MASTERARBEIT | MASTER'S THESIS

Titel | Title

Ligand-based and structure-based studies of connexin 36 modulators

verfasst von | submitted by

Ekaterina Danilkina-Timetskaia

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 606
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Drug Discovery and Development
Betreut von | Supervisor:	Assoz. Prof. Mag. Dr. Johannes Kirchmair

Acknowledgements

The two years of my master's program have flown by, and as I look back, I am amazed at the progress I have made—a journey that would not have been possible without the opportunity to study and the support of people around me.

I am immensely grateful to the staff at the Faculty of Pharmacy for this opportunity, to the professors whose lectures are still remembered with warmth, and to the Drug Discovery Programme group with whom we spent countless hours in the cafeteria doing group projects.

Special thanks to the entire COMP3D group, whose lunchtime conversations and events inspired me to become better. Within the framework of this project, I must particularly acknowledge Ya Chen (Anya), whose companionship during our tea time with experiments was both enjoyable and productive.

A huge thank you goes to Dr. Remy Hoffmann and Dr. Fernando Martin from Pharmacelera for providing us free access to their programs PharmScreen and exaScreen, as well as their support throughout the entire research project. Special thanks also go to Dr. Volodimir Korhov from the Paul Scherrer Institute (PSI) for providing essential information about connexin 36 and its structures.

Above all, my deepest gratitude and the very fact that I can write these lines are all thanks to the efforts of Professor Johannes Kirchmair. Your patient guidance, timely advice and also collaborative brainstorming always helped to set the research on the “right path”. I remember how, after your lecture at the beginning of last spring on toxicity prediction, I developed an interest in computational methods and never regretted it. Thank you for your unwavering support and for welcoming me to your exceptional team.

I also want to extend my heartfelt thanks to my moral support, without which I would not have been able to write this work during nights - my dear parents, who supported the fact that I would become the first master of science in the family, my friends and colleagues, as well as my dear fiancé Eddie, who cared about me even when I forgot about it myself.

These two wonderful years have given me so much, and I am immensely grateful for that. Thank you to everyone who reads this acknowledgement for being part of this remarkable journey.

Abstract

The pharmaceutical industry seeks to minimise financial costs while speeding up the drug development process. One solution for this problem is computer-aided drug design. It includes ligand-based and structure-based approaches, each of which, despite its advantages, also has disadvantages that are avoided by implementing combinational strategies.

Ligand-based and structure-based methods are presented here to investigate a potential inhibitor for the connexin 36 channel. This transmembrane protein is responsible for synaptic transmission and the activation of neurons in the brain. The compound serving as a starting point for these studies is mefloquine, a synthetic derivative of quinoline, an antimalarial drug.

As a ligand-based method, Pharmacelera's PharmScreen and exaScreen software packages were used, exploiting a new approach to virtual screening. The theory relies on the principles of molecular similarity and semi-empirical calculations based on quantum mechanics. This method allows obtaining a variety of hits structurally independent of the reference compound. The following libraries were used: Molport, Enamine, and DrugBank. GOLD was used as the structural-based method focusing on the Molport library.

The virtual screening identified 140,954 compounds using the ligand-based method and 202,467 hits from the structure-based method, which were then subjected to a thorough structural analysis. A list of 28 promising structures was compiled, which will subsequently be biologically tested for their ability to inhibit connexin 36.

Thus, this study demonstrates the importance of combining two virtual screening methods. This combination overcomes the individual limitations of each technique and allows to obtain the greatest accuracy in identifying new potential ligands.

Kurzfassung

In der Pharmaindustrie besteht der Wunsch, die finanziellen Kosten zu reduzieren und gleichzeitig den Entwicklungsprozess zu beschleunigen. Eine Lösung für dieses Problem ist ein computergestütztes Wirkstoffdesign. Es umfasst ligandenbasierte und strukturbasierte Ansätze, die trotz ihrer Vorteile jeweils auch Nachteile haben. Um Einschränkungen zu beseitigen, wurden verschiedene Kombinationsstrategien entwickelt.

Hier werden ligandenbasierte und strukturbasierte Methoden angewandt, um potenzielle Inhibitoren für den connexin 36-Kanal zu untersuchen. Dies ist ein Transmembranprotein, das für die synaptische Übertragung sowie die Aktivierung von Neuronen im Gehirn verantwortlich ist. Die Referenzkomponente war Mefloquin, ein synthetisches Derivat von Chinolin, einem Antimalariamittel.

Als ligandenbasierte Methode wurden die von Pharmacelera entwickelten Softwarepakete Pharmscreen und exaScreen verwendet, welche einen neuen Ansatz für das virtuelle Screening nutzen. Die Theorie basiert auf den Prinzipien der Ähnlichkeit sowie auf semiempirischen Berechnungen auf der Grundlage der Quantenmechanik. Mit dieser Methode können vielfältige Treffer erzielt werden, die strukturell unabhängig von der Referenzverbindung sind. Die folgenden Bibliotheken wurden verwendet: Molport, Enamine und Dragbank. Als strukturbasierte Methode wurde GOLD mit Schwerpunkt auf der Molport-Bibliothek verwendet.

Das virtuelle Screening identifizierte 140.954 Liganden mit der ligandenbasierten Methode und 202.467 Treffer mit der strukturbasierten Methode, die anschließend einer gründlichen Strukturanalyse unterzogen wurden. Durch die Durchführung einer vergleichenden Analyse jeder Methode wurde eine kombinierte Liste von 28 vielversprechenden Strukturen zusammengestellt. Anschließend würden sie biologisch auf ihre Fähigkeit getestet, connexin 36 zu hemmen.

Somit zeigt diese Studie die Bedeutung der Kombination zweier virtueller Screening-Methoden. Dies führt dazu, dass die individuellen Einschränkungen jeder Methode überwunden und neue potenzielle Liganden mit größter Genauigkeit identifiziert werden.

Contents

Introduction

1.1. Computer-aided drug design

Several types of research problems in medicine and pharmaceuticals can be distinguished. Many of these challenges are driven by the urgent need to develop new drugs that target diseases with the potential to become breakthrough therapies, as well as to improve the efficacy and safety of clinically tested drugs. These stages of drug testing, which take place in the laboratory and clinical settings, are considered a filtering process - complex selection, as well as preparation of potential candidates, sometimes taking decades. Therefore, it is important for the pharmaceutical industry to minimise financial costs in terms of profitability and speed up the process.

One advanced method is computer-aided drug design (CADD) or molecular modelling [1]. The drug development and analysis process is cyclical and is carried out in silico, that is, using computer technology to consider various probabilities of events. One advantage is the possibility of evaluating the biological activity, bioavailability, toxicity, and activity of the selected compound relative to the chemical structure long before clinical trials are carried out [2]. With this method, only a few pre-selected compounds will be evaluated instead of sending thousands of compounds for testing. This can reduce the costs of an investigation substantially.

As a starting point, compound identification was conducted experimentally using high-throughput screening (HTS) of virtual libraries [3]. This method aims to identify interactions between compounds and active binding sites. The process involves several stages, including sample preparation, processing, test development, analysis systems, and data collection [4]. Despite technological progress, which has made it possible to automate some production processes, this testing type is still equally defined as labour-intensive, time-consuming, and requiring special conditions. In addition, the results obtained require repeated testing to avoid false positive or false negative results. As a result, carrying out the HTS becomes almost impossible in an academic environment. This is why an additional method is generally preferred - targeted virtual screening or VS.

Virtual screening is used to identify potential targets during early screening, which minimises costs since, as a result, only 0.1 - 2.5% of the total amount, the highest rated substances according to function, are selected for testing. [5] With this method, it turns out that the difference between high-throughput screening and virtual screening is that the first is an experimental method, and “virtual”, based on its name, is theoretical. Therefore, several preparations should be done before:

- A collection of compounds (library);
- The structure of the proposed biological target, preferably 3D (crystal) [6];
- Choosing a feature to accurately estimate a suitable docking pose is important, as there is a risk of losing potential hits.

Virtual screening is usually divided into two types: ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS) [7].

1.1.1. Structure-based virtual screening

Each project is unique, so a method of analysis must be chosen depending on the available information regarding the interested structure.

For example, structure-based virtual screening relies on the actual 3D structure of the target [8]. At the same time, structure can be determined in different ways, but the main ones are X-ray crystallography and nuclear magnetic resonance. Information about structures can be found on relevant Internet resources, including the Protein Data Bank (PDB). The most popular method is molecular docking, which aims to predict the potential location of a ligand in the binding pocket based on the structural and chemical complementarity resulting from the interaction. For characterisation, evaluation functions are used, which are also supplemented with pharmacophore restrictions. The results can be further ranked for better interpretation, and the scoring systems themselves are divided into several categories: force-field-based, knowledge-based, and empirical scoring functions [9]. For better results, the structure of the intended target is also considered, along with water or metals, if any. Ideally, donor-acceptor interactions should be observed between hit and biological targets with hydrophobic properties and accessible surfaces for attachment.

1.1.2 Ligand-based virtual screening

Ligand-based virtual screening (LBVS) is based on structural information. Still, the primary role is played by the physicochemical properties of the constituent molecules, which are then considered in relation to the principle of similarity [10]. Within the framework of "similarity," it is assumed that the structures found must have similar properties to identify structure-activity relationships, obtain pharmacophores, and perform subsequent measurements. There are several options for calculating molecular similarity: using molecular descriptors (characteristics) and similarity functions (comparing pairs using numbers in the range from 0 to 1) [11].

Molecular descriptors are divided into several types, depending on the number of dimensions:

- 1D descriptors provide a simple representation that does not include information about atomic bonds. These include physicochemical or biological properties and chemical characteristics (for example, the number of functional groups) [12].
- 2D descriptors are considered one of the most common methods based on bond interactions, electrical properties of atoms in a topological coordinate system and topological descriptors (electron availability) [11].
- 3D molecular descriptors include calculations based on fields, shapes and volumes, as well as pharmacophores; such representations are considered closer to the ligand's

interaction with the receptor since they provide high-affinity structures based on a new basis [13].

Similarity functions represent a quantitative measure of chemical similarity from several points of view: a global one—if the whole molecule is considered—or a local one—when a specific region, for example, a functional group or a heteroatom, must be considered. The Tanimoto coefficient is the most widely used function, which combines similarity for binary fingerprints and continuous data. Such a condition is defined in Equation 1 [14].

$$Similarity\ coefficient\ a, b\ = \frac{c}{\alpha(a-c) + \beta(b-c)+c}$$

Equation 1. Measurement of similarity coefficient.

$\alpha = \beta = 1$. It is also known as the Dice ratio when $\alpha = \beta = \frac{1}{2}$, where *a* is the number of features in molecule A, *b* is the number of features in molecule B, and *c* is the number of common features in molecules A and B.

If partial similarity is needed, this equation could be used, but the number of features, for example, in molecule B, will be zero, leading to the Tversky coefficient [15]. The results can then be ranked to identify more promising hits. However, it is advised to be careful with this method since a high value may not always reflect an accurate picture of what is happening, so similarity-based methods are recommended depending on the direction of the research.

1.1.3. Combined virtual screening approaches

Despite the listed advantages and characteristics of the LB and SB methods, the accuracy of their predictive power is limited. For example, in the case of ligand-based methods, the result will depend entirely on the quality of the descriptors, the conformation of the reference and the parameters of the properties used for comparison [16]. If methods based on structure are considered, then they will also depend on the initial data - the structure of the protein, in which conformational states with structural water must be taken into account. Accordingly, combined approaches have been proposed to address the shortcomings of both methods.

One of the classifications of combined methods was proposed by Drval and Griffith, which distinguished three main categories: sequential, parallel, and hybrid [17].

1. The sequential approach attempts to reduce the cost of structural screening methods by applying pre-filtering by LB methods to obtain the most promising candidates, which are then further evaluated. However, this method can only use some available information, which does not affect already-known limitations [18].
2. The parallel approach uses two methods independently, and then two batches of the best candidates are selected and subjected to biological testing. The results obtained are then

aligned and ranked, providing increased reliability of the results, as well as reproducibility when compared with single-modal methods. However, it still depends on the sensitivity to the nature of the target and the binding pocket [19].

3. The hybrid approach is the combination of LB and SB, which is divided into two main ones:
    A. Interaction-based methods translate protein-ligand interactions into pharmacophore signatures and quantitative structure-activity relationship (QSAR) models, which can then be applied in ligand profiling or pseudoreceptor analysis [20].
    B. Similarity docking method - uses a combination of the concept of molecular similarity and docking. It assesses the reliability of proposed poses by connecting and aligning them to the existing templates [21].

1.2. PharmScreen and exaScreen: theory

Filtering is one of the essential steps used in the industry to reduce costs, as only several of the compounds would proceed for further testing. One of the methods used is molecular alignment, which assesses the similarity between compounds and determines pharmacophores. However, it is noted that the very concept of "molecular similarity" can be considered subjective since its assessment directly depends on descriptors, including functions. 3D-based similarity methods rely on molecular geometry. Non-superpositional methods analyse atomic distances to a set of reference positions. In contrast, superpositional methods depend on correctly positioning the molecules to maximise the overlap of their shapes and/or properties as represented by pharmacophores [22]. At the moment, quite a few programs use various specific properties, such as the superposition of chemical structures using the Gaussian function in ROCS [23], and molecular fields in spatial energy variations in the Field Screen [24]. Thus, it can be concluded that the shape criterion and electrostatic interactions are the most popular methods for calculating similarity, but this does not consider other parameters that may affect the binding. One of these criteria is usually identified as (de)solvation, which relates to both the ligand and the receptor [25]. Some studies confirm that the interaction of the ligand with the active site depends on the hydrophobic effect. In contrast, polar interactions model the kinetics and direction of the ligand acting as an "anchor" [26].

Based on these considerations, the next parameter of interest that should be touched upon is lipophilicity, which has a primary role in predicting the pharmacokinetics of a drug. It is common to use the molecular lipophilicity potential (MLP) and the hydrostatic interaction index (HINT). The first method is based on a quantitative description of lipophilicity relative to the structure and the surrounding space, making it attractive for 3D-QSAR and docking [27]. HINT provides an empirical yet quantitative estimate of the ligand-receptor complex formed through the sum of interactions derived from experimental octanol/water partition coefficient (logPo/w) data. In this case, the contributions of both enthalpy and entropy are considered, and the indicator

itself is found on the application pages in the bimolecular structure [28]. One of the new strategies for assessing molecular similarity is the use of 3D maps of the distribution of hydrophobicity and control of the overlap depending on the hydrophobic topology because, in this case, instead of empirical data, quantum mechanical theoretical calculations (solute water - n-octanol) are used. They are characterised by Miertus-Scrocco-Tomasi (MST) continuum solvation in self-consistent reaction fields [29].

If solvation in more detail is considered, obtaining 3D maps can be determined by atomic contributions to logP o/w. The free energy of solvation ($\Delta Gsol$) will be represented here as the sum of three compounds in Equation 2. The first is the cavitation term, which embodies the necessary work required to create a cavity ($\Delta Gcav$); the second is the Van der Wals term, characterising dispersion-repulsion in a solvent medium and solute (ΔGvw); the third is an electrostatic term for measuring the work expended to create a charge distribution in a solvent-solute system ($\Delta Gele$) [30].

$$\Delta Gsol = \Delta Gcav + \Delta Gvw + \Delta Gele$$

Equation 2. Solvation-free energy MST.

Subsequently, each of the parts was revealed separately. $\Delta Gele$ is transformed in Equation 3 since it affected the electrostatic effects of the solvent, taking into account *N* - the total number of atoms, *M* - the total number of charges of the reaction fields ($Q\,j,sol$), *Wo* - the wave function of the solute in the gas phase, and i is calculated based on the electrostatic potential the molecule itself [31].

$$\Delta Gele = \sum_{i=1}^{N} \Delta Gele, i = \sum_{i=1}^{N} \sum_{j=1, j \epsilon i}^{M} \left[\psi o \left(\frac{1}{2} \frac{Q\,j,sol}{Rj - Ri} \right) \psi o \right]$$

Equation 3. Expanding the meaning of the electrostatic term.

$\Delta Gcav$ and ΔGvw follow an expression that relates linearly from the surface of each atom and from the accessibility of the molecule to the solvent in Equations 4 and 5, where $\Delta Gp, i$ - is the free energy of cavitation, *i* is defined by the Pierotti formalism, which acts on the total surface - *ST*, and to the available one - *Si*, ξi - atomic surface tension [32].

$$\Delta Gcav = \sum_{i=1}^{N} \Delta Gcav,\ i = \sum_{i=1}^{N} \frac{Si}{ST} \Delta Gp, i$$

Equation 4. Expanding the meaning of the cavitation term.

$$\Delta Gvw = \sum_{i=1}^{N} \Delta Gvw,\ i = \sum_{i=1}^{N} \xi i\ Si$$

Equation 5. Expanding the meaning of the Van der Waals term.

Based on Equations 3, 4, and 5, the total free energy of solvation has the form presented in Equation 6 [33].

$$\Delta Gsol = \sum_{i=1}^{N} \Delta Gsol = \sum_{i=1}^{N} (\Delta Gcav,\ i + \Delta Gvw,\ i + \Delta Gele, i)$$

Equation 6. Total free energy of solvation.

If solvation of the compound in water and octanol is applied to Equation 6, then the energies change to $\Delta Go/w$, but all other parts remain the same. When the hydrophobicity of a molecule needs to be used, it will be expressed in terms of the logarithm of octanol/water (*log P*) in Equation 7, where *R* is Boltzmann's constant and *T* is the absolute temperature [34].

$$logPx = \sum_{i=1}^{N} logPxi = \sum_{i=1}^{N} - \frac{\Delta Gx, i\ (o/w)}{2.303RT}\ \ x\colon (ele,\ cav, vw)$$

Equation 7. Atomic contribution to LogP.

This method provides a fractional contribution to lipophilicity, including the influence of the entire molecule on the contribution of a single atom. These functions are independent of the existence of suitable chemical parameters for the studied groups, which may be absent in empirical databases.

1.2.1 PharmScreen

The PharmScreen software package from Pharmacelera was conceived to solve the problem of the need for more diversity of molecular scaffolds that chemists could reproduce. The main application is 3D virtual screening using a backbone of already-known ligands or potential targets. The principle is based on using a 3D representation of molecules derived from the interaction of electrostatic, steric, and hydrophobic fields obtained through semi-empirical calculations based on quantum mechanics, described in section 1.2.1 [35] [36] [37]. The resulting fields are highly sensitive to ligand-receptor interactions. In addition, since the descriptors used to be chemotype-independent, the results obtained are considered in terms of activity rather than structural similarity. This approach makes it possible to accurately identify descriptors, obtaining compounds with similar physicochemical properties despite the chemotype. The affinity index is determined by combining the Tanimoto/Tversky scores of the selected fields, resulting in various chemical structures.

The program was also calibrated using a set of ligands obtained from crystallographic complexes and literature from AstraZeneca, confirming the results' suitability [38].

1.2.2 exaScreen

exaScreen package depends on calculations based on a 3D molecular representation and non-chemotypical descriptors. However, it is already aimed at solving a different problem. Due to the expansion of libraries and the development of industry, a massive layer of information is being formed, which is quite challenging to investigate. It takes time and resources, so exaScreen was conceived as an innovative solution to an existing problem that also combines accuracy, speed, and the possibility of further synthesis of the results [39]. The program is suitable for the virtual screening of ultra-large libraries (currently, one is presented - Enamine REAL consisting of 31 Billion compounds) in a few hours.

1.3. Biological target: connexin 36

The connexin family are transmembrane proteins that, when oligomerised, form a transmembrane pore (half-channel) that ensures the exchange of molecules between the extracellular cavity and the cell's internal space. Moreover, two such half-channels, located between neighbouring cells, can dock to form an intercellular connection [40].

There are several types of connexins, each of which is also characterised by a different localisation: in the brain, neutrons and neuronal cells. For example, connexin 43 is present in astrocytes and plays a role in calcium-potassium metabolism, reaching certain brain areas through astrocytes. Consequently, if this circuit is disrupted, neurological problems, such as epilepsy, can be detected [41]. Connexin 36 is a more attractive target, as it is considered the main connexin expressed by neurons, forming interneuronal gap junctions or synapses. They are responsible for synaptic transmission, as well as for the activation of neurons in the brain [42]. However, the functionality is more comprehensive than this because protein is also involved in synchronous motor skills, learning, image processing and further learning, which makes connexin 36 a potential therapeutic target for disorders with synchrony, such as seizures, but the receptor relationship remains to be fully explored [43].

If the general structure of connexin 36 is considered, it consists of four transmembrane (TM) helices and two extracellular loops (ECL). The cytoplasmic region contains an N-terminal helix (NTH), a cytoplasmic loop (CL), and a C-terminal tail (CT) with varying diversity, demonstrating the differences of potential actions of this family depending on the regions. In addition, since this is a channel, it can have several states: closed (FN) and open (PLN) [44]. Due to phosphorylation, the state and permeability depend on voltage, pH, and the state of the isoforms.

The cryo-EM structure of connexin 36 is a dodecameric structure of subunits that correspond to the dimensions of 90 Å × 90 Å × 140 Å, as shown in Figure 1. In addition, while comparing it

with other members of the family (Cx26, Cx46), the structure and dodecameric interactions will be the same due to sequence similarity of approximately 52-54% [45].

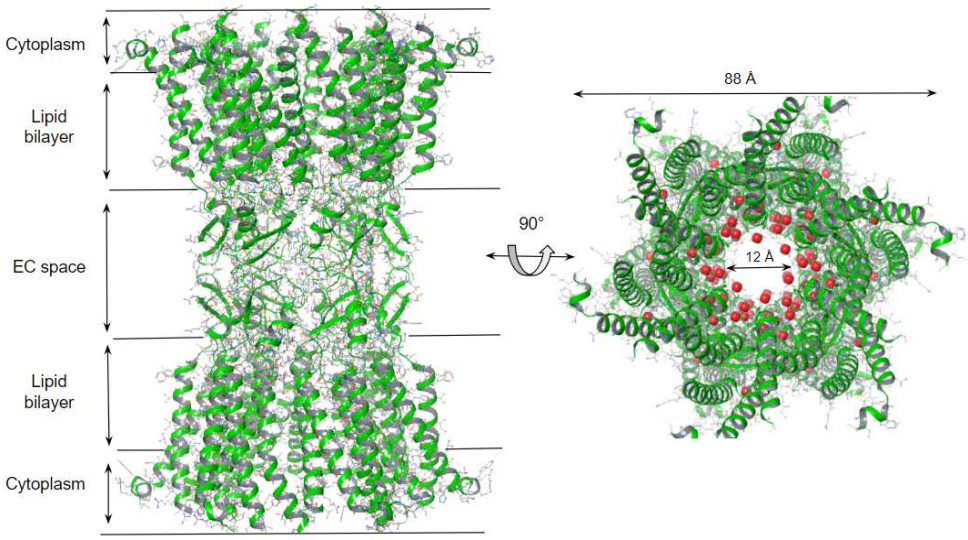

Figure 1. Cryo-EM structure of connexin 36 from two sides, where the red dots are water atoms.

Looking at more details, we can see that each protomer has extracellular loops (ECL) responsible for docking hemichannels. The linkage occurs through three disulfide bonds, with the residues present in ECL1 being constant across all members of the connexin family and those belonging to ECL2 being variable. They are shown in Figure 2 and are also called compatibility motifs due to their ability to dock not only with identical but also with other connexins [46].

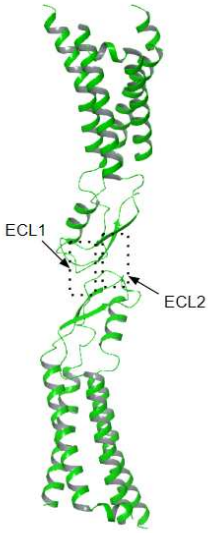

Figure 2. Two extracellular loops of connexin 36 with compatibility motifs indicated by dotted lines.

It has been proven that in these motifs, docking occurs through Lys238 and Glu239 and in ECL2, which is fundamentally different from other family members. This makes this connexin 36 unique due to the alternative intermolecular bridges formed [47]. They are considered to be responsible for the tight joining of the hemichannels and for the formation of a hydrophobic pocket. In addition, connexin 36 is located ~2 Å further from the channel entrance than other members, making the pore lumen much smaller [48].

Methods

This section describes the methods used in the following order: Pharmscreen, exaScreen and GOLD.

2.1. PharmScreen - libraries

Three main compound libraries were used for screening during the investigation:

- Molport is one of the libraries of chemical compounds that currently contains 4.6 million compounds, with isomers 6.5 million for screening and is constantly expanding, increasing the diversity of the spectrum of molecular structures and properties, which is not limited only to the pharmaceutical industry [49]. Searching for a specific function or similarity is also possible to find and filter compounds.
- The Enamine HTS Collection database is an extensive library of chemical compounds. Currently, the part of the screening used in this chapter accounts for 2.1 million compounds, 3.5 million with isomers [50].
- DrugBank is an online database that provides information on a selected compound's chemical, pharmaceutical, and pharmacological properties with bioinformatics resources. In this case, the idea was to screen among drugs already approved by the FDA, EMA, and other regulatory offices to find potential hits. Currently, this library includes approximately 4,408 approved drugs [51].

Pharmacelera already has Molport and Enamine libraries with calculated descriptors from its website. Approved Drugbank was downloaded from the official website. Mefloquine was used as a reference compound in all of the analyses.

2.2 PharmScreen - workflow

The PharmScreen program package served as a ligand-based virtual screening tool. The screening workflow with this program consisted of three main steps: library preparation, preparation of a reference ligand (already docked since the use of the bioactive conformation is important here), and virtual screening.

The library preparation step generated stereoisomers and conformations for its constituent molecules and accounted for flexibility in subsequent molecular descriptor calculations. As a result, the program forms a folder for each step, which contains the following files: a finished library file in .sdf format, a .log file that includes the selected parameters, and errors if they occur. Enamine and Molport libraries were downloaded from the company websites, where the desired parameters are already calculated. However, since the Drugbank library was obtained from the official website, it was required to prepare it using the following command:

$$pharmscreen\ -i\ Approved\_library.sdf\ -x\ lib\_preparation\ --\ genstereo$$

$-\ y\ gasteiger\ \ --\ logp\ at\ \ --\ single$

where *i* indicates the library used, --*genstereo* generates a variable amount of conformation for each molecule, -*y gasteiger* - calculation of partial charges, --*logp at* - calculation of a parameter at the atomic level, --single - unit library.

After retrieving a desired library in .sdf format, a reference compound (mefloquine) was prepared. This step was used to understand the molecule's flexibility and count the molecular descriptors that will be used for molecular similarity.

The command for ligand preparation was:

$pharmscreen\ \ --\ single\ \ -\ i\ Mfq\_reference.sdf\ \ --\ charges\ gasteiger\ \ --\ logp\ at\ \ --\ threads\ 16$

where --*single* means using one ligand as a reference; - *i* - points to the input file, mefloquine used in sdf format (Mfq_reference.sdf); --*charges gasteiger* - recommended method for calculating partial charges using the Gasteiger method, based on the type of fast atoms based on the Gasteiger Marsili parameterisation [52]; --*logp a*t - calculate the LogP parameter based on the type of atoms in a parameterisation from quantum mechanical calculations, without taking into account the conformation or types of atoms used; --*threads 16* - the number used for a software package with the threads option, depending on the load on the computer.

As a result, a folder with the experiment appears in the directory used. It will contain the prepared ligand itself in SDF format with calculations of partial charge and LogP parameters, as well as a Log file in which errors and a list of selected options can be displayed.

After retrieving a reference compound, the next step was a direct screening, identifying molecules similar to the reference structure and presenting a ranking of similarities. The software also assumes that each molecule in the library may have multiple conformations. For this step, the following command was used:

$pharmscreen\ \ -\ i\ (part\ of\ the\ library).sdf.gz\ \ -\ k\ mfq\_ref\_.sdf\ \ --\ charges\ userdefined\ \ --\ logp\ userdefined\ \ --\ threads\ 16$

This line contains two input files, where - i indicates the part of the library used, and -*k* indicates the reference structure. In this case, the screening was carried out relative to the charges, as well as the LogP indicator, which was calculated earlier; in addition, the --*threads* analysis was also carried out using 16 threads.

As a result, a folder is automatically generated with the following incoming files: a rating of molecules in .csv format based on the similarity index, the same rating, but in .sdf format, a ligand used as a standard in .sdf format, as well as a log file reporting on emerging errors and commands used.

2.3. exaScreen - workflow

exaScreen was used as a ligand-based screening tool to mine ultra-large chemical spaces. Enamine_REAL_202302, downloaded from the Pharmacelera website, was used as a library, and already-prepared mefloquine was used as a reference compound. Unlike PharmScreen, which performs 3D molecular field screenings on fully enumerated libraries, exaScreen performs a similarity search based on building block libraries. Here, the query molecule is partitioned into two query fragments. Since, in this case, there was a bicyclic structure, to avoid ring separation, a manual function was used at positions 10 - 14, presented in Figure 3. Numeration was obtained by uploading a reference compound in Pymol and adding labels.

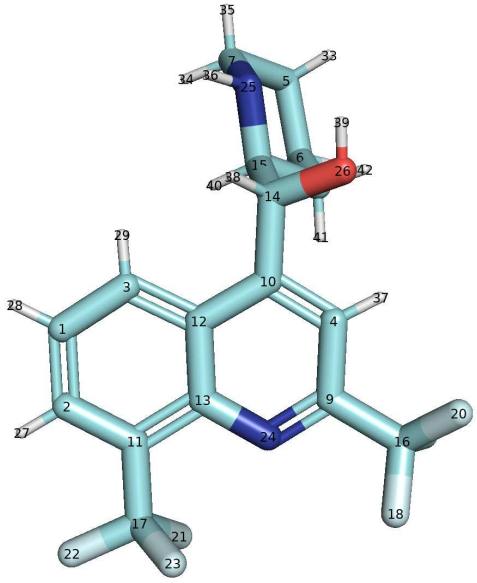

Figure 3. Visualisation of mefloquine with labels in Pymol.

The software will then perform a 3D field-based similarity search on a building block library, like Enamine REAL, exploring a potential chemical space of 31 billion of compounds. Selected building blocks are then used to enumerate de novo molecules based on reaction information, providing compounds that are synthesizable and purchasable to the commercial vendor.

The following command was used for screening:

$$exascreen \ -- \ input \ Enamine\_REAL\_202310.pha.gz \ - \ k \ Mfq\_reference.sdf \ -- \ manualpartition$$

exaScreen finds similar building blocks (more precisely, synthons) to the query molecules and then uses them to enumerate de novo molecules based on reaction information. The folder with the result contained de_novo_molecules. smi is a file with smiles containing all the listed structures with the molecules' names. Then, the same file, but in .sdf format, with a 3D structure of connections, as well as a .log file containing information about the commands used and errors.

Each hit found is evaluated by two indicators: PHA_EXA_SCORE, an assessment of the sum of the positions of building blocks relative to the rating of fragments. It can only be positive, and since not all building parts are readily amenable to combinatorics, the values can be different, but the smaller the better. PHA_FIELD_SCORE - the average value of similarity between the fields of two compounds, ranging from 0 to 1 (the higher, the better).

2.4 GOLD - workflow

Genetic Optimisation of Ligand Docking (GOLD) was used as a structure-based virtual screening tool. The genetic algorithm provides the basis for targeting a ligand into a binding pocket in a manner reminiscent of "natural selection," with populations of ligands depending on the potential and then evaluated using functions [53].

As a first step - the ligands were prepared using the LigPrep function from Schrodinger 20221-1. Molport library, obtained from the official website was prepared using this command line:

$$zcat\ INPUT.txt.gz \mid awk\ '\{print\ \$2,\ \$3\}'\ >\ INPUT.smi \mid /\ldots/ligprep\ -\ HOST{:}\ 16$$
$$-\ ismi\ \$file.smi\ -\ WAIT\ -\ ma\ 90\ -\ bff\ 16\ -\ pht\ 0.0\ epik\ -\ s\ 4\ -\ nt\ -\ osd\ \$file.sdf$$

, where *zcat* is used to decompress the content of the input file: | *awk '{print $2, $3}' > INPUT.smi* - extracts files and saves them in a *.smi; /../ligprep* - pathway of ligprep; *-HOST:16* - used computer and number of threads; *-WAIT* - completing job before existing; *-ma90* - maximum of number of conformers; *bff16* - number of torsional samples per rotatable bond; *pht 0.0.* - PH for protonation state; *epik* generating ionisation and tautomeric states; *-s4* - states for penalty for protonation state generation; - *nt* - generated structures are neutral; *-osd* - to have an .sdf format output.

The protein was also prepared using the Schrodinger Epik Protein Preparation Wizard 5.516149 in the Maestro program. Basic parameters were set to default (addition of hydrogen), except for removing water, which could interfere with identifying binding sites. The structure was also minimised and then moved to Hermes 1.10.551 software to retrieve a test configuration file, which contains all the parameters that determine the flow of the experiment, for example, an evaluation system, saving results and binding site. This file was used as a template for a massive screening process,

GoldScore was used as a rating function. It is empirical, considering hydrogen bonds, van der Waals interactions, and the internal energy of the ligand itself [54]. The use of each of these functions depends on the study as well as the complexity of the binding pocket. In this case, GoldScore was used because of its versatility and its focus not on optimisation but on calculating potential hits associated with the target.

Results

This section describes the results obtained in the following order: Ligand selection, Reference compound modification, Pharmscreen, exaScreen and GOLD.

3.1.1 Reference ligand selection

The reference compound was chosen from quinidine, quinine, and mefloquine, as recommended through a collaboration with a scientist, Dr. Volodimir Korhov, from PSI.

These three ligands presented in Figure 4 are antimalarial agents. However, quinidine is also known for its use in arrhythmias [55].

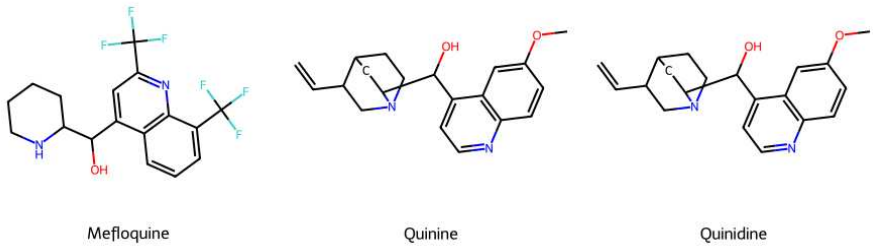

Figure 4. 2D-representation of structures of mefloquine, quinine, quinidine

If considered from a chemical point of view, then:

- Mefloquine is a synthetic derivative of quinoline because of a quinoline core with a trifluoromethyl group attached to position 8. It also has a piperidine ring attached to it at position 4 due to an oxygen bridge [56].
- Quinine is a natural alkaloid with a quinoline ring system. It is fused with a bicyclic quinuclidine ring, with a methoxy group in the sixth position and a vinyl group in the eighth position [57].
- Quinidine is an optical isomer of quinine; it has a similar structure (quinoline and quinuclidine rings) but in a different spatial arrangement, which causes the pharmacological activity to change [58].

Mefloquine has a trifluoromethyl structure attached to the quinoline ring, which is why an electron acceptor effect occurs, reducing the likelihood of metabolic degradation of the substance. Without this group, quinine and quinidine are less stable, and their groups can be metabolised, making mefloquine superior to other compounds.

The ligands are interesting due to the inhibitory effect of connexin 36 with high specificity. There are suggestions that the mechanism of this action depends on the 80S ribosome of Plasmodium falciparum, which plays the role of the main target in changing the dynamics of the

open-closed state of the channel, as well as on voltage. However, the full sequence still requires research [59]. In Figure 5, a significant reduction of the channel opening is observed compared to the unbound channel.

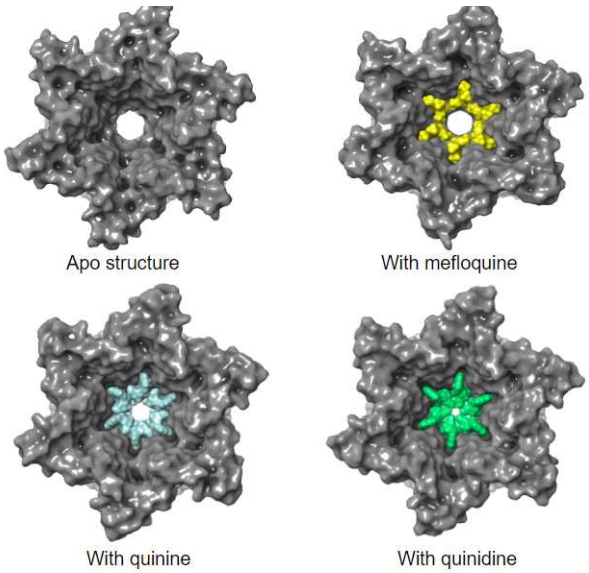

Figure 5. Structures of connexin 36, connexin 36 with ligands.

As a result, mefloquine was chosen as the reference compound due to its improved chemical stability and more accurate binding pocket.

3.1.2. Mefloquine in apo-structure

Mefloquine was examined in more detail. The structure of connexin 36 was prepared by optimising, removing excess water, and minimising. As a result, a single bond was observed to residue Glu43 and Tyr43 due to the protonated amine on the piperidine ring shown in Figure 6.

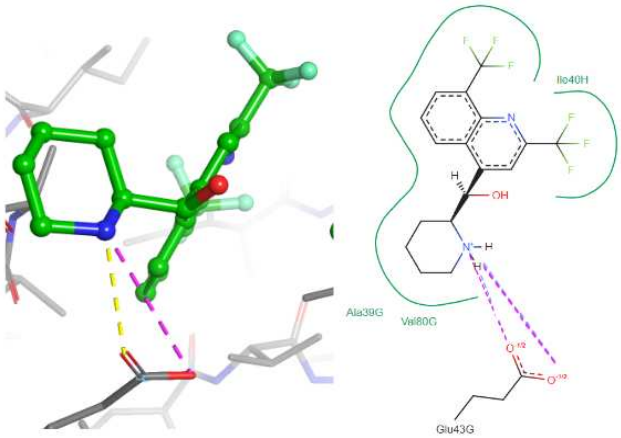

Figure 6. Mefloquine binding sites.

In addition, the green edging of the ligand that mefloquine forms on almost all hydrophobic contact surfaces makes this noticeable.

3.2. Reference ligand modification

After considering the ligand and the unique complex of six structures, the idea arose to add a link between them by adding a new group. As a result, a trial insertion of a keto group into the fifth position of the piperidine ring was made, shown in Figure 7.

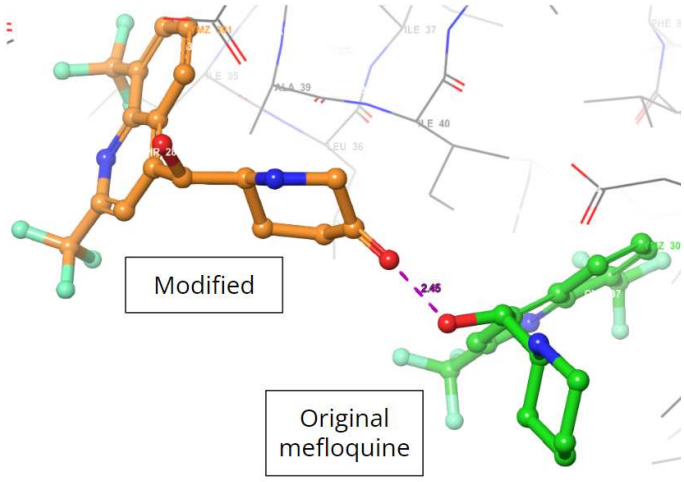

Figure 7. Measurement of the distance from the modified mefloquine with the introduced keto group (orange) to the hydroxy group of the original mefloquine (green).

The measurement revealed a donor-acceptor distance of 2.45 Å, corresponding to a short hydrogen bond. As a result, several modifications were suggested:

- Keto group was included since its high electronegativity as a strong acceptor of hydrogen bonds;
- Hydroxy group was included because of the donor-acceptor abilities;
- Fluorine group, similar to keto group, has strong electronegativity, making it an excellent acceptor.

However, as it is clear from Figure 8, such changes do not contribute to creating a bond with the subsequent ligand but rather, on the contrary, a decrease in hydrophobic contact.

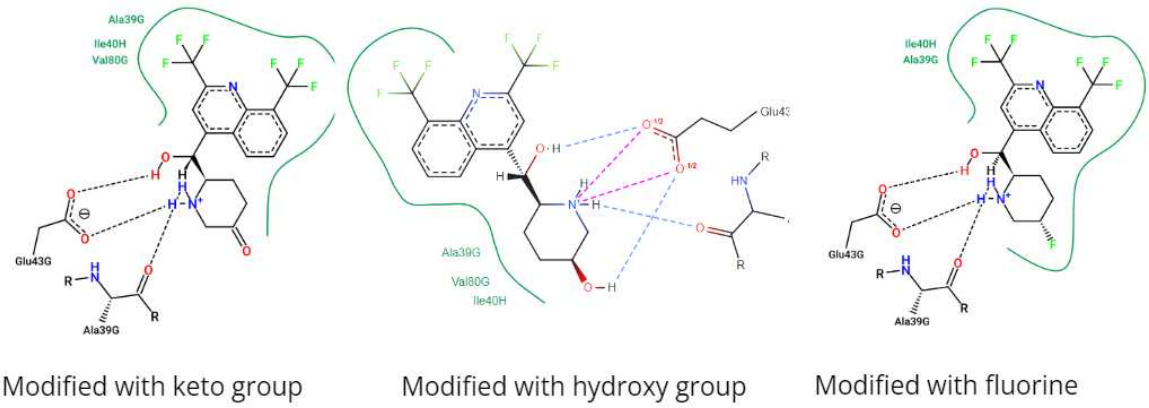

Figure 8. Interactions in a modified mefloquine version.

In addition, incorporating these groups into the piperidine ring is quite laborious from the point of view of chemical synthesis.

3.3. PharmScreen and exaScreen

This section describes the results obtained in the following order: from the Molport library, from the Enamine library, exaScreen and the final selection.

3.3.1. Pharmscreen: Molport library

From the screening of the Molport library, a list of 18,000 compounds was obtained ranked by pha_similarity_score (quantifying the similarity between a reference compound not in a structural form but due to the count of descriptors that affect the binding sites of the target) and the number of compounds. As a result, Figure 9 indicates where a characteristic break becomes noticeable according to the number of compounds, indicating an averaging of the similarity parameter.

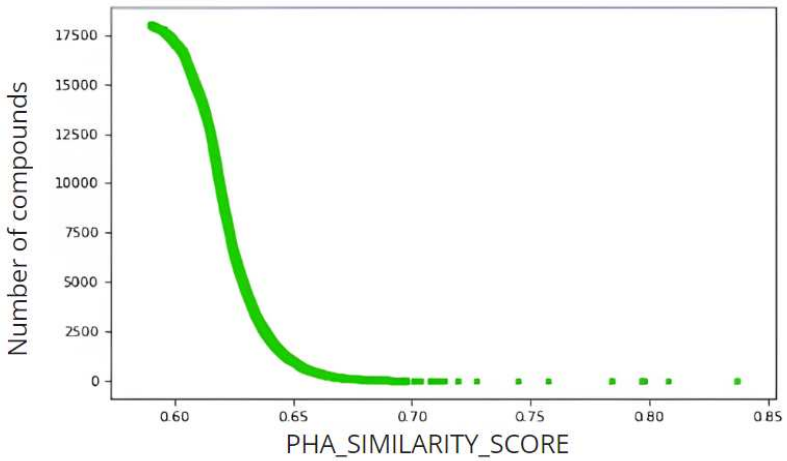

Figure 9. Graphical representation of the distribution of the similarity parameter in compounds from Molport library.

Compounds from different sections of the list were compared, and those with a similarity index of less than 0.641 had a significant difference in alignment from mefloquine. Sometimes, functional groups necessary for proper interaction with the target were absent. After this initial filtering, 2047 compounds remained and were subjected to a more detailed manual review. This review consisted of a step-by-step examination of each compound to determine its alignment with the original compound, structural adequacy, and the presence of an active reaction group.

Pharmscreen was able to provide a variety of scaffolds; however, some trends in the construction of molecules, which are demonstrated in Figure 9, were still possible to observe.

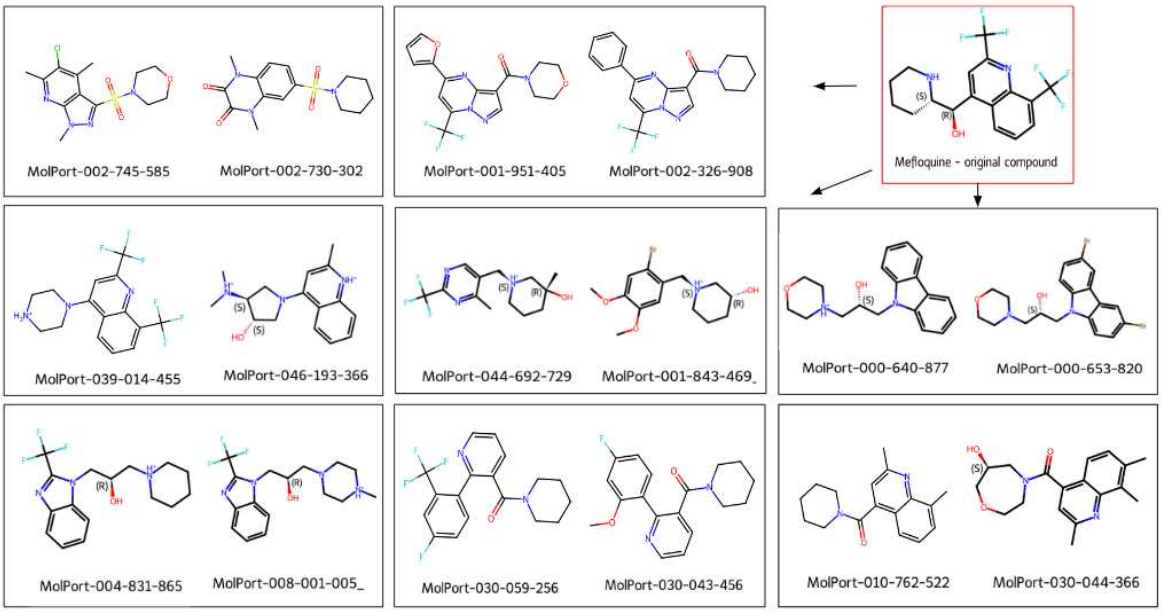

Figure 10. Examples of compounds obtained during the study of the Molport Library.

For example, the piperidine ring sometimes had its position changed to 3, or the addition chain was simply increased or shortened. Other groups, such as pyridine or pyrrole, could be used instead of this ring.

After considering a variety of structures, it was concluded that the ideal hit should be minor due to the geometry of the binding channel and having a protonated amine as the primary reactive group. After this filtering, 327 compounds remained, which were then subjected to docking with glide in three poses, where each was considered a separate compound.

3.3.2. Pharmscreen: Enamine library

After screening the Enamine library, 22,000 were obtained and reviewed. The resulting list was arranged in descending order (from highest to lowest) relative to the similarity indicator and number of compounds presented in Figure 11. It becomes noticeable that the hits start with a lower indicator of 0.72 this time.

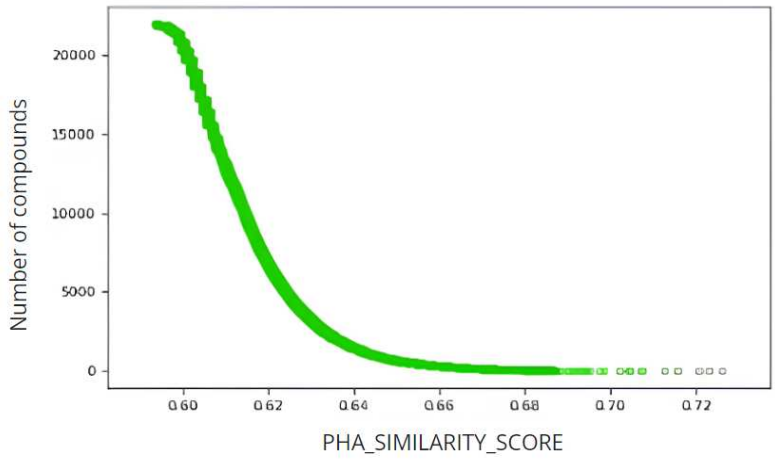

Figure 11. Graphical representation of the distribution of the similarity parameter in compounds from the Enamine library.

As a result, only 203 compounds were selected with the parameter accounting for more than 0.665, which were subjected to a more detailed manual examination, presented in Figure 12.

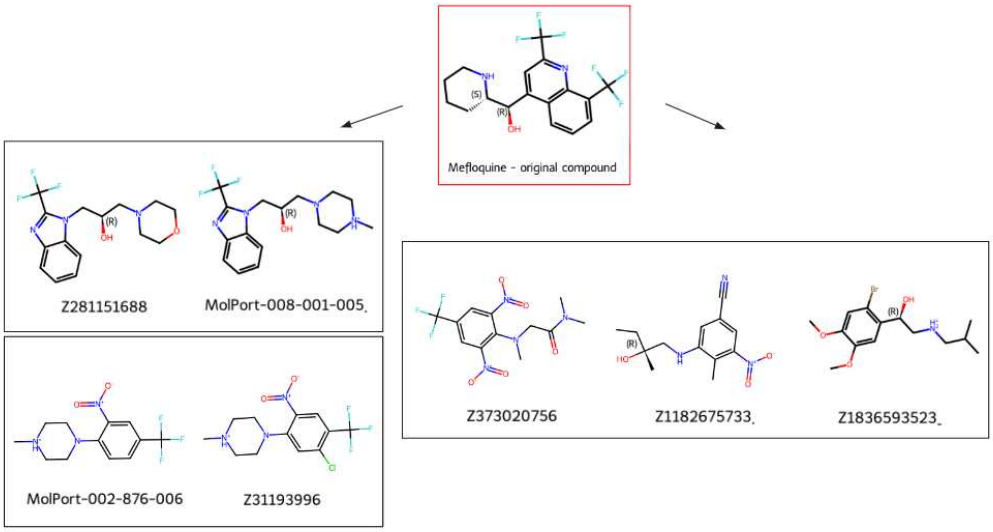

Figure 12. Examples of compounds obtained during the study of the enamine library.

For example, some of the hits were structurally similar to representatives from the Molport library, except for adding one functional group, and there was also a tendency to reduce the "skeleton" structure and replace the bicyclic ring with a single ring. After this selection, 144 compounds remained, and docking with glide was carried out in three poses, forming a final list of 432 hits.

3.3.3. Pharmscreen: DrugBank library

The prefiltered result consisted of 1000 molecules, which were then ranked according to the similarity score and number of compounds presented in Figure 13.

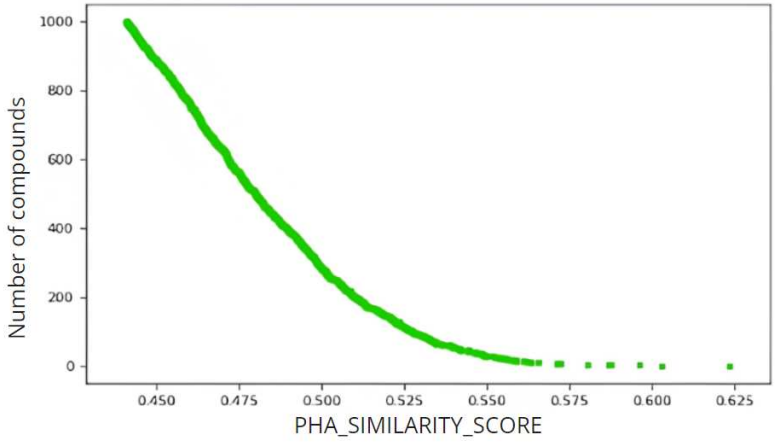

Figure 13. Graphical representation of the distribution of the similarity parameter in compounds from the DrugBank library.

The maximum value was 0.625, which is relatively low compared to the previous two libraries. Due to the small number of hits found, the entire library was subjected to manual examination of structures, examples of which are presented in Figure 14.

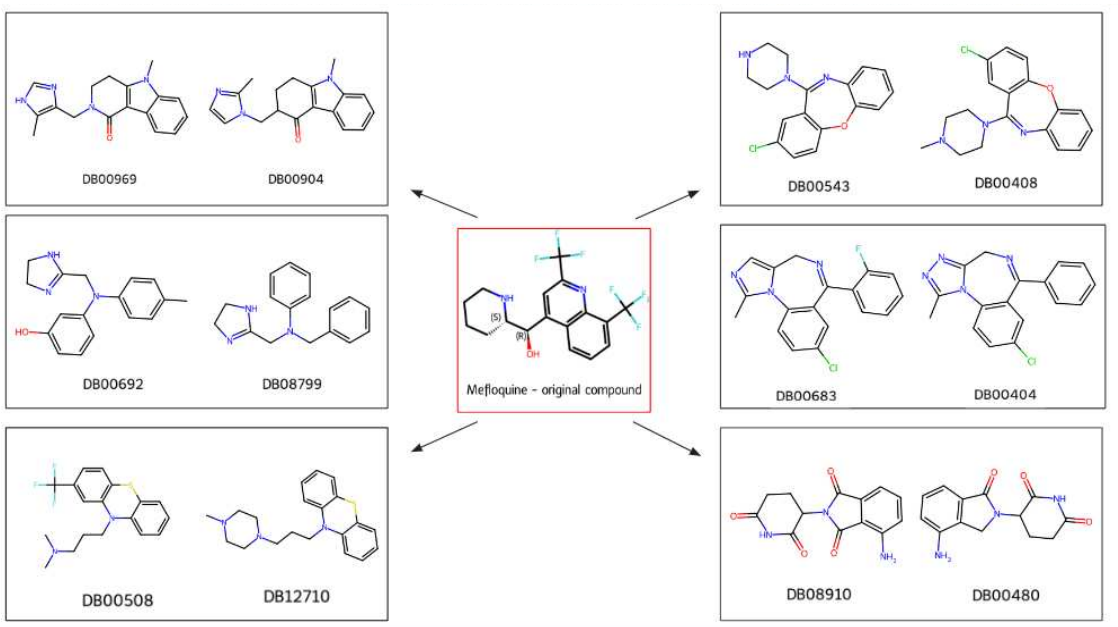

Figure 14. Examples of compounds obtained during the study of the DrugBank library.

From these examples, one can notice repeating basic structures among hits, sometimes differing only in position or the presence/absence of a functional group. As a result, after selection, 48 structures remained, in which docking with glide was used in three poses, presenting 144 options for consideration.

3.3.4. exaScreen: Enamine REAL library

After screening the Enamine REAL library consisting of 31 billion of compounds, 99,954 compounds were obtained and ranked against the two metrics. Each compound had a directly proportional score; that is, those with a lower PHA_EXA_SCORE also had a high PHA_FIELD_SCORE. Figure 15 shows the distribution of one of the functions relative to the number of compounds.

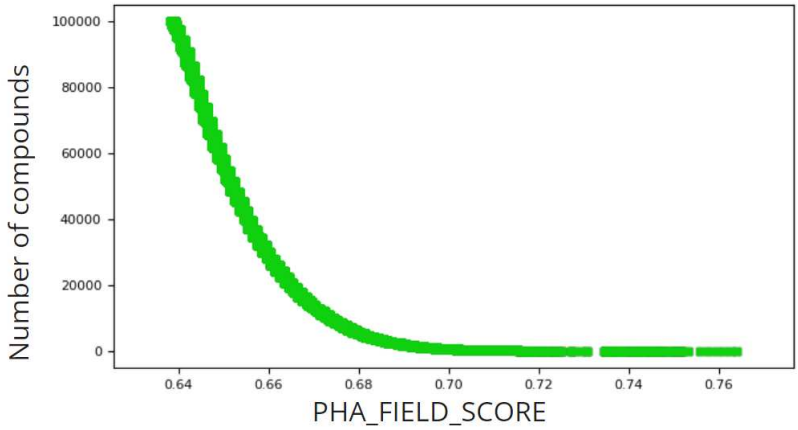

Figure 15. Graphical representation of the distribution of the similarity parameter in compounds from the Enamine REAL library.

Consequently, the first 4000 compounds were selected and reviewed. Figure 16 presents examples of structures obtained during the analysis.

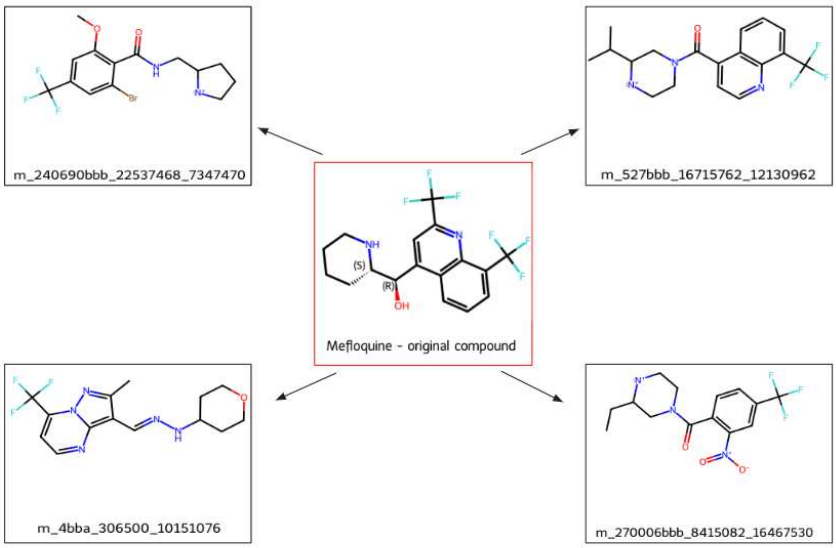

Figure 16. Examples of compounds obtained by using exaScreen.

Similar molecular scaffolds were observed in previous experiments, and the selected 236 compounds were subjected to the glide docking method.

3.3.5. Pharmacelera - selected results

After docking using the Glide method, 40 structures were selected from the Molport library. A final selection was carried out by comparing the alignment of the reference compound, resulting in 14 compounds, presented in Figure 17.

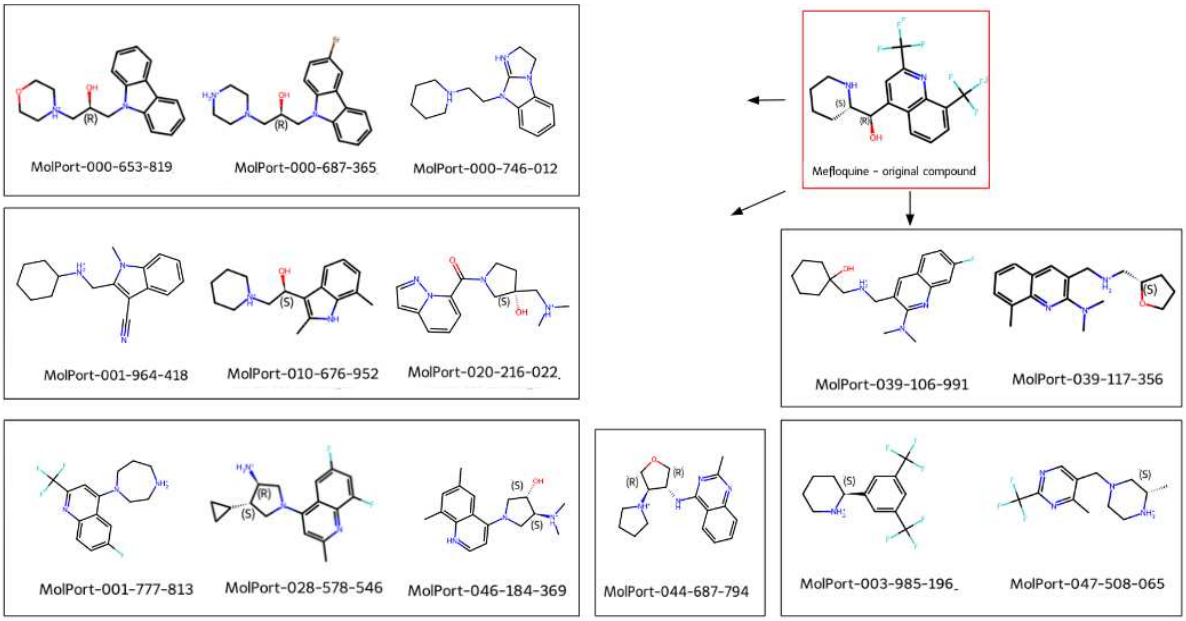

Figure 17. Selected hits from the Molport library.

The results from the Enamine library included five compounds presented in Figure 18.

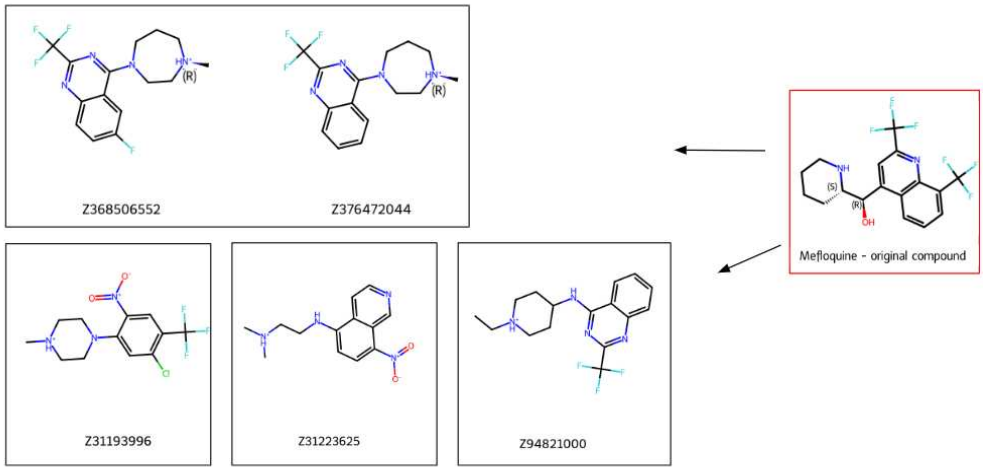

Figure 18. Selected hits from the Enamine library.

Due to the similarity of the obtained compounds, the exaScreen results were not included in the final selection indicating that it is close to a fully enumerated method, such as PharmScreen, but dedicating less resources and time. After inspection, nearly all of the hits can be divided into several groups depending on their structure base and the location of functional groups. However, they all possess the protonated amine required to bind to the pocket.

An example of a selected compound, MolPort-047-508-065 is shown in Figure 19.

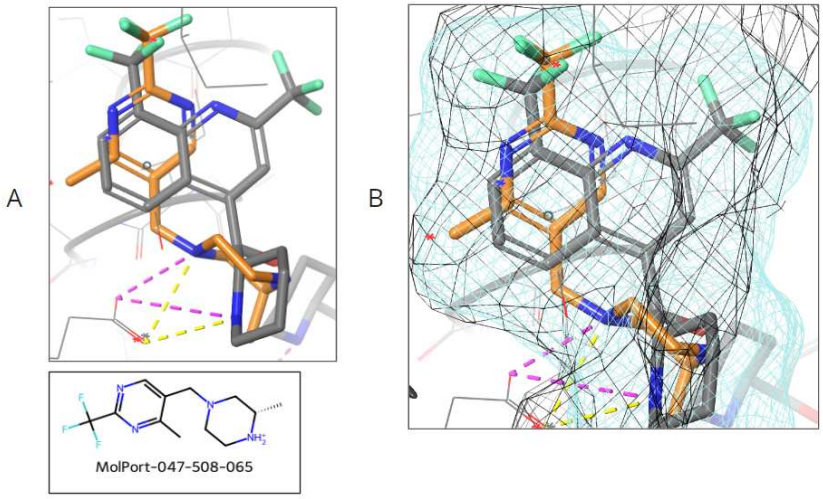

Figure 19. Example of the selected compound obtained from Molport aligned to the reference: A - alignment of reference (grey) and selected compound Molport-047-508-065 (orange), B - fitting in the grid of reference compound (light blue) and grid of the connexin 36 (black) of reference (grey) and selected compound Molport-047-508-065 (orange).

This compound aligns well with mefloquine while adapting to the structure of the binding pocket. Additionally, it has a protonated amine that interacts with the amino acids Gly46 and Tyr46, similar to the reference compound. Moreover, selected compounds from the Molport and Enamine libraries had Glide scoring functions ranging from -5.307 to -6.299.

3.5.1. GOLD: Molport library

After screening the Molport library with GOLD, 202,467 hits were obtained, which were then ranked in descending order using GoldScore. Due to the large number, only the first 10,000 were used for further investigation.

Unfortunately, some regularities emerged in the results: 1. About 98% of the compounds were quite bulky, without considering the channel geometry. 2. They contained long chains of carbon atoms, making the molecule too flexible. Figure 20 shows examples of the compounds obtained.

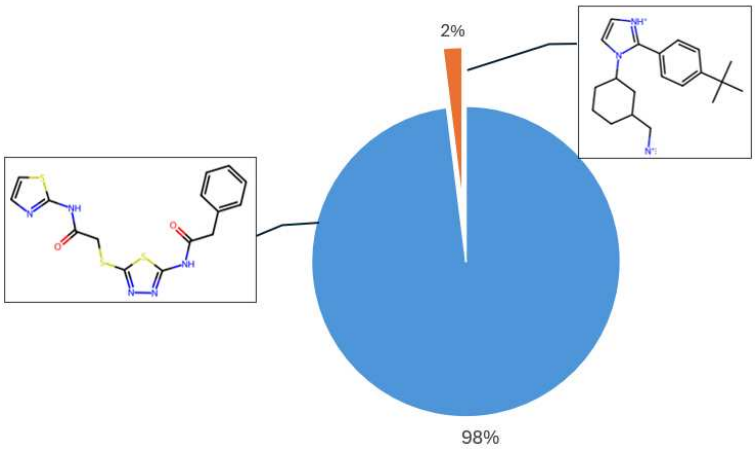

Figure 20. Composition of compounds obtained from docking with GOLD.

Such a situation happened due to the very mechanism of GOLD's action: when a complementary part is found, it tries to replicate it several times to increase the scoring functions. Critically, few compounds meet the criteria, which makes selection possible.

3.5.2. GOLD - selected results

Of the previously selected top 10,000 compounds, only 554 were found to be the right size to fit in the binding pocket, and they were carefully reviewed. As a result, potential structures were selected, presented in Figure 21.

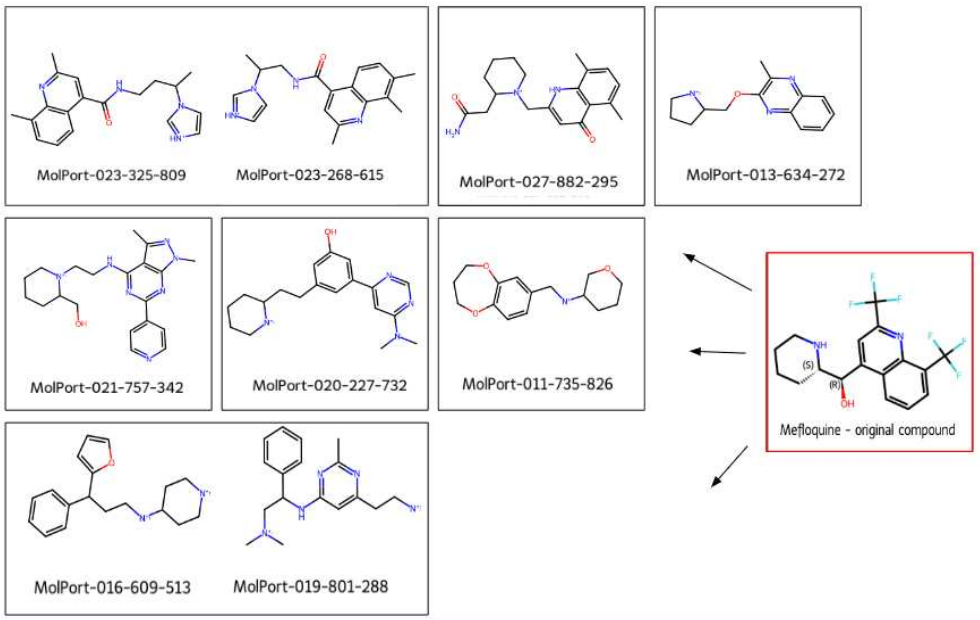

Figure 21. Molport compounds identified with GOLD as potential connexin 36 modulators.

The resulting compounds had extended chains larger than those obtained by the Pharmscreen method. The resulting structures had a GOLD.Goldscore.Fitness between 43.2 and 57.7, which is undoubtedly a lower result relative to more voluminous structures, but here, we also considered the geometrical features of the channel. However, this did not prevent aligning with the original compounds and, at the same time, folding so as not to violate the geometry of the channel, such as MolPort-021-757-342, presented in Figure 22. At the same time, there is another example, in the same Figure 22 - MolPort-010-856-704, which, even though the "base" is aligned with mefloquine, but due to the size of the connecting chain, occupies the entire space, making it impossible for six ligands to exist at the channel simultaneously.

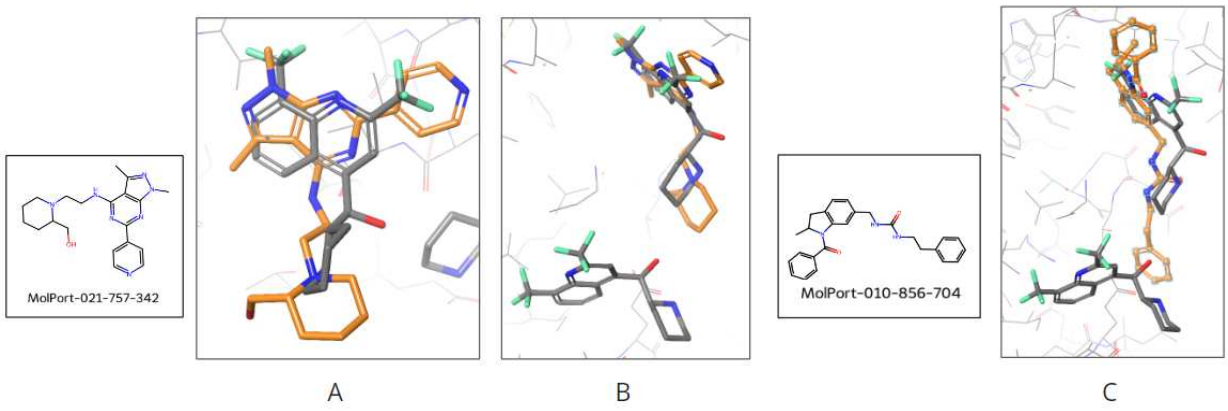

Figure 22. Examples of the compounds obtained from GOLD aligned to the reference compound: A - alignment of the reference compound (grey) with the selected MolPort-021-757-342 (orange), B - geometrical position in the channel of reference compounds(grey) and MolPort-021-757-342 (orange), C - geometrical position in the channel of reference compounds(grey) and MolPort-10-856-704 (orange).

3.6. Comparison of the selected compounds to the DrugBank

The results were divided according to the library used and the method obtained to compare them with already approved drugs. The achieved scatter plot is presented in Figure 23.

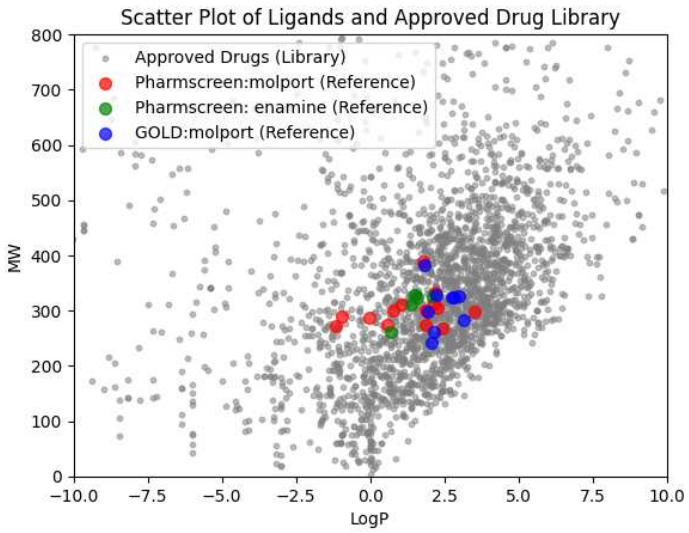

Figure 23. Scatter plot representing the comparison of the selected compounds from Pharmscreen - Molport (red), Pharmscreen - Enamine (green) and GOLD - Molport (blue) to already approved drugs from a DrugBank library according to the molecular weight and LogP parameter.

The correlation between molecular weight and the octanol-water partition coefficient was established, indicating the molecule's lipophilicity. As a result, all hits were in the molecular weight range from 200 to 300 Da, and on the x-axis, they tended to reach 0, except for two compounds obtained by the Pharmscreen method from the Molport library. Such comparison allows us to conclude that the properties of the chosen structures are similar to those already on the pharmaceutical market.

4. Conclusions

This study used a combination of structural and ligand screening methods to develop a new inhibitor of connexin 36. The antimalarial drug mefloquine was chosen over other quinoline derivatives due to its stability as a reference compound. The structure analysis revealed an unusual geometry because of the sixfold arrangement of ligands in the channel. Therefore, the potential ligand must be small enough for six copies to fit the diameter of the channel, have an adequate level of conformational rigidity, and have a leading interaction group - a protonated amine.

Modifying the reference compound was a primary method of investigating inhibitors, with the addition of a new functional group, fluorine, ketone, and hydroxy, to improve affinity. However, this idea did not affect the affinity of the compound, resulting in a decrease in lipophilicity. In addition, including such a group is challenging from the point of view of organic synthesis.

Structural-based methods and ligand-based methods are the basis for computational developments. However, despite their capabilities, they also have limitations that affect the sensitivity of the analysis. By evaluating each method separately, the study sought to combine them, producing a comprehensive list of the best potential inhibitors.

The results obtained using PharmScreen and exaScreen were characterised by diverse structures not structurally similar to the reference. After skimming through the first result, a specific pattern is recognisable: introductions, replacements of functional groups, or changes in positions. Both programs produced similar results, concluding that the selected structures also take advantage of exaScreen and can be synthesised.

GOLD was used as a structure-based method, and ligand was used as a reference for binding site determination. Unfortunately, the results were mainly represented by large, branched compounds that simply would not fit into the connexin 36 channel. Allowing us to conclude that this method did not take into account the geometric feature of the target and, having found a functional group that increases the binding index, began to replicate it repeatedly, which led to the situation where finding a hit of a suitable size became a challenging task.

The selected results were compiled into a list, which will be sent to biological trials to confirm their effectiveness as connexin 36 inhibitors.

References

1. Alves de Oliveira T, Silva M, Habib E, Silva A, Taranto A. Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. *Drugs Drug Candidates*. 2023;2:311-334. doi:10.3390/ddc2020017

2. Arya H, Coumar MS. Chapter 4 - Lead identification and optimization. In: Bhatt TK, Nimesh S, eds. *The Design & Development of Novel Drugs and Vaccines*. Academic Press; 2021:31-63. doi:10.1016/B978-0-12-821471-8.00004-0

3. Slater O, Kontoyianni M. The compromise of virtual screening and its impact on drug discovery. *Expert Opin Drug Discov*. Published online July 3, 2019. Accessed June 13, 2024. https://www.tandfonline.com/doi/abs/10.1080/17460441.2019.1604677

4. Entzeroth M, Flotow H, Condron P. Overview of high-throughput screening. *Curr Protoc Pharmacol*. 2009;Chapter 9:Unit 9.4. doi:10.1002/0471141755.ph0904s44

5. Kontoyianni M. Docking and Virtual Screening in Drug Discovery. In: Lazar IM, Kontoyianni M, Lazar AC, eds. *Proteomics for Drug Discovery: Methods and Protocols*. Springer; 2017:255-266. doi:10.1007/978-1-4939-7201-2_18

6. How were new medicines discovered? | Nature Reviews Drug Discovery. Accessed June 13, 2024. https://www.nature.com/articles/nrd3480

7. Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem*. 2013;20(23):2839-2860. doi:10.2174/09298673113209990001

8. Sperandio O, Miteva M, Villoutreix B. Combining Ligand- and Structure-Based Methods in Drug Design Projects. *Curr Comput - Aided Drug Des*. 2008;4:250-258. doi:10.2174/157340908785747447

9. Li J, Fu A, Zhang L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. *Interdiscip Sci Comput Life Sci*. 2019;11(2):320-328. doi:10.1007/s12539-019-00327-w

10. Concepts and Applications of Molecular Similarity | Wiley. Accessed June 13, 2024. https://www.wiley.com/en-us/Concepts+and+Applications+of+Molecular+Similarity-p-9780471621751

11. Duan J, Dixon SL, Lowrie JF, Sherman W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J Mol Graph Model*. 2010;29(2):157-170. doi:10.1016/j.jmgm.2010.05.008

12. FLAP: GRID Molecular Interaction Fields in Virtual Screening. Validation using the DUD Data Set | Journal of Chemical Information and Modeling. Accessed June 13, 2024. https://pubs.acs.org/doi/abs/10.1021/ci100221g

13. Abrahamian E, Fox PC, Naerum L, Christensen IT, Thøgersen H, Clark RD. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J Chem Inf Comput Sci*. 2003;43(2):458-468. doi:10.1021/ci025595r

14. Chemical Similarity Searching | Journal of Chemical Information and Modeling. Accessed June 13, 2024. https://pubs.acs.org/doi/10.1021/ci9800211

15. Muegge I, Mukherjee P. An overview of molecular fingerprint similarity search in virtual screening. *Expert Opin Drug Discov*. 2016;11(2):137-148. doi:10.1517/17460441.2016.1117070

16. Vázquez J, García R, Llinares P, Luque FJ, Herrero E. On the relevance of query definition in the performance of 3D ligand-based virtual screening. *J Comput Aided Mol Des*. 2024;38(1):18. doi:10.1007/s10822-024-00561-5

17. Drwal MN, Griffith R. Combination of ligand- and structure-based methods in virtual screening. *Drug Discov Today Technol*. 2013;10(3):e395-401. doi:10.1016/j.ddtec.2013.02.002

18. Banoglu E, Çalişkan B, Luderer S, et al. Identification of novel benzimidazole derivatives as inhibitors of leukotriene biosynthesis by virtual screening targeting 5-lipoxygenase-activating protein (FLAP). *Bioorg Med Chem*. 2012;20(12):3728-3741. doi:10.1016/j.bmc.2012.04.048

19. Structure- and Ligand-Based Virtual Screening on DUD-E+: Performance Dependence on Approximations to the Binding Pocket | Journal of Chemical Information and Modeling. Accessed June 13, 2024. https://pubs.acs.org/doi/10.1021/acs.jcim.0c00115

20. Protein-Ligand-Based Pharmacophores: Generation and Utility Assessment in Computational Ligand Profiling | Request PDF. Accessed June 13, 2024. https://www.researchgate.net/publication/223961942_Protein-Ligand-Based_Pharmacophores_Generation_and_Utility_Assessment_in_Computational_Ligand_Profiling

21. Anighoro A, Bajorath J. A Hybrid Virtual Screening Protocol Based on Binding Mode Similarity. *Methods Mol Biol Clifton NJ*. 2018;1824:165-175. doi:10.1007/978-1-4939-8630-9_9

22. Tresadern G, Bemporad D. Modeling approaches for ligand-based 3D similarity. *Future Med Chem*. 2010;2(10):1547-1561. doi:10.4155/fmc.10.244

23. Hawkins PCD, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem*. 2007;50(1):74-82. doi:10.1021/jm0603365

24. Cheeseright TJ, Mackey MD, Melville JL, Vinter JG. FieldScreen: virtual screening using molecular fields. Application to the DUD data set. *J Chem Inf Model*. 2008;48(11):2108-2117. doi:10.1021/ci800110p

25. Spyrakis F, Ahmed MH, Bayden AS, Cozzini P, Mozzarelli A, Kellogg GE. The Roles of Water in the Protein Matrix: A Largely Untapped Resource for Drug Discovery. *J Med Chem*. 2017;60(16):6781-6827. doi:10.1021/acs.jmedchem.7b00057

26. Schmidtke P, Luque FJ, Murray JB, Barril X. Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. *J Am Chem Soc*. 2011;133(46):18903-18910. doi:10.1021/ja207494u

27. Gaillard P, Carrupt PA, Testa B, Boudon A. Molecular lipophilicity potential, a tool in 3D QSAR: method and applications. *J Comput Aided Mol Des*. 1994;8(2):83-96. doi:10.1007/BF00119860

28. Carrupt P -A, Gaillard P, Billois F, et al. The Molecular Lipophilicity Potential (MLP): A New Tool for log P Calculations and Docking, and in Comparative Molecular Field Analysis (CoMFA). In: ; 2008:195-217. doi:10.1002/9783527614998.ch12

29. Solvation in octanol: Parametrization of the continuum MST model | Request PDF. Accessed June 13, 2024. https://www.researchgate.net/publication/220418164_Solvation_in_octanol_Parametrization_of_the_continuum_MST_model

30. (PDF) Hybrid QSPR models for the prediction of the free energy of solvation of organic solute/solvent pairs. Accessed June 13, 2024. https://www.researchgate.net/publication/333823552_Hybrid_QSPR_models_for_the_prediction_of_the_free_energy_of_solvation_of_organic_solutesolvent_pairs

31. Luque FJ, Barril X, Orozco M. Fractional description of free energies of solvation. *J Comput Aided Mol Des*. 1999;13(2):139-152. doi:10.1023/A:1008036526741

32. Conformational analysis and structural comparisons of (1R,3S)-(+)- and (1S,3R)-(-)-tefludazine, (S)-(+)- and (R)-(-)-octoclothepin, and (+)-dexclamol in relation to dopamine receptor antagonism and amine-uptake inhibition | Journal of Medicinal Chemistry. Accessed June 13, 2024. https://pubs.acs.org/doi/10.1021/jm00397a006

33. Robson B. Intermolecular Interactions: from diatomics to biopolymers Edited by B. Pullman Wiley-Interscience; New York, 1978 ix + 447 pages. £19.50. *Febs Lett - FEBS LETT*. 1978;91:383-383.

34. Curutchet C. Dispersion and repulsion contributions to the solvation energy: Refinements to a simple computational model in the continuum approximation. *J Comput Chem*. Published online January 1, 1991. Accessed June 13, 2024. https://www.academia.edu/32667556/Dispersion_and_repulsion_contributions_to_the_solvation_energy_Refinements_to_a_simple_computational_model_in_the_continuum_approximation

35. Vázquez J, Deplano A, Herrero A, et al. Development and Validation of Molecular Overlays Derived from Three-Dimensional Hydrophobic Similarity with PharmScreen. *J Chem Inf Model*. 2018;58(8):1596-1609. doi:10.1021/acs.jcim.8b00216

36. Ginex T, Muñoz-Muriedas J, Herrero E, Gibert E, Cozzini P, Luque FJ. Application of the quantum mechanical IEF/PCM-MST hydrophobic descriptors to selectivity in ligand binding. *J Mol Model*. 2016;22(6):136. doi:10.1007/s00894-016-2991-3

37. Ginex T, Muñoz-Muriedas J, Herrero E, Gibert E, Cozzini P, Luque FJ. Development and validation of hydrophobic molecular fields derived from the quantum mechanical IEF/PCM-MST solvation models in 3D-QSAR. *J Comput Chem*. 2016;37(13):1147-1162. doi:10.1002/jcc.24305

38. Giangreco I, Cosgrove DA, Packer MJ. An extensive and diverse set of molecular overlays for the validation of pharmacophore programs. *J Chem Inf Model*. 2013;53(4):852-866. doi:10.1021/ci400020a

39. exaScreen - Screening ultra-large synthesizable libraries. Pharmacelera | Pushing the limits of computational chemistry. Accessed June 13, 2024. https://pharmacelera.com/exascreen/

40. Bazzigaluppi P, Isenia SC, Haasdijk ED, et al. Modulation of Murine Olivary Connexin 36 Gap Junctions by PKA and CaMKII. *Front Cell Neurosci*. 2017;11. doi:10.3389/fncel.2017.00397

41. Wallraff A, Köhling R, Heinemann U, Theis M, Willecke K, Steinhäuser C. The Impact of Astrocytic Gap Junctional Coupling on Potassium Buffering in the Hippocampus. *J Neurosci*. 2006;26(20):5438-5447. doi:10.1523/JNEUROSCI.0037-06.2006

42. Analysis of connexin expression during seizures induced by 4-aminopyridine in the rat hippocampus | Journal of Biomedical Science | Full Text. Accessed June 13, 2024. https://jbiomedsci.biomedcentral.com/articles/10.1186/s12929-015-0176-5

43. Andrioli A, Fabene PF, Mudò G, et al. Downregulation of the Astroglial Connexin Expression and Neurodegeneration after Pilocarpine-Induced Status Epilepticus. *Int J Mol Sci*. 2022;24(1):23. doi:10.3390/ijms24010023

44. Oshima A, Tani K, Hiroaki Y, Fujiyoshi Y, Sosinsky GE. Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule. *Proc Natl Acad Sci U S A*. 2007;104(24):10034-10039. doi:10.1073/pnas.0703704104

45. Maeda S, Nakagawa S, Suga M, et al. Structure of the connexin 26 gap junction channel at 3.5 A resolution. *Nature*. 2009;458(7238):597-602. doi:10.1038/nature07869

46. Lee HJ, Cha HJ, Jeong H, et al. Conformational changes in the human Cx43/GJA1 gap junction channel visualized using cryo-EM. *Nat Commun*. 2023;14(1):931. doi:10.1038/s41467-023-36593-y

47. Marandykina A, Palacios-Prado N, Rimkutė L, Skeberdis VA, Bukauskas FF. Regulation of connexin36 gap junction channels by n-alkanols and arachidonic acid. *J Physiol*. 2013;591(Pt 8):2087-2101. doi:10.1113/jphysiol.2013.250910

48. Bargiello TA, Oh S, Tang Q, Bargiello NK, Dowd TL, Kwon T. Gating of Connexin Channels by transjunctional-voltage: Conformations and models of open and closed states. *Biochim Biophys Acta Biomembr*. 2018;1860(1):22-39. doi:10.1016/j.bbamem.2017.04.028

49. Compound Sourcing, Selling and Purchasing Platform. Molport. Accessed June 14, 2024. https://www.molport.com/shop/index

50. Home - Enamine. Accessed June 14, 2024. https://enamine.net/

51. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074-D1082. doi:10.1093/nar/gkx1037

52. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*. 1980;36(22):3219-3228. doi:10.1016/0040-4020(80)80168-2

53. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein–ligand docking using GOLD. *Proteins Struct Funct Bioinforma*. 2003;52(4):609-623. doi:10.1002/prot.10465

54. Nissink JWM, Murray C, Hartshorn M, Verdonk ML, Cole JC, Taylor R. A new test set for validating predictions of protein–ligand interaction. *Proteins Struct Funct Bioinforma*. 2002;49(4):457-471. doi:10.1002/prot.10232

55. Margineanu DG, Klitgaard H. The connexin 36 blockers quinine, quinidine and mefloquine inhibit cortical spreading depression in a rat neocortical slice model in vitro. *Brain Res Bull*. 2006;71(1-3):23-28. doi:10.1016/j.brainresbull.2006.07.011

56. Shao M, Yu H, Santhakumar V, Yu J. Antiepileptogenic and Neuroprotective effect of Mefloquine after Experimental Status Epilepticus. Published online October 19, 2023. doi:10.22541/au.166964432.25485110/v2

57. Quinine. Accessed June 14, 2024. https://go.drugbank.com/drugs/DB00468

58. Quinidine: Uses, Interactions, Mechanism of Action | DrugBank Online. Accessed June 14, 2024. https://go.drugbank.com/drugs/DB00908

59. Ding X, Aureli S, Vaithia A, et al. Structural basis of connexin-36 gap junction channel inhibition. Published online December 15, 2023:2023.12.09.570920. doi:10.1101/2023.12.09.570920

Appendix

Here is a list of selected compounds, grouped according to the methods and respective libraries.

Pharmacelera (molport):

Library identifier	Smiles
MolPort-000-653-819	O[C@@H](Cn1c2ccccc2c2ccccc21)C[NH+]1CCOCC1
MolPort-000-687-365	O[C@H](CN1CC[NH2+]CC1)Cn1c2ccccc2c2cc(Br)ccc21
MolPort-000-746-012	c1ccc2c(c1)N1CC[NH+]=C1N2CC[NH+]1CCCCC1
MolPort-001-777-813	Fc1ccc2nc(C(F)(F)F)cc(N3CCC[NH2+]CC3)c2c1
MolPort-001-964-418	Cn1c(C[NH2+]C2CCCCC2)c(C#N)c2ccccc21
MolPort-003-985-196	FC(F)(F)c1cc([C@@H]2CCCC[NH2+]2)cc(C(F)(F)F)c1
MolPort-010-676-952	Cc1[nH]c2c(C)cccc2c1[C@H](O)C[NH+]1CCCCC1
MolPort-020-216-022	C[NH+](C)C[C@]1(O)CCN(C(=O)c2cccc3ccnn23)C1
MolPort-028-578-546	Cc1cc(N2C[C@H](C3CC3)[C@@H]([NH3+])C2)c2cc(F)cc(F)c2n1
MolPort-039-106-991	CN(C)c1nc2cc(F)ccc2cc1C[NH2+]CC1(O)CCCCC1
MolPort-039-117-356	Cc1cccc2cc(C[NH2+]C[C@@H]3CCCO3)c(N(C)C)nc12
MolPort-044-687-794	Cc1nc(N[C@H]2COC[C@@H]2[NH+]2CCCC2)c2ccccc2n1
MolPort-046-184-369	Cc1cc(C)c2[nH+]ccc(N3C[C@H](O)[C@@H]([NH+](C)C)C3)c2c1
MolPort-047-508-065	Cc1nc(C(F)(F)F)ncc1CN1CC[NH2+][C@@H](C)C1

Pharmacelera (enamine):

Library identifier	Smiles
Z31193996	C[NH+]1CCN(c2cc(Cl)c(C(F)(F)F)cc2[N+](=O)[O-])CC1
Z31223625	C[NH+](C)CCNc1ccc([N+](=O)[O-])c2cnccc12
Z94821000	CC[NH+]1CCC(Nc2nc(C(F)(F)F)nc3ccccc23)CC1

Z368506552	C[N@@H+]1CCCN(c2nc(C(F)(F)F)nc3ccc(F)cc23)CC1
Z376472044	C[N@@H+]1CCCN(c2nc(C(F)(F)F)nc3ccccc23)CC1

GOLD (molport):

Library identifier	Smiles
MolPort-023-325-809	c1ccc(C)c(c12)nc(C)cc2C(=O)NCCC(C)n3cc[nH+]c3
MolPort-023-268-615	c1[nH+]ccn1C(C)CNC(=O)c2cc(C)nc(c23)c(C)c(C)cc3
MolPort-013-634-272	[N+]1CCCC1COc(c(C)n2)nc(c23)cccc3
MolPort-027-882-295	NC(=O)CC1CCCC[N+]1Cc(cc2=O)[nH]c(c23)c(C)ccc3C
MolPort-021-757-342	OCC1CCCC[N+]1CCNc2nc(-c3ccncc3)nc(c24)n(C)nc4C
MolPort-020-227-732	[N+]1CCCCC1CCc2cc(cc(c2)O)-c3cc(N(C)C)ncn3
MolPort-011-735-826	C1OCCCC1[N+]Cc(c2)ccc(c23)OCCCO3
MolPort-016-609-513	c1ccccc1C(c2ccco2)CC[N+]C3CC[N+]CC3
MolPort-019-801-288	c1ccccc1C(C[N+](C)C)Nc(nc(C)n2)cc2CC[N+]