven queries: three aromatics, two hydrophobics, one H-bond donor, and one H-bond acceptor. As a training set, we used our in-house database of compounds of around 7000 molecules, and we docked them at the DVL1 PDZ binding site. We collected 10 conformations for each molecule that were then fitted to the pharmacophore model, in order to score all of the suggested binding poses. The hydrophobic query, which may have driven selection between NHERF1 and DVL1, was considered as a required feature. The top 10 compounds from among the 500 derivatives, that fit the pharmacophore model the best, were sent to the biological assessment after being visually examined. The racemic derivative 1 had the highest EC₅₀ value ($0.74 \pm 0.08 \mu\text{M}$) among the investigated substances for blocking DVL binding to FZD4.

Identification of new inhibitors using Virtual Screening technique

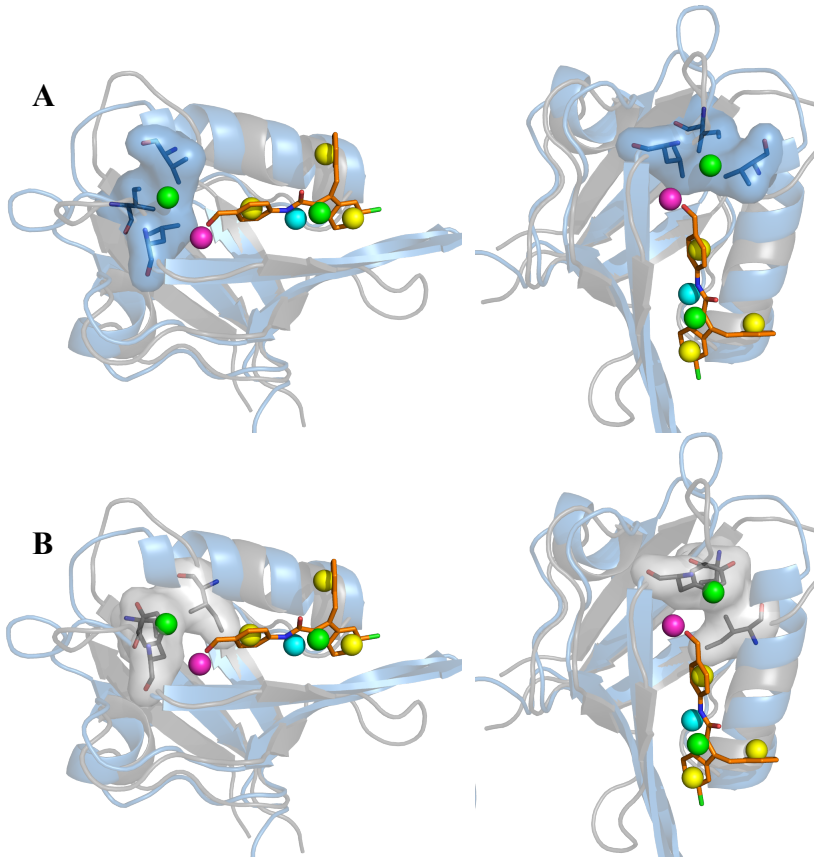


Figure 6. Map of the pharmacophore features superimposed with the RS5517 binding mode. The model has 7 features: aromatic (yellow), H-bond acceptor (magenta), H-bond donor (cyan) and hydrophobic (green). The PDZ domains are also reported: light blue for DVL1 PDZ and grey for NHERF1. The residues Leu12, Val75 and Ile81 that sized an additional sub-pocket for DVL1 were depicted as light blue stick and surface (panel A). The counterpart for NHERF1 DVL1 (panel B) is reported in grey stick and surface. The picture reported a frontal view (left side) and to better appreciate the space sized by the smaller residues of DVL1 PDZ a rotated (90 degrees) view (right side).

Identification of new inhibitors using Virtual Screening technique

2.3) HPLC separation of enantiomers

The racemate of RS4690 (**1**) was separated into the two enantiomers using the Chiralpak AD chiral stationary phase (CSP), a procedure that improved substantially the enantioselective HPLC protocol based on the Chiralpak IA CSP previously reported⁽³⁷⁾

2.4) Binding mode and molecular dynamics simulations

The two enantiomers showed quite different binding modes, so we tried to understand the relative interaction with the PDZ domain. For the (*S*)-**1** the dimethyl phenyl moiety was involved in π -cation interaction with Arg69 and Arg72, the indole ring formed hydrophobic interactions with Val68, the pyridine ring formed hydrophobic contacts with Ile14 and Val75; two H-bonds were also observed involving the indole and carboxamide NHs with the Ile16 and Ile14 backbones, respectively. The (*R*)-**1** enantiomer showed a different pattern of H-bonds and the pyridine moiety lies in a different sub-pocket (Figure 7).

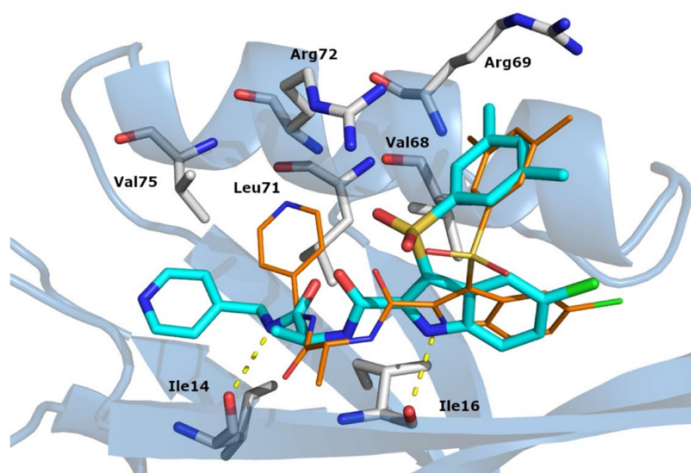


Figure 7. In silico docking results of (*S*)-**1** (cyan) and (*R*)-**1** (orange) in complex with the DVL1 PDZ binding site. Residues involved in interactions are reported as white sticks. The PDZ is depicted as light blue cartoon. H-bonds are reported as yellow dotted lines.

Identification of new inhibitors using Virtual Screening technique

Molecular dynamics was also performed to explore the relative stability of the enantiomers. The analyses led us to observe two additional H-bond between the nitrogen atom of the first carboxamide group and Ile16 backbone and another involving and the oxygen atom of the second carboxamide and the same Ile16. Trying to explain the differences between the two molecules, about the chiral center, we noticed that the (S)-enantiomer's methyl group created connections with Leu71 and Arg72 at the binding site, whereas the (R)-enantiomer's methyl group pointed toward the solvent-exposed region. Another significant distinction between the two enantiomers was that the (S)-enantiomer's pyridine nitrogen atom contacted the Leu12 backbone via a water molecule bridge, but the (R)-enantiomer did not exhibit this type of interaction (Figure 8 A and B).

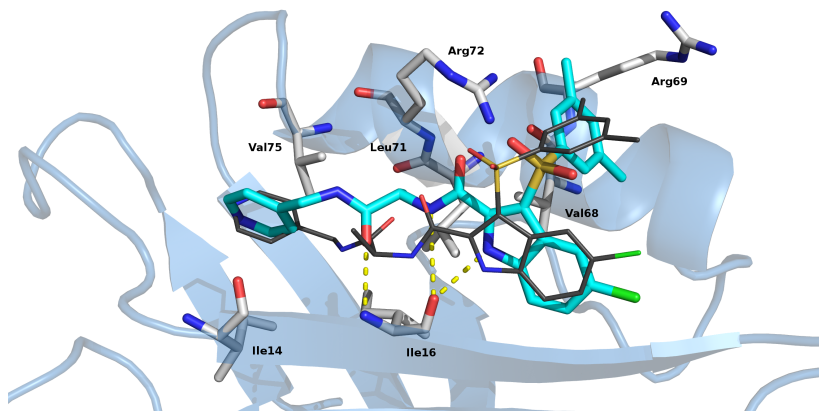


Figure 8. A Snapshot of the trajectory of RS4690 S with DVLI PDZ. The RS4690 docking pose is reported as grey lines; residues involved in interactions were reported as white stick; H-bonds are depicted as yellow dot lines.

Identification of new inhibitors using Virtual Screening technique

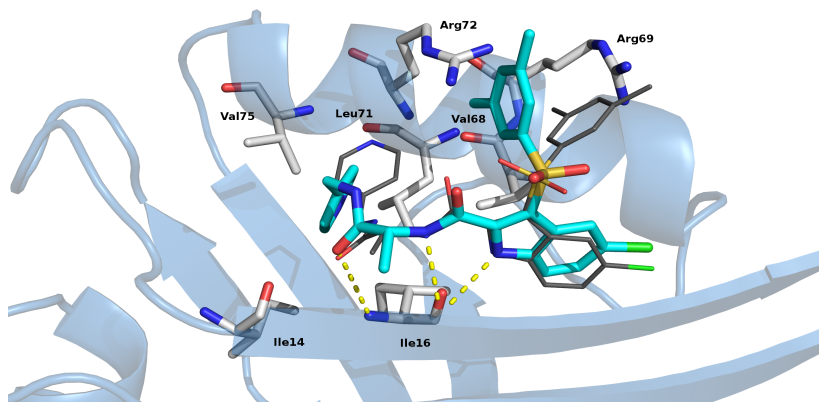


Figure 8. B Snapshot of the trajectory of RS4690 R with DVL1 PDZ. The RS4690 docking pose is reported as grey lines; residues involved in interactions were reported as white stick; H-bonds are depicted as yellow dot lines.

2.5) Binding assays

Equilibrium binding tests were conducted between the DVL1 PDZ domain and a peptide that mimicked the C-terminal region of the TMEM88 protein (TSGKVWV) and was dansylated at its N-terminus in order to validate the *in-silico* predictions. Both the (S)- and (R)-enantiomers eliminated the interaction between PDZ and TMEM88 at a concentration of 5 M. Only (S)-1 appeared to double the K_d of the complex at a concentration of 1 M (from 11.5 ± 0.5 to 20.8 ± 0.5 μM), whereas (R)-enantiomer did not impair the binding reaction (K_d of 12.3 ± 0.6 μM) (Figure 9). Compounds (S)-1 or (R)-1 failed to exhibit inhibitory efficacy when evaluated as NHERF1 PDZ1 domain inhibitors (data not shown). These tests showed that (S)-1 and (R)-1 are DVL1 PDZ domain-selective inhibitors.

Identification of new inhibitors using Virtual Screening technique

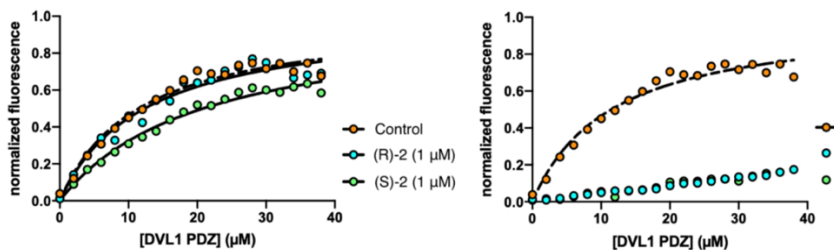


Figure 9. Equilibrium binding experiment between PDZ domain of DVL1 and the C-terminal portion of TMEM88 in the absence and presence of (*S*)-1 or (*R*)-1 at 1 μM (right panel) and 5 μM (left panel). Lines are the best fit for a hyperbolic function.

2.6) In vitro assays

The activity of racemic compound **1** and enantiomers (*S*)-**1** and (*R*)-**1** was assayed in HEK293 cells following DVL-GFP recruitment by FZD4. The racemic mixture of compound **1** reduced DVL-GFP recruitment at the PM with an EC₅₀ of $0.74 \pm 0.08 \mu\text{M}$. Dose-response experiments revealed for (*S*)-**1** and (*R*)-**1** an EC₅₀ of $0.49 \pm 0.11 \mu\text{M}$ and $29.5 \pm 0.9 \mu\text{M}$, respectively (Figure 10 and Table 1)

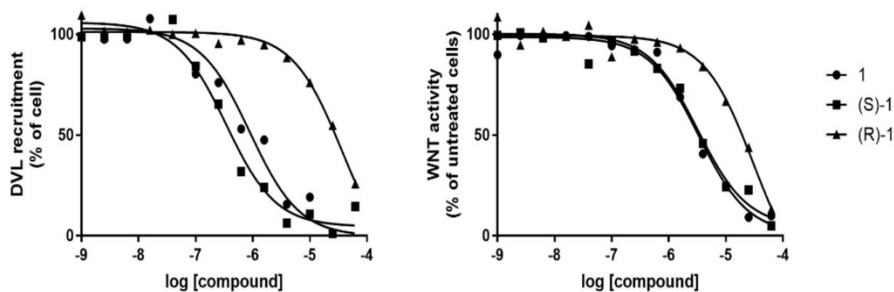


Figure 10. The dose response curves are representative for DVL recruitment inhibition experiment (left panel) and WNT pathway activity measurement (right panel).

Identification of new inhibitors using Virtual Screening technique

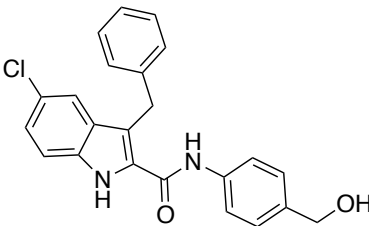
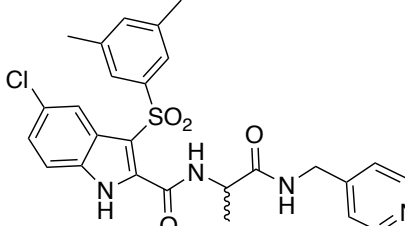
		
EC ₅₀ mM	EC ₅₀ mM	
Cmpd	DVL binding Inhibition	WNT Pathway Inhibition
1	0.74 ± 0.08	3.46 ± 0.07
(S)-1	0.49 ± 0.11	3.09 ± 0.05
(R)-1	29.5 ± 0.9	19.49 ± 0.06
2	5.10 ± 0.009	n.d.

Table 1. DVL Binding and WNT Pathway Inhibition by Compounds 1, (S)-1, (R)-1 and 2

Additionally, HEK293 cells transiently expressing both HA-FZD4 and the WNT reporter construct WRE-GFP were used to test the activity of the drugs⁽³⁸⁾. The WNT reporter construct WRE-GFP exposed the coding sequence of GFP under the control of an optimized WNT responsive element (WRE-wt). WRE-GFP enhances GFP transcription and translation when it is transiently transfected in cells. The amounts of intracellular WNT pathway activity determine how much GFP each cell produces (and, consequently, how fluorescent the entire cell population is). The racemic mixture **1** reduced overall GFP fluorescence, confirming its inhibitory activity on the WNT/ β -catenin pathway with an EC₅₀ of 3.46 ± 0.07 μ M. Dose-response experiments revealed for (S)-**1** and (R)-**1** an EC₅₀ of 0.49 ± 0.11 μ M and 19.49 ± 0.06, respectively (Figure 11 and Table 2).

They tested **1**, (S)-**1**, and (R)-**1** on a panel of colon cancer cell lines known to rely on the activity of the WNT pathway for their survival in order to further demonstrate the WNT-inhibitory

Identification of new inhibitors using Virtual Screening technique

activity. In vitro cultures of colon cancer cells often present mutations either in the APC gene or in the β -catenin gene.

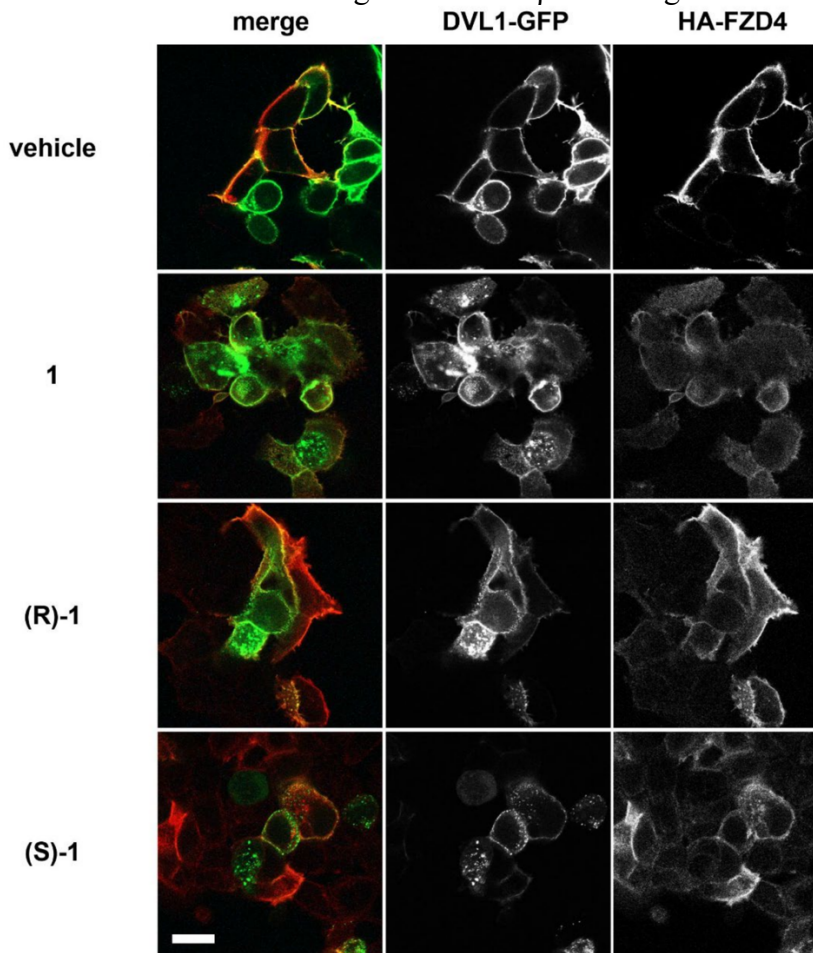


Figure 11. Confocal immunofluorescence showing the recruitment of DVL1-GFP (green) by HA-FZD4 (red) at the PM of HEK293 cells in the presence of 7 μ M of compound **1** (racemic mixture), (*S*)-**1**, (*R*)-**1** or the corresponding volume of vehicle (DMSO). (Magnification bar = to 14 μ m).

Colon cancer cells, SW480 and SW620, were used in dose-response experiments and (*S*)-**1** and (*R*)-**1** showed an EC_{50} for cell growth inhibition in the high micromolar range demonstrating that in cells where WNT pathway activity is independent of the

Identification of new inhibitors using Virtual Screening technique

DVL state, the three compounds do not exert their effect. They also examined their ability to prevent proliferation in the HCT116 colon cancer cell line. In terms of APC and β -catenin, these represent the wt counterparts of SW480 and SW260. HCT116 cells express wt APC and are heterozygous for β -catenin, harboring one wild type and one mutant (Ser45del) allele⁽³⁹⁾. WNT pathway activity and the survival of HCT116 cells thus depend on DVL and can be affected by FZD and DVL inhibitors. Interestingly, dose-response experiments on HCT116 cells revealed EC₅₀ values of cell growth inhibition for compounds **1**, (*S*)-**1** and (*R*)-**1** of 15.2 ± 1.1 , 7.1 ± 0.6 and 28.3 ± 1.2 μ M, respectively. These results confirm that the three compounds act on DVL to inhibit the WNT pathway. In vitro activities towards colon cancer cells further support our in vitro results for these compounds and point toward their eligibility as interesting lead compounds (Table 2).

Compound	EC ₅₀ (μ M)/Cell Lines		
	SW480 ^a	SW620 ^a	HCT116 ^b
1	39.17 ± 1.58	38.54 ± 1.6	15.2 ± 1.1
(<i>S</i>)- 1	54.95 ± 2.2	45.9 ± 2.0	7.1 ± 0.6
(<i>R</i>)- 1	59.47 ± 2.1	54.75 ± 1.9	28.3 ± 1.2

^a Incubation time was 72 h. ^b Incubation time was 48 h.

Table 2 Growth Inhibition of SW680, SW620 and HCT116 Human Colon Carcinoma Cell Lines by Racemate **1** and Enantiomers (*S*)-**1** and (*R*)-**1**.

2.7) ROS production

Reactive oxygen species (ROS) production was also evaluated; Although the mechanism is still mostly unclear, high amounts of ROS are known to cause cell death but are also implicated in other pathways, including WNT signaling^(40,41,42).

Using the fluorescent probe 2',7'-dichlorodihydrofluorescein diacetate DCFDA, the capacity of compounds (*S*)-**1** and (*R*)-**1** to

Identification of new inhibitors using Virtual Screening technique

produce ROS in HCT116 cells was assessed. With increasing concentrations of (S)-1 or (R)-1, it was possible to calculate the ROS production (%) after a 48-hour treatment (Figure 12). The findings demonstrate that ROS generation of (S)-1 was superior to (R)-1 at any chemical concentration evaluated. At 10 μM , the most ROS were produced; at higher concentrations, ROS production drastically declined. The stimulation of the mitochondrial pathway during treatment with (S)-1 10 μM for 48 h led to a large decrease in ROS generation, which may have contributed to the death of the HCT116 cells⁽⁴³⁾.

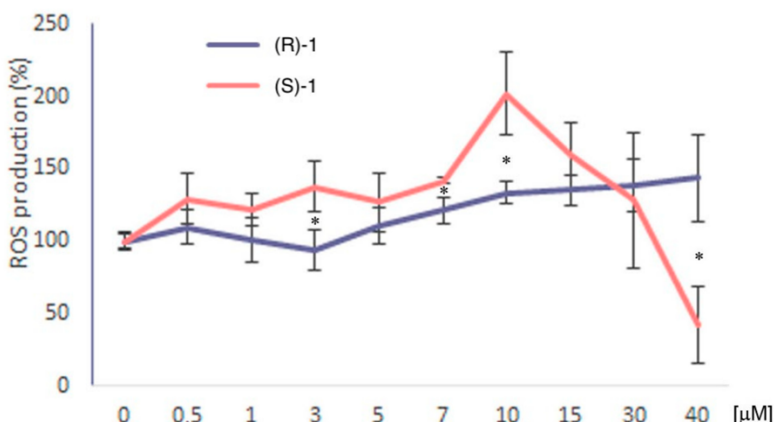


Figure 12. ROS production in HCT116 cells upon treatment with different concentrations of (R)-2 or (S)-2 for 48 h. The asterisks show the differences between the effects induced by the two drugs.

2.8) DVL1 binding interactions features

We also sum up key interactions with the DVL1 PDZ domain: (i) the presence of a 6-membered terminal ring, instead of a 5-membered ring, and (ii) the introduction of a methyl group at position 3 of the side chain (Figure 13). The interactions of both indole nucleus and 3,5-dimethylphenyl group were little affected by the modifications of the side chain. The terminal aromatic ring lay in a hydrophobic pocket surrounded by Leu12, Val75 and

Identification of new inhibitors using Virtual Screening technique

Ile81 residues; the 6-member rings filled the binding pocket ensuring, more stable interactions compared to the 5-membered heterocycles, mainly due to steric rather than electrostatic reasons. The methyl group at the C3' of the side chain was modeled for both (*S*) and (*R*) configurations. We observed that the methyl group of both enantiomers pointed toward the β 2 helix; only in the case of the (*S*)-configuration, the methyl formed hydrophobic contacts with Leu71 and Arg72, whereas the methyl of the (*R*)-enantiomer was too far from the closest residues to achieve positive contact. The absence of the pyridine basic nitrogen atom abolished the water-mediated H-bond with Leu12.

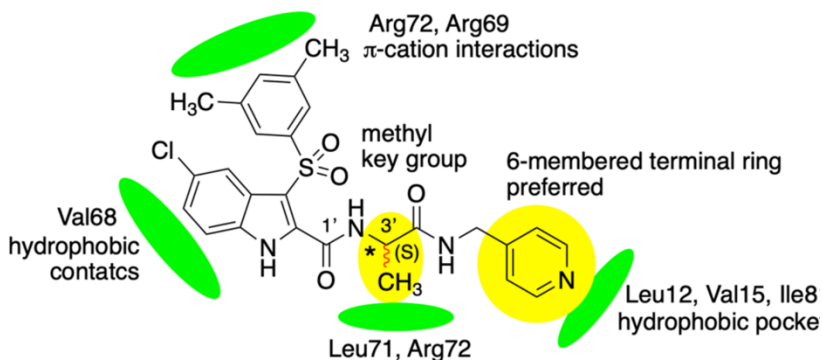


Figure 13. Key contacts of compound (*S*)-2 in the DVL1 PDZ binding pocket. H-bonds are not shown for the sake of clarity.

2.9) Conclusions

Structure-based virtual screening investigations identified the racemic compound 1 with preferential inhibition of DVL1 binding (EC_{50} of $0.74 \pm 0.08 \mu\text{M}$) toward NHERF1 and inhibition of the WNT pathway with an EC_{50} of $3.46 \pm 0.07 \mu\text{M}$. The molecular dynamic simulations revealed that enantiomer (*S*)-1 had a higher affinity for DVL1 than the (*R*)-enantiomer (EC_{50} of $0.49 \pm 0.11 \text{ M}$). At micromolar concentrations, all substances 1, (*S*)-1, and (*R*)-1 inhibited colorectal cancer cells HCT116, SW480, and SW620. These compounds inhibited the growth of

Identification of new inhibitors using Virtual Screening technique

HCT116, a cell line that depends on the activity of the WNT signaling pathway for its survival, expressing wild-type APC, with low micromolar EC50 values; compounds 1, (S)-1 and (R)-1 showed EC50 values of 15.2 ± 1.1 , 7.1 ± 0.6 μM and 28.3 ± 1.2 , respectively. HCT116 cells produced more ROS after being exposed to (S)-1 at a 20 M concentration for 48 hours. These results highlight that compound (S)-1 could be a lead compound with drug-like properties for the development of new therapeutic agents in the treatment of Wnt-dependent colon cancer. Additional research targeted at enhancing the inhibitory concentration and evaluating potential synergistic effects with known β -catenin inhibitors or FZD modulators will be assessed and may constitute a new weapon in the arsenal for the fight against colorectal cancer.

Coluccia A, Bufano M, La Regina G, Puxeddu M, Toto A, Paone A, Bouzidi A, Musto G, Badolati N, Orlando V, Biagioni S, Masci D, Cantatore C, Cirilli R, Cutruzzolà F, Gianni S, Stornaiuolo M, Silvestri R. Anticancer Activity of (S)-5-Chloro-3-((3,5-dimethylphenyl)sulfonyl)-N-(1-oxo-1-((pyridin-4-ylmethyl)amino)propan-2-yl)-1H-indole-2-carboxamide (RS4690), a New Dishevelled 1 Inhibitor. *Cancers* (Basel). 2022 Mar 7;14(5):1358. doi: 10.3390/cancers14051358.

Identification of new inhibitors using Virtual Screening technique

2.10) References

1. Niehrs C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol.* 2012 Dec;13(12):767-79. doi: 10.1038/nrm3470. Epub 2012 Nov 15. PMID: 23151663.
2. Stamos JL, Weis WI. The β -catenin destruction complex. *Cold Spring Harb Perspect Biol.* 2013 Jan 1;5(1):a007898. doi: 10.1101/cshperspect.a007898. PMID: 23169527; PMCID: PMC3579403..
3. Behrens J, von Kries JP, Kühl M, Bruhn L, Wedlich D, Grosschedl R, Birchmeier W. Functional interaction of beta-catenin with the transcription factor LEF-1. *Nature.* 1996 Aug 15;382(6592):638-42. doi: 10.1038/382638a0. PMID: 8757136.
4. Van der Flier LG, Sabates-Bellver J, Oving I, Haegebarth A, De Palo M, Anti M, Van Gijn ME, Suijkerbuijk S, Van de Wetering M, Marra G, Clevers H. The Intestinal Wnt/TCF Signature. *Gastroenterology.* 2007 Feb;132(2):628-32. doi: 10.1053/j.gastro.2006.08.039. Epub 2006 Aug 18. PMID: 17320548.
5. Steinhart Z, Angers S. Wnt signaling in development and tissue homeostasis. *Development.* 2018 Jun 8;145(11):dev146589. doi: 10.1242/dev.146589. PMID: 29884654.
6. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene.* 2017 Mar;36(11):1461-1473. doi: 10.1038/onc.2016.304. Epub 2016 Sep 12. PMID: 27617575; PMCID: PMC5357762.
7. Li X, Ortiz MA, Kotula L. The physiological role of Wnt pathway in normal development and cancer. *Exp Biol Med (Maywood).* 2020 Mar;245(5):411-426. doi: 10.1177/1535370220901683. Epub 2020 Jan 29. PMID: 31996036; PMCID: PMC7082880.
8. Sharma M, Castro-Piedras I, Simmons GE Jr, Pruitt K. Dishevelled: A masterful conductor of complex Wnt

Identification of new inhibitors using Virtual Screening technique

- signals. *Cell Signal*. 2018 Jul;47:52-64. doi: 10.1016/j.cellsig.2018.03.004. Epub 2018 Mar 17. PMID: 29559363; PMCID: PMC6317740.
9. Uematsu K, He B, You L, Xu Z, McCormick F, Jablons DM. Activation of the Wnt pathway in non small cell lung cancer: evidence of dishevelled overexpression. *Oncogene*. 2003 Oct 16;22(46):7218-21. doi: 10.1038/sj.onc.1206817. PMID: 14562050.
 10. Uematsu K, Kanazawa S, You L, He B, Xu Z, Li K, Peterlin BM, McCormick F, Jablons DM Wnt pathway activation in mesothelioma: Evidence of dishevelled overexpression and transcriptional activity of beta-catenin. *Cancer Res*. 2003, 63, 4547–4551.
 11. Lee YN, Gao Y, Wang HY Differential mediation of the Wnt canonical pathway by mammalian Dishevelleds-1, -2, and -3. *Cell Signal*. 2008, 20, 443–452
 12. Gao C, Chen YG. Dishevelled: The hub of Wnt signaling. *Cell Signal*. 2010 May;22(5):717-27. doi: 10.1016/j.cellsig.2009.11.021. Epub 2009 Dec 13. PMID: 20006983.
 13. Wharton KA Jr. Runnin' with the Dvl: proteins that associate with Dsh/Dvl and their significance to Wnt signal transduction. *Dev Biol*. 2003 Jan 1;253(1):1-17. doi: 10.1006/dbio.2002.0869. PMID: 12490194.
 14. Xu Y, Zhang C, Liang H, Hu S, Li P, Liu L, Duan X, Chen C, Zhang Y, Dai P. Dishevelled 1, a pivotal positive regulator of the Wnt signalling pathway, mediates 5-fluorouracil resistance in HepG2 cells. *Artif Cells Nanomed Biotechnol*. 2018;46(sup2):192-200. doi: 10.1080/21691401.2018.1453827. Epub 2018 Mar 27. PMID: 29583038.
 15. Luo K, Guc X, Liua J, Zenga G, Penga L, Huanga H, Jianga M, Yangc P, Lic M, Yanga Y et al. Inhibition of Disheveled-2 resensitizes cisplatin-resistant lung cancer cells through down-regulating Wnt/ β -catenin signaling. *Exp. Cell Res*. 2016, 347, 105–113

Identification of new inhibitors using Virtual Screening technique

16. Zhang Y, Appleton BA, Wiesmann C, Lau T, Costa M, Hannoush RN, Sidhu SS. Inhibition of Wnt signaling by Dishevelled PDZ peptides. *Nat Chem Biol.* 2009 Apr;5(4):217-9. doi: 10.1038/nchembio.152. Epub 2009 Mar 1. PMID: 19252499.
17. Fujii N, You L, Xu Z, Uematsu K, Shan J, He B, Mikami I, Edmondson LR, Neale G, Zheng J, Guy RK, Jablons DM. An antagonist of dishevelled protein-protein interaction suppresses beta-catenin-dependent tumor cell growth. *Cancer Res.* 2007 Jan 15;67(2):573-9. doi: 10.1158/0008-5472.CAN-06-2726. Erratum in: *Cancer Res.* 2007 Mar 1;67(5):2389. PMID: 17234765.
18. Lee HJ, Wang NX, Shi DL, Zheng JJ. Sulindac inhibits canonical Wnt signaling by blocking the PDZ domain of the protein Dishevelled. *Angew Chem Int Ed Engl.* 2009;48(35):6448-52. doi: 10.1002/anie.200902981. PMID: 19637179; PMCID: PMC2978498.
19. Grandy D, Shan J, Zhang X, Rao S, Akunuru S, Li H, Zhang Y, Alpatov I, Zhang XA, Lang RA, Shi DL, Zheng JJ. Discovery and characterization of a small molecule inhibitor of the PDZ domain of dishevelled. *J Biol Chem.* 2009 Jun 12;284(24):16256-16263. doi: 10.1074/jbc.M109.009647. Epub 2009 Apr 21. PMID: 19383605; PMCID: PMC2713547.
20. Hori K, Ajioka K, Goda N, Shindo A, Takagishi M, Tenno T, Hiroaki H. Discovery of Potent Disheveled/Dvl Inhibitors Using Virtual Screening Optimized With NMR-Based Docking Performance Index. *Front Pharmacol.* 2018 Sep 5;9:983. doi: 10.3389/fphar.2018.00983. PMID: 30233369; PMCID: PMC6134994.
21. Choi J, Ma S, Kim HY, Yun JH, Heo JN, Lee W, Choi KY, No KT. Identification of small-molecule compounds targeting the dishevelled PDZ domain by virtual screening and binding studies. *Bioorg Med Chem.* 2016

Identification of new inhibitors using Virtual Screening technique

- Aug 1;24(15):3259-66. doi: 10.1016/j.bmc.2016.03.026. Epub 2016 Mar 16. PMID: 27112452.
22. Kamdem N, Roske Y, Kovalsky D, Platonov MO, Balinskyi O Kreuchwig A, Saupe J, Fang L, Diehl A, Schmieder P et al. Small-molecule inhibitors of the PDZ domain of Dishevelled proteins interrupt Wnt signalling. *Magn. Reson.* 2021, 2, 355–374
 23. Liu Z, Wang P, Wold EA, Song Q, Zhao C, Wang C, Zhou J. Small-Molecule Inhibitors Targeting the Canonical WNT Signaling Pathway for the Treatment of Cancer. *J Med Chem.* 2021 Apr 22;64(8):4257-4288. doi: 10.1021/acs.jmedchem.0c01799. Epub 2021 Apr 6. PMID: 33822624.
 24. Generoso SF, Giustiniano M, La Regina G, Passacantilli S, Cassese H, Bruno A, Mallardo M, Silvestri R, Marinelli L, Bonatti S et al. Pharmacological folding chaperones act as allosteric ligands of Frizzled4. *Nat. Chem. Biol.* 2015, 11, 280–286
 25. Coluccia A, La Regina G, Naccarato V, Nalli M, Orlando V, Biagioni S, De Angelis ML, Baiocchi M, Gautier C, Gianni S, Di Pastena F, Di Magno L, Canettieri G, Coluccia AML, Silvestri R. Drug Design and Synthesis of First in Class PDZ1 Targeting NHERF1 Inhibitors as Anticancer Agents. *ACS Med Chem Lett.* 2019 Jan 14;10(4):499-503. doi: 10.1021/acsmedchemlett.8b00532. PMID: 30996786; PMCID: PMC6466550.
 26. Wheeler DS, Barrick SR, Grubisha MJ, Brufsky AM, Friedman PA, Romero G. Direct interaction between NHERF1 and Frizzled regulates β -catenin signaling. *Oncogene* 2011, 30, 32–42
 27. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des.* 2013 Mar;27(3):221-34. doi:

Identification of new inhibitors using Virtual Screening technique

- 10.1007/s10822-013-9644-8. Epub 2013 Apr 12. PMID: 23579614.
28. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model.* 2009 Jan;49(1):84-96. doi: 10.1021/ci800298z. PMID: 19125657.
 29. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des.* 2006 May;67(5):370-2. doi: 10.1111/j.1747-0285.2006.00384.x. PMID: 16784462.
 30. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem.* 2005 Dec;26(16):1668-88. doi: 10.1002/jcc.20290. PMID: 16200636; PMCID: PMC1989667.
 31. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput.* 2013 Jul 9;9(7):3084-95. doi: 10.1021/ct400341p. Epub 2013 Jun 25. PMID: 26583988.
 32. van Meerloo J, Kaspers GJ, Cloos J. Cell sensitivity assays: the MTT assay. *Methods Mol Biol.* 2011;731:237-45. doi: 10.1007/978-1-61779-080-5_20. PMID: 21516412
 33. *PyMOL*, version 1.2r1; DeLanoScientificLLC: SanCarlos, CA, USA, 2021
 34. Christensen NR, Čalyševa J, Fernandes EFA, Lüchow S, Clemmensen LS, Haugaard-Kedström LM, Strømgaard K. PDZ Domains as Drug Targets. *Adv Ther (Weinh).* 2019 Jul;2(7):1800143. doi: 10.1002/adtp.201800143. Epub 2019 Apr 24. PMID: 32313833; PMCID: PMC7161847.
 35. Wang NX, Lee HJ, Zheng JJ. Therapeutic use of PDZ protein-protein interaction antagonism. *Drug News*

Identification of new inhibitors using Virtual Screening technique

- Perspect. 2008 Apr;21(3):137-41. PMID: 18560611; PMID: PMC4055467.
36. Riccio G, Bottone S, La Regina G, Badolati N, Passacantilli S, Rossi GB, Accardo A, Dentice M, Silvestri R, Novellino E, Stornaiuolo M. A Negative Allosteric Modulator of WNT Receptor Frizzled 4 Switches into an Allosteric Agonist. *Biochemistry*. 2018 Feb 6;57(5):839-851. doi: 10.1021/acs.biochem.7b01087. Epub 2018 Jan 19. PMID: 29293331.
 37. Famiglini V, La Regina G, Coluccia A, Masci D, Brancale A, Badia R, Riveira-Muñoz E, Esté JA, Crespan E, Brambilla A, Maga G, Catalano M, Limatola C, Formica FR, Cirilli R, Novellino E, Silvestri R. Chiral Indolylarylsulfone Non-Nucleoside Reverse Transcriptase Inhibitors as New Potent and Broad Spectrum Anti-HIV-1 Agents. *J Med Chem*. 2017 Aug 10;60(15):6528-6547. doi: 10.1021/acs.jmedchem.6b01906. Epub 2017 Jul 5. PMID: 28628334.
 38. Wong HC, Bourdelas A, Krauss A, Lee HJ, Shao Y, Wu D, Mlodzik M, Shi DL, Zheng J. Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled. *Mol Cell*. 2003 Nov;12(5):1251-60. doi: 10.1016/s1097-2765(03)00427-1. PMID: 14636582; PMID: PMC4381837.
 39. Kaler P, Augenlicht L, Klampfer L. Activating mutations in β -catenin in colon cancer cells alter their interaction with macrophages; the role of snail. *PLoS One*. 2012;7(9):e45462. doi: 10.1371/journal.pone.0045462. Epub 2012 Sep 21. PMID: 23029025; PMID: PMC3448637.
 40. Korswagen HC. Regulation of the Wnt/beta-catenin pathway by redox signaling. *Dev Cell*. 2006 Jun;10(6):687-8. doi: 10.1016/j.devcel.2006.05.007. PMID: 16740470.

Identification of new inhibitors using Virtual Screening technique

41. Caliceti C., Nigro P., Rizzo P., Ferrari R. ROS, Notch, and Wnt signaling pathways: Crosstalk between three major regulators of cardiovascular biology. *Biomed. Res. Int.* 2014;2014:318714. doi: 10.1155/2014/318714
42. Rharass T, Lemcke H, Lantow M, Kuznetsov SA, Weiss DG, Panáková D. Ca²⁺-mediated mitochondrial reactive oxygen species metabolism augments Wnt/ β -catenin pathway activation to facilitate cell differentiation. *J Biol Chem.* 2014 Oct 3;289(40):27937-51. doi: 10.1074/jbc.M114.573519. Epub 2014 Aug 14. PMID: 25124032; PMCID: PMC4183826.
43. Redza-Dutordoir M, Averill-Bates DA. Activation of apoptosis signalling pathways by reactive oxygen species. *Biochim Biophys Acta.* 2016 Dec;1863(12):2977-2992. doi: 10.1016/j.bbamcr.2016.09.012. Epub 2016 Sep 17. PMID: 27646922.

3) LDHA (lactate dehydrogenase A)

Contrary to healthy cells, most cancer cells rely on an accelerated rate of glycolysis that frequently ferments glucose into lactate, even when the environment is aerobic. Otto Warburg, a German scientist, recognized the metabolic shift from oxidative phosphorylation (OXPHOS) to aerobic glycolysis (Warburg effect) and correlated it to the unusual glucose metabolism of cancer cells⁽¹⁾. Because cancer cells require more nutrients than healthy cells do, tumor glycolysis gives them a distinct advantage in terms of survival and growth^(2,3,4). Given how significantly cancer cells rely on this metabolism, there has recently been a rise in interest in the field of tumor glycolysis. Thus, attractive targets including those involved in glycolysis's transporters and enzymes are seen as promising targets. One of these received much interest for the creation of anti-cancer drugs: the lactate dehydrogenase enzyme (LDH), which is essential for tumor glycolysis.

The lactate dehydrogenase (LDH) enzyme produces lactate from pyruvate as the primary step in this glycolytic reprogramming. Almost all bodily tissues contain this enzyme, which is responsible for interconversion of pyruvate-NADH to lactate-NAD⁺. Structurally, LDH is a tetrameric enzyme composed of two main subunits, LDHA (muscle, M) and LDHB (heart, H), that are encoded by two distinct genes, *ldhA* and *ldhB*, respectively. These two subunits, LDHA and LDHB, combine into five different possible combinations, such as A₄, A₃B₁, A₂B₂, A₁B₃, and B₄, corresponding to five isoforms or isozymes, namely, LDH₅ to LDH₁, respectively. The homotetramer LDH₅ (A₄, LDHA) is assembled only from subunit A and is predominately found in skeletal muscle, whereas LDH₁ (B₄, LDHB), formed from only subunit B, primarily exists in heart muscle and other oxygenated (normoxia) tissues. The remaining three isoforms, LDH₂ (A₁B₃), LDH₃ (A₂B₂), and LDH₄ (A₃B₁), are hybrid tetramers. An additional sixth isoform, LDH-X or LDH-C₄, which is involved in male fertility, was identified in human testis and sperm. Pyruvate is converted to lactate by the A form,

Identification of new inhibitors using Virtual Screening technique

whereas the B form is responsible for the reverse conversion^(5,6) (Figure 14).

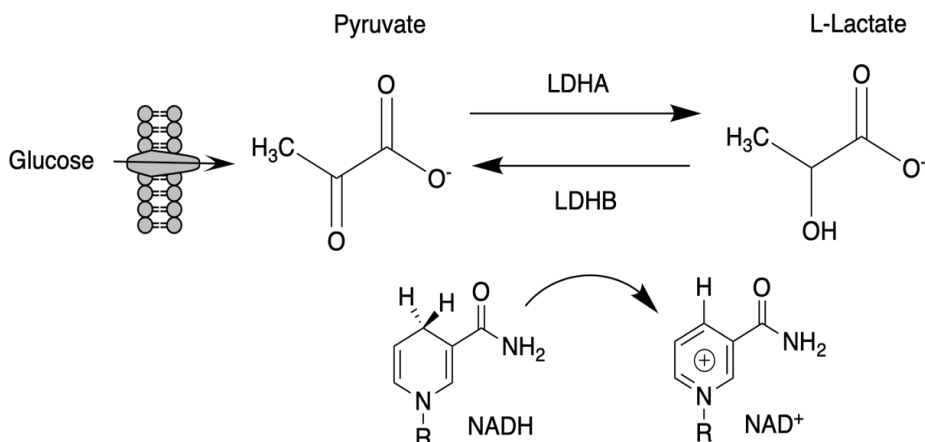


Figure 14. General reactions catalyzed by LDHA and LDHB.

The overexpression of LDHA has been frequently reported in a variety of highly glycolytic human cancers, such as non-Hodgkin lymphoma⁽⁷⁾, colorectal cancer^(8,9), melanoma⁽¹⁰⁾, pancreatic cancer⁽¹¹⁾, lung cancer^(12,13), prostate cancer⁽¹⁴⁾, gastric cancer⁽¹⁵⁾ and endometrial cancer⁽¹⁶⁾. Additionally, LDHA enzymatic activity lowers the extracellular pH by increasing lactate production and subsequent secretion outside the cells, favoring tumor invasion and metastasis^(17,18).

From a structural viewpoint LDH is featured by two domains: the larger (cofactor binding) domain formed by residues 20–162 and 248–266 and the mixed α/β substrate binding domain (smaller) comprising residues 163–247 and 267–331. The NADH cofactor binds to LDH at one end of the central β -sheet with residues His195, Asp168, Arg171, and Thr246 in an extended conformation, and its nicotinamide group partially overlaps the substrate-binding site⁽¹⁹⁾.

Taking a look at the progress made over time on the discovery of inhibitors (Figure 15), we can start from sodium oxamate (1) that inhibits aspartate aminotransferase and LDH (AAT). It inhibits

Identification of new inhibitors using Virtual Screening technique

the human LDHA in vitro in a pyruvate-competitive manner with a K_i of 136,3 μM ⁽²⁰⁾. Analysis of a ternary complex (NADH-LDH-oxamate) indicated that oxamate interacts with residues Gln99, Arg105, Asn137, Arg168, His192, and Thr247 via hydrogen-bonding, whereas Leu164 and Ala237 are engaged in hydrophobic contacts. The side chain of Arg168 interacts with the carboxylate group of the ligand^(19,21). Sodium oxamate enhances the expression of the tumor suppressor p53 while inhibiting aerobic glycolysis and the development of cancer cells.

The inhibition of **1** is observed only in glucose exposed cells, suggesting that this effect is related to the inhibition of aerobic glycolysis. Due to issues with its poor cellular absorption, weak efficacy, and restricted enzymatic selectivity, it has difficulty in being used clinically⁽²²⁾.

Another inhibitor is gossypol (**2**), a natural product (a polyphenolic binaphthyl disesquiterpene) first discovered in the cotton seeds of gossypium. It has a variety of biological properties that are promising, including anticancer, antioxidant, antiviral, and antiparasitic properties⁽²³⁾. It exists in two enantiomeric forms, (R) and (S) and the (R)-(+ form of **1** is more effective than the (S)-(+ isomer for dose-dependent cytotoxic activity, with mean IC50 values of 20 M in melanoma, lung cancer, breast cancer, cervical cancer, and leukemia cell lines⁽²⁴⁾. It also showed an ability to inhibit various LDH isoforms in an NADH-competitive manner^(25,26). Due to its chemical structure, the two aldehyde groups and catechol hydroxyl groups generate toxic metabolites⁽²⁷⁾. So, because of its lack of selectivity and substantial side effects, further development of gossypol as an anticancer agent was unsuccessful. As a result, various compounds with chemical structures that were closely related to this compound were developed. For example, FX11 (**3**) inhibits LDHA selectively with a K_i value of 8 μM , causes cell death, and promotes apoptosis with a simultaneous decrease in lactate production^(28,29).

Quinolinesulfonamide (**4**), NADH-competitive LDHA inhibitors decrease lactate production rates in multiple cancer cell lines,

Identification of new inhibitors using Virtual Screening technique

including hepatocellular and breast carcinomas. Lead optimization yielded QSA (**4**) with LDHA inhibitory potency as low as 2 nM and higher selectivity over LDHB⁽³⁰⁾.

Galloflavin (**5**), a tricyclic flavone-like compound, was found to be a strong *in vivo* anticancer drug and to be a nonselective inhibitor of LDHA with a K_i value of 5.46 μM for pyruvate and 15.06 μM for LDHB. The anticancer activity of **5** was observed in MCF-7, triple negative MDA-MB-231 and tamoxifen resistant MCF-tam cell lines^(31,32).

N-hydroxyindole (NHIs) (**6**) has been recently discovered as selective inhibitors; they are featured by a central indole scaffold contains a hydroxyl group on the nitrogen atom in the 1-position and a carboxyl group in the 2-position of the indole moiety. They showed inhibitory activity in the competition with both the cofactor NADH and the substrate pyruvate, with K_i values in the low micromolar range⁽³³⁾.

Compound NHI-1 (**6**) inhibited LDHA with $K_i = 8.9 \mu\text{M}$ and $K_i = 4.7 \mu\text{M}$ against cofactor NADH and pyruvate, respectively⁽³⁴⁾.

With compound **7** started the pyrazole series and quantitative high-throughput screening and structure-based optimization of this compound enabled the development of compound **8** that showed inhibition of the LDHA and LDHB with IC_{50} values of 0.032 and 0.027 μM , respectively⁽³⁵⁾. Compound **9** (NCI-006) too was generated as a further optimization of the pyrazole series which inhibited the LDHA with $\text{IC}_{50} = 0.06 \mu\text{M}$ and LDHB with IC_{50} of 0.03 μM , and showed potent on-target *in vivo* activity⁽³⁶⁾. Similarly, **10** (NCATS-SM1441) inhibited both LDHA and LDHB with an IC_{50} of 0.04 μM , and showed desirable attributes for further studying the effect of *in vivo* LDH inhibition⁽³⁷⁾.

Genentech **11** (Gen140) was identified as a promising LDHA inhibitor with IC_{50} of 8.9 nM. However, **11** did not appear to have progressed into clinical studies⁽³⁸⁾. Compound **12** inhibited the LDHA activity with an EC_{50} value of 0.090 μM and reduced MiaPaCa-2 cancer cell proliferation with an IC_{50} value of 2.1 μM ⁽³⁹⁾.

Identification of new inhibitors using Virtual Screening technique

Although certain LDHA inhibitors are already available, they have all demonstrated limitations that call for additional research to find innovative LDH inhibitors with enhanced potency, selectivity, and dependable safety and tolerance.

Here, we present the results of virtual screening investigations using docking and pharmacophore filters to find novel LDHA inhibitors.

Starting from the available crystal structures, each monomer is made up of 330 amino acids and has two binding pockets: a cofactor-binding site (NADH-binding site) and a substrate-binding site (pyruvate binding site), both of which are isolated from the solvent by a moveable "active-site loop" (residues 98–110). Both academia and the pharmaceutical sector have reported several LDHA inhibitors, but none of them were successful in the clinical study, mostly due to insufficient cellular and in vivo activity and inadequate pharmacokinetics parameters⁽⁵⁾. Pyrimidine derivatives showed, in preliminary tests, to significantly suppress the growth of either medulloblastoma or pancreatic tumor cells in vitro⁽⁴⁰⁾. Then, using molecular fitness techniques, more powerful pyrimidines have been found, including the patented tetrahydropyrimidine-5-carboxamide⁽⁴¹⁾. Here, we discuss the computational experiments that led to the discovery of derivative 18 as a potential lead compound inhibitor with submicromolar in vitro activity and a good drug-likeness profile.

Identification of new inhibitors using Virtual Screening technique

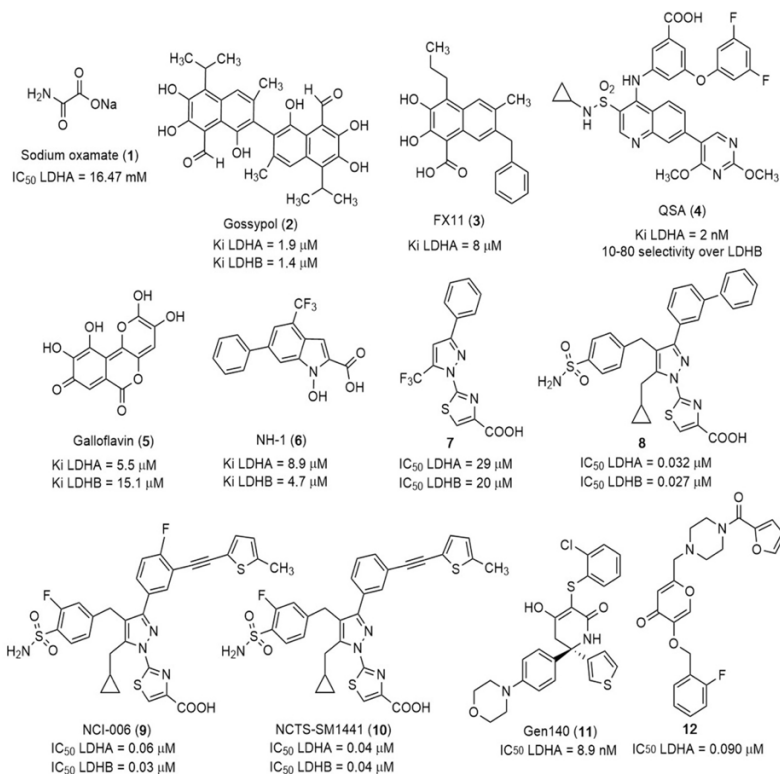


Figure 15. Chemical structures of LDHA inhibitors 1-12

Identification of new inhibitors using Virtual Screening technique

3.1) Materials and methods

3.1.1) Molecular modeling

Compounds 14–22 were purchased from Life Chemicals. The LDHA structures were downloaded from the PDB, pdb code 5W8K⁽³⁵⁾, 5W8I⁽³⁵⁾, 5IXS⁽³⁸⁾, 4QO7⁽⁴²⁾, 4ZVV⁽⁴³⁾ and 4M49⁽⁴⁴⁾. The proteins were prepared by adding hydrogen atoms, fixing ionization state and heteroatoms and removing crystallization co-factors using Maestro protein preparation wizard⁽⁴⁵⁾. The hydrogen atoms minimization was carried out keeping all the heavy atoms fixed until a rmsd gradient of 0.05 kcal/(mol·Å) was reached. Ligand structures were built with Maestro and minimized using the MMFF94x force field until a rmsd gradient of 0.05 kcal/(mol·Å) was reached. The 5W8I LDHA structure was selected for the following docking simulation by a Cross-docking approach. The docking simulations were performed by Plants⁽⁴⁶⁾. It was fixed a binding lattice of 12 Å radius, using all the default settings. The pharmacophore model was obtained by Phase^(47,48). The polar features had a tolerance of 2 Å while the hydrophobic/aromatic features had a tolerance of 2.5 Å. The commercially available compounds libraries Maybridge [www.Maybridge.com], Specs [www.SPECS.net] and Life Chemicals [www.lifechemicals.com] of about 2,000,000 derivatives were filtered out by the rule of five, and the obtained training set, about 1,400,000 compounds, was docked at the LDHA binding site. The docking simulations were used as a sort of tuned conformational search. All the docking proposed binding conformations (10 per molecule) were filtered out by the pharmacophore model and the 500 best ranked derivatives were visual inspected. The images in the manuscript were created with PyMOL⁽⁴⁹⁾.

3.1.2) Molecular dynamics

Molecular dynamics was performed with the AMBER 20 suite⁽⁵⁰⁾. Derivative 18 and NADH were parametrized by Leap of Amber. The structures of complex were solvated in a periodic octahedron simulation box using TIP3P water molecules, providing a

Identification of new inhibitors using Virtual Screening technique

minimum of 10 Å of water between the protein surface and any periodic box edge. Ions were added to neutralize the charge of the total system. The water molecules and ions were energy-minimized keeping the coordinates of the protein-ligand complex fixed (5000 cycle), and then the whole system was minimized (10000 cycles). Following minimization, the entire system was heated to 298 K (20 ps). The production (50 ns) simulation was conducted at 298 K with constant pressure and periodic boundary condition. Shake bond length condition was used ($ntc = 2$). A restraint (restraint wt = 3) was applied on NADH in order to let it stay in the cofactor binding site during the simulation time. Trajectories analysis was carried out by CPPTRAJ program⁽⁵¹⁾.

3.2) Virtual screening studies

We performed virtual screening experiments using a structure-based approach to find assessable inhibitors of LDH activity. Cross-docking was used to choose the 3-D structure of the enzyme from among those found in the Protein Data Bank (<https://www.rcsb.org>). The training set was created by combining the datasets Maybridge [www.Maybridge.com], Specs [www.SPECS.net], and Life Chemicals [www.lifechemicals.com], which are all commercially available until reaching a database of about 2.000.000 molecules. To guarantee that the examined compounds have drug-like properties, the Lipinski Rule of Five was used to pre-filter the training set⁽⁵²⁾.

We used Plants to dock the obtained training set into the LDH catalytic site of the enzyme surrounded by the activation loop and the NADH pocket. A pharmacophore model was created using the binding conformations of LDH inhibitors that are reported in Protein Data Bank (PDB). This model collected the primary binding interactions of the investigated inhibitors to LDHA in 6 queries, including 4 hydrophobic/aromatic, 1 H-bond acceptor, and 1 H-bond donor interactions. The proposed docking conformations (10 per molecule) were filtered out by the

Identification of new inhibitors using Virtual Screening technique

pharmacophore model, and the 500 best scored compounds were visually inspected. Then, nine compounds with the greatest potential were sent to the biological examination (Table 3).

Identification of new inhibitors using Virtual Screening technique

Compounds	Structure	Compounds	Structure
14 RS6208		15 RS6209	
16 RS6210		17 RS6211	
18 RS6212		19 RS6219	
20 RS6220		21 RS6221	
22 RS6110			

Table 3. Structures of compounds 14–22 selected by virtual screening studies

Identification of new inhibitors using Virtual Screening technique

3.3) **In vitro antitumor activity of the identified LDHA inhibitors**

Med1-MB, a cell line derived from Sonic Hedgehog (SHH) medulloblastoma (SHH MB), a pediatric tumor of the cerebellum characterized by dysregulation of the SHH signaling system, were used to conduct preliminary tests on the effects of the chosen drugs⁽⁵³⁾. In their earlier research, they discovered that through activating transcriptional targets, the anomalous activation of the Sonic Hedgehog pathway promotes a reprogramming of energy metabolism toward aerobic glycolysis. Because SHH MB cells rely on aerobic glycolysis, inhibitors of glycolysis such dichloroacetate, a PDK inhibitor, or bromopyruvate, a HK2 inhibitor, drastically slow down their proliferation⁽⁵⁴⁾. They examined how the selected compounds 14-22 at various concentrations affected the rate of proliferation of Med1-MB tumor cells, which are defined by mutations in the Ptch1 receptor gene⁽⁵⁵⁾.

The known LDHA inhibitor sodium oxamate inhibited the proliferation of these cells at millimolar doses, with an estimated IC50 of 15.94 mM, demonstrating their dependence on aerobic glycolysis (Figure 16).

Identification of new inhibitors using Virtual Screening technique

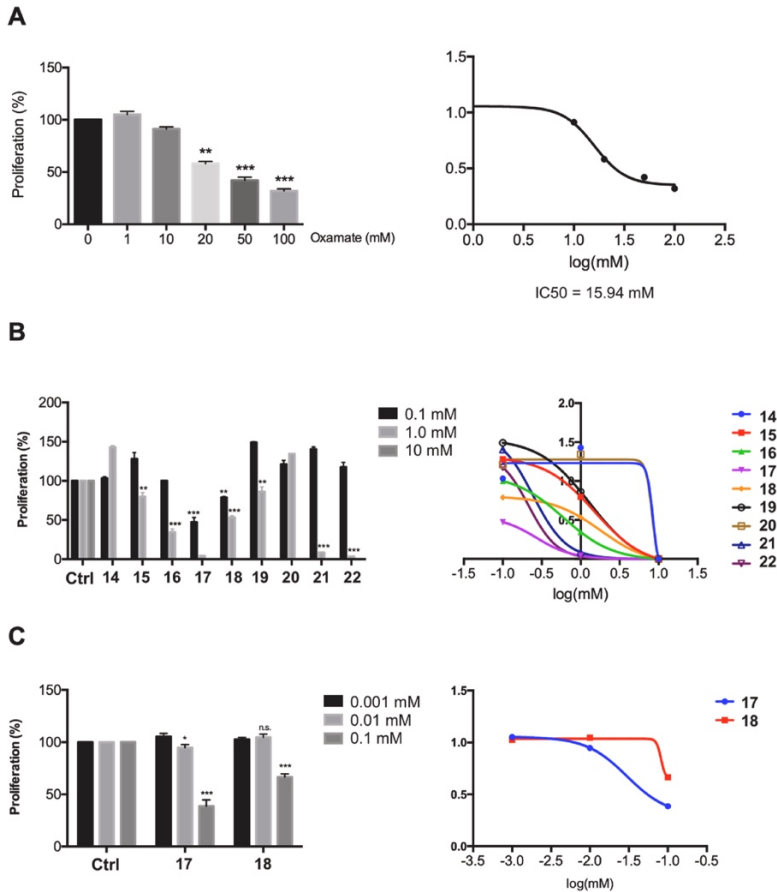


Figure 16. Testing of LDHA inhibitor in cancer cells. (A) (Left) Proliferation assay in Med1-MB cells treated with increasing concentrations of sodium oxamate for 24 h. IC_{50} values of sodium oxamate were determined by generating a dose-response curves by non-linear regression (Right). (B) (Left) Med1-MB cells were treated with or without compounds **14–21** at the indicated concentrations, and cell proliferation was evaluated after 24 h of treatment. (Right). Graphical representation of IC_{50} values. (C) (Left) Proliferation assay in Med1-MB cells treated with compound **17** or **18** or control (DMSO) at the indicated concentrations for 24 h (Right) A dose response curve was generated and IC_{50} values were calculated for both compounds. Data are represented as the mean \pm SD of three independent experiments, each performed in triplicate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. = not significant as determined by ANOVA test.

Identification of new inhibitors using Virtual Screening technique

Following a 24-hour incubation period with progressively lower concentrations of each of the nine compounds, the cells were stained with Trypan Blue before being counted. According to Figure 16B, at 10 mM, all compounds totally stopped cell proliferation, whereas at 1 mM, only compounds 13 and 20 had no meaningful effect. Only compounds 17 and 18 effectively suppressed cell proliferation at a dose of 100 μ M; their IC₅₀ values were 29 μ M and 81 μ M, respectively (Figure 16 C).

It should be noted that these values are significantly lower (534 and 196 fold, respectively) than the concentrations required to suppress cell proliferation by the reference inhibitor sodium oxamate (IC₅₀= 15.94 mM), demonstrating the efficacy of the found compounds (Table 4).

Table 4
Inhibition of growth of Med1-MB cell line by compounds **14-22**.^a

Compd	IC ₅₀ \pm SD ^b (mM)
	Med1-MB
14	8.463
15	1.510
16	1.157
17	0.0298
18	0.08177
19	1.295
20	8.309
21	0.964
22	1.139
Oxamate	15.94

^a Experiments were performed at least in duplicate.

^b SD: Standard deviations ranged from \pm 5% to \pm 10% of the indicated IC₅₀ values.

Table 5
Inhibition of LDHA by compound **18** and reference oxamate.^a

Compd	IC ₅₀ \pm SD ^b (mM)
	LDHA
18	0.01203
Oxamate	16.47

^a Experiments were performed at least in duplicate.

^b SD: Standard deviations ranged from \pm 5% to \pm 10% of the indicated IC₅₀ values.

Identification of new inhibitors using Virtual Screening technique

3.4) Enzymatic assay

By assessing the rate of NADH consumption at 340 nM after the addition of pyruvate by spectrophotometry, as previously described, they examined the capability of compounds 17 and 18 to block LDHA enzymatic activity⁽⁵⁶⁾.

Compound 18 showed a considerable inhibitory effect, with an IC₅₀ of 12.03 μ M (Table 5), in contrast to compound 17, which was unable to inhibit LDHA activity (Figure 17A). When compared to compound 18, the reference LDHA inhibitor sodium oxamate had an IC₅₀ of 16.47 mM, which was 1396 times lower. Treatment with compound 18 resulted in a dose-dependent decrease in lactate as measured by extracellular acidification rate (ECAR), consistent with the inhibition of pyruvate reduction and NADH oxidation, and a significant rise in NADH levels in Med1-MB cells (Figure 17B).

Identification of new inhibitors using Virtual Screening technique

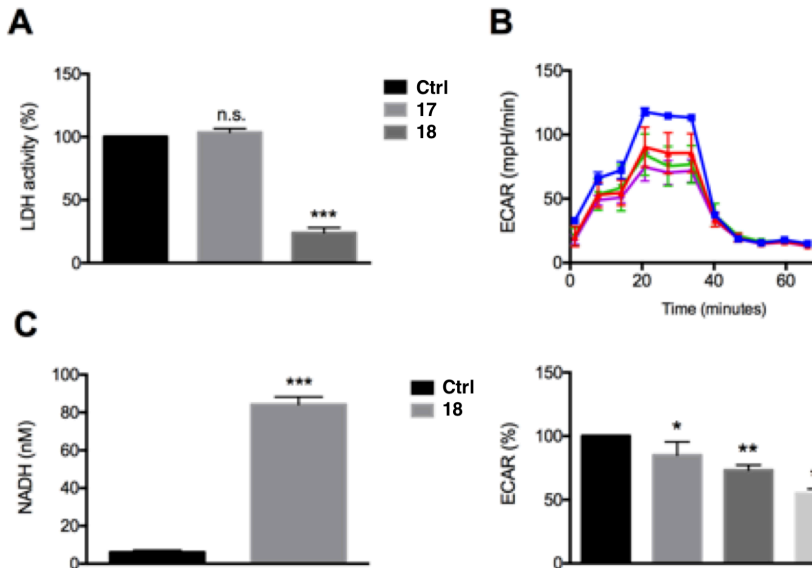


Figure 17. Measurement of LDH activity, glycolytic rate and NADH content after treatment with compound **18**. (A) Med1-MB cells were treated with or without compounds **17** (30 μ M) or **18** (80 μ M) and cellular extracts were prepared for determination of LDH enzyme activity. LDH activity is represented as % of the activity of LDH in the control group, which corresponds to 100%. (B) Cells were treated with increasing doses of compound **18**, as indicated, for 6 h. Analysis of glycolysis with the glycolysis rate assay test using the Seahorse extracellular flux analyzer revealed decreased levels of glycolysis (top) and ECAR (bottom) in compound **18**-treated cells. ECAR: extracellular acidification rate. Data are the mean \pm SD of six individual samples. (C) NADH content was determined in Med1-MB cells following treatment with compound **18** at its IC₅₀ value for 6 h. Data are represented as the mean \pm SD of three independent experiments, each performed in triplicate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. = not significant as determined by ANOVA and unpaired t test.

Identification of new inhibitors using Virtual Screening technique

3.5) Additional cancer cells

Then, we looked at how this recently discovered compound affected the growth of additional cancer cell lines. Pancreatic cancer (PANC-1), lung cancer (A549), and colorectal cancer (HCT116 and SW620) cell lines were examined, all of which have elevated glycolytic metabolism^(57,58,59). Molecule 18 considerably reduced the cell proliferation of all cancer cell lines examined, as shown in Figure 18A, indicating the potential therapeutic efficacy of this compound in a wide range of conditions.

3.6) Effect in HCT116 CRC cells lacking LDHA

In order to confirm the specificity of compound 18 even more, they examined its impact on HCT116 CRC cells deficient in LDHA. To achieve this, they generated an inducible LDHA-deficient cell line, using a lentiviral vectorexpressing shRNA targeting LDHA mRNA, under the control of a doxycycline-inducible promoter.

The knockdown was induced with 10 g/mL doxycycline for 24 hours after cells had been stably transduced with the lentiviral vector. Compound 18 was then applied to the cells for 24 hours. A robust 95% LDHA knockdown were achieved under these circumstances, as shown in Figure 18C, which was verified by quantitative RT-PCR. Notably, treatment of LDHA-deficient cells with compound 18 did not prevent cell proliferation, proving that the impact is only there when LDHA is present (Figure 18B). These findings show that compound 18 particularly suppresses LDHA activity and cancer cell growth.

3.7) Effect on apoptosis and autophagy

We assessed the cleavage of Poly (ADP-ribose) polymerase (PARP) and LC3B I conjugation to phosphatidylethanolamine

Identification of new inhibitors using Virtual Screening technique

(LC3B II), respectively, to establish whether the suppression of cell proliferation by compound 18 is connected to an increase in programmed cell death or an induction of autophagy. Compound 18 does not alone cause apoptosis or autophagy in HCT116 CRC cells, as evidenced by the fact that exposure to compound 18 in increasing concentrations did not significantly alter either PARP cleavage nor LC3B-I lipidation. (Figure 18D)

Identification of new inhibitors using Virtual Screening technique

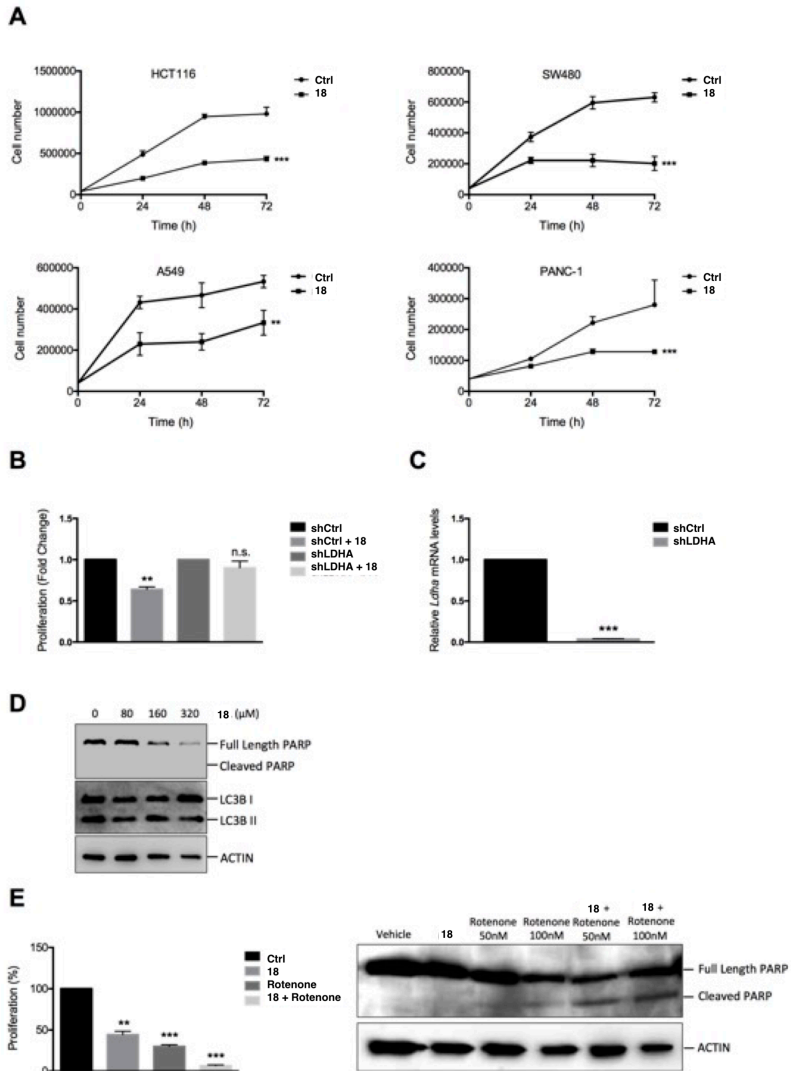


Figure 18. Effect of compound **18** on cell proliferation and apoptosis. (A) Growth curves in colorectal cancer (HCT116 and SW620), lung cancer (A549) and pancreatic cancer (PANC-1) cell lines treated with or without compound **18** (80 μ M) for the indicated time points. (B) Proliferation assay in HCT116 cells stably expressing shRNA targeting LDHA mRNA. Cells were pretreated with doxycycline (10 μ g/mL) to induce knockdown, and then treated with compound **18** at its IC₅₀ value for 24 h. (C) Knockdown levels of LDHA mRNA as described in (B) were determined by qPCR. (D)

Identification of new inhibitors using Virtual Screening technique

HCT116 cells were treated with compound **18** at the indicated concentrations for 48 h and cell lysates were analyzed by immunoblot for the indicated proteins. Actin was used as a loading control. (E) HCT116 cells were treated with compound **18** or rotenone at 50 or 100 nM alone or in combination for 24 h and cell proliferation was assessed. Lysates were analyzed by immunoblot for the indicated proteins. Data are represented as the mean \pm SD of three independent experiments, each performed in triplicate. **p < 0.01, ***p < 0.001 as determined by ANOVA and unpaired *t*-test.

3.8) Combined effect of compound **18** and complex I inhibition

Inhibiting glycolysis in cells with respiration defects has been shown to accelerate apoptosis by lowering ATP levels, phosphorylating BAD at Ser112, and relocating Bax to the mitochondria, according to earlier research⁽⁶⁰⁾. Because HCT116 cells do not have respiration defects, we hypothesized that compound **18** did not trigger apoptosis or autophagy because, upon LDHA inhibition, cells could transition their metabolism from glycolysis to mitochondrial respiration. We investigated the interaction of chemical **18** with the complex I and mitochondrial respiration inhibitor rotenone in order to test this hypothesis^(61,62). Treatment with rotenone alone resulted in an increase in PARP cleavage and an inhibition of cell proliferation, which is consistent with the reported action of rotenone on triggering apoptosis due to increased ROS generation⁽⁶¹⁾. Notably, the addition of rotenone and compound **18** together significantly increased the inhibition of cell proliferation and resulted in a significant cleavage of PARP (Figure 18E), showing that the combined inhibition of LDHA and Complex I result in robust activation of programmed cell death in these cancer cells. This robust activation of programmed cell death is most likely related to energy stress and excessive ROS production (Figure 19).

Identification of new inhibitors using Virtual Screening technique

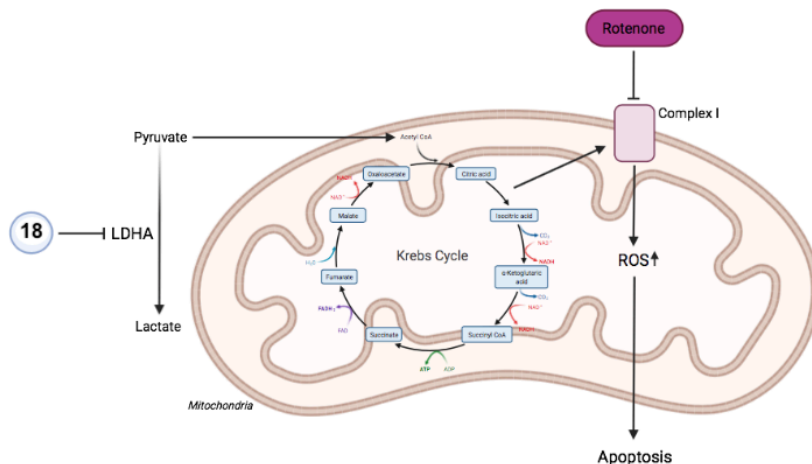


Figure 19. Schematic picture of the involvement of **18** and rotenone in apoptosis

3.9) Computational studies

3.9.1) Binding mode of compound **18**

The binding mode analyses revealed important interactions between compound **18** and the enzyme's catalytic site, including:

1. the sulfone oxygen atom forming H-bonds with Arg168 and Thr247 side chains;
2. the benzo[d]isothiazole dioxide bicyclic ring forming hydrophobic contacts with Ala237 and His192 and being involved in aromatic interactions with Tyr238;
3. the piperidine ring received hydrophobic stabilization by Asn137 side chain.
4. the terminal cyclopenta[c]pyridazine ring formed hydrophobic interactions with Pro138 and Asn99 side chains (Figure 20)

It is noteworthy that the nearest sulfonamide oxygen atom located at 3.4Å from the nicotinamide moiety to generate aromatic H-bond interactions.

Identification of new inhibitors using Virtual Screening technique

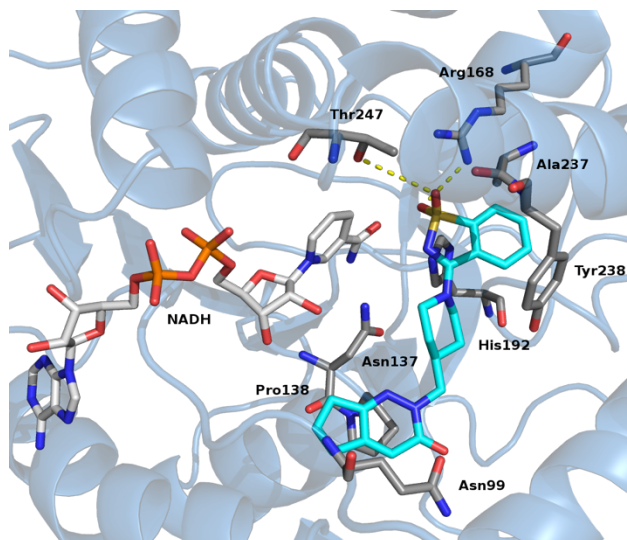


Figure 20. Proposed binding mode of **18**. LDHA is reported as cartoon; compound **18** is depicted as cyan stick, residues involved in interactions are reported as grey stick and NADH is represented as white stick. H-bonds contacts are reported as yellow dotted lines.

Then, this proposed binding mode of derivative 18 was compared with the co-crystallized structures binding of three LDHA inhibitors: pyrazoles **7** (5W8I, LDHA IC₅₀ = 29 μ M)⁽³⁵⁾ and **10** (6Q13, IC₅₀ = 0.04 μ M)⁽³⁷⁾ and dihydropyridinone **11** (4ZVV, IC₅₀ = 8.9 nM)⁽³⁸⁾. The results of the investigations demonstrated that derivative 18 and the pyrazole-containing compounds **7** and **10** had similar binding modes and occupied the same receptor sub-pockets. As opposed to 18, the dihydropyridinone derivative **11** showed a distinct binding mode that targeted various residues. However interactions with the catalytic residues Arg168, Thr247, Ala237, and Tyr238 remained the same. This investigation increased our confidence in the predicted binding mode and mechanism of action of derivative 18 (Figure 21 and 22).

Identification of new inhibitors using Virtual Screening technique

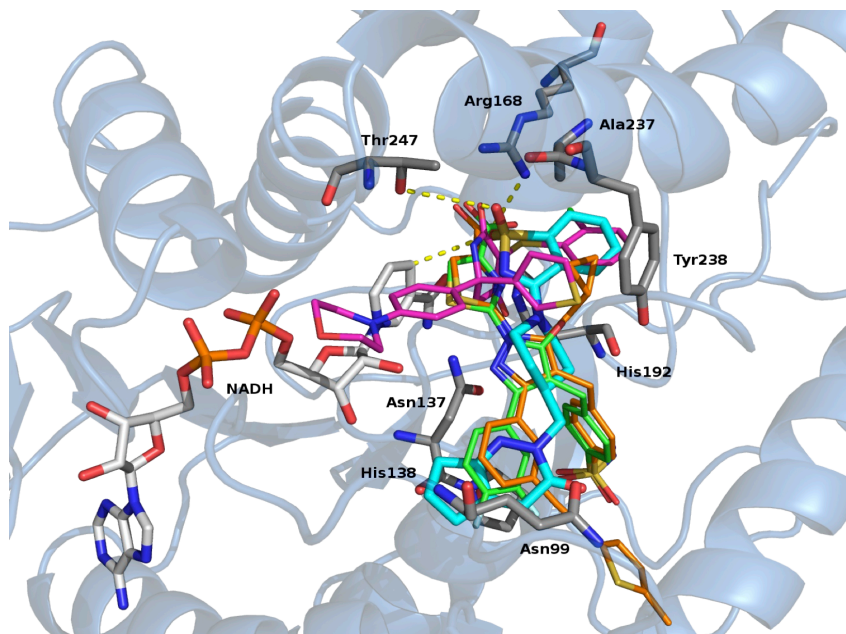


Figure 21 The proposed binding mode of derivative **18** (cyan) was superimposed with the crystallographic binding modes of pyrazole containing derivatives **7** (5W8I, green) ⁽³⁵⁾ and **10** (6Q13, orange) ⁽³⁷⁾ and dihydropyridinone **11** (4ZVV, magenta) ⁽³⁸⁾. The enzyme is depicted as light blue ribbon and NADH structure is reported in white. Residues involved in interactions with **18** are shown.

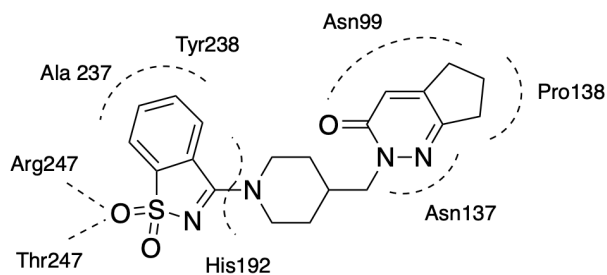


Figure 22. Summary of the binding interactions of compound **18** into the LDHA binding site

Identification of new inhibitors using Virtual Screening technique

3.9.2) Molecular dynamics simulations

Molecular dynamics analysis was used to examine the suggested binding mode of 18. The 18 had a steady binding mode, as evidenced by the 50 ns trajectory, and the majority of suggested interactions persisted throughout the simulation period. For example, the piperidine ring and the 1,1-dioxidobenzo[d]isothiazole ring both retained all of their connections, whereas the cyclopenta[c]pyridazin-3-one ring only kept its hydrophobic interactions with Asn99 (Figure 23).

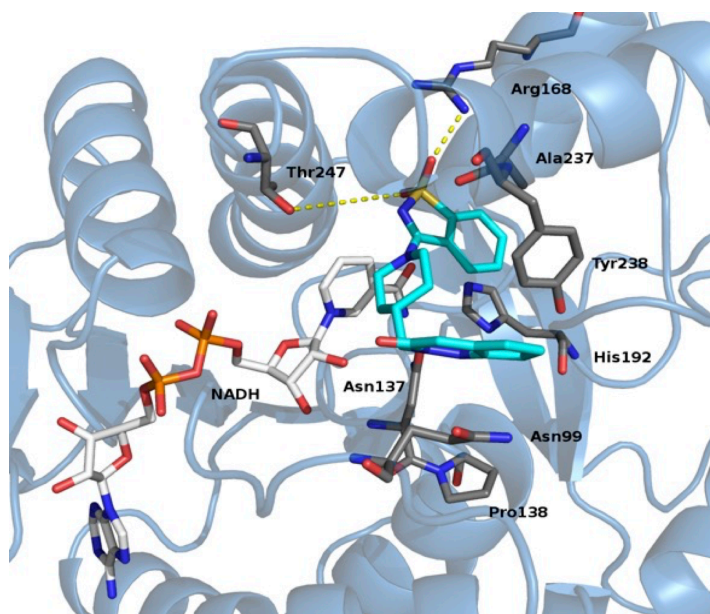


Figure 23. Snapshot of molecular dynamic simulations. LDHA is reported as cartoon, compound 18 is depicted as cyan stick, residues involved in interactions are reported as grey stick, and NADH is represented as white stick. H-bonds contacts are reported as yellow dotted lines.

The RMSD calculated over the whole simulation period demonstrated the general stability of the system, with the NADH and 18 RMSD values always being lower than 1 and 3, respectively (Figure 24). The RMSD calculation were obtained

Identification of new inhibitors using Virtual Screening technique

fitting the trajectories on the reference complex extracted from the last minimization step. NADH showed an RMSD close to 0 due to the applied restraint. These molecular dynamics analysis underlined the consistency of the predicted binding mode and indicated that the change in the cyclopenta[c]pyridazin-3-one ring would result in more advantageous binding interactions.

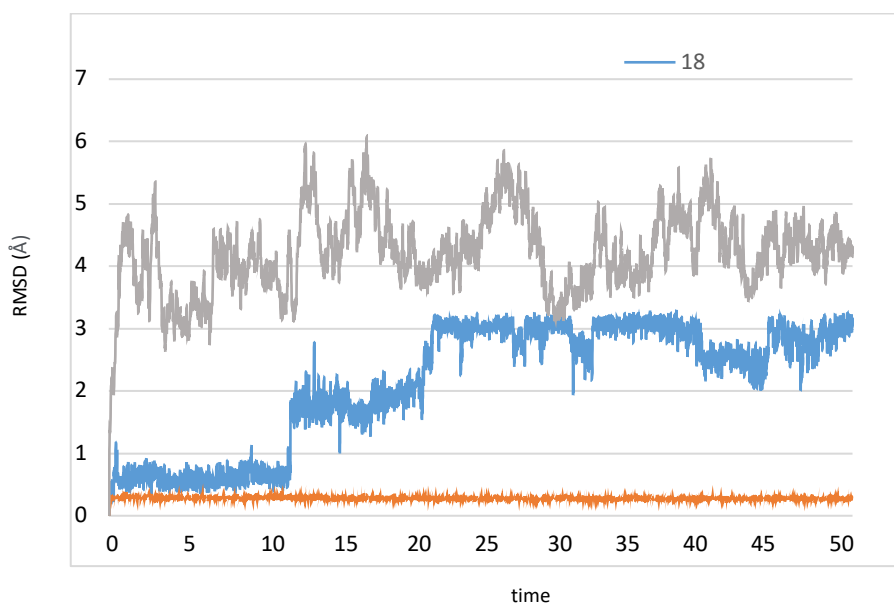


Figure 24. Calculated rmsd values during the whole simulation. The x axis reports the RMSD value in Å, the y axis reports the simulation time. NADH was always at lower distance of 1 Å; 18 was at lower distance of 3 Å; the jump of the distance values is related with the cyclopenta[c]pyridazine ring movements; the system (protein + NADH + 18) showed the higher values of RMSD, these higher distance values are mainly related with the C and N termini movements.

Identification of new inhibitors using Virtual Screening technique

3.9.3) Drug-like properties

We calculated drug-like properties of compound **18** using SwissADME server⁽⁶³⁾.

Compound **18** does not violate the Lipinski rule of five⁽⁵²⁾ and Veber⁽⁶⁴⁾ rule and was predicted to have a good bioavailability after oral administration (Table 6 and Figure 25 A). The 3/75 rule (logP >3 and topological polar surface area (PSA) < 75 Å²)⁽⁶⁵⁾ evaluated compound toxicity and displayed a low probability of in vivo toxicological effects (Figure 25 B).

Table 6. Drug like properties of compound **11**.

Comp	LogP ^a	MW ^b	LogSw ^c	tPSA ^d	HBA ^e	HBD ^f	Rot ^g	Lipinski ^h	Veber ⁱ	3/75 ^k
18	1.30	398.48	-3.25	93.01	5	0	3	0	0	Low

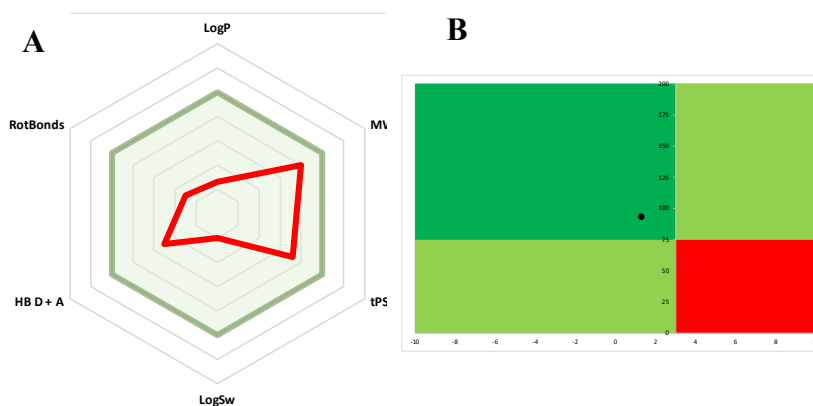


Figure 25. Estimated toxicity of compound **18**. (A) Radar Plot of drug-like properties. The light green colored zone represents the suitable physicochemical space for oral bioavailability. $-1 < \text{LogP} < 5$; $150 < \text{MW} < 500$; $20\text{\AA}^2 < \text{tPSA} < 130\text{\AA}^2$; $-10 < \text{LogSw} < 0$; $0 < \text{HB D+A} < 10$; $0 < \text{RotBonds} < 9$. The red line represents values for derivative 5. (B) 3/75 rule plot. Compounds locate in the red space are likely to cause toxicity; compounds locate in green space had a low likelihood of promiscuity and toxicity. The computed logP (horizontal axe) and tPSA (vertical axe) for derivative 5 (black circle) showed low likelihood of adverse effect and off-targets engagement.

Identification of new inhibitors using Virtual Screening technique

3.10) Conclusions

A novel, specific LDHA inhibitor 18 was found through structure-based virtual screening studies to have potent anticancer activity in the micromolar range in several cancer cell lines, including the aerobic glycolysis-dependent Med1-MB cell line, the CRC HCT116 and SW620, the lung cancer A549, and the pancreatic PANC-1 cancer cells. Compound 18 was 195 times more powerful than the reference inhibitor sodium oxamate at suppressing cell proliferation in the Med1-MB cell line, with an IC₅₀ of 81 M. With an IC₅₀ value that was 1396 times lower than sodium oxamate's, 18 at 100 M demonstrated substantial LDHA inhibition. As a result of LDHA inhibition, compound 18 consistently increased NADH content and decreased lactate levels in tumor cells. Compound 18 was unable to stop the proliferation of LDHA-deficient HCT116 cells, proving that the impact seen in cancer cells is caused by a particular suppression of LDHA activity. Additionally, when coupled with the complex I inhibitor rotenone, 18 demonstrated synergistic suppression of cell growth and activation of programmed cell death. When considered collectively, the findings presented here support the idea that compound 18 deserves more research as a starting point for the creation of LDH inhibitors and for innovative anticancer approaches based on the targeting of critical metabolic processes.

Di Magno L, Coluccia A, Bufano M, Ripa S, La Regina G, Nalli M, Di Pastena F, Canettieri G, Silvestri R, Frati L. Discovery of novel human lactate dehydrogenase inhibitors: Structure-based virtual screening studies and biological assessment. *Eur J Med Chem.* 2022 Oct 5;240:114605. doi: 10.1016/j.ejmech.2022.114605.

Identification of new inhibitors using Virtual Screening technique

3.11) References

1. WARBURG O. On the origin of cancer cells. *Science*. 1956 Feb 24;123(3191):309-14. doi: 10.1126/science.123.3191.309. PMID: 13298683.
2. Koppenol WH, Bounds PL, Dang CV. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*. 2011 May;11(5):325-37. doi: 10.1038/nrc3038. Epub 2011 Apr 14. Erratum in: *Nat Rev Cancer*. 2011 Aug;11(8):618. PMID: 21508971.
3. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*. 2009 May 22;324(5930):1029-33. doi: 10.1126/science.1160809. PMID: 19460998; PMCID: PMC2849637.
4. Fiume L, Manerba M, Vettraino M, Di Stefano G. Inhibition of lactate dehydrogenase activity as an approach to cancer therapy. *Future Med Chem*. 2014 Mar;6(4):429-45. doi: 10.4155/fmc.13.206. PMID: 24635523.
5. Zhang SL, He Y, Tam KY. Targeting cancer metabolism to develop human lactate dehydrogenase (hLDH)5 inhibitors. *Drug Discov Today*. 2018 Jul;23(7):1407-1415. doi: 10.1016/j.drudis.2018.05.014. Epub 2018 May 8. PMID: 29750903.
6. Farhana A, Lappin SL. Biochemistry, Lactate Dehydrogenase. 2022 May 8. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. PMID: 32491468.
7. Lu R, Jiang M, Chen Z, Xu X, Hu H, Zhao X, Gao X, Guo L. Lactate dehydrogenase 5 expression in Non-Hodgkin lymphoma is associated with the induced hypoxia regulated protein and poor prognosis. *PLoS One*. 2013 Sep 23;8(9):e74853. doi: 10.1371/journal.pone.0074853. PMID: 24086384; PMCID: PMC3781153.
8. Koukourakis MI, Giatromanolaki A, Simopoulos C, Polychronidis A, Sivridis E. Lactate dehydrogenase 5 (LDH5) relates to up-regulated hypoxia inducible factor pathway and

Identification of new inhibitors using Virtual Screening technique

- metastasis in colorectal cancer. *Clin Exp Metastasis*. 2005;22(1):25-30. doi: 10.1007/s10585-005-2343-7. PMID: 16132575.
9. Wang J, Wang H, Liu A, Fang C, Hao J, Wang Z. Lactate dehydrogenase A negatively regulated by miRNAs promotes aerobic glycolysis and is increased in colorectal cancer. *Oncotarget*. 2015 Aug 14;6(23):19456-68. doi: 10.18632/oncotarget.3318. PMID: 26062441; PMCID: PMC4637298.
 10. Zhuang L, Scolyer RA, Murali R, McCarthy SW, Zhang XD, Thompson JF, Hersey P. Lactate dehydrogenase 5 expression in melanoma increases with disease progression and is associated with expression of Bcl-XL and Mcl-1, but not Bcl-2 proteins. *Mod Pathol*. 2010 Jan;23(1):45-53. doi: 10.1038/modpathol.2009.129. Epub 2009 Oct 16. PMID: 19838163.
 11. Cui J, Shi M, Xie D, Wei D, Jia Z, Zheng S, Gao Y, Huang S, Xie K. FOXM1 promotes the warburg effect and pancreatic cancer progression via transactivation of LDHA expression. *Clin Cancer Res*. 2014 May 15;20(10):2595-606. doi: 10.1158/1078-0432.CCR-13-2407. Epub 2014 Mar 14. PMID: 24634381; PMCID: PMC4024335.
 12. Li XB, Gu JD, Zhou QH. Review of aerobic glycolysis and its key enzymes - new targets for lung cancer therapy. *Thorax Cancer*. 2015 Jan;6(1):17-24. doi: 10.1111/1759-7714.12148. Epub 2015 Jan 7. PMID: 26273330; PMCID: PMC4448463.
 13. Xie H, Hanai J, Ren JG, Kats L, Burgess K, Bhargava P, Signoretti S, Billiard J, Duffy KJ, Grant A, Wang X, Lorkiewicz PK, Schatzman S, Bousamra M 2nd, Lane AN, Higashi RM, Fan TW, Pandolfi PP, Sukhatme VP, Seth P. Targeting lactate dehydrogenase--a inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells. *Cell Metab*. 2014 May 6;19(5):795-809. doi: 10.1016/j.cmet.2014.03.003. Epub 2014 Apr 10. PMID: 24726384; PMCID: PMC4096909.

Identification of new inhibitors using Virtual Screening technique

14. Xian ZY, Liu JM, Chen QK, Chen HZ, Ye CJ, Xue J, Yang HQ, Li JL, Liu XF, Kuang SJ. Inhibition of LDHA suppresses tumor progression in prostate cancer. *Tumour Biol.* 2015 Sep;36(10):8093-100. doi: 10.1007/s13277-015-3540-x. Epub 2015 May 16. PMID: 25983002; PMCID: PMC4605959.
15. Kolev Y, Uetake H, Takagi Y, Sugihara K. Lactate dehydrogenase-5 (LDH-5) expression in human gastric cancer: association with hypoxia-inducible factor (HIF-1 α) pathway, angiogenic factors production and poor prognosis. *Ann Surg Oncol.* 2008 Aug;15(8):2336-44. doi: 10.1245/s10434-008-9955-5. Epub 2008 Jun 3. PMID: 18521687.
16. Giatromanolaki A, Sivridis E, Gatter KC, Turley H, Harris AL, Koukourakis MI; Tumour and Angiogenesis Research Group. Lactate dehydrogenase 5 (LDH-5) expression in endometrial cancer relates to the activated VEGF/VEGFR2(KDR) pathway and prognosis. *Gynecol Oncol.* 2006 Dec;103(3):912-8. doi: 10.1016/j.ygyno.2006.05.043. Epub 2006 Jul 11. PMID: 16837029.
17. Brand A, Singer K, Koehl GE, Kolitzus M, Schoenhammer G, Thiel A, Matos C, Bruss C, Klobuch S, Peter K, Kastenberger M, Bogdan C, Schleicher U, Mackensen A, Ullrich E, Fichtner-Feigl S, Kesselring R, Mack M, Ritter U, Schmid M, Blank C, Dettmer K, Oefner PJ, Hoffmann P, Walenta S, Geissler EK, Pouyssegur J, Villunger A, Steven A, Seliger B, Schreml S, Haferkamp S, Kohl E, Karrer S, Berneburg M, Herr W, Mueller-Klieser W, Renner K, Kreutz M. LDHA-Associated Lactic Acid Production Blunts Tumor Immunosurveillance by T and NK Cells. *Cell Metab.* 2016 Nov 8;24(5):657-671. doi: 10.1016/j.cmet.2016.08.011. Epub 2016 Sep 15. PMID: 27641098.
18. Rani R, Kumar V. Recent Update on Human Lactate Dehydrogenase Enzyme 5 (hLDH5) Inhibitors: A Promising Approach for Cancer Chemotherapy. *J Med Chem.* 2016 Jan 28;59(2):487-96. doi: 10.1021/acs.jmedchem.5b00168. Epub 2015 Sep 4. PMID: 26340601.

Identification of new inhibitors using Virtual Screening technique

19. Read JA, Winter VJ, Eszes CM, Sessions RB, Brady RL. Structural basis for altered activity of M- and H-isozyme forms of human lactate dehydrogenase. *Proteins*. 2001 May 1;43(2):175-85. doi: 10.1002/1097-0134(20010501)43:2<175::aid-prot1029>3.0.co;2-#. PMID: 11276087.
20. Thornburg JM, Nelson KK, Clem BF, Lane AN, Arumugam S, Simmons A, Eaton JW, Telang S, Chesney J. Targeting aspartate aminotransferase in breast cancer. *Breast Cancer Res*. 2008;10(5):R84. doi: 10.1186/bcr2154. Epub 2008 Oct 15. PMID: 18922152; PMCID: PMC2614520.
21. Shi Y, Pinto BM. Human lactate dehydrogenase a inhibitors: a molecular dynamics investigation. *PLoS One*. 2014 Jan 17;9(1):e86365. doi: 10.1371/journal.pone.0086365. PMID: 24466056; PMCID: PMC3895040.
22. GOLDBERG EB, NITOWSKY HM, COLOWICK SP. THE ROLE OF GLYCOLYSIS IN THE GROWTH OF TUMOR CELLS. IV. THE BASIS OF GLUCOSE TOXICITY IN OXAMATE-TREATED, CULTURED CELLS. *J Biol Chem*. 1965 Jul;240:2791-6. PMID: 14342296.
23. Vander Jagt DL, Deck LM, Royer RE. Gossypol: prototype of inhibitors targeted to dinucleotide folds. *Curr Med Chem*. 2000 Apr;7(4):479-98. doi: 10.2174/0929867003375119. PMID: 10702620.
24. Shelley MD, Hartley L, Fish RG, Groundwater P, Morgan JJG, Mort D, Mason M, Evans A. Stereo-specific cytotoxic effects of gossypol enantiomers and gossypolone in tumour cell lines. *Cancer Lett*. 1999;135:171-180
25. Tuszynski GP, Cossu G. Differential cytotoxic effect of gossypol on human melanoma, colon carcinoma, and other tissue culture cell lines. *Cancer Res*. 1984 Feb;44(2):768-71. PMID: 6581864.
26. Wu YW, Chik CL, Knazek RA. An in vitro and in vivo study of antitumor effects of gossypol on human SW-13 adrenocortical carcinoma. *Cancer Res*. 1989 Jul 15;49(14):3754-8. PMID: 2736516.

Identification of new inhibitors using Virtual Screening technique

27. Jaroszewski JW, Kaplan O, Cohen JS. Action of gossypol and rhodamine 123 on wild type and multidrug-resistant MCF-7 human breast cancer cells: 31P nuclear magnetic resonance and toxicity studies. *Cancer Res.* 1990 Nov 1;50(21):6936-43. PMID: 2208159.
28. Rajeshkumar NV, Dutta P, Yabuuchi S, de Wilde RF, Martinez GV, Le A, Kamphorst JJ, Rabinowitz JD, Jain SK, Hidalgo M, Dang CV, Gillies RJ, Maitra A. Therapeutic Targeting of the Warburg Effect in Pancreatic Cancer Relies on an Absence of p53 Function. *Cancer Res.* 2015 Aug 15;75(16):3355-64. doi: 10.1158/0008-5472.CAN-15-0108. Epub 2015 Jun 25. PMID: 26113084; PMCID: PMC4537812.
29. Xian ZY, Liu JM, Chen QK, Chen HZ, Ye CJ, Xue J, Yang HQ, Li JL, Liu XF, Kuang SJ. Inhibition of LDHA suppresses tumor progression in prostate cancer. *Tumour Biol.* 2015 Sep;36(10):8093-100. doi: 10.1007/s13277-015-3540-x. Epub 2015 May 16. PMID: 25983002; PMCID: PMC4605959.
30. Billiard J, Dennison JB, Briand J, Annan RS, Chai D, Colón M, Dodson CS, Gilbert SA, Greshock J, Jing J, Lu H, McSurdy-Freed JE, Orband-Miller LA, Mills GB, Quinn CJ, Schneck JL, Scott GF, Shaw AN, Waitt GM, Wooster RF, Duffy KJ. Quinoline 3-sulfonamides inhibit lactate dehydrogenase A and reverse aerobic glycolysis in cancer cells. *Cancer Metab.* 2013 Sep 6;1(1):19. doi: 10.1186/2049-3002-1-19. PMID: 24280423; PMCID: PMC4178217.
31. Farabegoli F, Vettraino M, Manerba M, Fiume L, Roberti M, Di Stefano G. Galloflavin, a new lactate dehydrogenase inhibitor, induces the death of human breast cancer cells with different glycolytic attitude by affecting distinct signaling pathways. *Eur J Pharm Sci.* 2012 Nov 20;47(4):729-38. doi: 10.1016/j.ejps.2012.08.012. Epub 2012 Aug 30. PMID: 22954722.
32. Manerba M, Vettraino M, Fiume L, Di Stefano G, Sartini A, Giacomini E, Buonfiglio R, Roberti M, Recanatini M. Galloflavin (CAS 568-80-9): a novel inhibitor of lactate dehydrogenase.

Identification of new inhibitors using Virtual Screening technique

- ChemMedChem. 2012 Feb 6;7(2):311-7. doi: 10.1002/cmdc.201100471. Epub 2011 Nov 4. PMID: 22052811.
33. Rani R, Granchi C. Bioactive heterocycles containing endocyclic N-hydroxy groups. *Eur J Med Chem.* 2015 Jun 5;97:505-24. doi: 10.1016/j.ejmech.2014.11.031. Epub 2014 Nov 18. PMID: 25466924; PMCID: PMC4436092.
 34. Granchi C, Roy S, Giacomelli C, Macchia M, Tuccinardi T, Martinelli A, Lanza M, Betti L, Giannaccini G, Lucacchini A, Funel N, León LG, Giovannetti E, Peters GJ, Palchadhuri R, Calvaresi EC, Hergenrother PJ, Minutolo F. Discovery of N-hydroxyindole-based inhibitors of human lactate dehydrogenase isoform A (LDH-A) as starvation agents against cancer cells. *J Med Chem.* 2011 Mar 24;54(6):1599-612. doi: 10.1021/jm101007q. Epub 2011 Feb 18. PMID: 21332213.
 35. Rai G, Brimacombe KR, Mott BT, Urban DJ, Hu X, Yang SM, Lee TD, Cheff DM, Kouznetsova J, Benavides GA, Pohida K, Kuenstner EJ, Luci DK, Lukacs CM, Davies DR, Dranow DM, Zhu H, Sulikowski G, Moore WJ, Stott GM, Flint AJ, Hall MD, Darley-Usmar VM, Neckers LM, Dang CV, Waterson AG, Simeonov A, Jadhav A, Maloney DJ. Discovery and Optimization of Potent, Cell-Active Pyrazole-Based Inhibitors of Lactate Dehydrogenase (LDH). *J Med Chem.* 2017 Nov 22;60(22):9184-9204. doi: 10.1021/acs.jmedchem.7b00941. Epub 2017 Nov 9. PMID: 29120638; PMCID: PMC5894102.
 36. Oshima N, Ishida R, Kishimoto S, Beebe K, Brender JR, Yamamoto K, Urban D, Rai G, Johnson MS, Benavides G, Squadrito GL, Crooks D, Jackson J, Joshi A, Mott BT, Shrimp JH, Moses MA, Lee MJ, Yuno A, Lee TD, Hu X, Anderson T, Kusewitt D, Hathaway HH, Jadhav A, Picard D, Trepel JB, Mitchell JB, Stott GM, Moore W, Simeonov A, Sklar LA, Norenberg JP, Linehan WM, Maloney DJ, Dang CV, Waterson AG, Hall M, Darley-Usmar VM, Krishna MC, Neckers LM. Dynamic Imaging of LDH Inhibition in Tumors Reveals Rapid In Vivo Metabolic Rewiring and Vulnerability to Combination Therapy. *Cell Rep.* 2020 Feb 11;30(6):1798-1810.e4. doi:

Identification of new inhibitors using Virtual Screening technique

- 10.1016/j.celrep.2020.01.039. PMID: 32049011; PMCID: PMC7039685.
37. Rai G, Urban DJ, Mott BT, Hu X, Yang SM, Benavides GA, Johnson MS, Squadrito GL, Brimacombe KR, Lee TD, Cheff DM, Zhu H, Henderson MJ, Pohida K, Sulikowski GA, Dranow DM, Kabir M, Shah P, Padilha E, Tao D, Fang Y, Christov PP, Kim K, Jana S, Muttill P, Anderson T, Kunda NK, Hathaway HJ, Kusewitt DF, Oshima N, Cherukuri M, Davies DR, Norenberg JP, Sklar LA, Moore WJ, Dang CV, Stott GM, Neckers L, Flint AJ, Darley-Usmar VM, Simeonov A, Waterson AG, Jadhav A, Hall MD, Maloney DJ. Pyrazole-Based Lactate Dehydrogenase Inhibitors with Optimized Cell Activity and Pharmacokinetic Properties. *J Med Chem.* 2020 Oct 8;63(19):10984-11011. doi: 10.1021/acs.jmedchem.0c00916. Epub 2020 Sep 27. PMID: 32902275; PMCID: PMC7830743.
 38. Purkey HE, Robarge K, Chen J, Chen Z, Corson LB, Ding CZ, DiPasquale AG, Dragovich PS, Eigenbrot C, Evangelista M, Fauber BP, Gao Z, Ge H, Hitz A, Ho Q, Labadie SS, Lai KW, Liu W, Liu Y, Li C, Ma S, Malek S, O'Brien T, Pang J, Peterson D, Salphati L, Sideris S, Ultsch M, Wei B, Yen I, Yue Q, Zhang H, Zhou A. Cell Active Hydroxylactam Inhibitors of Human Lactate Dehydrogenase with Oral Bioavailability in Mice. *ACS Med Chem Lett.* 2016 Aug 26;7(10):896-901. doi: 10.1021/acsmedchemlett.6b00190. PMID: 27774125; PMCID: PMC5066143.
 39. Zhou Y, Tao P, Wang M, Xu P, Lu W, Lei P, You Q. Development of novel human lactate dehydrogenase A inhibitors: High-throughput screening, synthesis, and biological evaluations. *Eur J Med Chem.* 2019 Sep 1;177:105-115. doi: 10.1016/j.ejmech.2019.05.033. Epub 2019 May 17. PMID: 31129449.
 40. Frati L. Reconsidering Otto Warburg's glycolytic shift: pyrimidine derivatives are effective for the treatment of tumors exerting aerobic glycolysis. *Panminerva Med.* 2022 Dec;64(4):567-568. doi: 10.23736/S0031-0808.22.04658-4. Epub 2022 Feb 4. PMID: 35119248.

Identification of new inhibitors using Virtual Screening technique

41. Italian Patent Office Patent N. 102021000022682 (September 1, 2021) PCT/IB2022/055143, date of receipt 01 June 2022
42. Dragovich PS, Fauber BP, Boggs J, Chen J, Corson LB, Ding CZ, Eigenbrot C, Ge H, Giannetti AM, Hunsaker T, Labadie S, Li C, Liu Y, Liu Y, Ma S, Malek S, Peterson D, Pitts KE, Purkey HE, Robarge K, Salphati L, Sideris S, Ultsch M, VanderPorten E, Wang J, Wei B, Xu Q, Yen I, Yue Q, Zhang H, Zhang X, Zhou A. Identification of substituted 3-hydroxy-2-mercaptocyclohex-2-enones as potent inhibitors of human lactate dehydrogenase. *Bioorg Med Chem Lett*. 2014 Aug 15;24(16):3764-71. doi: 10.1016/j.bmcl.2014.06.076. Epub 2014 Jul 1. PMID: 25037916.
43. Boudreau A, Purkey HE, Hitz A, Robarge K, Peterson D, Labadie S, Kwong M, Hong R, Gao M, Del Nagro C, Pusapati R, Ma S, Salphati L, Pang J, Zhou A, Lai T, Li Y, Chen Z, Wei B, Yen I, Sideris S, McClelland M, Firestein R, Corson L, Vanderbilt A, Williams S, Daemen A, Belvin M, Eigenbrot C, Jackson PK, Malek S, Hatzivassiliou G, Sampath D, Evangelista M, O'Brien T. Metabolic plasticity underpins innate and acquired resistance to LDHA inhibition. *Nat Chem Biol*. 2016 Oct;12(10):779-86. doi: 10.1038/nchembio.2143. Epub 2016 Aug 1. PMID: 27479743.
44. Fauber BP, Dragovich PS, Chen J, Corson LB, Ding CZ, Eigenbrot C, Giannetti AM, Hunsaker T, Labadie S, Liu Y, Liu Y, Malek S, Peterson D, Pitts K, Sideris S, Ultsch M, VanderPorten E, Wang J, Wei B, Yen I, Yue Q. Identification of 2-amino-5-aryl-pyrazines as inhibitors of human lactate dehydrogenase. *Bioorg Med Chem Lett*. 2013 Oct 15;23(20):5533-9. doi: 10.1016/j.bmcl.2013.08.060. Epub 2013 Aug 22. PMID: 24012183.
45. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013 Mar;27(3):221-34. doi: 10.1007/s10822-013-9644-8. Epub 2013 Apr 12. PMID: 23579614.
46. Korb O, Stützel T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf*

Identification of new inhibitors using Virtual Screening technique

- Model. 2009 Jan;49(1):84-96. doi: 10.1021/ci800298z. PMID: 19125657.
47. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des.* 2006 Oct-Nov;20(10-11):647-71. doi: 10.1007/s10822-006-9087-6. Epub 2006 Nov 24. PMID: 17124629.
 48. Dixon SL, Smondyrev AM, Rao SN. PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des.* 2006 May;67(5):370-2. doi: 10.1111/j.1747-0285.2006.00384.x. PMID: 16784462.
 49. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
 50. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem.* 2005 Dec;26(16):1668-88. doi: 10.1002/jcc.20290. PMID: 16200636; PMCID: PMC1989667.
 51. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput.* 2013 Jul 9;9(7):3084-95. doi: 10.1021/ct400341p. Epub 2013 Jun 25. PMID: 26583988.
 52. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2001 Mar 1;46(1-3):3-26. doi: 10.1016/s0169-409x(00)00129-0. PMID: 11259830.
 53. Di Magno L, Coni S, Di Marcotullio L, Canettieri G. Digging a hole under Hedgehog: downstream inhibition as an emerging anticancer strategy. *Biochim Biophys Acta.* 2015 Aug;1856(1):62-72. doi: 10.1016/j.bbcan.2015.06.003. Epub 2015 Jun 12. PMID: 26080084.
 54. Di Magno L, Manzi D, D'Amico D, Coni S, Macone A, Infante P, Di Marcotullio L, De Smaele E, Ferretti E, Screpanti I,

Identification of new inhibitors using Virtual Screening technique

- Agostinelli E, Gulino A, Canettieri G. Druggable glycolytic requirement for Hedgehog-dependent neuronal and medulloblastoma growth. *Cell Cycle*. 2014;13(21):3404-13. doi: 10.4161/15384101.2014.952973. PMID: 25485584; PMCID: PMC4613849.
55. Di Magno L, Manni S, Di Pastena F, Coni S, Macone A, Cairoli S, Sambucci M, Infante P, Moretti M, Petroni M, Nicoletti C, Capalbo C, De Smaele E, Di Marcotullio L, Giannini G, Battistini L, Goffredo BM, Iorio E, Agostinelli E, Maroder M, Canettieri G. Phenformin Inhibits Hedgehog-Dependent Tumor Growth through a Complex I-Independent Redox/Corepressor Module. *Cell Rep*. 2020 Feb 11;30(6):1735-1752.e7. doi: 10.1016/j.celrep.2020.01.024. PMID: 32049007.
56. Miskimins WK, Ahn HJ, Kim JY, Ryu S, Jung YS, Choi JY. Synergistic anti-cancer effect of phenformin and oxamate. *PLoS One*. 2014 Jan 21;9(1):e85576. doi: 10.1371/journal.pone.0085576. PMID: 24465604; PMCID: PMC3897486.
57. Sánchez-Aragó M, Chamorro M, Cuezva JM. Selection of cancer cells with repressed mitochondria triggers colon cancer progression. *Carcinogenesis*. 2010 Apr;31(4):567-76. doi: 10.1093/carcin/bgq012. Epub 2010 Jan 15. PMID: 20080835.
58. Chan AK, Bruce JI, Siriwardena AK. Glucose metabolic phenotype of pancreatic cancer. *World J Gastroenterol*. 2016 Mar 28;22(12):3471-85. doi: 10.3748/wjg.v22.i12.3471. PMID: 27022229; PMCID: PMC4806205.
59. Goodwin J, Neugent ML, Lee SY, Choe JH, Choi H, Jenkins DMR, Ruthenborg RJ, Robinson MW, Jeong JY, Wake M, Abe H, Takeda N, Endo H, Inoue M, Xuan Z, Yoo H, Chen M, Ahn JM, Minna JD, Helke KL, Singh PK, Shackelford DB, Kim JW. The distinct metabolic phenotype of lung squamous cell carcinoma defines selective vulnerability to glycolytic inhibition. *Nat Commun*. 2017 May 26;8:15503. doi: 10.1038/ncomms15503. PMID: 28548087; PMCID: PMC5458561.

Identification of new inhibitors using Virtual Screening technique

60. Xu RH, Pelicano H, Zhou Y, Carew JS, Feng L, Bhalla KN, Keating MJ, Huang P. Inhibition of glycolysis in cancer cells: a novel strategy to overcome drug resistance associated with mitochondrial respiratory defect and hypoxia. *Cancer Res.* 2005 Jan 15;65(2):613-21. PMID: 15695406.
61. Li N, Ragheb K, Lawler G, Sturgis J, Rajwa B, Melendez JA, Robinson JP. Mitochondrial complex I inhibitor rotenone induces apoptosis through enhancing mitochondrial reactive oxygen species production. *J Biol Chem.* 2003 Mar 7;278(10):8516-25. doi: 10.1074/jbc.M210432200. Epub 2002 Dec 20. PMID: 12496265.
62. Mohammed F, Gorla M, Bisoyi V, Tammineni P, Sepuri NBV. Rotenone-induced reactive oxygen species signal the recruitment of STAT3 to mitochondria. *FEBS Lett.* 2020 May;594(9):1403-1412. doi: 10.1002/1873-3468.13741. Epub 2020 Feb 18. PMID: 31981230.
63. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep.* 2017 Mar 3;7:42717. doi: 10.1038/srep42717. PMID: 28256516; PMCID: PMC5335600.
64. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem.* 2002 Jun 6;45(12):2615-23. doi: 10.1021/jm020017n. PMID: 12036371.
65. Hughes JD, Blagg J, Price DA, Bailey S, Decrescenzo GA, Devraj RV, Ellsworth E, Fobian YM, Gibbs ME, Gilles RW, Greene N, Huang E, Krieger-Burke T, Loesel J, Wager T, Whiteley L, Zhang Y. Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg Med Chem Lett.* 2008 Sep 1;18(17):4872-5. doi: 10.1016/j.bmcl.2008.07.071. Epub 2008 Jul 24. PMID: 18691886.

Application of molecular dynamics techniques to the study of ligand-protein binding and protein-protein binding

The use of molecular dynamics simulations is now an essential technique that let to observe relevant events for the rationalization of experimental data, indeed one of the aims of computer simulation is to “simulate” experiments to light up the invisible microscopic details and thus explain the results. Also, this technique could drive us though the investigations about rare events to notice.

Here I will show the topics where I applied these approaches: the first examine the possible binding site for the AM-001 allosteric inhibitor of EPAC1 using co-solvent molecular dynamics and the second uses another enhanced technique, accelerated molecular dynamics, to understand the CCRL2-chemerin main interactions.

1) Exploring Epac1 interactions with the allosteric inhibitor AM-001

One of the most prevalent and ubiquitous second messengers, cAMP is produced when a series of extracellular ligands connect to Gs protein-coupled receptors. Following this activation, membrane-bound adenylyl cyclases (ACs) produce cAMP from adenosine triphosphate⁽¹⁾. cAMP regulates fundamental physiologic processes including metabolism, secretion, calcium homeostasis, muscle contraction, cell fate, and gene transcription. Four major downstream effectors, including the protein kinase A (PKA), exchange proteins directly activated by cAMP proteins (Epac), cyclic nucleotide gated (CNG) ion channels, and Popeye domain-containing (POPDC) proteins, ensure the biological effects of cAMP^(2,3,4,5). Additionally, the cellular activities

Application of molecular dynamics simulations

and specificity of this second messenger are controlled by a molecular scaffold, A-kinase anchoring proteins (AKAPs), which sequester PDEs and cAMP effectors into distinct subcellular compartments⁽⁶⁾. cAMP effectors can be activated differently as a result of this compartmentalization. According to recent data, pathological cardiac remodeling and heart failure (HF), a significant cause of death globally, are influenced by abnormal cAMP signaling through dysregulation of cAMP compartmentalization⁽⁷⁾.

Guanine-nucleotide-exchange factors (GEFs) known as EPAC proteins were first identified 20 years ago by two separate research teams as new cAMP downstream effectors that regulated the PKA-independent activation of Rap1 and Rap2 small G-proteins.

Two different genes, which result in a variety of transcripts, encode the Epac proteins, Epac1 and Epac2⁽⁵⁾. Human kidney and heart have significant amounts of Epac1 mRNA expression, but Epac2 isoforms are mostly expressed in the brain and endocrine tissues^(8,9). Indeed, EPAC1 is ubiquitously expressed whereas EPAC2 and its splice variants are localized in the brain (EPAC2A), pancreatic cells (EPAC2B) testis and liver (EPAC2C)⁽¹⁰⁾. Epac1 and Epac2 are multidomain proteins and share a similar structural organization⁽¹¹⁾. They are featured by two main domains: an N-terminal regulatory region and a C-terminal catalytic domain. The amino-terminal regulation region contains a Disheveled/Egl-10/pleckstrin (DEP) domain followed by a cyclic nucleotide-binding domain (CNBD); in the absence of cAMP the N-terminal region serves as an auto-inhibitory domain to suppress Epac catalytic activity. The longer Epac2 isoform (Epac2A) has an extra CNB domain (CNBD-A) that binds cAMP with a low affinity, it is not necessary for Epac2 activation by cAMP, but it is involved in the localization of Epac2A in the subcellular space. CNBD prevents Rap

Application of molecular dynamics simulations

proteins from accessing the Epac catalytic site. When cAMP binds to Epac, it causes the protein to undergo significant structural changes and loses its auto-inhibitory function, activating Rap1⁽¹²⁾.

The catalytic region consists of a Ras exchange motif (REM), a Ras association domain (RA) and a CDC25 homology domain (CDC25HD). The CDC25HD catalyses GDP-GTP exchange for Rap, while the REM domain contributes to stabilize a catalytic helix, and the RA domain influences EPAC subcellular localization.

The Ras-exchange motif (REM) domain is involved in the stabilization of the active conformation of Epac⁽¹²⁾. The Ras-association (RA) domain is a protein interaction motif involved in the EPAC subcellular localization⁽¹²⁾ and the cell division cycle 25 homology domain (CDC25-HD) promotes the exchange of GDP for GTP on Rap GTPases. The relative orientations of the RR and CR with regard to one another are controlled by the regulatory CNBD at the RR C-terminus^(13,14). The two areas acquire a closed conformation in the absence of cAMP, limiting the Rap GTPases' ability to enter the CR and causing constitutive inhibition. The two areas take on an open (active) configuration when cAMP binds to the conserved CNBD, allowing the Rap substrate to connect with the CR and induce GDP-GTP exchange^(14,15,16). (Figure 26)

Application of molecular dynamics simulations

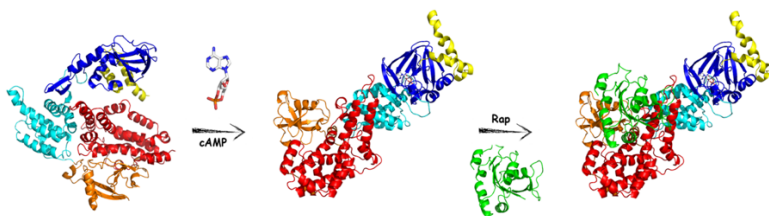


Fig 26 Graphical representation of EPAC activation upon cAMP binding. CNBD, DEP, REM, RA, and CDC25-HD domains are coloured in yellow, blue, cyan, orange and red respectively. cAMP is reported as white stick, RAP is reported as green cartoon. The cyclic nucleotide binding allows the regulatory domain to open leading the catalytic region exposed for binding of Rap

The majority of biological action of cAMP in cardiomyocytes have traditionally been attributed to PKA, which facilitates the acute effect of the β -adrenergic receptor (β -AR) on heart contractility⁽¹⁷⁾. Although β -AR activation is a crucial component of the fight-or-flight response in the heart, prolonged activation, as seen in heart failure (HF), leads to pathological cardiac remodeling, contractile failure, and arrhythmia⁽¹⁸⁾. The negative effects of prolonged β -AR stimulation are consistent with the finding that plasma norepinephrine levels in HF patients is correlated with the degree of cardiac dysfunction and mortality.

Interestingly, new research demonstrates that genetic or pharmacological suppression of EPAC1 improves cardiac function by preventing myocardial hypertrophy and fibrosis induced by persistent β -AR activation⁽¹⁹⁾. Consequently, focusing on EPAC1 may offer therapeutic advantages for heart disorders^(19,20).

The research for Epac-selective pharmacological modulators was motivated by the data showing that Epac proteins are engaged in numerous important pathophysiological processes in different tissues (agonists and inhibitors). These compounds are expected not to

Application of molecular dynamics simulations

affect the cAMP/PKA-dependent pathway, and, ideally, be able to distinguish between the two Epac isoforms.

These compounds can be categorized as cAMP analogues and non-nucleotidic small molecules based on their chemical structure. The main shortcomings of reported modulators were their inability to distinguish between PKA or EPAC isoforms, as well as their inadequate drug-like properties. Also, the orthosteric cAMP site is abundantly conserved in cAMP-regulated protein, giving difficulties concerning selectivity⁽²¹⁾. Finding a non-competitive allosteric modulator that is anticipated to have less risks of cross-reactivity with other receptors that signal via cAMP, such as PKA or CNG ion channels, may be the solution to this problem⁽²²⁾. Only CE3F4^(23,24) and, to our knowledge, AM-001⁽¹⁹⁾ have been shown to selectively inhibit EPAC1.

The small molecule AM-001 is a thieno[2,3-b]pyridine derivative that inhibits EPAC1 specifically and non-competitively⁽¹⁹⁾. During long-term β -adrenergic receptor activation, AM-001 reduces heart hypertrophy, inflammation, and fibrosis and enhances cardiac function. After mouse myocardial ischaemia/reperfusion injury, AM-001 lowers the infarct size in in-vivo tests⁽¹⁹⁾.

At this point we tried to understand the possible allosteric binding site of the AM-001 derivative since there is no data regarding its location. Considering that the biological evaluation showed a non-competitive mechanism of action and given the critical involvement of the hinge region of the EPAC in the close to open conformation shift, this region has recently been found to be a druggable area^(25,26).

Both catalytic and allosteric sites can be highlighted very effectively using the cosolvent molecular dynamics (CMD) approach⁽²⁷⁾. The multiple solvent crystal structures (MSCS) methodology⁽²⁸⁾ served as inspiration for this technique. The MSCS is based on solving protein

Application of molecular dynamics simulations

crystal structures while using a variety of organic cosolvents. It was discovered that there was a strong correlation between the overlapping sites of several cosolvents and biologically significant regions⁽²⁷⁾. The CMD might be seen as the in-silico analog of MSCS experiments. In order to identify the regions where organic solvents bind more preferentially, the CMD approach simulates the dynamics of a protein being solvated by a solution of organic solvents and water.

In our research, we used CMD to search for the EPAC1 allosteric binding site. As a result, the cosolvent occupancy maps for the EPAC1 active and inactive conformations were created and examined after the EPAC1 active and inactive conformations were solvated with a solution of 20% (w/w) isopropanol, ethanol, and dimethylsulfoxide in water. ISO is the most used solvent for CMD as it is able to interact with hydrophobic protein sites⁽²⁹⁾. Due to their relative abilities as hydrogen-bond acceptors and donors, DMSO and ETOH were selected^(30,31). Additionally, the cosolvents were selected to be entirely water miscible to prevent the phenomenon of molecules aggregating and the concentration was set to prevent protein denaturation⁽³²⁾.

By using homology from the EPAC2 crystal structure, models of the active and inactive EPAC1 conformations were created. The RMSD of the C α backbone was calculated to determine whether the unfolding phenomenon did not have an impact on the protein during the simulations⁽³³⁾. The RMSD results demonstrated that the cosolvents molecules did not interfere with protein folding (Figures 27 A and B).

Application of molecular dynamics simulations

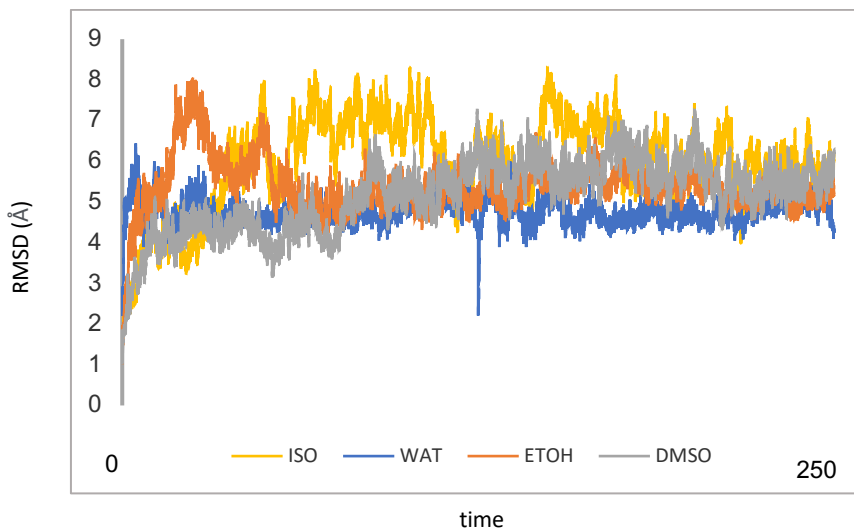


Figure 27. (A) RMSD of Ca atom of Epac1 inactive conformation simulations. ISO is for isopropanol, Wat is for water, ETOH is for ethanol and DMSO is for dimethyl sulfoxide.

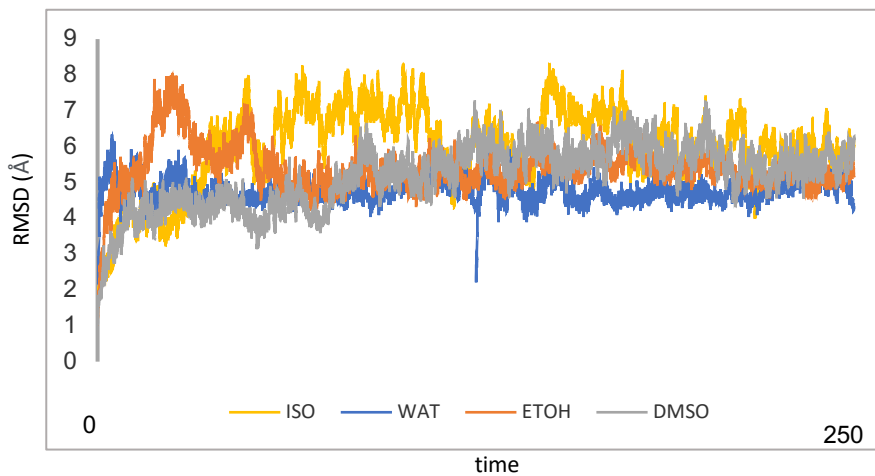


Figure 27. (B) RMSD of Ca atom of Epac1 inactive conformation simulations. ISO is for isopropanol, Wat is for water, ETOH is for ethanol and DMSO is for dimethyl sulfoxide.

Application of molecular dynamics simulations

The cosolvent occupancy maps were determined by analyzing the simulation's trajectory. The pockets where the cosolvent molecules were found most frequently are shown on each map. Protein and cosolvent molecule interactions were not considered. To make comprehension of the maps easier, a size cut-off was used to remove the smaller disconnected portion of the maps. As a result, the regions of the proteins where the cosolvent occupancy maps could be superimposed represented the locations that would be ideal for partner binding.

1.1) Materials and methods

1.1)1. EPAC structures

The EPAC structure was downloaded from the PDB data bank (<http://www.rcsb.org/>), inactive state: PDB code 2BYV⁽³⁴⁾; active state PDB code 3CF6⁽¹²⁾.

Homology model were carried out by homology model PRIME⁽³⁵⁾ of MAESTRO⁽³⁶⁾. The obtained models were examined to avoid steric clashes, bond length deviations and angles, etc., by using the Ramachandran plot and the Protein Preparation Wizard⁽³⁷⁾. The missing loops were obtained by Raptor X^(38,39). The primary sequence of EPAC1 and EPAC2 was downloaded by UniProtKB (<https://www.uniprot.org>) code: O95398NCBI and UniProtKB code: Q8WZA2 respectively.

1.1)2. Molecular dynamics

Molecular dynamics was performed with the AMBER 12 suite⁽⁴⁰⁾. For simulation in water the protein was solvated in a periodic octahedron simulation box using TIP3P water molecules, providing a minimum of 10 Å of water between the protein surface and any periodic box edge. The organic cosolvent/water box were prepared by packmol by fixing a 20% w/w ratio⁽⁴¹⁾. Parameters for cosolvents were obtained by pyMDMix.

Application of molecular dynamics simulations

(<http://mdmix.sourceforge.net>) Then, ions were added to neutralize the charge of the total system. The solvents molecules and the ions were energy-minimized keeping the coordinates of the protein-ligand complex fixed (1000 cycle), and then the whole system was minimized (5000 cycle). Following minimization, the entire system was heated to 298 K (20 ps). The production simulation was conducted at 298 K with constant pressure and periodic boundary condition. Shake bond length condition was used ($ntc = 2$). Production was carried out on GeForce gtx780 gpu. The cosolvent occupancy maps were computed with trajectories length of 100 ns and 250 ns. The obtained maps were similar regardless the length of the trajectories.

The production length of the simulations to evaluate the AM-001 stability at the studied sites was 10 ns. The production time to run PCA was 200 ns. All simulations were repeated five times. The occupancy maps were calculated by Chimera Volume viewer⁽⁴²⁾. Chimera we also used for the representation of the maps by volume viewer range 50%, and for the size filtering by the Hide Dust tool size 5.0⁽⁴³⁾. Compounds were parametrized by Antechamber^(44,45) using BCC charges. Trajectories analysis were carried out by cpptraj program⁽⁴⁶⁾.

1.1)3. PCA

The PCA analyses were carried out by cpptraj program⁽⁴⁶⁾. The trajectory was firstly superimposed to an average structure obtained by cpptraj. Then, the trajectory of the α -carbon was extracted, and the covariance matrix calculated. The diagonalization of the covariance matrix generated a set of eigenvectors that gave a vectorial description of the motion. The computed eigenvectors and eigenvalues were used to draw the porcupine plot by means of the Normal Mode Wizard module (NMWiz)^(47,48) of the VMD gui⁽⁴⁹⁾.

Application of molecular dynamics simulations

1.1)4. Molecular docking

Ligand structures were built with Maestro Schrodinger⁽³⁶⁾ and minimized using the OPLS3 force field until a rmsd gradient of 0.05 kcal/(mol Å) was reached. The docking simulations were performed using Gold⁽⁵⁶⁾, Plants⁽⁵⁷⁾ and Autodock⁽⁵⁵⁾. We set a binding lattice of radius or side large enough to cover the whole studied maps, then all default settings were used. The Docking of the AM-001 analogues were carried out by Plants⁽⁵⁷⁾, using as receptor a representative structure obtained by the trajectory of the EPAC1/posa3_6. The representative structure was the average structure extracted by cpptraj⁽⁴⁶⁾ after a stable rmsd was reached. The poses binding free energy was calculated by MAESTRO prime MMGB-SA module⁽⁶⁴⁾. The pictures reported in the manuscript were done with Pymol⁽⁶³⁾.

1.2) Analyses of the active conformation

In order to evaluate the effectiveness of the approach, the active conformation of the enzyme was primarily investigated. BRET tests were previously used to establish if AM-001 inhibited EPAC1⁽²³⁾. Indeed, the experiment determined the degree of EPAC1 activation, or the change from an inactive to active state of the enzyme following antagonist binding. Cosolvent occupancy maps were generated using 100 ns (Figures 28, 29) and 250 ns trajectory times.

Application of molecular dynamics simulations

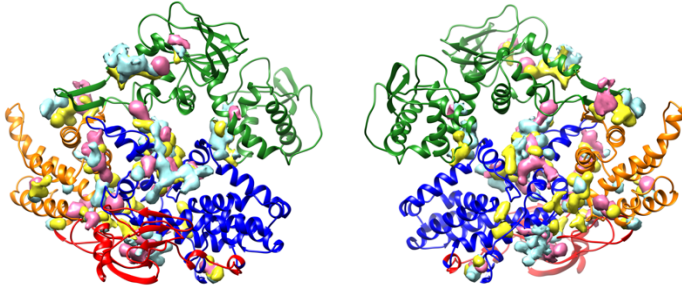


Figure 28. Epac1 Csolvents occupancy maps for Active conformation. Epac1 is reported as cartoon colored by domains: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. Yellow maps are for ETA; cyan for ISO and pink for DMSO. The pictures are rotated of 180 degrees.

Regardless of the length of the trajectories, the occupancy maps studies for the EPAC1 active conformation resulted in the discovery of four intriguing zones labeled as regions from 1 to 4 (Figure 29).

Application of molecular dynamics simulations

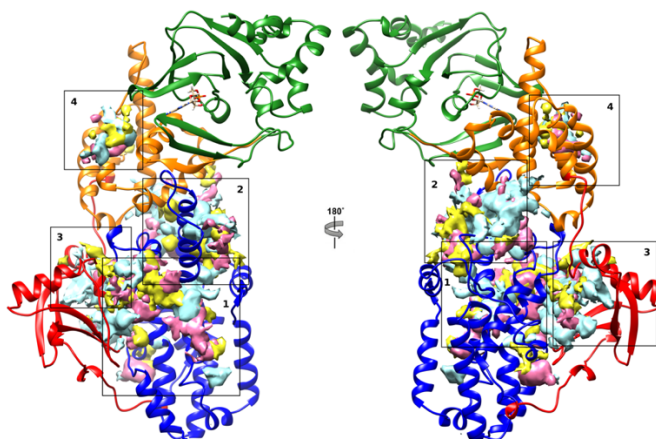


Figure 29 Epac1 Active conformation cosolvent occupancy maps selected areas. Epac1 is reported as cartoon: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. Yellow maps are for ETA; cyan for ISO and pink for DMSO

The Rap binding site corresponded to areas 1 and 2, which were situated at the CDC25-HD domain⁽¹²⁾. Indeed, the areas 1 and 2 were occupied by this domain, as demonstrated by the superimposition EPAC2 crystal structure, containing the Rap-interacting domain, with the EPAC1 model (PDB code: 4MGI)⁽⁵⁰⁾. We observed that the area 1 involved 4 α helices of the CDC25-HD namely α 1 (671–680), α 3 (708–713), α 5 (750–756) and α 9 (839–844). This area bound the β 4 and the loop connecting β 4 to α 3 of the Rap domain. (Figure 30). The area 2 involved the CDC25 Helical Hairpin (α 9 and α 10 from 823 to 854) that accommodated the loop between α 1 and β 2 of Rap⁽⁵¹⁾ (Figure 30). A possible interaction between EPAC1 and RanPB2 was also suggested to occur in region 2⁽⁵²⁾. The site 3 contained a loop that was not present in EPAC2 and was situated at the interface between the CDC25-HD and RA domains. A portion of the loop (550–557), the CDC25-HD α 1 (678–682) and the β 1 (558–562) of RA domain were all included in this wide surface. This area

Application of molecular dynamics simulations

was reported to be suitable for a protein-protein interaction⁽⁵³⁾.

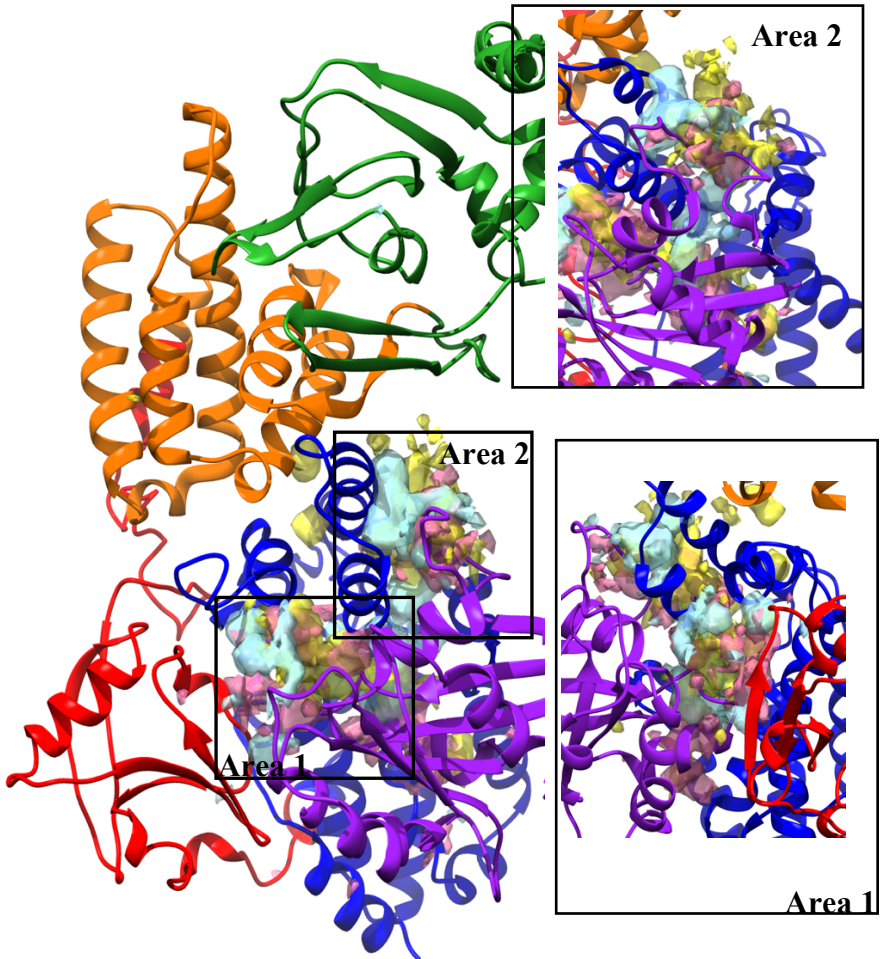


Figure 30. Epac1 active conformation with Rap domain. Epac1 is reported as cartoon colored by domains: CNBD and DEP green; REM orange; RA red and CDC25-HD blue. Rap domain is reported as purple cartoon. Yellow maps are for ETA; cyan for ISO and pink for DMSO. cAMP was reported as brown stick.

The area 4 was located at the REM domain. It comprised the $\alpha 1$ (394–398) and $\alpha 3$ (434–440) helices of REM and a

Application of molecular dynamics simulations

small loop of CNBD (294–299). This zone was found to be suitable for partners binding by FTPmap experiments⁽²⁶⁾.

EPAC1 was regulated and/or influenced in the subcellular localization by various proteins with unclear binding pockets⁽⁵²⁾. Indeed, it was reported that the residues 838–881 were involved in the compartmentalization of EPAC1, but it was not clarified how these residues drove this activity⁽⁵³⁾. Furthermore, tubulin was also reported as a direct partner of EPAC1, but its binding site was not yet identified⁽⁵⁴⁾. Our analyses of the occupancy maps revealed that specific regions were appropriate for partner protein binding and were compatible with the existing structural information, proving the suitability of the method.

1.3) Occupancy maps analysis of inactive conformation

The EPAC1 auto-inhibited form (inactive conformation) was next submitted to the CMD procedure to locate a potential AM-001 binding site. The cosolvent occupancy maps were generated using 100 ns and 250 ns trajectory calculations (Figure 31). The evaluations of the cosolvent occupancy maps did not reveal any appreciable variations depending on the length of the trajectories and help in the identification of several suitable locations (areas). The proposed mechanism of cAMP-induced EPAC1 activation involves EPAC1 regulatory domain that moves away from the catalytic domain. The lid (first -strand of REM 384-391 and the tip of CDC25-HD HP832-837) and the hinge helix play a fundamental role in this transition (CNBD 299–308). So, this cAMP-dependent conformational change should be impaired as a result of AM-001 binding to its allosteric pocket⁽²⁶⁾. These results were utilized to exclude the area surrounding the CDC25-

Application of molecular dynamics simulations

HD and RA domains. Indeed, the regions seem to be too far away from the lid and hinge helix to have any effect on the transition from an inactive to an active conformation state. This finding led us to manage 7 areas (Figure 31).

The cAMP binding pocket was in Area 1⁽¹²⁾. The identification of the cAMP site stressed the validity of the method used.

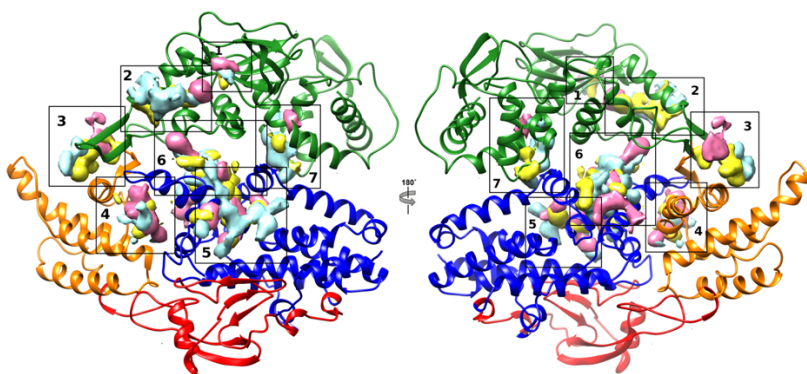


Figure 31 Epac1 Inactive conformation cosolvent occupancy maps selected zones. Epac1 inactive conformation is reported as cartoon: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. Yellow maps are for ETA; cyan for ISO and pink for DMSO

Docking and molecular dynamics were used to inspect the remaining locations. Area 2 was located at the CNBD near the catalytic site, Near the catalytic site, area 2 was located at the CNBD, which included CNBD 3 and the following loops (245-250), $\alpha 4$ (251-260), and $\alpha 6$ (1-6). (344-350). Area 3 was located at the interface between CNBD and REM domains. REM $\beta 1$ (387-392) and $\alpha 1$ (398-402) together with the CNBD $\beta 9$ (353-359) and $\beta 10$ (363-368) surround this area. Area 4 is featured by the HP loop of CDC25-HD and the REM $\alpha 2$ (423-430). The only area completely located at the CDC25-HD was Area 5, which

Application of molecular dynamics simulations

included $\alpha 2$ (689–692), $\alpha 3$ (704–708), $\alpha 5$ (761–765) and $\alpha 9$ (815–827). The area 6 was at the interface of CDCD25-HD and REM. The area was defined by the $\alpha 8$ (800–805), $\alpha 9$ (817–822) and $\alpha 10$ (844–851) of CDC25-HD and the loop between $\alpha 2$ and $\alpha 3$ (247–251) of the CNDB. Area 7 was found between the CDC25-HD and CNDB domains, and it included the CDC25-HD's helix $\alpha 8$ (793–801) and the long helix (206–216) connecting CNDB to the Dep domain (Figure 32).

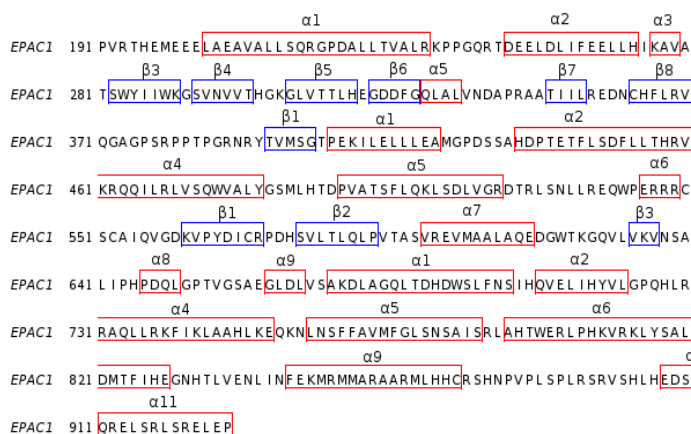


Figure 32. Epac1 sequence and secondary structure. EPAC1 UniProtKB code: O95398. The DEP domain is not included in the sequence secondary assignment. CNBD residues 191-380; REM residues 381-525; RA residues 526-661; CDC25-HD residues 662-923. For each domain the number of the α and β begins from 1.

Thus, AM-001 was docked to each of the mentioned areas. The docking investigations were conducted using Autodock⁽⁵⁵⁾, Gold⁽⁵⁶⁾ and Plants⁽⁵⁷⁾.

Among the docking software, at least one consistent AM-001 binding pose (RMSD < 1.5 Å) for each location was chosen and submitted to molecular dynamic simulations (10 ns).

Application of molecular dynamics simulations

The analysis of the molecular dynamic trajectories revealed that the ligand went out of the pocket early in the simulation for areas 2, 3, 4, and 7. Indeed, these sites were evidently solvent exposed, and it was assumed that these sites were more suitable as protein-protein interaction pockets⁽⁵⁸⁾ than as a small molecule binding sites⁽⁵⁹⁾.

Managing area 5 and 6, the analysis of the trajectory revealed two stable ($\text{rmsf} < 1.0 \text{ \AA}$) binding modes for area 5 (denoted as pose4_5 and pose9_5) (Figures 33, 34) and one for area 6 (denoted as pose3_6).

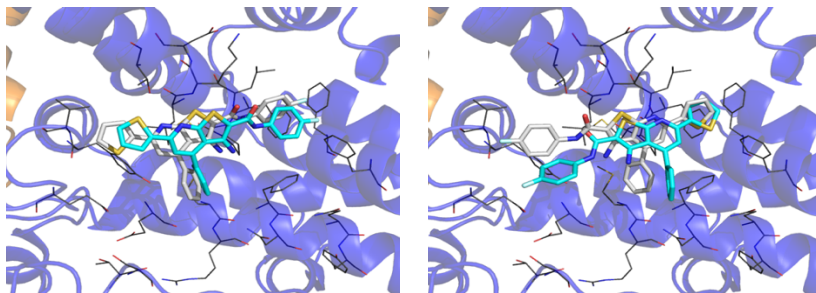


Figure 33. Epac1 Area5 pose4_5 and pose9_5 binding modes. Left panel pose4_5; right panel pose9_5. Docking proposed binding mode (white) and molecular dynamic trajectories snapshots (cyan) of AM-001 at Epac1 Area5. The enzyme is reported as cartoon. The domains reported in the picture are, REM orange and CDC25-HD blue. Residues of the pocket are reported as grey lines.

Application of molecular dynamics simulations

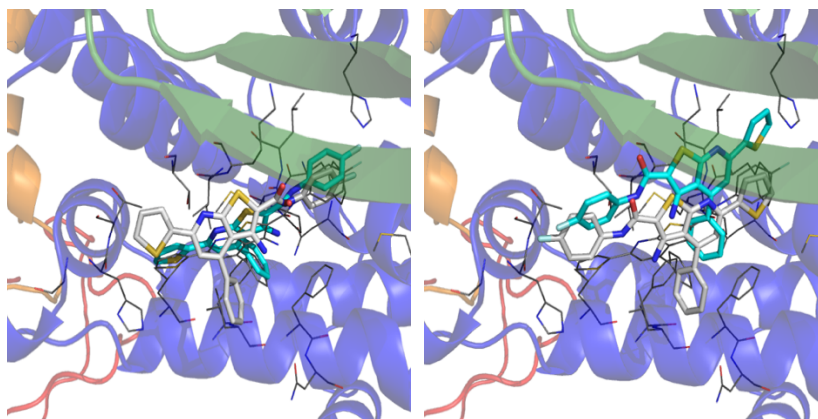


Figure 34 Epac2 Area5 pose4_5 and pose9_5 binding mode. Left panel pose4_5; right panel pose9_5. Docking proposed binding mode (white) and molecular dynamic trajectories snapshots (cyan) of AM-001 at Epac2 Area5. The enzyme is reported as cartoon. The domains reported in the picture are, CNBD and DEP green; REM orange; RA red and CDC25-HD blue. Residues of the pocket are reported as grey lines

Then, we investigated whether the three chosen binding modes may be appropriate for EPAC2 because AM-001 predominantly bound to the EPAC1 isoforms ($IC_{50}EPAC1$ 48.5 μ M) and was ineffective in inhibiting EPAC2 activity ($IC_{50}EPAC2 > 1000$ μ M)⁽¹⁹⁾. So, we ran several molecular dynamics simulations of the three binding modes placed in the corresponding pockets of EPAC2.

We calculated the binding free energy^(60,61) and we compared the trajectories with those resulting from the EPAC1 simulations. The trajectories inspection showed that the poses for area 5 were stable for EPAC2 (pose4_5 rmsf 0.49 Å pose9_5 rmsf 0.39 Å) as observed for EPAC1 (pose4_5 rmsf 0.41 Å pose9_5 rmsf 0.60 Å). The computed binding free energies for EPAC1 and EPAC2 were very similar (EPAC1 ΔG pose4_5 = -53.9, pose9_5 = -56.1; EPAC2 ΔG pose4_5 = -63.3 pose9_5 = -53.1). The difference was just 3 Kcal/mol for pose9_5, and a difference of 10 Kcal/mol for pose4_5 with the best ΔG

Application of molecular dynamics simulations

value for EPAC2. These findings would suggest that area 5 is not suitable for binding of AM-001 as a selective allosteric inhibitor of EPAC1. However, this site might correlate to a biologically significant area for both EPAC isoforms' enzymatic activity. The area 6 pose underwent the same investigation once more. A degree of instability was evident from the trajectory inspection. The compound moved toward the solvent, showing a rmsd higher than 3 Å and a calculated ΔG of -45.9 kcal/mol. The complex EPAC1/pose3 6 trajectory was examined in a similar manner. The AM-001 pose displayed overall stability with a rmsd of 1.19 Å and an estimated ΔG of -65.1 kcal/mol, favoring the EPAC1 binding with about 20 kcal/mol.

According to our findings, region 6 would be most suited for the allosteric inhibition of EPAC1. To verify our hypothesis, we ran a 200-ns molecular dynamic simulation and tested whether this binding site could match the known SAR profile of the chemical analogues (Table 7).

The trajectory analysis demonstrated the stability of the chosen binding mode and allowed us to identify a series of pharmacophoric interactions: the fluorinated phenyl group interacted with the R801 side chain through a pi-cation interaction and hydrophobic contacts with L235 and M; the primary amine moiety and the amidic nitrogen atom were involved in H-bond with D234 side chain; the pyridine thiophene fused ring had hydrophobic contacts with F237, R255, I825 and M844 side chains; the unsubstituted phenyl ring was trapped by pi-cation contacts with R847 and R850; the thiophene ring lay in a hydrophobic pocket mainly formed by R377, P378, N260 and N838 side chains; additionally, the thiophen aromatic ring behaved as weak H-bond acceptor⁽⁶²⁾ for the asparagine side chain amide (Figure 35).

Application of molecular dynamics simulations

The stability of these interactions throughout the simulation time supported the validity of the suggested binding site.

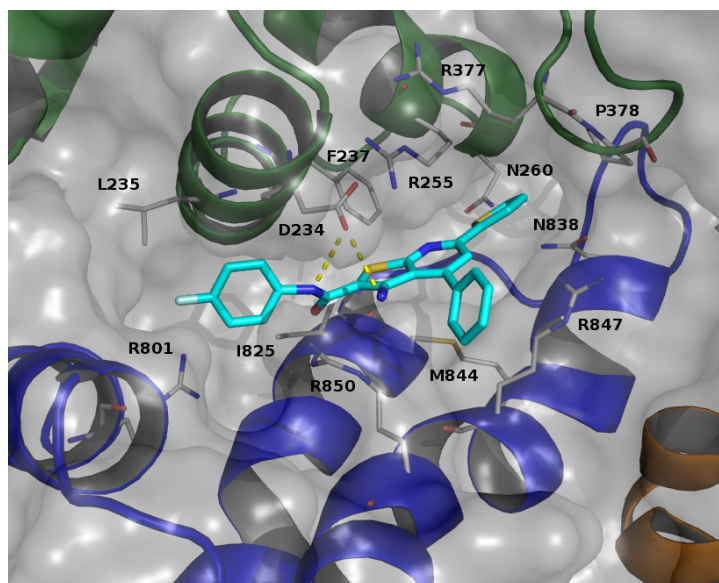


Figure 35. Proposed binding mode for AM-001 (cyan). Epac1 is reported as cartoon: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. Surface is reported in grey. Residues involved in interactions were depicted as grey stick. H-bond was reported as yellow dotted lines.

Then, we analyzed the two binding locations with the intention of understanding the reasons that contributed to AM-001's selectivity between the two EPAC isoforms under study. The main difference deals with the residues of the loop between the CNBD b10 and REM b1. For EPAC1 and EPAC2, the loop had 19 and 18 residues, respectively. Even though the two isoforms' lengths were comparable, just two residues were shared. Additionally, the four residues facing EPAC1's binding pocket (RPPT) were considerably different from those in the same places

Application of molecular dynamics simulations

in EPAC2 (NQGN). Consequently, it might be suggested that the selectivity is caused by how this loop behaves in terms of conformation and interactions with AM-001. Furthermore, R847 and I825, involved in contacts with AM-001 are not the same in the EPAC2 isoform but they are replaced by N (956) and T (934) respectively, weakening the binding stability. Additionally, E(345) in EPAC2 corresponds to D239 in EPAC1. When combined, these results offer a potential justification for the affinity of AM-001 for EPAC1.

The nine AM-001 analogues (Table 7 and Figure 36) were then docked into this binding site to assess how well the site might fit the experimental data⁽¹⁸⁾.

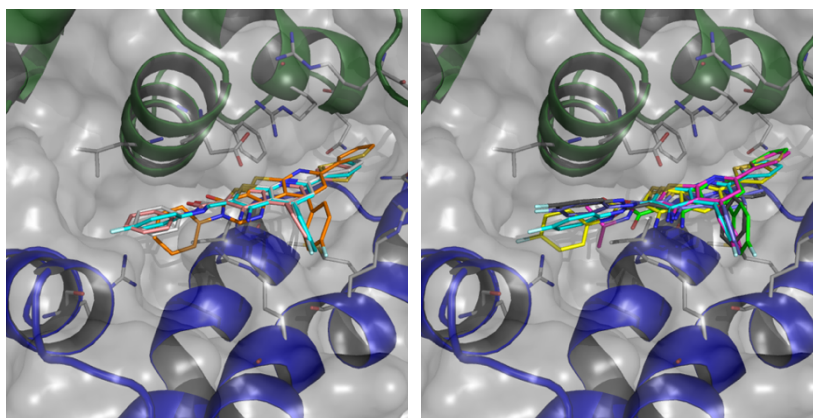


Figure 36. Plants proposed binding mode of AM-001 analogues. Left panel: AM-001 (cyan) AM-002 orange, AM-004 pink and AM-005 white; Right panel: AM-001 (cyan), AM-003 magenta, AM-006 light green, AM-007 grey, AM-008 violet and AM-009 yellow. AM-010 is not reported because of unrelated binding mode. Epac1 is reported as cartoon: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. Surface is also reported in grey. Residues involved in interactions were depicted as white stick

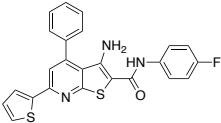
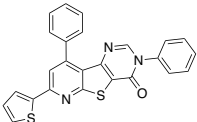
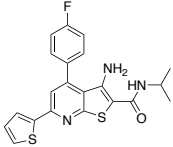
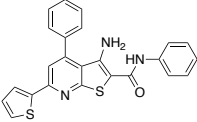
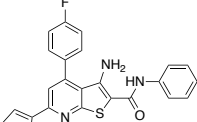
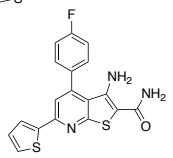
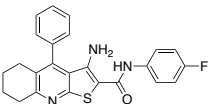
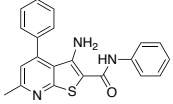
The compounds AM-004 and AM-005, which varied only in the fluorine atom location, shared the same interactions with AM-001 and had a very comparable binding mode. Although there were some obvious differences, the

Application of molecular dynamics simulations

binding mode for the compounds AM-003, AM-006, AM-007, AM-008, and AM-009 was similar to AM-001, we did not observe the contacts of the fluorinated phenyl ring (AM-003 and AM-006), the thiophene (AM-007 and AM-008), and the unsubstituted phenyl ring (AM-009) of AM-001. We found a binding mode for compound AM-002 that was comparable to AM-001, but we did not see any H-bonds, and the pyrimidone-linked phenyl ring was positioned farther away from the R801 position (Figure 34).

Moreover, we found a strong correlation between the experimental biological activity and the estimated binding energy for the docking poses^(36,64), with a significant correlation coefficient, R , of 0.74 and R_s of 0.83 p (2-tailed) 0.53%. (Table 8). The chosen binding mode essentially matches the SAR for analogs of AM-001. In conclusion, the chosen binding mode rationally fits the known structure-activity relationship for the AM-001 analogues and appears appropriate as a selective allosteric binding site for the inhibitor AM-001.

Application of molecular dynamics simulations

Compounds	Structures	BRET ratio variation ^a
AM-001		46.7 ± 0.8
AM-002		93.6 ± 0.9
AM-003		100.4 ± 2.1
AM-004		47.0 ± 0.7
AM-005		46.9 ± 1.3
AM-006		74 ± 2.8
AM-007		80.5 ± 3.9
AM-008		82.1 ± 5.5

Application of molecular dynamics simulations

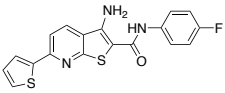
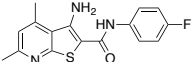
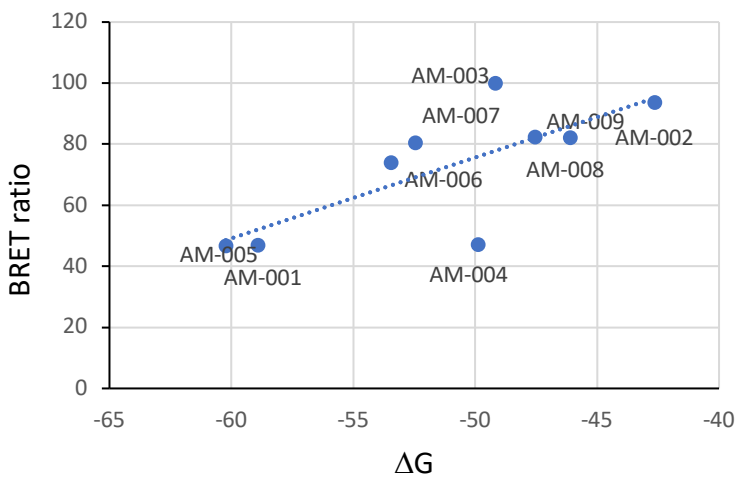
AM-009		82.4 ± 1.8
AM-010		101.3 ± 2.8

Table 7 Structures and Bret ratio variation of AM-001 analogues.^a
^a Data from reference 1



Application of molecular dynamics simulations

Compd	DG ^a	BRET ratio ^b
AM-001	-60.23	46.7
AM-002	-42.65	93.6
AM-003	-49.17	100
AM-004	-49.88	47
AM-005	-58.92	46.9
AM-006	-53.45	74
AM-007	-52.46	80.5
AM-008	-46.1	82.1
AM-009	-47.54	82.4

Table 8. Correlation between BRET-ratio values and calculated DG (Kcal/mol) of docking proposed binding poses for AM-001 analogues. ^aDG values were calculated by MMGB-SA module of Maestro suite [2, 3]; ^bdata are from reference 1. Compound AM-010 was omitted in the computation by its unrelated binding mode.

The Spearman coefficient R_s was 0.83, p (2-tailed) 0.53%;

The Pearson coefficient R was 0.74

At the end, we analysed the EPAC1/AM-001 complex trajectory in order to understand how the AM-001 binding impaired the activating transition of the enzyme. To underline the dominant mode of motion of the protein over the simulation time, PCA analysis was conducted^(65,66,67). Porcupine plots^(68,69), that display the direction and magnitude of the top two eigenvectors (Figure 37) for each of the backbone Ca atoms, were used to represent these protein motions (Figure 37 and 38). The most prominent observed motion was related to the loops at the CDC25-HD. Focusing on the helices that shaped the binding site we observed that $\alpha 7$ (794–805), $\alpha 8$ (821–830) and $\alpha 9$ (840–850) helices of the CDC25-HD moved toward the CNBD, as well as the CNBD $\alpha 2$ (230–242) moved toward the CDC25-HD. The observed movements involved the enzyme loop and helices that are crucial in the open -to-close conformational transition. As a result,

Application of molecular dynamics simulations

AM-001 binding to region 6 in the closed conformation of EPAC1 may improve the interaction between the CDC25-HD and CNBD domains, stabilizing an inactive-like conformation and avoiding the activation transition that would otherwise be triggered by cAMP binding to the CNBD domain.

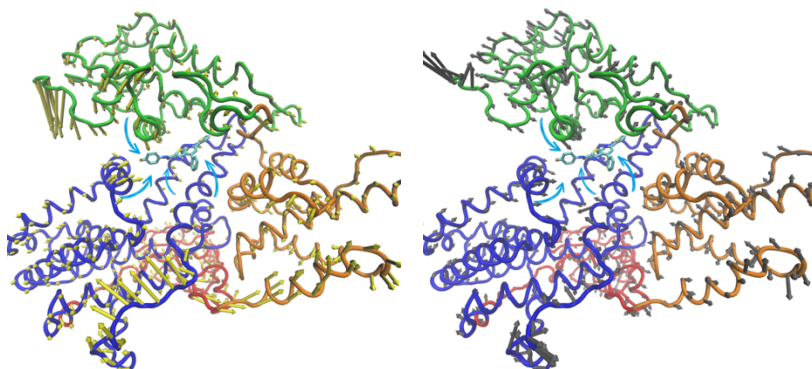


Figure 37 Porcupine plot of the top two eigenvectors. Right panel: eigenvector 1, left panel: eigenvector 2. Epac1 is reported as tube: CNBD and DEP green; REM orange; CDC25-HD blue and RA red. AM-001 is reported as cyan stick. The yellow and grey arrows attached to each α -carbon atom indicate the direction of the movement; the size of each arrow shows the magnitude of the corresponding movement.

Application of molecular dynamics simulations

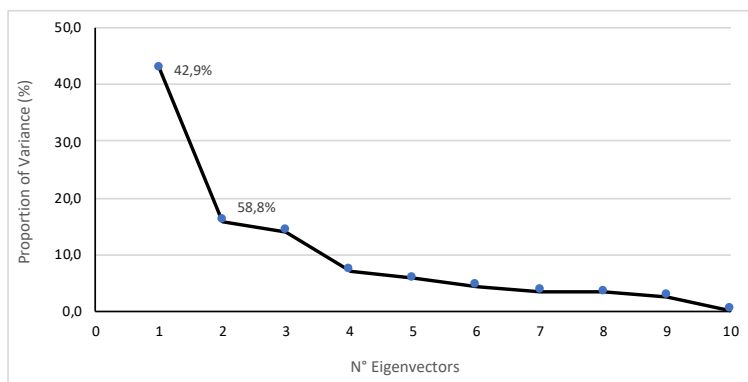


Figure 38. PCs cumulative contribution (%), Y-axis) of variance for the 10 eigenvectors (X-axis) calculated by PCA. The contribution of first and second eigenvectors is also reported to the corresponding point.

N° Eigenvectors	Proportion of Variance (%)	Eigenvalues
1	42.9	982.07
2	15.9	363.37
3	14.1	321.60
4	7.2	165.06
5	5.8	133.42
6	4.4	101.89
7	3.5	81.88
8	3.4	77.59
9	2.6	60.75
10	0.2	5.25

Table 9 Values of the ten eigenvectors, eigenvalues and relative proportion of variance

1.4) Conclusions

To identify a potential AM-001 EPAC1 allosteric inhibitor binding site, cosolvent molecular dynamics was employed. By analyzing cosolvent occupancy maps, we were able to determine the EPAC1 sites that would be best for partner binding. The docking and molecular dynamics of AM-001 evaluated each location. The pocket located at the interface between $\alpha 8$, $\alpha 9$ and $\alpha 10$ of CDC25-HD and $\alpha 2$ and $\alpha 3$ of CNBD appears to be suitable as a selective allosteric binding site for the AM-001. The available AM-001 analogues were docked to this site showing a good match with the already reported SAR⁽¹⁸⁾. The EPAC1/AM-001 trajectory was subjected to PCA analysis, which revealed that the AM-001 binding may have strengthened the contact between CDC25-HD of the catalytic region and CNBD of the regulatory region, stabilizing an inactive "like" conformation. Therefore, despite the binding of cAMP to EPAC1, AM-001 binding to the predicted binding pocket may force the highly dynamic protein in an inhibited conformation.

Bufano M, Laudette M, Blondeau JP, Lezoualc'h F, Nalli M, Silvestri R, Brancale A, Coluccia A. Modeling Epac1 interactions with the allosteric inhibitor AM-001 by co-solvent molecular dynamics. *J Comput Aided Mol Des.* 2020 Nov;34(11):1171-1179. doi: 10.1007/s10822-020-00332-y.

1.5) References

1. Poppinga WJ, Muñoz-Llancao P, González-Billault C, Schmidt M. A-kinase anchoring proteins: cAMP compartmentalization in neurodegenerative and obstructive pulmonary diseases. *Br J Pharmacol*. 2014 Dec;171(24):5603-23. doi: 10.1111/bph.12882. PMID: 25132049; PMCID: PMC4290705.
2. Schmidt M, Dekker FJ, Maarsingh H. Exchange protein directly activated by cAMP (epac): a multidomain cAMP mediator in the regulation of diverse biological functions. *Pharmacol Rev*. 2013 Feb 27;65(2):670-709. doi: 10.1124/pr.110.003707. Erratum in: *Pharmacol Rev*. 2013 Jul;65(3):1134. PMID: 23447132.
3. Brand T, Schindler R. New kids on the block: The Popeye domain containing (POPDC) protein family acting as a novel class of cAMP effector proteins in striated muscle. *Cell Signal*. 2017 Dec;40:156-165. doi: 10.1016/j.cellsig.2017.09.015. Epub 2017 Sep 20. PMID: 28939104; PMCID: PMC6562197.
4. Sartiani L, Mannaioni G, Masi A, Novella Romanelli M, Cerbai E. The Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels: from Biophysics to Pharmacology of a Unique Family of Ion Channels. *Pharmacol Rev*. 2017 Oct;69(4):354-395. doi: 10.1124/pr.117.014035. PMID: 28878030.
5. Robichaux WG 3rd, Cheng X. Intracellular cAMP Sensor EPAC: Physiology, Pathophysiology, and Therapeutics Development. *Physiol Rev*. 2018 Apr 1;98(2):919-1053. doi: 10.1152/physrev.00025.2017. PMID: 29537337; PMCID: PMC6050347.
6. Ercu M, Klussmann E. Roles of A-Kinase Anchoring Proteins and Phosphodiesterases in the Cardiovascular System. *J Cardiovasc Dev Dis*. 2018 Feb 20;5(1):14. doi: 10.3390/jcdd5010014. PMID: 29461511; PMCID: PMC5872362.

Application of molecular dynamics simulations

7. Lezoualc'h F, Fazal L, Laudette M, Conte C. Cyclic AMP Sensor EPAC Proteins and Their Role in Cardiovascular Function and Disease. *Circ Res.* 2016 Mar 4;118(5):881-97. doi: 10.1161/CIRCRESAHA.115.306529. PMID: 26941424.
8. de Rooij J, Zwartkruis FJ, Verheijen MH, Cool RH, Nijman SM, Wittinghofer A, Bos JL. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature.* 1998 Dec 3;396(6710):474-7. doi: 10.1038/24884. PMID: 9853756.
9. Kawasaki H, Springett GM, Mochizuki N, Toki S, Nakaya M, Matsuda M, Housman DE, Graybiel AM. A family of cAMP-binding proteins that directly activate Rap1. *Science.* 1998 Dec 18;282(5397):2275-9. doi: 10.1126/science.282.5397.2275. PMID: 9856955.
10. Niimura M, Miki T, Shibasaki T, Fujimoto W, Iwanaga T, Seino S. Critical role of the N-terminal cyclic AMP-binding domain of Epac2 in its subcellular localization and function. *J Cell Physiol.* 2009 Jun;219(3):652-8. doi: 10.1002/jcp.21709. PMID: 19170062.
11. Lezoualc'h F, Fazal L, Laudette M, Conte C. Cyclic AMP Sensor EPAC Proteins and Their Role in Cardiovascular Function and Disease. *Circ Res.* 2016 Mar 4;118(5):881-97. doi: 10.1161/CIRCRESAHA.115.306529. PMID: 26941424.
12. Rehmann H, Arias-Palomo E, Hadders MA, Schwede F, Llorca O, Bos JL. Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature.* 2008 Sep 4;455(7209):124-7. doi: 10.1038/nature07187. Epub 2008 Jul 27. PMID: 18660803.
13. Selvaratnam R, Mazhab-Jafari MT, Das R, Melacini G. The auto-inhibitory role of the EPAC hinge helix as mapped by NMR. *PLoS One.* 2012;7(11):e48707. doi: 10.1371/journal.pone.0048707. Epub 2012 Nov 21. PMID: 23185272; PMCID: PMC3504058.

Application of molecular dynamics simulations

14. VanSchouwen B, Selvaratnam R, Fogolari F, Melacini G. Role of dynamics in the autoinhibition and activation of the exchange protein directly activated by cyclic AMP (EPAC). *J Biol Chem*. 2011 Dec 9;286(49):42655-42669. doi: 10.1074/jbc.M111.277723. Epub 2011 Aug 26. PMID: 21873431; PMCID: PMC3234915.
15. Mazhab-Jafari MT, Das R, Fotheringham SA, SilDas S, Chowdhury S, Melacini G. Understanding cAMP-dependent allostery by NMR spectroscopy: comparative analysis of the EPAC1 cAMP-binding domain in its apo and cAMP-bound states. *J Am Chem Soc*. 2007 Nov 21;129(46):14482-92. doi: 10.1021/ja0753703. Epub 2007 Oct 31. PMID: 17973384.
16. Rehmann H, Prakash B, Wolf E, Rueppel A, de Rooij J, Bos JL, Wittinghofer A. Structure and regulation of the cAMP-binding domains of Epac2. *Nat Struct Biol*. 2003 Jan;10(1):26-32. doi: 10.1038/nsb878. PMID: 12469113.
17. Bers DM. Calcium cycling and signaling in cardiac myocytes. *Annu Rev Physiol*. 2008;70:23-49. doi: 10.1146/annurev.physiol.70.113006.100455. PMID: 17988210.
18. El-Armouche A, Eschenhagen T. Beta-adrenergic stimulation and myocardial function in the failing heart. *Heart Fail Rev*. 2009 Dec;14(4):225-41. doi: 10.1007/s10741-008-9132-8. PMID: 19110970.
19. Laudette M, Coluccia A, Sainte-Marie Y, Solari A, Fazal L, Sicard P, Silvestri R, Mialet-Perez J, Pons S, Ghaleh B, Blondeau JP, Lezoualc'h F. Identification of a pharmacological inhibitor of Epac1 that protects the heart against acute and chronic models of cardiac stress. *Cardiovasc Res*. 2019 Oct 1;115(12):1766-1777. doi: 10.1093/cvr/cvz076. PMID: 30873562.
20. Bouvet M, Blondeau JP, Lezoualc'h F. The Epac1 Protein: Pharmacological Modulators, Cardiac Signalosome and Pathophysiology. *Cells*. 2019 Nov 29;8(12):1543. doi:

Application of molecular dynamics simulations

- 10.3390/cells8121543. PMID: 31795450; PMCID: PMC6953115.
21. Berman HM, Ten Eyck LF, Goodsell DS, Haste NM, Kornev A, Taylor SS. The cAMP binding domain: an ancient signaling module. *Proc Natl Acad Sci U S A*. 2005 Jan 4;102(1):45-50. doi: 10.1073/pnas.0408579102. Epub 2004 Dec 23. PMID: 15618393; PMCID: PMC544069.
 22. Westley A. M., and Westley J. (1996) Enzyme inhibition in open systems. Superiority of uncompetitive agents. *J. Biol. Chem.* 271, 5347–5352 10.1074/jbc.271.10.5347
 23. Courilleau D, Bissierier M, Jullian JC, Lucas A, Bouyssou P, Fischmeister R, Blondeau JP, Lezoualc'h F. Identification of a tetrahydroquinoline analog as a pharmacological inhibitor of the cAMP-binding protein Epac. *J Biol Chem*. 2012 Dec 28;287(53):44192-202. doi: 10.1074/jbc.M112.422956. Epub 2012 Nov 8. PMID: 23139415; PMCID: PMC3531735.
 24. Courilleau D, Bouyssou P, Fischmeister R, Lezoualc'h F, Blondeau JP. The (R)-enantiomer of CE3F4 is a preferential inhibitor of human exchange protein directly activated by cyclic AMP isoform 1 (Epac1). *Biochem Biophys Res Commun*. 2013 Oct 25;440(3):443-8. doi: 10.1016/j.bbrc.2013.09.107. Epub 2013 Oct 4. PMID: 24099776.
 25. Tsalkova T, Blumenthal DK, Mei FC, White MA, Cheng X. Mechanism of Epac activation: structural and functional analyses of Epac2 hinge mutants with constitutive and reduced activities. *J Biol Chem*. 2009 Aug 28;284(35):23644-51. doi: 10.1074/jbc.M109.024950. Epub 2009 Jun 24. PMID: 19553663; PMCID: PMC2749139.
 26. Brown LM, Rogers KE, Aroonsakool N, McCammon JA, Insel PA. Allosteric inhibition of Epac: computational modeling and experimental validation to identify allosteric sites and inhibitors. *J Biol Chem*. 2014 Oct 17;289(42):29148-57. doi: 10.1074/jbc.M114.569319.

Application of molecular dynamics simulations

- Epub 2014 Sep 2. PMID: 25183009; PMCID: PMC4200268.
27. Ghanakota P, Carlson HA. Driving Structure-Based Drug Discovery through Cosolvent Molecular Dynamics. *J Med Chem*. 2016 Dec 8;59(23):10383-10399. doi: 10.1021/acs.jmedchem.6b00399. Epub 2016 Aug 17. PMID: 27486927; PMCID: PMC5217181.
 28. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, Ringe D. An experimental approach to mapping the binding surfaces of crystalline proteins. *J Phys Chem*. 1996;100:2605–2611.
 29. Uehara S, Tanaka S. Cosolvent-Based Molecular Dynamics for Ensemble Docking: Practical Method for Generating Druggable Protein Conformations. *J Chem Inf Model*. 2017 Apr 24;57(4):742-756. doi: 10.1021/acs.jcim.6b00791. Epub 2017 Apr 7. PMID: 28388074.
 30. Yu W, Lakkaraju SK, Raman EP, Fang L, MacKerell AD Jr. Pharmacophore modeling using site-identification by ligand competitive saturation (SILCS) with multiple probe molecules. *J Chem Inf Model*. 2015 Feb 23;55(2):407-20. doi: 10.1021/ci500691p. Epub 2015 Feb 6. PMID: 25622696; PMCID: PMC4339487.
 31. Lexa KW, Goh GB, Carlson HA. Parameter choice matters: validating probe parameters for use in mixed-solvent simulations. *J Chem Inf Model*. 2014 Aug 25;54(8):2190-9. doi: 10.1021/ci400741u. Epub 2014 Aug 1. PMID: 25058662; PMCID: PMC4144759.
 32. Lakkaraju SK, Mbatia H, Hanscom M, Zhao Z, Wu J, Stoica B, MacKerell AD Jr, Faden AI, Xue F. Cyclopropyl-containing positive allosteric modulators of metabotropic glutamate receptor subtype 5. *Bioorg Med Chem Lett*. 2015 Jun 1;25(11):2275-9. doi: 10.1016/j.bmcl.2015.04.042. Epub 2015 Apr 20. PMID: 25937015; PMCID: PMC4690453.

Application of molecular dynamics simulations

33. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J Chem Theory Comput.* 2012 Jul 10;8(7):2435-2447. doi: 10.1021/ct300117j. Epub 2012 Jun 5. PMID: 22798729; PMCID: PMC3392909.
34. Rehmann H, Das J, Knipscheer P, Wittinghofer A, Bos JL. Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature.* 2006 Feb 2;439(7076):625-8. doi: 10.1038/nature04468. PMID: 16452984.
35. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins.* 2004 May 1;55(2):351-67. doi: 10.1002/prot.10613. PMID: 15048827.
36. Schrödinger Release 2018-1: Prime, Schrödinger, LLC, New York, NY, 2018
37. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des.* 2013 Mar;27(3):221-34. doi: 10.1007/s10822-013-9644-8. Epub 2013 Apr 12. PMID: 23579614.
38. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.* 2012 Jul 19;7(8):1511-22. doi: 10.1038/nprot.2012.085. PMID: 22814390; PMCID: PMC4730388.
39. Ma J, Wang S, Zhao F, Xu J. Protein threading using context-specific alignment potential. *Bioinformatics.* 2013 Jul 1;29(13):i257-65. doi: 10.1093/bioinformatics/btt210. PMID: 23812991; PMCID: PMC3694651.
40. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods

Application of molecular dynamics simulations

- RJ. The Amber biomolecular simulation programs. *J Comput Chem.* 2005 Dec;26(16):1668-88. doi: 10.1002/jcc.20290. PMID: 16200636; PMCID: PMC1989667.
41. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem.* 2009 Oct;30(13):2157-64. doi: 10.1002/jcc.21224. PMID: 19229944.
 42. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004 Oct;25(13):1605-12. doi: 10.1002/jcc.20084. PMID: 15264254.
 43. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J Struct Biol.* 2010 Jun;170(3):427-38. doi: 10.1016/j.jsb.2010.03.007. Epub 2010 Mar 23. PMID: 20338243; PMCID: PMC2874196.
 44. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 2006 Oct;25(2):247-60. doi: 10.1016/j.jm gm.2005.12.005. Epub 2006 Feb 3. PMID: 16458552.
 45. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem.* 2004 Jul 15;25(9):1157-74. doi: 10.1002/jcc.20035. Erratum in: *J Comput Chem.* 2005 Jan 15;26(1):114. PMID: 15116359.
 46. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput.* 2013 Jul 9;9(7):3084-95. doi: 10.1021/ct400341p. Epub 2013 Jun 25. PMID: 26583988.

Application of molecular dynamics simulations

47. Bakan A, Meireles LM, Bahar I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*. 2011 Jun 1;27(11):1575-7. doi: 10.1093/bioinformatics/btr168. Epub 2011 Apr 5. PMID: 21471012; PMCID: PMC3102222.
48. Bakan A, Dutta A, Mao W, Liu Y, Chennubhotla C, Lezon TR, Bahar I. Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics*. 2014 Sep 15;30(18):2681-3. doi: 10.1093/bioinformatics/btu336. Epub 2014 May 21. PMID: 24849577; PMCID: PMC4155247.
49. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996 Feb;14(1):33-8, 27-8. doi: 10.1016/0263-7855(96)00018-5. PMID: 8744570.
50. Rehmann H, Selective activation of Epac1 and Epac2. To be published. (PDB code: 4MGI)
51. Peng W, Xu J, Guan X, Sun Y, Zhang XC, Li X, Rao Z. Structural study of the Cdc25 domain from Ral-specific guanine-nucleotide exchange factor RalGPS1a. *Protein Cell*. 2011 Apr;2(4):308-19. doi: 10.1007/s13238-011-1036-z. Epub 2011 Apr 14. PMID: 21494904; PMCID: PMC4875206.
52. Baameur F, Singhmar P, Zhou Y, Hancock JF, Cheng X, Heijnen CJ, Kavelaars A. Epac1 interacts with importin β 1 and controls neurite outgrowth independently of cAMP and Rap1. *Sci Rep*. 2016 Nov 3;6:36370. doi: 10.1038/srep36370. PMID: 27808165; PMCID: PMC5093460.
53. Parnell E, Smith BO, Yarwood SJ. The cAMP sensors, EPAC1 and EPAC2, display distinct subcellular distributions despite sharing a common nuclear pore localisation signal. *Cell Signal*. 2015 May;27(5):989-96. doi: 10.1016/j.cellsig.2015.02.009. Epub 2015 Feb 12. PMID: 25683912; PMCID: PMC4372255.

Application of molecular dynamics simulations

54. Mei FC, Cheng X. Interplay between exchange protein directly activated by cAMP (Epac) and microtubule cytoskeleton. *Mol Biosyst.* 2005 Oct;1(4):325-31. doi: 10.1039/b511267b. Epub 2005 Sep 26. PMID: 16880999.
55. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009 Dec;30(16):2785-91. doi: 10.1002/jcc.21256. PMID: 19399780; PMCID: PMC2760638.
56. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol.* 1997 Apr 4;267(3):727-48. doi: 10.1006/jmbi.1996.0897. PMID: 9126849.
57. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model.* 2009 Jan;49(1):84-96. doi: 10.1021/ci800298z. PMID: 19125657.
58. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 2007 Dec 13;450(7172):1001-9. doi: 10.1038/nature06526. PMID: 18075579.
59. Campbell SJ, Gold ND, Jackson RM, Westhead DR. Ligand binding: functional site location, similarity and docking. *Curr Opin Struct Biol.* 2003 Jun;13(3):389-95. doi: 10.1016/s0959-440x(03)00075-7. PMID: 12831892.
60. Cui Q, Sulea T, Schrag JD, Munger C, Hung MN, Näim M, Cygler M, Purisima EO. Molecular dynamics-solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex. *J Mol Biol.* 2008 Jun 13;379(4):787-802. doi: 10.1016/j.jmb.2008.04.035. Epub 2008 Apr 20. PMID: 18479705.
61. Näim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO. Solvated interaction energy (SIE) for scoring protein-ligand binding affinities. 1. Exploring the

Application of molecular dynamics simulations

- parameter space. *J Chem Inf Model.* 2007 Jan-Feb;47(1):122-33. doi: 10.1021/ci600406v. PMID: 17238257.
62. Kumar S, Das A. Effect of acceptor heteroatoms on π -hydrogen bonding interactions: a study of indole \cdots thiophene heterodimer in a supersonic jet. *J Chem Phys.* 2012 Sep 7;137(9):094309. doi: 10.1063/1.4748818. PMID: 22957571.
 63. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC
 64. Jacobson MP, Friesner RA, Xiang Z, Honig B. On the role of the crystal environment in determining protein side-chain conformations. *J Mol Biol.* 2002 Jul 12;320(3):597-608. doi: 10.1016/s0022-2836(02)00470-9. PMID: 12096912.
 65. Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins.* 1993 Dec;17(4):412-25. doi: 10.1002/prot.340170408. PMID: 8108382.
 66. Amadei A, Linssen AB, de Groot BL, van Aalten DM, Berendsen HJ. An efficient method for sampling the essential subspace of proteins. *J Biomol Struct Dyn.* 1996 Feb;13(4):615-25. doi: 10.1080/07391102.1996.10508874. PMID: 8906882.
 67. Yamaguchi H, van Aalten DM, Pinak M, Furukawa A, Osman R. Essential dynamics of DNA containing a cis.syn cyclobutane thymine dimer lesion. *Nucleic Acids Res.* 1998 Apr 15;26(8):1939-46. doi: 10.1093/nar/26.8.1939. PMID: 9518486; PMCID: PMC147494.
 68. Tai K, Shen T, Börjesson U, Philippopoulos M, McCammon JA. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys J.* 2001 Aug;81(2):715-24. doi: 10.1016/S0006-3495(01)75736-0. PMID: 11463620; PMCID: PMC1301548.

Application of molecular dynamics simulations

69. Tai K, Shen T, Henchman RH, Bourne Y, Marchot P, McCammon JA. Mechanism of acetylcholinesterase inhibition by fasciculin: a 5-ns molecular dynamics simulation. *J Am Chem Soc.* 2002 May 29;124(21):6153-61. doi: 10.1021/ja017310h. PMID: 12022850.

2) Exploring CCRL2-Chemerin binding

Chemokines are soluble mediators that control immunological development and function, primarily via enhancing leukocyte trafficking. Chemokines interact with seven-transmembrane domain (7-TMD) receptors. The two main functional categories of chemokine receptors are the "classical" G protein-coupled signaling receptors (GPCR) and the atypical chemokine receptors (ACKRs)⁽¹⁾. ACKRs did not behave like GPCR for signal transduction, but rather they induce ligand internalization and degradation. In various circumstances ACKRs have the ability to move the ligand throughout the cell⁽²⁾. Similar to ACKRs, CCRL2 has a changed amino acidic sequence (QGYRVFS) in the cytoplasmic extension of the third transmembrane helix in place of the conserved motif DRYLAIV, which is necessary for the induction of G protein-mediated responses⁽¹⁾. Therefore, CCRL2 does not promote cell motility or activate traditional G protein-mediated signaling. CCRL2 has been suggested to function as a receptor for chemokines including CCL2, CCL5, CCL7, CCL8, CCL19, and CCL21 in the past^(3,4,5), although other research teams have doubted these findings. Instead, it is commonly accepted that CCRL2 bind the non-chemokine chemotactic factor chemerin. According to the widely accepted theory, CCRL2 binds chemerin at its N-terminus while leaving the C-terminus free for interaction with cells expressing the functional chemerin receptor, chemokine-like receptor 1 (CMKLR1). So, it is suggested that CCRL2 function as a chemerin-presenting molecule at the surface of barrier cells^(6,7). The recruitment of CMKLR1-expressing cells, such as monocytes/macrophages, dendritic cells, plasmacytoid dendritic cells, and NK cells, may therefore be aided by CCRL2^(8,9,10). It has been demonstrated that this

Application of molecular dynamics simulations

CCRL2/CMKLR1 axis controls the trafficking of dendritic cells, mast cells, and NK cells *in vivo*^(6,11,12,13,14).

This concept, which assumes that CCRL2 differs from all other ACKRs in not internalizing, recycling, and scavenging the ligand, is supported by *in vitro* and *in vivo* studies^(13,14). Therefore, the biological role of CCRL2 as well as its scavenging and recycling properties are still up for debate.

As explain previously, the only recognized CCRL2 ligand up to this point is chemerin, which is encoded by the RARRES2 gene. Mammalian cells synthesize chemerin, a 163 amino acid (aa) pro-precursor. A precursor form of chemerin is secreted as a result of the N-terminal processing of the 20 aa⁽⁷⁾. Depending on the extent of processing, the additional C-terminal cleavage produces both active and deactivated chemerin forms. For instance, chemerin is activated by proteases such as plasmin, elastase, and cathepsin G, resulting in multiple chemerin isoforms with varying affinities for CMKLR1, the active chemerin receptor. Chemerin becomes inactive when chymase further cleaves the bioactive form⁽¹⁵⁾. Therefore, the C-terminal proteolytic processing functions as a regulatory mechanism to manage the concentration of active chemerin.

To date, two more G protein-coupled receptors, CMKLR1 and GPR1, that show high affinity for chemerin have been identified. Most innate immune cells, including dendritic cells, macrophages, and Natural Killer (NK) cells, express CMKLR1.

Chemerin triggers a poor Ca²⁺ mobilization and ERK1/2 activation in GPR1-expressing cell lines, but the agonist-induced internalization of the receptor is effective. Given that no GPR1-mediated activity has been identified on primary cells or in living organisms, it is possible that GPR1 functions as a chemerin decoy receptor. GPR1 is found to be expressed in the central nervous system,

Application of molecular dynamics simulations

skeletal muscle, skin, and adipose tissue, where it may control the action of chemerin, rather than by leukocyte populations^(7,16).

Only CCRL2, between these three receptors, lacks the capacity to initiate an intracellular signaling cascade.

It has been demonstrated that CCRL2 controls the recruitment of NK cells in pathological situations^(11,17) as well as inflammation-related disorders as experimental autoimmune encephalitis, hypersensitivity, and inflammatory arthritis^(16,18,19,20).

Given the numerous physiological roles of the ccl2-chemerin complex we choose to run protein-protein docking followed by accelerated molecular dynamic⁽²¹⁾ (aMD) simulation of the potential binding conformations in order to further define the CCRL2–chemerin interaction.

It was setting a protocol of aMD as opposed to a standard molecular dynamic approach, which is forced by kinetic trapping effects and a restricted sampling of the conformational space. By applying a potential energy boost, this technique lowers the energy barrier between two different low-energy states, raising the transition probabilities between two distinct conformations⁽²²⁾. Principal component analysis (PCA) were used to analyze the 5.5 μM second trajectories that we collected and identify the variables with the most variance^(23,24). Then, the hot-spot residues of the CCRL2 chemerin complex were highlighted. These were the residues that were more frequently implicated in binding interactions.

2.1) Materials and methods

2.1.1) Structural comparison of modeled proteins

The optimized CCRL2 and chemerin models were compared to their respective AlphaFold

Application of molecular dynamics simulations

(<https://alphafold.ebi.ac.uk/>, entry: O00421 CCRL2_HUMAN) conformations using Matchmaker function of UCSF Chimera⁽²⁵⁾ and RMSD values were obtained.

2.1.2) Protein–protein docking

Protein–protein docking was performed on HADDOCK server⁽²⁶⁾ using the 3D models of CCRL2 and chemerin proteins. We set up as “active” residues (residues expected to be involved in the interaction between the two molecules) the N-terminal residues 1–100 of chemerin and the extracellular loops of the CCRL2 receptor (from UniProtKB, 25–67; 120–128; 190–122; 284–310) following literature data⁽²⁶⁾; the other residues were defined as “passive” (residues accessible to the solvent closed of the active residues). This docking protocol consisted of three stages: rigid body (it0), semi-flexible refinement (it1), and explicit solvent refinement (water). The docking experiments were carried out by using default parameters. Just MD steps in the TAD and cooling stage were increased to 2000 for it0⁽²⁶⁾. HADDOCK produced 1000 models in the first step, then refined to 200 best model in the following steps. The final models were automatically clustered based on the fraction of common contacts that measures the similarity of the intermolecular contacts. At the end, we obtained 12 clusters, and for each of them, it was selected as representative structure the conformation with the best score value.

2.1.3) Accelerated molecular dynamics

The membrane embedded complexes structures⁽²⁷⁾ and the Amber parameters⁽²⁸⁾ were obtained by CHARMM-GUI server through Membrane builder module⁽²⁹⁾.

The structural information to have a reliable placement of the bilayer membrane at CCRL2 were obtained by

Application of molecular dynamics simulations

Uniprot (UniProtKB: O00421) at the subcellular localization section. Indeed, in this section were defined both the nonmembrane region (topological domain), and the extent of the membrane-spanning regions (transmembrane) of CCRL2.

In the first step of the CHARMM membrane builder we upload the pdb files of the representative structures from HADDOCK cluster. In the second step the protein was aligned in order to orient it with respect to the membrane normal (the Z-axis). In the third step the system size was determined calculating the cross-sectional area along the Z-axis and the cross-sectional areas of lipid molecules. So the systems were inserted in a membrane of 20% cholesterol and 80% POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine); the system was solvated with TIP3P water model and ionized up to a concentration of 0.15 M NaCl still using the CHARMM-GUI, additional Cl⁻ ions were added to neutralize the systems⁽²⁷⁾. All the system (proteins + membrane + solvent) consists of 67 447 atoms. Each system was then submitted to aMD, carried out on Cineca supercomputer using Amber20. The whole system was minimized (5000 cycle) using restraints for CCRL2 and membrane (10 and 2.5, respectively); then, the CHARMM-steps equilibration protocol with progressive removal of position restraints was applied to the membrane and protein atoms (http://www.charmm-gui.org/demo/amber_ff/2).

Several restraints are set on the protein, water, ion, and lipid molecules throughout the equilibration process to ensure gradual equilibration of the initially constructed system.⁽⁴⁴⁾ Harmonic restraints are applied to ions and heavy atoms of the protein, repulsive planar restraints to prevent water from entering into the membrane hydrophobic region, and planar restraints to hold the position of head groups of membranes along the Z-axis. As the equilibration process continues, these restraining

Application of molecular dynamics simulations

forces gradually decrease. To permit a successful equilibration, i.e., to avoid instability of dynamics integrations during equilibration, the NVT dynamics (constant volume and temperature) is used for the first and second steps, and the NPAT (constant pressure, area, and temperature) dynamics for the rest at 303.15 K (DMPC and POPC) and 323.15 K (DPPC) (Table 10).

Step	Ensemble ¹	Time steps	Equilibration time	Force constant for harmonic restraint ²				
				Protein backbone ³	Protein sidechain ³	Water ⁴	Lipid ⁵	Ion ³
1	NVT	1 ps	125 ps	10.0	5.0	2.5	2.5	10.0
2	NVT	1 ps	125 ps	5.0	2.5	2.5	2.5	0.0
3	NPAT	2 ps	125 ps	2.5	1.0	1.0	1.0	0.0
4	NPAT	2 ps	250 ps	1.0	0.5	0.5	0.5	0.0
5	NPAT	2 ps	250 ps	0.5	0.1	0.1	0.1	0.0
6	NPAT	2 ps	250 ps	0.1	0.0	0.0	0.0	0.0

¹NVT stands for constant volume and temperature, NPAT stands for constant pressure, area, and temperature.

²Force constants are in kcal/(mol Å)

³positional armonic restraint

⁴Harmonic restraint to keep water molecule away from the membrane hydrophobic region.

⁵Harmonic restraint to keep the lipid tail in $-5 \text{ \AA} < Z < 5 \text{ \AA}$ and lipid head groups close to the membrane surface ($Z \pm 17 \text{ \AA}$ for DMPC and $Z \pm 19 \text{ \AA}$ for DPPC and POPC)

Table 10. Information on each equilibration steps

This equilibration protocol was carried out by Amber and consists of two NVT (constant number of particles (N), volume (V), and temperature (T)) steps to heat the system to 303.15 K employing as thermostat Langevin dynamics (collision frequency 1 ps) and four NPAT (constant number of particles (N), pressure (P), area (A) and temperature (T)) steps (Table 10) with SHAKE algorithm and the particle mesh Ewald (PME)⁽³⁰⁾ (with a cutoff of 9 Å). The required average dihedral energy and average total potential energy were computed during 5 ns classical molecular dynamics for each studied complex⁽³¹⁾. The aMD production (500 ns) was conducted at 315 K with constant pressure (1 bar) and periodic boundary

Application of molecular dynamics simulations

condition, Shake ($ntc = 2$) and PME with cut of 10 Å were set, each simulation was repeated three times.

Total Accessible Surface Area and Buried Surface Area (BSA) were computed by Pisa server (<http://pdbe.org/pisa/>). The property maps were calculated by Coco server⁽³²⁾.

2.1.4) Trajectories analyses

Trajectories analyses were carried out by mdtraj⁽³³⁾. The PCA analyses was carried out with scikit-learn using the decomposition module⁽³⁴⁾. Scipy library⁽³⁵⁾ was used to calculate Gaussian Kernel density estimation (KDE). Graphics were done with Matplotlib⁽³⁶⁾.

2.2) Modeling of CCRL2 and chemerin

A methodology based on protein-protein docking and aMDs was used to determine the potential residues implicated in the CCRL2 chemerin binding.

Since the crystal structure of CCRL2 was not available, it was decided to derive CCRL2 using a homology modeling approach and chemerin using ab-initio computations.

CCRL2 has two conformational states, active and inactive, like other chemokine receptors. It was decided to model exclusively the active state of the receptor because interaction with the ligands puts GPCRs in the active state. The CCRL2 model was specifically based on the chemokine receptor that shared the most similarities with it (PDB: 5WB1⁽³⁷⁾) and was lacking an N-terminal tail. The absence of highly conserved homologous proteins led to the ab-initio modeling of chemerin.

Application of molecular dynamics simulations

Notably, both the CCRL2 and chemerin structures became accessible at the AlphaFold database in the meantime, as the designed computations were completed (alphafold.ebi.ac.uk). By calculating the RMSD, the AlphaFold and our models were compared. For CCRL2, it was calculated a C α RMSD of 1.02 Å and the great amount of this distance was related with the extracellular loop 2 (ECL2, residues 169–192) and the TM6 helix (Figure 39).

TM6 was embedded in the membrane, far from the chemerin binding site. As a result, we predicted that it would only slightly affect the ligand binding. A long loop (23 residues) is difficult to predict with accuracy, indeed also AlphaFold listed this the ECL loop as having poor confidence (per-residue confidence scores between 70 and 50).

Furthermore, the bias associated with the various loop conformations was decreased by using aMD in place of classical MD. Indeed, aMD provided a significant benefit in simulating rare events necessary for protein conformational change and modeling conformational change without prior knowledge of conformational states⁽³⁸⁾. The superimposition of our model and the one proposed by AlphaFold resulted in a C α RMSD of 1.12 Å for chemerin. The C terminal helix 2 was the least well-fitting domain (Figure 40) but It was claimed that chemerin binding to the CCRL2 was not affected by this area. From this information we could tell that a good superimposition between the AlphaFold and our models was appreciated.

Application of molecular dynamics simulations

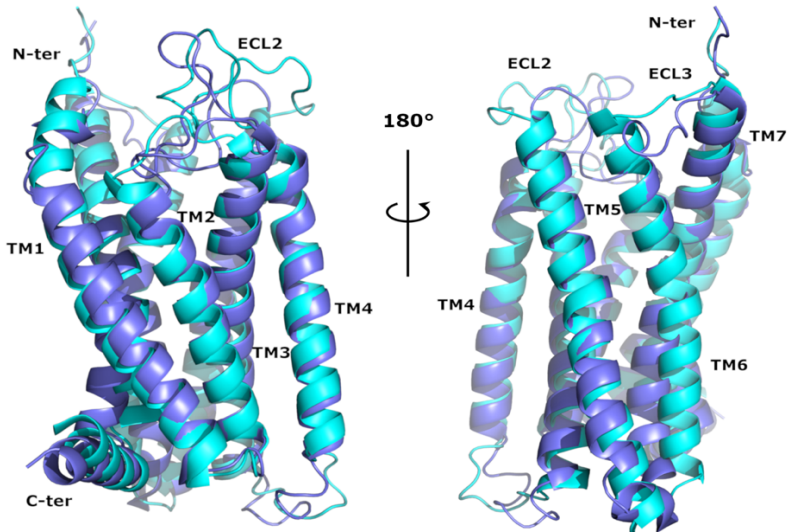


Figure 39. Superimposition of CCRL2 alphafold (cyan) and in-house homology model (blue). It is reported a frontal (left) and a 180 degrees rotated view (right) of the receptor.

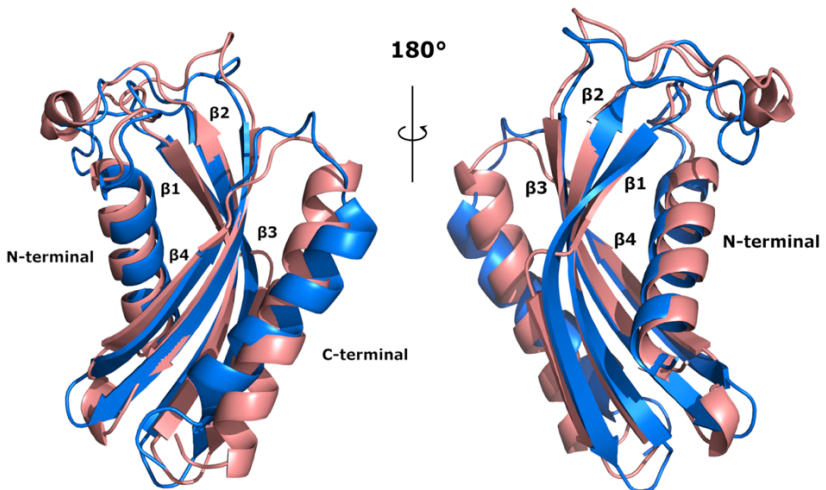


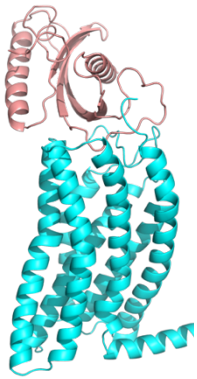
Figure 40. Superimposition of chemerin alphafold (blue) and in-house model (red). It is reported a frontal (left) and a 180 degrees rotated view (right).

2.3) CCRL2-chemerin protein–protein docking

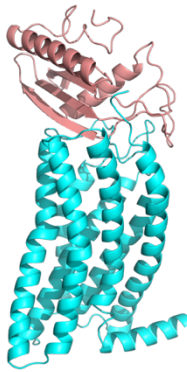
The CCRL2-chemerin complex was docked using HADDOCK server⁽²⁶⁾ and the results gave us 12 different clusters. Haddock post docking quality evaluation algorithms determined a representative binding conformation for each cluster, designated as complexes 1–12 (Figure 41). To the chosen models weren't given any more refinement. Similarly, in order to be as objective as possible, we didn't consider the docking score and the energy of the complexes. We generated a membrane belyer for each of the representative complexes and then we submitted them to aMD; each MD had a length of 500 ns. In order to assess the system stability, the resulting trajectories were first subjected to RMSD analysis.

Among the simulations we observed some complexes with peculiar trends. The C-terminal helix, for complex 9, moved up to the binding site during the simulation time and this behavior led us to don't consider this complex. Indeed, as told in the introduction, a model in which Chemerin C-terminal binds CMKLR is commonly accepted⁽³⁹⁾. Another complex that we don't investigate was the complex 10 in which Chemerin assumes a distorted conformation with the C-terminal moving far away from the N-terminal.

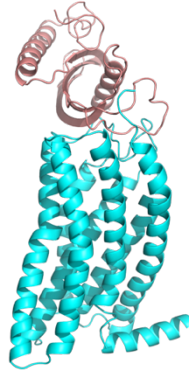
Application of molecular dynamics simulations



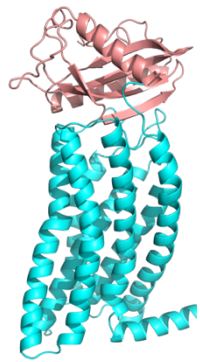
Complex 1



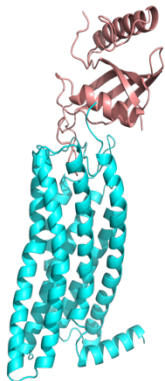
Complex 2



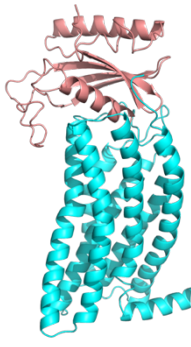
Complex 3



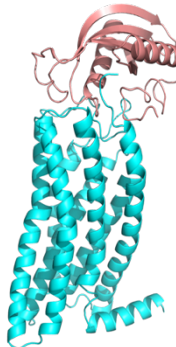
Complex 4



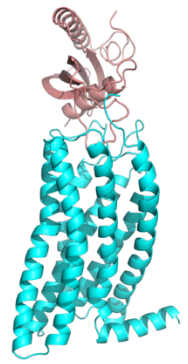
Complex 5



Complex 6

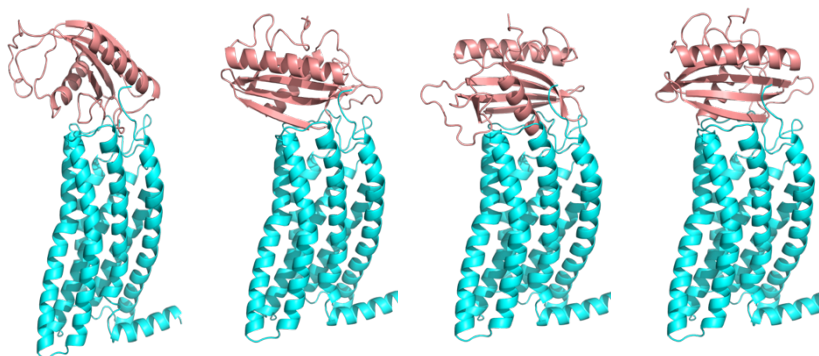


Complex 7



Complex 8

Application of molecular dynamics simulations



Complex 9 Complex 10 Complex 11 Complex 12
Figure 41. Docking proposed binding conformation for Chemerin (pink) and CCRL2 (cyan) aligned on CCRL2

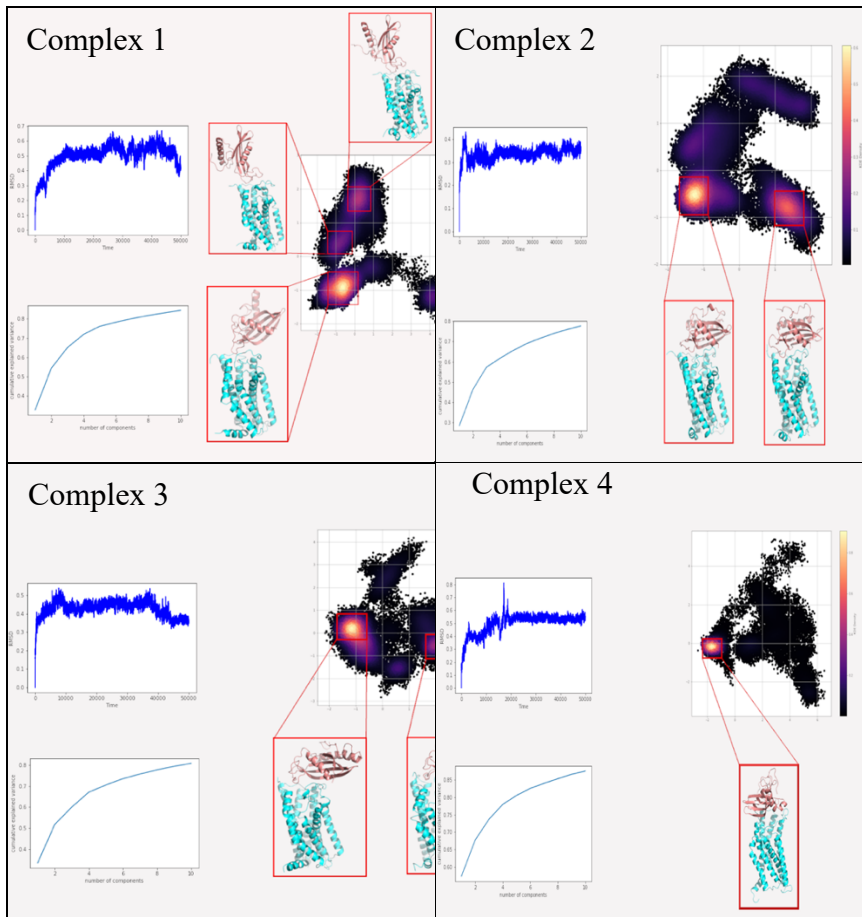
2.4) Selection of CCRL2-chemerin binding models

The analyses of the trajectories started from a PCA analyses. The obtained matrixes were investigated by Gaussian kernel (KDE) to produce probability distribution functions in the subspaces covered by principal components 1 and 2 (PC1 vs. PC2). It was demonstrated that the method for analyzing MD trajectories data that coupled a dimension reduction step (PCA) with a subsequent clustering step (KDE) was able to reduce noise and producing more compact and well-separated clusters of conformations⁽⁴⁰⁾. Furthermore, for each trajectory, the KDE plots facilitated the identification of conformational basins that presented the population with high density; from each basin we extracted the representative conformations. This method offered the more frequently present conformations believed to be most significant (Table 10).

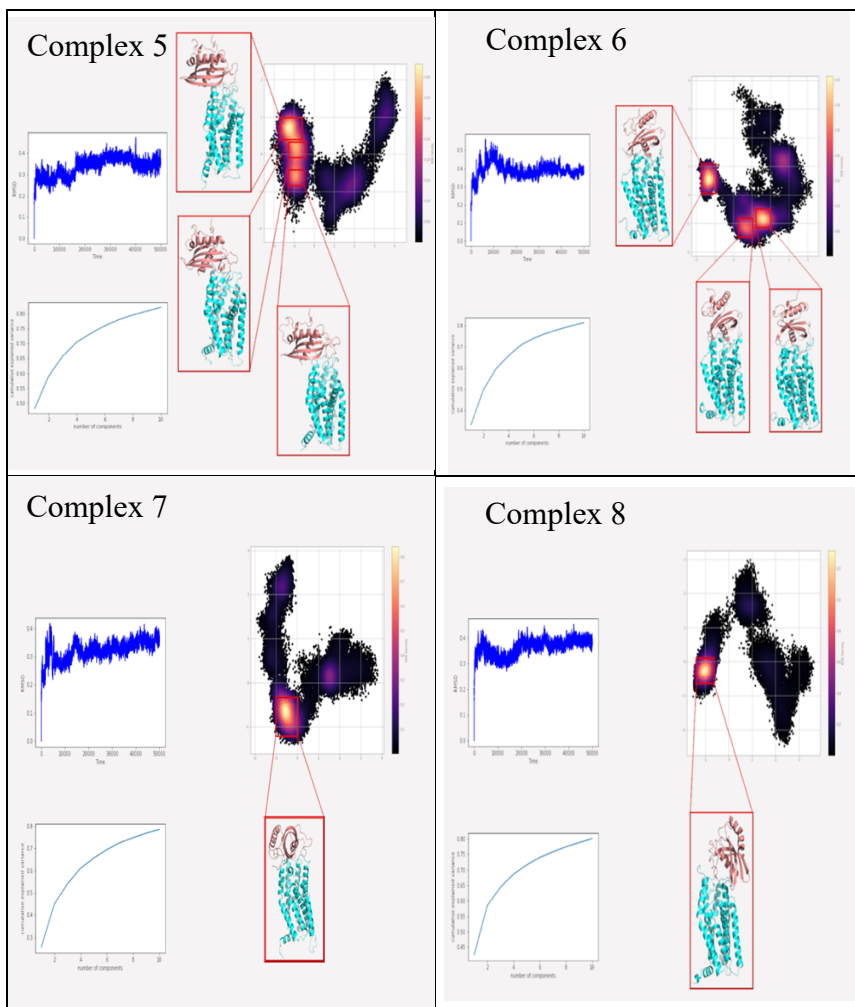
Then, we computed the BSA for each of these conformations. Because of the small dimensions of the interface between CCRL2 and chemerin, conformations with BSA smaller than 600 \AA^2 were excluded⁽⁴¹⁾. The 22

Application of molecular dynamics simulations

remaining conformations were analyzed by visual inspection focusing on the salt bridge interactions. Indeed, the specificity of how proteins interact with other



Application of molecular dynamics simulations



Application of molecular dynamics simulations

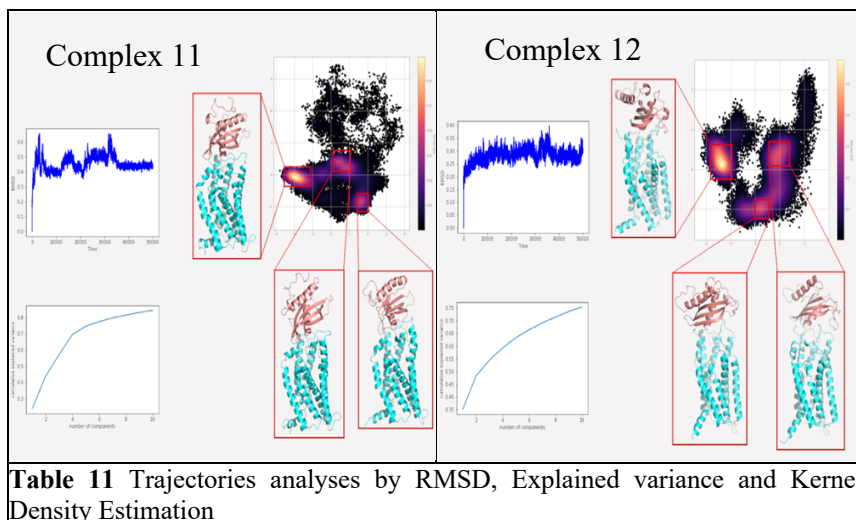


Table 11 Trajectories analyses by RMSD, Explained variance and Kernel Density Estimation

biomolecules is greatly influenced by this type of contact^(42,43).

The examined conformations have been divided into two broad chemerin binding modes. At the same time, the CCRL2 and Chemerin areas that are more frequently engaged in the binding have been identified: the two extracellular loops ECL2 (residues169–192) and ECL3 (residues264–270) for CCRL2, as well as the residue lining the receptor channel's opening; the α 1 helix, the β 1 sheet, and the loop between β 2 and β 3 sheets (β 2 β 3-loop residues 49–73) were the three areas of Chemerin that were most crucial for interaction with the cognate receptor.

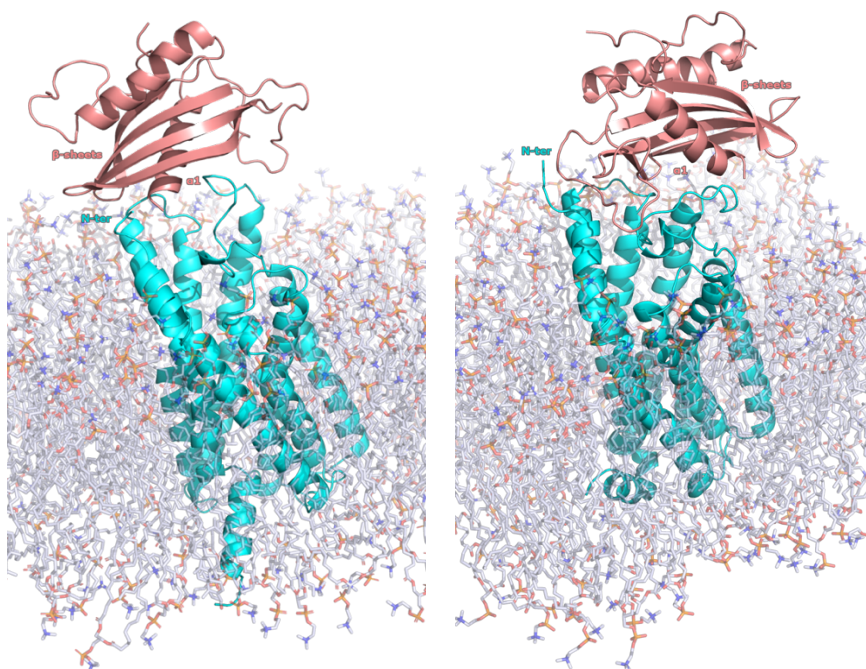
Twelve of the twenty-two examined conformations shared the first binding mode (defined as BM1). The interactions between the chemerin β 2 β 3-loop and ECL3 (6 conformations of12) and ECL2 were indicative of this binding mode (6 conformations of 12). Additionally, the chemerin α 1 helix touch the entrance of the channel (9 conformations of 12).

Application of molecular dynamics simulations

For BM2, which was present in seven of the examined conformations, the chemerin $\beta 2\beta 3$ -loop made contact with the CCRL2 ECL2 and ECL3 in seven of those conformations, the $\alpha 1$ helix made contact with the CCRL2 ECL2 in seven of those conformations, and the $\beta 1$ sheet made contact with the ECL3 and the residues lining in the receptor channel entrance in four of those conformations. The significant contribution of the Chemerin C-terminal domain to the binding to CCRL2 distinguished the three remaining conformations. These three conformations were discarded because it was claimed that the C-terminal was solely involved in the binding of CMKLR1⁽³⁹⁾.

An interesting point to note is that the 180° rotation of the chemerin conformation was the principal difference between the two binding modes, BM1 and BM2. The chemerin $\alpha 1$ helix for the BM1 was located behind the β sheets, as opposed to the BM2 where it was placed in front of the β sheets (Figure 42).

Application of molecular dynamics simulations



Conformation 2 as representative of BM1 Conformation 14 as representative of BM2
Figure 42. Docking proposed chemerin binding modes.

2.5) Proposed interaction models for CCRL2-chemerin binding

The binding residues, the types and frequencies of the identified interactions were examined to acquire additional knowledge. Two patterns of interaction were seen for the BM1. For the first one, we found that the chemerin $\beta 2\beta 3$ -loop make contacts with the ECL2 residues of CCRL2.

The majority of the chemerin $\beta 2\beta 3$ -loop residues were polar, for this reason salt bridges and H-bonds were the most often seen interactions. Indeed, we found a conserved array of polar contacts (6 conformation of 12)

Application of molecular dynamics simulations

Lys60chem with Asp271CCRL2, Lys61chem with Glu265CCRL2, Glu63chem with Lys197CCRL2, and Lys72chem with Asp176CCRL2. Additionally, a hydrophobic interaction between Val66chem and Phe188CCRL2 was noticed (Figure 43 and 44). For the conformation falling within BM1, the second pattern of contacts included a salt bridge between CCRL2 Lys30 and Glu175 and the chemerin α 1 helix residues Glu1 and Arg4, respectively. Chemerin Arg5 also made polar contact with Glu26 or Asp29 of CCRL2. Notably, the chemerin β 1 sheet residues were also involved in contacts with the CCRL2 ECL2, and a polar contact was seen between Glu26chem and Arg185CCRL2. Another polar contact between the chemerin β 2 β 3 loop Lys61 and Glu192 of the CCRL2 ECL2 was observed (Figure 45 and 46).

As a result, the BM1 conformation analysis revealed two key sites known as the first and second pattern of interactions.

We hypothesized that the chemerin β 2 β 3-loop loop might interact with the CCRL2 TM6–TM7 loop, moving the far away from the CCRL2 entrance channel and allowing the chemerin α 1 helix to move toward this channel, even though we did not observe any shifting of one position to the other during the simulation time.

The chemerin β 2 β 3-loop created significant polar connections and hydrophobic contacts in the BM2. Indeed, the chemerin residues Lys60, Lys65, Arg67, and Lys72 established salt bridge with Glu175 of ECL2, Asp32 and Glu26 of TM1, and Asp271 of ECL3, respectively (five conformations of seven).

Application of molecular dynamics simulations

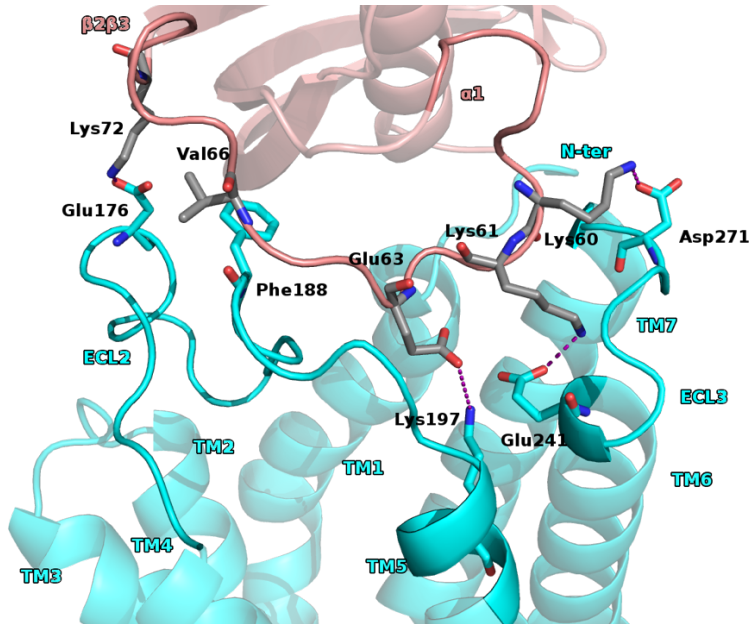


Figure 43. First proposed pattern of interactions for BM1. Proteins are reported as cartoon: pink for chemerin, and cyan for CCRL2. Residues involved in contacts were reported as stick: grey for chemerin and cyan for CCRL2. Salt bridges are depicted as purple dot lines. Membrane is not reported for clarity

Application of molecular dynamics simulations

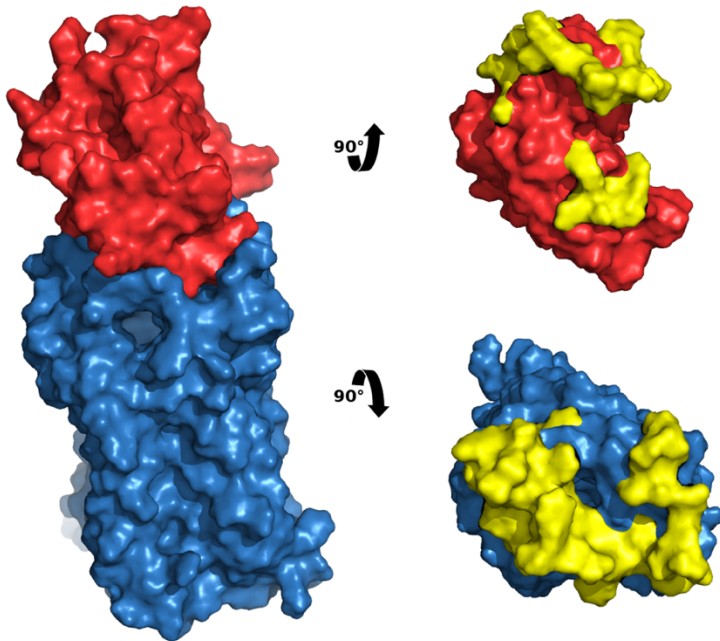


Figure 44. BM1 first proposed pattern of interactions as surface. Chemerin is reported as red surface, CCRL2 is reported as blue surface. In the right panel are depicted the interacting areas, as yellow surfaces, for both chemerin (top) and CCRL2 (down)

Application of molecular dynamics simulations

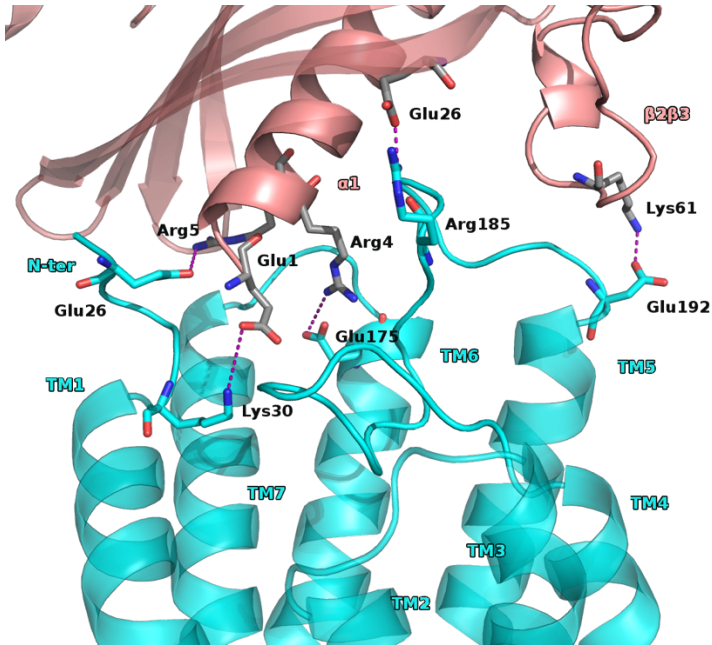


Figure 45. Second proposed patten of interactions for BM1. Proteins are reported as cartoon: pink for chemerin, and cyan for CCRL2. Residues involved in contacts were reported as stick: grey for chemerin and cyan for CCRL2. Salt bridges are depicted as purple dot lines. Membrane is not reported for clarity

Application of molecular dynamics simulations

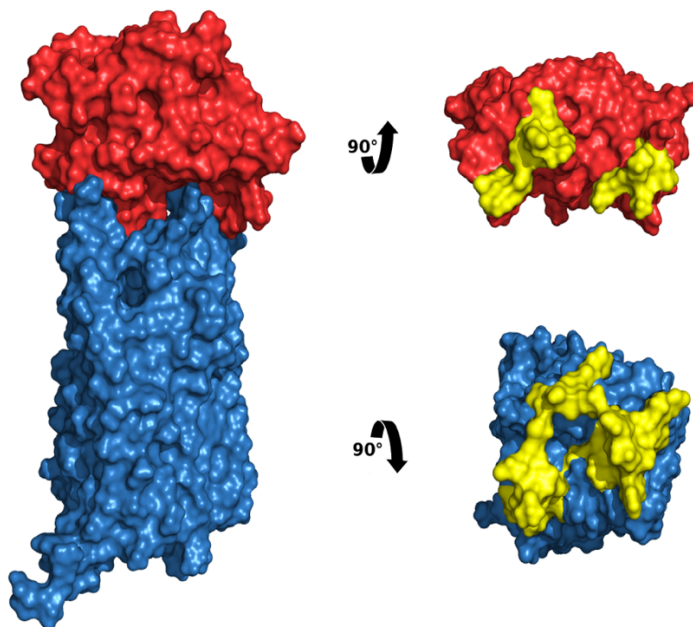


Figure 46. BM1 second proposed pattern of interactions as surface. Chemerin is reported as red surface, CCRL2 is reported as blue surface. In the right panel are depicted the interacting areas, as yellow surfaces, for both chemerin (top) and CCRL2 (bottom).

It is noteworthy that interactions between the chemerin $\beta 2\beta 3$ -loop and the CCRL2 ECL2 appeared to push the latter away from the receptor entrance channel, resulting in a space filled by $\beta 1$ sheet residues (QETSV), forming a salt bridge between Glu322chem and Arg161ECL2 and hydrophobic contacts between Gln321chem and Phe159ECL2 (Figures 47 and 48).

Application of molecular dynamics simulations

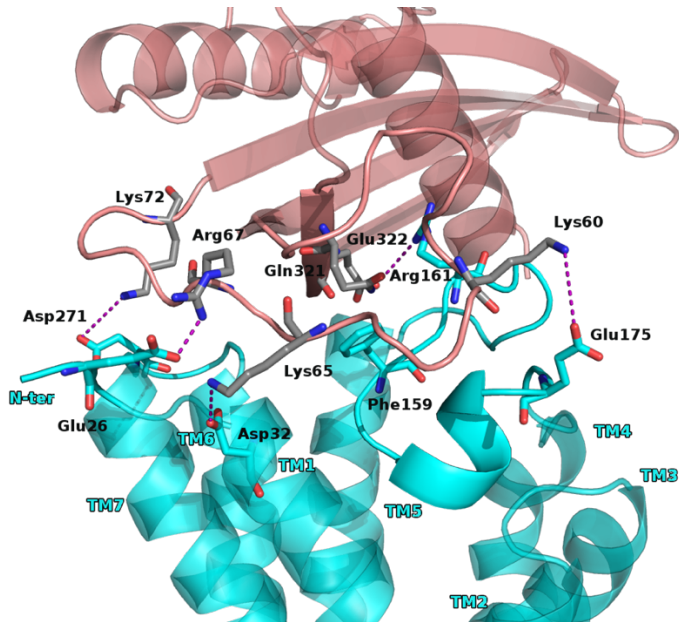


Figure 47. Proposed interaction for BM2. Proteins are reported as cartoon: pink for chemerin, and cyan for CCRL2. Residues involved in contacts were reported as stick: grey for chemerin and cyan for CCRL2. Salt bridges are depicted as purple dot lines. Membrane is not reported for clarity.

2.6) Conclusions

The accomplished simulations allowed us to learn more about how chemerin binds to CCRL2. The findings that have been provided are the earliest efforts to explain the CCRL2 chemerin interaction. Two binding modes for chemerin were found after 5.5 μ s of simulations, and both BMs point to the importance of the chemerin helix α 1, β 1 sheet and β 2 β 3-loop. Additionally, it was proposed that the CCRL2 chemerin complex might arise because of the CCRL2 ECL2 moving away from the receptor entry channel under the influence of the chemerin approach, thus making binding easier. Additionally, a short list of hotspot residues that may be essential for facilitating complex formation and chemotactic activity was obtained by the analysis of the trajectory data. Indeed, we identify for chemerin the α 1 helix Glu1, Arg4, and Arg5, at the β 2 β 3-loop three lysine residues (60, 61, and 65), and for the β 1 sheet Gln25 and Glu26. Furthermore, the ECL2 and ECL3 areas for CCRL2 were identified: the residues Glu175, Asp176, and Asp271 appeared to be important for ECL3. Although these findings still require experimental confirmation, they may aid in elucidating the CCRL2—chemerin interaction. The presented models may also help elucidate the physio-pathological roles of both CCRL2 and chemerin as well as their potential utility as targets for therapeutic intervention. These studies in medicinal chemistry could pave the way for the discovery of modulators of the CCRL2 chemerin interaction.

Bufano M, Laffranchi M, Sozzani S, Raimondo D, Silvestri R, Coluccia A. Exploring CCRL2 chemerin binding using accelerated molecular dynamics. *Proteins*. 2022 Sep;90(9):1714-1720. doi: 10.1002/prot.26348.

2.7) References

1. Bachelierie F, Ben-Baruch A, Burkhardt AM, Combadiere C, Farber JM, Graham GJ, Horuk R, Sparre-Ulrich AH, Locati M, Luster AD, Mantovani A, Matsushima K, Murphy PM, Nibbs R, Nomiyama H, Power CA, Proudfoot AE, Rosenkilde MM, Rot A, Sozzani S, Thelen M, Yoshie O, Zlotnik A. International Union of Basic and Clinical Pharmacology. [corrected]. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors. *Pharmacol Rev.* 2013 Nov 11;66(1):1-79. doi: 10.1124/pr.113.007724. Erratum in: *Pharmacol Rev.* 2014 Apr;66(2):467. PMID: 24218476; PMCID: PMC3880466.
2. Vacchini A, Locati M, Borroni EM. Overview and potential unifying themes of the atypical chemokine receptor family. *J Leukoc Biol.* 2016 Jun;99(6):883-92. doi: 10.1189/jlb.2MR1015-477R. Epub 2016 Jan 6. PMID: 26740381.
3. Hartmann TN, Leick M, Ewers S, Diefenbacher A, Schraufstatter I, Honczarenko M, Burger M. Human B cells express the orphan chemokine receptor CRAM-A/B in a maturation-stage-dependent and CCL5-modulated manner. *Immunology.* 2008 Oct;125(2):252-62. doi: 10.1111/j.1365-2567.2008.02836.x. Epub 2008 Apr 4. PMID: 18397265; PMCID: PMC2561131.
4. Biber K, Zuurman MW, Homan H, Boddeke HW. Expression of L-CCR in HEK 293 cells reveals functional responses to CCL2, CCL5, CCL7, and CCL8. *J Leukoc Biol.* 2003 Aug;74(2):243-51. doi: 10.1189/jlb.0802415. PMID: 12885941.
5. Leick M, Catusse J, Follo M, Nibbs RJ, Hartmann TN, Veelken H, Burger M. CCL19 is a specific ligand of the constitutively recycling atypical human chemokine receptor CRAM-B. *Immunology.* 2010 Apr;129(4):536-

Application of molecular dynamics simulations

46. doi: 10.1111/j.1365-2567.2009.03209.x. Epub 2009 Dec 2. PMID: 20002784; PMCID: PMC2842500.
6. Zabel BA, Nakae S, Zúñiga L, Kim JY, Ohyama T, Alt C, Pan J, Suto H, Soler D, Allen SJ, Handel TM, Song CH, Galli SJ, Butcher EC. Mast cell-expressed orphan receptor CCRL2 binds chemerin and is required for optimal induction of IgE-mediated passive cutaneous anaphylaxis. *J Exp Med*. 2008 Sep 29;205(10):2207-20. doi: 10.1084/jem.20080300. Epub 2008 Sep 15. PMID: 18794339; PMCID: PMC2556791.
 7. Bondue B., Wittamer V., Parmentier M. Chemerin and its receptors in leukocyte trafficking, inflammation and metabolism. *Cytokine Growth Factor Rev*. 2011;22:331–338. doi: 10.1016/j.cytogfr.2011.11.004.
 8. Del Prete A, Locati M, Otero K, Riboldi E, Mantovani A, Vecchi A, Sozzani S. Migration of dendritic cells across blood and lymphatic endothelial barriers. *Thromb Haemost*. 2006 Jan;95(1):22-8. PMID: 16543957.
 9. Sozzani S, Vermi W, Del Prete A, Facchetti F. Trafficking properties of plasmacytoid dendritic cells in health and disease. *Trends Immunol*. 2010 Jul;31(7):270-7. doi: 10.1016/j.it.2010.05.004. Epub 2010 Jun 25. PMID: 20579936.
 10. Tiberio L, Del Prete A, Schioppa T, Sozio F, Bosisio D, Sozzani S. Chemokine and chemotactic signals in dendritic cell migration. *Cell Mol Immunol*. 2018 Apr;15(4):346-352. doi: 10.1038/s41423-018-0005-3. Epub 2018 Mar 21. PMID: 29563613; PMCID: PMC6052805.
 11. Parolini S, Santoro A, Marcenaro E, Luini W, Massardi L, Facchetti F, Communi D, Parmentier M, Majorana A, Sironi M, Tabellini G, Moretta A, Sozzani S. The role of chemerin in the colocalization of NK and dendritic cell subsets into inflamed tissues. *Blood*. 2007 May 1;109(9):3625-32. doi: 10.1182/blood-2006-08-038844. Epub 2007 Jan 3. PMID: 17202316.

Application of molecular dynamics simulations

12. Otero K, Vecchi A, Hirsch E, Kearley J, Vermi W, Del Prete A, Gonzalvo-Feo S, Garlanda C, Azzolino O, Salogni L, Lloyd CM, Facchetti F, Mantovani A, Sozzani S. Nonredundant role of CCRL2 in lung dendritic cell trafficking. *Blood*. 2010 Oct 21;116(16):2942-9. doi: 10.1182/blood-2009-12-259903. Epub 2010 Jul 6. PMID: 20606167; PMCID: PMC3389732.
13. Monnier J, Lewén S, O'Hara E, Huang K, Tu H, Butcher EC, Zabel BA. Expression, regulation, and function of atypical chemerin receptor CCRL2 on endothelial cells. *J Immunol*. 2012 Jul 15;189(2):956-67. doi: 10.4049/jimmunol.1102871. Epub 2012 Jun 13. PMID: 22696441; PMCID: PMC3428203.
14. Gonzalvo-Feo S, Del Prete A, Pruenster M, Salvi V, Wang L, Sironi M, Bierschenk S, Sperandio M, Vecchi A, Sozzani S. Endothelial cell-derived chemerin promotes dendritic cell transmigration. *J Immunol*. 2014 Mar 1;192(5):2366-73. doi: 10.4049/jimmunol.1302028. Epub 2014 Jan 27. PMID: 24470498.
15. Mattern A, Zellmann T, Beck-Sickinger AG. Processing, signaling, and physiological function of chemerin. *IUBMB Life*. 2014 Jan;66(1):19-26. doi: 10.1002/iub.1242. Epub 2014 Jan 20. PMID: 24446308.
16. Del Prete A, Bonocchi R, Vecchi A, Mantovani A, Sozzani S. CCRL2, a fringe member of the atypical chemoattractant receptor family. *Eur J Immunol*. 2013 Jun;43(6):1418-22. doi: 10.1002/eji.201243179. Epub 2013 May 10. PMID: 23580473.
17. Del Prete A, Sozio F, Schioppa T, Ponzetta A, Vermi W, Calza S, Bugatti M, Salvi V, Bernardini G, Benvenuti F, Vecchi A, Bottazzi B, Mantovani A, Sozzani S. The Atypical Receptor CCRL2 Is Essential for Lung Cancer Immune Surveillance. *Cancer Immunol Res*. 2019 Nov;7(11):1775-1788. doi: 10.1158/2326-6066.CIR-19-0168. Epub 2019 Sep 4. PMID: 31484658; PMCID: PMC7176487.

Application of molecular dynamics simulations

18. Mazzon C, Zanotti L, Wang L, Del Prete A, Fontana E, Salvi V, Poliani PL, Sozzani S. CCRL2 regulates M1/M2 polarization during EAE recovery phase. *J Leukoc Biol.* 2016 Jun;99(6):1027-33. doi: 10.1189/jlb.3MA0915-444RR. Epub 2016 Jan 7. PMID: 26744451.
19. Del Prete A, Martínez-Muñoz L, Mazzon C, Toffali L, Sozio F, Za L, Bosisio D, Gazzurelli L, Salvi V, Tiberio L, Liberati C, Scanziani E, Vecchi A, Laudanna C, Mellado M, Mantovani A, Sozzani S. The atypical receptor CCRL2 is required for CXCR2-dependent neutrophil recruitment and tissue damage. *Blood.* 2017 Sep 7;130(10):1223-1234. doi: 10.1182/blood-2017-04-777680. Epub 2017 Jul 25. PMID: 28743719.
20. Salvi V, Sozio F, Sozzani S, Del Prete A. Role of Atypical Chemokine Receptors in Microglial Activation and Polarization. *Front Aging Neurosci.* 2017 May 26;9:148. doi: 10.3389/fnagi.2017.00148. PMID: 28603493; PMCID: PMC5445112.
21. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys.* 2004 Jun 22;120(24):11919-29. doi: 10.1063/1.1755656. PMID: 15268227.
22. Bucher D, Pierce LC, McCammon JA, Markwick PR. On the Use of Accelerated Molecular Dynamics to Enhance Configurational Sampling in Ab Initio Simulations. *J Chem Theory Comput.* 2011 Apr 12;7(4):890-897. doi: 10.1021/ct100605v. Epub 2011 Mar 4. PMID: 21494425; PMCID: PMC3074571.
23. Maisuradze GG, Liwo A, Scheraga HA. Principal component analysis for protein folding dynamics. *J Mol Biol.* 2009 Jan 9;385(1):312-29. doi: 10.1016/j.jmb.2008.10.018. Epub 2008 Oct 15. PMID: 18952103; PMCID: PMC2652707.
24. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intel Lab Syst.* 1987;2(1-3):37-52

Application of molecular dynamics simulations

25. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004 Oct;25(13):1605-12. doi: 10.1002/jcc.20084. PMID: 15264254.
26. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc.* 2003 Feb 19;125(7):1731-7. doi: 10.1021/ja026939x. PMID: 12580598.
27. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem.* 2008 Aug;29(11):1859-65. doi: 10.1002/jcc.20945. PMID: 18351591.
28. Lee J, Hitznerberger M, Rieger M, Kern NR, Zacharias M, Im W. CHARMM-GUI supports the Amber force fields. *J Chem Phys.* 2020 Jul 21;153(3):035103. doi: 10.1063/5.0012280. PMID: 32716185.
29. Lee J, Patel DS, Stähle J, Park SJ, Kern NR, Kim S, Lee J, Cheng X, Valvano MA, Holst O, Knirel YA, Qi Y, Jo S, Klauda JB, Widmalm G, Im W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J Chem Theory Comput.* 2019 Jan 8;15(1):775-786. doi: 10.1021/acs.jctc.8b01066. Epub 2018 Dec 28. PMID: 30525595.
30. Darden T, York D, Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys.* 1993;98:10089-10092
31. Gedeon PC, Thomas JR, Madura JD. Accelerated molecular dynamics and protein conformational change: a theoretical and practical guide using a membrane embedded model neurotransmitter transporter. *Methods Mol Biol.* 2015;1215:253-87. doi: 10.1007/978-1-4939-1465-4_12. PMID: 25330967; PMCID: PMC4518716.

Application of molecular dynamics simulations

32. Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. COCOMAPS: a web application to analyse and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*. 2011;27(20):2915-2916.
33. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*. 2015 Oct 20;109(8):1528-32. doi: 10.1016/j.bpj.2015.08.015. PMID: 26488642; PMCID: PMC4623899.
34. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Machine Learn Res*. 2011;12:2825-2830
35. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods*. 2020;17(3):261-272
36. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95
37. Miles TF, Spiess K, Jude KM, Tsutsumi N, Burg JS, Ingram JR, Waghray D, Hjorto GM, Larsen O, Ploegh HL, Rosenkilde MM, Garcia KC. Viral GPCR US28 can signal in response to chemokine agonists of nearly unlimited structural degeneracy. *Elife*. 2018 Jun 8;7:e35850. doi: 10.7554/eLife.35850. PMID: 29882741; PMCID: PMC5993540.
38. Allison JR. Computational methods for exploring protein conformations. *Biochem Soc Trans*. 2020 Aug 28;48(4):1707-1724. doi: 10.1042/BST20200193. PMID: 32756904; PMCID: PMC7458412.
39. De Henau O, Degroot GN, Imbault V, Robert V, De Poorter C, Mcheik S, Galés C, Parmentier M, Springael JY. Signaling Properties of Chemerin Receptors CMKLR1, GPR1 and CCRL2. *PLoS One*. 2016 Oct 7;11(10):e0164179. doi: 10.1371/journal.pone.0164179. PMID: 27716822; PMCID: PMC5055294.

Application of molecular dynamics simulations

40. Wolf A, Kirschner KN. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J Mol Model*. 2013 Feb;19(2):539-49. doi: 10.1007/s00894-012-1563-4. Epub 2012 Sep 8. PMID: 22961589; PMCID: PMC3592554.
41. Chen J, Sawyer N, Regan L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci*. 2013 Apr;22(4):510-5. doi: 10.1002/pro.2230. Epub 2013 Feb 26. PMID: 23389845; PMCID: PMC3610057.
42. Bush J, Makhatadze GI. Statistical analysis of protein structures suggests that buried ionizable residues in proteins are hydrogen bonded or form salt bridges. *Proteins*. 2011 Jul;79(7):2027-32. doi: 10.1002/prot.23067. Epub 2011 May 10. PMID: 21560169.
43. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci*. 1994 Feb;3(2):211-26. doi: 10.1002/pro.5560030206. PMID: 8003958; PMCID: PMC2142793.
44. Im W, Roux B. Ions and counterions in a biological channel: a molecular dynamics simulation of OmpF porin from *Escherichia coli* in an explicit membrane with 1 M KCl aqueous salt solution. *J Mol Biol*. 2002 Jun 21;319(5):1177-97. doi: 10.1016/S0022-2836(02)00380-7. PMID: 12079356

Laboratory equipment

Laboratory equipment

In our lab we have two SuperMicro computers with Intel® Xeon R Silver 4314 CPU @ 2.40GHz x 64 and one GPU Nvidia A30 running Ubuntu 20.04.5 LTS.

We used also two MacPro dual 2.66 GHz Xeon running Ubuntu 14 LTS.

Conclusions and future perspectives

Conclusions and future perspectives

This thesis focuses on the application of molecular modeling during the early stages of drug discovery and the use of molecular dynamics simulations to predict ligand-protein or protein-protein stable interactions.

Because of their comparatively higher efficiency and cheaper costs compared to traditional experiments, *in silico* methods, particularly docking and scoring in virtual screening, are crucial during this lead compound design phase.

Molecular docking is one of the most widely utilized techniques due to its capacity to predict the conformation of small molecule ligands within the proper target binding site. For instance, investigations involving important molecular events, such as ligand binding modes and the intermolecular interactions that stabilize the ligand-receptor complex, can be easily carried out. This technique is part of the virtual screening procedures that aims to identify novel bioactive compounds from huge chemical libraries using knowledge about the protein target (structure-based VS) or already-existing bioactive ligands (ligand-based VS).

We developed a protocol featured by 5 steps: identification of the target, validation of docking software, docking of compounds libraries, pharmacophore generation and filter, visual inspection. It was applied on targets involved in the development of cancer disease.

The first VS focus the attention on a pharmacological approach used to treat cancer that tries to stop the interaction between Gab2 and C-SH3 domain of Grb2. We identify a compound, AN-465-J137-985, that significantly lowers the affinity of the C-SH3 for Gab2 in binding and displacement experiments. Also, in cellula experiments, on A549 and H1299 overexpressing the C-SH3 fragment of Grb2, validate its antitumorigenic activity.

We applied structure-based VS on PDZ domain of DVL1 and we found that the (S)-RS4690 showed inhibitory activity on

Conclusions and future perspectives

colorectal cancer cell HCT116, SW480, and SW620 demonstrating the inhibition of the WNT pathway and a selectivity toward the PDZ domain of NHERF1.

With the same protocol we identify compound 18, that specific inhibits LDHA in several cancer cells, CRC HCT116, SW620, A549 and PANC-1. Moreover, compound 18 consistently increased NADH content and decreased lactate levels in tumor cells.

Furthermore, molecular dynamics simulations could be complementary to classical molecular docking techniques with the possibility to examine the binding process to shed light on stable state reached by the ligand or conformational rearrangements occurring before, during, or after binding. Indeed, we used Co-solvent molecular dynamics (CMD) to search for the allosteric binding site of the AM-001 inhibitor of EPAC1 analyzing the active and inactive conformations and generating their cosolvent occupancy maps in 100ns and 250 ns trajectories. Occupancy maps and PCA analyses helped us to identify the pocket at the interface between CDC25-HD and CNBD as suitable binding site.

Additional methods in MD helped us to increase the accessible timescale, increase the sampling and overcoming high energy barriers. This led us to explore infrequent events like protein-protein binding interactions. We used accelerated molecular dynamics on the CCRL2-Chemerin complex to let the system exit from the local minima adding a bias potential. Trajectories analyses highlighted hot-spot residues more frequently implicated in binding interactions.

In conclusions, several techniques are available for researching the dynamic properties of pharmacological targets, predicting and characterizing binding sites, finding new active molecules, and optimizing existing ones. One of these, structure-based Virtual Screening, gave us the opportunity to identify molecules of pharmacological interest that are considered as lead compounds and will be optimized to enhance their inhibitory activity and improve their pharmacokinetics profiles (drug like properties).

Conclusions and future perspectives

Figures index

Figures index

Figure 1 Pharmacophore model for Grb2 C-SH3:Gab2 complex	54
Figure 2 AN-465-J137-985 proposed binding mode	56
Figure 3 Equilibrium binding titration of 5 μ M AN-465-J137-985	58
Figure 4 Kinetic binding experiments at different concentrations of Gab2	60
Figure 5 Effect of AN-465-J137-985 in A549 and H1299 lung cancer cells	63
Figure 6 Map of the pharmacophore features superimposed with the RS5517 binding mode	77
Figure 7 In silico docking results of (<i>S</i>)- 1 and (<i>R</i>)- 1 in complex with the DVL1 PDZ binding site	78
Figure 8 A Snapshot of the trajectory of RS4690 S with DVL1 PDZ	79
Figure 8 B Snapshot of the trajectory of RS4690 R with DVL1 PDZ	80
Figure 9 Equilibrium binding experiment between PDZ domain of DVL1 and the C-terminal portion of TMEM88	81
Figure 10 The dose response curves are representative for DVL recruitment inhibition experiment and WNT pathway activity measurement	81
Figure 11 Confocal immunofluorescence in the presence of 7 μ M of compound 1 (racemic mixture), (<i>S</i>)- 1 , (<i>R</i>)- 1	83
Figure 12 ROS production in HCT116 cells	85
Figure 13 Key contacts of compound (<i>S</i>)- 2 in the DVL1 PDZ binding pocket	86
Figure 14 General reactions catalyzed by LDHA and LDHB	96
Figure 15 Chemical structures of LDHA inhibitors 1-12	100
Figure 16 Testing of LDHA inhibitor in cancer cells	106
Figure 17 Measurement of LDH activity, glycolytic rate and NADH content after treatment with compound 18	109
Figure 18 Effect of compound 18 on cell proliferation and apoptosis	112
Figure 19 Schematic picture of the involvement of 18 and rotenone in apoptosis	114
Figure 20 Proposed binding mode of 18	115

Figures index

Figure 21 Superimposition of 18 and derivative 7, 10, 11	116
Figure 22 Summary of the binding interactions of compound 18 into the LDHA binding site	116
Figure 23 Snapshot of molecular dynamic simulations	117
Figure 24 RMSD values during molecular dynamics simulation	118
Figure 25 Estimated toxicity of compound 18	119
Figure 26 Graphical representation of EPAC activation upon cAMP binding	135
Figure 27 A, B RMSD of Ca atom of Epac1 active and inactive conformation simulations	138
Figure 28 Epac1 Cosolvents occupancy maps for active conformation	142
Figure 29 Epac1 Active conformation cosolvent occupancy maps selected areas	143
Figure 30 Epac1 active conformation with Rap domain	144
Figure 31 Epac1 Inactive conformation cosolvent occupancy maps selected zones	146
Figure 32 Epac1 sequence and secondary structure	147
Figure 33 Epac1 Area5 pose4_5 and pose9_5 binding modes	148
Figure 34 Epac2 Area5 pose4_5 and pose9_5 binding mode	149
Figure 35 Proposed binding mode for AM-001	151
Figure 36 Plants proposed binding mode of AM-001 analogues	152
Figure 37 Porcupine plot of the top two eigenvectors	157
Figure 38 PCA analyses	158
Figure 39 Superimposition of CCRL2 alphafold and in-house homology model	179
Figure 40 Superimposition of chemerin alphafold and in-house Model	179
Figure 41 Docking proposed binding conformation for Chemerin and CCRL2	182
Figure 42 Docking proposed chemerin binding modes	187
Figure 43 First proposed patter of interactions for BM1	189
Figure 44 BM1 first proposed pattern of interactions as surface	190
Figure 45 Second proposed patter of interactions for BM1	191
Figure 46 BM1 second proposed pattern of interactions as Surface	192
Figure 47 Proposed interaction for BM2	197

Figures index

Figures index

Tables index

Table 1 DVL Binding and WNT Pathway Inhibition by Compounds 1 , (<i>S</i>)- 1 , (<i>R</i>)- 1 and 2	82
Table 2 Growth Inhibition of SW680, SW620 and HCT116 cell lines	84
Table 3 Structures of compounds 14–22 selected by virtual screening studies	104
Table 4 Inhibition of growth of Med1-MB cell line by compounds 14-22	107
Table 5 Inhibition of LDHA by compound 18 and reference oxamate	107
Table 6 Drug like properties of compound 11	119
Table 7 Structures and Bret ratio variation of AM-001 analogues	155
Table 8 Correlation between BRET-ratio values and calculated DG (Kcal/mol) of docking proposed binding poses for AM-001 analogues	156
Table 9 Values of the ten eigenvectors, eigenvalues and relative proportion of variance	158
Table 10 Information on each equilibration steps	176
Table 11 Trajectories analyses by RMSD, Explained variance and Kernel Density Estimation	185

Tutti i diritti riservati

“Il presente documento è distribuito secondo la licenza Tutti i diritti riservati.”