Virtual screening of the potential hCA XII inhibitors based on the pharmacophore modelling, molecular docking, MMGBSA and molecular dynamics simulations studies

Mohammad M. Al-Sanea

Jouf University

Garri Chilingaryan (✉ g_chilingaryan@mb.sci.am)

Institute of Molecular Biology of NAS RA

Narek Abelyan

Russian-Armenian University

Grigor Arakelov

Institute of Molecular Biology of NAS RA

Harutyun Sahakyan

Institute of Molecular Biology of NAS RA

Vahram Arakelov

Institute of Molecular Biology of NAS RA

Shaimaa Hussein

Jouf University

Gharam M. Alazmi

Jouf University

Haifa E. Alsharari

Jouf University

Waad M. Al-faraj

Jouf University

Faten S. Alruwaili

Jouf University

Nouf Q. Albilasi

Jouf University

Tahani S. Alsharari

Jouf University

Abdulaziz A. S. Alsaleh

Jouf University

Turki M. Alazmi

Jouf University

Atiah H. Almalki

Jouf University

Nasser H. Alotaibi

Jouf University

Mohamed A. Abdelgawad

Jouf University

Research Article

Keywords: human carbonic anhydrase XII, virtual screening, simulations studies

Posted Date: April 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-394808/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on July 30th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-94809-x>.

Virtual screening of the potential hCA XII inhibitors based on the pharmacophore modelling, molecular docking, MMGBSA and molecular dynamics simulations studies

Mohammad M. Al-Sanea^{1,†}, Garri Chilingaryan^{2,3*,†}, Narek Abelyan^{3,4†}, Grigor Arakelov², Harutyun Sahakyan^{2,4}, Vahram G. Arakelov², Shaimaa Hussein⁵, Gharam M. Alazmi¹, Haifa E. Alsharari¹, Waad M. Al-faraj¹, Faten S. Alruwaili¹, Nouf Q. Albilasi¹, Tahani S. Alsharari¹, Abdulaziz A. S. Alsaleh¹, Turki M. Alazmi¹, Atiah H. Almalki^{7,8}, Nasser H. Alotaibi⁶, Mohamed A. Abdelgawad¹

¹ Pharmaceutical Chemistry Department, College of Pharmacy, Jouf University, Sakaka, Aljouf 72341, Saudi Arabia; mmalsanea@ju.edu.sa

² Institute of Molecular Biology of NAS RA, 0014, Yerevan, Armenia.

³ Institute of Biomedicine and Pharmacy, Russian-Armenian University, 0051, Yerevan, Armenia.

⁴ Foundation for Armenian Science and Technology, 0033, Yerevan, Armenia.

⁵ Department of Pharmacology, College of Pharmacy, Jouf University, Sakaka, Aljouf 72341, Saudi Arabia.

⁶ Department of Clinical Pharmacy, College of Pharmacy, Jouf University, Sakaka, Aljouf 72341, Saudi Arabia.

⁷ Department of Pharmaceutical Chemistry, College of Pharmacy, Taif University, Taif 21944, Saudi Arabia

⁸ Addiction and Neuroscience Research Unit, Taif University, Al-Hawiah, Taif 21944, Saudi Arabia

[†]These authors contributed equally to this work.

*g_chilingaryan@mb.sci.am

ABSTRACT

Human carbonic anhydrase XII (hCA XII) has gained therapeutic value in the field of medicinal chemistry as a pharmacological target and biomarker for different types of cancer. Despite the fact that interaction features of hCAs inhibitors with the catalytic site of the enzyme are well described, lack in the selectivity of the traditional hCA inhibitors is an urgent issue. This study was devoted to the identification of novel potential hCA XII inhibitors using comprehensive set of computational approaches, including generation of structure- and ligand-based pharmacophore models, molecular docking, MMGBSA calculations and molecular dynamics simulations. As the results of the study multiple potential hCA inhibitors were identified. Along with the identified derivatives of the classical hCA inhibitors (have sulfonamide motifs), several compounds with alternative chemical scaffolds, that can be considered and further investigated as potential non-classical hCA XII inhibitors, were identified.

Introduction

Carbonic anhydrases (CAs, EC 4.2.1.1) are zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide into bicarbonate and a proton¹. Sixteen CA isoforms are found in humans and all vary in kinetic properties, subcellular localization and distribution to the tissues. Many essential physiological processes such as pH regulation, electrolyte secretion, respiration, bone resorption and tumorigenesis are

involved in the various hCAs, so irregular levels and/or activities of these enzymes are typically associated with many diseases, such as glaucoma, neurological diseases, osteoporosis and metabolic disorders².

In general, CA inhibitors (CAIs) are compounds equipped with an effective zinc-binding group (ZBG) capable of chelating the prosthetic zinc ion placed inside the hCA binding site that is essential for these enzymes' catalytic action. Sulfonamide moieties or related structural motifs represent the most common ZBGs shared by typical CAIs (such as sulfamides and sulfamates). These groups are especially effective in endowing high-affinity small-molecule ligands with hCAs, as they enable not only the proper coordination of the catalytic zinc ion, but also the forming of associations of H-bonds with key protein residues located in the zinc-binding cavity area. In particular, the most significant groups of CAIs are these molecules, and multiple ligands with high inhibitory potency have been recorded to date³.

However, because of the high amino acid conservation found at the level of their catalytic site and adjacent regions in the various hCA isoforms, these ligands also present the downside of being insufficiently selective against particular hCAs that are therapeutically interesting. In particular, in the field of medicinal chemistry, hCA IX and hCA XII have gained special interest, as these isoforms have been designated as biomarkers and pharmacological targets for different types of cancer. As a result, the inhibition of these two enzymes is currently the subject of a vast field of study⁴.

Notwithstanding, the development of hCA IX/XII selectivity inhibitors over hCA I/II, which are ubiquitously distributed and involved in key physiological processes, is still a difficult challenge, although some examples of selective ligands have been reported^{5,6}.

Non-classical CAIs are an especially useful resource in this sense for finding isoform specificity and avoiding possible off-target events, side effects and or adverse reactions (such as sulfur allergies) associated with the use of sulfurized ligands⁷.

This study was devoted to identification of novel potential hCA XII inhibitors using computational approaches: pharmacophore modeling (structure-, and ligand-based), molecular docking, MMGBSA and molecular dynamics simulations.

Materials and Methods

Preparation of the pharmacophore validation datasets

3D structures of the 2500 compounds with $K_i < 10$ against hCA XII enzyme were downloaded from PubChem⁸ and ChEMBL⁹ databases. 100 of them were used for generation of 5000 decoys using a Database of Useful Decoys: Enhanced (DUD-E)¹⁰ and were not included in the validation datasets. Two validation datasets were prepared and each of them consisted of 5% active compounds and 95% decoys. First validation dataset included active top 135 active compounds based on the K_i values. Second dataset included randomly picked 135 active compounds. 3D structures of decoys were generated by Open Babel v3.1.1¹¹ software based on the smiles obtained by using DUD-E. Hydrogens were added to all compounds using Open Babel software. The iConBest method, implemented in LigandsScout software v4.4, was used for the generation of the conformations for all compounds.

Structure-based approach pharmacophore generation

Both, structure-based and ligand-based pharmacophore models were generated using LigandScout v4.4¹². For structure-based pharmacophore generation, 22 crystal structures of the hCA XII in complexes with inhibitors (resolution < 2 Å) available in the Protein Data Bank (PDB)¹³. Several approaches for the structure-based pharmacophore model generation were tested: 1) based on the separate crystal structures, 2) based on the combination of separate crystal structures, 3) merged and 4) shared pharmacophores. Best pharmacophore model was generated using crystal structure of the catalytic domain of human carbonic anhydrase isozyme XII with benzenesulfonamide derivative (PDB ID : 4QJW with resolution 1.55 Å, chain C). All features that characterize interactions with the water molecules and hydrophobic interactions were removed from the obtained pharmacophore model. Also exclusion volumes coat feature was added.

Ligand-based pharmacophore generation

Clusterization of the top 135 active compounds with the best K_i values was performed using pharmacophore alignment score as similarity measure. Maximum number of the conformations and cluster distance values were set to 3 and 0.4, respectively. The average method was used for cluster distance calculations. Based on the clusterization of the 135 active compounds 14 clusters were identified and representative compounds of those clusters were used for the creation of ligand-based pharmacophore models (Table S1). The best ligand-based pharmacophore model was obtained based on the biggest cluster. “H bond donor” feature was removed from the selected ligand-based pharmacophore model.

Validation metrics

ROC AUC - represents the area value under the ROC (receiver operating characteristic). Values range from 0 to 1, where 1 is perfect classification and 0.5 and below is random classification.

ROC, widely used to evaluate virtual screening and pharmacophore modeling methods^{14,15}, is defined as a graphical representation of the test sensitivity in relation to its specificity or false-positive rate. The AUC is the probability of active compounds being ranked earlier than decoy compounds. The classifier “precision” represents the share of true positives (TP) among all hits ($TP/(TP+FP)$, where FP is false positive compounds). The classifier “specificity” represents the ratio of the active compounds found in the list of compounds identified as “true positives” (TP/A , where A is active compounds). $TP/(TP+FP)$). “Specificity” classifier is the ratio of true negative compounds to all compounds in the database, excluding active ones.

Enrichment Factor (EF) measures the fraction of active compounds found in a specific percentage, solving the problem of comparing the results for datasets with different active/inactive compound ratios¹⁶. The EF for 1, 5, 10 and 100 % was calculated for the share of true positives among the molecules identified as hit compounds using generated pharmacophore models.

Virtual screening

ZINCPharmer¹⁷ was used for filtering of the ZINC purchasable (21,777,093 compounds) and ZINC natural products and natural derivatives (197, 488 compounds) datasets using generated structure-based and ligand-based pharmacophore models. Identified hit compounds were used for the molecular docking against the active site of the hCA XII enzyme (PDB ID 4QJW). AutoDock Tools¹⁸ was used for the estimation of the grid box ($< 27 \text{ \AA}$), calculation of the protein’s and compounds’ charges and the addition of the polar hydrogens. Virtual screening was performed using AutoDock Vina software¹⁹. AutoDock Vina was regarded as highly efficient software for molecular docking and virtual screening procedures based on the recent benchmark studies among both academic and commercial software²⁰. AutoDock Vina uses “Iterated Local Search global optimizer” similar to that by Abagyan *et al* and Broyden-Fletcher-Goldfarb-Shanno (BFGS), which is quasi-Newton method for the local optimization as a search algorithm and hybrid scoring function (empirical + knowledge-based function) inspired in the X-Score function¹⁹. Standard, recommended by the developers, parameters were used for the virtual screening procedure.

Re-scoring using MMGBSA method

The algorithm for the molecular mechanics-generalized Born surface area (MMGBSA) calculations is based on a freely-available AmberTools suite²¹. The algorithm can be described in three stages, 1) receptor and ligands parametrization 2) minimization and 3) MM/GBSA and MM/PBSA calculations. At the first stage, the ff14SB force field²² is used to describe protein parameters, and General Amber Force Field (GAFF)²³ with AM1-BCC charge model²⁴ is used for small molecule parametrization. Next, the algorithm prepares necessary input files (coordinates and topologies) with mbondi3 radii using LEaP. Minimization is performed in generalized Born implicit solvent models (igb=8) using the sander engine. Finally, for the free energy calculations, the algorithm uses the MMPBSA.py program²⁵ for the MMGBSA calculations. The algorithm is implemented in a bash script and can be run in parallel in most Linux distributions without any additional libraries for parallelization. The full code is available at the following link: <https://github.com/sahakyanhk/iPBSA>

Molecular Dynamics simulations

The molecular dynamics simulations were carried out using AMBER20 with ff14SB force field for protein and GAFF for the ligand parameterization according to the AM1-BCC scheme order to calculate the atomic point charges. Minimized conformations from the previous step (re-scoring of docked complexes with MMGBSA) were used as starting positions for corresponding simulations. The complexes were solvated with TIP3P water and Na⁺ /Cl⁻ ions at 150 mM concentration²⁶. The Monte Carlo barostat²⁷ with reference pressure at 1 bar and Langevin thermostat²⁸ with collision frequency (γ_{in}) 2 ps⁻¹ to keep the temperature at 310.15 K were used. Particle Mesh Ewald (PME) with electrostatic interactions cut off at 1.0 nm was used for the long-range electrostatic interactions. Bonds involving hydrogen were constrained using the SHAKE algorithm and 2fs integration step was used²⁹. Finally, for every simulation, binding free energies were re-calculated using the same MMGBSA method and MMPBSA.py program, using 250 snapshots with equal intervals collected from the last 20 ns of every trajectory for each simulation.

Results and Discussion

Ligand-based pharmacophore generation

As a result of clusterizations, 14 clusters have been obtained. The best ligand-based pharmacophore model was obtained based on the biggest cluster (fig. 1). “H bond donor” feature was removed from the pharmacophore. Therefore, the final ligand-based pharmacophore model included three “H-bond (acceptor)” features, two “H-bond (donor)” features, one “hydrophobic” and “aromatic” features of the LigandScout v4.4.

Figure 1. The best ligand-based pharmacophore model and active compounds (with PubChem CIDs) of the selected cluster used for its generation.

Validation tests of the generated ligand-based pharmacophore model demonstrated high results on two separate datasets. First dataset: 105 hits (97 TP, 8 FP), AUC 1, 5, 10, 100 - 1, 1, 1, 0.86; EF 1, 5, 10, 100 - 19.8, 18.3, 18.3, 18.3. Second dataset: 74 hit compounds (63 TP, 11 FP) AUC 1, 5, 10, 100 - 1, 1, 1, 0.73; EF 1, 5, 10, 100 - 19.8, 16.9, 16.9, 16.9 (fig. 2).

Figure 2. Validation of the generated ligand-based pharmacophore on the two datasets (A and B) Identified hit compounds out of 2635 total compounds (135 actives, 2500 decoys).

Receptor-ligand pharmacophore generation (Structure-based approach)

Structure-based pharmacophore model generated using crystal structure of catalytic domain of human carbonic anhydrase isozyme XII with benzenesulfonamide derivative (PDB ID: 4QJW) demonstrated LigandScout's binding affinity score of -17.18, which includes interaction and desolvation energies. The selected structure-based pharmacophore model included two “H-bond (acceptor)” that are important for the interaction of the ligand with ASN64 and THR198 of the active site of hCA XII enzyme, and two “H-bond (donor) feature that are important for interaction with PRO200 and GLU104 (fig. 3).

Figure 3. The best generated structure-based pharmacophore model. 3D (A) and 2D (C) representations of the pharmacophore model with all initial features. Final pharmacophore models after removing

features responsible for the interactions with water molecules and hydrophobic interactions (B, D) and addition of the exclusion volumes coat (B).

Validation of structure-based pharmacophore model in the case of the first dataset demonstrated following values: 103 hit compounds (38 TP, 65 FP), AUC 1, 5, 10, 100 - 0.72, 0.93, 0.95, 0.63; EF 1, 5, 10, 100 - 7.6, 7.3, 7.3, 7.3. In the case of the second validation dataset: 82 hit compounds (24 TP, 58 FP), AUC 1, 5, 10, 100 - 0.97, 0.92, 0.93, 0.58; EF 1, 5, 10, 100 - 3.8, 5.8, 5.8, 5.8 (fig. 4).

Figure 4. Validation of the generated structure-based pharmacophore on the two datasets (A and B). Identified hit compounds out of 2635 total compounds (135 actives, 2500 decoys).

Filtration of zinc database using generated pharmacophore models

Both selected best ligand-based and structure-based pharmacophore models were applied using ZINCPharmer, separately, to filter ZINC datasets of purchasable and natural and derivatives compounds datasets (total of 21.970.000 compounds). As the result 6357 and 193 hit compounds were identified from purchasable and natural (including derivatives) datasets, respectively, using a structure-based pharmacophore model. Other 8415 and 1514 hit compounds were identified using a ligand-based pharmacophore model. All of those compounds have been passed to molecular docking against the active site of the hCA XII enzyme.

Molecular Docking

As the result of the molecular docking of compounds, obtained by application of the ligand-based pharmacophore, against ligand binding site of the hCA XII enzyme, 1361 “purchasable” and 264 natural derivatives compounds with similar or higher docking scores compared to the reference molecule (co-crystallized ligand of 4QJW, benzenesulfonamide derivative: -8 kJ/mol Vina core) were identified (Table S2).

As the result of the molecular docking of compounds, obtained by application of the structure-based pharmacophore, against ligand binding site of the hCA XII enzyme, 1027 “purchasable” and 17 natural derivatives compounds with similar or higher docking scores compared to the reference molecule (co-crystallized ligand of 4QJW, benzenesulfonamide derivative: -8 kJ/mol Vina core) were identified (Table S3).

All of the compounds with similar or higher docking scores in comparison to reference ligand, were passed to the additional stages of minimization and re-scoring using MM-GBSA method.

MMGBSA re-scoring results

As the results of the minimization procedures and following calculations of the binding energies of the complexes of the identified compounds with hCA XII, obtained as the result of performed molecular docking only 38 compounds from the “purchasable” (Table 1) and 2 from “natural and derivatives” (Table 2) datasets demonstrated similar or higher binding energy values than reference ligand. 24 out of aforementioned 38 compounds were obtained with the use of structure-based pharmacophore and the rest

15 with the use of ligand-based pharmacophore. Both compounds identified from “naturals and derivatives” dataset were obtained with the use of a ligand-based pharmacophore model.

Structure-based approach				Ligand-based approach			
ZINC ID	Energy (kJ/mol)	ZINC ID	Energy (kJ/mol)	ZINC ID	Energy (kJ/mol)	ZINC ID	Energy (kJ/mol)
ZINC66466630	-157,83	ZINC76941861	-129,56	ZINC82980951	-159,25	ZINC21761334	-133,43
ZINC16137455	-153,97	ZINC39147130	-128,51	ZINC68025286	-147,68	ZINC10514459	-130,47
ZINC90089846	-152,20	ZINC75669379	-127,26	ZINC27522612	-144,36	ZINC38671716	-129,54
ZINC76965500	-146,49	ZINC39252273	-127,24	ZINC12555593	-141,68	ZINC55678727	-128,52
ZINC05699310	-139,63	ZINC12551609	-126,40	ZINC29565674	-139,19	ZINC27929386	-128,16
ZINC09126577	-139,14	ZINC89392154	-126,25	ZINC59456846	-137,50	ZINC13056037	-127,18
ZINC44547064	-138,05	ZINC12085595	-125,75	ZINC49448410	-136,54	ZINC40897288	-124,45
ZINC09562497	-135,91	ZINC39252405	-125,39	ZINC06510447	-136,40	ligand_4qjw	-124,20
ZINC22239311	-134,62	ZINC12983599	-124,99				
ZINC47251290	-134,41	ZINC58304576	-124,67				
ZINC58247763	-132,89	ZINC06142500	-124,23				
ZINC40218576	-131,14	ligand_4qjw	-124,20				
ZINC58263892	-129,77						

Table1. Binding energies of the top “purchasable” compounds after re-scoring using MM-GBSA

In order to analyse chemical diversity of identified compounds, an agglomerative hierarchical clustering method based on the “ECFP” fingerprint was performed (fig. 5). As a result of the clusterization of the identified 38 compounds, 18 clusters were obtained. As expected, most of the identified compounds (fig. 5, clusters 1-10) have sulfonamide moieties or related structural motifs, that are widely known to promote small-molecules interaction with hCAs active site. However, several of the identified compounds (fig. 5, clusters 11-18) have different from sulfonamide chemical moieties. Those compounds include cyclohexanecarboxamide, propanamide, acetamide, cyclopropanecarbohydrazide, thiadiazolidine, cyclopropane, carbohydrazide, chromen-7, tetrahydrofuran derivatives and analogues.

Figure 5. Results of the clusterization of the identified potential hcAXII inhibitors as the result of MM-GBSA re-scoring procedure. 2D structures of representative compounds (colored in red) are presented for all clusters.

Remarkably one of those compounds, identified with the use of ligand-based pharmacophore (ZINC82980951), showed lowest binding energy among all compounds as the result of MM-GBSA re-scoring (Table 1). Thus, compounds that are different from the traditional hCA XII inhibitors (which include sulfonamide chemical group) are of special interest for further analysis and investigation as potential alternative hCA XII inhibitors.

In the case of the natural and natural-derivatives compounds only two compounds (ZINC49181869 and ZINC49181861, Table 2), which are stereoisomers of the same oxychromen-2-one derivative, showed higher binding energies than reference ligand.

<i>Structure-based approach</i>		<i>Ligand-based approach</i>	
ZINC ID	Energy(kJ/mol)	ZINC ID	Energy(kJ/mol)
ZINC70704873	-105,68	ZINC49181869	-139,62
ZINC04221765	-93,03	ZINC49181861	-133,30
ZINC02131655	-85,34	ZINC49181866	-120,74
ZINC70699917	-77,17	ZINC08829478	-111,70
ZINC15958674	-76,88	ZINC08792367	-109,86

Table 2. Binding energies of the top “natural and derivatives” compounds after re-scoring using MM-GBSA.

Molecular dynamics

From a list of compounds identified from the “purchasable” dataset, two compounds with lowest binding energy (one identified using structure-based pharmacophore (ZINC66466630) and another one using ligand-based pharmacophore model (ZINC82980951) were selected for additional MD simulations. In the case of the “natural and derivatives” dataset, two compounds with the lowest binding energies identified using structure-based (ZINC70704873) and ligand-based pharmacophore (ZINC49181869) models were also selected for MD simulations.

Based on the results of conducted 100 ns MD simulation of the reference compound (co-crystallized ligand of 4QJW structure), it was stable during the whole simulation and has maintained conformation close to its initial crystal state (fluctuations around 0.05 nm, fig. 6).

Figure 6. RMSD values of the selected four compounds and reference ligand.

From the four tested compounds, only ZINC49181869 was stable during the whole simulation and also maintained conformation close to the one predicted by molecular docking fluctuations around 0.05 nm. Compound ZINC70704873 stabilized after ~55 ns of the simulation and maintains its conformation with fluctuations around ~ 0.05 nm. ZINC66466630 stabilized after ~70 ns and maintains its conformation with fluctuations around ~ 0.025nm. ZINC82980951 stabilized after ~50 ns and maintains its conformation with fluctuations around ~ 0.1 nm.

In its stable conformation, during molecular dynamics simulation, the reference ligand has five hydrogen bonds with the following amino acid residues of active site of the hCA XII enzyme: HIE 117, HID 91, GLN 89, ASN 64, THR 198. In the presence of the reference ligand, HID 93, HIE 117 and THR 198 also have hydrogen bonds with Zn ion (fig. 7).

Figure 7. A. ZINC66466630 B. ZINC70704873 C. ZINC49181869 D. ZINC82980951 E. Reference Common interacting amino acid residues of the hCA XII active site with tested compounds and reference ligand: THR 198 (ZINC66466630, ZINC82980951), THR 198 and GLN 89 (ZINC70704873), GLN 89 and HID 91 (ZINC49181869).

Common amino acid residues that interact with the Zn in the presence of tested compounds and reference ligand: HID 93, THR 198 (ZINC66466630, ZINC70704873, ZINC49181869), HID 93 (ZINC82980951). Unique, in comparison to the reference ligand, interacting amino acid residues with the tested compounds: HID 93 GLU 104, SER 133 (ZINC66466630), PRO 200, LYS 69 (ZINC70704873), HID 93, GLU 104, LYS 71 (ZINC49181869), SER 30 (ZINC82980951).

Unique, in comparison to the reference ligand, amino acid residues that interact with the Zn in the presence of the tested compounds: GLU 104 (ZINC70704873), HID 91 and GIU 104 (ZINC49181869), HID 93 (ZINC82980951).

Conclusion

The hCA XII enzyme has gained special interest in the drug design and discovery field, as a biomarker and pharmacological target for different types of cancer. Given the medical relevance, the discovery and development of selective inhibitors for hCA XII have been under continuous efforts. However typical hCAs binding ligands, that have sulfonamide functional groups, have insufficient selectivity against particular hCAs that are therapeutically interesting, such as hCA XII.

In this study, ligand- and structure-based pharmacophore modeling approaches were combined with molecular docking, MMGBSA, molecular dynamic simulations and applied with the purpose of finding new potential hCA XII inhibitors from a vast number commercially available natural, natural derivatives and synthetic compounds. As the result of the study, a number of new compounds with typical sulfonamide functional group were identified. Number of identified compounds have non-classical chemical scaffolds and groups and are of special interest as alternative potential hCA XII inhibitors and matter of further investigations.

References

1. Supuran, C. T. Structure and function of carbonic anhydrases. *Biochem. J.* **473**, 2023–2032 (2016).
2. Supuran, C. T. Advances in structure-based drug discovery of carbonic anhydrase inhibitors. *Expert Opin. Drug Discov.* **12**, 61–88 (2017).
3. Carta, F., Supuran, C. T. & Scozzafava, A. Sulfonamides and their isosters as carbonic anhydrase inhibitors. *Future Med. Chem.* **6**, 1149–1165 (2014).
4. Lomelino, C. & McKenna, R. Carbonic anhydrase inhibitors: a review on the progress of patent literature (2011-2016). *Expert Opin. Ther. Pat.* **26**, 947–956 (2016).
5. Krasavin, M. *et al.* Continued exploration of 1,2,4-oxadiazole periphery for carbonic anhydrase-targeting primary arene sulfonamides: Discovery of subnanomolar inhibitors of membrane-bound hCA IX isoform that selectively kill cancer cells in hypoxic environment. *Eur. J. Med. Chem.* **164**, 92–105 (2019).
6. Andreucci, E. *et al.* The carbonic anhydrase IX inhibitor SLC-0111 sensitises cancer cells to conventional chemotherapy. *J. Enzyme Inhib. Med. Chem.* **34**, 117–123 (2019).
7. Lomelino, C. L., Supuran, C. T. & McKenna, R. Non-Classical Inhibition of Carbonic Anhydrase. *Int. J. Mol. Sci.* **17**, (2016).
8. Kim, S. *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res.* **44**, D1202-1213 (2016).
9. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–D1107 (2012).
10. Mysinger, M. M., Carchia, M., Irwin, John. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).
11. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
12. Wolber, G. & Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **45**, 160–169 (2005).
13. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
14. Empereur-mot, C. *et al.* Predictiveness curves in virtual screening. *J. Cheminformatics* **7**, (2015).
15. Pascual, R., Almansa, C., Plata-Salamán, C. & Vela, J. M. A New Pharmacophore Model for the Design of Sigma-1 Ligands Validated on a Large Experimental Dataset. *Front. Pharmacol.* **10**, (2019).
16. Krüger, D. M. & Evers, A. Comparison of Structure- and Ligand-Based Virtual Screening Protocols Considering Hit List Complementarity and Enrichment Factors. *ChemMedChem* **5**, 148–158 (2010).
17. Koes, D. R. & Camacho, C. J. ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Res.* **40**, W409–W414 (2012).

18. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30**, 2785–2791 (2009).
19. 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.* **31**, 455–461 (2010).
20. Wang, Z. *et al.* Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **18**, 12964–12975 (2016).
21. Macke, T. J. *et al.* NAB (Nucleic Acid Builder). 260.
22. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
23. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
24. Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **23**, 1623–1641 (2002).
25. Miller, B. R. *et al.* MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **8**, 3314–3321 (2012).
26. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
27. Oliver, D. S., Cunha, L. B. & Reynolds, A. C. Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. *Math. Geol.* **29**, 61–91 (1997).
28. Wu, X. & Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chem. Phys. Lett.* **381**, 512–518 (2003).
29. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

Additional information

The authors declare no competing interests.

Figures

Figure 1

The best ligand-based pharmacophore model and active compounds (with PubChem CIDs) of the selected cluster used for its generation.

Figure 2

Validation of the generated ligand-based pharmacophore on the two datasets (A and B) Identified hit compounds out of 2635 total compounds (135 actives, 2500 decoys).

Figure 3

The best generated structure-based pharmacophore model. 3D (A) and 2D (C) representations of the pharmacophore model with all initial features. Final pharmacophore models after removing features responsible for the interactions with water molecules and hydrophobic interactions (B, D) and addition of the exclusion volumes coat (B).

Figure 4

Validation of the generated structure-based pharmacophore on the two datasets (A and B). Identified hit compounds out of 2635 total compounds (135 actives, 2500 decoys).

Figure 5

Results of the clusterization of the identified potential hcAXII inhibitors as the result of MMGBSA re-scoring procedure. 2D structures of representative compounds (colored in red) are presented for all

clusters.

Figure 6

RMSD values of the selected four compounds and reference ligand.

Figure 7

A. ZINC66466630 B. ZINC70704873 C. ZINC49181869 D. ZINC82980951 E. Reference Common
interacting amino acid residues of the hCA XII active site with tested compounds and reference ligand: THR 198 (ZINC66466630, ZINC82980951), THR 198 and GLN 89 (ZINC70704873), GLN 89 and HID 91 (ZINC49181869).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xls](#)
- [Table2.xls](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)