J Comput Aided Mol Des (2010) 24:803–817
DOI 10.1007/s10822-010-9375-z

Fragment-guided approach to incorporating structural information into a CoMFA study: BACE-1 as an example

Lívia Barros Salum · Napoleão Fonseca Valadares

Received: 6 May 2010 / Accepted: 15 July 2010 / Published online: 27 July 2010

Abstract Alzheimer's disease is an ultimately fatal neurodegenerative disease, and BACE-1 has become an attractive validated target for its therapy, with more than a hundred crystal structures deposited in the PDB. In the present study, we present a new methodology that integrates ligand-based methods with structural information derived from the receptor. 128 BACE-1 inhibitors recently disclosed by GlaxoSmithKline R&D were selected specifically because the crystal structures of 9 of these compounds complexed to BACE-1, as well as five closely related analogs, have been made available. A new fragment-guided approach was designed to incorporate this wealth of structural information into a CoMFA study, and the methodology was systematically compared to other popular approaches, such as docking, for generating a molecular alignment. The influence of the partial charges calculation method was also analyzed. Several consistent and predictive models are reported, including one with $r^2 = 0.88$, $q^2 = 0.69$ and $r^2_{pred} = 0.72$. The models obtained with the new methodology performed consistently better than those obtained by other methodologies, particularly in terms of external predictive power. The visual analyses of the contour maps in the context of the enzyme drew attention to a number of possible opportunities for the development of analogs with improved potency. These results suggest that 3D-QSAR studies may benefit from the additional structural information added by the presented methodology.

Keywords CoMFA · 3D-QSAR · BACE-1 · Alzheimer · Fragment-based

Introduction

Alzheimer's disease (AD) is a devastating and ultimately fatal neurodegenerative disease. This neurological disorder is the most common cause of dementia, affecting more than 18 million people worldwide [1]. It is estimated that in the United States 4.5 million Americans suffer from AD [2], and that the current direct and indirect annual costs of caring are at least $100 billion [3, 4]. Moreover, AD places physical and emotional stress on caregivers, usually family and friends, causing great impact on their quality of life [1].

The therapeutic options for AD are severely limited, and the drugs currently marketed for AD treatment do not prevent or reverse disease progression. Despite being approved only for the management of symptoms, these drugs collectively generated $3 billion in sales in 2006 [5]. It is expected that in the coming years the AD drug market will consistently expand driven by the aging population and the approval of the first disease modifying drugs and new symptom management agents.

AD is characterized by the progressive formation of insoluble extracellular amyloid plaques and intracellular neurofibrillary tangles of hyperphosphorylated microtubule associated tau protein in the brain. The amyloid plaques are mainly composed of a 4 kDa amyloid β peptide (Aβ) consisting of 42 amino acids, generated by the proteolytic processing of a large transmembrane amyloid precursor protein (APP) [6]. The β-site APP cleaving enzyme

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9375-z) contains supplementary material, which is available to authorized users.

L. B. Salum · N. F. Valadares (✉)
Instituto de Física de São Carlos, Centro de Biotecnologia Molecular Estrutural, Universidade de São Paulo, Av. Trabalhador são-carlense 400, São Carlos-SP 13560-970, Brazil
e-mail: napo@ifsc.usp.br

(BACE-1, also known as β-secretase) is the membrane bound enzyme that cleaves APP, releasing an extracellular APP fragment. The remaining membrane bound C-terminal 99 amino acid long fragment of APP is heterogeneously cleaved by the γ-secretase, producing mainly Aβ peptides of 40 and 42 amino acids in length. APP can also be processed in the middle of the Aβ domain by a metalloproteinase, α-secretase, which precludes the formation of Aβ [7, 8].

The most widely accepted theory to explain the origin of AD is the amyloid cascade hypothesis, which is supported by strong genetic, histopathological and clinical evidence, and states that Aβ causes progressive synaptic and neuritic injury [9–11]. Due to the findings that BACE-1 knockout mice are healthy, fertile, appear normal in numerous analyses and produce much less Aβ [12, 13], BACE-1 has become an attractive target for the therapy of AD [4]. In the last few years, hundreds of publications and patents concerning the development of BACE-1 inhibitors have been disclosed, particularly by the pharmaceutical companies [14].

As expected for any promising validated pharmaceutical target, 10 years after the resolution of its first X-ray structure more than a hundred BACE-1 crystal structures have been made available [4, 14, 15], and there are many more in the vaults of pharmaceutical companies. The ever increasing pace at which X-ray structures are made available generates an enormous pressure for the development of methodologies that integrate ligand-based methods with structural information for their use in the medicinal chemistry field.

To address the aforementioned necessity, in the present work we selected 128 BACE-1 inhibitors recently disclosed by GlaxoSmithKline R&D specifically because the crystal structures of 9 of these compounds and 5 closely related analogs have been made available [16–22]. A new fragment-guided approach was designed to incorporate this wealth of structural information into a comparative molecular field analysis (CoMFA) study [23], and this procedure allowed us to derive consistent three-dimensional quantitative structure–activity relationship (3D-QSAR) models. The methodology was also systematically compared to other popular approaches, and the analysis of the contour maps in the context of the enzyme highlighted a number of possibilities for developing analogs with improved potency.

Experimental section

Data set

The data set used for the QSAR studies contains 128 BACE-1 inhibitors that were synthesized and assayed by GlaxoSmithKline R&D [16–22]. Their potency values (as measured by IC_{50}) were determined under standardized experimental conditions, rendering these data comparable between all molecules in the data set, a fundamental requirement for successful QSAR analyses. The data set also comprises 14 X-ray structures disclosed by the same group (PDB codes 2VIY, 2VIZ, 2VJ6, 2VJ9, 2VIE, 2VJ7, 2VNM, 2VNN, 2WEZ, 2WF0, 2WF1, 2WF2, 2WF3, 2WF4) that were used for fragment-based similarity analyses, for generation of the data set alignments and in the analyses of the CoMFA results in the context of the enzyme. The IC_{50} values were converted into the corresponding negative logarithmic units, pIC_{50} ($-\log IC_{50}$), and then used in the QSAR studies. The chemical structures and corresponding pIC_{50} values for the 9 data set compounds for which co-crystal structures with BACE-1 were disclosed are listed in Table 1, and those for the complete set of compounds are listed in Supplementary Table 1.

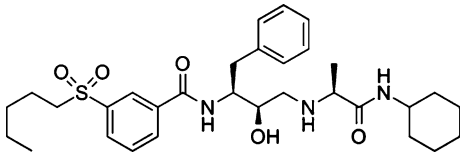
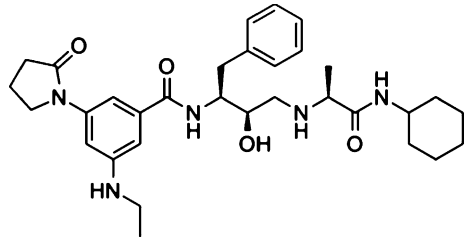
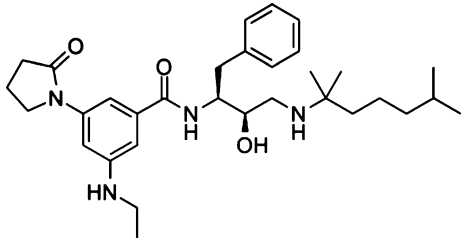
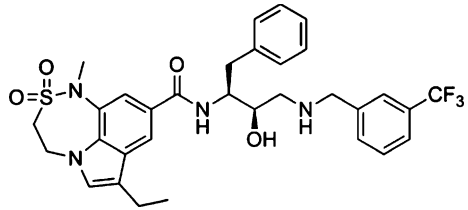
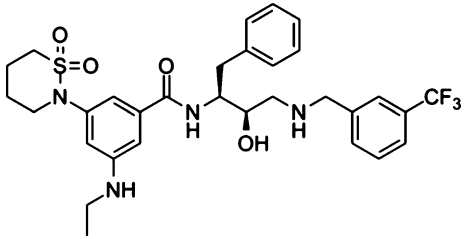
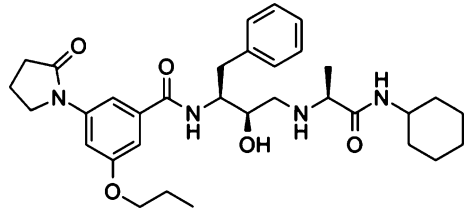
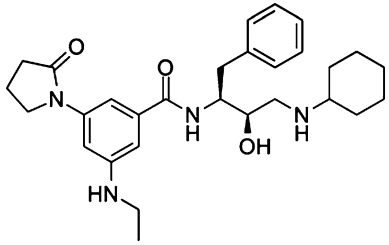
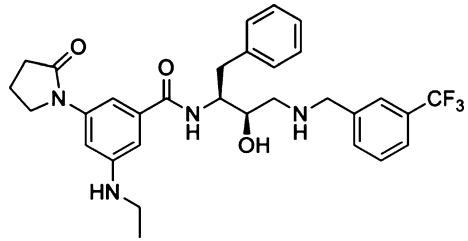
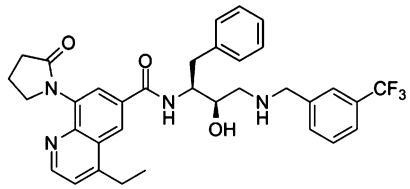
Computational approach

Hologram QSAR (HQSAR) and CoMFA analyses, calculations, and visualizations were performed using the SYBYL 8.0 package (Tripos Inc., St. Louis, USA) and Pymol [24], running on Red Hat Enterprise Linux and Slackware 13 Linux workstations. For the HQSAR fragment-based similarity search the 3D structures of the BACE-1 inhibitors were constructed using the standard geometric parameters of the molecular modeling software package SYBYL 8.0 and minimized using CONCORD. Before the calculation of the net formal charges, the partial charges and the final geometry used in the CoMFA studies, basic amines were modeled in their protonated states. Subsequently each single optimized conformation of each molecule in the data set was energy minimized employing the Tripos force field and Gasteiger–Huckel charges. Hierarchical cluster analyses of the data set were carried out within SYBYL 8.0.

Hologram-based fragment similarity analysis

Molecular holograms were generated for each molecule of the data set, from different sets of combinations of parameters concerning hologram generation (hologram length, fragment size, and fragment distinction) [25]. Results of HQSAR modeling were employed to determine the best combinations of hologram generation parameters in describing the molecular structures of the data set compounds. Accordingly, several combinations of fragment distinction parameters were considered during the QSAR modeling runs—holograms were generated using distinct combinations of atoms (A), bonds (B), connections (C),

Table 1 Chemical structures and corresponding pIC_{50} of the BACE-1 inhibitors used for CoMFA studies for which X-ray structures are available

Cpd	Structure	pIC_{50}	PDB code	Cpd	Structure	pIC_{50}	PDB code
3		5.74	**2VIY**	**8**		6.22	**2VIZ**
9		7.89	**2VJ6**	**13**		6.74	**2VJ9**
17		7.48	**2VIE**	**20**		7.40	**2VJ7**
52		8.70	**2VNN**	**67**		6.68	**2WF0**
111		8.52	**2VNM**				

Cpd compound, *PDB code* PDB code of the X-ray structure, $pIC_{50} = -\log(IC_{50})$

hydrogen atoms (H), chirality (Ch), and donor and acceptor (DA) as fragment distinctions. The HQSAR analyses were performed by screening the 12 default series of hologram length values ranging from 53 to 401 bins using distinct fragment sizes (2–5, 3–6, 4–7, 5–8 and 6–9 atoms) [25]. The patterns of fragment counts were then related to the experimental biological parameters using Partial Least Squares (PLS) regression analyses to derive the HQSAR models. Cross-validation procedures (q^2 leave-one-out and leave-many-out with either 10 or 5 randomly selected groups) were used to assess model stability and statistical significance.

The parameters employed to generate the molecular holograms as independent variables for the most significant four HQSAR models were retrieved as the fragment-based descriptors that best capture the structural characteristics of the data set of inhibitors [26]. Hierarchical cluster analyses (HCA) were carried out employing the complete linkage clustering method with no scaling of the hologram variables to define the families of related compounds in the data set. Molecular holograms generated from the four different combinations of the parameters previously selected were used as descriptors of the molecular structures of the data set inhibitors in the individual unsupervised analyses to define the trees of 2D similarity. The resulting dendrograms were analyzed and the ten best levels (taking the linkage distance to the next level as reference) were selected to divide the compounds in similar groups in each different HCA model generated (one for each type of molecular hologram selected).

The 14 data set X-ray structures (Table 1 and Supplementary Table 1) were referenced in the generated HCA individual dendrograms. Subsequently, we attributed to each inhibitor in the data set the most 2D similar small-molecule structure for which a 3D crystallographic complex was available. The consensus X-ray structure, selected among the four models generated, was used for the structural alignment of the data set.

Alignment rules

QSAR methods have been widely used as valuable tools to assist the design of enzyme ligands with improved properties and potential utility in clinical medicine [27]. Among well-established 3D QSAR methods, such as CoMSIA [28], COMBINE [29], GRID [30] and AFMoC [31], CoMFA is one of the most popular, having a number of promising attributes, including the possibility of visualizing the regions in space responsible for increases or decreases in the values of a particular biological property. The first requisite to perform a CoMFA study is the alignment of the data set compounds, which is accomplished by the assignment of the conformation and relative position of each molecule in the data set. The molecular alignment is the most important input parameter in a CoMFA study, the quality of the results depends on it, and different methodologies to achieve a meaningful alignment have been proposed and described in the literature [32–34]. The present study employs and compares 4 different alignment procedures.

(A) The first procedure is a new strictly structure-based method. The present data set was particularly selected because the crystal structures of BACE-1 bound to 9 of its compounds and five closely related analogs were disclosed by GlaxoSmithKline R&D [16–22]. One of the highest resolution structures (PDB code 2VIZ, 1.6 Å resolution) was chosen as template, and all other X-ray structures were subject to a structural alignment of the Cα atoms in Pymol [24] followed by up to five refinement cycles in order to reject structural outliers found during the fit. The maximum fit RMSD in the first cycle was 0.57 Å, and the maximum fit RMSD in the last cycle was 0.22 Å, indicative of highly similar structures involving little global alteration to the fold.
Subsequently, all the ligand structures were extracted, and a.mol2 file for each ligand with correct atom type assignments, added hydrogen atoms and correct protonation states was generated in SYBYL 8.0. Finally, these structures were energy minimized with all non-hydrogen atoms constrained employing the Tripos force field and Gasteiger–Huckel charges.
Guided by the fragment-based similarity analyses, a ligand conformation extracted from a crystal structure was assigned to each molecule in the data set, which was assumed to be close to the bioactive conformation of that inhibitor (Supplementary Table 1). For each compound in the data set the assigned structure was loaded, and the substituents that were not present in the X-ray template were generated by modifying the assigned crystal structure scaffold using conformations in accordance with the other available X-ray structures. The geometry of each resulting compound was then optimized only in the modified positions of the assigned crystal structure, and in all hydrogen atoms. This protocol yielded an alignment that, by definition, agrees extremely well with the X-ray structures.

(B) The second alignment was obtained by submitting the first alignment to a rigid fit of the common substructure in SYBYL (Align Database).

(C) The third alignment was generated by the program GOLD 4.1. The reference crystal structure (PDB code 2VIZ) was used as template after removal of the

ligand and water molecules. Hydrogen atoms were then added in standard geometry and the structure was energy minimized with all non-hydrogen atoms constrained. The docking site within BACE-1 was defined by the ligand in the structure. The GoldScore scoring function was employed, and the search efficiency was set to the maximum (200%). Each ligand was docked up to 20 times with early termination allowed if the top 3 best ranked solutions were within 1.2 Å RMSD. The best ranked conformation of each compound was then selected for the CoMFA studies.

(D) Since a visual inspection of the alignment described in **C** in the context of the crystallographic structures suggested that the GOLD generated docking pose of some compounds were incorrect, a fourth alignment was proposed. This last alignment was essentially alignment **C**, except that 8 of the compounds with docking poses flagged as incorrect in the visual inspection were replaced by new GOLD results obtained by a slightly different protocol. Each of these eight ligands was docked accordingly to the protocol described for alignment **A**, except that each ligand was docked 40 times with no early termination allowed. All the 40 docking poses of each of these eight compounds were visually inspected in the context of the crystallographic structures, and a visually acceptable docking pose was selected for each of them.

Prior to CoMFA field calculations, both MMFF94 and Gasteiger-Huckel charges were assigned to the aligned molecules in each of these four molecular alignments, generating eight different data sets. Other alignment rules were attempted, like a plain rigid fit of the common substructure of the minimized molecules and a pharmacophore based approach, but both yielded visibly poor alignments, presumably due to the large size and flexibility of the data set ligands.

CoMFA studies

The influence of the electrostatic and steric contributions to the potency of the data set inhibitors was studied by using the CoMFA approach as previously described [34]. Briefly, a 3D grid with 2.0 Å spacing containing all training set molecules and with an additional 4.0 Å extension in each direction was created. The steric and electrostatic field energies between the molecules and an sp3 hybridized carbon atom with a charge of +1 were calculated according to Lennard-Jones and Coulomb potentials at each lattice intersection of the 3D grid. The maximum steric and electrostatic energy cutoff was set to 30 kcal mol^{-1}. CoMFA descriptors were used as independent variables, and pIC_{50} values were considered as dependent variables in the PLS. All CoMFA models were evaluated by the leave-one-out (LOO) and leave-many-out (LMO) methods, and the sensitivity of the 3D-QSAR models to chance correlation was assessed by the progressive scrambling method. The test set compounds, which were not used in the generation of any of the CoMFA models, were used exclusively in the external validation procedure. In this validation method, the CoMFA models were used to predict the pIC_{50} values for the test set compounds, allowing the calculation of the predictive r^2 values (r^2_{pred}) for the QSAR models.

The statistical parameters, non-cross-validated correlation coefficient (r^2), standard error of estimate (SEE), LOO cross-validated correlation coefficient (q^2) and r^2_{pred} were calculated accordingly to the following definitions (1–4):

$$r^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i^{\text{fit}} - y_i\right)^2}{\sum_{i=1}^{N} \left(y_i - y_{\text{mean}}\right)^2} \tag{1}$$

$$\text{SEE} = \sqrt{\frac{\sum_{i=i}^{N} \left(y_i^{\text{fit}} - y_i\right)^2}{N - \text{comp} - 1}} \tag{2}$$

$$q^2\text{LOO} = 1 - \frac{\sum_{i=1}^{N} \left(y_i^{\text{pred(N-1)}} - y_i\right)^2}{\sum_{i=1}^{N} \left(y_i - y_{\text{mean}}^{(N-1,i)}\right)^2} \tag{3}$$

$$r^2_{\text{pred}} = 1 - \frac{\sum_{i=1}^{N} \left(y_i^{\text{pred}} - y_i\right)^2}{\sum_{i=1}^{N} \left(y_i - y_{\text{mean}}^{\text{test}}\right)^2} \tag{4}$$

where N is the number of compounds; y_i is the observed value of potency; y_i^{fit} is the fitted value of y_i; y_{mean} is the arithmetic mean of the potency values of all data set compounds; *comp* is the number of principal components; y_i^{pred} is the predicted potency value of the i-th compound; $y_{\text{mean}}^{\text{test}}$ is the mean of the potency values of the test set compounds; $y_i^{\text{pred(N-1)}}$ is the i-th predicted potency value generated through application of a subset model trained on $N-1$ compounds where the i-th compound was left out, and $y_{\text{mean}}^{(N-1,i)}$ is the arithmetic mean of that subset [35].

It is worthy of note that Schuurmann and coworkers presented an elegant mathematical demonstration that the use of the training set activity mean instead of the mean of the test set biological property values yields a systematic overestimation of the external prediction capability [35]. The overestimation happens because the mean of the test set activities yields the minimum value for the denominator on Eq. 4, therefore any other value increases this denominator, artificially increasing the r^2_{pred}.

Results and discussion

Data set characterization

3D-QSAR CoMFA models were derived for a series of 128 BACE-1 inhibitors for which potency was determined under the same experimental conditions [16–22] (Supplementary Table 1). The compounds of the data set present molecular weights between 450 and 650 Daltons, and their associated IC_{50} values are adequately distributed over more than three orders of magnitude potency range, varying from 2 nM to 5.89 μM. Non supervised HCA performed within SYBYL employing the complete linkage clustering method with no scaling of the CoMFA descriptors, which were used as variables to define the families of related compounds in the data set. This methodology was used to guide the appropriate division of the 128 BACE-1 inhibitors into a training set for model construction containing 102 compounds (1–102, Supplementary Table 1) and a test set for external model validation containing the remaining 26 compounds (103–128, Supplementary Table 1). This analysis guaranteed that both training and test sets represent the structural diversity and cover the whole data set potency space, rendering the data set appropriate for the purpose of QSAR model development, as depicted in Fig. 1.

Structural background

The first X-ray structure of the protease domain of BACE-1 was determined in the year 2000, and it was crystallized as a complex with an eight-residue transition-state inhibitor [15]. Since then many X-ray structures of BACE-1 in its apo form or in complex with structurally diverse inhibitors have been made available [36, 37]. It is important to note that the BACE-1 residue numbering nomenclature used in the present study starts from the beginning of the signal peptide, following the same standard used in the description of the structures in the present data set [16–22]. The signal peptide and pro-peptide are removed during purification [16], therefore Val61 is the first residue in the final purified protein, and Glu62 is the first residue with actual electron density. The nomenclature used by the seminal BACE-1 structures [15, 36, 37] is different, numbering this glutamate (Glu62) as the first residue.

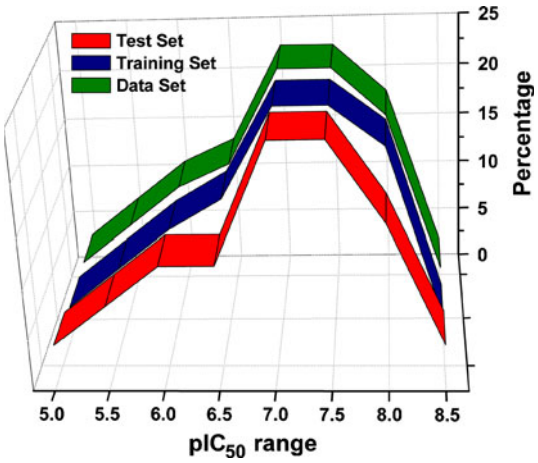


Fig. 1 Distribution of the biological property values for the training set (*blue*), test set (*red*), and complete data set (*green*)

The comparison between BACE-1 in its apo form and its complexes with data set ligands reveals that the front hairpin domain (flap, residues 128–138) undergoes a considerable conformational change upon ligand binding, moving up to 7 Å. This conformational flexibility is further exemplified in co-crystal structures where the flap electron density is too weak to be observed, indicating an open and highly disordered conformation [38]. Another flexible region near the inhibitor binding site is the 10 s loop (residues 70–75), which can adopt distinct conformations in the apo form and when bound to different inhibitors [36]. Moreover, many crystal structures lack electron density for the residues 217–229, some of which are spatially near the 10 s loop [16–22].

The BACE-1 active site region is mostly hydrophilic and superficial (Supplementary Fig. 1), presenting only two deeper pockets (S1 and S3, Fig. 2a) on its vast surface. BACE-1 crystal structures were resolved in complex with a variety of structurally diverse ligands, ranging from transition-state analogues of peptides [15, 37] having molecular weights above 900 Da to small inhibitors presenting high ligand efficiencies and molecular weights below 300 Da [38]. Most inhibitors interact directly with the side chains of the catalytic aspartates, but some do not [39]. These features illustrate some of the challenging aspects in the design of BACE-1 inhibitors, especially when taking into account that these compounds should display good oral absorption and blood–brain barrier permeability in order to be considered viable drug candidates.

The compounds in the present data set, as with many other inhibitors currently being developed [4, 14], present a hydroxy ethylamine non-cleavable transition-state isostere core, and make extensive interactions with the enzyme (Fig. 2a, c). These inhibitors present substituents in the P3, P2 and P1 positions on the non-prime side and the P1′ and P2′ positions on the prime side, which are accommodated chiefly in 5 interaction pockets, termed S3, S2, S1, S1′ and S2′, respectively (Fig. 2a). The data set compounds also interact directly with the flap (Fig. 2b), which is in a closed conformation due to hydrogen bonds to both the main chain amide nitrogen atoms and the side chains of Thr133 and Gln134 (Fig. 2d). Moreover, the benzyl group in the S1 pocket is within face-to-edge stacking distance of Tyr132 [38]. Additional interaction pockets, such as the $S3_{subpocket}$,

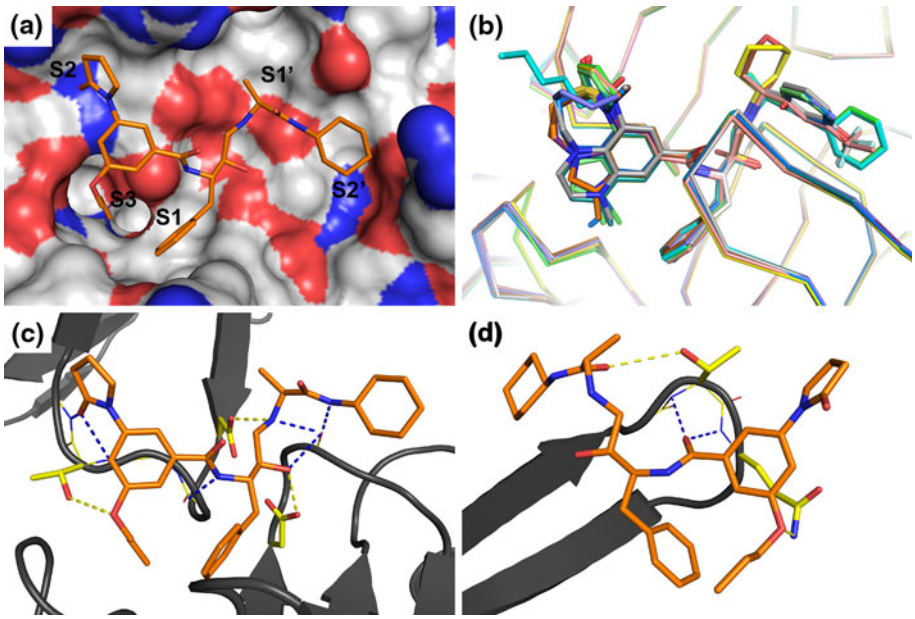


Fig. 2 BACE-1 interaction pockets and the positioning of the data set compounds into the catalytic site. **a** Labeled interaction pockets of BACE-1 with compound **8** shown as reference (PDB code 2VIZ), the flap region and water molecules were omitted for clarity. **b** Superposition of all 14 crystal structures used in this study, with the flap region shown as reference. For each structure the protein is depicted as a carbon alpha trace and the complexed inhibitor is shown as sticks. **c** and **d** Interactions of BACE-1 with compound **8** (PDB code 2VIZ). BACE-1 overall fold is shown as a cartoon, the amino acids side chains are shown as sticks and the backbone residues that interact with the inhibitor are shown as lines. Hydrogen bonds to side chains are colored in yellow and hydrogen bonds to the main chain atoms are depicted in blue. In **c** the flap region was omitted and **d** shows only the flap region and its interactions

may be explored by some inhibitors, and this subsite in particular may accommodate larger substituents due to an induced conformational change to the 10 s loop.

The 14 crystal structures used in the present study were obtained to a reasonably high resolution (ten at 1.8 Å or below, and only one above 1.9 Å), and parameters such as the *R* value, free *R* value, mean *B* values and the Ramachandran plot indicate that all of them are reliable. An inspection of the electron density maps demonstrates that the positioning of the ligands is unequivocal (Supplementary Fig. 2). Overall the structures are very similar, especially in the inhibitor binding sites (Fig. 2b), even including the side chain positions in this region. Likewise, their superposition produces low values for RMSD (0.22 Å maximum RMSD for the superposition of all structures accordingly to the superposition protocol described in the Experimental Section). In fact, the region that exhibits the most noticeable difference in this data set is the 10 s loop, for which three structures present a different conformation. However, there is no evident relationship between the structure of the ligands in these three structures and the conformation of the 10 s loop.

CoMFA analyses

The molecular structures of bioactive compounds can be comprehended as specific arrangements of functional groups or fragments that are able to establish intermolecular interactions with a specific receptor cavity. Indeed, similarity of compound chemical structures, and consequently of their constituent fragments, often indicates related binding modes and pharmacological profiles [26]. We resorted to molecular holograms, which are two dimensional frequency-based fragment descriptors, to correlate each data set compound to a crystal structure with a closely related analog, in order to obtain an X-ray structure-based alignment of all data set compounds. It is worth mentioning that, as a direct consequence of the availability of structures presenting close related analogs (PDB code 2WF1 and 2VNN), one of the fourteen available crystal structures was not selected by the fragment-guided method, and therefore was not employed in the generation of the molecular alignments.

Figure 3 depicts the four molecular alignments displaying all 128 BACE-1 inhibitors used in this study. Important features are noticeable when comparing the molecular alignments by visual inspection in the context of the enzyme. In contrast to the alignments obtained by rigorous docking protocols using the program GOLD 4.1 (Fig. 3c, d), the alignments derived directly from X-ray structures resulted in a closer general superposition for all 128 compounds (Fig. 3a, b). Despite the large common core shared by the aligned compounds and the similarities in all 14 crystal

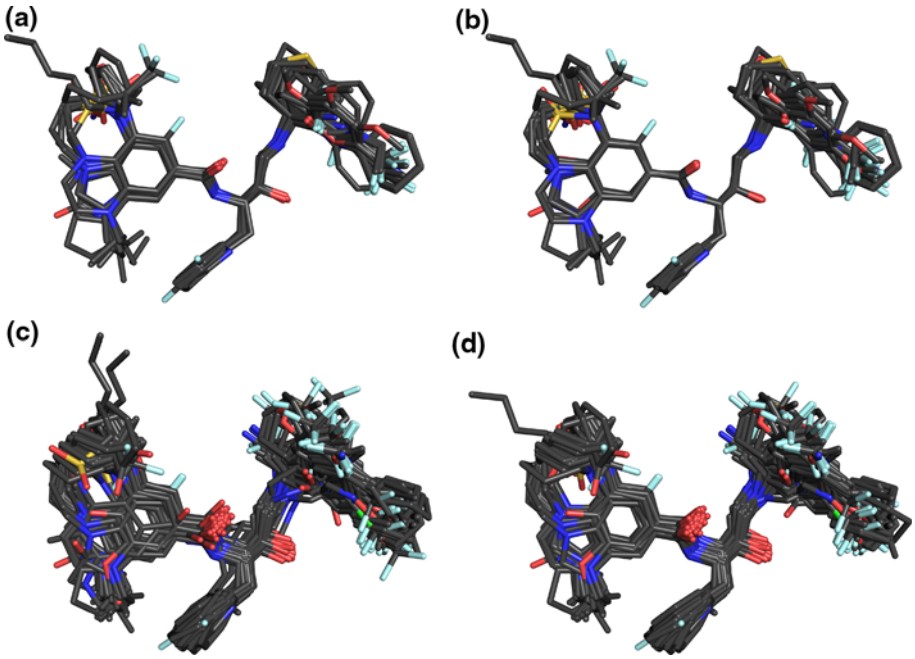


Fig. 3 Molecular alignments employed in this study. Each alignment presents all 128 data set compounds (training set and test set). **a** Alignment generated by the fragment-guided structural alignment methodology described in the Experimental Section. **b** Alignment obtained by a database alignment procedure of the alignment **A** in SYBYL. **c** Alignment generated by a rigorous docking protocol using GOLD 4.1. **d** Alignment obtained after a visual inspection of alignment **C** and the replacement of the docking poses for 8 compounds by alternatives in better agreement with the binding mode observed for the closely related analogs. See the Experimental Section for a detailed description of each particular alignment

structures, which indicate that these analogs share the same binding pattern, the docking procedures failed to yield equivalent spatial positions for identical substituents in different compounds (compare alignments **A** and **C**). Since the CoMFA method is highly sensitive to the relative position of the compounds in the molecular alignment, the misalignment might result in poor correlation between the fields and the biological property, yielding models with lower predictive power. Additionally, inadequately aligned groups may propitiate the appearance of contour fields with no physical meaning, possibly leading to a wrong interpretation. These artifacts might be difficult to detect, particularly in a model with artificially inflated statistical parameters.

In an attempt to reduce this bias, alignment **D** was generated by the replacement of docking poses for eight compounds by alternatives that were in better agreement with the binding mode observed for the closely related analogs. It must be made clear that this time consuming and work demanding procedure did not eliminate the misalignment of the data set molecules, but only reduced it by a substantial amount. Since the value of using docking procedures to generate molecular alignments for their use in 3D-QSAR studies has been extensively established, the program GOLD being particularly popular [23, 40], our purpose was studying the effects of substituents in discordance with X-ray determined positions of groups in the CoMFA predictive capability and contour maps.

The same 102 compounds of each alignment were assigned to the training set (Supplementary Table 1) and several PLS analyses were carried out using the CoMFA method and only the training set compounds. The models were optimized using the CoMFA region focusing method, which was weighted by StDev × Coefficient values ranging from 0.3 to 1.5, and grid spacing ranging from 0.5 to 1.5. This strategy not only increased q^2 values during the process of model generation, but also resulted in the refinement of 3D contour maps. The best statistical results are presented in Table 2.

A recent study by Mittal and coworkers [41] showed that the partial charge calculation method may affect CoMFA prediction accuracy, although this interesting study was focused mainly in the effect of different partial charge calculation methods on LOO q^2 and LMO q^2. The present study sought to evaluate and compare the effect of the Gasteiger-Huckel and the MMFF94 partial charge calculation methods on all statistical parameters, and most particularly on the q^2 and r^2_{pred}, which together are a better indicator of the external predictive potential of the CoMFA models.

Leave-Many-Out (LMO) procedures were performed as a more rigorous test of the stability and statistical significance of the model. The results confirmed the stability of the 3D QSAR models, since the statistical values obtained for the LMO analyses were comparable to those of the

Table 2 CoMFA results

Alignment	Charges	q^2	N	r^2	SEE	r^2_{pred}	Fraction	
							S	*E*
A	MMFF94	0.69	6	0.88	0.31	0.72	0.66	0.34
	Gast-Huck	0.66	6	0.86	0.32	0.70	0.76	0.24
B	MMFF94	0.66	6	0.87	0.31	0.69	0.66	0.34
	Gast-Huck	0.67	6	0.86	0.32	0.70	0.79	0.21
C	MMFF94	0.66	6	0.88	0.30	0.40	0.64	0.36
	Gast-Huck	0.67	6	0.90	0.28	0.46	0.69	0.30
D	MMFF94	0.69	5	0.91	0.27	0.48	0.71	0.29
	Gast-Huck	0.70	5	0.90	0.28	0.51	0.70	0.30

Gast–Huck Gasteiger–Huckel charges, q^2 leave-one-out (LOO) cross-validated correlation coefficient, N optimum number of components, r^2 non-cross-validated correlation coefficient, *SEE* standard error of estimate, r^2_{pred} predictive r^2, *S steric field*, *E* electrostatic field

LOO analyses (Supplementary Table 2). The progressive scrambling method, applied to check for the possibility of correlations and to test the stability of the models, further confirmed their consistency.

The predictive ability of the internally consistent CoMFA models generated employing the 102 training set molecules (compounds 1–102, Supplementary Table 1) was assessed by predicting the biological activities of the 26 external test set analogs (compounds 102–128, Supplementary Table 1), which were completely excluded during the generation of the models. The generation and alignment of the test set compounds followed the exact same procedures used to obtain the training set and described in the Experimental Section. The results of the external validation process are listed in the Supplementary Table 3, and the graphic results for the eight CoMFA models simultaneously displayed in Fig. 4.

The present results allow the identification of several interesting remarks. As presented in Table 2, statistically robust 3D QSAR models with significant correlation coefficients ($q^2 > 0.65$ and an $r^2 > 0.85$) were obtained for each combination of alignment and partial charge calculation method. However, the models differ significantly in their external predictive power, as measured by r^2_{pred}. This feature can also be observed in the Fig. 4, where the dispersion presented by the test set compounds in the case of alignments **C** and **D** is visibly more prominent. It is clear that the alignment has a greater influence on r^2_{pred} than r^2, SEE and q^2, which may lead to the delusion that a model is reliable unless external validation procedures are carefully considered [42]. As expected, the influence of the alignment in the external predictive power is greater than that of the partial charges, but the data suggests that the influence of different partial charge calculation methods is more noticeable for the alignments yielding lower r^2_{pred} values.

For the QSAR model employing alignment **A** and MMFF94 partial charges, the predicted values fall close to the experimental pIC_{50} values, not deviating by more than 0.75 logarithm unit, except for compound 105, for which the predicted value is still below one logarithm unit (Fig. 4 and Supplementary Table 3). Additionally, an r^2_{pred} value of 0.72 was obtained for this model (Table 2), indicating that accurate predictions can be achieved with the present methodology. The good agreement between actual and predicted pIC_{50} for the test set compounds in the 3D QSAR suggests that this particular model is reliable and can be used for the design of inhibitors with improved potency.

These results demonstrate the worth of the fragment-guided approach described here to align flexible and structurally diverse compounds provided that crystal structures are available. When compared to the popular docking approaches, the method performs particularly well, especially regarding its external prediction power. The use of crystallographic information in 3D QSAR studies is not a recent trend, but the most common approaches (i.e. docking) usually employ only one X-ray structure. The methodology described in this paper appears to suit better the cases when multiple crystal structures are available, which is expected for most relevant pharmaceutical targets.

CoMFA contour maps

The CoMFA method can be employed to develop models reliable enough to predict the activities of untested molecules, and highlight structural and chemical features that are directly related to biological activity. The CoMFA contour maps provide information about regions in space responsible for increases or decreases in the studied biological property. Unless otherwise stated, the 3D contour maps presented here were derived from the model generated using alignment **A** and MMFF94 charges. This model presents very good statistics (Table 2), performing particularly well in the external validation procedures ($r^2_{pred} = 0.72$, Table 2), and the contour maps identified a good set of the already established structure–activity relationships for the present data set. However, the contour maps of every model in Table 2 were rigorously inspected, and a comparison between representative contour maps is presented later. The contour maps are displayed as PLS standard deviation × coefficient.

The first peculiar feature of the contour maps is their distribution around the space regions where the data set ligands present structural variations, indicative of an accurate alignment (Figs. 5 and 6). The pattern of field distribution is a consequence of the CoMFA methodology, which correlates variations in biological property to differences in shape. Consequently, a common substructure

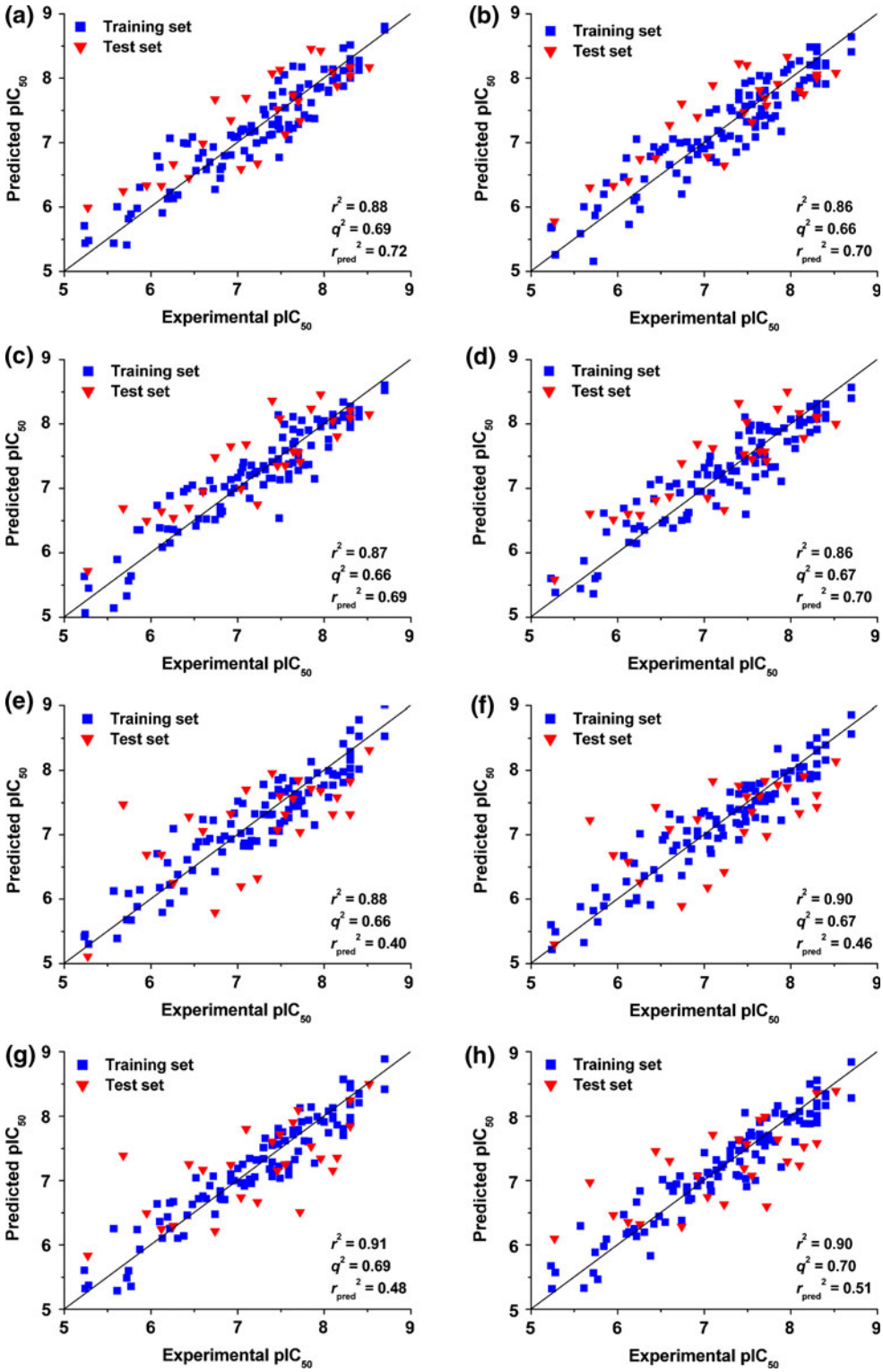


Fig. 4 Plot of predicted values of pIC_{50} *versus* the corresponding experimental values for the 102 training set compounds (*blue squares*) and the 26 test set compounds (*red triangles*). **a** Alignment **A** and MMFF94 partial charges, **b** Alignment **A** and Gasteiger–Huckel partial charges, **c** Alignment **B** and MMFF94 partial charges, **d** Alignment **B** and Gasteiger–Huckel partial charges, **e** Alignment **C** and MMFF94 partial charges, **f** Alignment **C** and Gasteiger–Huckel partial charges, **g** Alignment **D** and MMFF94 partial charges, **h** Alignment **D** and Gasteiger–Huckel partial charges. Consult the Experimental Section for a detailed description of each particular alignment

shared by all data set compounds is not expected to present strong field contributions, particularly for a well superposed data set.

The green contour region on the right delimits the ideal location of the substituent on the prime side, more specifically at the $S2'_{subpocket}$, which is frequently occupied by a *meta*-CF_3- or a *meta*-OCH_3-benzyl substituent for optimal potency (Supplementary Fig. 3 and 4) [17]. The detrimental steric field (Fig. 5, red arrow) indicates that a benzyl substituent is disfavored in this location, despite the

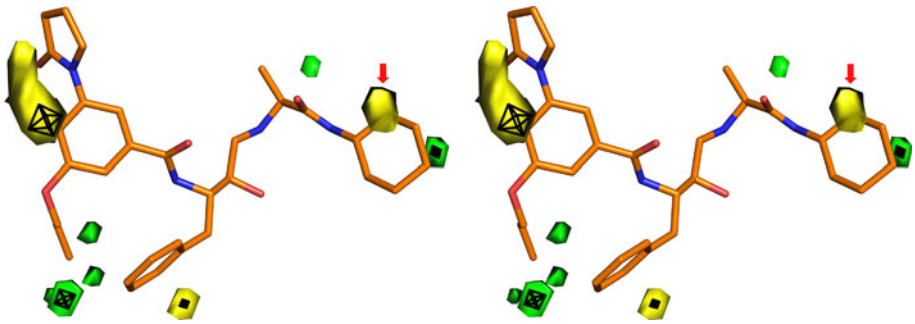


Fig. 5 Stereoview of the 3D-QSAR CoMFA steric contour maps. Compound **8** is shown as reference in *orange*, and the steric contour maps are shown in *yellow* (detrimental) and *green* (beneficial). The small meshes inside the colored contour surfaces were generated by applying a higher contour level, and can be used as a reference (contour levels surface −0.015 (*yellow*) and 0.05 (*green*), mesh −0.03 and 0.1)

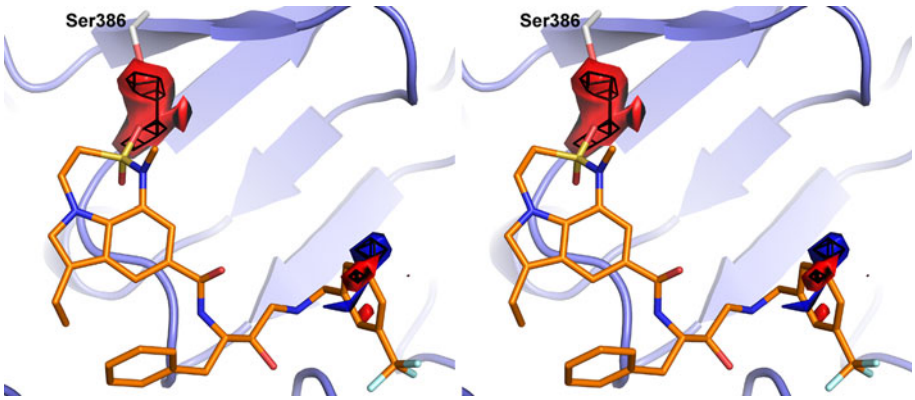


Fig. 6 Stereoview of the 3D-QSAR CoMFA electrostatic contour maps. Compound **52** is shown as reference, and Ser386 is labeled. The electrostatic contour maps are shown in *red* (electronegative groups in these positions are favored) and *blue* (electropositive groups in these positions are favored). The small meshes inside the colored contour surfaces were generated by applying a higher contour level, and can be used as a reference (contour levels surface −0.02 (*red*) and 0.02 (*blue*), mesh −0.04 and 0.05)

most potent compounds in the dataset presenting a *meta*-substituted benzyl group in this position. The key to understanding this apparent paradox lies in the *meta*-substituent of the benzyl group, which contacts an important favorable steric contour (green contour with a mesh, on the right side), compensating the detrimental contribution of the benzyl group. This observation suggests that less voluminous substituents yet still capable of filling the $S2'_{subpocket}$ may present improved potency.

The large detrimental contribution depicted in the upper left side of Fig. 5 delimits the location of the group occupying the S2 pocket, which is positioned between Gln134 from the flap and the loop formed by Thr292, Thr293 and Asn294 to which it makes extensive contact (Supplementary Fig. 3). Another detrimental steric contour is situated below the substituent in the P1 position. As depicted in the Supplementary Fig. 3, this field is located near the region occupied by the Phe169, demarcating the S1 pocket.

The positioning of two favorable steric contours at the very bottom of the S3 pocket is intriguing, since most compounds in the data set present small substituents in this interaction pocket, including the most potent analogs. This data set particularity derives from structure–activity relationships suggesting that increasing the size of the substituent at this position may decrease the selectivity towards other enzymes [16]. However, the addition of two carbon atoms to the P3 position allowed for an 18-fold potency increase in compound **6** in comparison to compound **8**, illustrating the role of this lipophilic interaction in enhancing potency. Moreover, the crystal structure of compound **8** in complex with BACE-1 presents a different conformation of the 10 s loop, leading to an increase in the volume of the $S3_{subpocket}$. These observations and the predictions for compounds presenting increased alkyl - chains suggest that analogs with longer alkyl substituents could achieve improved potency by occupying the $S3_{subpocket}$.

As depicted in Fig. 6, the most prominent feature of the electrostatic contour maps is a large region favorable for electronegative groups near Ser386. All data set compounds hydrogen bond Asn294 and abolishing this interaction leads to a one hundred-fold loss of potency [21]. Many analogs present a sulfonamide group, or a corresponding tricyclic derivative, allowing the essential hydrogen bond to Asn294 as well as an additional interaction with Ser386, as does the compound **52** (Fig. 6). The electrostatic contours on the prime side are exposed to the solvent and difficult to rationalize, but the only blue contour region may be related to a detrimental contribution of the amide carbonyl oxygen of some of the data set analogs presenting a higher peptidic character.

A careful analysis of the steric and electrostatic contour maps allowed the observation that the sultam derivatives present better electrostatic complementarity to the inhibitor binding site than the pyrrolidinone derivatives, probably due to an additional interaction with Ser386. However, as depicted in Fig. 7, the aliphatic portion of the sultam group is in contact with a detrimental steric contour, indicating an unfavorable steric contribution to the potency. The CoMFA contour maps are in agreement with the approach used by GlaxoSmithKline R&D to explore the chemical space around this position to reduce the detrimental steric contributions while retaining the better electrostatic complementarity [21, 22].

The contour maps derived from the fragment-guided approach (alignment **A**) and MMFF94 partial charges yielded results that are consistent with the experimental data. To address the question of how the other models would perform, representative contour maps generated from models employing different alignment strategies and partial charge calculation methods were compared (Fig. 8). It is worthy of note that the contour fields are also dependent on the focusing parameters and the relative contributions from the steric and electrostatic fields. For that reason, partial charges not only influence electrostatic contours, but may also exert an indirect effect on steric maps. Therefore, even models deriving from the same alignment and partial charges may not present the same contours, and comparison of maps obtained by different models must be made with caution. Another consequence is that contour levels may not be directly compared.

The steric contours from the models derived from the alignment **A** employing either MMFF94 or Gasteiger-Huckel partial charges share more common features than the electrostatic contours (Fig. 8a, b). The larger apparent influence of the partial charges in the electrostatic contour maps is a recurrent feature, as might be expected. As a general trend, the GOLD derived alignments presented more fragmented contour maps, with contour regions associated to positions where the data set ligands do not present structural variation (Fig. 8c, d). Although all these models present acceptable statistics ($q^2 > 0.65$, Table 2), their contour maps present almost no common features (Fig. 8a–d), which further reinforces the influence of the alignment on the CoMFA results, as well as the role of external validation in model selection. In conclusion, the models derived from the fragment-guided approach presented better external prediction power than the docking approach, and the analyses of its contour maps highlighted relevant structure–activity relationships.

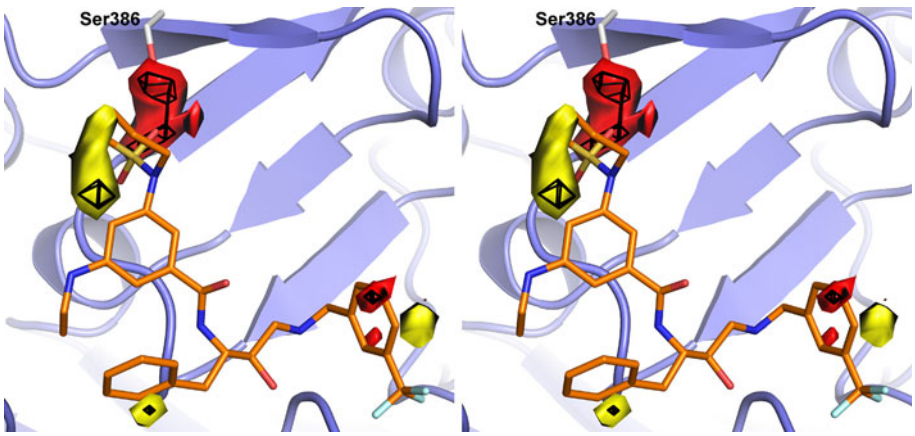


Fig. 7 Stereoview of the 3D-QSAR CoMFA contour maps. Compound **111** is shown as reference, and Ser386 is labeled (PDB code 2VNM). The electrostatic contour regions referent to spatial positions where electronegative groups are favored are colored in *red*, and the detrimental steric contour regions are colored in *yellow*. The small meshes inside the colored contour surfaces were generated by applying a higher contour level, and can be used as a reference (contour levels surface −0.02 (*red*) and −0.015 (*yellow*), mesh −0.04 and 0.03)

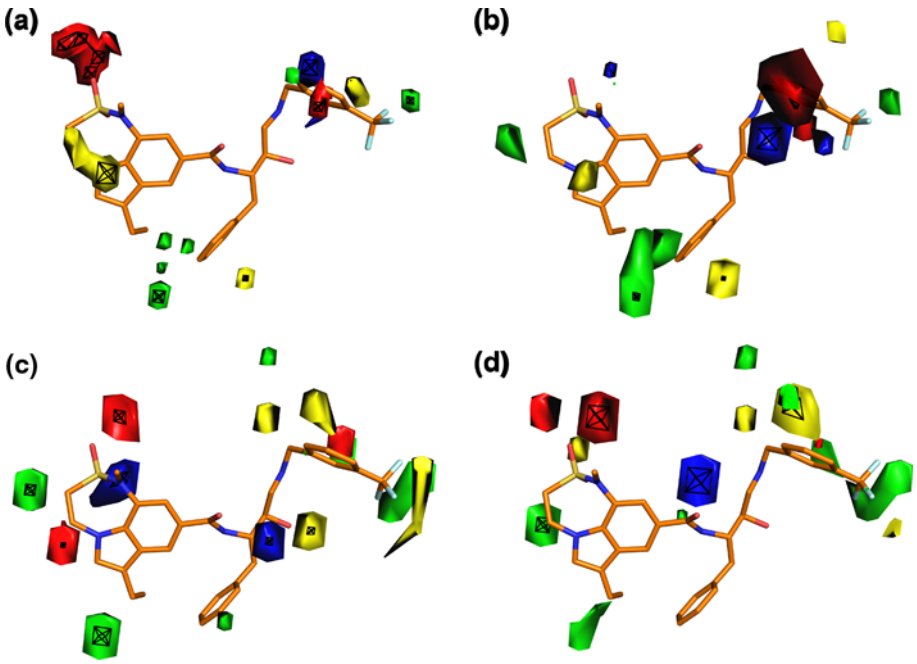


Fig. 8 Contour maps generated by models obtained by different combinations of alignments and partial charges calculation methods. **a** Alignment **A** and MMFF94 partial charges (steric contour levels 0.05 and −0.015; electrostatic contour levels 0.02 and −0.02), **b** Alignment **A** and Gasteiger–Huckel partial charges (steric contour levels 0.12 and −0.05; electrostatic contour levels 0.04 and −0.04), **c** Alignment **C** and MMFF94 partial charges (steric contour levels 0.07 and −0.5; electrostatic contour levels 0.07 and −0.08), **d** Alignment **D** and MMFF94 partial charges (steric contour levels 0.7 and −0.5; electrostatic contour levels 0.02 and −0.08). In all cases the mesh contour is twice the surface contour

Conclusion

Aiming to take advantage of the ever increasing number of X-ray structures for validated pharmaceutical targets, we have developed a new fragment-guided approach to incorporate structural information from multiple crystal structures into CoMFA studies. The methodology, which has the advantage of requiring only the SYBYL software platform to be implemented, employs two dimensional frequency-based fragment descriptors to assign a reference X-ray co-crystal structure to each inhibitor in the data set, allowing the generation of crystal structure derived alignments. CoMFA models obtained using this novel structure-based approach were generated for a large data set containing 128 BACE-1 inhibitors and presented substantially better external validation power ($r^2_{pred} = 0.72$) than our best docking derived models (program GOLD, $r^2_{pred} = 0.51$). Additionally, comparisons between different partial charge calculation methods suggest that their influence may be more preeminent when the molecular alignment yields models with lower r^2_{pred}. The visual analysis of the resultant contour maps corroborated previously known structure–activity relationships, and allowed the identification of new opportunities for potency optimization of BACE-1 inhibitors. These results suggest that the fragment-guided approach presented here represents a valuable option to integrate additional structural information into CoMFA studies.

Acknowledgment We would like to thank Dr. Emmanuel Demont for kindly and promptly clarifying diverse aspects about the data set inhibitors and crystal structures, and Dr. Richard Charles Garratt for critical reading of the manuscript. This work was supported by The State of São Paulo Research Foundation (FAPESP, grants 2008/58316-5 and 2007/07294-9).

References

1. Mount C, Downton C (2006) Alzheimer disease: progress or profit? Nat Med 12:780–784
2. Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA (2003) Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 60:1119–1122
3. Ernst RL, Hay JW (1994) The US economic and social costs of Alzheimer's disease revisited. Am J Public Health 84:1261–1264
4. Stachel SJ (2009) Progress toward the development of a viable BACE-1 inhibitor. Drug Dev Res 70:101–110
5. Melnikova I (2007) Therapies for Alzheimer's disease. Nat Rev Drug Discov 6:341–342
6. Vassar R (2002) Beta-secretase (BACE) as a drug target for Alzheimer's disease. Adv Drug Deliv Rev 54:1589–1602
7. Haass C, Selkoe DJ (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nat Rev Mol Cell Biol 8:101–112
8. Hunt CE, Turner AJ (2009) Cell biology, regulation and inhibition of beta-secretase (BACE-1). FEBS J 276:1845–1859
9. Hardy JA, Higgins GA (1992) Alzheimer's disease: the amyloid cascade hypothesis. Science 256:184–185
10. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297:353–356
11. Hardy J (2009) The amyloid hypothesis for Alzheimer's disease: a critical reappraisal. J Neurochem 110:1129–1134
12. Luo Y, Bolon B, Kahn S, Bennett BD, Babu-Khan S, Denis P, Fan W, Kha H, Zhang J, Gong Y, Martin L, Louis JC, Yan Q, Richards WG, Citron M, Vassar R (2001) Mice deficient in

BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation. Nat Neurosci 4:231–232
13. Roberds SL, Anderson J, Basi G, Bienkowski MJ, Branstetter DG, Chen KS, Freedman SB, Frigon NL, Games D, Hu K, Johnson-Wood K, Kappenman KE, Kawabe TT, Kola I, Kuehn R, Lee M, Liu W, Motter R, Nichols NF, Power M, Robertson DW, Schenk D, Schoor M, Shopp GM, Shuck ME, Sinha S, Svensson KA, Tatsuno G, Tintrup H, Wijsman J, Wright S, McConlogue L (2001) BACE knockout mice are healthy despite lacking the primary beta-secretase activity in brain: implications for Alzheimer's disease therapeutics. Hum Mol Genet 10: 1317–1324
14. Silvestri R (2009) Boom in the development of non-peptidic beta-secretase (BACE1) inhibitors for the treatment of Alzheimer's disease. Med Res Rev 29:295–338
15. Hong L, Koelsch G, Lin X, Wu S, Terzyan S, Ghosh AK, Zhang XC, Tang J (2000) Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. Science 290:150–153
16. Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 1: identification of novel hydroxy ethylamines (HEAs). Bioorg Med Chem Lett 18:1011–1016
17. Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Hawkins J, Hussain I, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Soleil V, Smith KJ, Stanway S, Stemp G, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 2: identification of hydroxy ethylamines (HEAs) with reduced peptidic character. Bioorg Med Chem Lett 18: 1017–1021
18. Beswick P, Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Faller A, Gleave R, Hawkins J, Hussain I, Johnson CN, MacPherson D, Maile G, Matico R, Milner P, Mosley J, Naylor A, O'Brien A, Redshaw S, Riddell D, Rowland P, Skidmore J, Soleil V, Smith KJ, Stanway S, Stemp G, Stuart A, Sweitzer S, Theobald P, Vesey D, Walter DS, Ward J, Wayne G (2008) BACE-1 inhibitors part 3: identification of hydroxy ethylamines (HEAs) with nanomolar potency in cells. Bioorg Med Chem Lett 18:1022–1026
19. Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, East P, Hawkins J, Howes C, Hussain I, Jeffrey P, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2008) Second generation of hydroxyethylamine BACE-1 inhibitors: optimizing potency and oral bioavailability. J Med Chem 51:3313–3317
20. Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, Hawkins J, Howes C, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009) Second generation of BACE-1 inhibitors. Part 1: the need for improved pharmacokinetics. Bioorg Med Chem Lett 19:3664–3668
21. Charrier N, Clarke B, Demont E, Dingwall C, Dunsdon R, Hawkins J, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009) Second generation of BACE-1 inhibitors part 2: optimisation of the non-prime side substituent. Bioorg Med Chem Lett 19:3669–3673
22. Charrier N, Clarke B, Cutler L, Demont E, Dingwall C, Dunsdon R, Hawkins J, Howes C, Hubbard J, Hussain I, Maile G, Matico R, Mosley J, Naylor A, O'Brien A, Redshaw S, Rowland P, Soleil V, Smith KJ, Sweitzer S, Theobald P, Vesey D, Walter DS, Wayne G (2009) Second generation of BACE-1 inhibitors part 3: towards non hydroxyethylamine transition state mimetics. Bioorg Med Chem Lett 19:3674–3678
23. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (Comfa) 0.1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110: 5959–5967
24. The PyMOL molecular graphics system, version 1.2r3pre, Schrödinger, LLC
25. Valadares NF, Castilho MS, Polikarpov I, Garratt RC (2007) 2D QSAR studies on thyroid hormone receptor ligands. Bioorg Med Chem 15:4609–4617
26. Salum LB, Andricopulo AD (2010) Fragment-based QSAR strategies in drug design. Expert Opin Drug Discov 5:405–412
27. Tropsha A, Golbraikh A (2007) Predictive QSAR modeling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13:3494–3504
28. Klebe G, Abraham U, Mietzner T (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. J Med Chem 37:4130–4146
29. Ortiz AR, Pisabarro MT, Gago F, Wade RC (1995) Prediction of drug binding affinities by comparative binding energy analysis. J Med Chem 38:2681–2691
30. Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T (2000) GRID/CPCA: a new computational tool to design selective ligands. J Med Chem 43:3033–3044
31. Gohlke H, Klebe G (2002) DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. J Med Chem 45:4153–4170
32. Sutherland JJ, O'Brien LA, Weaver DF (2004) A comparison of methods for modeling quantitative structure-activity relationships. J Med Chem 47:5541–5554
33. Salum LB, Polikarpov I, Andricopulo AD (2008) Structure-based approach for the study of estrogen receptor binding affinity and subtype selectivity. J Chem Inf Model 48:2243–2253
34. Valadares NF, Salum LB, Polikarpov I, Andricopulo AD, Garratt RC (2009) Role of halogen bonds in thyroid hormone receptor selectivity: pharmacophore-based 3D-QSSR studies. J Chem Inf Model 49:2606–2616
35. Schüürmann G, Ebert RU, Chen J, Wang B, Kühne R (2008) External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. J Chem Inf Model 48:2140–2145
36. Patel S, Vuillard L, Cleasby A, Murray CW, Yon J (2004) Apo and inhibitor complex structures of BACE (beta-secretase). J Mol Biol 343:407–416
37. Hong L, Tang J (2004) Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. Biochemistry 43:4689–4695
38. Zhu Z, Sun ZY, Ye Y, Voigt J, Strickland C, Smith EM, Cumming J, Wang L, Wong J, Wang YS, Wyss DF, Chen X, Kuvelkar R, Kennedy ME, Favreau L, Parker E, McKittrick BA, Stamford A, Czarniecki M, Greenlee W, Hunter JC (2010) Discovery of cyclic acylguanidines as highly potent and selective beta-site amyloid cleaving enzyme (BACE) inhibitors: Part I-inhibitor design and validation. J Med Chem 53:951–965
39. Steele TG, Hills ID, Nomland AA, de León P, Allison T, McGaughey G, Colussi D, Tugusheva K, Haugabook SJ, Espeseth AS, Zuck P, Graham SL, Stachel SJ (2009) Identification of a small molecule beta-secretase inhibitor that binds without catalytic aspartate engagement. Bioorg Med Chem Lett 19:17–20
40. Xie A, Odde S, Prasanna S, Doerksen RJ (2009) Imidazole-containing farnesyltransferase inhibitors: 3D quantitative

structure-activity relationships and molecular docking. J Comput Aided Mol Des 23:431–448
41. Mittal RR, Harris L, McKinnon RA, Sorich MJ (2009) Partial charge calculation method affects CoMFA QSAR prediction accuracy. J Chem Inf Model 49:704–709
42. Golbraikh A, Tropsha A (2002) Beware of q2!. J Mol Graph Model 20:269–276