Homology Modeling and Receptor-Based 3D-QSAR Study of Carbonic Anhydrase IX

Tiziano Tuccinardi,*,‡ Gabriella Ortore,‡ Armando Rossello,‡ Claudiu T. Supuran,† and Adriano Martinelli‡

Dipartimento di Scienze Farmaceutiche, Università di Pisa, via Bonanno 6, 56126 Pisa, Italy, and Polo Scientifico, Laboratorio di Chimica Bioinorganica, Room 188, Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Florence, Italy

Received June 19, 2007

Carbonic anhydrase (CA) IX is a very interesting subject for study due to its overexpression in cancer and its expression in very few normal tissues. There are not yet experimental 3D structures of the catalytic domain of this isozyme, and only a few computational studies have been reported. A homology model of CA IX was developed, and using Gold software 124 CA IX inhibitors were docked. The best poses of the ligands were then used as an alignment tool for the development of the first reported CA IX 3D-QSAR model. The obtained results confirm the reliability of the constructed CA IX model and the proposed computational strategy for investigating CAs.

INTRODUCTION

α-Carbonic anhydrases (CAs) are widespread, distributed zinc metalloenzymes. They are implicated in a variety of physiological functions such as pH regulation, CO_2 and HCO_3^- transport, bone resorption, production of biological fluids, ureagenesis, gluconeogenesis, and lipogenesis.[1]

To date, 16 CA isozymes have been identified and mainly differ in their subcellular localization and catalytic activity. Five CAs are cytosolic (CA I−III, VII, and XIII), two are mitochondrial (CA VA and VB), one is secreted (CA VI), and the others are membrane-bound (CA IV, IX, XII, and XIV). Three noncatalytic forms are also reported and defined as carbonic anhydrase-related proteins,[2] and recently, a new isoform which is not expressed in humans, CA XIII, has been reported.[3]

Many CA subtypes constitute interesting targets for the design of pharmacological agents useful, for example, as antiglaucoma, anticonvulsant, antiurolithic, and antiepileptic and for the treatment of obesity.[4−6]

It has been known for many years that inhibitors of CAs are able to inhibit the growth of tumor cells in vitro and in vivo to various degrees,[7] and, among the various CA subtypes, CA IX and CA XII were discovered to be overexpressed and associated with many tumors, where they are involved in processes linked with cancer progression and response to therapy.[8,9]

CA IX is expressed in only a few normal tissues but is ectopically induced and highly overexpressed in many tumor types, mainly due to its transcriptional activation by hypoxia via the transcription factor hypoxia-inducible factor. These properties make CA IX a useful marker for hypoxia as well as a prognostic indicator for many cancers. Furthermore, this CA subtype is also involved in both pH regulation and cell-adhesion processes caused by tumor metabolism.[10]

Therefore, CA IX constitutes an interesting target for novel approaches to anticancer therapy. Further research is needed to better understand its exact role in cancer. In order to elucidate this mechanism, the design of new highly active and selective CA IX inhibitors could be very helpful.

From a computational point of view, CAs (and, in particular, CA II) have been intensely investigated. There are many published examples of docking of CA inhibitors,[11,12] virtual screening studies,[13,14] QSAR,[15−17] and 3D-QSAR studies.[18−20]

Very recently we evaluated, by means of the cross-docking approach, the reliability of the Gold program for the docking of CA II inhibitors.[21] The best docking procedure was then used for docking almost 300 CA II ligands, and the best poses were used as an alignment tool for the development of a 3D-QSAR model.

Presently, no experimental 3D structures of the catalytic domain of CA IX have been reported, and the docking of inhibitors into CA IX homology models has been recently published.[22,23] With regards to other computational studies, Jaiswal and co-workers in 2004 reported the first QSAR study of CA IX inhibitors,[24] whereas till now no 3D-QSAR models have been reported.

In this paper we constructed a homology model of CA IX, then using our recently reported procedure[21] we docked 124 CA IX inhibitors, and using the docking results as an alignment tool we report a 3D-QSAR model of CA IX.

MATERIALS AND METHODS

Homology Modeling. All the primary sequences were obtained from the SWISS-PROT protein sequence database.[25] Sequence similarity searches were carried out using BlastP.[26] The crystal structure of the murine CAXIV (1RJ6) was taken from the Protein Data Bank.[27,28] The sequence alignment of CA IX, CA XIV, CA XII, and CA II was performed by CLUSTAL W,[29] with a gap open penalty of 10 and a gap extension penalty of 0.05.

* Corresponding author fax: ++39 050 2219605; e-mail: tuccinardi@farm.unipi.it.
† Università degli Studi di Firenze.
‡ Università di Pisa.

10.1021/ci700214j CCC: $37.00
Published on Web 10/20/2007

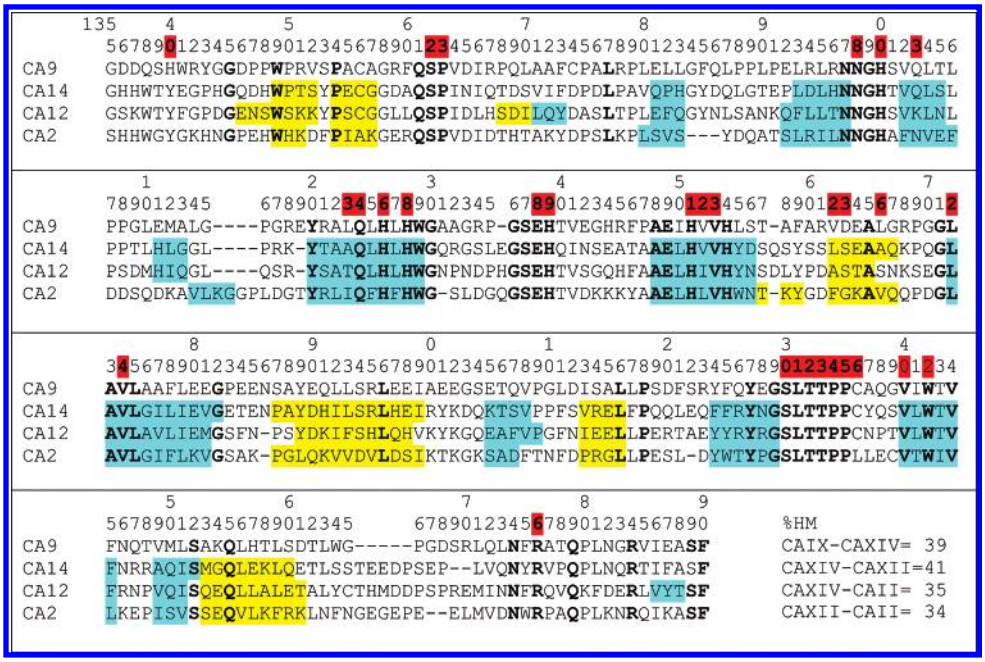


Figure 1. Alignment of human CA IX with murine CA XIV, human CA XII, and CA II. Conserved residues are in bold. The positions of secondary structure elements are colored yellow (α-helix) and sky blue (*β*-sheet). In the first line of the alignment scheme the numeration of CA IX is reported; the residue numbers of the CA IX binding site are in bold and colored red.

The extracellular domain of CA IX (residue numbers 135−390) was constructed using Modeller 9v1.[30] The loop region comprised between residues L363-D368 was constructed by means of the "Loop optimization method" of Modeller, applying the "very_slow" loop refinement method. Starting from this protein, ten structures were generated by means of the "very slow MD annealing" refinement method, as implemented in Modeller, and the best receptor model was chosen on the basis of the DOPE (Discrete Optimized Protein Energy) assessment method. The backbone conformation of the resulting receptor structure was evaluated by inspection of the Ramachandran plot using PROCHECK (see Figure 1 in the Supporting Information).[31]

The chosen receptor model was complexed with acetazolamide (see below for the docking procedure) and minimized by means of AMBER 9,[32] using the parm94 force field at 300 K. The complex was placed in a rectangular parallelepiped water box, an explicit solvent model for water, TIP3P, was used, and the complex was solvated with a 10 Å water cap. Sodium ions were added as counterions to neutralize the system. Two steps of minimization were then carried out: in the first stage, we kept the protein fixed with a position restraint of 500 kcal/mol/Å^2, and we solely minimized the positions of the water molecules. In the second stage, we minimized the entire system through 5000 steps of Steepest Descent followed by Conjugate Gradient (CG) until a convergence of 0.05 kcal/ Å·mol. For the zinc ion, the parameters developed by Merz and co-workers were used.[33] The Ramachandran plot of the resulting complex highlighted that no residue had a disallowed geometry (see Figure 2 in the Supporting Information).

Molecular Dynamics simulations. Molecular dynamics (MD) simulations were performed using AMBER 9,[32] and they were carried out using the same conditions described above for the preliminary minimization steps. MD was done on a 1 CPU/node, 4 node, 1024MB RAM/node cluster Intel Core 2 Duo E6600 2.4 GHz, and the average wall clock time for a 1 ns simulation was about 19 h.

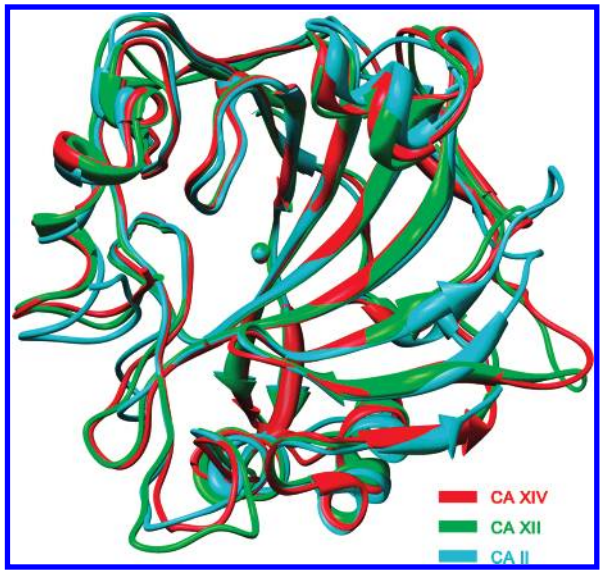


Figure 2. Superimposition between the X-ray structures of CA XIV, CA XII, and CA II.

Particle mesh Ewald (PME)[34] electrostatics and periodic boundary conditions were used in the simulation, MD trajectory was run using the minimized structure as the starting conformation, the time step of the simulations was 2.0 fs with a cutoff of 10 Å for the nonbonded interaction, and SHAKE was employed to keep all bonds involving hydrogen atoms rigid. The nonbonded list was updated whenever any atom moved more than 1 Å (half of the default width of the nonbonded "skin" parameter) since the last list update. Constant-volume periodic boundary MD was carried out for 200 ps, during which the temperature was raised from 0 to 300 K; then 6 ns of constant-pressure periodic boundary

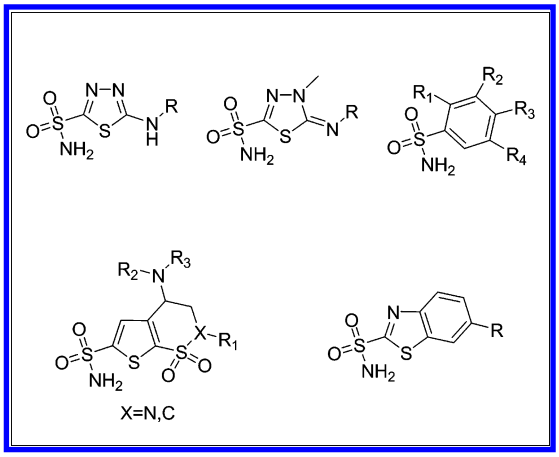


Figure 3. General scaffolds of the compounds used for the docking study.

MD was carried out at 300 K using the Langevin thermostat to maintain constant the temperature of our system.

The α carbons of the receptor were blocked with a harmonic force constant, which was decreased from 20 to 5 kcal/mol/Å^2 during the simulation. The Ramachandran plot of the complex obtained at the end of the MD simulations highlighted that no residue had a disallowed geometry, and only R7 and N247 were in generously allowed regions (see Figure 3 in the Supporting Information).

The final structure of the complex was obtained as the average of the last 3 ns of MD minimized by the CG method until a convergence of 0.05 kcal/mol·Å. The average structure of the last 3 ns of MD was obtained using the ptraj program implemented in AMBER 9. General Amber Force Field (GAFF) parameters were assigned to acetazolamide, while partial charges were calculated using the AM1-BCC method as implemented in the Antechamber suite of AMBER 9.

Docking of the CA IX Inhibitors. The ligands were built by means of Maestro[35] and were then subjected to a conformational search of 1000 steps in a water environment (using the Generalized-Born/Surface-Area model) by means of Macromodel.[36] The algorithm used was the Monte Carlo method with the MMFFs and a distance-dependent dielectric constant of 1.0. The lowest energy conformation of each ligand was then minimized using the CG method until a convergence value of 0.05 kcal/Å·mol, using the same force field and parameters as for the conformational search. Using QuACPAC,[37] special caution was given to the protonation state of ionizable groups of all ligands that were assumed to be ionized at a physiological pH (e.g., carboxylates deprotonated, amines protonated).

The ligands minimized in this manner were docked into the CA IX model by means of Gold 3.0.1.[38] The region of interest used by Gold was defined to contain the residues which stay within 7 Å of acetazolamide in the receptor model resulting from MD simulations, our previous studies suggested that this binding site radius, which takes in consideration about 30 residues, was a good value.[21]

Three water molecules that interacted in the binding site were used in the docking studies, GOLD automatically determined whether they should be bound or displaced by toggling it on and off during the docking run, and, furthermore, the orientation of the water hydrogen atoms were optimized by GOLD during docking.

Metal coordination in GOLD is modeled as 'pseudo-hydrogen bonding' in which metals can be considered to bind to H-bond acceptors and the metal will compete with H-bond donors for interaction. The zinc ion was set as a tetrahedral atom; the docking program recognized the coordination of three histidines with the zinc ion, and then in order to satisfy the tetrahedral geometry the missing coordination point was used as a fitting point that could bind to ligand acceptor groups. The "allow early termination" command was deactivated, while the possibility for the ligand to flip ring corners was activated. All the other parameters were used as Gold default values, and the ligands were submitted to 30 Genetic Algorithm runs.

The docking analysis was carried out using the ChemScore fitness function imposing the formation of an H bond between the ligands and the hydroxy group of the highly conserved residue T332.[21] Gold software is able to apply a protein hydrogen bond constraint, which can be used to specify that a particular protein atom should be hydrogen-bonded to the ligand but without specifying to which ligand atom.

The best docked conformation was then used for further studies.

This procedure has already been shown to predict the ligand poses in CA II with good reliability,[21] and the H bond constraint between the ligands and T332 was applied because our previous cross-docking studies on CA II highlighted that the ChemScore function showed good results but was unable to predict the right interaction of the sulfonamido group with the zinc ion. The addition of the H bond constraint at T199 (T332 in the CA IX) determined the correct prediction of the sulfonamido interaction.[21]

The binding site in the preliminary CA IX model was determined by aligning the 3D structure of CA IX to the X-ray structure of CA II complexed with acetazolamide (1YDA[39]) and using this ligand position as the center of the binding site. For all the other parameters, the ones described above were used.

Receptor-Based 3D-QSAR Model. *Alignment of the Molecules.* One hundred twenty-four compounds were docked inside the optimized CA IX homology model using the ChemScore fitness function imposing the formation of an H bond between the ligands and the hydroxy group of residue T332. For each ligand, the best docked structure was chosen, and this receptor-based alignment was used for further studies.

Data Set. The GOLPE program[40] was used to define a 3D-QSAR model, using GRID interaction fields[41] as descriptors (see below). The training set was composed of 87 compounds, characterized by affinity values spanning about 4 orders of magnitude. Similarly, compounds belonging to the first test set showed a pK_I affinity value ranging from 5.9 to 8.9 and were uniformly distributed along the activity range. For the second test set, we used the compounds published by Puccetti et al.[42] and Wilkinson et al.[43] Regarding the series of compounds reported by Puccetti et al. the zinc complexes of these derivatives were not taken into account.

Probe Selection. The GRID program[41] was used to describe the previously superimposed molecular structure. Interaction energies between the selected probes and each molecule were calculated using a grid spacing of 1 Å. Initially, many probes were tested, and the preliminary PLS

analyses suggested that the use of a combination of C3 (corresponding to a methyl group), N2 (corresponding to neutral flat NH_2), and O (carbonyl oxygen) probes best described the system.

Variable Selection. The MIFs of the training set were imported into the GOLPE program; it is well-known that many of the variables deriving from GRID analysis could be considered as noise, which decreases the quality of the model. For this reason, variable selection was operated by zeroing values with absolute values smaller than 0.06 kcal/mol and removing variables with a standard deviation below 0.1. Moreover, variables which either exhibited only two values or had a skewed distribution were also removed.

The smart region definition algorithm[44] was applied with 10% of the active variables as the number of seed (selected in the PLS weight space), using a critical distance cutoff of 2.5 Å and a collapsing distance cutoff of 4.0 Å. The groups were then used in the Fractional Factorial Design (FFD) procedure. FFD selection was applied until the r^2 and q^2 values did not significantly increase, using the cross-validation routine with five random sets of compounds.

Ligand-Based 3D-QSAR Model. The alignment of the compounds was developed using the ROCS 2.2 software,[45] that is a shape-similarity method based on the Tanimoto-like overlap of volumes. The alignment was developed combining the Tanimoto shape score with the color score that added the score for the appropriate overlap of groups with similar properties (donor, acceptor, hydrophobe, cation, anion, and ring). All the other parameters were used as ROCS default values. The maximum number of conformations per molecule was ~1000, generated using Omega 2.1.0.[46] Many alignments were developed using as reference structures different CA IX inhibitors; however, preliminary PLS studies revealed that the alignment obtained using compound **1** as a reference ligand gave the best results. With regards to the data set, probe, and variable selection, the same approach and parameters applied for the receptor-based 3D-QSAR study were used.

RESULTS AND DISCUSSION

Homology Modeling. As highlighted above, at present there are no experimental structures of CA IX; therefore, in order to analyze the ligand−proteins key interactions we constructed a homology model of this CA subtype.

The search for the best template for modeling was carried out by choosing X-ray structures of carbonic anhydrases possessing a high degree of sequence similarity with CA IX. The murine CA XIV proved to have the best sequence similarity with CA IX. Beyond the highest sequence similarity it also showed the higher sequence identity (39%) followed by human CA XII and CA II. This identity value could be considered of marginal value; however, as shown in Figure 1, the percentage of identities between CA XIV, CA XII, and CA II is only 34−41%, but the overall secondary structure of these three CAs is very similar.

As shown in Figure 2, the superimposition of the X-ray structures of CA XIV, CA XII, and CA II highlights that the folding of these CAs is highly similar and the 3D alignment between the Cα of these three anhydrases give rmsd values lower than 1.0 Å, suggesting the possibility of obtaining good homology models for this type of protein.

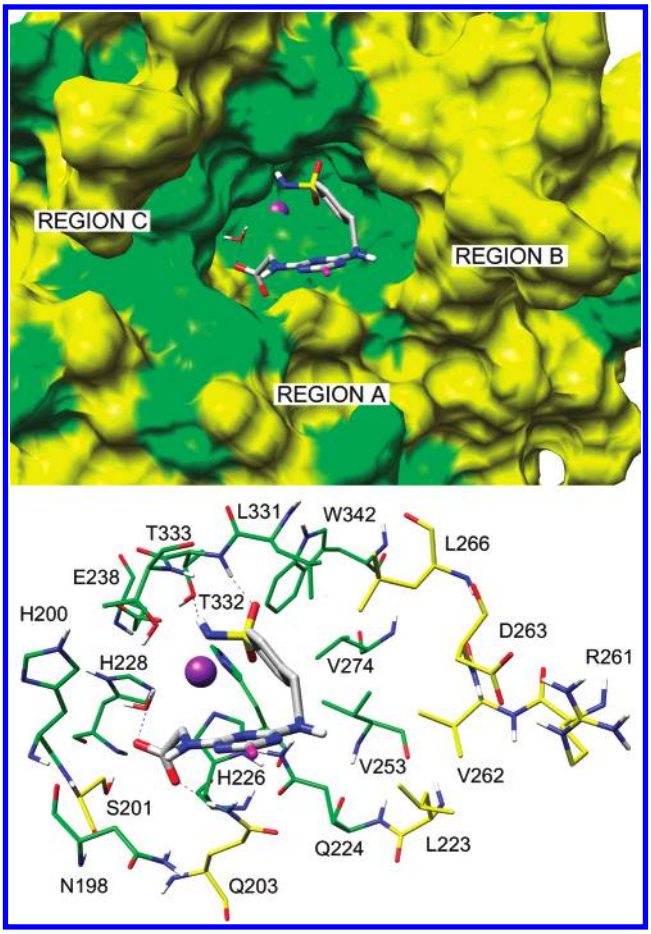


Figure 4. Molecular surface and residue analysis of the CA IX binding site. The CA IX/CA II nonconserved residues are colored yellow.

The binding site is highly conserved, among the four reported CAs the percentage of identity is about 70%, the 3D alignment between the Cα of the residues of the binding sites give rmsd values lower than 0.5 Å, and the alignment of the heavy atoms of the binding sites (for the conserved residues only) gives rmsd values of 0.6−0.7 Å.

Following the alignment shown in Figure 1 a preliminary model of CA IX was built using CA XIV as a template and was subjected to a simulated annealing protocol by means of the Modeller program.[30]

The best scored structure was analyzed, investigating in particular its backbone conformations by means of PROCHECK.[31] An analysis of the Psi/Phi Ramachandran plot of the investigated structure indicated that only residue R142 had a disallowed geometry (see Figure 1 in the Supporting Information); however, this residue is far away from the binding site and belongs to a loop region; therefore, we considered its influence to be negligible.

The CA IX inhibitor acetazolamide was then docked into the constructed model (see the Material and Methods section for the description of the docking procedure used), and the complex was then subjected to 6.2 ns of molecular dynamics (MD) simulation (see the Material and Methods section for details). Six nanoseconds of a constant pressure MD simulation should be a relative long simulation time that should ensure the achievement of a stabilized system.

As shown in Figure 4 in the Supporting Information, after about 300 ps of MD, the system reached an apparent

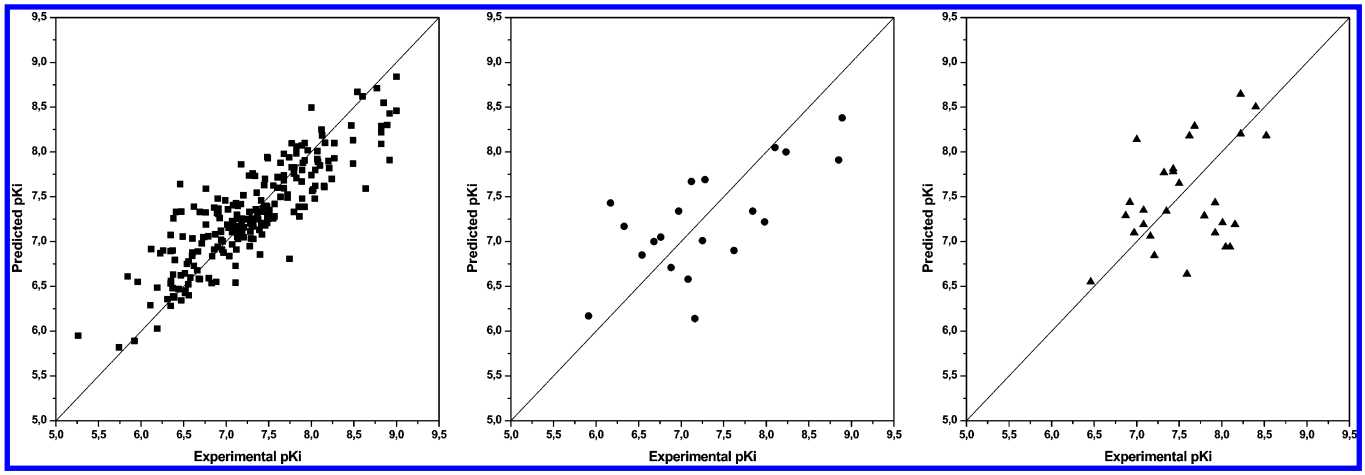


Figure 5. Plots of the 3D-QSAR model: experimental/predicted pK_I is reported. On the left side the training set (■) is reported, in the center the first test set (●) is reported, and on the right side the second test set (▲) is reported.

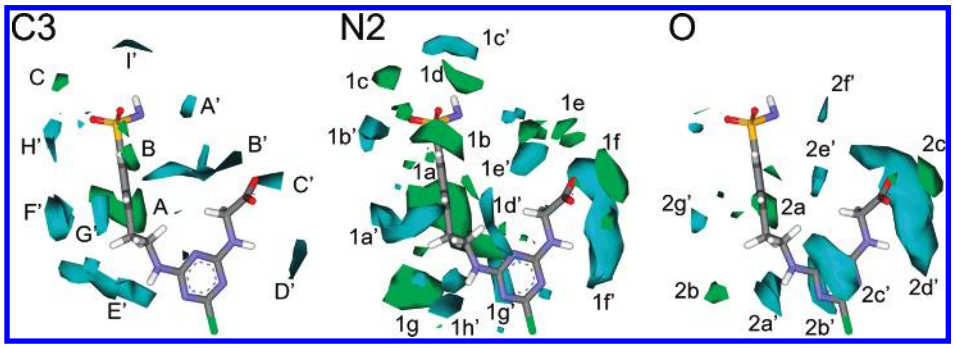


Figure 6. Negative (cyan) and positive (green) regions of the PLS coefficient plot obtained with the C3, N2, and O probes. Compound **1** is also displayed as a reference structure.

equilibrium since the total energy for the remaining picoseconds remained constant. Analyzing the root-mean-square deviation (rmsd) from the initial model of all the heavy atoms of the proteins, we observed that after an initial increase, during the last 3 ns, the rmsd remained approximately constant, around the range of 1.2−1.4 Å (see Figure 5 in the Supporting Information). Therefore this analysis justifies the relative long length of the MD simulation; however, the acetazolamide disposition inside the binding site during the whole MD simulation appeared to change slightly since the rmsd of the heavy atom of the ligand between the starting and the final structure of the MD simulation was 0.5 Å.

The analysis of the Psi/Phi Ramachandran plot of the minimized average structure of the last 3.0 ns of MD indicated that no residue had a disallowed geometry, and only R7 was in a generously allowed region (see Figure 6 in the Supporting Information). Furthermore this residue is far away from the binding site and belongs to a loop region.

Automated Docking of CA IX Inhibitors. In order to investigate the ligand−protein interactions, 124 CA IX inhibitors were retrieved from literature sources and docked into the receptor model.[47−50]

Their inhibitory activity spanned about 4 orders of magnitude (the inhibition data ranged from 1 nM to 5 μM, expressed as K_I) and, as shown in Figure 3, these inhibitors are characterized by five different central scaffolds.

Very recently, we reported an exhaustive computational analysis of CA II.[21] In that paper, using docking, cross-docking, and "receptor-based" 3D-QSAR methodologies, we evaluated the reliability of Gold software to predict the binding orientation of CA inhibitors, and we obtained particularly good results using a constrained ChemScore fitness function (see the Material and Methods section for details).

Therefore, in this study all the analyzed compounds were docked into CA IX using the same procedure that was shown to be the best one among those tested for CA II.

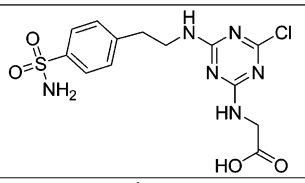
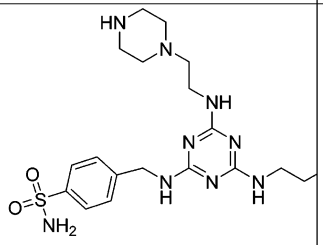
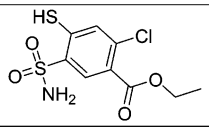
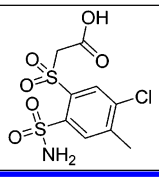
Figure 4 shows the CA IX model with the highlighted conserved region (green) and the nonconserved region (yellow) between the constructed model and CA II; this picture also shows the best docking pose for compound **1** (see Table 1) into the CA IX binding site. This compound is the most CA IX active and CA IX/CA II selective inhibitor among those analyzed.

From this analysis, we observed that the region near the zinc ion is highly conserved, whereas at the entrance of the binding site cavity two regions, indicated as A and B, are not conserved. A third region (indicated as region C in the figure) could also be cited even if it is far from the zinc binding site.

The region A is constituted by the presence in the CA IX of the not conserved residues S201 and Q203 which are substituted by alanine and asparagine in CA II. Region B is mainly formed by residues L223 (isoleucine in CA II), R261 (aspartate in CA II), V262 (phenylalanine in CA II), D263 (glycine in CA II), and L266 (valine in CA II).

Figure 4 also illustrates the binding site interaction of compound **1**: in agreement with the published X-ray structures of CAs complexed with sulfonamido derivatives, the nitrogen of the sulfonamido group interacts with the zinc

Table 1. Compounds Reported in the 3D-QSAR Analysis

Cpd.	Structure	Ki (nM) hCA IX
1[47]		1.0
2[47]		274
3[50]		56
4[50]		5440

ion and forms an H bond with the hydroxy group of T332, and the sulfonyl group forms an H bond with the backbone nitrogen of T332. The phenyl ring is stabilized by lipophilic interactions with V253 and L331, and the chloro-1,3,5-triazine group is directed toward the entrance of the binding site cavity. Finally, very interestingly, the carboxylate group forms an H bond with a structural water molecule and with the nonconserved residue Q203. This unique interaction could partially explain the high CA IX/CA II selectivity of this compound since this residue is not conserved in CA II.

Receptor-Based 3D-QSAR Analysis. In order to obtain quantitative results, the docking poses of the analyzed ligands were used as a receptor-based alignment for 3D-QSAR analysis.

The reliability of the 3D-QSAR model was characterized by its correlation coefficient (r^2), predictive correlation coefficient (q^2), and the random groups cross-validated standard deviation of errors of prediction ($SDEP_{CV}$). Compounds of the original set were divided into a training set and a test set, according to the usual guidelines. In detail, either the training or the test set should contain compounds in such a way to maximize their structural diversity (that is, compounds belonging to the training set and to the test set should be representative of the molecular diversity of all the compounds under study) and uniformly span over the whole range of activity. As a result, the training set was composed of 87 compounds,[47−50] while an external validation set (test set) of 20 molecules[47−50] was selected to test the predictive power of the model. Furthermore we also used a second external test set of 17 compounds that corresponded to sulfonamide compounds published by Puccetti et al.[42] and Wilkinson et al.,[43] and none of these compounds was included in the training set.

Table 2. Statistical Results of the Receptor-Based 3D-QSAR Model

Vars	PC	r^2	q^2	$SDEP_{cv}$	$SDEP_{TS1}$	r^2_{TS1}	$SDEP_{TS2}$	r^2_{TS2}
1884	2	0.82	0.63	0.54	0.63	0.41	0.50	0.18
1884	3	0.89	0.74	0.46	0.58	0.46	0.52	0.15
1884	4	0.93	0.74	0.46	0.64	0.38	0.55	0.15

As reported in the Material and Methods section, the final 3D-QSAR model was developed using the C3, N2, and O probes.

Table 2 lists the main data of the 3D-QSAR analysis. The third PLS component explained 89% of variance and was predictive ($q^2 = 0.74$ SDEP of test sets = 0.58 and 0.52); the fourth PLS component seemed to slightly improve only the correlation coefficient (r^2), whereas the ability of predicting the activity of the two external test sets was made worse ($r^2 = 0.93$, $q^2 = 0.74$ and the SDEP of the test set = 0.64 and 0.55). Thus the model's optimal dimensionality was given by three components (see Figure 5 and Tables 1 and 2 in the Supporting Information).

One important feature of 3D-QSAR analysis is the graphic representation of the model, usually used to make its interpretation easier. In the GOLPE program[40] there are several options for displaying the final model. Among these, the PLS pseudocoefficient and the activity contribution plots are very useful. The PLS coefficient plot makes it possible to visualize favorable and unfavorable interactions between the probes and the molecules under study, while the activity contribution plot is different for every molecule within the training set, displaying spatial regions that are individually important for the selected molecule.

Figure 6 illustrates the PLS coefficient plots of the model for the C3, N2, and O probes.

The left part of Figure 6 shows the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the C3 probe which describes the lipophilic interactions. There are three principal regions (A−C) with positive values, in which a favorable interaction between a substituent and the probe determines increased activity, whereas an unfavorable interaction between a substituent and the probe determines a decrease in activity. Moreover there are nine main regions (A′−I′) with negative values, in which a favorable interaction between a substituent and the probe determines a decrease in activity, whereas an unfavorable interaction between a substituent and the probe determines an increase in activity.

In the center part of Figure 6, the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the N2 probe describe polar interactions with proton donor characteristics. There are eight principal regions (1a′−1h′) with negative values, in which a favorable interaction between a substituent and the probe determines an increase in activity, whereas an unfavorable interaction between a substituent and the probe determines a decrease in activity, and there are seven positive regions (1a−1g) where a favorable interaction between a substituent and the probe determines a decrease in activity. Regions 1a and 1b correspond to the A and B regions of the C3 probe suggesting the important lipophilic role of these two regions. Finally the right part of Figure 6 reports the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the O2 probe which describes polar interactions

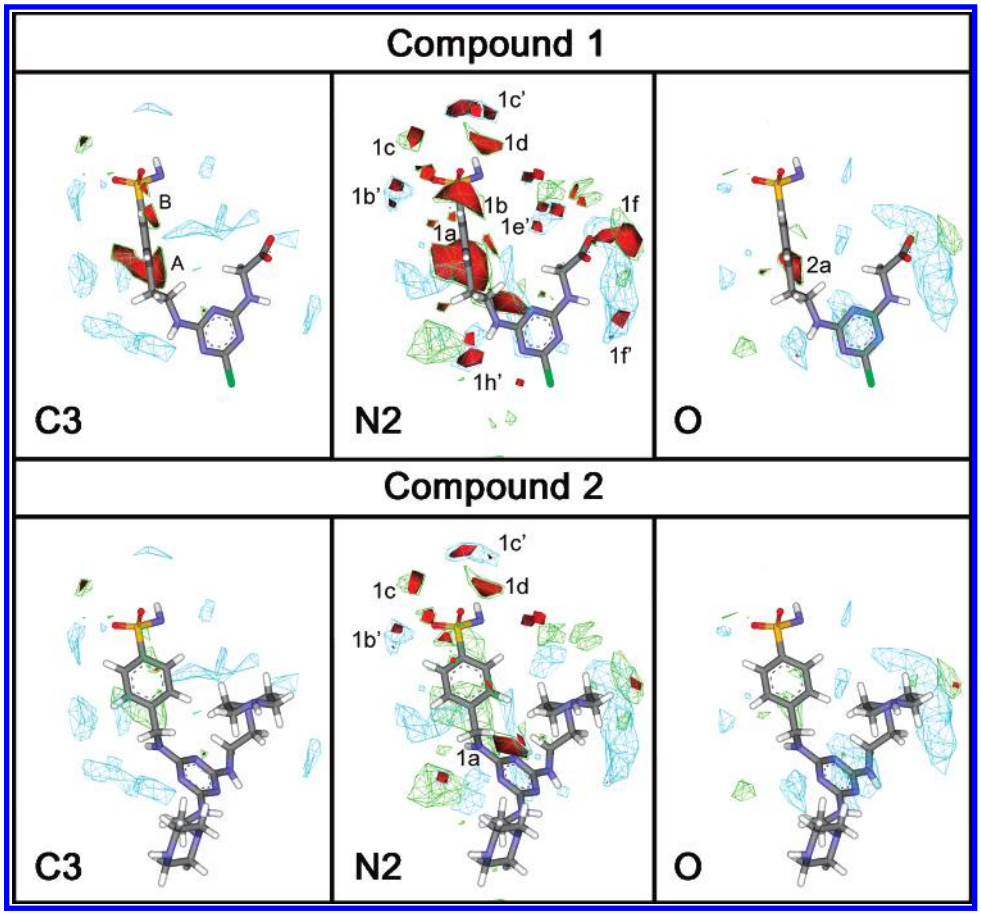


Figure 7. Positive activity contribution plots (red polyhedrons) for C3, N2, and O probes of compounds **1**−**2** embedded in the negative (cyan) and positive (green) regions of the PLS coefficient plot.

with proton acceptor characteristics. There are seven principal regions (2a′−2g′) with negative values and three regions (2a−2c) with positive values; in particular, the 2a region corresponds to the A region of the C3 probe and to the 1a region of the N2 probe.

Figure 7 illustrates the PLS coefficient for the C3, N2, and O probes as a triangle mesh, and compounds **1** and **2** (see Table 1), embedded in their positive activity contribution plots, are displayed as red polyhedrons. While the PLS coefficient plots give a global interpretation of the 3D QSAR model, the activity contribution plots allow display of spatial regions that are individually important for the selected molecules.

The phenyl ring of compound **1** favorably interacts with regions A and B of the C3 probe, the sulfonamide group favorably interacts with regions 1b′ and 1c′, the aminoacetyl substituent interacts with 1e′ and 1f′ regions, and the aminoethyl group interacts with 1h′ of the N2 probe. Finally, the interaction with the regions 1a and 1b (N2 probe) and 2a (O probe) confirm the importance of the lipophilic interaction of the phenyl ring.

Compound **2** is 274-fold less active than **1** and mainly differs with the presence of the 4,6-bis(2-(piperazin-1-yl)-ethylamino)-1,3,5-triazine instead of the 4-(2-carboxylato-ethylamino)-6-chloro-1,3,5-triazine core. The activity contribution plots for this compound highlight a lack of lipophilic interaction (see the C3 plot in Figure 7); with regards to the N2 probe, only interactions with the sulfonamido group were observed (regions 1b′, 1c′, 1c, and 1d), whereas no activity contribution plots were shown for the O probe.

Compound **3** (see Table 1) shows a CA IX inhibition activity of 56 nM, the C3 activity contribution plot (see Figure 8) suggests that the phenyl ring interacts with region B, and this lipophilic interaction was also confirmed by the interaction with regions 1a and 1b of the N2 probe. Moreover, for the N2 probe, the sulfonamido group interacts with regions 1c′ and 1d′, and the ester function interacts with the 1f′ region. There are no interactions with the O probe.

Compound **4** (see Table 1) is the least active inhibitor (K_I = 5440 nM) among those tested and, as shown in Figure 8, does not possess any lipophilic interaction. The sulfonamide group appears to be upset with respect to other common inhibitors, and this fact determines the lack of the interaction with the region of the N2 probe near this group; the only N2 region important for this compound was the 1f′ that interacts with the carboxylate substituent. In contrast, the O probe shows in Figure 8 that no regions are important for this compound.

As the alignment of the ligands was performed using the structures docked into the CA IX homology model, it would be useful to check for matching between the CA IX protein and the 3D-QSAR maps.

In Figure 9, the binding site of the CA IX overlaps with the positive 3D PLS coefficient maps of C3 probe and negative coefficient maps of N2 and O probes. As regards the C3 probe, regions A and B are in the proximity of V253 and L331, respectively, purporting an important role for these residues (see Figure 9A).

As shown in Figure 9B for the N2 probe, the negative region 1d′ is situated between Q203 and Q224, region 1e′

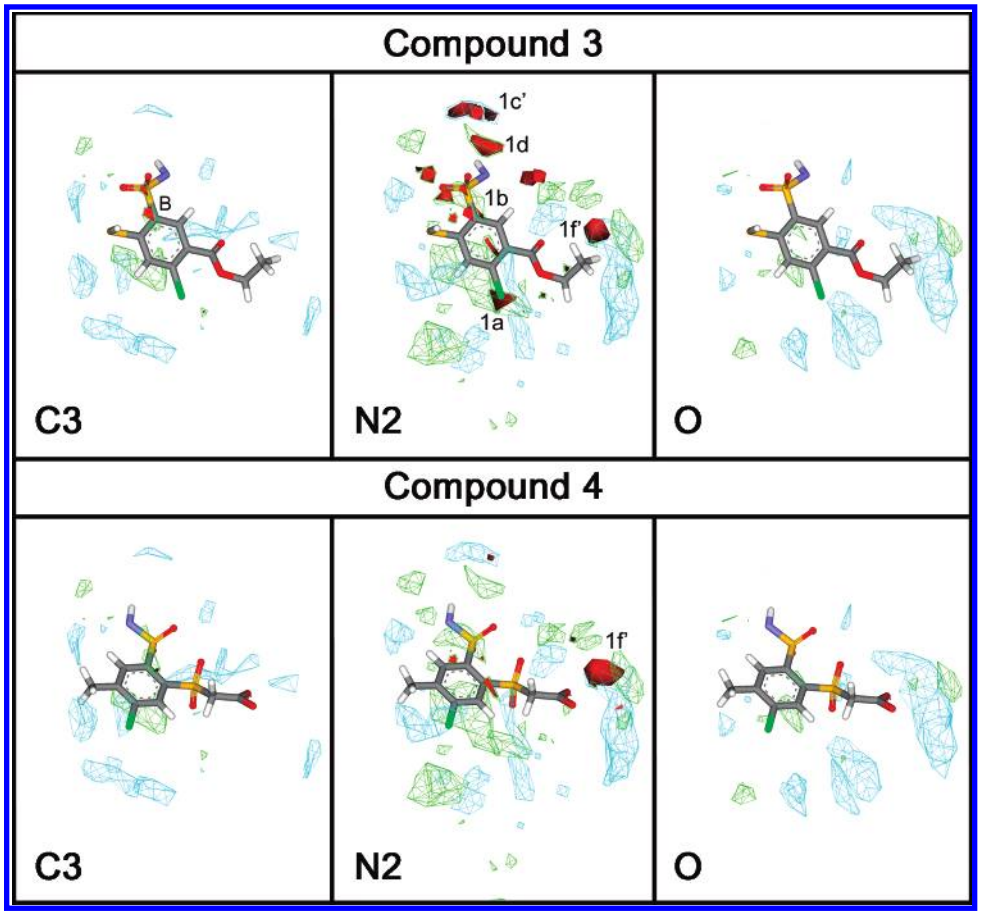


Figure 8. Positive activity contribution plots (red polyhedrons) for C3, N2, and O probes of compounds **3**−**4** embedded in the negative (cyan) and positive (green) regions of the PLS coefficient plot.

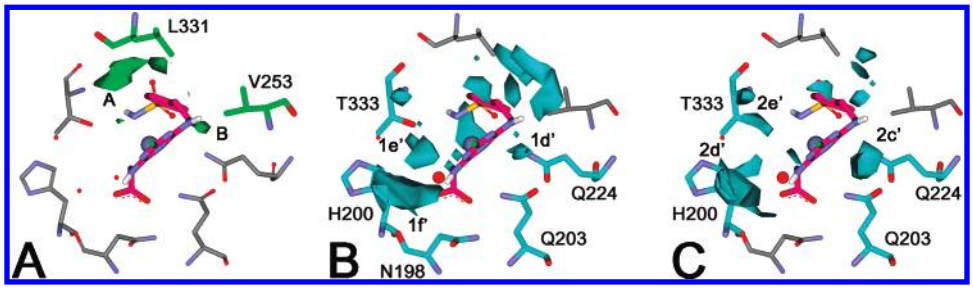


Figure 9. PLS coefficient plots obtained with the C3, N2, and O probes superimposed on the CA IX binding site.

corresponds to T333 and a structural water molecule, and, finally, region 1f′ corresponds to N198 and H200. On the other hand, the probe O region 2c′ is in proximity of Q203 and Q224, 2e′ corresponds to T333, and 2d′ corresponds to H200.

Ligand-Based 3D-QSAR Analysis. The results obtained for the receptor-based 3D-QSAR study were then compared with a ligand-based model. The alignment of the molecules was developed using ROCS.[45] This software, starting from a reference structure, is able to compare the shape of the molecules using a smooth Gaussian function. The alignment was developed using compound **1** as a reference structure. The 3D-QSAR model was developed using C3, N2, and O probes; as displayed in Table 3, the ligand-based model showed good correlation coefficients (r^2 = 0.85, 0.91 and 0.94), but with respect to the receptor-based 3D-QSAR model it showed worse predictive correlation coefficients (q^2 = 0.61, 0.71, 0.71) and a worse ability for the prediction of the activity of the external test sets.

Table 3. Statistical Results of the Ligand-Based 3D-QSAR Model

Vars	PC	r^2	q^2	$SDEP_{cv}$	$SDEP_{TS1}$	r^2_{TS1}	$SDEP_{TS2}$	r^2_{TS2}
3658	2	0.85	0.61	0.56	0.64	0.42	0.52	0.05
3658	3	0.91	0.71	0.48	0.67	0.37	0.56	0.05
3658	4	0.94	0.71	0.48	0.69	0.34	0.56	0.05

CA IX/CA II Selectivity Analysis. The receptor-based 3D-QSAR study herein reported can be compared with the one recently developed for CA II, in order to analyze the binding site characteristics important for the selectivity. However, this analysis should only give indications about the selectivity since the two 3D-QSAR studies were not developed using the same training set of inhibitors.

For CA II the 3D-QSAR model suggests an important role for three residues in particular: F131 for lipophilic interactions and Q92 and T200 for electrostatic contributions.

The lipophilic 3D-QSAR region corresponding in the CA II to the interaction of F131 was not displayed by the CA IX 3D-QSAR model, and this lack is probably due to the

substitution in this subtype of phenylalanine with V262.

The interaction of Q92 and T200 (Q217 and T333 in CA IX) was also shown by the CA IX 3D-QSAR model, which also highlights the important role of the nonconserved residue Q203 (asparagine in CA II) localized in the A region of Figure 4.

In the B region of Figure 4, our results highlight that except for V262 (F131 in the CA II), the other nonconserved residues did not seem to heavily influence the ligand hydrophobic interactions. The MD simulation highlights that during the last 3 ns of the MD simulation the two nonconserved polar residues of region B (R261 and D263) interact each other forming two H bonds, therefore, limiting their interaction with the ligands (see Figure 7 in the Supporting Information).

CONCLUSIONS

CA IX has become an interesting target for the imaging and treatment of tumors; however, drug design of new inhibitors for this target have been aided by only a few computational studies.

In this paper, using the X-ray structure of murine CA XIV as a template, we constructed a homology model of CA IX. One hundred twenty-four ligands were then docked into the CA model applying a computational approach that has already been shown to give good results for CA II,[21] and the best poses were used for developing a receptor-based 3D-QSAR.

The results obtained allowed us to extract both qualitative and quantitative information about the ligand−receptor interactions.

The superposition of the CA IX model with the X-ray structure of CA II revealed, in particular, the presence in the binding site of two nonconserved regions that could be considered useful for the developing of selective CA IX/CA II inhibitors.

The 3D-QSAR model suggests that both lipophilic and electrostatic interactions play a fundamental role in determining the ligand affinity. The superimposition between the PLS coefficient surfaces and the CA IX structure pointed out that the highly conserved residues V253 and L331 showed key hydrophobic interactions, while the electrostatic interactions of five residues (N198, H200, Q203, Q224, T333) were suggested as important for the ligand affinity.

These results enriched the computational data about CA IX, suggesting that the structural peculiarities are useful for the design of new CA IX active and CA IX/CA II selective ligands; furthermore, the results confirm the reliability of our computational approach for studying CA systems.

ACKNOWLEDGMENT

Many thanks are due to Prof. Gabriele Cruciani and Prof. Sergio Clementi (Molecular Discovery and MIA srl) for the use of the GOLPE program in their chemometric laboratory (University of Perugia, Italy) and for having provided the GRID program. A particular thanks goes to Prof. Gabriele Cruciani for valuable discussions and suggestions.

Supporting Information Available: Ramachandran plots of the CA IX model (Figures 1, 2, 3, and 6), analysis of the MD simulation of the CA IX model (Figures 4 and 5), hydrogen bonds analysis of the R261-D263 system (Figure 7), and actual versus predicted data of the receptor-based 3D-QSAR model (Tables 1 and 2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Sly, W. S.; Hu, P. Y. Human carbonic anhydrases and carbonic anhydrase deficiencies. *Annu. Rev. Biochem.* **1995**, *64*, 375−401.
(2) Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. Carbonic anhydrase inhibitors and activators and their use in therapy. *Expert Opin. Ther. Pat.* **2006**, *16*, 1627−1664.
(3) Hilvo, M.; Tolvanen, M.; Clark, A.; Shen, B.; Shah, G. N.; Waheed, A.; Halmi, P.; Hanninen, M.; Hamalainen, J. M.; Vihinen, M.; Sly, W. S.; Parkkila, S. Characterization of CA XV, a new GPI-anchored form of carbonic anhydrase. *Biochem. J.* **2005**, *392*, 83−92.
(4) Supuran, C. T.; Scozzafava, A.; Casini, A. Carbonic anhydrase inhibitors. *Med. Res. Rev.* **2003**, *23*, 146−189.
(5) Supuran, C. T. Carbonic anhydrase inhibitors in the treatment and prophylaxis of obesity. *Expert Opin. Ther. Pat.* **2003**, *13*, 1545−1550.
(6) Winum, J. Y.; Scozzafava, A.; Montero, J. L.; Supuran, C. T. Sulfamates and their therapeutic potential. *Med. Res. Rev.* **2005**, *25*, 186−228.
(7) Cecchi, A.; Hulikova, A.; Pastorek, J.; Pastorekova, S.; Scozzafava, A.; Winum, J. Y.; Montero, J. L.; Supuran, C. T. Carbonic anhydrase inhibitors. Design of fluorescent sulfonamides as probes of tumor-associated carbonic anhydrase IX that inhibit isozyme IX-mediated acidification of hypoxic tumors. *J. Med. Chem.* **2005**, *48*, 4834−4841.
(8) Tureci, O.; Sahin, U.; Vollmar, E.; Siemer, S.; Gottert, E.; Seitz, G.; Parkkila, A. K.; Shah, G. N.; Grubb, J. H.; Pfreundschuh, M.; Sly, W. S. Human carbonic anhydrase XII: cDNA cloning, expression, and chromosomal localization of a carbonic anhydrase gene that is overexpressed in some renal cell cancers. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7608−7613.
(9) Thiry, A.; Dogne, J. M.; Masereel, B.; Supuran, C. T. Targeting tumor-associated carbonic anhydrase IX in cancer therapy. *Trends Pharmacol. Sci.* **2006**, *27*, 566−573.
(10) Pastorekova, S.; Parkkila, S.; Pastorek, J.; Supuran, C. T. Carbonic anhydrases: current state of the art, therapeutic applications and future prospects. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 199−229.
(11) Vicker, N.; Ho, Y.; Robinson, J.; Woo, L. L.; Purohit, A.; Reed, M. J.; Potter, B. V. Docking studies of sulphamate inhibitors of estrone sulphatase in human carbonic anhydrase II. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 863−865.
(12) Ho, Y. T.; Purohit, A.; Vicker, N.; Newman, S. P.; Robinson, J. J.; Leese, M. P.; Ganeshapillai, D.; Woo, L. W.; Potter, B. V.; Reed, M. J. Inhibition of carbonic anhydrase II by steroidal and non-steroidal sulphamates. *Biochem. Biophys. Res. Commun.* **2003**, *305*, 909−914.
(13) Gruneberg, S.; Wendt, B.; Klebe, G. Subnanomolar Inhibitors from Computer Screening: A Model Study Using Human Carbonic Anhydrase II. *Angew. Chem., Int. Ed. Engl.* **2001**, *40*, 389−393.
(14) Gruneberg, 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*, 3588−3602.
(15) Clare, B. W.; Supuran, C. T. QSAR studies of sulfonamide carbonic anhydrase inhibitors. In *Carbonic Anhydrase, its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC Press: 2004; pp 149−182.
(16) Agrawal, V. K.; Singh, J.; Khadikar, P. V.; Supuran, C. T. QSAR study on topically acting sulfonamides incorporating GABA moieties: a molecular connectivity approach. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2044−2051.
(17) Thakur, A.; Thakur, M.; Khadikar, P. V.; Supuran, C. T.; Sudele, P. QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index. *Bioorg. Med. Chem.* **2004**, *12*, 789−793.
(18) Hillebrecht, A.; Supuran, C. T.; Klebe, G. Integrated approach using protein and ligand information to analyze selectivity- and affinity-determining features of carbonic anhydrase isozymes. *Chem. Med. Chem.* **2006**, *1*, 839−853.
(19) Weber, A.; Bohm, M.; Supuran, C. T.; Scozzafava, A.; Sotriffer, C. A.; Klebe, G. 3D QSAR Selectivity Analyses of Carbonic Anhydrase Inhibitors: Insights for the Design of Isozyme Selective Inhibitors. *J. Chem. Inf. Model.* **2006**, *46*, 2737−2760.
(20) Huang, H.; Pan, X.; Tan, N.; Zeng, G.; Ji, C. 3D-QSAR study of sulfonamide inhibitors of human carbonic anhydrase II. *Eur. J. Med. Chem.* **2007**, *42*, 365−372.
(21) Tuccinardi, T.; Nuti, E.; Ortore, G.; Supuran, C. T.; Rossello, A.; Martinelli, A. Analysis of human carbonic anhydrase II: docking reliability and receptor-based 3D-QSAR study. *J. Chem. Inf. Model.* **2007**, *47*, 515−525.
(22) Thiry, A.; Ledecq, M.; Cecchi, A.; Dogne, 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*, 2743−2749.
(23) Alterio, V.; Vitale, R. M.; Monti, S. M.; Pedone, C.; Scozzafava, A.; Cecchi, A.; De Simone, G.; Supuran, C. T. Carbonic anhydrase inhibitors: X-ray and molecular modeling study for the interaction of a fluorescent antitumor sulfonamide with isozyme II and IX. *J. Am. Chem. Soc.* **2006**, *128*, 8329−8335.
(24) Jaiswal, M.; Khadikar, P. V.; Scozzafava, A.; Supuran, C. T. Carbonic anhydrase inhibitors: the first QSAR study on inhibition of tumor-associated isoenzyme IX with aromatic and heterocyclic sulfonamides. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3283−3290.
(25) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784−3788.
(26) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(27) Whittington, D. A.; Grubb, J. H.; Waheed, A.; Shah, G. N.; Sly, W. S.; Christianson, D. W. Expression, assay, and structure of the extracellular domain of murine carbonic anhydrase XIV: implications for selective inhibition of membrane-associated isozymes. *J. Biol. Chem.* **2004**, *279*, 7223−7228.
(28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235−242.
(29) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(30) Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753−1773.
(31) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thorton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(32) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
(33) Merz, K. M., Jr; Murcko, M. A.; Kollman, P. A. Inhibition of Carbonic Anhydrase. *J. Am. Chem. Soc.* **1991**, *113*, 4484−4490.
(34) Essman, U.; Perela, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8592.
(35) *Maestro, version 7.5*; Schrödinger Inc.: Portland, OR, 1999.
(36) *Macromodel, version 8.5*; Schrödinger Inc.: Portland, OR, 1999.
(37) *QuACPAC 1.1*; OpenEye Scientific Software Inc.: 9 Bisbee Ct, Suite D, Santa Fe, NM 87508, 2007.
(38) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(39) Nair, S. K.; Krebs, J. F.; Christianson, D. W.; Fierke, C. A. Structural basis of inhibitor affinity to variants of human carbonic anhydrase II. *Biochemistry* **1995**, *34*, 3981−3989.
(40) *GOLPE 4.5*; Multivariate Infometric Analysis Srl.: Viale dei Castagni 16, Perugia, Italy, 1999.
(41) *GRID version 22a*; Molecular Discovery Ltd.: West Way House, Elms Parade, Oxford, U.K., 2004.
(42) Puccetti, L.; Fasolis, G.; Vullo, D.; Chohan, Z. H.; Scozzafava, A.; Supuran, C. T. Carbonic anhydrase inhibitors. Inhibition of cytosolic/tumor-associated carbonic anhydrase isozymes I, II, IX, and XII with Schiff's bases incorporating chromone and aromatic sulfonamide moieties, and their zinc complexes. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3096−3101.
(43) Wilkinson, B. L.; Bornaghi, L. F.; Houston, T. A.; Innocenti, A.; Vullo, D.; Supuran, C. T.; Poulsen, S. A. Carbonic anhydrase inhibitors: inhibition of isozymes I, II, and IX with triazole-linked O-glycosides of benzene sulfonamides. *J. Med. Chem.* **2007**, *50*, 1651−1657.
(44) Pastor, M.; Cruciani, G.; Clementi, S. Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J. Med. Chem.* **1997**, *40*, 1455−1464.
(45) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and its Application to a Bacterial Protein-Protein Interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.
(46) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graphics Modell.* **2003**, *21*, 449−462.
(47) Garaj, V.; Puccetti, L.; Fasolis, G.; Winum, J. Y.; Montero, J. L.; Scozzafava, A.; Vullo, D.; Innocenti, A.; Supuran, C. T. Carbonic anhydrase inhibitors: novel sulfonamides incorporating 1,3,5-triazine moieties as inhibitors of the cytosolic and tumour-associated carbonic anhydrase isozymes I, II and IX. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3102−3108.
(48) Winum, J. Y.; Dogne, J. M.; Casini, A.; de Leval, X.; Montero, J. L.; Scozzafava, A.; Vullo, D.; Innocenti, A.; Supuran, C. T. Carbonic anhydrase inhibitors: synthesis and inhibition of cytosolic/membrane-associated carbonic anhydrase isozymes I, II, and IX with sulfonamides incorporating hydrazino moieties. *J. Med. Chem.* **2005**, *48*, 2121−2125.
(49) Wilkinson, B. L.; Bornaghi, L. F.; Houston, T. A.; Innocenti, A.; Supuran, C. T.; Poulsen, S. A. A novel class of carbonic anhydrase inhibitors: glycoconjugate benzene sulfonamides prepared by "click-tailing". *J. Med. Chem.* **2006**, *49*, 6539−6548.
(50) Saczewski, F.; Slawinski, J.; Kornicka, A.; Brzozowski, Z.; Pomarnacka, E.; Innocenti, A.; Scozzafava, A.; Supuran, C. T. Carbonic anhydrase inhibitors. Inhibition of the cytosolic human isozymes I and II, and the transmembrane, tumor-associated isozymes IX and XII with substituted aromatic sulfonamides activatable in hypoxic tumors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4846−4851.

CI700214J