Analysis of Human Carbonic Anhydrase II: Docking Reliability and Receptor-Based 3D-QSAR Study

Tiziano Tuccinardi,[†] Elisa Nuti,[†] Gabriella Ortore,[†] Claudiu T. Supuran,[‡] Armando Rossello,[†] 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 October 30, 2006

The ability of Gold software to predict the binding disposition of carbonic anhydrase (CA) inhibitors was evaluated using CA II as a case study. The best procedure was subsequently used for docking almost 300 CA II ligands, and the best poses were used as an alignment tool for the development of a 3D quantitative structure−activity relationship (QSAR) study. Evaluation of the resulting 3D-QSAR model allowed us to indicate the ligand properties and residues important for CA II inhibition. Since CAs are an important target involved in many pathologies such as glaucoma, obesity, and tumors, the results obtained could accurately predict the binding affinity of newly designed CA II inhibitors. Furthermore, it is reasonable that this strategy could be profitably used also for the investigation of other CAs.

INTRODUCTION

α-Carbonic anhydrases (CAs) are ubiquitously distributed zinc metalloenzymes that are implicated in a variety of physiological functions. A total of 16 CA isozymes are presently known in humans: four are cytoplasmic (CA I−III and VII) and two are mitochondrial (CA VA and VB), whereas others are secreted (CA VI) or are membrane-bound (CA IV, IX, XII, and XIV).[1] These enzymes catalyze the reversible hydration of CO_2 to form HCO_3^-, which is involved in many biosynthetic reactions, among which are gluconeogenesis, lipogenesis, the synthesis of certain amino acids, and pyrimidine nucleotide biosynthesis. Moreover, these enzymes are involved in pH homeostasis, bone resorption, ion transport, electrolyte secretion in a variety of tissues, calcification, and tumorigenesis.[2] Thus, it is not surprising that many of their isozymes have been discovered as important targets for inhibitors with clinical applications. Several inhibitors of human CAs have been utilized as antiglaucoma, anticonvulsant, antiurolithic, antiepileptic, and anticancer agents.

Most of these drugs (such as acetazolamide, methazolamide, or brinzolamide clinically used for the treatment of glaucoma) contain a sulfonamide group as an anchoring system to coordinate the catalytic zinc.[3,4] X-ray crystallographic structures are available for many adducts of sulfonamide inhibitors with various CA isozymes. In all these complexes, the deprotonated sulfonamide is coordinated to the Zn(II) ion and its NH moiety forms a hydrogen bond with the Oγ of Thr199, which in turn is engaged in another hydrogen bond with the carboxylate group of Glu106. Moreover, one of the SO_2 oxygen atoms participates in a hydrogen bond with the backbone NH moiety of Thr199.[5]

Typically, a CA inhibitor must possess (i) a zinc-binding function (ZBF), such as sulfonamide, sulfamate, hydroxamate, and so forth; (ii) an organic scaffold (usually an aromatic or heterocyclic moiety); and (iii) a tail attached to the scaffold that increases solubility in water.

Human CA II (hCA II) is the physiologically most relevant and widespread CA isoform that has been most extensively studied from the structural and inhibitor design points of view.[5−7] This enzyme, together with hCA IV, is involved in the pathogenesis of glaucoma, since it is present in the ciliary process of the eye, where it is responsible for an increased secretion of bicarbonate and aqueous humor that causes an elevated intraocular pressure. CA II is present also in tissues other than that of the eye, such as the blood, kidneys, lungs, central nervous system, and so forth. The available pharmacological agents possess many undesired side effects, mainly due to their lack of selectivity for the different isozymes. Thus, development of isozyme-specific or at least organ-selective inhibitors would be highly beneficial.[8]

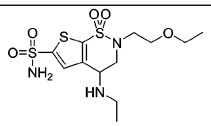
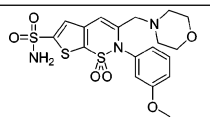
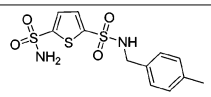
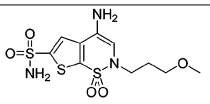
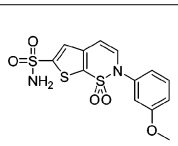
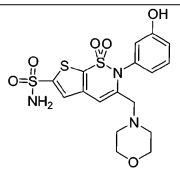
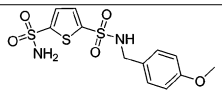
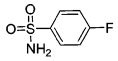
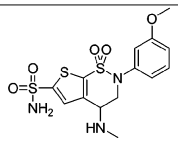
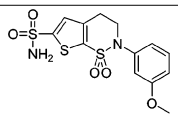
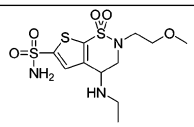
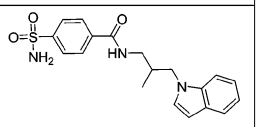
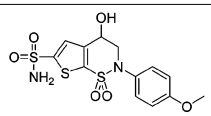
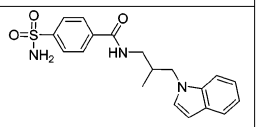
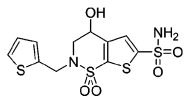
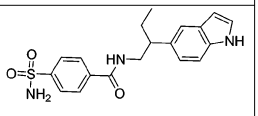
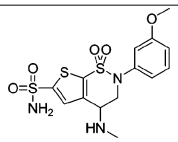
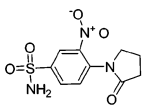
From a computational point of view, many 3D quantitative structure−activity relationship (QSAR) studies of CA inhibitors have been developed,[9,10] and furthermore, in 2001−2002, Grüneberg and co-workers used a new pharmacophore-docking strategy for successfully virtual screening novel CA II inhibitors.[11,12]

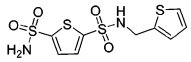
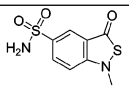
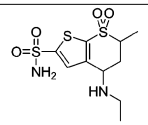
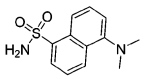
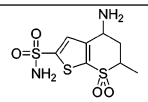
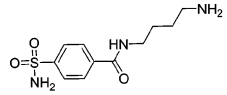
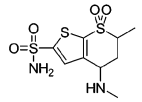
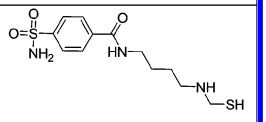
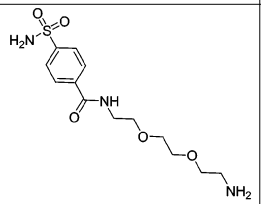
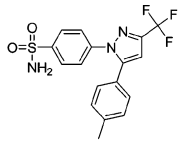
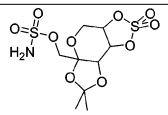
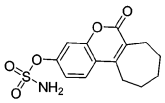
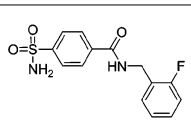
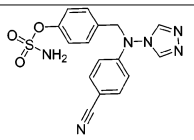
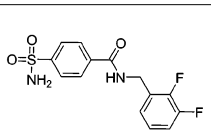
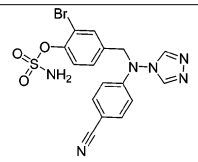
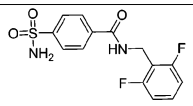
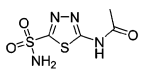
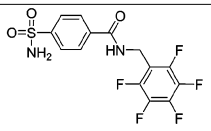
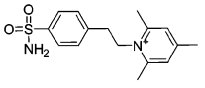
During the development of our project, Hillebrecht and co-workers explored the selectivity and affinity-determining features of CA I, II, and IV using a ligand-based 3D-QSAR and the novel AFMoc approach,[13] whereas Weber and co-workers combined ligand-based 3D-QSAR models of CA I, II, and IV with the docking methods for screening novel CA inhibitors.[14]

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

10.1021/ci600469w CCC: $37.00
Published on Web 02/13/2007

Table 1. Ligand Structures of the hCA II X-ray Complexes Used in This Study

PDB code	Ligand	PDB code	Ligand
1A42		1I8Z	
1BN1		1I90	
1BN3		1I91	
1BN4		1IF4	
1BNM		1IF5	
1BNN		1IF6	
1BNQ		1IF7	
1BNT		1IF8	
1BNU		1IF9	
1BNV		1KWQ	

PDB code	Ligand	PDB code	Ligand
1BNW		1KWR	
1CIL		1OKL	
1CIM		1OKM	
1CIN		1OKN	
1CNX		1OQ5	
1EOU		1TTM	
1G1D		1XPZ	
1G52		1XQ0	
1G53		1YDA	
1G54		1ZE8	

In this paper, using the Gold program,[15] we evaluated the reliability of various docking procedures to predict the binding disposition of CA II inhibitors. Then, using almost 300 compounds, we developed a receptor-based 3D-QSAR model of CA II.

MATERIALS AND METHODS

CA−Inhibitor Complex Structures and Docking. The following hCA II structures, taken from the Protein Data Bank (PDB),[16] were used in this work for a total of 40 proteins (see Table 1). Hydrogen atoms were added by means of Maestro 7.5[17] using the all-atom model. Three water residues in the region of Asn62−Glu106, which are present in most of the X-ray proteins analyzed, were maintained also during docking calculations.

The ligands were extracted from the X-ray complexes and were then subjected to a conformational search (CS) of 1000 steps in a water environment (using the generalized-Born/surface-area model) by means of Macromodel.[18] The algorithm used was the Monte Carlo method with the MMFFs forcefield and a distance-dependent dielectric constant of 1.0. The ligands were then minimized using the conjugated gradient method until a convergence value of 0.05 kcal/Å

Table 2. RMSD Evaluation of the Binding Disposition for the Ligands Using the Three Docking Procedures, GoldScore (Gscore), ChemScore (Cscore), and ChemScore with the Thr199 Constraint (Cscore_c)[a]

PDB	Gscore	Cscore	Cscore_c
1A42	3.3	0.8	**4.3**
1BN1	1.3	1.8	1.7
1BN3	**5.3**	3.5	3.3
1BN4	**3.8**	1.7	1.7
1BNM	3.3	**4.5**	**4.9**
1BNN	3.5	2.0	2.3
1BNQ	2.2	**4.0**	**4.0**
1BNT	3.0	1.5	3.3
1BNU	2.5	1.4	3.3
1BNV	3.0	2.7	2.7
1BNW	3.5	1.9	2.0
1CIL	0.3	3.4	3.4
1CIM	2.6	1.0	1.2
1CIN	3.1	3.1	3.0
1CNX	**5.8**	1.3	1.3
1EOU	**4.7**	**5.4**	**4.5**
1G1D	**4.4**	1.4	1.2
1G52	**4.3**	1.1	0.9
1G53	**4.6**	1.7	0.8
1G54	**5.6**	2.2	1.9
1I8Z	**4.3**	**4.0**	2.1
1I90	2.5	**4.1**	**4.0**
1I91	3.2	2.1	2.0
1IF4	1.6	0.8	1.4
1IF5	3.0	1.5	2.2
1IF6	**3.6**	2.5	2.7
1IF7	3.2	2.4	2.5
1IF8	2.8	2.5	2.3
1IF9	**5.8**	2.5	2.6
1KWQ	**4.1**	2.7	2.5
1KWR	**4.4**	2.5	2.5
1OKL	0.8	2.6	0.5
1OKM	**4.4**	1.7	1.6
1OKN	**5.5**	2.1	2.8
1OQ5	**5.9**	**5.4**	**5.4**
1TTM	2.7	2.5	2.5
1XPZ	**5.0**	2.6	1.9
1XQ0	**3.9**	2.7	2.7
1YDA	2.6	1.5	1.6
1ZE8	0.9	1.0	0.9
RMSD	3.5	2.4	2.5

[a] The values greater than 3.5 Å are reported in bold.

Table 3. RMSD Evaluation of the ZBF Disposition Using the Three Docking Procedures, GoldScore (Gscore), ChemScore (Cscore) and ChemScore with the Thr199 Constraint (Cscore_c)[a]

PDB	Gscore	Cscore	Cscore_c
1A42	0.65	0.57	0.36
1BN1	0.54	1.00	0.63
1BN3	0.78	**1.90**	0.53
1BN4	0.56	**2.00**	0.63
1BNM	0.45	1.39	0.57
1BNN	0.47	0.63	0.56
1BNQ	0.60	0.61	0.44
1BNT	0.38	0.76	0.53
1BNU	0.55	0.71	0.45
1BNV	0.46	0.66	0.49
1BNW	0.83	0.80	0.49
1CIL	0.34	0.70	0.41
1CIM	0.38	0.64	0.75
1CIN	0.38	1.28	0.47
1CNX	0.60	1.25	0.45
1EOU	0.68	**3.00**	0.66
1G1D	0.67	1.28	0.45
1G52	0.69	1.26	0.58
1G53	0.44	1.18	0.55
1G54	0.59	1.11	0.59
1I8Z	0.40	0.69	0.92
1I90	0.80	1.19	0.81
1I91	0.58	0.94	0.78
1IF4	0.81	0.89	0.71
1IF5	0.61	0.97	0.98
1IF6	**2.30**	**2.20**	**2.00**
1IF7	0.44	0.93	0.56
1IF8	0.42	0.61	0.55
1IF9	0.57	0.80	0.49
1KWQ	0.65	1.30	0.44
1KWR	0.94	0.62	0.33
1OKL	0.51	0.88	0.68
1OKM	0.60	0.96	0.59
1OKN	1.27	**2.10**	0.95
1OQ5	0.75	1.17	0.55
1TTM	**1.70**	1.26	1.11
1XPZ	1.34	**2.20**	1.50
1XQ0	0.88	**2.34**	1.30
1YDA	1.28	**1.86**	0.55
1ZE8	0.69	0.78	0.40
RMSD	0.70	1.20	0.70

[a] The values greater than 1.5 Å are reported in bold.

mol was reached, using the same force field and parameters as those used for the conformational search.

The ligands minimized in this way were docked in their corresponding proteins by means of Gold 3.0.1.[15] The region of interest used by Gold was defined in such a manner that it contains the residues which stay within 7 Å from the ligand in the X-ray structures; the zinc ion was set as a tetrahedral atom. The "allow early termination" command was deactivated while the possibility for the ligand to flip ring corners was activated. As regards all the other parameters, the Gold default ones were used, and the ligands were submitted to 30 genetic algorithm runs.

Three docking analyses were carried out: In the first two cases, the two fitness functions implemented in Gold, GoldScore, and ChemScore, were used. In the third one, the formation of an H bond between the hydroxy group of Thr199 and the ligands was also imposed, and the ChemScore function was used again. The best docked conformation was then used for further studies.

The docking results were evaluated through the comparison of the found docked positions of the ligand and the experimental ones: as a measure of docking reliability, the root-mean-square deviation (RMSD) between the positions of heavy atoms of the ligand in the calculated and experimental structures was taken into account.

Cross-Docking. Each of the 40 sulfonamide inhibitors was docked in turn into all the 40 available X-ray structures, therefore performing a total of 1600 GOLD docking calculations for each of the three above-mentioned procedures (GoldScore, ChemScore, and ChemScore plus the H bond imposed on the Thr199).

In order to evaluate these procedures, the CA II X-ray structure 1A42 was selected as reference structure; all the other X-ray complexes were aligned to it through the superimposition of the CA II backbone atoms; in such a manner, it was possible to obtain an experimental reference position for each of the 40 ligands into all of the 40 CA II X-ray structures.

The docking reliability was then evaluated by calculating for each ligand and each of the 40 X-ray CA II structures the RMSD (heavy atoms) between the above-mentioned experimental reference position and that calculated by GOLD.

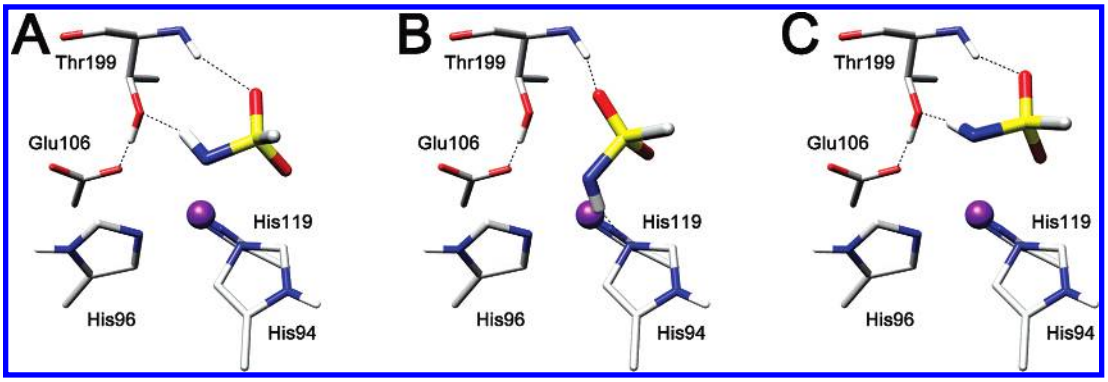


Figure 1. ZBF interaction scheme using GoldScore (A), ChemScore (B), and ChemScore with the Thr199 constraint (B).

Beyond the RMSD value, to quantify the ligand docking quality, the docking accuracy (DA) was also reported.[19,20] This parameter makes use of RMSD values and measures how accurately the ligands are docked:

$$DA = frmsd \leq 2 + 0.5\,(frmsd \leq 3 - frmsd \leq 2)$$

where the frmsd ≤ 2 indicates the fraction of ligands docked with an RMSD value less than or equal to 2 Å, whereas frmsd ≤ 3 indicates the fraction of ligands docked with an RMSD value less than or equal to 3 Å.

For the AUTODOCK 3.0[21] calculations, AUTODOCK TOOLS was used to identify the torsion angles in the ligands, add the solvent model, and assign the partial atomic charges to the protein and the ligands. As regards the catalytic zinc atom, the parameters published by Hu et al.[22,23] were used.

The regions of interest used by AUTODOCK were defined by considering the 1A42 ligand as the central group; in particular, a grid of 52, 52, and 52 points in the *x*, *y*, and *z* directions was constructed centered on the center of the mass of this inhibitor. A grid spacing of 0.375 Å and a distance-dependent function of the dielectric constant were used for the energetic map calculations.

Using the Lamarckian genetic algorithm, we subjected the docked compounds to 100 runs of the AUTODOCK search, in which the default values of the other parameters were used (see also the Supporting Information).

3D-QSAR. *Alignment of the Molecules.* Almost 300 compounds were docked inside the 1OKN CA II binding site using the ChemScore fitness function and applying the constraint at Thr199. Furthermore, the template similarity constraint function of Gold was applied for the 1,3,4-thiadiazole derivatives using the acetazolamide of 1YDA as a template, in order to prevent the overturning of the thiadiazole ring.

As regards the protein, the 1OKN X-ray structure was chosen since it displays the best average of RMSD (2.2 Å) and DA (0.69) values in the cross-docking analysis.

For each ligand, the best docked structure was chosen, and this receptor-based alignment was used for further studies.

Data Set. The GOLPE program[24] was used to define a 3D-QSAR model, using GRID interaction fields[25] as descriptors (see below). The training set was composed of 220 compounds, characterized by affinity values spanning about 3 orders of magnitude. Similarly, compounds belonging to the first test set showed a pK_i affinity value ranging from 6.3 to 8.7 and were uniformly distributed along the activity range. As regards the second and third test sets, we used the compounds published by Ilies et al.[26] and Scozzafava et al.[27]

Probe Selection. The GRID program[25] 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, the combination of C3 (corresponding to a methyl

Table 4. Backbone Superimposition of the Whole Proteins (WP RMSD) and Heavy Atom Superimposition of the Binding Sites (BS RMSD) Using the X-ray Structure 1A42 as Reference Protein[a]

PDB	WP RMSD	BS RMSD
1A42	0.00	0.00
1BN1	0.22	0.29
1BN3	0.22	0.28
1BN4	0.22	0.43
1BNM	0.25	0.50
1BNN	0.23	0.43
1BNQ	0.16	0.17
1BNT	0.17	0.20
1BNU	0.17	0.62
1BNV	0.24	0.30
1BNW	0.19	0.25
1CIL	0.26	0.29
1CIM	0.24	0.61
1CIN	0.24	0.56
1CNX	0.25	0.57
1EOU	0.39	0.61
1G1D	0.25	0.37
1G52	0.27	0.36
1G53	0.24	0.37
1G54	0.25	0.48
1I8Z	0.30	0.36
1I90	0.30	0.36
1I91	0.31	0.42
1IF4	0.31	0.40
1IF5	0.35	0.42
1IF6	0.32	0.46
1IF7	0.26	0.61
1IF8	0.32	0.38
1IF9	0.31	0.60
1KWQ	0.30	0.40
1KWR	0.27	0.38
1OKL	0.26	0.45
1OKM	0.23	0.55
1OKN	0.31	0.65
1OQ5	0.33	0.42
1TTM	0.38	0.46
1XPZ	0.38	0.49
1XQ0	0.39	0.48
1YDA	0.28	1.61
1ZE8	0.39	0.34

[a] All the reported values are in angstroms.

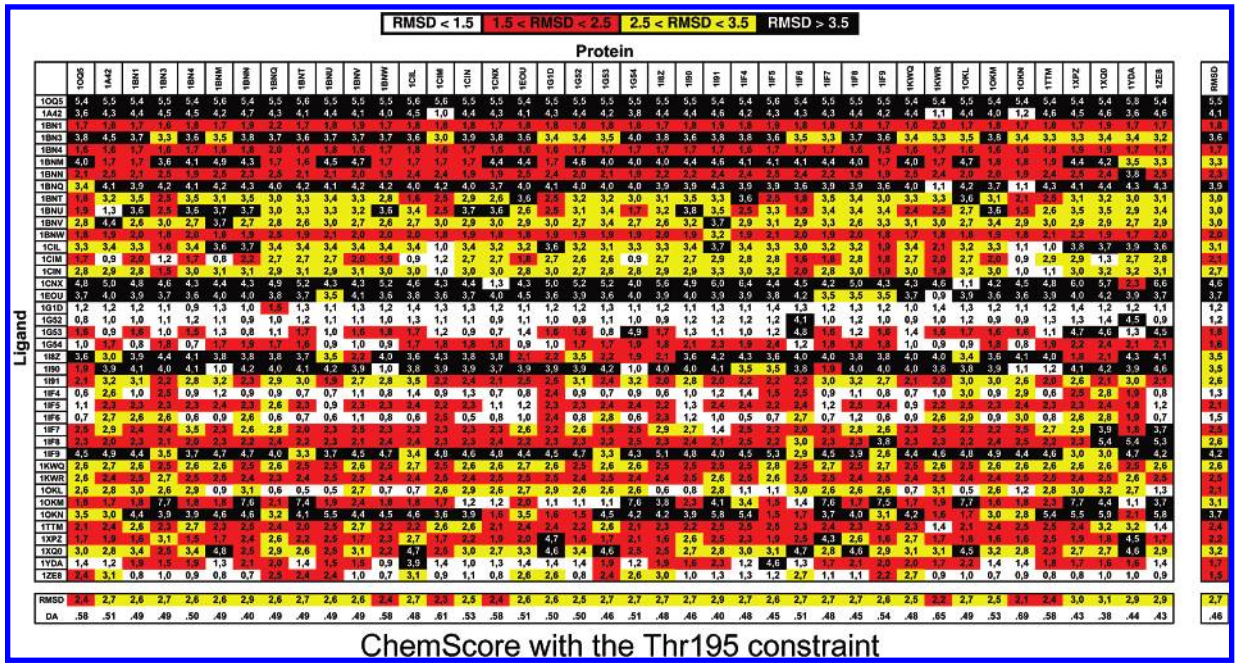


Figure 2. Matrix of RMSDs obtained by the cross-docking study using ChemScore with the Thr199 constraint.

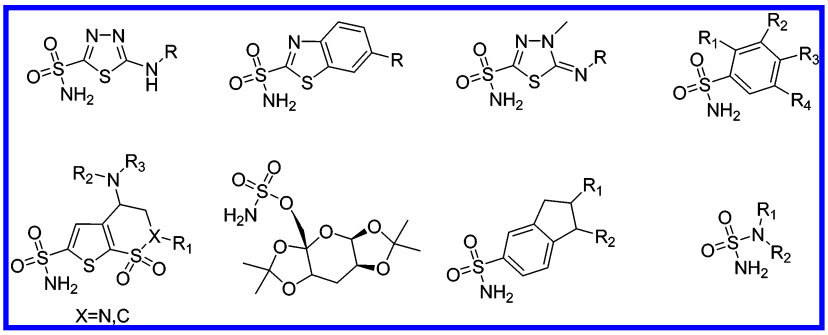


Figure 3. General scaffolds of the compounds used for the 3D-QSAR study.

group), N2 (corresponding to neutral, flat NH_2), and O (carbonyl oxygen) probes was used to calculate the molecular interaction fields (MIFs), in order to evaluate the lipophilic and H-bond donor/acceptor properties of the ligands. Preliminary partial least-squares (PLS) analyses suggested that the use of the O probe did not improve the model; thus, it was not taken into account.

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 exhibited only two values and had a skewed distribution were also removed.

The smart region definition algorithm[28] was applied with 10% of the active variables as the number of seeds (selected in the PLS weights space), 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 twice, until the r^2 and q^2 values did not increase significantly, using the cross-validation routine with five random sets of compounds.

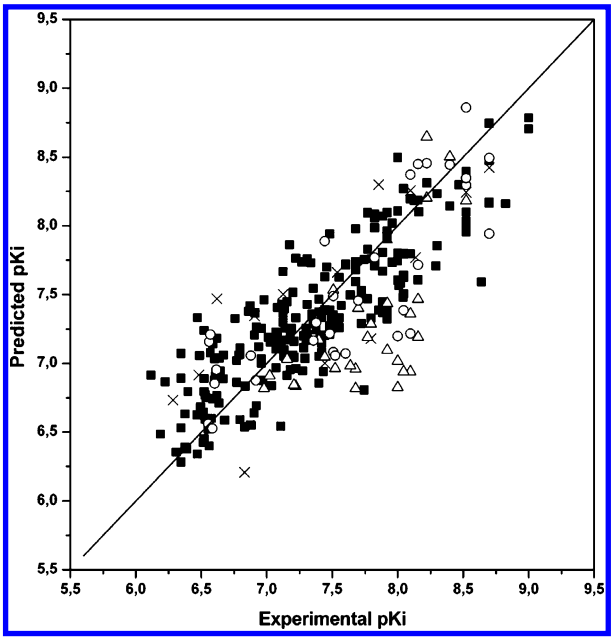


Figure 4. Plot of the 3D-QSAR model: experimental/predicted pK_i is reported; the test sets are represented as ×, △, and ○.

RESULTS AND DISCUSSION

A total of 40 CA II inhibitors were docked into their crystal structures through the Gold program[15] after an extensive CS

Table 5. Statistical Results of the 3D-QSAR Model

vars	PC	r^2	q^2	$SDEP_{cv}$	$SDEP_{TS1}$	$SDEP_{TS2}$	$SDEP_{TS3}$
1357	3	0.86	0.70	0.34	0.44	0.62	0.43
1357	4	0.88	0.72	0.33	0.44	0.63	0.41
1357	5	0.89	0.71	0.34	0.44	0.63	0.42

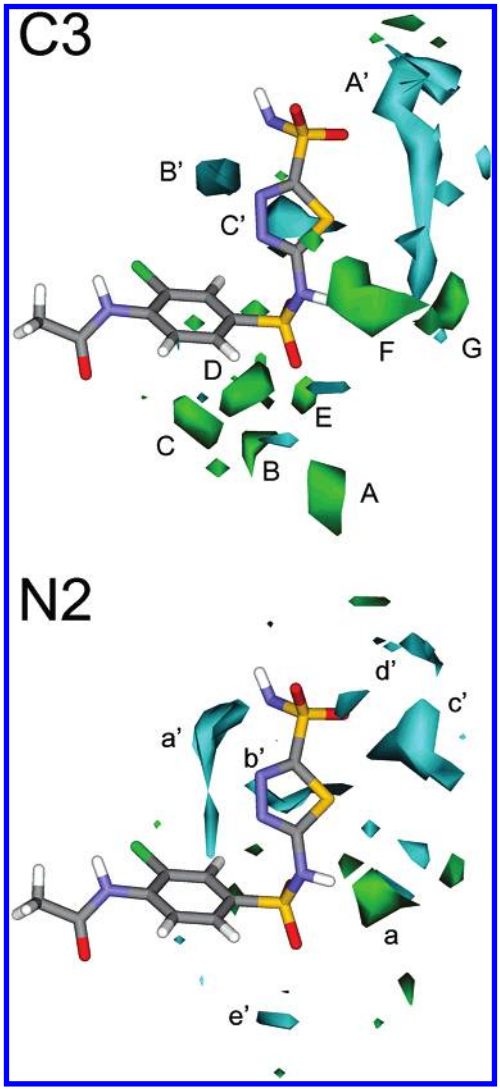


Figure 5. Negative (cyan) and positive (green) regions of the PLS coefficient plot obtained with the C3 and N2 probes.

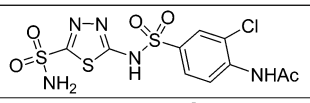
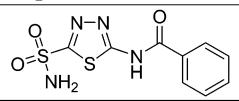
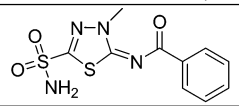
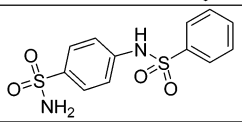
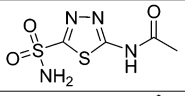
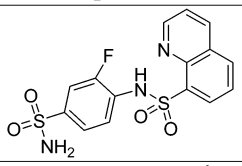
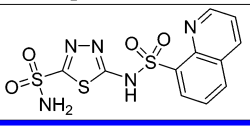
using both the available GoldScore and ChemScore fitness functions.

As shown in Table 2 using the ChemScore function, the Gold program was able to predict the binding disposition of the ligands with an average RMSD error of 2.4 Å, and only 6 ligands out of 40 possessed a predicted disposition that, compared with their X-ray one, showed an RMSD greater than 3.5 Å.

As regards GoldScore, the docking results seemed to be worse than those obtained by means of ChemScore, as they showed an average RMSD of 3.5 Å with 18 ligands out of the 40 possessing an RMSD value greater than 3.5 Å.

However, the analysis of the ZBF disposition revealed that GoldScore was able to better predict its interaction with the protein, as the RMSD evaluation highlighted an average RMSD of 0.7 Å with only compounds 1IF6 and 1TTM showing a value greater than 1.5 Å (see Table 3). As regards ChemScore, the docking results showed an average RMSD of the ZBF disposition of 1.2 Å with six compounds that showed an RMSD greater than 1.5 Å.

Table 6. Compounds Reported in the 3D-QSAR Analysis

Cpd.	Structure	Ki (nM) hCA II
1[29]		1
2[29]		15
3[29]		17
4[29]		49
5[29]		12
6[31]		32
7[31]		2

In agreement with previously reported studies. Otherwise, the ChemScore function favored the interaction between the sulfonamido group and the nitrogen atom of His94, which could not interact with the ZBF as it coordinated the zinc atom.

In order to combine the good ZBF geometry and the good ligand disposition shown respectively by GoldScore and ChemScore, a third docking analysis was performed using the ChemScore fitness function and also applying as a constraint the formation of an H bond between the hydroxy group of Thr199 and the docked ligands.

As shown in Table 2, the average RMSD of the ligand dispositions was very similar to the one obtained by the ChemScore fitness function without any constraint. Furthermore only 6 ligands out of the 40 possessed a predicted disposition with an RMSD greater than 3.5 Å.

As regards the ZBF, the RMSD analysis revealed that the use of the constraint improved its geometry showing results similar to those obtained through GoldScore (average RMSD of 0.7 Å and only compound 1IF6 that showed a value greater than 1.5 Å). As shown in Figure 1C, the sulfonamido nitrogen atom correctly formed an H bond with Thr199.

Cross-Docking. In the previous section, we evaluated the ability of Gold to reproduce the known binding mode of several CA II inhibitors. To further verify the reliability of the docking, we applied a cross-docking approach, docking all the above-mentioned known ligands inside all the X-ray

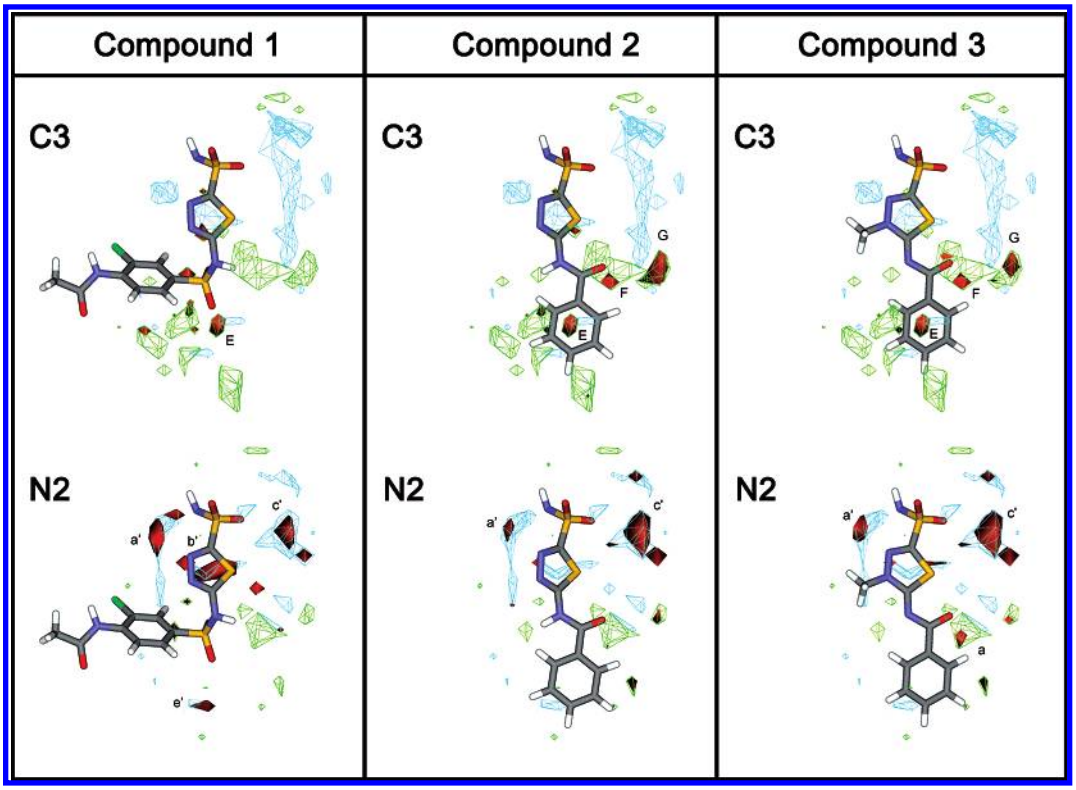


Figure 6. Positive activity contribution plots (red polyhedrons) for C3 and N2 probes of compounds **1**−**3** embedded in the negative (cyan) and positive (green) regions of the PLS coefficient plot.

proteins and comparing the docking binding disposition with the one observed in the X-ray structures.

As a first step of this method, the flexibility of CA II was evaluated aligning all the X-ray proteins (using 1A42 as a reference protein) and measuring the RMSD value of both the whole protein backbone and the heavy atom binding site. For this purpose, the binding site was defined as all the residues within 6 Å from the 1A42 inhibitor.

As shown in Table 4, CA II seemed to possess a low degree of movement upon the binding of different ligands. The RMSD of the protein backbone never exceeded the value of 0.5 Å, and as regards the binding site, excluding 1YDA, for all the other proteins, the RMSD value was lower than 1 Å.

Figure 2 shows the cross-docking results obtained through GoldScore, ChemScore, and ChemScore with the Thr199 constraint. The results confirmed that GoldScore was not able to highly predict the binding of CA II inhibitors, as it showed a global average RMSD of 3.4 Å (see the last column of Figure 2) and the best RMSD value for all 40 of the ligands was 2.9 Å, obtained with 1BNQ, 1IF7, and 1IF9 X-ray proteins. The ChemScore function showed better results, with a global average RMSD value of 2.6 Å (DA = 0.51); however, this fitness function was unable to predict the right interaction of the ZBF since for most of the complexes the nitrogen of the sulfonamido group formed an H bond with His94.

The addition of the constraint at Thr199 determined the right prediction of the ZBF interaction and also low RMSD values since the global average RMSD value was 2.7 Å (DA = 0.46).

Furthermore, from these results, it appeared that the program was able to correctly predict all the analyzed X-ray ligand binding dispositions, with the exception of Celecoxib (1OQ5), since, in this case, for all three of the methods applied to the 40 proteins, the RMSD value of Celecoxib was always greater than 5 Å.

The cross-docking analysis of the acetazolamide (1YDA) revealed also that the docking procedure in many cases was not able to arrange the thiadiazole ring in the correct disposition, overturning it by 180°.

In order to evaluate some alternative docking method, we cross-docked the 40 inhibitors also using the AUTODOCK 3.0 program.[21] However, the preliminary calculations did not show any improvement of the results, since the RMSD average values were worse than the GOLD ones for all the tested proteins (see the Materials and Methods section for more details).

Automated Docking and 3D-QSAR Analysis. The above-described results highlighted that, among the three procedures, the use of the ChemScore fitness function together with the constraint at Thr199 was the only one that showed good results for the binding geometry of the ZBF and also for the disposition of the ligands inside the receptor.

Thus, using this procedure, 294 CA II sulfonamide inhibitors[26,27,29−36] were docked into the binding site of 1OKN. Their inhibitory activity spanned about 3 orders of magnitude (the inhibition data ranged from 1 nM to 1 μM, expressed as K_i), and as shown in Figure 3, they were characterized by eight different central scaffolds.

The docking results were then scored by means of the ChemScore, GoldScore, DrugscoreCSD, and DrugscorePDB scoring functions, but the correlation between the calculated free energy of binding and the experimental one was not good since the quadratic correlation (R^2) showed low values ($R^2 < 0.3$). Thus, with the aim of obtaining quantitative results, the docking poses were used as a receptor-based alignment for a 3D-QSAR analysis.

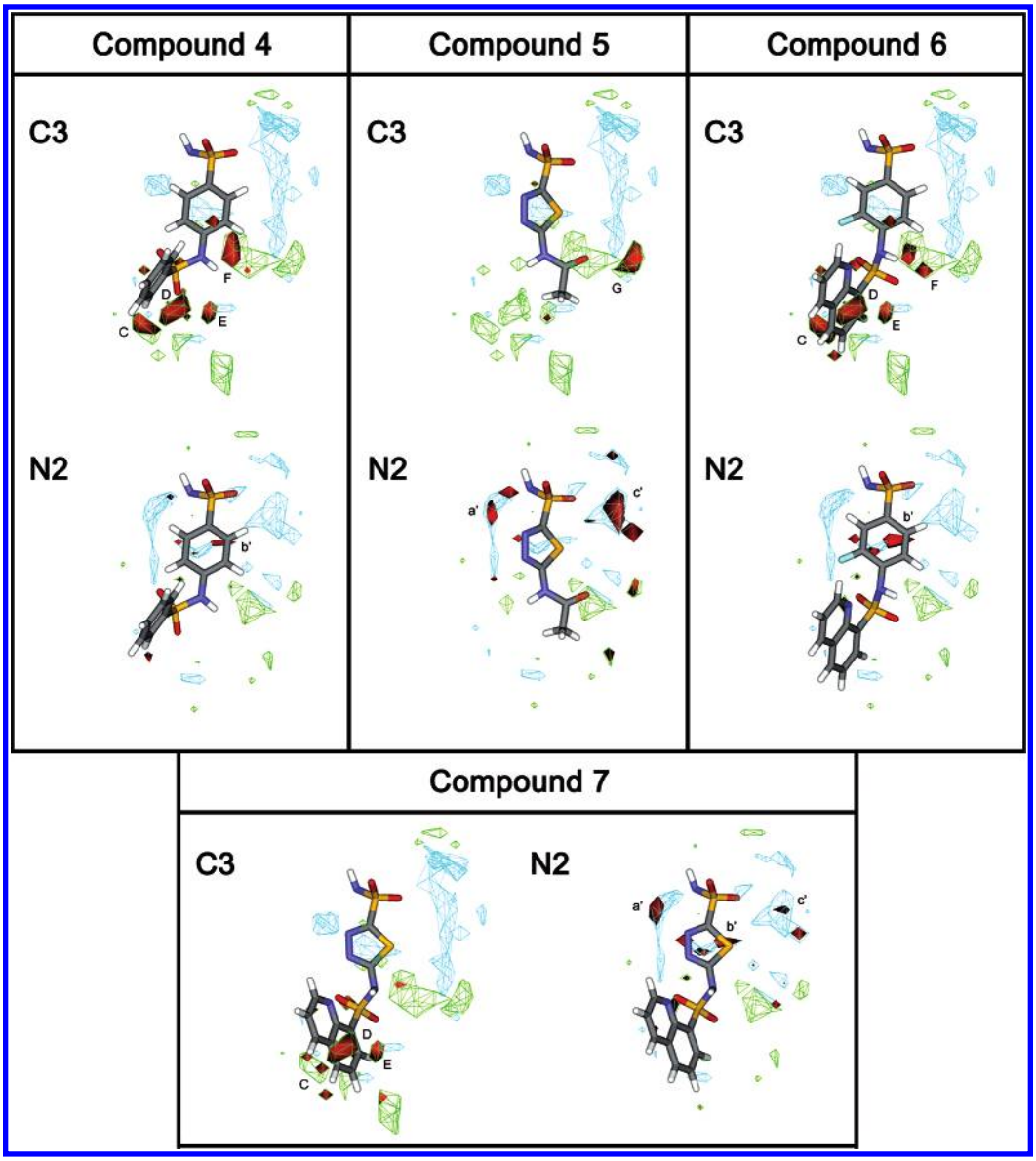


Figure 7. Positive activity contribution plots (red polyhedrons) for C3 and N2 probes of compounds **4**−**7** embedded in the negative (cyan) and positive (green) regions of the PLS coefficient plot.

The reliability of the 3D-QSAR model was characterized by its correlation coefficient (r^2), predictive correlation coefficient (q^2), and cross-validated standard deviation of errors of prediction ($SDEP_{CV}$). The training set was constituted of 220 compounds,[29−36] and furthermore, three external tests were used: the first one of 15 compounds,[29−36] extracted from the training set, whose activity values ranged from 2 to 520 nM (K_i), and a second and third test set of 28 and 31 compounds, respectively, that corresponded to the compounds published by Ilies et al.[26] and Scozzafava et al.[27]

Table 5 reported the main data of the 3D-QSAR analysis. The third PLS component explained 86% of the variance and was predictive ($q^2 = 0.70$ and SDEP of test sets = 0.44, 0.62, and 0.43); the fourth PLS component improved both the fitting and the predictive ability of the model ($r^2 = 0.88$, $q^2 = 0.72$, and the SDEP of the test set = 0.44, 0.63, and 0.41), whereas the fifth PLS component had no further predictivity improvement. Thus, the model's optimal dimensionality was given by four components (see Figure 4).

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,[24] 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 and makes it possible to display spatial regions that are individually important for the selected molecule.

Figure 5 illustrates the PLS coefficient plots of the model for the C3 and N2 probes. Compound **1** is also reported.

The upper part of Figure 5 shows the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the C3 probe; in particular, there are seven principal regions (A−G) with positive 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. On the contrary, the negative PLS coefficients show areas where 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 this picture, three main regions (A′−C′) can be recognized.

In the lower part of Figure 5 are reported the positive (green polyhedrons) and negative (cyan polyhedrons) PLS coefficients for the N2 probe. There are five principal regions (a′−e′) 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 is only one positive region (a) where a favorable interaction between a substituent and the probe determines a decrease in activity.

Figure 6 illustrates the PLS coefficient for the C3 and N2 probes as a triangle mesh and compounds **1**−**3** (see Table 6) embedded in their positive activity plot contributions, displayed as red polyhedrons.

Compound **1** (characterized by the best K_i value, 1 nM) does not favorably interact with the lipophilic regions (C3 probe), except for the small region E. Otherwise, the thiadiazole ring favorably interacts with the electrostatic region a′, the ZBF interacts with the c′ regions, and the carbonyl oxygens of the central sulfonamido group favorably interact with the b′ and e′ regions (N2 probe).

Compound **2**, which is 15-fold less active than **1**, differs from **1** for the substitution of the SO_2 with the carbonyl group and for the lack of phenyl ring substituents. Analysis of the activity contribution plot highlights a better lipophilic interaction with the G and F regions (C3 probe); however, the absence of the SO_2 determines the lack of the electrostatic interaction with b′ and e′ (N2 probe).

The activity contribution plot of compound **3** appears to be very similar to that of **2**. These compounds possess a similar activity, and **3** is characterized by the presence of the 4-methyl-4,5-dihydrothiadiazole nucleus; thus, the N^4 atom of the ring seems not to be important.

Compound **4** possesses an activity 49-fold lower than that of **1**, and as displayed in Figure 7, it possesses lipophilic interactions with regions C, D, E, and F (C3 probe). However, the presence of the phenyl ring instead of the thiadiazole one causes the loss of all the electrostatic interactions (N2 probe). Thus, the electrostatic interaction seems to be more important than the lipophilic one.

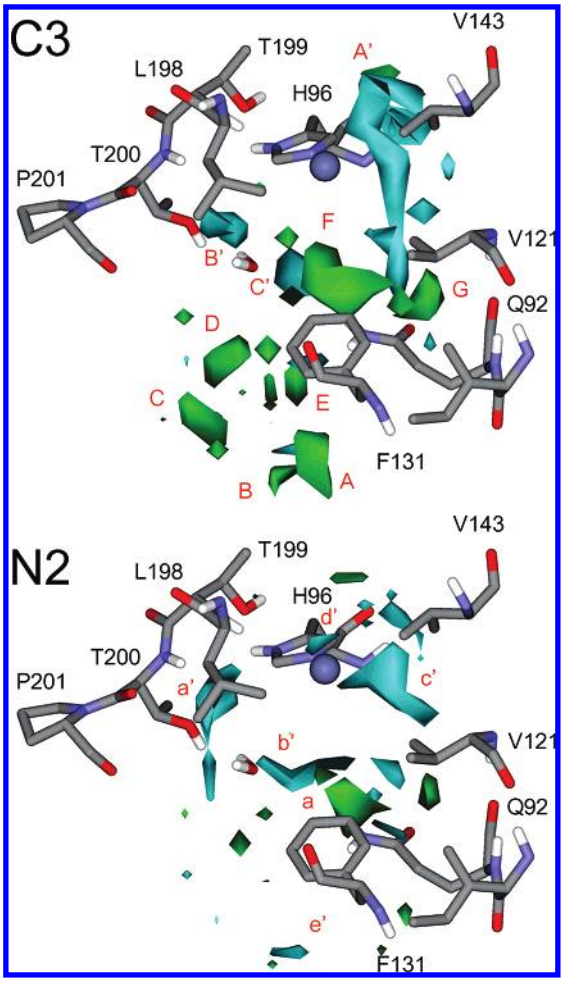
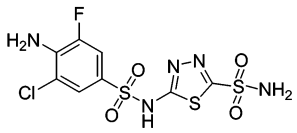
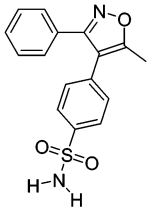
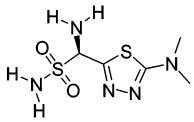
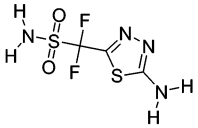
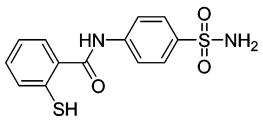


Figure 8. PLS coefficient plots obtained with the C3 and N2 probes superimposed on the CA II binding site.

This fact is confirmed by the analysis of the positive activity plot contributions of compound **5**, which possesses an activity similar to that of **2**. The lipophilic interactions, due to the small size of **5**, are restricted to the G region (C3 probe), but like **2**, compound **5** shows the electrostatic interactions with the a′ and c′ regions (N2 probe).

Table 7. Latest Deposited hCA II Inhibitors' X-ray Complexes and RMSD Values Calculated between the Docked and Experimental Ligand Dispositions

		
2HOC – RMSD 2.3 Å	**2AW1** – RMSD 2.5 Å	**2EU2** – RMSD 2.4 Å
		
2EU3 – RMSD 3.3 Å	**2HD6** – RMSD 2.2 Å	

A further confirmation of the importance of the thiadiazole, and in particular of the N^3 nitrogen, is given by the comparison between **6** (K_i = 32 nM) and **7** (K_i = 2 nM). Compound **6** displays lipophilic interactions with regions C, D, E, and F, while **7** interacts only with the C, D, and E regions (C3 probe). However, differently from compound **6**, the presence of the thiadiazole ring allows the electrostatic interaction with the a′ region beyond the interaction of the SO_2 with the b′ area (N2 probe).

As the alignment of the ligands was performed using the structures docked into the X-ray CA II structure, it was likely to be useful to check for matching between the CA II protein and the 3D-QSAR maps.

In Figure 8, the binding site of the CA II overlaps with the 3D PLS coefficient maps of the C3 and N2 probes. As regards the C3 probe, regions D, E and F are in the proximity of Phe131, thus supporting an important role for this residue.

As regards the N2 probe, the electrostatic surface a′ corresponds to Thr200, and region b′ is situated between Gln92 and a structural water molecule.

Newly Published X-ray Structures. During the progress of our studies, new hCA II−inhibitor complexes have been published. These compounds were docked inside the 1OKN protein using the same computational procedure applied for the ligands used in the 3D-QSAR study.

As shown in Table 7, the ligand disposition was in agreement with the X-ray structures since the RMSD between the docked dispositions and the one observed in the X-ray complexes showed a value comprised between 2.2 and 3.3 Å.

CONCLUSIONS

The CA II structure is one of the most studied carbonic anhydrases: more than 40 X-ray structures have been published, and the activity of hundreds of inhibitors has been reported.

In this paper, using the CA II as a case study, we tested the possibility of developing a receptor-based 3D-QSAR model.

First of all, we analyzed the ability of the Gold program to predict the binding disposition of known ligands; then, the best procedure was used to dock almost 300 inhibitors, and the best poses were used as an alignment tool for the development of a 3D-QSAR model.

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

The 3D-QSAR analysis mainly revealed that the electrostatic interactions play a fundamental role in determining the ligand binding; in particular, the presence of an H-bond acceptor in the α position with respect to the ZBF seemed to be very important. In the case of CA II inhibitors with a thiadiazole structure such as compounds **1**, **2**, **5**, and **7**, it is the N^3 atom that plays this role.

Furthermore, the superimposition between the PLS coefficient surfaces and the CA II structure pointed out that residues Gln92, Phe131, and Thr200 showed a key function.

The aim of this paper was the development and the validation of a computational approach useful for studying CA ligands. For this purpose, a large amount of compounds were taken into account for defining a significant training set. Moreover, three different test sets were also considered for the evaluation of the predictive ability of the model. Therefore, the 3D-QSAR model should be suitable for the design of new CA II inhibitors, and the obtained results encourage us to apply this method also for other CAs, including the systems for which X-ray data are not available

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.

Note Added after ASAP Publication. This article was released ASAP on February 13, 2007 with minor errors in ref 10. The correct verion was posted on February 26, 2007.

Supporting Information Available: Matrix of RMSDs obtained by the cross-docking study using GOLD (GoldScore and ChemScore methods) and AUTODOCK 3.0. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Smith H. J.; Simons C. *Enzymes and their Inhibition: Drug Development*; CRC Press: Boca Raton, FL, 2005; pp 1−308.

(2) Chegwidden, W. R.; Dodgson, S. J.; Spencer, I. M. The Roles of Carbonic Anhydrase in Metabolism, Cell Growth and Cancer in Animals. In *The Carbonic Anhydrases: New Horizons*; Chegwidden, W. R., Carter, N. D., Edwards, Y. H., Eds.; Birkhauser Verlag: Basel, Switzerland, 2000; pp 343−363.

(3) Schuman, J. S. Short- and Long-Term Safety of Glaucoma Drugs. *Expert Opin. Drug Saf.* **2002**, *1*, 181−194.

(4) Supuran, C. T.; Scozzafava, A. Applications of Carbonic Anhydrase Inhibitors and Activators in Therapy. *Expert Opin. Ther. Pat.* **2002**, *12*, 217−242.

(5) Supuran, C. T.; Scozzafava, A.; Casini, A. Carbonic Anhydrase Inhibitors. *Med. Res. Rev.* **2003**, *23*, 146−189.

(6) 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.

(7) Fisher, Z.; Hernandez Prada, J. A.; Tu, C.; Duda, D.; Yoshioka, C.; An, H.; Govindasamy, L.; Silverman, D. N.; McKenna, R. Structural and Kinetic Characterization of Active-Site Histidine as a Proton Shuttle in Catalysis by Human Carbonic Anhydrase II. *Biochemistry* **2005**, *44*, 1097−1105.

(8) Winum, J. Y.; Scozzafava, A.; Montero, J. L.; Supuran, C. T. New Zinc Binding Motifs in the Design of Selective Carbonic Anhydrase Inhibitors. *Mini Rev. Med. Chem.* **2006**, *6*, 921−936.

(9) 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: Boca Raton, FL, 2004; pp 149−182.

(10) Melagraki, G.; Afantitis, A.; Sarimveis, H.; Igglessi-Markopoulou, O.; Supuran, C. T. QSAR Study on Para-Substituted Aromatic Sulfonamides as Carbonic Anhydrase II Inhibitors Using Topological Information Indices. *Bioorg. Med. Chem.* **2006**, *14*, 1108−1114.

(11) Gruneberg, S.; Wendt, B.; Klebe, G. Subnanomolar Inhibitors from Computer Screening: A Model Study Using Human Carbonic Anhydrase II. *Angew. Chem., Int. Ed.* **2001**, *40*, 389−393.

(12) 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.

(13) 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.

(14) 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.

(15) 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.
(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(17) *Maestro*, version 7.5; Schrödinger Inc.: Portland, OR, 2005.
(18) *Macromodel*, version 8.5; Schrödinger Inc.: Portland, OR, 1999.
(19) Ragno, R.; Frasca, S.; Manetti, F.; Brizzi, A.; Massa, S. HIV-Reverse Transcriptase Inhibition: Inclusion of Ligand-Induced Fit by Cross-Docking Studies. *J. Med. Chem.* **2005**, *48*, 2002−2012.
(20) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks, C. L., III. Comparative Study of Several Algorithms for Flexible Ligand Docking. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 755−763.
(21) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.
(22) Hu, X.; Shelver, W. H. Docking Studies of Matrix Metalloproteinase Inhibitors: Zinc Parameter Optimization to Improve the Binding Free Energy Prediction. *J. Mol. Graphics Modell.* **2003**, *22*, 115−126.
(23) Hu, X.; Balaz, S.; Shelver, W. H. A Practical Approach to Docking of Zinc Metalloproteinase Inhibitors. *J. Mol. Graphics Modell.* **2004**, *22*, 293−307.
(24) *GOLPE 4.5*; Multivariate Infometric Analysis Srl.: Perugia, Italy, 1999.
(25) *GRID*, version 22a; Molecular Discovery Ltd.: Oxford, U. K., 2004.
(26) Ilies, M.; Supuran, C. T.; Scozzafava, A.; Casini, A.; Mincione, F.; Menabuoni, L.; Caproiu, M. T.; Maganu, M.; Banciu, M. D. Carbonic Anhydrase Inhibitors: Sulfonamides Incorporating Furan-, Thiophene- and Pyrrole-Carboxamido Groups Possess Strong Topical Intraocular Pressure Lowering Properties as Aqueous Suspensions. *Bioorg. Med. Chem.* **2000**, *8*, 2145−2155.
(27) Scozzafava, A.; Menabuoni, L.; Mincione, F.; Briganti, F.; Mincione, G.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Synthesis of Water-Soluble, Topically Effective, Intraocular Pressure-Lowering Aromatic/Heterocyclic Sulfonamides Containing Cationic or Anionic Moieties: Is the Tail More Important Than the Ring? *J. Med. Chem.* **1999**, *42*, 2641−2650.
(28) 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.
(29) Vullo, D.; Franchi, M.; Gallori, E.; Antel, J.; Scozzafava, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Inhibition of Mitochondrial Isozyme V with Aromatic and Heterocyclic Sulfonamides. *J. Med. Chem.* **2004**, *47*, 1272−1279.
(30) Ilies, M. A.; Vullo, D.; Pastorek, J.; Scozzafava, A.; Ilies, M.; Caproiu, M. T.; Pastorekova, S.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Inhibition of Tumor-Associated Isozyme IX by Halogenosulfanilamide and Halogenophenylaminobenzolamide Derivatives. *J. Med. Chem.* **2003**, *46*, 2187−2196.
(31) Borras, J.; Scozzafava, A.; Menabuoni, L.; Mincione, F.; Briganti, F.; Mincione, G.; Supuran, C. T. Carbonic Anhydrase Inhibitors: Synthesis of Water-Soluble, Topically Effective Intraocular Pressure Lowering Aromatic/Heterocyclic Sulfonamides Containing 8-Quinoline-sulfonyl Moieties: Is the Tail More Important Than the Ring? *Bioorg. Med. Chem.* **1999**, *7*, 2397−2406.
(32) 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.
(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: 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.
(34) Chazalette, C.; Masereel, B.; Rolin, S.; Thiry, A.; Scozzafava, A.; Innocenti, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors. Design of Anticonvulsant Sulfonamides Incorporating Indane Moieties. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5781−5786.
(35) Casini, A.; Winum, J. Y.; Montero, J. L.; Scozzafava, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors: Inhibition of Cytosolic Isozymes I and II with Sulfamide Derivatives. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 837−840.
(36) Mincione, F.; Starnotti, M.; Menabuoni, L.; Scozzafava, A.; Casini, A.; Supuran, C. T. Carbonic Anhydrase Inhibitors: 4-Sulfamoyl-benzenecarboxamides and 4-Chloro-3-sulfamoyl-benzenecarboxamides with Strong Topical Antiglaucoma Properties. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1787−1791.

CI600469W