Article

Complementary Dual Approach for In Silico Target Identification of Potential Pharmaceutical Compounds in Cystic Fibrosis

Liza Vinhoven [1], Frauke Stanke [2,3], Sylvia Hafkemeyer [4] and Manuel Manfred Nietert [1,5,*]

[1] Department of Medical Bioinformatics, University Medical Center Göttingen, Goldschmidtstraße 1, 37077 Göttingen, Germany
[2] Clinic for Pediatric Pneumology, Allergology and Neonatology, Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
[3] Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), German Center for Lung Research, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
[4] Mukoviszidose Institut, In den Dauen 6, 53117 Bonn, Germany
[5] CIDAS Campus Institute Data Science, Goldschmidtstraße 1, 37077 Göttingen, Germany
* Correspondence: manuel.nietert@med.uni-goettingen.de; Tel.: +49-551-39-14920

Abstract: Cystic fibrosis is a genetic disease caused by mutation of the CFTR gene, which encodes a chloride and bicarbonate transporter in epithelial cells. Due to the vast range of geno- and phenotypes, it is difficult to find causative treatments; however, small-molecule therapeutics have been clinically approved in the last decade. Still, the search for novel therapeutics is ongoing, and thousands of compounds are being tested in different assays, often leaving their mechanism of action unknown. Here, we bring together a CFTR-specific compound database (CandActCFTR) and systems biology model (CFTR Lifecycle Map) to identify the targets of the most promising compounds. We use a dual inverse screening approach, where we employ target- and ligand-based methods to suggest targets of 309 active compounds in the database amongst 90 protein targets from the systems biology model. Overall, we identified 1038 potential target–compound pairings and were able to suggest targets for all 309 active compounds in the database.

Keywords: cystic fibrosis; docking; ligand-based drug design; target-based drug design/target identification; virtual screening

1. Introduction

Cystic fibrosis (CF) is one of the most common genetic diseases prevalent among the population of Caucasian ancestry, where it affects approximately 1 in 3000 newborns [1–3]. It is caused by mutations of the cystic fibrosis transmembrane conductance regulator (CFTR) gene [4], which encodes a membrane protein that serves as a chloride and bicarbonate channel in the exocrine epithelia of various organs, thereby regulating the viscosity of the mucus lining [5]. Defective CFTR, therefore, has severe implications throughout the body, its major hallmarks being recurrent pulmonary infections and pancreatic insufficiency [4,5]. Of the currently known 2100 mutations of the CFTR gene, several hundred have been shown to be disease causing [6–8]. These mutations cause disturbances throughout CFTR's intricate and delicately balanced biogenesis, which, even in its wild-type (wt) form, only converts 20–40% of the transcripts into a fully functional protein [9]. In order to make handling and working with the multitude of mutations easier, they are categorized into different classes, depending on what kind of defect they cause. Originally, four major mutation classes were proposed, and over the years, this has been expanded to seven classes (I: no CFTR synthesis; II: CFTR trafficking defect; III: CFTR dysregulation; IV: defective gating; V: reduced CFTR transcription; VI: less stable protein; and VII: no CFTR mRNA) [10–14]; however, as many mutations cause multiple defects, an expanded, combinatorial classification system

was proposed [15]. Here, mutations are assigned to groups of all possible combinations of the single-defect mutation classes, resulting in a more comprehensive grouping. In addition to the multi-defect mutations, patients often carry more than one mutation. These factors lead to a vast range of geno- and phenotypes, which makes the development of effective causative therapeutics especially challenging. In recent years, different small-molecule therapeutics have been developed for clinical applications, which improve the CFTR function by directly targeting the CFTR protein and not just alleviating symptoms of CF-patients. Currently, four pharmaceutical drugs, different combinations of four compounds, are approved and available as causative therapy to some CF patients [16–19]. The active compound in the first approved drug, Kalydeco, is Ivacaftor, a CFTR potentiator which is approved mainly for gating mutations [20,21]. The other drugs are combination therapies [18,19], i.e., they contain two or more active compounds and thereby target multiple defects. Orkambi contains Ivacaftor and Lumcaftor, a CFTR corrector, which acts as small-molecule chaperone to correct the folding defect of some class II mutations [22,23], and Symdeco, which contains Ivacaftor and Tezacaftor, an alternative CFTR corrector [24]. The most recent drug, Kaftrio (known as Trikafta in the US), is a triple combination of Ivacaftor, Tezacaftor and Elaxacaftor, which also acts as CFTR corrector [25,26]. Still, for about 10% of patients, especially those with rare mutations, there is no causative medication available [27]. In an effort to find effective treatments for all patients, the search for CFTR modulators, especially synergistic compound combinations, is ongoing [28,29]. Thousands of compounds are being tested in different cell-assays, often in high-throughput screens, leading to large amounts of data and various candidate substances [20,30–41]. In order to structure and collect the compounds tested as CFTR modulators, we previously developed the publicly available database CandActCFTR (candactcftr.ams.med.uni-goettingen.de), where compounds are annotated and categorized according to various characteristics, including their mode of action and order of interaction with CFTR [42,43]. When analysing the data, it becomes apparent that for about 70% of the active compounds, it is unknown whether they affect CFTR directly through physical interaction, or indirectly through its interactome. To support the elucidation of their mode of action, we previously developed the CFTR Lifecycle Map (cf-map.uni-goettingen.de), where we used a systems biology approach to create a human- and machine-readable model of the CFTR maturation pathway in cells [44,45]. The CFTR Lifecycle Map is written in the standardized SBGN Process Description format and comprises detailed representations of the molecular interactions and pathways CFTR undergoes during its entire biogenesis. It contains 156 reactions with 262 different molecular entities, including 170 biomacromolecules, mainly proteins. Here, we now connect the two resources in order to shed light on the mechanism of action of active compounds by identifying their targets. For this purpose, we are using and combining two different reverse screening approaches. Traditionally, in drug discovery, virtual screening approaches are applied to find bioactive compounds that bind to a specific target protein. In reverse screening, the opposite approach is employed to identify the target proteins of active compounds [46,47]. This is becoming increasingly important in order to predict drug side effects and for drug repositioning, where existing drugs are repurposed for other disorders. Several computational methods exist for reverse screening, which, similarly to traditional virtual screening, can generally be divided into three categories [47]. The first class is target-based approaches, mainly docking, which requires high-quality protein structures and has high computational costs. The two other classes are ligand based, either on their shape or their pharmacological features. These require a solid data foundation of known ligand–target interactions as reference databases. In the last decade, reverse screening approaches have been applied to a range of use cases, especially to find targets of natural compounds [46,47]. For example, a molecular shape-based method was employed to identify the cyclin dependent kinase (CDK2) as target of the phytochemical curcumin as possible explanation for its cancer-preventive properties [48]. Furthermore, inverse docking, i.e., a target-based approach, has been used to, amongst others, study different helicases in Zika viruses as targets of ligands from a flowering plant [49], inves-

tigate the effect of thyme derived thymol on fat deposition [50], and shed light on the antitumor targets of a library of natural bioactive compounds [51]. While most studies use either one or the other approach, we here use both a target- and ligand-based approach independently and combined to identify possible targets in the CFTR Lifecycle Map of the active compounds from CandActCFTR. By using the two approaches, we were able to include more potential protein targets than by exclusively using docking or shape-based approaches. Ultimately, we tested 309 active compounds against 90 targets, resulting in almost 30,000 possible combinations. Of these, 1038 unique target–compound pairings were identified with graduated confidence levels in the range of 1–5. Importantly, we could suggest at least one target for each active compound in the database, thereby bringing the elucidation of their mechanism of action one step closer and serving as a basis for finding novel compounds (classes) and predicting synergistic compound combinations. This application shows how systems medicine disease maps, such as the CFTR Lifecycle Map, can be utilized for in silico target identification and to provide the means to fill knowledge gaps and support drug design.

2. Results

2.1. PDB Targets

In order to suggest possible mechanisms of actions of active compounds in the CandActCFTR database [43], possible protein targets that are involved in CFTR biogenesis were selected from the CFTR Lifecycle Map [45]. For this purpose, the Protein Data Bank [52] was queried for experimental X-ray or CryoEM structures of all proteins within the CFTR Lifecycle Map. The list was then filtered according to different criteria such as structure completeness, resolution and experimental method. This list was narrowed down to 35 PDB structures, including that of wt-CFTR, which are listed in Supplementary Table S1, together with their experimental specifications and their role in the CFTR biogenesis. Overall, for 35 of the 170 proteins present in the CFTR Lifecycle Map, appropriate PDB structures could be found, which cover a diverse range of functions and stages in the CFTR biogenesis. This will be especially important to develop combination therapies with synergistic effects that influence CFTR at different stages of its lifecycle. However, for some steps in the CFTR lifecycle, specifically at the transcription stage, no docking-appropriate PDB structures of CFTR interactors could be found. To remedy this situation, structure predictions from the AlphaFold database (https://alphafold.ebi.ac.uk/, accessed on 9 September 2022) [53–55] were considered as alternatives. Unfortunately, when comparing the results of the blind cross-docking between a set of reference protein structures with co-crystallized ligands and their predicted counterparts, the predicted structures resulted in much less accurate docking results. Since the aim of this study was to find potential targets for our query ligands, rather than of putative ligands for specific proteins, the inaccurate docking results from predicted structures could potentially bias the overall results. It was therefore decided not to use the predicted structures from the AlphaFold database and continue with the subset of 35 experimental structures from the PDB database. Nonetheless, except for the transcription stage, the 35 targets are well distributed across the CFTR Lifecycle Map. More precisely, in addition to CFTR itself, 9 potential targets are involved in translation, folding and ER quality control, 7 are associated with the secretory pathway, 5 with endocytosis, and 13 are involved in CFTR activity.

2.2. Docking

All active compounds from the CandActCFTR [42,43] database were docked against all 35 targets using the Smina [56] and QuickVina-W (qvina-W) [57] docking programs. After docking, two methods (2.1) were used for post-processing. The first method [51] (method I) filters out false positives and ultimately results in a list of the overall best-scoring target–ligand pairings, while the second method [58] (method II) also filters out false positives but then calculates the most likely target for each ligand. A comparison of the results from all four approaches can be seen in Figure 1A. Method I resulted in a list of

21 target–ligand pairings for the docking results from the qvina-w docking program. Of the 48 high-scoring pairings, 19 were also among the ones identified by method II.
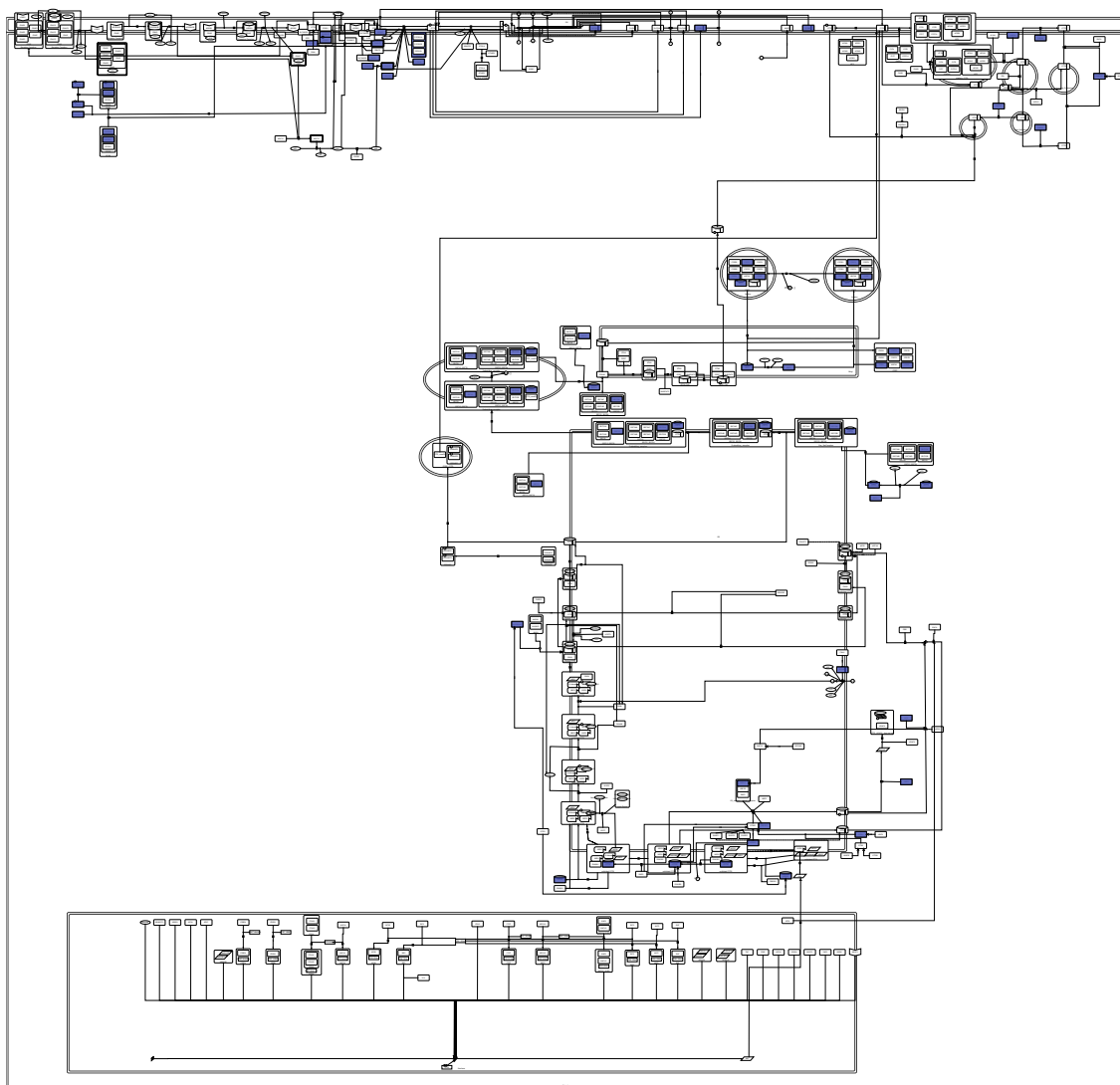
Figure 1. Available protein structures of targets in the CFTR Lifecycle Map. Highlighted in blue are all proteins of which an appropriate PDB structure could be found and which were used as targets in the target-based approach.

Using the results from the Smina docking program, 45 target–ligand pairings were identified using the method I, 30 of which were also identified by the second approach.

When comparing the results from the two docking programs, method II identified 43 common target–ligand pairings, while there were no common pairings found by both method I approaches.

Figure 1 shows the number of ligands attributed to the targets by all methods combined. On average, the mean Tanimoto similarity amongst ligands associated with the same target was 0.25 on average, indicating that there is no bias towards a specific compound class for one target. As can be seen, PRKACA (the catalytic subunit α of protein kinase A) (108 ligands) and CSNK2A1 (Casein kinase II subunit α) (60 ligands) have a lot of compounds ascribed to them by at least one approach. As both analysis methods take into account average scores for each ligand over all targets, the docking results for PRKACA and CSNK2A1 were removed from the data, and the entire analysis of the docking results was repeated, in order to eliminate bias from these two targets.

Figure 2B shows that the number of ligands identified per target are much more evenly distributed after removing the PRKACA and CSNK2A1 from the docking data. The target with the most ligands associated with it here is RAB5A (Ras-related protein Rab-5A) (47 ligands). On average, the mean Tanimoto similarity amongst ligands associated with the same target was 0.21, so they appear to be structurally different.
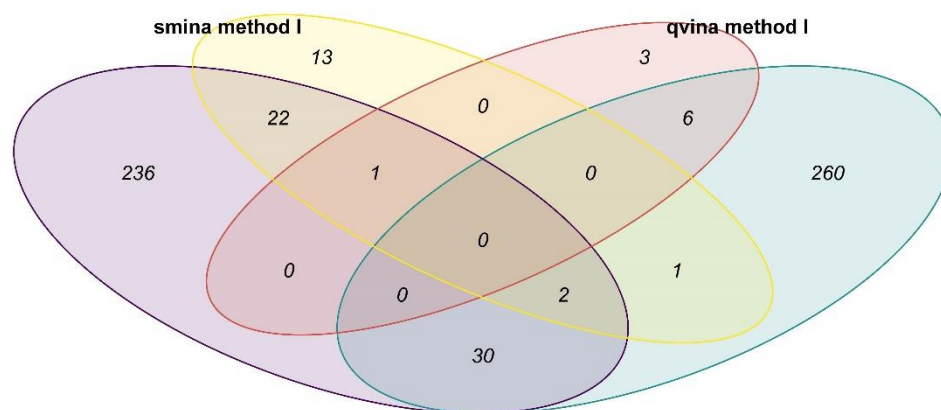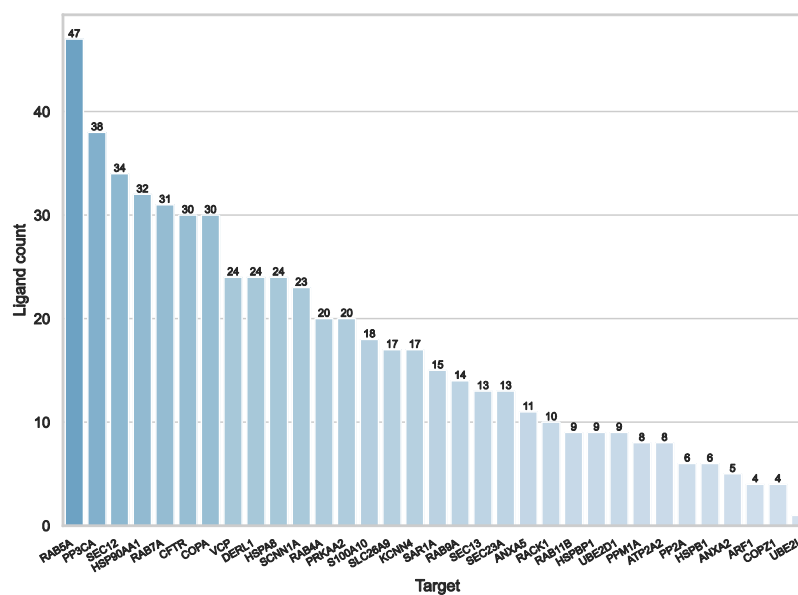

(**A**)
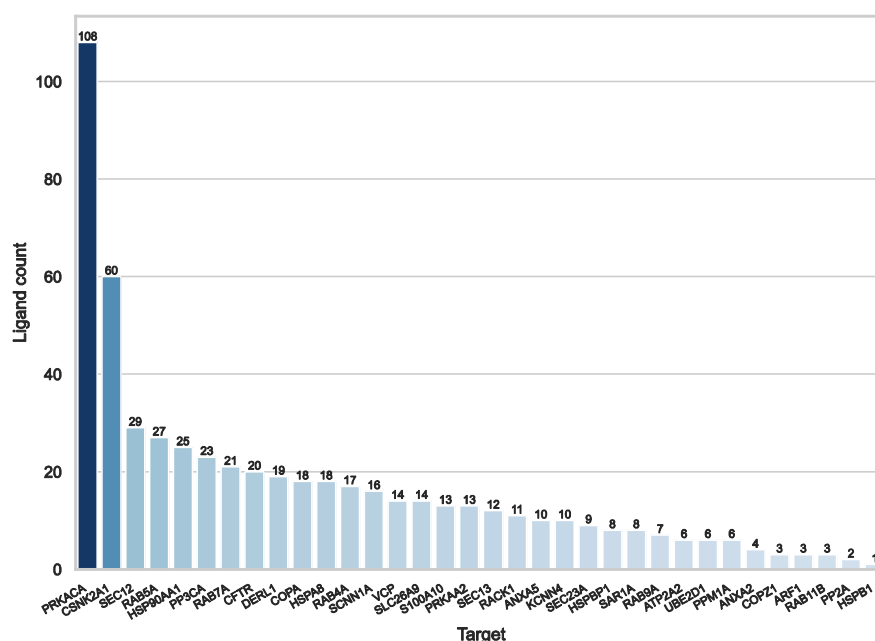
(**B**)

Figure 2. Results of the target-based approach using all targets. (**A**): Venn diagram representing the number of target–ligand pairings identified by both methods with both docking programs and the overlap in results. The yellow ellipse shows the pairings identified by method I using the docking results by Smina, the red ellipse shows the results of method I using qvina-w docking results. Shown in purple and teal are the pairings identified by method II using the Smina and qvina-w docking results, respectively. (**B**): The number of ligands associated with each target by all four approaches combined.

The number of targets per ligand of all pairings identified by method I (including and excluding the results with PRKACA and CSNK2A1) can be seen in Figure 3. As can be seen, the number of potential targets for each ligand is relatively low, indicating specific interactions. Most ligands are only associated with one or two targets, one is associated with three and four targets, respectively, two to five targets, and only one ligand is associated with six targets. The ligand with the most targets associated (ligand 2144) is the deoxyribose dATP, which is predicted to bind to the Ras-related proteins RAB4A, RAB5A, RAB7A, RAB9A, and RAB11B and the secretion-associated Ras-related GTPase 1A (SAR1A), all of which are GTP-binding proteins.


(**A**)

(**B**)

Figure 3. Results of the target-based approach using excluding PRKACA and CSNK2A1. (**A**): Venn diagram representing the number of target–ligand pairings identified by both methods with both docking programs and the overlap in results. The yellow ellipse shows the pairings identified by method I using the docking results by Smina, the red ellipse shows the results of method I using qvina-w docking results. Shown in purple and teal are the pairings identified by method II using the Smina and qvina-w docking results, respectively. (**B**): The number of ligands associated with each target by all four approaches combined.

Figure 3B shows the Protein–Ligand-Interaction network for all pairings identified by method I (including and excluding the results with PRKACA and CSNK2A1). Each edge between two targets stands for a compound found for both of them, the width of the

edge representing the number of compounds and the size of the node (targets) representing its degree, i.e., how many other nodes it is connected to. Most targets have only one or two potential compounds in common, except PRKACA and CSNK2A1, which share four compounds.

As can be seen in the Venn diagram in Figure 2A, the consensus amongst the different approaches was similar compared to when all targets were used. For the qvina-w docking program, method I led to ten target–ligand parings, six of which were also identified by method II. The results from the Smina docking program led to 39 pairings using method I, 25 of which were also found with method II. Comparing the two docking programs, method II identified 32 common pairings, and method I identified only one common pairing. Interestingly, the ligand in this pairing is Lumacaftor (InChI Key: UFSKUSARDNFIRC-UHFFFAOYSA-N), also known as VX-809, a well-known and clinically approved CF-drug, known to bind to CFTR directly. Here, however, it is predicted to bind to the Ras-related protein Rab-7a (RAB7A), a protein involved in endocytosis.

Two different binding pockets and binding poses were identified by both docking programs, Smina and qvina-w, in accordance. The binding pockets can be seen in Figure 4, where the poses predicted by Smina are coloured in blue, and the ones predicted by qvina-w are green. Binding pocket A corresponds to the GTP binding site of RAB7A. The two docking poses calculated by Smina and qvina-w, respectively, have an RMSD of 2.30 Å in pocket A and 2.24 Å in pocket B. Figure 5 shows the 2D poseview of Lumacaftor docked into both pockets by both docking programs. The black dotted lines represent hydrogen bonds, the green spline sections represent hydrophobic interactions and the green dotted lines show π-π stacking or π–cation interactions. As can be seen in Figure 5, in pocket A, for both predicted binding poses, Lumacaftor interacts with various residues via hydrogen bonds, and it undergoes hydrophobic interactions with tyrosine-37 and an π–cation interaction with lysine-126. However, while the binding pose calculated by Smina shows a π–cation interaction with the Magnesium ion in the GTP binding site, the binding pose predicted by qvina-w suggests a π-π stacking interaction with tyrosine-37. In binding pocket B, both binding poses suggest Lumacaftor undergoes hydrogen bonding with serine-72. The binding pose predicted by qvina-w suggests an additional hydrogen bond with glutamine-71, while the one predicted by Smina shows π-π stacking with tryptophane-102, as well as hydrophobic interactions. Overall, both docking programs predict a higher binding affinity for pocket A (Smina 13.4 kcal/mol, qvina-w 12.4 kcal/mol) than for pocket B (both 10.8 kcal/mol).


(A)

Figure 4. *Cont.*

(B)
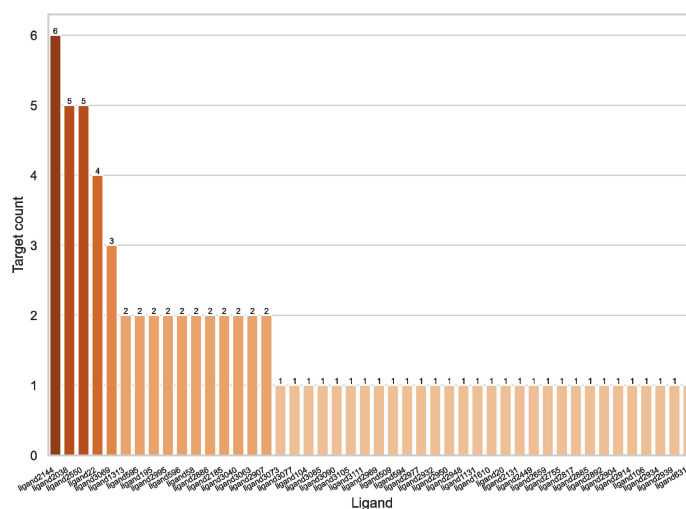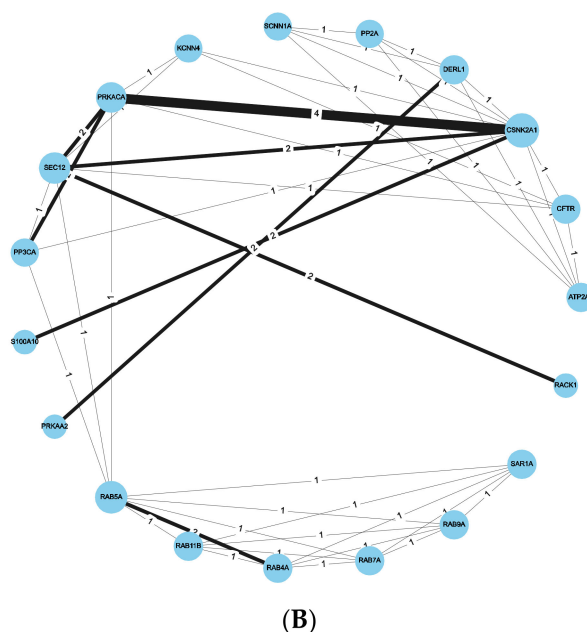
Figure 4. Results of method I using both docking programs combined, including and excluding PRKACA and CSNK2A1. (**A**): Number of targets associated with each ligand. (**B**): Protein–Ligand-Interaction network. Each nodes (circle) stands for one target, while the edges (lines) connecting them represent the number of common ligands (=number on edge) associated with them by either method. The thicker the edge, the more ligands two targets share. The size of the nodes represents their degree, i.e., the number of other nodes they are connected to.
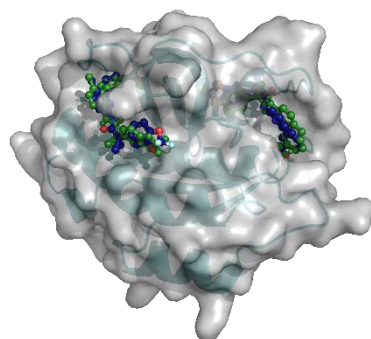
Figure 5. Docking results of Lumcaftor to RAB7A (PDB 1T91). Pocket A is shown on the right side, pocket B on the left. The docking poses of Lumacaftor predicted by qvina-w are coloured in green, the ones by Smina in blue.

Two other target–ligand pairings were identified by three of the four approaches after the exclusion of PRKACA and CSNK2A1 from the docking data. The Ras-related protein Rab-4a (RAB4A) was predicted to interact with ligand2995 (InChI Key: PSPRNONTFBJUDQ-SCFUHWHPSA-N), an ATP analogue. Figure 6A shows the ligand at the GTP binding site of RAB4A. There it is suggested to be coordinated via hydrogen bonding with leucine-155, serine-158, alanine-154 and lysine-124, as well as a π-π stacking interaction with phenylalanine-35. For this pairing, qvina-w calculated a binding affinity of −10 kcal/mol, and Smina calculated a binding affinity of −12.7 kcal/mol.
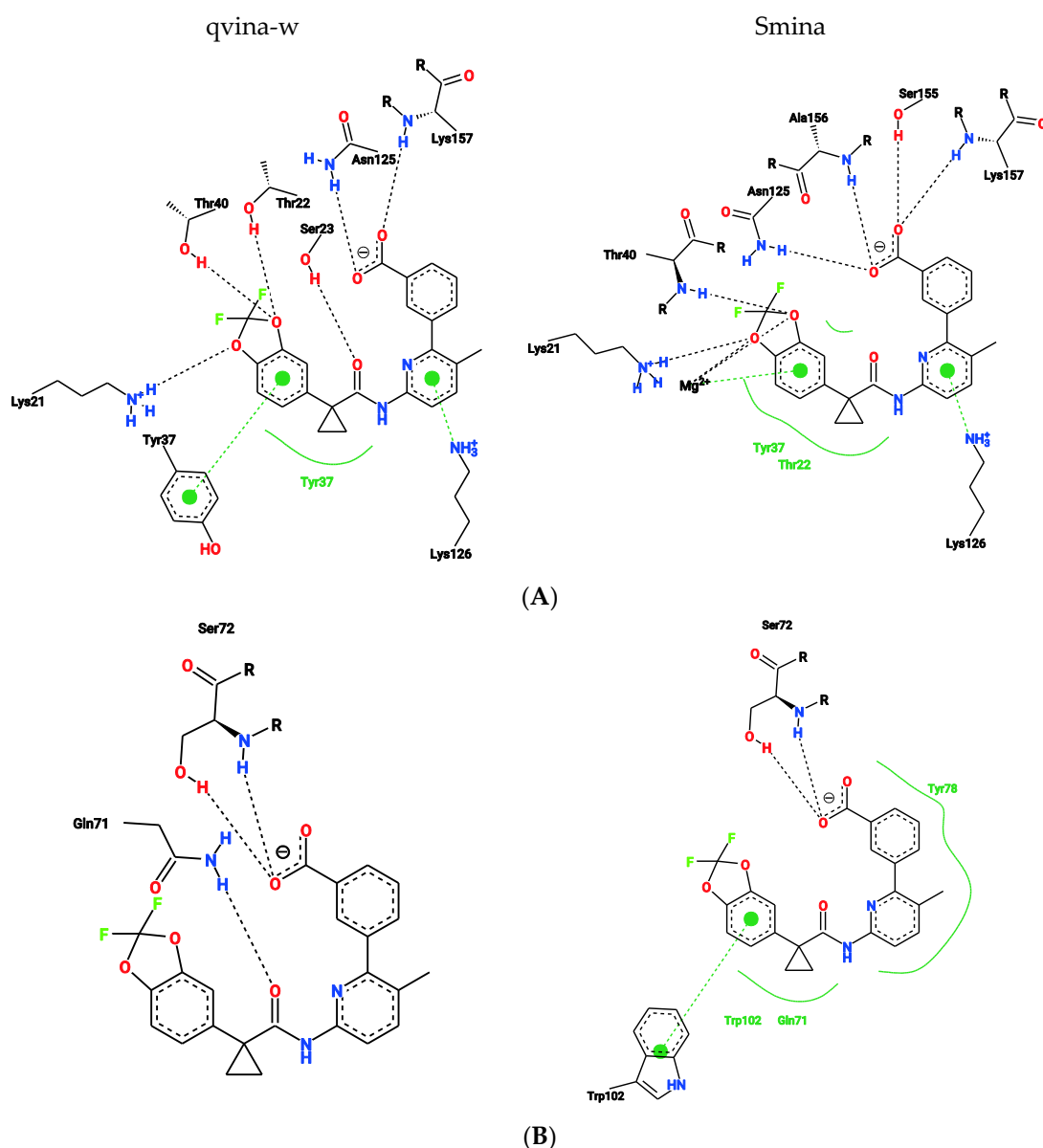
Figure 6. Two-dimensional representation of the docking poses of Lumacaftor to RAB7A and predicted interactions between them. Black dashed lines represent hydrogen bonds, green spline sections show hydrophobic interactions and green dashed lines represent π-π stacking or π−cation interactions. (**A**) Docking poses calculated by qvina-w and Smina for binding pocket A. (**B**) Docking poses calculated by qvina-w and Smina for binding pocket B.

Furthermore, Derlin-1 (DERL1), also known as degradation in endoplasmic reticulum protein 1, was predicted to interact with Equol (ligand104, InChIKey: ADFCQWZHKCXPAJ-GFCCVEGCSA-N). Figure 6B shows the predicted binding pocket of Equol in DERL1, where it is predicted to undergo π-π stacking interactions with tryptophane-106 and 18. Here, qvina-w calculated a binding affinity of −8.9 kcal/mol, while Smina calculated a binding affinity of −11.1 kcal/mol.

2.3. Ligand Similarity Approach

For the ligand-based similarity approach, known ligands for the 35 selected targets were collected from the databases BindingDB [59] (https://www.bindingdb.org, accessed on 9 September 2022) and ChEMBL (https://www.ebi.ac.uk/chembl/, accessed on 9 September 2022) [60–62]. Overall, 6787 target–ligand interactions could be found in the

BindingDB and 8536 were collected from ChEMBL. When merging the two datasets, 3631 of target–ligand interactions could be found in both databases, resulting in a combined dataset of 11,692 unique interactions. From ChEMBL, ligands could be found for 22 targets, of which 15 also had ligands listed in the BindingDB. The targets and the number of ligands per target from each database are listed in Supplementary Table S2. The most ligands could be found for the chaperone HSP90AA1 (3112 ligands), followed by the protein kinases PRKACA (1730 ligands), CSNK2A1 (1710 ligands) and PRKAA2 (1551 ligands), which are well-known pharmacological targets. For CFTR, 1012 ligands could be found, while less than 1000 compounds were found for the remaining targets. Hence, ligand-based similarity comparisons were conducted for 22 of the 35 targets.

For this purpose, the molecular fingerprints of the reference compounds and the query compounds were calculated and compared pairwise to each other by calculating the Tanimoto similarity. Using a similarity cut-off of 0.75, similar compounds between reference and query ligands could be found for eight targets (Supplementary Table S3). Figure 7A shows that the most compounds were found for CFTR (52 compounds), followed by the catalytic subunit of the Serine/threonine-protein phosphatase 2B (PPP3CA; 34 compounds), and the kinases PRKACA (11 compounds), PRKAA2 (11 compounds) and CSNK2A1 (10 compounds). Two compounds could be found for each of the chaperones HSPB1 (also known as Hsp27) and HSP90AA1, and one for ATPase 2 (ATP2A1). Of the 52 compounds identified for CFTR, 13 were previously known to interact with CFTR directly. Reversely, not all compounds reported as directly interacting with CFTR in the CandActCFTR database were also identified via ligand similarity, showing that the reference ligands from ChEMBL and the BindingDB do not exhaustively include all known ligands.

Figure 7B shows the Protein–Ligand-Interaction network of the results. As can be seen, the kinases PRKACA, PRKAA2 and CSNK2A1 share the most compounds with each other. All 11 compounds associated with them are shared between PRKACA and PRKAA2, 9 of which they also share with CSNK2A1. All three kinases share the same seven compounds with CFTR and the same two with HSPB1. PRKACA and PRKAA2 additionally share two identical compounds with PPP3CA, which also shares one compound with each CFTR and HSP90AA1, who share another compound amongst them.
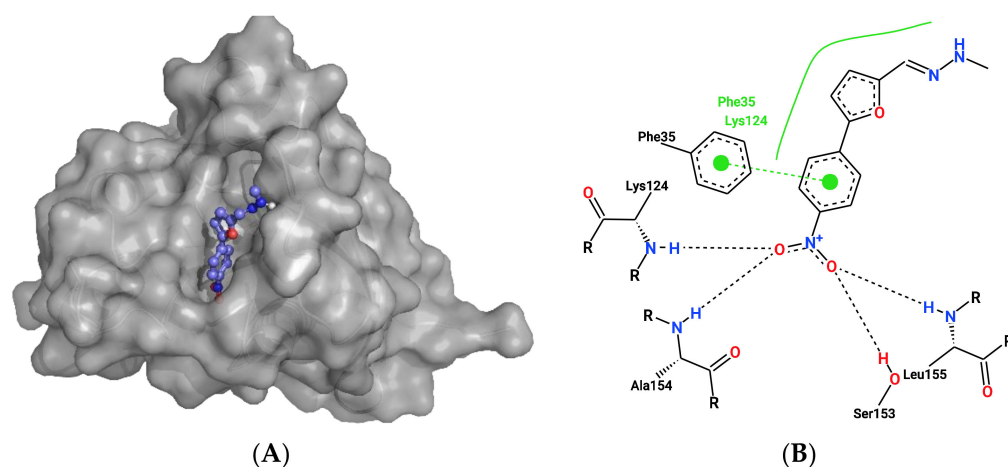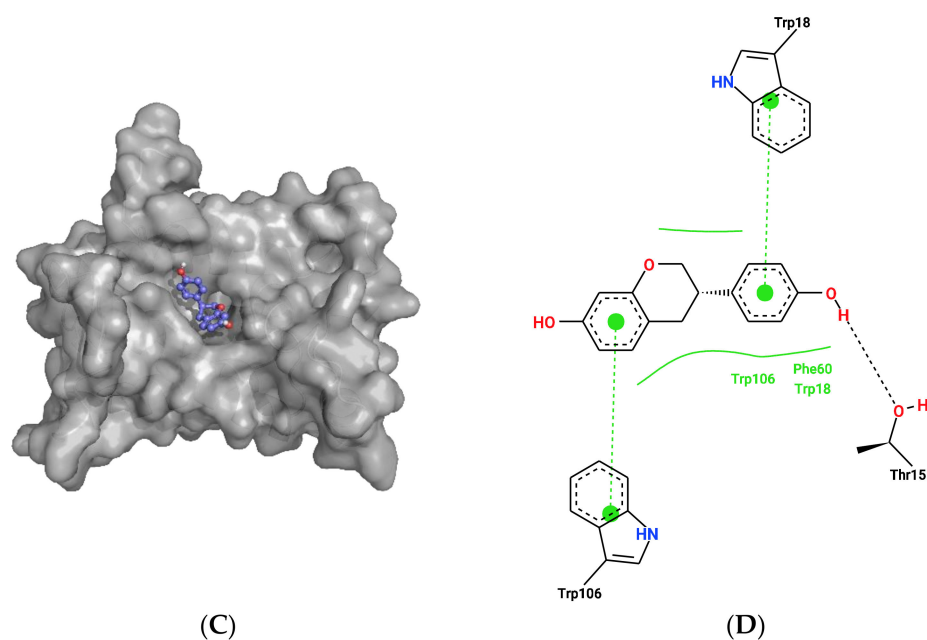

(**A**)   (**B**)

Figure 7. *Cont.*

(C)   (D)

Figure 7. Docking results for ligand2995 to RAB4A and Equol to DERL1. The upper panel shows the docking results of ligand2995 to RAB4A (PDB 2BME). (**A**) Ligand2995 (coloured in purple) docked to RAB4A (PDB:2BME) crystal structure at the GTP (coloured in grey) biding site. (**B**) Two-dimensional representation of the binding pose of ligand2995 to RAB4A and predicted interactions between them. Black dashed lines represent hydrogen bonds, green spline sections show hydrophobic interactions and green dashed lines represent π-π stacking or π–cation interactions. (**C**): Equol (coloured in purple) docked to DERL1 (PDB:7CZB) crystal structure. (**D**) Two-dimensional representation of the binding pose of Equol to DERL1 and predicted interactions between them. Black dashed lines represent hydrogen bonds, green spline sections show hydrophobic interactions and green dashed lines represent π-π stacking or π–cation interactions.

As the ligand-based similarity approach does not depend on protein structures, the dataset was extended by including all remaining potential targets from the CFTR Lifecycle Map, resulting in a list of 168 proteins. The same procedure used for the small dataset was used here to identify potential targets of active compounds in the CandActCFTR database. For all targets combined, 36,851 target–ligand interactions could be found in the BindingDB, and 29,020 were found in ChEMBL. There was an overlap of 15,984 target–ligand interactions, resulting in a combined dataset of 49,887 unique interactions. Ligands for 75 targets were collected from ChEMBL and ligands for 54 targets were collected from the BindingDB, with an overlap of 52 targets between them. More than 5000 ligands could be found for the histone deacetylase 6 (HDAC6, 6815 ligands), the phosphodiesterase 4D (PDE4D, 6769 ligands), the glucocorticoid receptor NR3C1 (5063 ligands) and the adenosine A2B receptor (ADORA2B, 5032 ligands). Again, the high number of ligands present in the database indicates that they are pharmacological interesting targets. For 8 different targets, more than 1000 ligands could be collected, and less than 100 ligands were found for 46 targets.

The ligand similarity comparisons resulted in potential targets for 108 compounds, distributed across 25 targets. Figure 8A shows the number of compounds identified per target. Again, most compounds were found for CFTR (52 compounds), due to the general bias towards CFTR-relevant compounds amongst the query ligands from CandActCFTR. More than 40 compounds were linked to NR3C1 (42 compounds), the beta-2 adrenergic receptor (ADRB2, 40 compounds), followed again by the catalytic subunit of the Serine/threonine-protein phosphatase 2B (PPP3CA; 34 compounds). For the remainder of the targets, less than 15 compounds were found to be similar to the known ligands.

(**A**)

(**B**)
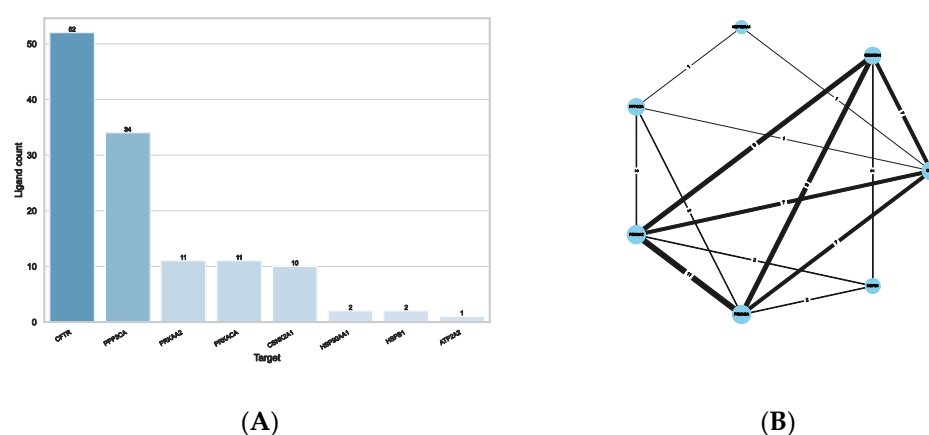
Figure 8. Results from the ligand-based approach using the 35 selected targets. (**A**) Number of ligands associated with each target. (**B**) Protein–Ligand-Interaction network. Each nodes (circle) stands for one target, while the edges (lines) connecting them represent the number of common ligands (=number on edge) associated with them by either method. The thicker the edge, the more ligands two targets share. The size of the nodes represents their degree, i.e., the number of other nodes they are connected to.

Figure 8B shows the Protein–Ligand-Interaction network of the results. As can be seen, the targets with the most shared compounds are NR3C1, ADRB2 and PPP3CA. PPP3CA shares all of its 34 compounds with ADRB2 and NR3C1. The next targets with the most shared compounds are again PRKACA, PRKAA2 and CSNK2A1, which, as already described above, have 11 and 9 shared compounds, respectively. Additionally, CFTR and NR3C1 share an overlap on 9 compounds. Overall, these seven targets (ADRB2, CFTR, CSNK2A1, NR3C1, PPP3CA, PRKAA2 and PRKACA) are the most connected between each other. The remaining targets share a smaller number of compounds with any of the other targets.

2.4. Combined Approach

In order to obtain the most comprehensive overview of potential target–compound interactions, the results from both the structure-based docking and the ligand-based approach were combined.

Overall, a total of 1038 unique target–compound pairings were found, 757 via the structure-based approach and 290 via the ligand-based approach (Supplementary Table S4). The pairings were assigned confidence scores in the range of 1–5, depending on the number of approaches they were identified by. Hence, a score of 5 is assigned when pairings are identified by all methods, i.e., all four target-based approaches as well as the ligand-based approach. The score therefore represents the consensus amongst the different approaches used. The highest score reached was 4, by the RAB7A–Lumacaftor pairing. Overall, 8 pairings have a score of 3, 117 pairings have a score of 2, and the remaining 912 pairings were identified by only 1 approach. No significant structural similarities could be found between the compounds with higher confidence levels, indicating that there is no bias present amongst them.

Furthermore, nine target–ligand pairings were identified by both the target- and ligand-based approach. Of the nine pairings, five ligands were associated with CFTR and two to each of PRKACA and PRKAA2. Four of the five compounds *(Corr4a, InChIKey RDOBOPJBMQURAT-UHFFFAOYSA-N; CHEMBL4471507, InChIKey DBGTUHVUTOSJOJ-UHFFFAOYSA-N; CHEMBL4574818, InChIKey KHNUPLUQJFLSLN-UHFFFAOYSA-N; CHEMBL4435663, InChIKey XTWGHYUHSFTZLL-UHFFFAOYSA-N)* associated with CFTR are similar in structure, but the fifth (InChIKey NOGRVEQYQUCZIW-UHFFFAOYSA-N) differs substantially. One compound, apigenin, was associated with both PRKACA and PRKAA2. Apart from apigenin, PRKAA2 was associated with Kaempferol (InChIKey

IYRMWMYZSQPJKC-UHFFFAOYSA-N) and PRKACA was associated with Biochanin (InChIKey WUADCCWRTIWANL-UHFFFAOYSA-N).

Potential targets were found for all 309 active compounds from the CandActCFTR database (Supplementary Table S5). The distribution of compounds across the targets is visualized in Figure 9. Targets which were associated with at least one compound are displayed in colour, the colour itself representing how many compounds they were associated to, ranging from 1 compound (yellow) to 120 compounds (red). The target with by far the most compounds associated with it was PRKACA (120 compounds), followed by CFTR with 76 compounds, CSNK2A1 with 72 compounds and RAB5A with 52 compounds. Of the remaining targets, 27 had between 10 and 50 ligands associated with them, and 20 targets had less than 10 compounds. When looking at the binding sites of the potential CFTR ligands identified by the target-based approach, mostly five main binding sites were identified (Figure 10). Two of these correspond to the ATP-binding sites in the nucleotide binding domains 1 and 2. One was close to the experimentally resolved binding site of Lumacaftor (coloured in blue), one was close to the one of Ivacaftor (coloured in purple) [63], and there was one additional binding site in the transmembrane domain 2.

(A)   (B)
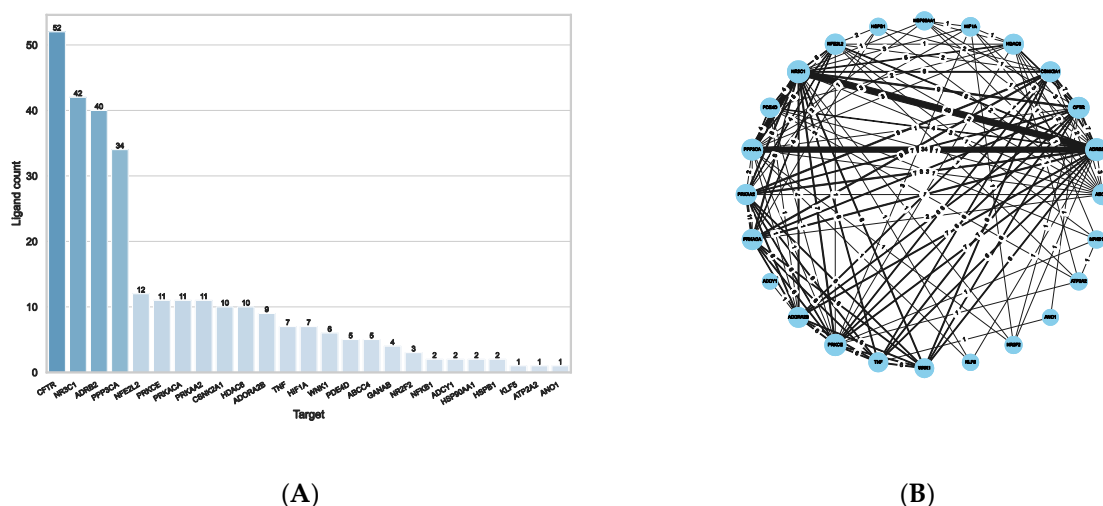
Figure 9. Results from the ligand-based approach using all targets. (**A**) Number of ligands associated with each target. (**B**) Protein–Ligand-Interaction network. Each nodes (circle) stands for one target, while the edges (lines) connecting them represent the number of common ligands (=number on edge) associated with them by either method. The thicker the edge, the more ligands two targets share. The size of the nodes represents their degree, i.e., the number of other nodes they are connected to.

Conversely, for the majority of compounds, 2–4 targets were identified, while a single target was identified for only 15 compounds. Between 5 and 10 targets were identified for 49 compounds, 1 compound has 11 targets associated with it and 6 compounds have 12 targets associated with them. Interestingly, all of these 6 compounds are almost identical in structure and vary only in stereochemistry or side chain. These compounds were all predicted to target the kinases CSNK2A1, PRKAA2, PRKACA, PRKCE and WNK1, but also other targets, namely, ADOR2B, ADRB2, CFTR, NR3C1 and TNF.
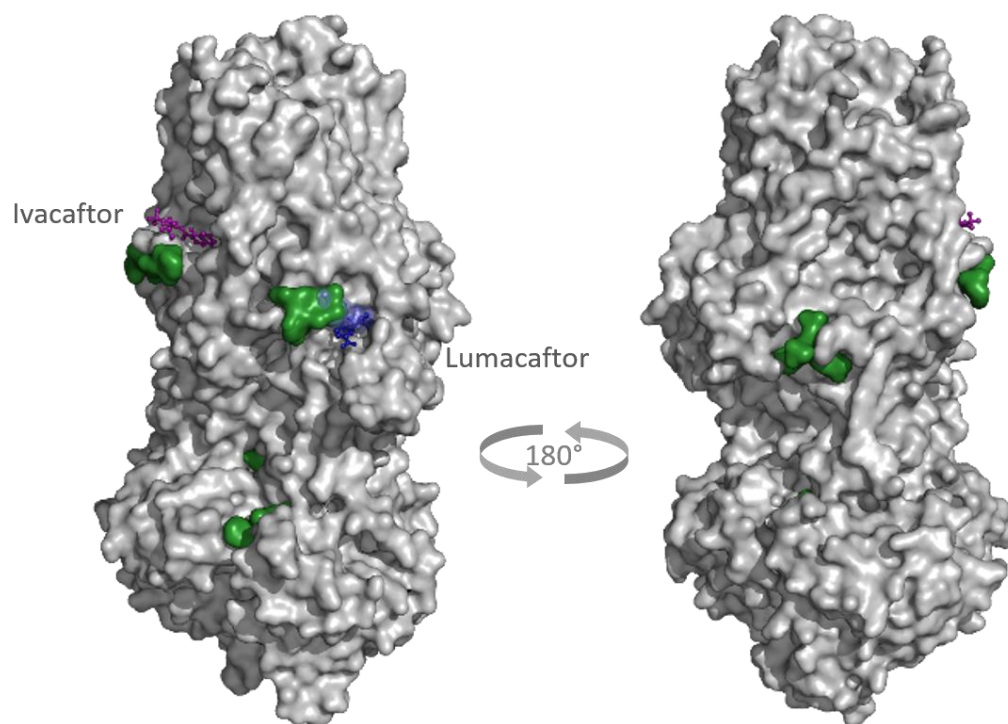
Figure 10. Most predicted binding sites in CFTR by target-based approach (PDB 7SVD and 6O2P). Coloured in green are the potential binding sites. The positions of Lumacaftor (blue) and Ivacaftor (purple) were shown experimentally by Fiedorczuk and Chen, 2022 [63]. *Shown in light blue is the binding site of Lumacaftor predicted by blind docking in the target-based approach, which is in agreement with the experimentally predicted binding site.*

3. Discussion

When searching for new drugs experimentally, especially in high-throughput screens, where thousands of compounds are tested simultaneously, the mechanism of action of promising compounds often remains unclear. Knowing the mechanism of action, however, is helpful and important for a number of reasons. It is useful in order to identify novel drug targets, find new classes of potentially active compounds, and to aid lead optimization. Furthermore, it can be helpful for the early identification of potential side effects. Elucidating the mechanism of action of promising compounds is especially important in the drug discovery for cystic fibrosis, as it is crucial for the development of combination therapies. Here, we bring together two previously developed resources of community-derived knowledge on cystic fibrosis, namely, the CandActCFTR database and the CFTR Lifecycle Map, to aid in the elucidation of modes of action for known active compounds. For this purpose, we use two different approaches to in silico target identification, a target-based approach and a ligand-based approach.

Both approaches have different advantages and disadvantages when it comes to prerequisites. The ligand-based approach is much faster and less computationally expensive. However, in order to be most effective, it requires comprehensive ligand libraries of the targets in question. The target-based approach, on the other hand, requires complete and high-resolution structures of the target proteins. From the 170 potential targets in the CFTR Lifecycle Map, quality-sufficient protein structures could be found for 35 targets, and ligand collections could be found for 77 targets, with an overlap of 22 targets. By combining the approaches, it was thus possible to cover more than 50% of the potential targets in the CFTR Lifecycle Map.

For the target-based approach, all active compounds from the CandActCFTR database were docked blindly into the 35 target structures, meaning that no prior binding site was defined, but the whole protein was searched. In order to produce comprehensive results,

two different docking programs, QuickVina-W (qvina-w) and Smina, were used, which differ in their method to calculate binding affinities. Due to their different scoring functions, qvina-w and Smina produced different results, but with overall similar trends. Generally, the binding affinities calculated by Smina were slightly higher than those calculated by qvina-w. This is most likely caused by their different approaches to estimating ligand–receptor-based affinity. While qvina-w uses the empirical Vina scoring function [57], which is based on machine-learning [64], Smina uses a more physics-based approach [65]. However, the exact binding affinities are of secondary importance in this case, as the aim in target identification is not to calculate the most precise scores, but rather compare the likeliness of different matches to find potential target–ligand pairings. Therefore, in order to normalize the results and remove false positives, two post-processing methods were applied to the docking data independently. Method I results in a list of high-ranking target–ligand pairings, while method II identifies the most likely target for each ligand. Overall, however, the two methods produce similar results. Due to the high number of ligands, method II results in a lot of more potential pairings; however, a high portion of pairings identified by method I are also suggested by method II, underpinning both their validity, while still remaining non-redundant. Interestingly, two proteins, PRKACA and CSNK2A1, were identified in pairings significantly more often than the other ones. Both proteins are protein kinases, which have been previously found to be promiscuous targets [66,67]. With respect to the high number of ligands associated with the PRKACA and CSNK2A1, the results from the ligand-based approach support the results from the target-based approach. Again, the two kinases were amongst those with the most compounds associated with them. Interestingly, a majority of these compounds could be associated with both proteins, and additionally the kinase PRKAA2, which suggests a common, possibly promiscuous, binding motif amongst these kinases. In order to remove bias from PRKACA and CSNK2A1 from the docking calculations, the analysis was repeated without the data for these two kinases.

Overall, the majority of the predicted interactions (506 of the 1038 predicted interactions) involve proteins that play a role in the activity and regulation of CFTR at the plasma membrane. One explanation for this are the readouts of the experimental assays used to identify the active compounds in their original publications. Most assays use the ion conductance directly as a readout, so compounds that affect the activity of CFTR directly at the membrane are readily detected by these methods. Of the remaining predicted interactions, 161 involve proteins that play a role in CFTR translation and folding, 120 involve proteins of endocytosis and 108 involve protein of the secretory pathway. At total of 76 interactions involve CFTR directly, and only 67 interactions involve transcriptions factors or other proteins that influence CFTR transcription.

Remarkably, one of the protein-ligand pairings with the highest consensus amongst the different approaches was for the compound Lumacaftor (VX-809) with RAB7A. Lumacaftor is a well-known, clinically approved drug for patients with F508del mutations, as it acts as a small-molecule chaperone to correct the folding defect of F508del-CFTR. It is known to bind directly to CFTR, with its exact binding site elucidated by CryoEM (PDBs 7SVD and 7SVR) [63]. Here, however, it was suggested to potentially also bind to the Ras-related protein RAB7A, a GTP-binding protein involved in endocytosis, including that of CFTR [68–70]. When looking close at the predicted binding mode, two different binding sites are suggested for Lumacaftor in the RAB7A protein by both docking programs. In both pockets, Lumacaftor is predicted to interact with the protein via different hydrogen bonds, hydrophobic interactions and π-interactions. When looking at the binding mode for Lumacaftor to CFTR as shown in the CryoEM structure (7SVD) [63], the compound is coordinated mainly by hydrophobic interactions and only one hydrogen bond (Figure 11). Hydrophobic interactions are rather weak intermolecular interactions, which is probably why the binding affinity for Lumacaftor was calculated to be higher for RAB7A than CFTR by the docking programs. However, when looking at the raw docking results for CFTR, Lumacaftor is nonetheless amongst the top 10 highest ranking active compounds calculated by both docking programs. Furthermore, Hou et al. previously showed how

RAB7A inhibition increases apical CFTR stability [70]. Hence, while certainly intriguing and requiring further investigation before confirmation, the pairing of Lumacaftor with RAB7A, rather than with CFTR, therefore does not undermine the potential validity of the results, as the scores calculated for the CFTR–Lumacaftor (Figure 12) pairing are only slightly lower than the one for RAB7A–Lumacaftor.
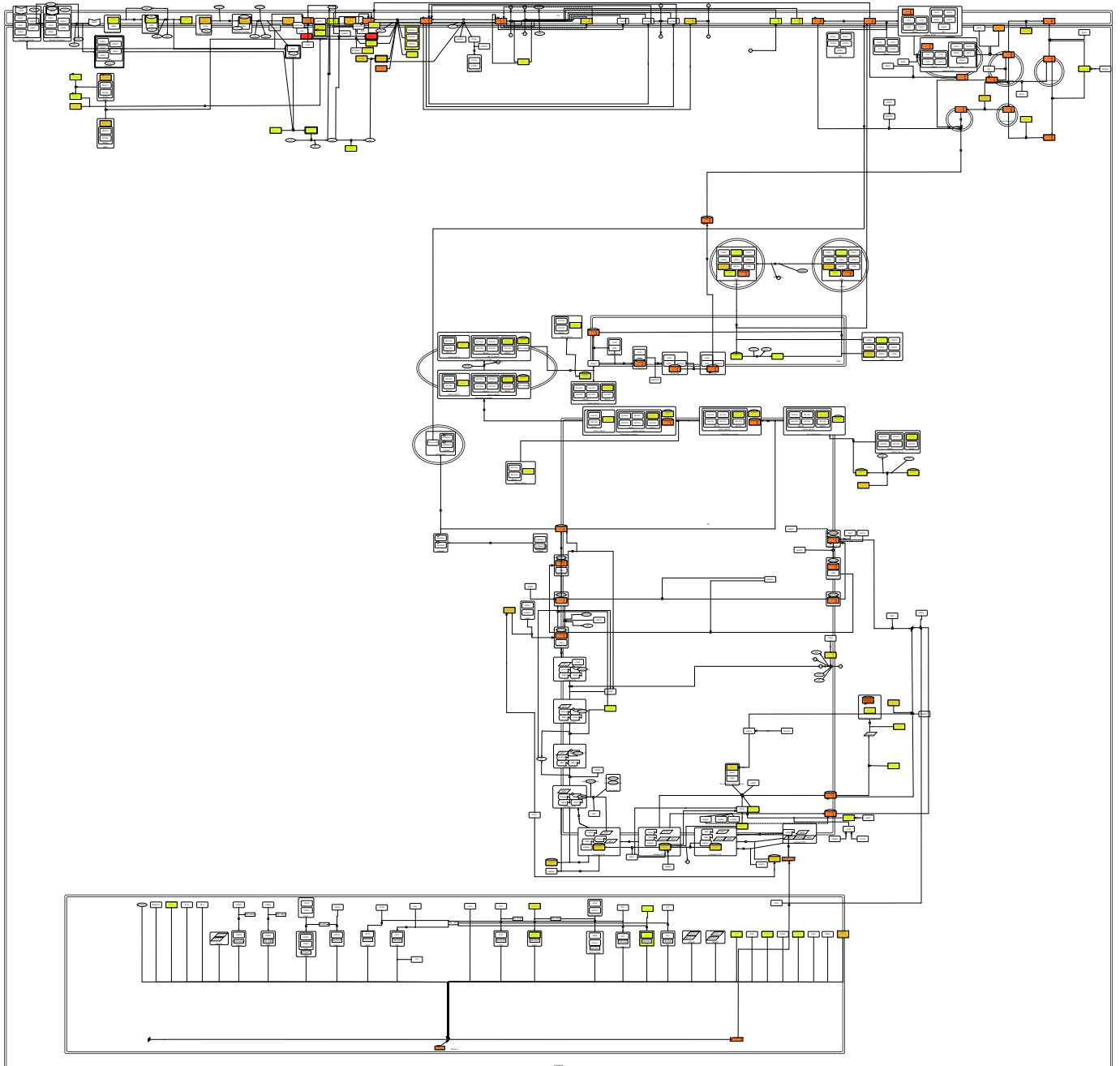
Figure 11. Combined results of all target identification approaches visualized in CFTR Lifecycle Map. The CFTR Lifecycle Map is an SBGN (Systems Biology Graphical Notation) representation of CFTR biogenesis in the cell. All proteins are shown as rounded rectangles. The colour represents the number of compounds associated with each target by all four approaches combined, ranging from one compound (yellow) to 120 compounds (red).
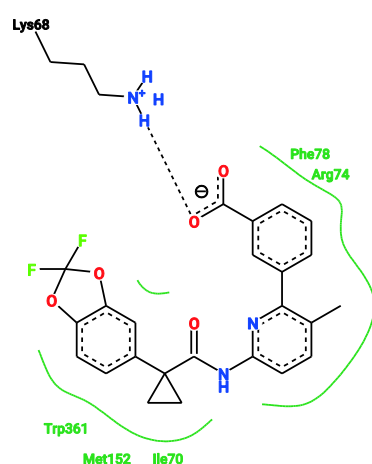
Figure 12. Two-dimensional representation of the experimentally shown binding site of Lumacaftor to CFTR (PDB 7SVD) [63] and predicted interactions between them. Black dashed lines represent hydrogen bonds, green spline sections show hydrophobic interactions and green dashed lines represent π-π stacking or π–cation interactions.

Another high-consensus protein–ligand pairing from the target-based approach was for the compound Equol with DERL1. Equol is an isoflavandiol oestrogen, known to bind to oestrogen receptor β [71]. It was also shown to be on the one hand a potent activator of F508del-CFTR, but on the other hand, it affects its misprocessing [72]. DERL1 is an ER membrane protein mediating the degradation of misfolded proteins such as misfolded CFTR [73,74]. Thus, the effect Equol has on F508del-CFTR misprocessing might be mediated by its binding to DERL1.

Of the 170 different proteins in the map, either protein structures or ligand data could be found for 90 of them, which amounts to a coverage of 53%. This means that 80 proteins (47% of proteins in the map) could not be considered as targets, due to the lack of available data. As is often the case in in silico approaches, this effort to elucidate the mechanism of action of candidate compounds was hampered by the lack of a comprehensive data corpus. Nonetheless, by combining the ligand- and the target-based approach, potential targets were suggested for all 309 active compounds in the CandActCFTR database, distributed across 53 of the potential targets from the CFTR Lifecycle map.

It has to be noted that not all proteins involved in the CFTR biogenesis are equally suitable as targets for CFTR modulators. For example, chaperones and other proteins that are important for CFTR folding, ER quality control, and trafficking play a quite general role in cells and are not specific to CFTR alone. Targeting them can therefore potentially lead to unwanted side effects. Therefore, when selecting compounds and targets for experimental testing, more research is needed on which proteins can be targeted and which should rather not be interfered with.

Of course, experimental validation of protein–ligand interactions is always necessary and preferred to docking- and ligand-similarity approaches. However, where experimental data is not available or its generation not feasible, these in silico approaches are a good way to fill knowledge gaps, suggest potential interactors, and thereby narrow down the candidates for testing in the wet lab. These methods should therefore not be viewed as replacement for wet-lab experiments, but as preceding step to support experimental work and, in the meantime, at least provide indications of possible modes of action.

This project is part of the CandActCFTR project and serves as an essential connecting piece between the chemistry centric CandActCFTR database, which collects compounds tested as CFTR modulators, and the cell biology centric CFTR Lifecycle map, which is a systems biology model of the CFTR biogenesis. By bringing the two parts together here, we not only use and repurpose data from both resources, but are able to generate new knowledge. The structured knowledge in systems medicine disease maps can therefore directly be applied to drug design approaches. By using this comprehensive approach,

we can now suggest mechanisms of actions for active compounds, propose potential drug targets and predict possible additive effects of different substance combinations.

4. Materials and Methods

4.1. Reverse Docking

In order to identify targets for the active compounds of the CF-specific compound database CandActCFTR [42,43], 37 potentially relevant protein targets were used for reverse docking. The proteins were selected using the CFTR lifecycle map [45], a systems biology model of the CFTR lifecycle. For this purpose, a KNIME [75] workflow was used to search all PDB [52] entries for all proteins in the map and then filter them according different criteria such as structure completeness, resolution and experimental method. This list was narrowed down to 35 PDB structures, including that of wt-CFTR, belonging to targets evenly distributed across the CFTR lifecycle. All target structures were prepared for docking using AutoDockTools 1.5.7 [76].

The active compounds from the CandActCFTR database [43] were used as ligand library. Structures were obtained in SMILES notation and converted and prepared for docking using Open Babel [77,78].

Docking of all compounds against all ligands was performed using Virtual Flow [79] with two different docking programs. Specifically, it was carried out using the Smina docking program [56], with the Vinardo scoring function [65] to calculate binding scores between ligands and targets, and the QuickVina-W [57] program, with the AutoDock Vina scoring function [64]. Docking was carried out as blind docking, meaning that no specific binding pocket was defined for each protein, but the entire protein was searched in order to account for structures without known binding site and targets with several, potentially unknown, binding sites. The search space in blind docking is significantly higher than when using pre-defined binding sites, which leads to higher probabilities of not finding the optimal conformation. Nevertheless, to obtain comprehensive results, the exhaustiveness, i.e., the number of runs calculated per ligand and target was increased to 100.

All docking calculations were run on the Scientific Compute Cluster at the joint data centre of Max Planck Society for the Advancement of Science (MPG) and University of Göttingen with the SLURM Workload Manager [80].

Post-processing of the docking was performed using KNIME [75,81] and Pyrhon. In order to remove false positives and obtain more reliable consensus results, two different approaches were used to evaluate target–ligand pairings.

The first approach proposed by Lauro et al. [51] normalizes the binding energies of each target–ligand pairing in a way that eliminates systematic errors and exclude false positives. To do so, the docking results are written in a matrix format and equation (1) was applied, where V is the new value, V_0 is the binding energy from the docking calculate, M_L is the average binding energy of each ligand, and M_T is the average binding energy of each target.

$$V = \frac{V_0}{[M_L + M_T]/2} \tag{1}$$

Promising pairings were then selected by setting a lower threshold at $V = +3\sigma$, where M is the average of the whole matric, and σ is its standard deviation.

The second approach by Kim et al. [58] uses a 2-directional Z-transformation on the docking-score matrix. Here, a target- and ligand-specific Z-score is calculated using Equations (2) and (3), where is x_i the docking score specific to the query ligand i, $\overline{x}_t$ and $\overline{x}_l$ are the average scores of all targets and ligands, respectively, and SD_T and SD_L are the respective standard deviations.

$$Z_T = \frac{(x_i - \overline{x}_T)}{SD_T} \tag{2}$$

$$Z_L = \frac{(x_i - \overline{x}_L)}{SD_L} \tag{3}$$

The combined Z_{comb} is then calculated through $Z_{comb} = 0.7 * Z_T + 0.3 * Z_L$. In order to select the most likely target, the receptor with the lowest Z_{comb} is selected for each ligand. The 2-directional Z-transformation was carried out using the Python script provided by Kim et al. [58].

Results from both filtering approaches were then compared to check for consensus and find the most promising target–ligand pairings.

4.2. Ligand Based Approach

For the ligand-based approach, known ligands of the 35 proteins also used for the reverse docking were collected from different databases. Databases used were the BindingDB [59] (https://www.bindingdb.org, accessed on 9 September 2022) and ChEMBL (https://www.ebi.ac.uk/chembl/, accessed on 9 September 2022) [60–62]. Interaction data from BindingDB were collected using their API via the KNIME workflow provided by the BindingDB team, and from ChEMBL using the ChEMBL websource client [60] for Python. All ligands were collected in the SMILES notation [82], which was converted to the InChIKey [83] using OpenBabel [78].

In order to perform similarity comparisons according to the structural properties of the molecules, for each reference ligand from the databases, as well as each active compound from the CandActCFTR database [42,43], the Morgan fingerprint [84] was computed using the RDKit [85]. Next, the similarity between each reference ligand and each query compound from CandActCFTR was calculated using the Tanimoto coefficient. If a query compound shared a similarity of ≥0.75 with at least one of the target-specific reference compounds, it was considered a potential ligand of the respective target. The threshold of 0.75 was chosen to be less restrictive than the commonly used threshold of 0.85.

5. Conclusions

This study used a complementary approach of target- and ligand-based in silico target identification methods to predict potential targets of compounds that were shown to affect CFTR activity. By using a dual approach, we were able to reap the benefits of both methods and increase the number of potential target proteins to get a better coverage of the CFTR biogenesis. Overall, the number of potential pairings could be narrowed down to 1038 predicted interactions, which are assigned a score depending on how high the consensus is amongst the methods employed. We were thereby able to predict potential targets for all 309 active compounds from the CandActCFTR database. These results help elucidate the mechanism of action of promising compounds and can be used to select compounds and targets to predict synergistically acting compound combinations for testing in the wet lab.

Conflicts of Interest: The authors declare no conflict of interest.

References and Note

1. Bobadilla, J.L.; Macek, M.; Fine, J.P.; Farrell, P.M. Cystic fibrosis: A worldwide analysis of CFTR mutations—Correlation with incidence data and application to screening. *Hum. Mutat.* **2002**, *19*, 575–606. [CrossRef] [PubMed]
2. Farrell, P.M. The prevalence of cystic fibrosis in the European Union. *J. Cyst. Fibros.* **2008**, *7*, 450–453. [CrossRef] [PubMed]
3. Bell, S.C.; Mall, M.A.; Gutierrez, H.; Macek, M.; Madge, S.; Davies, J.C.; Burgel, P.R.; Tullis, E.; Castaños, C.; Castellani, C.; et al. The future of cystic fibrosis care: A global perspective. *Lancet Respir. Med.* **2020**, *8*, 65–124. [CrossRef]
4. O'Sullivan, B.P.; Freedman, S.D. Cystic fibrosis. *Lancet* **2009**, *373*, 1891–1904. [CrossRef]
5. Elborn, J.S. Cystic fibrosis. *Lancet* **2016**, *388*, 2519–2531. [CrossRef]
6. Cystic Fibrosis Mutation Database. Available online: http://www.genet.sickkids.on.ca/ (accessed on 26 January 2021).
7. Welcome to CFTR2 | CFTR2. Available online: https://www.cftr2.org/ (accessed on 26 January 2021).
8. Sosnay, P.R.; Siklosi, K.R.; van Goor, F.; Kaniecki, K.; Yu, H.; Sharma, N.; Ramalho, A.S.; Amaral, M.D.; Dorfman, R.; Zielenski, J.; et al. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat. Genet.* **2013**, *45*, 1160–1167. [CrossRef]
9. Pranke, I.M.; Sermet-Gaudelus, I. Biosynthesis of cystic fibrosis transmembrane conductance regulator. *Int. J. Biochem. Cell Biol.* **2014**, *52*, 26–38. [CrossRef]
10. Welsh, M.J.; Smith, A.E. Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell* **1993**, *73*, 1251–1254. [CrossRef]
11. Rowe, S.M.; Miller, S.; Sorscher, E.J. Cystic fibrosis. *N. Engl. J. Med.* **2005**, *352*, 1992–2001. [CrossRef]
12. Zielenski, J.; Tsui, L.C. Cystic fibrosis: Genotypic and phenotypic variations. *Annu. Rev. Genet.* **1995**, *29*, 777–807. [CrossRef]
13. Zielenski, J. Genotype and Phenotype in Cystic Fibrosis. *Respiration* **2000**, *67*, 117–133. [CrossRef] [PubMed]
14. De Boeck, K. Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatr.* **2020**, *109*, 893–899. [CrossRef] [PubMed]
15. Veit, G.; Avramescu, R.G.; Chiang, A.N.; Houck, S.A.; Cai, Z.; Peters, K.W.; Hong, J.S.; Pollard, H.B.; Guggino, W.B.; Balch, W.E.; et al. From CFTR biology toward combinatorial pharmacotherapy: Expanded classification of cystic fibrosis mutations. *Mol. Biol. Cell* **2016**, *27*, 424–433. [CrossRef] [PubMed]
16. Gentzsch, M.; Mall, M.A. Ion Channel Modulators in Cystic Fibrosis. *Chest* **2018**, *154*, 383–393. [CrossRef]
17. Zaher, A.; ElSaygh, J.; Elsori, D.; ElSaygh, H.; Sanni, A. A Review of Trikafta: Triple Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Modulator Therapy. *Cureus* **2021**, *13*, e16144. [CrossRef]
18. Drug Development Pipeline: CFF Clinical Trials Tool. Available online: https://www.cff.org/Trials/Pipeline (accessed on 26 January 2021).
19. Clinical Pipeline. Available online: https://www.glpg.com/clinical-pipelines (accessed on 28 June 2022).
20. Van Goor, F.; Hadida, S.; Grootenhuis, P.D.J.; Burton, B.; Cao, D.; Neuberger, T.; Turnbull, A.; Singh, A.; Joubran, J.; Hazlewood, A.; et al. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18825–18830. [CrossRef]
21. Ramsey, B.W.; Davies, J.; McElvaney, N.G.; Tullis, E.; Bell, S.C.; Dřevínek, P.; Griese, M.; McKone, E.F.; Wainwright, C.E.; Konstan, M.W.; et al. A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation. *N. Engl. J. Med.* **2011**, *365*, 1663–1672. [CrossRef]
22. Clancy, J.P.; Rowe, S.M.; Accurso, F.J.; Aitken, M.L.; Amin, R.S.; Ashlock, M.A.; Ballmann, M.; Boyle, M.P.; Bronsveld, I.; Campbell, P.W.; et al. Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the F508del-CFTR mutation. *Thorax* **2012**, *67*, 12–18. [CrossRef]
23. Wainwright, C.E.; Elborn, J.S.; Ramsey, B.W.; Marigowda, G.; Huang, X.; Cipolli, M.; Colombo, C.; Davies, J.C.; de Boeck, K.; Flume, P.A.; et al. Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del CFTR. *N. Engl. J. Med.* **2015**, *373*, 220–231. [CrossRef]
24. Taylor-Cousar, J.L.; Munck, A.; McKone, E.F.; van der Ent, C.K.; Moeller, A.; Simard, C.; Wang, L.T.; Ingenito, E.P.; McKee, C.; Lu, Y.; et al. Tezacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del. *N. Engl. J. Med.* **2017**, *377*, 2013–2023. [CrossRef]
25. Voelker, R. Patients with Cystic Fibrosis Have New Triple-Drug Combination. *JAMA* **2019**, *322*, 2068. [CrossRef] [PubMed]
26. Ridley, K.; Condren, M. Elexacaftor-tezacaftor-ivacaftor: The first triple-combination cystic fibrosis transmembrane conductance regulator modulating therapy. *J. Pediatr. Pharmacol. Ther.* **2020**, *25*, 192–197. [CrossRef] [PubMed]
27. Goetz, D.M.; Savant, A.P. Review of CFTR modulators 2020. *Pediatr. Pulmonol.* **2021**, *56*, 3595–3606. [CrossRef] [PubMed]
28. Martiniano, S.L.; Sagel, S.D.; Zemanick, E.T. Cystic fibrosis: A model system for precision medicine. *Curr. Opin. Pediatr.* **2016**, *28*, 312–317. [CrossRef] [PubMed]
29. Southern, K.W.; Patel, S.; Sinha, I.P.; Nevitt, S.J. Correctors (specific therapies for class II CFTR mutations) for cystic fibrosis. *Cochrane Database Syst. Rev.* **2018**, *2018*, CD010966. [CrossRef]
30. Pedemonte, N.; Lukacs, G.L.; Du, K.; Caci, E.; Zegarra-Moran, O.; Galietta, L.J.V.; Verkman, A.S. Small-molecule correctors of defective ΔF508-CFTR cellular processing identified by high-throughput screening. *J. Clin. Investig.* **2005**, *115*, 2564–2571. [CrossRef]

31. Berg, A.; Hallowell, S.; Tibbetts, M.; Beasley, C.; Brown-Phillips, T.; Healy, A.; Pustilnik, L.; Doyonnas, R.; Pregel, M. High-Throughput Surface Liquid Absorption and Secretion Assays to Identify F508del CFTR Correctors Using Patient Primary Airway Epithelial Cultures. *SLAS Discov.* **2019**, *24*, 724–737. [CrossRef]

32. De Wilde, G.; Gees, M.; Musch, S.; Verdonck, K.; Jans, M.; Wesse, A.S.; Singh, A.K.; Hwang, T.C.; Christophe, T.; Pizzonero, M.; et al. Identification of GLPG/ABBV-2737, a novel class of corrector, which exerts functional synergy with other CFTR modulators. *Front. Pharmacol.* **2019**, *10*, 514. [CrossRef]

33. Merkert, S.; Schubert, M.; Olmer, R.; Engels, L.; Radetzki, S.; Veltman, M.; Scholte, B.J.; Zöllner, J.; Pedemonte, N.; Galietta, L.J.V.; et al. High-Throughput Screening for Modulators of CFTR Activity Based on Genetically Engineered Cystic Fibrosis Disease-Specific iPSCs. *Stem Cell Rep.* **2019**, *12*, 1389–1403. [CrossRef]

34. Van Goor, F.; Hadida, S.; Grootenhuis, P.D.J.; Burton, B.; Stack, J.H.; Straley, K.S.; Decker, C.J.; Miller, M.; McCartney, J.; Olson, E.R.; et al. Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 18843–18848. [CrossRef]

35. Phuan, P.W.; Veit, G.; Tan, J.A.; Finkbeiner, W.E.; Lukacs, G.L.; Verkman, A.S. Potentiators of defective DF508-CFTR gating that do not interfere with corrector action. *Mol. Pharmacol.* **2015**, *88*, 791–799. [CrossRef] [PubMed]

36. Carlile, G.W.; Robert, R.; Goepp, J.; Matthes, E.; Liao, J.; Kus, B.; Macknight, S.D.; Rotin, D.; Hanrahan, J.W.; Thomas, D.Y. Ibuprofen rescues mutant cystic fibrosis transmembrane conductance regulator trafficking. *J. Cyst. Fibros.* **2015**, *14*, 16–25. [CrossRef] [PubMed]

37. Liang, F.; Shang, H.; Jordan, N.J.; Wong, E.; Mercadante, D.; Saltz, J.; Mahiou, J.; Bihler, H.J.; Mense, M. High-Throughput Screening for Readthrough Modulators of CFTR PTC Mutations. *SLAS Technol.* **2017**, *22*, 315–324. [CrossRef]

38. Giuliano, K.A.; Wachi, S.; Drew, L.; Dukovski, D.; Green, O.; Bastos, C.; Cullen, M.D.; Hauck, S.; Tait, B.D.; Munoz, B.; et al. Use of a High-Throughput Phenotypic Screening Strategy to Identify Amplifiers, a Novel Pharmacological Class of Small Molecules That Exhibit Functional Synergy with Potentiators and Correctors. *SLAS Discov.* **2018**, *23*, 392–399. [CrossRef] [PubMed]

39. Van der Plas, S.E.; Kelgtermans, H.; de Munck, T.; Martina, S.L.X.; Dropsit, S.; Quinton, E.; de Blieck, A.; Joannesse, C.; Tomaskovic, L.; Jans, M.; et al. Discovery of N-(3-Carbamoyl-5,5,7,7-tetramethyl-5,7-dihydro-4H-thieno[2,3-c]pyran-2-yl)-lH-pyrazole-5-carboxamide (GLPG1837), a Novel Potentiator Which Can Open Class III Mutant Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Channels to a High Extent. *J. Med. Chem.* **2018**, *61*, 1425–1435. [CrossRef] [PubMed]

40. Veit, G.; Xu, H.; Dreano, E.; Avramescu, R.G.; Bagdany, M.; Beitel, L.K.; Roldan, A.; Hancock, M.A.; Lay, C.; Li, W.; et al. Structure-guided combination therapy to potently improve the function of mutant CFTRs. *Nat. Med.* **2018**, *24*, 1732–1742. [CrossRef] [PubMed]

41. Wang, X.; Liu, B.; Searle, X.; Yeung, C.; Bogdan, A.; Greszler, S.; Singh, A.; Fan, Y.; Swensen, A.M.; Vortherms, T.; et al. Discovery of 4-[(2R,4R)-4-({[1-(2,2-Difluoro-1,3-benzodioxol-5-yl)cyclopropyl]carbonyl}amino)-7-(difluoromethoxy)-3,4-dihydro-2H-chromen-2-yl]benzoic Acid (ABBV/GLPG-2222), a Potent Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Corrector for the Treatment of Cystic Fibrosis. *J. Med. Chem.* **2018**, *61*, 1436–1449. [CrossRef]

42. Welcome to CandActCFTR. Available online: https://candactcftr.ams.med.uni-goettingen.de/ (accessed on 26 January 2021).

43. Nietert, M.M.; Vinhoven, L.; Auer, F.; Hafkemeyer, S.; Stanke, F. Comprehensive Analysis of Chemical Structures That Have Been Tested as CFTR Activating Substances in a Publicly Available Database CandActCFTR. *Front. Pharmacol.* **2021**, *12*, 689205. [CrossRef]

44. CF-Map. Available online: https://cf-map.uni-goettingen.de/ (accessed on 29 June 2022).

45. Vinhoven, L.; Stanke, F.; Hafkemeyer, S.; Nietert, M.M. CFTR Lifecycle Map—A Systems Medicine Model of CFTR Maturation to Predict Possible Active Compound Combinations. *Int. J. Mol. Sci.* **2021**, *22*, 7590. [CrossRef]

46. Xu, X.; Huang, M.; Zou, X. Docking-based inverse virtual screening: Methods, applications, and challenges. *Biophys. Reports* **2018**, *4*, 1–16. [CrossRef]

47. Huang, H.; Zhang, G.; Zhou, Y.; Lin, C.; Chen, S.; Lin, Y.; Mai, S.; Huang, Z. Reverse screening methods to search for the protein targets of chemopreventive compounds. *Front. Chem.* **2018**, *6*, 138. [CrossRef] [PubMed]

48. Lim, T.G.; Lee, S.Y.; Huang, Z.; Lim, D.Y.; Chen, H.; Jung, S.K.; Bode, A.M.; Lee, K.W.; Dong, Z. Curcumin suppresses proliferation of colon cancer cells by targeting CDK2. *Cancer Prev. Res.* **2014**, *7*, 466–474. [CrossRef] [PubMed]

49. Buendia-Atencio, C.; Pieffet, G.P.; Montoya-Vargas, S.; Martínez Bernal, J.A.; Rangel, H.R.; Muñoz, A.L.; Losada-Barragán, M.; Segura, N.A.; Torres, O.A.; Bello, F.; et al. Inverse Molecular Docking Study of NS3-Helicase and NS5-RNA Polymerase of Zika Virus as Possible Therapeutic Targets of Ligands Derived from Marcetia taxifolia and Its Implications to Dengue Virus. *ACS Omega* **2021**, *6*, 6134–6143. [CrossRef] [PubMed]

50. Ban, F.; Hu, L.; Zhou, X.H.; Zhao, Y.; Mo, H.; Li, H.; Zhou, W. Inverse molecular docking reveals a novel function of thymol: Inhibition of fat deposition induced by high-dose glucose in Caenorhabditis elegans. *Food Sci. Nutr.* **2021**, *9*, 4243–4253. [CrossRef] [PubMed]

51. Lauro, G.; Romano, A.; Riccio, R.; Bifulco, G. Inverse virtual screening of antitumor targets: Pilot study on a small database of natural bioactive compounds. *J. Nat. Prod.* **2011**, *74*, 1401–1407. [CrossRef]

52. RCSB Research Collaboratory for Structural Bioinformatics (RCSB).

53. AlphaFold Protein Structure Database. Available online: https://alphafold.ebi.ac.uk/ (accessed on 3 June 2022).

54. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

55. Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, D439–D444. [CrossRef]

56. Koes, D.R.; Baumgartner, M.P.; Camacho, C.J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893–1904. [CrossRef]

57. Hassan, N.M.; Alhossary, A.A.; Mu, Y.; Kwoh, C.K. Protein-Ligand Blind Docking Using QuickVina-W with Inter-Process Spatio-Temporal Integration. *Sci. Rep.* **2017**, *7*, 1–13. [CrossRef]

58. Kim, S.S.; Aprahamian, M.L.; Lindert, S. Improving inverse docking target identification with Z-score selection. *Chem. Biol. Drug Des.* **2019**, *93*, 1105–1116. [CrossRef]

59. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045–D1053. [CrossRef] [PubMed]

60. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J.P. ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* **2015**, *43*, W612–W620. [CrossRef] [PubMed]

61. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; de Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [CrossRef]

62. CHEMBL Database Release 30. 2022. Available online: http://chembl.blogspot.com/2022/03/chembl-30-released.html (accessed on 9 September 2022).

63. Fiedorczuk, K.; Chen, J. Mechanism of CFTR correction by type I folding correctors. *Cell* **2022**, *185*, 158.e11–168.e11. [CrossRef] [PubMed]

64. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31*, 455–461. [CrossRef] [PubMed]

65. Quiroga, R.; Villarreal, M.A. Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLoS ONE* **2016**, *11*, e0155183. [CrossRef] [PubMed]

66. Hanson, S.M.; Georghiou, G.; Thakur, M.K.; Miller, W.T.; Rest, J.S.; Chodera, J.D.; Seeliger, M.A. What Makes a Kinase Promiscuous for Inhibitors? *Cell Chem. Biol.* **2019**, *26*, 390.e5–399.e5. [CrossRef]

67. Cerisier, N.; Petitjean, M.; Regad, L.; Bayard, Q.; Réau, M.; Badel, A.; Camproux, A.C. High impact: The role of promiscuous binding sites in polypharmacology. *Molecules* **2019**, *24*, 2529. [CrossRef]

68. Ameen, N.; Silvis, M.; Bradbury, N.A. Endocytic trafficking of CFTR in health and disease. *J. Cyst. Fibros.* **2007**, *6*, 1–14. [CrossRef]

69. Farinha, C.M.; Matos, P. Rab GTPases regulate the trafficking of channels and transporters—A focus on cystic fibrosis. *Small GTPases* **2018**, *9*, 136–144. [CrossRef]

70. Hou, X.; Wu, Q.; Rajagopalan, C.; Zhang, C.; Bouhamdan, M.; Wei, H.; Chen, X.; Zaman, K.; Li, C.; Sun, X.; et al. CK19 stabilizes CFTR at the cell surface by limiting its endocytic pathway degradation. *FASEB J.* **2019**, *33*, 12602–12615. [CrossRef]

71. Muthyala, R.S.; Ju, Y.H.; Sheng, S.; Williams, L.D.; Doerge, D.R.; Katzenellenbogen, B.S.; Helferich, W.G.; Katzenellenbogen, J.A. Equol, a natural estrogenic metabolite from soy isoflavones: Convenient preparation and resolution of R- and S-equols and their differing binding and biological activity through estrogen receptors alpha and beta. *Bioorganic Med. Chem.* **2004**, *12*, 1559–1567. [CrossRef] [PubMed]

72. Pyle, L.C.; Fulton, J.C.; Sloane, P.A.; Backer, K.; Mazur, M.; Prasain, J.; Barnes, S.; Clancy, J.P.; Rowe, S.M. Activation of the cystic fibrosis transmembrane conductance regulator by the flavonoid quercetin: Potential use as a biomarker of ΔF508 cystic fibrosis transmembrane conductance regulator rescue. *Am. J. Respir. Cell Mol. Biol.* **2010**, *43*, 607–616. [CrossRef] [PubMed]

73. Younger, J.M.; Chen, L.; Ren, H.-Y.; Rosser, M.F.N.; Turnbull, E.L.; Fan, C.-Y.; Patterson, C.; Cyr, D.M. Sequential quality-control checkpoints triage misfolded cystic fibrosis transmembrane conductance regulator. *Cell* **2006**, *126*, 571–582. [CrossRef]

74. Grove, D.E.; Fan, C.-Y.; Ren, H.Y.; Cyr, D.M. The endoplasmic reticulum-associated Hsp40 DNAJB12 and Hsc70 cooperate to facilitate RMA1 E3-dependent degradation of nascent CFTRDeltaF508. *Mol. Biol. Cell* **2011**, *22*, 301–314. [CrossRef] [PubMed]

75. Berthold, M.R.; Cebron, N.; Dill, F.; Gabriel, T.R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME—The Konstanz information miner. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 26–31. [CrossRef]

76. Morris, G.M.; Ruth, H.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [CrossRef]

77. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]

78. Open Babel. Available online: http://openbabel.org/wiki/Main_Page (accessed on 20 May 2022).

79. Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.F.; Fischer, P.D.; Coote, P.W.; Padmanabha Das, K.M.; Malets, Y.S.; Radchenko, D.S.; Moroz, Y.S.; Scott, D.A.; et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **2020**, *580*, 663–668. [CrossRef]

80. SLURM. Available online: https://slurm.schedmd.com (accessed on 9 September 2022).

81. Data Analytics Platform: Open Source Software Tools | KNIME. Available online: https://www.knime.com/knime-analytics-platform (accessed on 31 May 2022).

82. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36. [CrossRef]

83. Southan, C. InChI in the wild: An assessment of InChIKey searching in Google. *J. Cheminform.* **2013**, *5*, 10. [CrossRef] [PubMed]
84. Morgan, H.L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113. [CrossRef]
85. RDKit. Available online: https://www.rdkit.org/ (accessed on 2 June 2022).