CARBONIC ANHYDRASE INHIBITORS: IDENTIFYING THERAPEUTIC CANCER AGENTS THROUGH VIRTUAL SCREENING

Giulio Poli,[a] Claudiu T. Supuran[b,] Tiziano Tuccinardi,[a,*]

[a] Department of Pharmacy, University of Pisa, 56126 Pisa, Italy

[b] Neurofarba Department, Universita degli Studi di Firenze, Florence, Italy

Virtual screening in drug discovery

Computer-aided drug design (CADD) strategies are nowadays widely employed in different areas of medicinal chemistry along the different steps of the drug development process, from hit identification to lead optimization campaigns. The expensive and time consuming trial-and-error approach can be often replaced by a rational drug design approach, which can benefit the guide and support of molecular modelling and computational studies. Hit identification is undoubtedly the drug design stage to which *in silico* techniques can offer the most valuable help, thanks to the application of virtual screening (VS) protocols enabling the discovery of novel and structurally diverse active ligands from libraries of commercially available compounds.

VS strategies can be divided in receptor-based and ligand-based approaches. Receptor-based VS methods are the most used and profitable strategies for *in silico* hit identification, since they exploit structural knowledge about the target protein.[1] The availability of at least an X-ray crystallographic or NMR structure of the target receptor is thus an essential prerequisite for the development of receptor-based VS protocols, although homology modelling can sometimes supply with reliable receptor models when no structural information about the target of interest is available.[2] When possible, chemical knowledge of one or more reference ligands with experimentally confirmed activity towards the target of interest can also be employed in receptor-based approaches. In the ideal situation, a ligand-protein co-crystal structure providing both types of information and displaying the bioactive conformation of the ligand bound to the target receptor can be used as a reference for receptor-based methods, as in the case of receptor-based pharmacophore modeling. By following this

approach, the fundamental ligand-protein interactions observed in the X-ray complex, known to be the main responsible for the activity of bound ligand and related compounds, can be used to generate one or multiple pharmacophore models to be employed in pharmacophore-based VS studies. In these studies, libraries of up to millions of commercial compounds can be filtered with the aim of identifying molecules that can reproduce the pattern of fundamental interactions represented by the pharmacophore model and are thus supposed to show affinity for the target receptor.[3] In the recent years, receptor-based pharmacophore screenings have been widely applied and proved to be efficient hit finding strategies for several different target receptors including *h*CAs.[4–11]

The principal and most popular computational approach belonging to receptor-based strategies is however represented by docking, which constitutes the gold standard technique for predicting the potential binding mode of small-molecule ligands within a target receptor. Docking studies are thus widely used in medicinal chemistry for evaluating the most energetically favored disposition of experimentally active compounds within their protein targets[12] and for rationalizing structure-activity relationships (SAR) among series of related compounds.[13] Nevertheless, docking approaches proved to be a very profitable computational tool also in VS campaigns, allowing the discovery of novel compounds active towards many different types of protein targets and expanding the chemical space of their known ligands.[14] However, docking reliability often raised some criticism.[15] In fact, the performance of docking procedures should be generally tested through self-docking (at least) or possibly with cross-docking studies before applying them for VS workflows, while the use of special parametrizations, force-fields or post-docking procedures seem to be valuable strategies for improving docking accuracy when dealing with metalloenzymes such as *h*CAs.[16–18] Anyway, classic high-throughput molecular docking demonstrated to be a useful VS strategy for identifying novel *h*CAs inhibitors even when used alone, without employing other structure-based or ligand-based approaches.[19–21]

Molecular dynamics (MD) simulations constitute another receptor-based technique that can be employed in VS workflows. For instance, MD simulations can be performed to refine the target

protein structures to be used for docking and other receptor-based methods, relaxing the side chains of the protein residues and eliminating possible steric clashes.[21] However, MD studies can be employed within VS workflows to perform a thorough evaluation of single ligand binding modes predicted by docking for few top-scored hit compounds selected through the previous VS steps. The behavior of the corresponding ligand-protein complexes is studied by creating a solvated system with explicit water molecules where both the receptor and the bound ligand are provided with sufficient energy to move freely as in a cellular environment. Through such analysis, it is possible to evaluate the stability of the ligand disposition and the ligand-protein interactions predicted by docking for the selected compounds and to assess the reliability of their binding mode. MD simulations studies can thus be employed as a qualitative post-docking filter in the final step of VS workflows to discard those hit compounds whose key interactions with the protein are not properly conserved during the simulations.[22–24]

Ligand-based similarity strategies, uniquely based on the molecular structure and properties of known reference compounds experimentally confirmed as active ligands towards the target of interest, are typically the only possible option when no structural information about the target receptor is available. Ligand-based, and in particular 2D-similarity methods such as topological and fingerprint searches, showed to be useful approaches for identifying *h*CAs inhibitors,[25,26] probably because they are constituted by three main functional moieties: a) a zinc-binding group (ZBG, often represented by a sulfonamide group) coordinating the prosthetic zinc ion of the enzyme, b) a central core (often an aromatic ring) and c) a terminal tail that can show either lipophilic or hydrophilic character. Ligand-based techniques are often used in combination with structure-based approaches, especially in the initial steps of the VS workflow, since they are generally less time consuming and can be employed to rapidly filter large compound databases with the aim of discarding those ligands that are too structurally different with respect to the reference active ligands and thus less likely to show affinity for the target receptor.[27,28] Interestingly, ligand-based pharmacophores and substructure searches

proved not only to be valuable for selecting potential *h*CA ligands endowed with known ZBGs, but also to identify novel inhibitors with atypical zinc-chelating moieties.[2,5]

Virtual screening studies identifying novel carbonic anhydrase inhibitors

A remarkable example of how ligand-based and receptor-based strategies can be successfully combined into a hierarchic VS workflow that led to the discovery of new *h*CA ligands was reported by Klebe and collaborators in 2002.[29] The study was aimed at assessing the general potential of VS for lead discovery and *h*CA II was selected as a suitable target for this purpose. As a first step of the VS protocol, a dataset of about 99'000 drug-like compounds belonging from Maybridge and LeadQuest databases and respecting the Lipinski rules was created and subjected to a substructure filter, in order to retain only compounds bearing a series of functional moieties already described as ZBGs in other zinc proteases, including sulfonamide, amide, hydroxamic and carboxylic groups. In this way, only about 5900 potential zinc chelators were retained from the initial dataset. After this preliminary filter, the selected compounds were subjected to a receptor-based pharmacophore screening. The pharmacophore model was generated through an analysis of the *h*CA II binding site regions in which a potential ligand could form energetically favorable interactions. The mapping of the binding hot spots into *h*CA II catalytic site was performed using four different and complementary methods (i.e. LUDI, GRID, SuperStar and DrugScore), which essentially searched for favorable H-bond acceptor, H-bond donor and hydrophobic interactions and using 13 different ligand-protein X-ray reference structures. By considering the hot spots and contours maps obtained using the different computational methods, as well as the experimental disposition of the 13 sulfonamide ligands co-crystalized with *h*CA II, a receptor-based pharmacophore model including five features was eventually generated using UNITY software. The model included an H-bond donor and an H-bond acceptor feature representing the interactions of the ligand sulfonamide group with the key anchoring residue T199, a couple of hydrophobic features representing the central core of the ligands and a further H-bond acceptor group representing a possible interaction with the side chain of Q92. By

considering this latter feature and one of the hydrophobic features as optional, the pharmacophore model successfully retrieved 35 known *h*CA II ligands (the 13 crystallographic compounds and other 22 *h*CA II inhibitors selected from literature) that were used to enrich the subset of about 5900 potential zinc binders. However, more than a half of these compounds (about 3300) passed the pharmacophore-based filter and the retrieved ligands were thus subjected to a further ligand-based filter through the flexible superposition with FlexS software on two reference co-crystallized inhibitors selected for having a small and a big molecular shape. Out of the 2237 compounds for which a superposition could be computed by the software, the 100 top-scored ligands were then docked into the *h*CA II binding site with FlexX. A final set of 13 compounds was eventually selected after visual inspection of the docking poses, based on the pharmacophore features matched, the key ligand-protein interactions maintained and docking score calculated by both FlexX and DrugScore scoring functions. The final compounds were purchased and subjected to enzymatic assays, together with the reference inhibitor acetazolamide, and the results revealed an inhibitory activity for 11 out of the 13 tested compounds. In fact, only two hydroxamic acid derivatives showed no *h*CA II inhibition, while all other sulfonamide ligands presented an inhibitory potency ranging from low micromolar to subnanomolar values (pIC_{50} values between 5.26 and 9.21). Two of the active compounds, including the most potent one, were then co-crystallized with *h*CA II and the two X-ray structures (PDB codes 1KQW and 1KQR) further demonstrated the reliability of the VS workflow, since the two ligands respected the pharmacophore model in their experimental binding modes, which differed of less than 1.5 Å in terms of root mean squared deviation (RMSD) from the best docking poses selected by DrugScore.

Another example of receptor-based pharmacophore screening focused on *h*CA II was reported in 2011 by Supuran, Sechi and co-workers, which identified a novel low micromolar inhibitor incorporating an unusual ZBG.[5] In this work, high-resolution X-ray structures of *h*CA II in complex with sulfonamide inhibitors were aligned and used as input ligand-protein complexes for generating a receptor-based pharmacophore model using MOE software. The model obtained included four

different features: a metal ligator feature and two H-bond acceptor features, representing the sulfonamide group of the reference experimental ligands, and a big hydrophobic/aromatic feature representing both their central aromatic scaffold and the terminal moieties constituting the ligands' tails. Moreover, excluded volume constraints were generated using the binding site of *h*CA II in the X-ray structure in complex with 3-[4-(aminosulfonyl)phenyl]propanoic acid (PDB code 2NNO). The pharmacophore model was used to screen the ZINC lead-like database,[30] collecting about 970'000 drug-like compounds, from which about 37'400 ligands were retained as hits. The subset of selected compounds was then further filtered by excluding all molecules bearing sulphur atoms, with the aim of excluding from the following VS step all sulfonamide ligands and compounds bearing sulfonamide-like groups, so that to retain only potential *h*CA II with different ZBGs. The about 4600 retained compounds were then docked into the structure of *h*CA II using FlexX and only the best 29 compounds were subjected to a further docking evaluation using Autodock4. Among these final compounds, a small molecule with a unique original structure was selected to be purchased and subjected to enzymatic assays. The results showed that the compound was able to inhibit *h*CA II with low micromolar potency (K_i = 9.0 µM) and demonstrated selectivity over *h*CA I (K_i = 410.0 µM). Despite the activity of the identified ligand was not high, the compound was characterized by an unusual ZBG, represented by a gem-dihydroxil-keto moiety that was supposed to coordinated the zinc ion and to form H-bonds with the key anchoring residues T199 and T200 of *h*CA II (Figure X). The presence of a trifluoromethyl group connected to the germinal diol was also supposed to enhance the polarization of the two hydroxyl groups, thus favoring metal chelation.

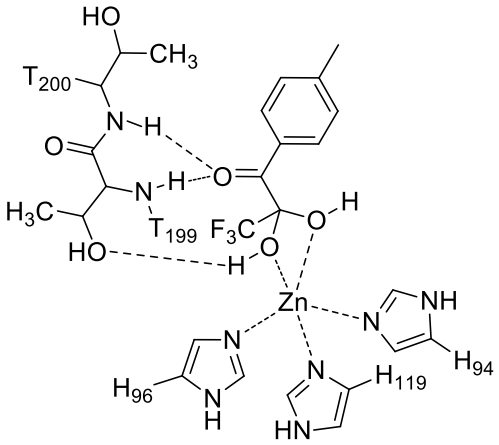
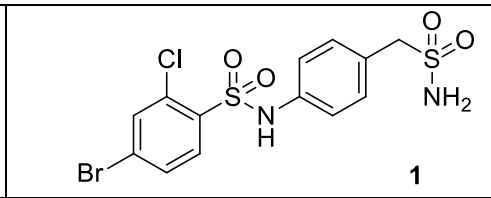
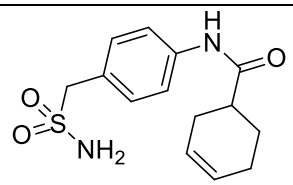


Figure 1. Schematized binding mode of the VS hit identified by Supuran and Sechi.

Receptor-based pharmacophore modelling and screening were also successfully applied for the identification of novel *h*CA VII inhibitors by De Luca and co-workers, which discovered two novel nanomolar sulphonamide inhibitors.[6] The pharmacophore model was generated based on the X-ray structures of *h*CA VII in complex with the well-known inhibitors acetazolamide (PDB code 3ML5) and ethoxzolamide (PDB code 3MDZ). The two aligned co-crystal structures were used as input structures for LigandScout software to generate two corresponding receptor-based pharmacophore models, which were then merged into a single final model including six total features: three H-bond acceptor, one H-bond donor and two hydrophobic features. The H-bond donor and a closely placed H-bond acceptor represented the sulfonamide groups of the co-crystallized ligands, a second H-bond acceptor represented the interaction with the side chain of T200 shared by both inhibitors through their endocyclic nitrogen, while the other three features corresponded to moieties present in either one or the other inhibitor. The merged model was validated using a set of 22 *h*CA inhibitors reported in literature and used to filter a focused library of about 6300 compounds bearing sulfonamide moieties retrieved from the ZINC database. The screening was performed with the software Catalyst, which assigned a fit score valued to reference and database compounds based on the geometric matching between their structures and the pharmacophore features of the model. Since 13 out of the 22 reference actives showed a fit score value higher than 3, this threshold was employed to filter the

database compounds, thus obtaining 299 best fitting compounds. After a visual inspection of the pharmacophore hits, 34 ligands selected based on their structural diversity were then docked into the X-ray structure of *h*CA VII in complex with acetazolamide (PDB code 3ML5) using GOLD software. Considering the fitness score calculated by GoldScore scoring function for the docked compounds, the ligand-protein interactions predicted in their docking poses and the commercial availability of the molecules, two final compounds were purchased and subjected to enzymatic assays. The ligands showed inhibitory activities against *h*CA VII in the nanomolar range (K_i values of 62.9 and 39.4 nM, see Table 1), thus demonstrating the reliability of the VS workflow. Nevertheless, the ligands also showed comparable or even higher potency against other *h*CA isoforms, particularly *h*CA I (K_i values of 8.9 and 8.6 μM), as the ligands potency for this isoform was higher than that reported for the reference inhibitors.

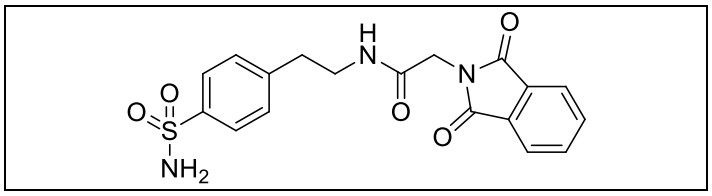
Table 1. K_i values against *h*CAs showed by compounds 1 and 2

	 **1**	 **2**			
	K_i (nM)				
	*h*CA I	*h*CA II	*h*CA VII	*h*CA IX	*h*CA XIV
1	8.6	6.3	62.9	66.0	19.4
2	8.9	73.2	39.4	53.2	7.6

A ligand-based pharmacophore screening, in combination with other techniques, was instead used in a VS study reported in 2009 by Thiry and co-workers, aimed at identifying new *h*CA IX sulfonamide inhibitors.[2] The model was generated using the MOE software, on the basis of the structures of 10 known inhibitors reported in literature, with strong activity against *h*CA IX and selectivity over *h*CA II. Although being structurally different, the 10 active ligands all shared a sulfonamide moiety as ZBG; therefore, the compounds were flexibly aligned one another so that the coordinates of their sulfonamide groups were properly superimposed. Based on the aligned reference ligands, a ligand-

based pharmacophore model including six different features was obtained: the sulfonamide fragment was represented by a metal ligator and two H-bond acceptor features, while the central aromatic core and the terminal tails of the active ligands were represented, respectively, by an aromatic feature and two mixed features (hydrophobic/H-bond donor and H-bond donor/acceptor). Once generated the model, a dataset of about 1200 compounds was obtained by applying molecular properties and substructure filters to about 4.6 million commercial compounds belonging to the ZINC database. The compounds retained in the dataset were characterized by the presence of a sulfonamide ZBG group and satisfied the following molecular properties: -2 < logP < +4; rotatable bonds < 12; H-bond donor < 5, Hbond acceptor < 10; polar surface area < 140. The selected compounds were then screened using the pharmacophore model and only the 500 hits respecting all pharmacophore features were subjected to molecular docking using the software GOLD. In this case, the 500 compounds were docked into a *h*CA IX homology model previously developed by the same research group.[31] After visual inspection of the top-score 100 compounds, six potential new inhibitors were eventually selected for enzymatic assays and all ligands showed *h*CA IX inhibitory activities, with K_i values between 2.75 and 0.29 µM. Furthermore, the compounds demonstrated selectivity over *h*CA I (being 3- to 28-fold less active against this *h*CA isoform) but showed comparable or higher activities against *h*CA II. For instance, the most interesting compound (compound **3** of Table 2), with the highest activity for *h*CA IX and selectivity over *h*CA I was also found to be 60-fold more potent against *h*CA II, with a K_i value of 5 nM. These results demonstrated the reliability of the VS workflow in the discovery of new *h*CA IX ligands but confirmed that the identification of *h*CA ligands also endowed with good a selectivity profile represents a more challenging task.

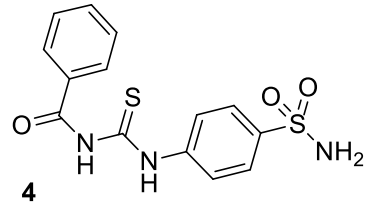
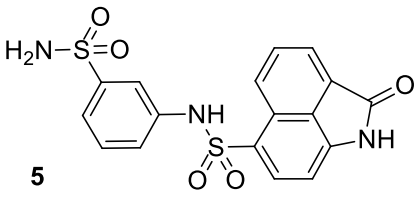
Table 2. K_i values against *h*CAs showed by compounds **3**.



K_i (nM)		
*h*CA I	*h*CA II	*h*CA VII
8350	5	300

A further VS campaigns focused on *h*CA IX allowed the identification of thirteen novel sulfonamide inhibitors with nanomolar potency.[19] In this study, a pure docking-based approach was followed: the whole SPECS database, collecting about 280'000 commercial compounds, was docked into the X-ray structure of *h*CA IX in complex with acetazolamide (PDB code 3IAI) by using Glide software with the standard-precision method (SP). Subsequently, the top-scored 10'000 compounds were subjected to an additional docking evaluation into the same receptor performed with the extra-precision method (XP). The 500 compounds showing the highest score values after this second docking step were visually inspected and 49 structurally different ligands were eventually selected to be purchased and tested to evaluate their inhibitory activity. Although the final set of selected compounds included molecules bearing different types of ZBGs, only sulfonamide ligands showed activity against *h*CA IX. However, most of the 13 hit compounds showed single- to double-digit nanomolar potency and considerable selectivity over *h*CA II and/or *h*CA I. For instance, compound **4** of Table 3 was probably the most interesting, with an IC_{50} value for *h*CA I inhibition of 2.86 nM, a 21-fold selectivity over *h*CA II and 69-fold selectivity over *h*CA I, whereas compound **5** (IC_{50} for *h*CA I = 28.35) showed the highest selectivity over *h*CA II (about 137-fold).

Table 3. IC_{50} values against *h*CAs showed by compounds **4** and **5**.

	 **4**	 **5**	
	IC_{50} (nM)		
	*h*CA I	*h*CA II	*h*CA VII
4	197.64	60.93	2.86
5	231.41	3873	28.35

A similar example of *h*CA IX-targeted VS study entirely relying on high-throughput docking was recently reported by Durdagi and collaborators in 2016.[21] In this work, the X-ray co-crystal structure of *h*CA IX in complex with acetazolamide (PDB code 3IAI) was used. Prior to docking studies, the reference ligand-protein complex was subjected to a 10 ns molecular dynamics (MD) simulation performed with NAMD in an explicit water environment, in order to relax the system and the protein side chains. The average protein structure derived from the MD simulation was used for the docking evaluation of 7 million drug-like commercial compounds from ZINC database, performed with the high-throughput virtual screening (HTVS) method of Glide software. Out of the whole database of docked compounds, only the top-scored 70 molecules were selected and subjected to a more thorough docking calculation employing Glide XP method. Out of the 19 compounds showing a docking score better than -8 kcal/mol and forming H-bond interactions with key anchoring residues of *h*CA IX binding site, such as T199 and T200, three compounds were chosen, purchased and tested for their *h*CA IX inhibitory activity. Surprisingly, none of the top-scored compounds presented a sulfonamide moiety as ZBG. The three tested compounds, all characterized by a central α-hydroxylactam core, showed *h*CA IX inhibitory activity in the low micromolar to submicromolar range, with K_i values between 1.58 and 0.85 μM. Interestingly, none of them was predicted to directly interact with the zinc ion in their proposed binding mode.

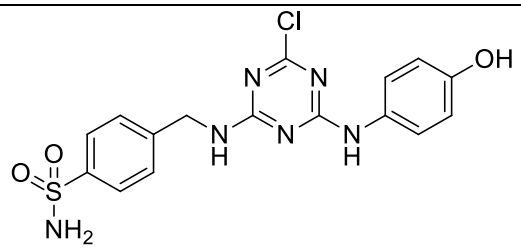
Pure molecular docking screens were also successfully employed for the identification of novel *h*CAs inhibitors from virtual focused libraries of natural compounds or structurally related derivatives designed through combinatorial chemistry. Alcaro and collaborators performed a docking-based virtual screening on a pool of natural compounds belonging from both the volatile and non-volatile fractions of the typical Calabrian products Bergamot and Tropea red onion.[20] Considering different possible tautomeric and protomeric forms of the specific natural compounds, a library of 280 molecules was generated and the standard-precision method of Glide was used to dock the compounds into the X-ray structures of five different CA isoforms, i.e. *h*CAs I, II, IX and XII (PDB

codes 1AZM, 4CQ0, 3IAI and 4HT2 , respectively), as well as *m*CA VA (PDB code 1DMY). Self-docking studies of the ligands co-crystallized with the different CA isoforms in the selected X-ray structures were used to assess the reliability of the docking protocol; moreover, the estimated binding energies associates to the docking results of acetazolamide into the different X-ray structures were used as reference values for selecting potential *h*CA ligands to be tested for inhibitory activity. Precisely, the ten flavonoid ligands that showed a docking score value better that that obtained for acetazolamide in at least one of the five CA isoforms were selected and subjected to enzymatic assays. The experimental results demonstrated that all tested compounds were endowed with inhibitory activity against all human isoforms of the five CAs with low micromolar to nanomolar potencies (K_i values ranging from 4.24 to 0.11 μM). Two of the identified hits, namely eritrocitin and apigenin, showed very interesting activities against *h*CA VA, since they were found to be more potent *h*CA VA inhibitors (K_i for *h*CA VA of 0.15 and 0.30 μM, respectively) than the reference ligand acetazolaminde (K_i for *h*CA VA of 0.38 μM).

Finally, in 2018 Supuran and collaborators reported on the identification of novel nanomolar inhibitors of *h*CA I, II and IX through a VS study performed on a focused library of sulfonamide compounds with a 1,3,5-triazine core,[32] as a continuation of previous works in which they discovered that sulfonamides bound to a triazine skeleton had high potency and usually also specificity for *h*CA IX over *h*CA I or II.[33,34] The software CombiGlide from Schrodinger suite was used to create a combinatorial library comprising 2200 unique compounds (9766 molecules considering multiple possible isomers and ionization states for each ligand) by combining three different arylsulfonamide moieties and a set of 21 fragments (including aminoalcohols, aryltriazoles and others) with cyanuric chloride (2,4,6-trichloro-1,3,5-triazine). The library was built by considering the generations of the possible triazine-containing products obtained by substituting cyanuric chloride with an arylsulfonamide group and at least one among the other 21 fragments. The whole set of molecules was docked into the X-ray structure of *h*CA IX (PDB code 3IAI) using Glide HTVS method; the top-scored 2000 compounds were then docked into the same protein structure using the SP method and

the best 400 ligands according to this second step were subjected to further docking calculation using the XP method. Finally, the 80 top-scored molecules based on the XP score (corresponding to 66 unique ligands) were analysed through an additional docking evaluation using the quantum-polarized ligand docking (QPLD) protocol, which computes ligands partial charges with the semiempirical RM6 method. The same method was used to dock the 66 ligands into the X-ray structure of *h*CA II (PDB code 3MMF) in order to evaluate their potential selectivity over this *h*CA isoform. Among the 20 top-scored molecules resulting from QPLD calculations, eleven compounds and their synthetic intermediates (for a total of 24 ligands) were selected to be synthesized and tested for *h*CA I, II and IX inhibitory activity. A final set of newly synthesized compounds was thus subjected to enzymatic assays, revealing nanomolar *h*CA IX inhibitory activity for all compounds, among which 14 ligands showed a high potency (K_i < 50 nM). All tested compounds were generally selective for *h*CA IX over *h*CA I, but most of them showed either comparable or higher potency against *h*CA II. The best selectivity over *h*CA II was shown by compound **6** of Table 4, which represented the most promising compound of the series, presenting subnanomolar activity against *h*CA IX (K_i = 0.4 nM) and being 18.5-fold and about 42-fold less active against *h*CA II and I, respectively.

Table 4. K_i values against *h*CAs showed by compounds **6**.



K_i (nM)		
*h*CA I	*h*CA II	*h*CA IX
16.7	7.4	0.4

Retrospective virtual screening studies using carbonic anhydrases as target receptors

Due to the number of known inhibitors and ligand-protein X-ray structures reported in literature, some well-studied *h*CA isoforms such as *h*CA I and *h*CA II has been used as reference target proteins for retrospective analyses aimed at assessing the hit identification performance of several different VS approaches, including docking, 2D fingerprint screenings, 3D shape-based ligand similarity and machine learning techniques.

In an interesting work reported in 2005, Shoichet and collaborators evaluated the performance of their molecular docking software Dock in identifying potential ligands of metalloenzymes using a standard noncovalent scoring function, i.e. employing molecular mechanics parameters for treating the prosthetic metal ion but without considering the covalent-like interaction between metal and ligands.[35] The analysis was performed using a dataset of about 95'000 compounds belonging from the MDL Drug Data Report (MDDR) version 2000.2, a licensable database of biologically relevant compounds from patent literature. The study considered reference metalloenzymes with a minimum of 20 known ligands included in the database and for which various ligand-protein co-crystal structures were available, i.e. xanthine oxidase (XO), neutral endopeptidase (NEP), peptide deformylase (PDF), matrix metalloproteinase 3 (MMP-3) and *h*CA II. In particular, the retrospective analysis focused on *h*CA II was carried out using the X-ray structure of *h*CA II bound to dorzolamide (PDB code 1CIL) and considering the 241 *h*CA II ligands included in the MDDR database as reference actives. The zinc ion was parametrized with van der Waals radius of 1.09 Å and a well-depth minimum of 0.25 kcal/mol[36] as well as a net charge of +1.4 (considering a net charge transfer of +0.2 to each of the three coordinating histidine residues). The enrichment plot analysis performed on the docking results showed that, using these parameters, a satisfying screening performance could be obtained. In fact, the enrichment factor (EF) calculated for the top-scoring 0.1% of the ranked database ($EF_{1\%}$) was 82-fold better than random selection and 25% of known actives were identified within the top 1.6% of the ranked database. Moreover, 95% of the reference *h*CA II ligands showed docking poses resembling the experimental disposition of dorzolamide.

McGaughey and co-workers performed an extensive analysis in which they assessed the VS performance of three different docking software (Flog, Fred and Glide) in comparison with both 2D ligand-based (Daylight, Toposim) and 3D shape-based (SQW, Rocs) similarity strategies.[25] The seven total VS approaches were tested using a set of eleven different target enzymes that also included *h*CA I. In this study, the X-ray structure of the enzyme in complex with acetazolamide (PDB code 1AZM) was used for docking studies and the co-crystallized ligand was employed as the query structure for ligand-based methods. Two screening databases were used: a dataset of about 24'500 compounds obtained by clustering the MDDR database and a dataset of less than 10'000 elements including compounds randomly selected from Merck's corporate database (MCIDB) and a set of reference active compounds carefully selected for each target. In both datasets, compounds with more than 80 heavy atoms were discarded. For *h*CA I, 80 and 241 reference actives were present in the final MDDR and MCIDB datasets, respectively. This analysis revealed that, on average, the ligand-based VS approaches outperformed the docking methods, as demonstrated by comparing the EF values calculated for the top 1% of both ranked databases. This was particularly true for *h*CA I, which resulted to be one of the most challenging targets for docking-based VS among the 11 enzymes tested in the study, showing $EF_{1\%}$ values between 0 and 2.5, compared to mean values ranging from 4.0 to 13.4 (calculated considering all target enzymes). The fact that no special parametrization for the zinc ion was used in docking calculations may have contributed to the low enrichments obtained for *h*CA I, although similar results were showed by other targets with no metal prosthetic groups within the ligand binding site (namely HIV-rt and COX2). Conversely, the 2D ligand similarity methods performed the best on *h*CA I, showing the highest EF values among those obtained for all targets. Particularly, $EF_{1\%}$ values of 50.2-56.4 were obtained for MDDR dataset and 17.6-25.5 for MCIDB dataset compared to mean values of 24.5-29.0 and 7.4-10.6 respectively. These results can be however rationalized considering that most of the *h*CA inhibitors present a sulfonamide moiety acting as ZBG, which can facilitate the retrieval of active compounds using VS approaches based on the structural similarity of the ligands. In fact, pure shape-based methods such as Rocs showed a lower performance

compared to 2D similarity, but when an atom type-based similarity component was included in the screening (as in Rocs-*color*) an about 3-fold improvement in the EF values were obtained for *h*CA I. An in-depth analysis focused on 2D fingerprint methods was reported by Sherman et al. in 2010.[26] In this work, the VS performance of eight different fingerprint types available in the software Canvas (linear, dendritic, radial, pairwise, triplet, torsion, Molprint2D and MACCS fingerprints) was evaluated using a dataset of about 24'500 MDDR compounds obtained as described by McGaughey and co-workers[25] and considering the same group of eleven target enzymes, including *h*CA I. The same reference ligand-protein co-crystal structures were also employed for each target. Therefore, acetazolamide was used as the query structure for the fingerprint similarity analyses aimed at retrieving the 80 reference actives of *h*CA I out of the MDDR compounds dataset. Since for each fingerprint type the performance evaluation was carried out using different atom typing schemes, bit scaling rules and similarity/distance metrics, a total number of almost 160'000 different parameter combinations were evaluated based on the $EF_{1\%}$. The analysis demonstrated the robustness of Molprint2D method, which showed the best overall performance across all 11 target enzymes considering both default settings and the best combination of parameters tested. Again, *h*CA I was found to be the target protein for which the best enrichments were obtained, on average. In particular, considering the best settings for each fingerprint type, Molprint2D and radial fingerprints achieved an $EF_{1\%}$ value of 80.0 and even MACCS fingerprints, which produced the worst overall results, showed an $EF_{1\%}$ of 50.0 for *h*CA I. These results are in agreement with the above reported considerations about the structural similarity of *h*CAIs; in fact, although the average tanimoto similarity of the 80 reference *h*CA I ligands was considerably low (around 0.1), the authors themselves pointed out that all ligands presented a sulfonamide group in a similar environment, and this may have contributed to the generally high performance of the fingerprints methods. Notably, atom typing schemes such as Mol2 and Daylight performed better than less specific atom typing (less likely to discriminate specific structural moieties).

The same author reported a further analysis concerning ligand-based techniques performed using the same set of target enzymes, reference actives, query structures and approximately the same dataset of MDDR compounds. This study was focused on shape-based flexible ligand superposition methods, as the Phase Shape tool implemented in Schrödinger suite was tested for its VS performance, analyzing the impact of multiple parameters such as conformer generation method and atom/feature typing method.[37] The results shown in this study were consistent with those reported by McGaughey et. al, which were also used to compare the performance of Phase Shape with Rocs-color and SQW. Overall, the $EF_{1\%}$ obtained with *h*CA I were above the average values calculated across all 11 target enzymes, but not the highest as observed for the 2D similarity methods, and the use of MacroModel atom types or better Phase pharmacophore feature types performed substantially better than more generic atom typing-based and shape-only scoring. However, the 3D alignments obtained with shape-only scoring method demonstrated to perform qualitatively well for *h*CA I target. Surprisinlgy, the use of a low-energy conformer of acetazolamide generated with the conformer generator implemented in Phase Shape as the query structure produced better results than those obtained using the experimental disposition of the ligand.

In summary, the results of these analyses highlighted that a proper parametrization of the prosthetic zinc ion in docking studies focused on *h*CAs can improve the quality and accuracy of the docking results and that ligand-based techniques, particularly 2D-similarity approaches, can show superior performance in identifying new potential *h*CA inhibitors, provided that either no or minimal structural diversity at the level of the ZBG is required or expected.

References

(1) Ripphausen, P.; Stumpfe, D.; Bajorath, J. Analysis of Structure-Based Virtual Screening Studies and Characterization of Identified Active Compounds. *Future Med. Chem.* **2012**, *4* (5), 603–613.

(2) Thiry, A.; Ledecq, M.; Cecchi, A.; Frederick, R.; Dogné, J.-M.; Supuran, C. T.; Wouters, J.; Masereel, B. Ligand-Based and Structure-Based Virtual Screening to Identify Carbonic Anhydrase IX Inhibitors. *Bioorg. Med. Chem.* **2009**, *17* (2), 553–557.

(3) Vuorinen, A.; Schuster, D. Methods for Generating and Applying Pharmacophore Models as Virtual Screening Filters and for Bioactivity Profiling. *Methods* **2015**, *71*, 113–134.

(4) Tanrikulu, Y.; Rau, O.; Schwarz, O.; Proschak, E.; Siems, K.; Müller-Kuhrt, L.; Schubert-Zsilavecz, M.; Schneider, G. Structure-Based Pharmacophore Screening for Natural-Product-Derived PPARγ Agonists. *ChemBioChem* **2009**, *10* (1), 75–78.

(5) Pala, N.; Dallocchio, R.; Dessì, A.; Brancale, A.; Carta, F.; Ihm, S.; Maresca, A.; Sechi, M.; Supuran, C. T. Virtual Screening-Driven Identification of Human Carbonic Anhydrase Inhibitors Incorporating an Original, New Pharmacophore. *Bioorg. Med. Chem. Lett.* **2011**, *21* (8), 2515–2520.

(6) De Luca, L.; Ferro, S.; Damiano, F. M.; Supuran, C. T.; Vullo, D.; Chimirri, A.; Gitto, R. Structure-Based Screening for the Discovery of New Carbonic Anhydrase VII Inhibitors. *Eur. J. Med. Chem.* **2014**, *71*, 105–111.

(7) Wang, Y.; Zhao, Y.; Sun, R.; Kong, W.; Wang, B.; Yang, G.; Li, Y. Discovery of Novel Antagonists of Glycoprotein IIb/IIIa-Mediated Platelet Aggregation through Virtual Screening. *Bioorg. Med. Chem. Lett.* **2015**, *25* (6), 1249–1253.

(8) Tuccinardi, T.; Poli, G.; Corchia, I.; Granchi, C.; Lapillo, M.; Macchia, M.; Minutolo, F.; Ortore, G.; Martinelli, A. A Virtual Screening Study for Lactate Dehydrogenase 5 Inhibitors by Using a Pharmacophore-Based Approach. *Mol. Inform.* **2016**, *35* (8–9), 434–439.

(9) Poli, G.; Gelain, A.; Porta, F.; Asai, A.; Martinelli, A.; Tuccinardi, T. Identification of a New

STAT3 Dimerization Inhibitor through a Pharmacophore-Based Virtual Screening Approach. *J. Enzyme Inhib. Med. Chem.* **2016**, *31* (6), 1011–1017.

(10) Chiarelli, L. R.; Mori, M.; Barlocco, D.; Beretta, G.; Gelain, A.; Pini, E.; Porcino, M.; Mori, G.; Stelitano, G.; Costantino, L.; Lapillo, M.; Bonanni, D.; Poli, G.; Tuccinardi, T.; Villa, S.; Meneghetti, F. Discovery and Development of Novel Salicylate Synthase (MbtI) Furanic Inhibitors as Antitubercular Agents. *Eur. J. Med. Chem.* **2018**, *155*, 754–763.

(11) Poli, G.; Scarpino, A.; Aissaoui, M.; Granchi, C.; Minutolo, F.; Martinelli, A.; Tuccinardi, T. Identification of Lactate Dehydrogenase 5 Inhibitors Using Pharmacophore- Driven Consensus Docking. *Curr. Bioact. Compd.* **2018**, *14* (2), 197–204.

(12) Luger, D.; Poli, G.; Wieder, M.; Stadler, M.; Ke, S.; Ernst, M.; Hohaus, A.; Linder, T.; Seidel, T.; Langer, T.; Khom, S.; Hering, S. Identification of the Putative Binding Pocket of Valerenic Acid on GABAA Receptors Using Docking Studies and Site-Directed Mutagenesis. *Br. J. Pharmacol.* **2015**.

(13) Aghazadeh Tabrizi, M.; Baraldi, P. G.; Ruggiero, E.; Saponaro, G.; Baraldi, S.; Poli, G.; Tuccinardi, T.; Ravani, A.; Vincenzi, F.; Borea, P. A.; Varani, K. Synthesis and Structure Activity Relationship Investigation of Triazolo[1,5-a]Pyrimidines as CB2 Cannabinoid Receptor Inverse Agonists. *Eur. J. Med. Chem.* **2016**, *113*, 11–27.

(14) Irwin, J. J.; Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **2016**, *59* (9), 4103–4120.

(15) Chen, Y.-C. Beware of Docking! *Trends Pharmacol. Sci.* **2015**, *36* (2), 78–95.

(16) Bai, F.; Liao, S.; Gu, J.; Jiang, H.; Wang, X.; Li, H. An Accurate Metalloprotein-Specific Scoring Function and Molecular Docking Program Devised by a Dynamic Sampling and Iteration Optimization Strategy. *J. Chem. Inf. Model.* **2015**, *55* (4), 833–847.

(17) Poli, G.; Jha, V.; Martinelli, A.; Supuran, C.; Tuccinardi, T. Development of a Fingerprint-Based Scoring Function for the Prediction of the Binding Mode of Carbonic Anhydrase II Inhibitors. *Int. J. Mol. Sci.* **2018**, *19* (7), E1851.

(18) Pecina, A.; Brynda, J.; Vrzal, L.; Gnanasekaran, R.; Hořejší, M.; Eyrilmez, S. M.; Řezáč, J.; Lepšík, M.; Řezáčová, P.; Hobza, P.; Majer, P.; Veverka, V.; Fanfrlík, J. Ranking Power of the SQM/COSMO Scoring Function on Carbonic Anhydrase II-Inhibitor Complexes. *ChemPhysChem* **2018**, *19* (7), 873–879.

(19) Wang, L.; Yang, C.; Lu, W.; Liu, L.; Gao, R.; Liao, S.; Zhao, Z.; Zhu, L.; Xu, Y.; Li, H.; Huang, J.; Zhu, W. Discovery of New Potent Inhibitors for Carbonic Anhydrase IX by Structure-Based Virtual Screening. *Bioorg. Med. Chem. Lett.* **2013**, *23* (12), 3496–3499.

(20) Gidaro, M.; Alcaro, F.; Carradori, S.; Costa, G.; Vullo, D.; Supuran, C.; Alcaro, S. Eriocitrin and Apigenin as New Carbonic Anhydrase VA Inhibitors from a Virtual Screening of Calabrian Natural Products. *Planta Med.* **2015**, *81* (06), 533–540.

(21) Salmas, R. E.; Senturk, M.; Yurtsever, M.; Durdagi, S. Discovering Novel Carbonic Anhydrase Type IX (CA IX) Inhibitors from Seven Million Compounds Using Virtual Screening and in Vitro Analysis. *J. Enzyme Inhib. Med. Chem.* **2016**, *31* (3), 425–433.

(22) Poli, G.; Giuntini, N.; Martinelli, A.; Tuccinardi, T. Application of a FLAP-Consensus Docking Mixed Strategy for the Identification of New Fatty Acid Amide Hydrolase Inhibitors. *J. Chem. Inf. Model.* **2015**, *55* (3), 667–675.

(23) Liu, K.; Kokubo, H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-Docking Study. *J. Chem. Inf. Model.* **2017**, *57* (10), 2514–2522.

(24) Zhang, C.; Feng, L.-J.; Huang, Y.; Wu, D.; Li, Z.; Zhou, Q.; Wu, Y.; Luo, H.-B. Discovery of Novel Phosphodiesterase-2A Inhibitors by Structure-Based Virtual Screening, Structural Optimization, and Bioassay. *J. Chem. Inf. Model.* **2017**, *57* (2), 355–364.

(25) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W. D. Comparison of Topological, Shape, and Docking Methods in Virtual Screening. *J. Chem. Inf. Model.* **2007**, *47* (4), 1504–1519.

(26) Sastry, M.; Lowrie, J. F.; Dixon, S. L.; Sherman, W. Large-Scale Systematic Analysis of 2D

Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J. Chem. Inf. Model.* **2010**, *50* (5), 771–784.

(27) Poli, G.; Tuccinardi, T.; Rizzolio, F.; Caligiuri, I.; Botta, L.; Granchi, C.; Ortore, G.; Minutolo, F.; Schenone, S.; Martinelli, A. Identification of New Fyn Kinase Inhibitors Using a FLAP-Based Approach. *J. Chem. Inf. Model.* **2013**, *53* (10), 2538–2547.

(28) Broccatelli, F.; Brown, N. Best of Both Worlds: On the Complementarity of Ligand-Based and Structure-Based Virtual Screening. *J. Chem. Inf. Model.* **2014**, *54* (6), 1634–1641.

(29) Grüneberg, S.; Stubbs, M. T.; Klebe, G. Successful Virtual Screening for Novel Inhibitors of Human Carbonic Anhydrase: Strategy and Experimental Confirmation. *J. Med. Chem.* **2002**, *45* (17), 3588–3602.

(30) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**.

(31) Thiry, A.; Ledecq, M.; Cecchi, A.; Dogné, J.-M.; Wouters, J.; Supuran, C. T.; Masereel, B. Indanesulfonamides as Carbonic Anhydrase Inhibitors. Toward Structure-Based Design of Selective Inhibitors of the Tumor-Associated Isozyme CA IX. *J. Med. Chem.* **2006**, *49* (9), 2743–2749.

(32) Havránková, E.; Csöllei, J.; Vullo, D.; Garaj, V.; Pazdera, P.; Supuran, C. T. Novel Sulfonamide Incorporating Piperazine, Aminoalcohol and 1,3,5-Triazine Structural Motifs with Carbonic Anhydrase I, II and IX Inhibitory Action. *Bioorg. Chem.* **2018**, *77*, 25–37.

(33) Garaj, V.; Puccetti, L.; Fasolis, G.; Winum, J.-Y.; Montero, J.-L.; Scozzafava, A.; Vullo, D.; Innocenti, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors: Synthesis and Inhibition of Cytosolic/Tumor-Associated Carbonic Anhydrase Isozymes I, II, and IX with Sulfonamides Incorporating 1,2,4-Triazine Moieties. *Bioorg. Med. Chem. Lett.* **2004**, *14* (21), 5427–5433.

(34) Carta, F.; Garaj, V.; Maresca, A.; Wagner, J.; Avvaru, B. S.; Robbins, A. H.; Scozzafava, A.; McKenna, R.; Supuran, C. T. Sulfonamides Incorporating 1,3,5-Triazine Moieties Selectively and Potently Inhibit Carbonic Anhydrase Transmembrane Isoforms IX, XII and XIV over

Cytosolic Isoforms I and II: Solution and X-Ray Crystallographic Studies. *Bioorg. Med. Chem.* **2011**, *19* (10), 3105–3119.

(35) Irwin, J. J.; Raushel, F. M.; Shoichet, B. K. Virtual Screening against Metalloenzymes for Inhibitors and Substrates †. *Biochemistry* **2005**, *44* (37), 12316–12328.

(36) Stote, R. H.; Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins Struct. Funct. Genet.* **1995**, *23* (1), 12–31.

(37) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume Overlap Scoring. *J. Chem. Inf. Model.* **2011**, *51* (10), 2455–2466.