UNIVERSITÀ DEGLI STUDI DI NAPOLI
"FEDERICO II"

FACOLTÀ DI FARMACIA
DIPARTIMENTO DI CHIMICA FARMACEUTICA E TOSSICOLOGICA

Dottorato di Ricerca in
Scienza del Farmaco XX ciclo

Ensemble-docking approach on BACE-1: Pharmacophore Perception and Guidelines for Drug Design

Coordinatore
Chiar.ma Prof.ssa
MARIA VALERIA D'AURIA

Supervisore
Chiar.mo Prof.
ETTORE NOVELLINO

Candidato
VITTORIO LIMONGELLI

Cover Picture

Vittorio Limongelli, Luciana Marinelli*, Sandro Cosconati, Hannes A. Braun, Boris Schmidt, and Ettore Novellino

The cover picture shows a model of the BACE-1 enzyme represented in ribbon mode with the catalytic site highlighted as a grey Connolly surface. The binding site is filled with six BACE-1 inhibitors in their docked conformations and coloured according to conserved interaction points. Theoretical studies of inhibitor binding allowed us to delineate an exhaustive nine-point pharmacophore model (upper right), which captures both the common geometric and the electronic features essential for enzyme inhibition. Interestingly, five of these points are present in all the inspected ligands (blue spheres), thus they can be referred to as essential features for inhibitor recognition and binding, whereas the other four points (light-blue spheres) contribute to BACE-1 selective inhibition. For more details, see the Full Paper by L. Marinelli et al. on p. 667 ff.

“One day the fear knocked at the door, went the courage to open and did not find anyone.”

(Martin Luther King)

A mio fratello, il mio Migliore Amico

RINGRAZIAMENTI

Ringrazio il mio tutor Prof. Ettore Novellino per avermi dato la possibilità di svolgere la mia attività di ricerca in questi anni senza farmi mai mancare il suo supporto e la fiducia nei miei mezzi.

Un ringraziamento particolare va al Prof. Enrico Abignente, coordinatore del dottorato in Scienza del Farmaco, in lui noi dottorandi abbiamo sempre trovato grande disponibilità e fonte inesauribile di consigli.

Ringrazio molto sentitamente la Dott.ssa Luciana Marinelli senza la quale il mio percorso umano prima che scientifico sarebbe stato sicuramente molto più impervio e di certo meno prolifico. Devo a lei gran parte di quello che sono oggi. Semplicemente GRAZIE Luci.

Ringrazio il Dott. Sandro Cosconati, in lui ho trovato di sicuro un amico su cui contare oltre che un validissimo “compagno” di lavoro.

Contents

Computational Methods

CHAPTER I

INTRODUCTION

1.1 Alzheimer disease (AD)

Alzheimer disease (AD) is a cerebral neurodegenerative pathology that is characterized by the progressive formation of insoluble amyloid plaques and fibrillary tangles, first described by Alois Alzheimer nearly a century ago.[1] The major protein component of the plaques is the amyloid-β peptides (Aβ), and the tangles are composed of filaments of the microtubule-associated protein tau.

Although AD is the most common cause of dementia in western industrialized countries, up to now, there is no approved causal treatment. The available symptomatic treatments or disease modifiers provide only limited benefits to the affected people. The approved drugs, such as Vitamin E or AChE inhibitors, slow down, but do not stop the disease progression.[2] Thus, a growing need exists for new effective therapies with a specific mode of action, which allows to control the onset and the progression of the disease. Over the last decade, great attention has been paid to the cascade of physiological events that contribute or accompany AD.[3-5]

Many pathological analysis at cytological and biochemical levels, reveals that, in AD, fibrillar forms of Aβ densely deposit forming neuritic

plaques in cerebral and midbrain regions associated with cognition and memory. Moreover, it has been demonstrated that fibrillary tangles, also localized in the brain regions critical to higher brain function, were found to be characterized by the filamentous form of the hyperphosphorylated tau protein. Tau hyperphosphorylation renders insoluble this otherwise highly soluble cytosolic protein, and this modification seems to be essential for the pathogenic profile of these tangles. Interestingly, tau-containing neurofibrillary tangles occur in a number of other, uncommon neurodegenerative diseases, while the amyloid containing neuritic plaques are unique to AD.

A major point of debate in the last recent years was if these lesions are causative or simply “tombstones” or markers of regions that have degenerated due to unknown pathogenic events. After the identification of genes associated with AD, it has been unequivocally demonstrated that alterations in the proteolytic processing that produces Aβ protein can cause the disease. As a consequence, Aβ is either the molecular culprit or an intimately linked epiphenomenon, in either event, inhibiting the responsible proteases would be a fruitful strategy to deal with AD.

The amyloid hypothesis.

The above reported observations embody the amyloid hypothesis of AD pathogenesis, which states that production and deposition of Aβ fibrils leads to neuronal cell death and eventually to the clinical onset and progression of AD (Figure 1).

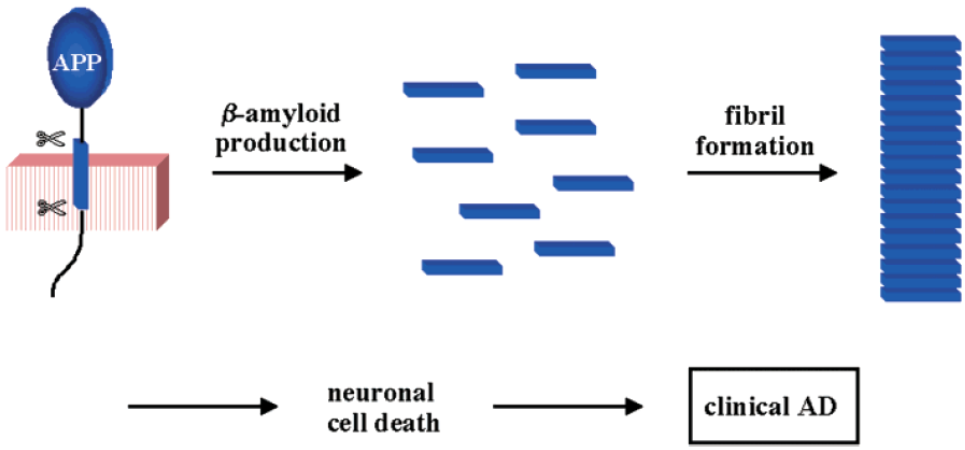


Figure 1. The amyloid hypothesis of AD pathogenesis.

In the recent years, key pieces in the puzzle of AD have been falling into place. Particularly, the secretases responsible of the APP processing are no longer completely mysterious and it is generally accepted that, the β-amyloid precursor protein (APP) is cleaved by two proteases to generate the

40/42 amino acid long amyloid-β peptides (Aβ). The increased Aβ formation results in extracellular amyloid plaques deposition and it is accompanied by the intracellular formation of neurofibrillary tangles in the brain.[5,6] The neurotoxicity associated with the Aβ• oligomerisation is supposed to cause neuronal death, brain inflammation, and finally AD.[7] The APP is processed via the major α- or the minor β-secretase pathway; both produce fragments further processed by γ-secretase.[8]

Differently from the non pathogenic products of α/γ-secretase pathway, the β/γ-secretase pathway produces pathogenic Aβ peptides.

Certainly, an understanding of the mechanism of γ-secretase specificity may suggest strategies to selectively decrease the production of the more deleterious Aβ peptide. Even though γ-secretase is an aspartyl protease, its peculiar active site topology makes this enzyme less tractable with respect to other common proteases. Indeed, the two key aspartates crucial for proteolytic activity seems to be embedded in the membrane, and other members of what is likely to be a larger γ-secretase complex have not been yet identified. For all these reasons, the X-ray structure of this secretase is far to be detected soon. Thus, only the information derived from affinity

labelling with transition-state analogue inhibitors and site directed mutagenesis may be used to characterize its active site.

Alternative processing of APP by α-secretase precludes Aβ production, since this proteolysis occurs within the Aβ sequence at Lys686-Leu689 generating non pathogenic peptides. Two missense mutations in APP near the α-secretease cleavage site, cause disease and an increased Aβ production, as more APP is shunted into the β-secretase pathway. In addition, evidence suggests that α-APPs may be neuroprotective with general beneficial effects on learning and cognition, so augmenting α-secretase processing of APP may be a valuable strategy in treating AD.

After the demonstration that β-secretase (BACE-1), a member of the pepsin family of aspartyl proteases, is the rate-limiting enzyme in the production of Aβ,[9] and that its genetic depletion in mice abolishes the β-amyloid formation without major side effects,[10] BACE-1 has emerged as a leading target for the therapeutic treatment of Alzheimer disease.[11] Recently, BACE-1 was shown to control the myelination of the peripheral nerves in the late foetal development, the relevance of this finding to chronic treatment of adults will have to be considered.[12]

1.2 The current state of art on BACE-1

So far, two β-secretases are known, BACE-1 and BACE-2, and they exhibit 52% amino acid sequence identity and 68% similarity and cleave APP and short peptides in a very similar manner. Interestingly, BACE-2 is mainly expressed in highly vascularized systemic tissues and not well in the brain, with the implication that it has not a key role in the AD plaque formation. Furthermore, while BACE-1 preferentially cleaves APP on the lumenal/extracellular side at approximately 30 residues from the transmembrane domain at the sequence EVKM*DAEF (the asterisk denotes the cleavage site), BACE-2 makes an additional proteolysis in the middle of the Aβ region, suggesting that BACE-2 might limit the production of pathogenic forms of Aβ. If such were the case, the achievement of a selective inhibition of BACE-1, leaving BACE-2 active, is of great interest.

Up to now, several X-ray structures of BACE-1 (hereinafter "BACE") have been reported, either in the apo form (PDB codes: 1SGZ and 1W50), either in complex with large-size peptidomimetic ligands (1FKN, 1M4H, 1XN2, 1XN3, 1XS7, 2F3E, 1YM2, 1YM4, 2B8L, 2B8V and 2FDP), or with rather small inhibitors (1W51, 1TQF, 2G94). An important advance in the elucidation of the inhibitor-BACE recognition process has been provided by

the 1W51 structure, where the enzyme has been cocrystallized with the inhibitor **1**.[13] Figure 2 highlights the main interactions between **1** and the BACE-1 enzyme.

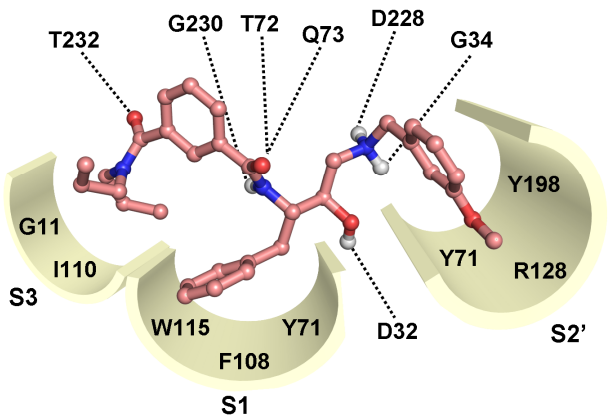


Figure 2. Schematic representation of the main interactions of **1** with the BACE catalytic site.

A detailed comparison of the available X-ray structures suggests that BACE can assume at least two major conformations mainly differing in the FLAP region, which can adopt an open and a closed conformation in the

ligand free and ligand bound enzyme, respectively. Thanks to the availability of all these structures, great strides in the development of new BACE inhibitors have been made by both academic and industrial research groups. As a result, Vertex, and recently Wyeth, reported new inhibitors bound to BACE with the flap open conformation.[14]

Many noncleavable transition state isosters were designed as new inhibitors on the basis of initial kinetics and substrate specificity data.[15] Most of these peptide ananlogues mimic the scissile amide bond of the endogenous substrates.[16,17] The hydroxyethylenes derivates, such as OM99-2 and OM00-3, represented the first class of highly potent BACE inhibitors.[15], The employment of the statine moiety led to the peptidomimetic compound **2** (IC50 = 20 nM), which features non peptidic portions at both C- and N-termini.[19] In the effort to reduce the peptidic character of the first inhibitors, several hydroxyethylamine-containing compounds were investigated as new BACE inhibitors. Among them, inhibitors **3** and **4** are of particular interest for their low nanomolar activity (IC50 = 1 and 1.4 nM, respectively) and for the originality of their structures, being the secondary amine of HEA arranged in a six-membered cycle.[20] With the aim of achieving selectivity for BACE over the other

aspartic proteases, a sulphonylamide group has been introduced in an HEA derivative leading to compound **6**.[21] Recently, high BACE-1 selective compounds, (e.g. compound **5**, IC50 = 4 nM) featuring a Ψ(CH2NH) reduced amide bond, were reported by Coburn et al.[22]

1.3 Future Perspective *via* new Computational Methodologies

Despite this considerable progress, it is worth noting that the majority of the reported peptidomimetics with low nanomolar activity in BACE-1 enzymatic assays are poorly active in cell-based assays because of the limited penetration across the cell membranes. Thus, non-peptidic inhibitors with a lower molecular weight, suitable for oral delivery and transport through the cell membranes and the blood-brain barrier are still in great demand. In spite of all efforts made by the pharmaceutical companies and academic groups, non-peptidic leads for BACE inhibition are still few.[23] Thus, BACE turns out to be a structurally challenging target having on one hand multiple sites for effective binding and, on the other hand, an high homology with other aspartyl proteases such as cathepsin D, pepsin or renin. Currently, medicinal chemists can choose among a number of novel approaches for drug discovery such as high-throughput in vitro screening,

combinatorial chemistry, focused library or pharmacophore-based and/or target structure-based virtual screening (VS), being the latter two approaches increasingly used having the advantage to avoid long and expensive experimental efforts. However, the pharmacophore-based VS is exclusively possible when a trustworthy pharmacophore model exists, while the structure-based VS needs the design of a proper protocol as it is well-known that several features, as the charge state of potential-interactive residues, the protein conformations (apo and ligand bound), the docking methods and the scoring functions can all deeply affect the success rate. In this regard, the works of Polgàr and Keserü[24] are particularly helpful, as the influence of protonation state of catalytic Asp residues of BACE (D32 and D228) as well as the enzyme conformations were investigated in a comparative VS. From these studies, it emerges that the monoprotonated form (D228-, D32) of the BACE catalytic site gave better enrichment factor compared to the default protonation state (D32 and D228 deprotonated), and ligands can find proper poses easier in a ligand-bound structure (FLAP closed), than in the unbound form (FLAP open). Interestingly, the introduction of pharmacophore constraints in the docking calculations improved enrichment factors for both structures (bound and unbound),

reducing ligands false positive poses and increasing the inactive drop-out rate.

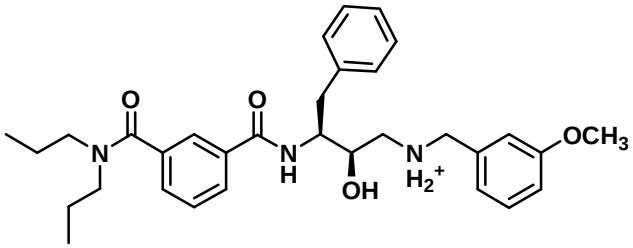
In structure-based VS, a 3-D pharmacophore model can be used either to constrict the number of the possible ligands poses either to pre-screen compounds databases, both helping to pursue a more accurate and saving-time simulations. The pharmacophore constraints used in the study of Polgàr and Keserü were retrieved from a patent document, where a congeneric series of BACE-1 inhibitors were presented.[14]

Here, with the aim to extend our understanding of inhibitors binding at the BACE-1 catalytic site and to provide an exhaustive structure-based pharmacophore model, the most active and selective (whenever it was possible) compound for each class of BACE inhibitors (see above) (**2**-**6**) was selected and subjected to an ensemble molecular docking process into five BACE X-ray structures. The superimposition of the calculated bioactive conformations of these inhibitors allowed us to capture both the common geometric and electronic features essential for the ligand recognition and the enzyme inhibition. Furthermore, in order to achieve BACE-1 selective inhibition, a comparison of the X-ray structures of

BACE-1 and the cathepsin D was made to better understand the structural and chemical differences in their respective catalytic sites.

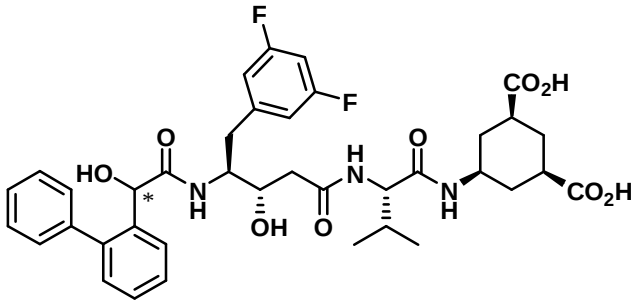
The elucidation of the different binding modes of the diverse ligands on one hand, and the development of a pharmacophore model on the other, are intended to extend our knowledge in the BACE field furnishing a support for pharmacophore- and/or structure-based VS techniques and a source for the optimization of the screened compounds as well as of the already known leads. Moreover, the proposed pharmacophore hypothesis can be of help in the common target-based and ligand-base drug design approaches as well as in the setting of a focused-library of BACE inhibitors.

Chart 1. Chemical structures of docked compounds.



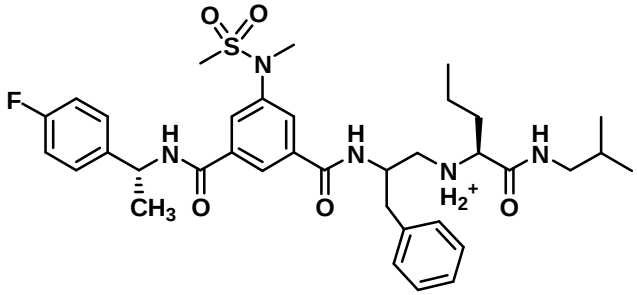
1

(IC_{50} = 200 nM)



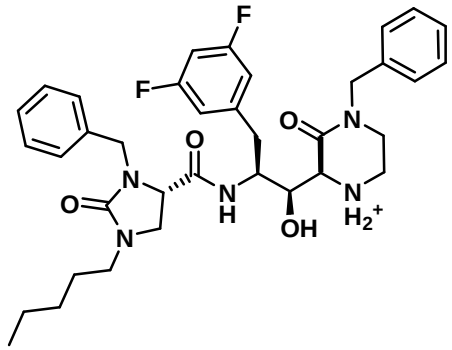
2

(IC_{50} = 20 nM)



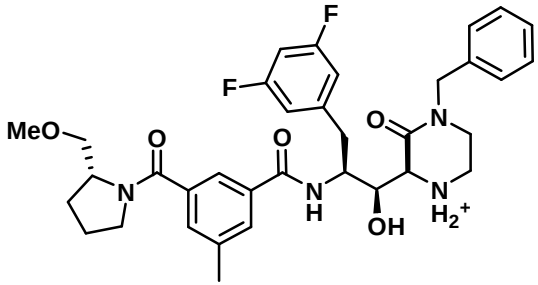
5

(IC_{50} = 4 nM)



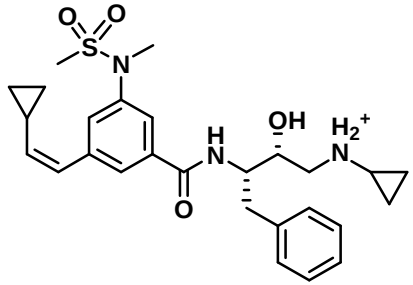
3

(IC_{50} = 1 nM)



4

(IC_{50} = 1.4 nM)



6

(IC_{50} = 35 nM)

CHAPTER II

RESULTS AND DISCUSSION

2.1 X-ray Structures Selection for the Ensemble Docking Studies

Even today, a major hurdle for a successful molecular docking is the protein flexibility. At the present time, many effective methods are available for docking a flexible ligand into a rigid protein, while docking calculations including target flexibility still remain problematic, both in terms of computational time and efficiency. In this respect, BACE shows the type of flexibility that can pose challenging problems in docking simulations as the enzyme is known to undergo a massive rearrangement of the FLAP region (residues 68-74) during association with ligands and a certain mobility is expected for the “10S loop” (residues 9-14). To the best of our knowledge, no docking program, that attempts to include wide flexibility of the protein, has been extensively validated so far. Fortunately, in the case of BACE, numerous crystallographic structures exist enabling us to use an ensemble of enzyme conformations for our docking calculations. Docking a ligand into a battery of binding pockets is a strategy to deal with the protein flexibility, although it is still far from perfection.

A comparative structural analysis of all available BACE structures revealed that the FLAP region is always in the closed conformation whenever an inhibitor, either peptidomimetic or non-peptidomimetic, is

bound. Because ligands **1**-**6** are substrate analogues that interact with the FLAP-closed form of BACE, we limited our studies to all BACE structures with a FLAP closed conformation (PDB codes: 1FKN, 1M4H, 1TQF, 1XN2, 1XN3, 1XS7, 1YM2, 1YM4, 2B8L, 2B8V, 2FDP, 2G94, 2F3E, 2F3F). With the aim of reducing the BACE structure redundancy, only the most divergent structures were considered for our docking calculations.

To assess the differences among the BACE structures, they were superimposed on the alpha carbon atoms (Cα) using 1W51 structure as reference. Interestingly, the FLAP closed conformations are all surprisingly similar regardless of the inhibitor type bound, while some differences came out in the side chains conformations of few residues, with the most notable one residing in Q73 (Figure 3). Additionally, some expected flexibility was also found in various residues lining the binding site cleft such as R128, T231, D307 and D235.

From the comparative analysis of BACE X-ray structures, it clearly emerges that the “10S loop”, a short loop located between two strands at the base of the S3 subpocket, shows mainly three low-energy conformations, an open (1FKN, 1XN3, 1XN2, 1XS7, 1YM2, 1M4H, 2F3F, 2G94) a closed

(1W51, 1FDP, 2B8L, 2B8V, 1YM4, 2F3E) and an outlier conformation (1TQF) (Figure 3).

In view of the capability of “10S loop” in affecting the shape of the S3 subpocket (S3sp) and thus the ligand binding,[25] the inclusion of such structural variability in a docking study becomes of fundamental importance. Consequently, 1FKN,[26] 1W51,[13] and 1TQF,[27] were chosen for our docking experiments since they cover the experimentally observed motions of the “10S loop” as well as of some residues in the catalytic site such as Q73, R128, R307, R235, and T231. Two additional structures were considered (1XN3 and 2G94)[28,29] so as to include additional conformers of Q73 and R235 residues. As a result, each ligand was docked in a total of five BACE structures (1FKN, 1W51,1TQF, 2G94, 1XN3).

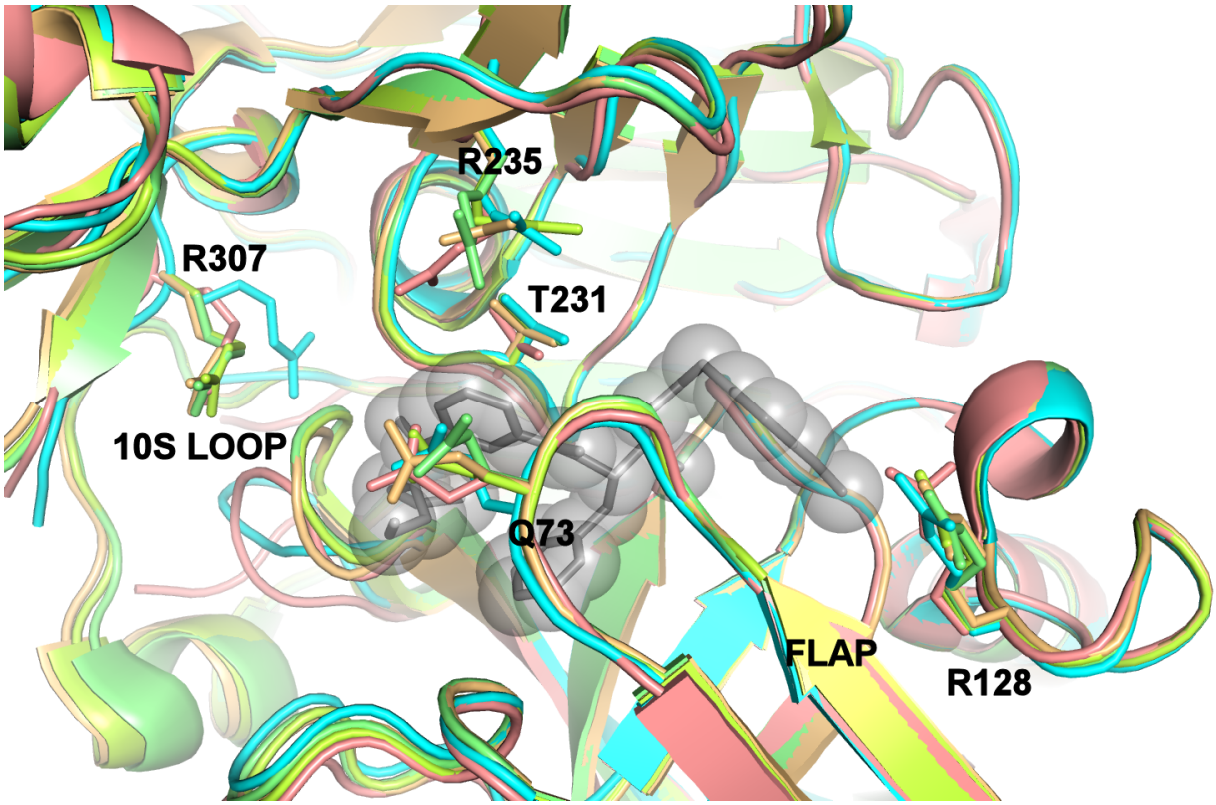


Figure 3. Cartoon representation of the BACE X-ray structures used for the ensemble-docking study (1W51, 1FKN, 1TQF, 1XN3, 2G94 are in pink, green, cyan, yellow and orange, respectively) with **1** shown as grey sticks and transparent grey spheres; hydrogen atoms are omitted for clarity reason. The most flexible residues, emerged from the BACE X-ray structures superposition, are represented in stick mode and coloured according to the BACE structures colour code.

2.2 Assessment of the Docking Program

Although AutoDock program is the most widely used docking program[30] and has been extensively validated, it is well-known that each docking algorithm performs better for certain protein systems than for others, thus the reliability of a docking program towards the target of interest has always to be assessed. Furthermore, testing a program by docking a ligand into its native protein is intrinsically biased because the protein has already changed its shape to better accommodate the ligand and this unavoidably affects positively the docking results. Here, with the aim to accurately evaluate the program performances on the studied system, a cross-docking experiment of **1** in its native enzyme (1W51) and in four non-native enzyme conformations (1FKN,1TQF, 2G94 and 1XN3) was conducted. It is generally accepted that, a successful docking result reproduces the crystallographic conformation of a ligand in the complex structure within ~2 Å of RMSD on all ligand atoms, and that the first-ranked docked conformation (herein referred as ranking conformation) is the preferable one. On the other hand, from our experience, in the case of Autodock program, the lowest energy conformation of the most populated cluster (herein referred as cluster conformation) has to be taken into account

as well. In the case of BACE, in two complexes, 1W51 and 2G94, the conformation calculated by AutoDock with the lowest free energy of binding belongs to the most populated cluster, thus, no ambiguity exists for the selection of the "best" binding pose. Autodock program well reproduced the experimental binding mode of **1** (Figure 2) both in native (1W51) and non native (2G94) enzyme structure, with a RMSD values of 0.4 and 0.6 Å, respectively. For the other three enzyme structures, good results were obtained considering the cluster conformations (0.59 Å for both 1FKN and 1TQF, and 1.20 Å for 1XN3) while the accuracy in reproducing the X-ray conformation lowered when the ranking conformation was considered only (2.9 Å for 1FKN, 3.4 Å for 1TQF and 2.8 Å for 1XN3).

In order to make the test independent from the single ligand used (**1**), we carried out an additional cross-docking experiment using the inhibitor referred as compound **5** in the paper of Ghosh et al. (complex PDB code: 2G94).[28] The experimental binding conformation of Ghosh ligand was well reproduced in four out of five BACE structures when either the ranking or the cluster conformation is considered. Specifically, in 1TQF and in the native 2G94 BACE structures, AutoDock perfectly predicted the experimental pose (RMSD value of 1.45 and 1.12 Å, respectively) and

provided one unambiguous solution since the ranking conformation belongs to the most populated cluster. In 1FKN and 1W51 structures, the ranking conformation well reproduces the correct pose having an RMDS value of 1.82 and 1.46 Å, respectively. Only in the case of 1XN3, AutoDock did not exactly reproduce the X-ray conformation of the inhibitor (RMSD value of 3.10 Å). After a visual inspection of the docking results, it clearly emerged that the cluster conformation places the ligand in a very similar way to the X-ray conformation apart from the diazole branch that fills the S2 region in the place of the sulfonyl moiety. This exchange might be due to the different conformation of Arg235 with respect to the others X-ray BACE structures. All in all, our test experiments clearly proves that Autodock program can be successfully applied to BACE-1 field, although whenever the ranking conformation does not correspond to the cluster one, both solutions have to be taken in consideration. The final choice between the ranking and the cluster conformations will be governed by their coherency with experimental data, when available (e.g. SARs).

2.3 Docking Results

Docking of **2**. Due to the undetermined absolute stereochemistry of the carbon atom attached to the biphenyl ring of compound **2**[19] both stereoisomers were subjected to docking calculations.

Docking of **2** with the (R) absolute stereochemistry revealed that in four out of five calculations (1TQF, 1W51, 1XN3 and 2G94) comparable results were found for all the predicted ranking conformations. Using as receptor 1FKN, docking of **2** did not succeed in predicting a plausible binding mode, therefore it was omitted from the comparison. As depicted in Figure 4, the (S) statine isoster places the hydroxyl group in between the catalytic dyad allowing the simultaneous interaction with D32 and D228 as previously observed in other X-ray complexes (e.g. 1FKN).

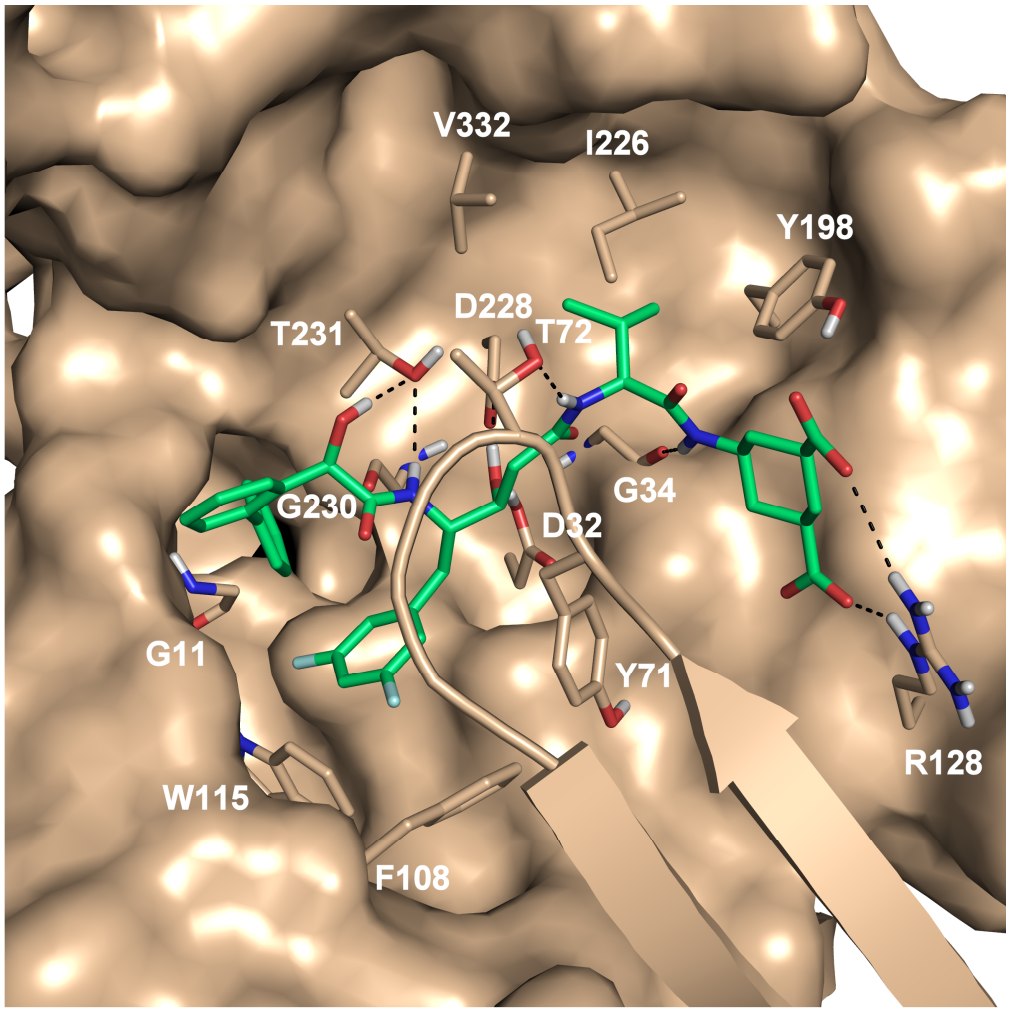


Figure 4. Binding mode of compound **2** (green) into BACE catalytic site represented as Connolly surface. The ligand and the interacting residues are shown in stick representation and coloured by atom type, while the FLAP region is represented as cartoon. Hydrogen bonds are represented with dashed black lines. All nonpolar hydrogens were removed for clarity.

The difluorobenzyl moiety of **2** (P1) occupies the aromatic pocket S1, analogously to the corresponding benzyl group of **1**. Noteworthy, due to the withdrawing property of the fluorine atoms present on the aromatic system of **2**, the charge transfer interactions of the P1 branch in **2** with the Y71, W108 and P115 aromatic rings are expected to be stronger with respect to those observed for compound **1**. Analogously to 1, the calculated binding mode preserves the H-bond with G34 backbone CO, while two additional H-bonds with T72 and T231 side chains are present (Figure 4).

Unfortunately, a direct comparison between the potency of **2** (IC50 = 20 nM) and **1** (IC50 = 200 nM) is not meaningful, this is due to the different biological assays. Differently from **1**, compound **2** features an isopropyl group (P1') which establishes hydrophobic contacts with I226 and V332 side chains (S1' pocket) while the biphenyl moiety deepens into a narrow passage (S3sp) mainly formed by two glycines (G13 and G230). An interesting feature of **2**, which certainly contributes to its great potency (IC50 = 20 nM), is the aminocyclohexanedicarboxylate moiety (P2'), which was inserted to mimic the C-termini of the first known peptidic inhibitors.[15,18] Indeed, docking results confirm that this P2' moiety

entirely fills the S2' hydrophobic pocket with both the carboxylate groups engaging a charged-reinforced H-bond with the guanidine group of R128.

Docking of **2** with biphenyl unit attached in the (S) configuration gave for three out of five docking calculations (1FKN, 1XN3 and 2G94) a slightly different binding conformation in comparison to that found for the (R) isomer. Indeed, the main interactions with the enzyme are well conserved for this isomer while some differences come out for the N-terminal moiety (P3 branch). Here, the biphenyl group points into the S3sp, similarly to the (R) isomer, while the hydroxyl function due to its (S) stereochemistry is now incapable to interact with T231. Although it is unknown which is the most active diastereoisomer, it has been reported that one isomer is 100 fold more active than the other.[19] Our docking results do not clearly discriminate between the two analyzed isomers. Nevertheless, the low convergence of docking results for the (S)-isomer allows us to hypothesize a weaker binding to BACE if compared to the (R)-isomer.

However, the proposed binding modes are in alignment with the available SARs data.[19] Indeed, analogues of **2**, featuring non acidic aminocyclohexanedicarboxylate derivatives, do not interact with R128, this results in a loss of activity.[19]

Furthermore, the replacement of the aminocyclohexanedicarboxylic moiety by 4-aminomethylbenzoic acid, presenting only one acidic function, caused a 10-fold loss of activity thus demonstrating the contribution of both acidic groups for the enzyme binding. Interestingly, the aminocyclohexanedicarboxylic methylester derivative displays only a 10 fold decrease in the inhibitory activity. These data are in accordance with our results, which place the aminocyclohexanedicarboxylic near to R128, where the carbonyl of the methyl ester forms an H-bond with the guanidine side chain.

Docking of **3** and **4**. The binding pose of **3** does not substantially change when different enzyme structures are used and basically resembles the binding position found for **1** and **2**. As depicted in Figure 5a, the hydroxyl group of the HEA core H-bonds with D32, while the protonated secondary amine, which differs from **1** by the locked conformation of the 6-membered ring. It engages a salt bridge with D228 and H-bonds with G34. Interestingly, all HEA derivatives feature an unusual stereochemistry at the secondary alcohol (R absolute configuration). A secondary amine in the HEA derivatives causes the interaction with D228, which would be lost by the inversion of the stereochemistry at the secondary alcohol. The benzyl

ring (P2') is stacked in between the Y198 and Y71 residues (S2' pocket), while the ligand amide group H-bonds with G230 and Q73 backbones. Analogously to **2**, the difluorobenzyl branch (P1) fills the S1 pocket shaped by Y71, F108 and W115 residues (Figure 5a).

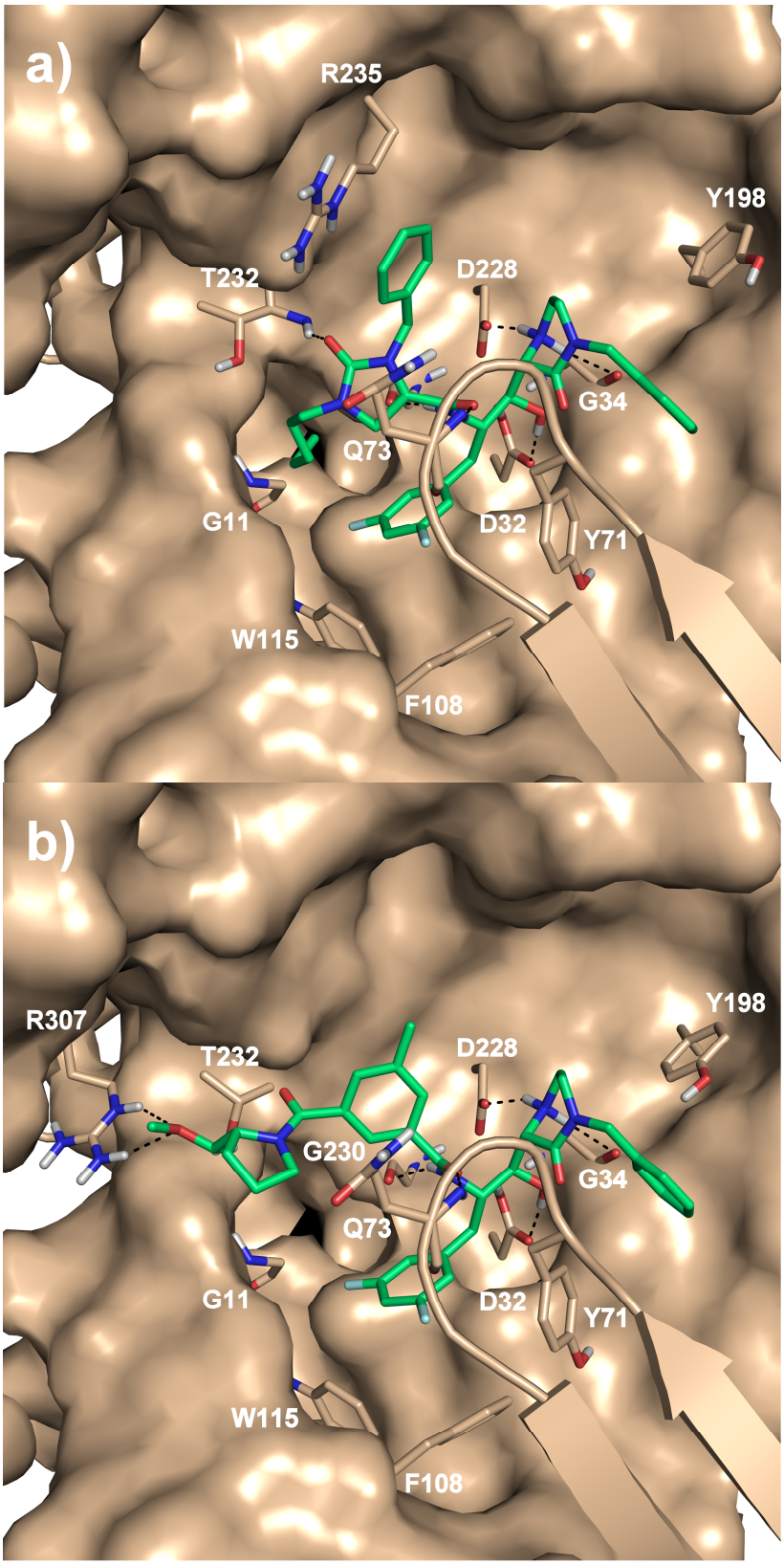


Figure 5. Binding modes of compound **3** (a) and **4** (b) into BACE catalytic site.

While the above-described interactions are conserved in all five BACE structures, an ambiguity was encountered for the relative position of the imidazolidinone moiety. Ligand docking into 1W51 and 1XN3 structures placed the imidazolidinone so as to allow the carbonyl group to H-bond with T232 backbone, the benzyl group (P2) to establish a cation-π interaction with the guanidine group of R235 and the N-alkyl substituent to thread into the narrow S3 subpocket (S3sp) (Figure 5a). However, the ligand docking into 1FKN and 1TQF structures, placed the imidazolidinone ring so that the phenyl ring pointed to the S3sp, while the alkyl chain pointed out of the enzyme. In this case, ensemble docking leads to two comparable but not equal conformations. A subsequent analysis of BACE structures suggests that the different conformations of the R235 side chain are mainly responsible for the divergent results. More precisely, in 1W51 and 1XN3 structures, the R235 side chain is optimally oriented to engage a cation-π interaction with the phenyl ring of **3** (Figure 5a), while in 1FKN and 1TQF structures, the R235 guanidine group partially occludes the catalytic site so as to prevent the placement of the phenyl group. Although both ligand conformations are feasible, the orientation of the N-alkyl substituent into the S3sp and the benzyl moiety towards the external part of the enzyme

maximizes the interaction with the protein by allowing the formation of a H-bond with T232 and a cation-π interaction with R235. Further support for this hypothesis is provided by recent X-ray studies outlining the importance of the interactions with T232 and R235.[31]

Similarly to **3**, the HEA isoster in **4** well interacts with both the catalytic aspartates as well as with G34 (Figure 5b). The benzyl and difluorobenzyl moieties of the ligand optimally fill S2' and the S1 pockets, respectively, while the isophthalamide group lies in the S2 open region with one of the two amide functions H-bonding with G230 and Q73. Interestingly, the methoxymethyl substituent of the pyrrolidine protrudes above the S3sp, where polar interactions occur with R307 and T232 side chains. The high potencies of **3** and **4** (IC50 = 1 nM, IC50 = 1.4 nM, respectively) suggest that T232, R235 and R307 are further points of ligand attachment strengthening the inhibitor binding.

Docking of **5** and **6**. Ensemble docking experiments on **5** and **6** showed for both of them convergence of results. In fact, all the docking calculations apart from one (1W51) detected a single solution which is at the same time the ranking and cluster conformation.

The protonated nitrogen of **5** was found to interact with D228 while the isobutylamide branch (P2') engages H-bonds with T72 and G34 placing the alkyl chain into the hydrophobic S2' pocket (Figure 6a). The n-propyl branch (P1') lies in S1' pocket shaped by hydrophobic aminoacids such as I226, V332 and Y198. The benzyl group (P1) is placed into the aromatic cage S1 with the adjacent amide group forming two H-bonds with G230 and Q73. Comparing the binding modes of the docked ligands, we noticed that H-bonds with G230, T72 or Q73 backbones are frequently present and this seems to be important for high BACE inhibitory activity.[32] Interestingly, in all the five BACE structures used for the ensemble docking, the N-methyl methylsulfonamide group (P2) was found in a polar region among N233, S325 and R235, mostly interacting with the latter residue. The proposed location of the sulfonamide function is in line with the recently reported X-ray structures of BACE complexed with some sulfonamide-containing ligands.[21,27,29,32]

The difluorobenzyl branch deepens inside the narrow channel in the S3sp engaging a T-shape interaction with Tyr14 (Figure 6a). It is interesting to note that this channel constitutes the access to an additional small pocket

lined by hydrophobic residues (L152, L154, V31 and Y14) and up to date no inhibitor has entirely filled this newly identified pocket.

The hydrophobic interactions of the n-propyl branch (P1') in the S1' pocket are supported by SAR data which show a slight decrease of activity for the ethyl and/or methyl (P1') substituent.[22] Furthermore, additional SARs suggest that an H-bond donor on P2' substituent can be important for BACE activity and this is in agreement with our finding of an H-bond interaction between the isobutylamide branch and the G34 backbone.

Due to the structural similarity, compound **6** docked in a mode similar to **5**. Nevertheless, being **6** an HEA derivate, it contacts both D32 and D228, as described for all the other HEA derivates. While the benzyl group (P1) deepens into the S1 pocket, the sulfonamide group engages an electrostatic interaction with R235 (Figure 6b).

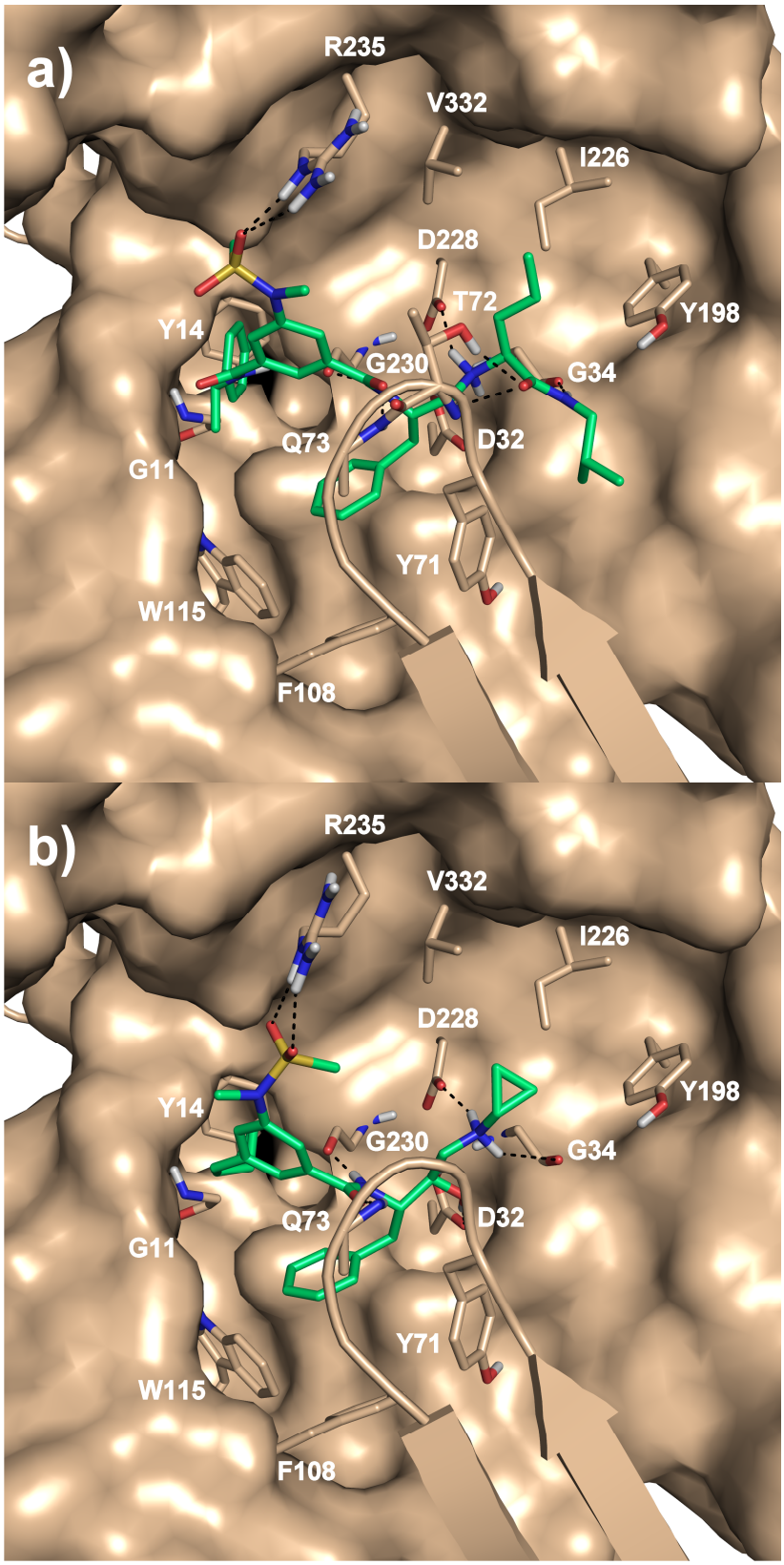


Figure 6. Binding modes of compound **5** (a) and **6** (b) into BACE catalytic site.

The two cyclopropyl branches of **6** fit in the hydrophobic S1' pocket and S3sp, respectively. Despite the structural similarity, compounds **5** and **6** show different activities (IC50 = 4 and 35 nM, respectively). According to our docking results, this difference in potency has to be ascribed to the additional interactions established by **5**, which occupies S3sp, S1' and S2' pocket, while **6** just partially occupies the S3sp and S1'pocket and does not fill the S2' pocket at all.

2.4 Pharmacophore Fingerprints and Guidelines for Drug Design

The superimposition of the calculated bioactive conformations of inhibitors **1**-**6** (Figure 7a) allowed us to capture both the common geometric and electronic features essential for the ligand recognition and the enzyme inhibition. From the analysis of the interactions established between the ligands and the enzyme, it is apparent that both polar and hydrophobic interactions are equally important in the inhibitor-enzyme recognition process.

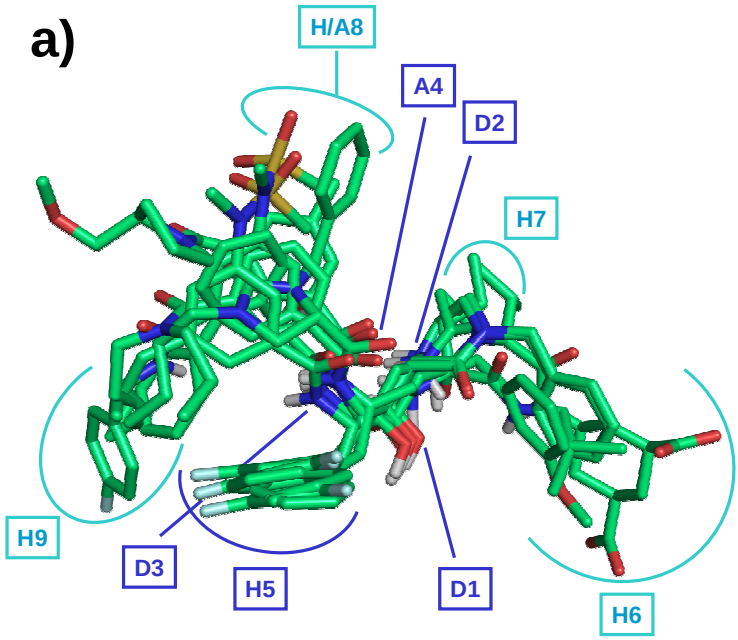
Despite the structural diversity, compounds **1**-**6** are linked by five highly conserved pharmacophoric points (blue spheres in Figure 7b and 7c): three H-bond donors (D1, D2 and D3), one acceptor (A4) and one hydrophobic centre (H5). All compounds, with the exception of **5**, feature an interaction point with D32 (D1) and with the other catalytic aspartate (D228) through the D2 point. This observation confirms the importance of the interaction with the two catalytic aspartates of the binding site for an effective enzyme inhibition. Furthermore, all compounds, apart from **2**, present the D2 point. This highlights the convenient insertion of a protonable amine in this position to achieve a simultaneous interaction with D228 and G34 residues. As shown in Figure 7c, the D3 point donates an H-bond to G230 backbone

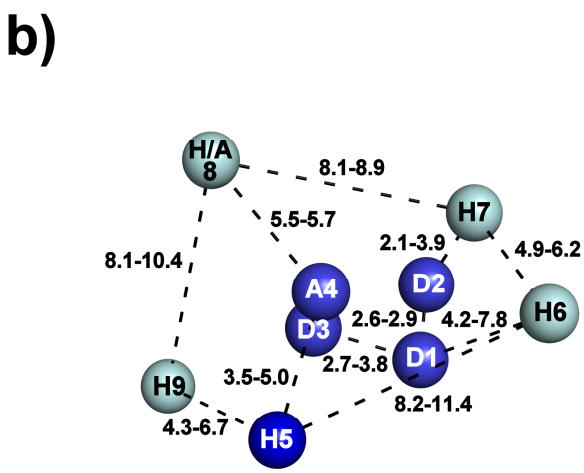
CO while the H-bond acceptor A4 interacts with either T72 or Q73 backbone NHs. It is noteworthy how these two latter points (D3 and A4) represent an ancestral inheritance of the endogenous ligands of BACE where these points are normally filled by an amide moiety of the peptidic backbone.

Despite the relevance of such polar features, the hydrophobic point H5, constantly present in **1**-**6**, underlines the essential role of an hydrophobic group in this position, interacting with the S1 pocket residues Y71, W108 and F115. (Figure 7c).

The frequent occurrence of D1, D2, D3, A4 and H5 pharmacophoric points in the analysed compound set (Table 1) suggests that these are the indispensable features for ligand recognition. Unfortunately these interactions do not offer the key to selective BACE inhibition. This is due to the conservation of the majority of their corresponding interacting residues in other proteases such as cathepsin D as discussed hereafter.

From our docking results, four additional pharmacophore points, represented as cyan spheres in Figure 7b and 7c, emerge. They are mostly hydrophobic (H6, H7 and H9), with the exception of one, that can be either hydrophobic or a polar (H/A8).


a)
H/A8
A4
D2
H7
H9
D3
H5
D1
H6


b)
H/A 8
8.1-8.9
H7
5.5-5.7
2.1-3.9
4.9-6.2
8.1-10.4
A4
D2
D3
2.6-2.9
4.2-7.8
H6
2.7-3.8
D1
H9
3.5-5.0
8.2-11.4
4.3-6.7
H5

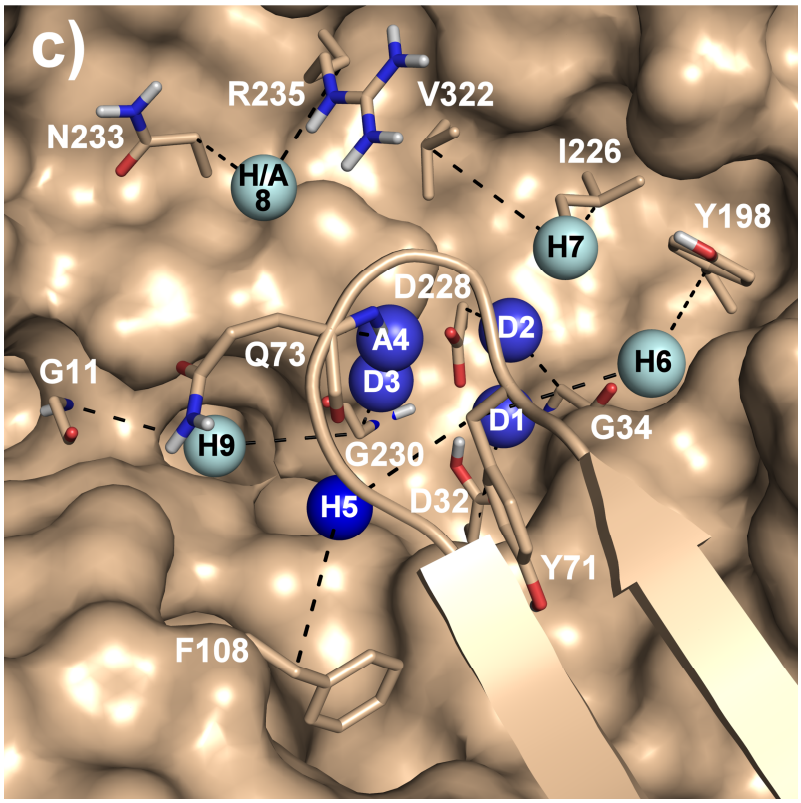


Figure 7. a) Overlay of bioactive conformations of compounds **1-6** on the experimentally determined bound conformation of **1**. The pharmacophoric points are colour coded as conservative (blue) or additional (cyan). The non polar hydrogens are omitted for clarity. The letter "A" corresponds to an H-bond acceptor group, "D" to an H-bond donor, while "H" to an hydrophobic group. "A/H" means that an H-bond acceptor or an hydrophobic group is tolerated. b) Tridimensional representation of the distances between the identified pharmacophoric points. The distances are reported in Å and represent the minimum and the maximum value found in the proposed bioactive conformations of **1-6**. The distances were calculated considering the nitrogen and oxygen atoms of the H-bond acceptor and donor groups, the sulphur atom of the sulphonamide function, the centroids of the aromatic rings, the centre of mass of the alkyl and cycloalkyl groups. c) Mapping of the pharmacophoric points into the BACE catalytic site represented as Connolly surface. Interactions between the pharmacophoric points and some BACE residues are highlighted by black dashed lines.

The H6 point represents an hydrophobic feature able to reach the S2' pocket (Figure 7c). The structural variability of this feature (Figure 7a) demonstrates that, although an aromatic substituent in H6 is not mandatory (see compounds **2** and **5**), an hydrophobic group is required due to the nonpolar character of the S2' pocket. It is worth noting that in all analysed compounds, only **2** features the optimally functionalized H6 point having two carboxyl groups that establish a double salt bridge with R128. Such an interaction is particularly interesting being R128 an unique feature of BACE enzyme. In order to achieve a better pharmacokinetic profile preserving a good inhibitory potency, the acidic functions present in the H6 point could be methylated not preventing the ability to H-bond with R128.

The additional point H7 finds place in the hydrophobic S1' pocket (I226, V332 and Y198) where, due to its limited dimension, only an alkyl or cycloalkyl chain with at maximum three carbon atoms, seems to be tolerated. This position was recently employed to achieve selective BACE inhibition and should be investigated further.[33]

Another pharmacophoric point, which may be exploited to achieve BACE-selectivity, is represented by the H/A8 point, which is located in the S2 open region and can be either hydrophobic or polar. An H-bond acceptor

in this position, such as a sulfonamide (compounds **5** and **6**), or an hydrophobic phenyl ring (compound **3**) can interact with the surrounding residues such as N233, R235 and S325. These three residues are peculiar to BACE in comparison to other proteases, thus this interaction is expected to have an important role in the BACE selectivity. Our observation is in accordance with the recently reported X-ray analysis of the co-crystallized complexes of some selective BACE inhibitors featuring a sulfonamide group as H/A8 point.[27,29]

After the superposition of all compounds in their predicted binding poses, it clearly emerges a different sized branch as H9 point indicating that aliphatic, as well as aromatic branches are well tolerated in the S3sp. The extension of this pocket mainly depends on the conformation of the 10S loop, but it has to be pointed out that among all analysed compounds, only inhibitor **5** goes across the S3sp, reaching with its aromatic system the inner hydrophobic pocket made by L152, L154, Vl31 and Y14 residues. In the design of new BACE inhibitors, this cavity should be further explored, as demonstrated for the renin inhibitor aliskiren.[34] Comparing the BACE and renin cavities, we noticed that they have chemical and structural

differences, which offer additional chances to improve the inhibitor selectivity.

Table 1. Pharmacophoric points present in ligands **1-6**.

Compounds	Points
1	D1-D2-D3-A4-H5-H6-H9
2	D1-H5-H6-H7-H9
3	D1-D2-D3-A4-H5-H6-H/A8-H9
4	D1-D2-D3-A4-H5-H6
5	D2-D3-A4-H5-H6-H7-H/A8-H9
6	D1-D2-D3-A4-H5-H7-H/A8-H9

The compounds **1-6** are characterized by five to eight identified pharmacophoric points (Table 1); appropriate chemical modifications can result in more potent analogues. For instance, the BACE binding affinity of compound **5** may be improved by the addition of an hydroxyl group on the

carbon of the Ψ(CH2NH) reduced amide bond so as to present the D1 point and complete the nine-point pharmacophore.

Regarding compound **1**, the substitution of the n-propyl chain with a benzyl moiety may optimize the interactions with the BACE S3sp and the addition of an H-bond acceptor in position 2 on the isophthalamide group such as a sulfonate or carbonyl group may provide the basis for BACE selectivity.

So far, only one pharmacophore model derived from a congeneric series of BACE inhibitors has been disclosed via a patent application by Vertex.[14] The authors proposed that the flap is shifted and stabilised in an open conformation in the presence of their inhibitors.

Comparing the Vertex pharmacophore model with ours, we found that the two models are rather similar regarding the pharmacophoric points interacting with the residues unaffected by the flap movement (Table 2).

Table2. Distances (Å) between the pharmacophoric points and the BACE interacting residues.

Pharmacophoric Points	Interacting Residues	Distances[a]
D1	D32	4.9
D2	G34 D228	3.7 4.5
D3	G230	4.4
A4	Q73	4.5
H5	Y71 P108	5.4 8.9
H6	Y71 Y198	5.4 8.1
H7	I226 V332	6.0 6.4
H/A8	R235 N233	5.6 5.3
H9	G11 G230	5.1 4.5

[a] The distances were calculated considering the pharmacophoric points and each Cβ of the corresponding residues, Cα were taken into account for glycine residues. As pharmacophoric points the nitrogen and oxygen atoms of the H-bond acceptor and donor groups, the sulphur atom of the sulphonamide function, the centroids of the aromatic rings, the centre of mass of the alkyl and cycloalkyl groups, were considered.

In particular, the Vertex model shares a pattern of three H-bond donors, and two hydrophobic points corresponding to D1, D2 and D3 and H5 and H6 in our model. Despite the general coherency of the chemical features of these points in both pharmacophoric models, the reported distances between them diverge to a large extent. For instance, in our model, the distance between D1 and D2 and D1 and D3 is maximum 2.9 and 3.8 Å respectively, (see Figure 7b) while in the Vertex model both range from 4 to 5 Å. Moreover, the D1-H6 distance is calculated as a range of 4.2-7.8 Å in our pharmacophore model and this value is very low in comparison to the minimum distance of 8 Å reported for the corresponding points in the Vertex model (HB-1 and HPB-3, respectively). This discrepancy may be due to the fact that Vertex model places the HPB-3 point in a different pocket of the S2' region.

The discrepancies found between the two pharmacophoric models can be assigned to the different compounds used for model generation. The Vertex pharmacophore derived from a congeneric series based on a piperazine scaffold which are thought to stabilize the flap in an open conformation. Consequently, in their pharmacophoric model, an additional hydrophobic point referred as HPB-2 is involved in the interactions with the

flap pocket (W76, F108, F109, W115 and I102). Here, we have used BACE with the flap in a closed conformation, thus this pocket is no longer present. Consequently the HPB-2 point has to be considered a typical feature of the open flap pharmacophore model.

With respect to the Vertex model, our model offers an accurate description of three new pharmacophoric points: A4, H7 and H9, which are particularly important for a selective BACE inhibition.

The structural diversity of the compounds used in our study contributes to the value of our pharmacophore model, which is also substantiated by the X-ray structure of the binding conformation of compound **1**, perfectly filling the seven pharmacophoric points.

The design of new BACE inhibitors has to consider the other human aspartic proteases which could be potentially inhibited by BACE ligands, such as renin, napsin-A, and B, cathepsin-E, pepsinogen-A, and C and cathepsin-D. Indeed, the catalytic domain of BACE is similar to that of other aspartyl proteases and the interactions of these enzymes with their inhibitors do not diverge too much from those we observed in the case of BACE. For instance, most of the human aspartyl proteases accept a phenylalanine analogue in P1. The selectivity versus BACE over other

human aspartic proteases is required to avoid adverse side effects and is thus mandatory for clinical development of BACE inhibitors. For instance, inhibition of cathepsin D, which is largely expressed in all cells controlling their protein catabolism,[35] would mean a likely consuming of that BACE inhibitor as well as the occurrence of probable toxicity.

Therefore we performed a structure-based sequence alignment of BACE and cathepsin D to investigate the differences in their binding sites (Table 3).

Table 3. Main dissimilarities in BACE and Cathepsin-D catalytic sites.

Binding Site Location	BACE residues	Cathepsin D residues
S2 Open Region	N233 R235 S325 S327	L236 V238 M307 M309
S1' pocket	K224	E227
Loop between S1' and S2'	S328, T329 and G330	D310, I311, P312, P313, P314, S315, G316, P317 and L318
S2' pocket	R128	V144
Flap Region	T72	G79
Others	R307 K321	L292 L303

The superposition of BACE and cathepsin D three-dimensional structures reveals that the two enzymes display very similar residues in their binding sites and consequently possess a similar shape, which is visualised by their Connolly surfaces in Figure 8. In addition to the catalytic dyad, several residues important for the ligands recognition such as G34 or G230 (G35 and G233 in cathepsin D) are conserved. As shown in Figure 8, the shape of the hydrophobic S1 and S2' pockets is equivalent in BACE and cathepsin D. However, a careful comparison of the two binding sites reveals several important points of diversification (Figure 8).

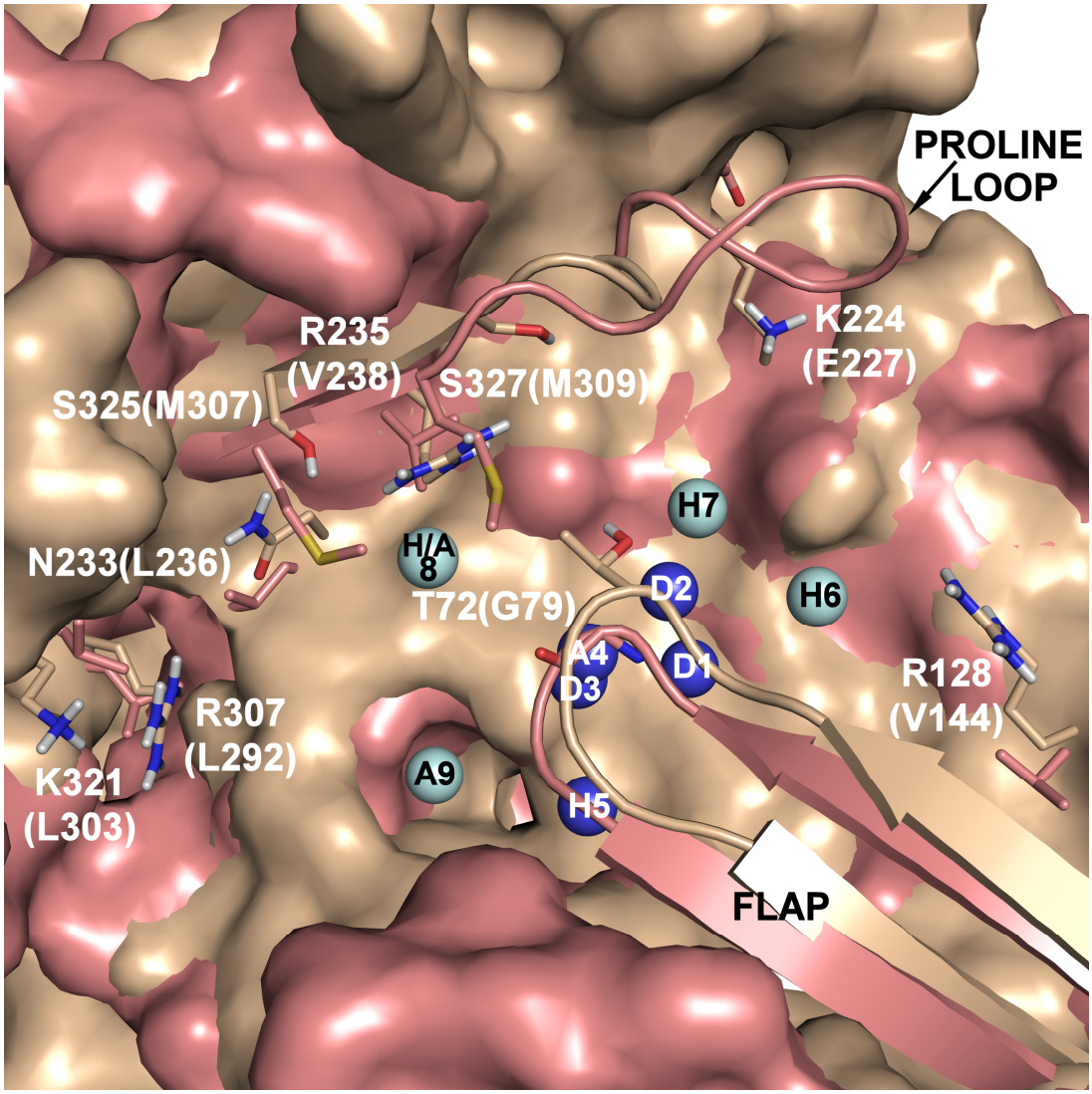


Figure 8. Superposition of BACE (pink) and cathepsin D (brown) structures represented as Connolly surfaces. The pharmacophoric points were mapped into the BACE enzyme. Mutated residues in the binding pocket are shown in stick representation and labelled with the one-letter amino acid code. The letters and the number in parentheses refer to the cathepsin D enzyme.

Probably the most striking difference is located near to the catalytic dyad and thus easily accessible to the inhibitors; it concerns the S2 pocket, which presents the polar triplet N233-R235-S325 in BACE and the hydrophobic triplet L236-V238-M307 in cathepsin D. In line with this observation, a series of highly selective inhibitors containing a sulfonyl group in the P2 branch interacts with the residues of the BACE S2 region.[27,29] Besides the different character of the S2 region in BACE and cathepsin D, the space available for the ligand binding in cathepsin D is limited by the M307 and M309 side chains; these are replaced by less space spacious residues in BACE (S325 and S327). This finding suggests the incorporation of bulky P2 branches, functionalized with polar groups capable to interact with the BACE triplet (N233-R235-S325).

The superposition of BACE and cathepsin D reveals significant differences in the length and sequence of the loop defining the S1'/S3' pocket. Indeed, this loop is shorter in BACE (S327, S328, T329 and G330) than in cathepsin D, which presents a long loop of ten residues containing a rigid section called proline loop (P312, P313, P314, S315, G316 and P317).

This relevant difference indicates an alternative way to achieve BACE/cathepsin D selectivity. Indeed, a properly oriented bulky moiety on

the P1' substituent will occupy the BACE S1' pocket and will not be tolerated by cathepsin D. Moreover, the mutation of BACE-K224 in E227 in cathepsin D, suggests the insertion of an hydrogen bond acceptor or a positively charged group on the bulky moiety. To our knowledge, only few peptidic inhibitors targeted an interaction with K224.[15,18]

Another dissimilarity between BACE and cathepsin D resides in the S2' pocket, where R128 is replaced by V144 in cathepsin D. Therefore, inhibitors including one or more acidic functions on the P2' branch are expected to favour the interaction with BACE. The cathepsin D flap region, where G79 replaces the BACE-T72, offers another opportunity for enhancing the selectivity. The last divergence of BACE/cathepsin D resides in the region above the S3 pocket. Here, as shown in Figure 8, two basic amino acids (R307 and K321) are replaced by two hydrophobic residues (L292 and L303) in cathepsin D. Thus, compounds presenting an interaction with R307 and/or K321 (e.g. **4**) may contribute to the selective BACE inhibition.

Mapping our pharmacophore model into the BACE/cathepsin D superposed structures, two main issues can be inferred. Firstly, it is apparent how the conserved pharmacophoric points (blue spheres in Figure 8) are

essential for the ligand recognition in both the enzymes. These pharmacophoric points are placed in a region where all residues are conserved, therefore these points cannot confer selectivity.

On the contrary, some additional points (cyan spheres in Figure 8) are located in regions, which are dissimilar in BACE and cathepsin D. In particular, the H/A8 has to be considered critical for the improvement of ligand potency and selectivity by allowing the interaction with the basic triplet (N233, R235 and S325) present in BACE and not in the homologous cathepsin D. Similarly, the H6 and H7 points both offer the opportunity to obtain compounds featuring an acidic groups or an H-bond acceptor on the P1' and P2' branches so as to allow an interaction with R128 and K224 residues, which are only expressed in BACE.

In conclusion, despite the high sequence homology between BACE and cathepsin D, we have identified a distinctive fingerprint of the BACE catalytic site that is worth targeting in the effort to achieve potent and selective inhibitors.

CONCLUSION

In the present paper an ensemble-docking approach was undertaken on six highly potent BACE inhibitors identifying for all of them plausible binding modes. A common pharmacophore model linking the multiple structural classes of inhibitors was derived. This allowed us to capture both the geometric and electronic features essential for the ligand recognition and the enzyme inhibition. In particular, we identified a nine points pharmacophore model outlining the relative distances among them. Interestingly, five of these points are present in all the inspected ligands; they can be referred to as essential features for the ligand recognition. Whereas the other four points have been defined as accessory points of interaction. An accurate structural comparison of BACE and cathepsin D was made to support the rational design of BACE-selective inhibitors. Despite the high degree of similarity, many structural differences were identified and highlighted; these can be used to achieve or enhance a selective BACE inhibition.

Both, the elucidation of the binding modes of the diverse ligands, and the development of an exhaustive structure-based pharmacophore model are expected to provide a support for pharmacophore- and structure-based VS techniques and a source for the optimization of screen derived hits as well as of established leads. Moreover, the pharmacophore hypotesis can be of help

in the common target-based and ligand-base drug design approaches as well as in the setting of a focused-library of BACE inhibitors.

COMPUTATIONAL METHODS

Molecular modeling calculations and graphics manipulations were performed on a Silicon Graphics Octane2 workstation equipped with two 2600 MHz R14000 processors using the SYBYL7.2 software package.[36] Automated docking calculations were performed using version 3.0.5 of the AutoDock software package.[37]

The Docking Program.

AutoDock program combines a rapid energy evaluation through pre-calculated grid maps of affinity potentials, one for each atom type present in the ligand being docked, with a variety of search algorithms to find suitable binding positions for a ligand on a given protein. While the protein is required to be rigid, the program allows torsional flexibility in the ligand. The calculation of these maps helps to make the docking process extremely fast. They are calculated by the AutoGrid procedure where the protein is embedded in a 3-D grid with a probe atom placed at each grid point. The interaction energy between the probe atom and the protein is computed for each grid point *via* a smoothed Lennard-Jones pairwise potential, with the effect of widening the region of maximum affinity at ε, and also reducing the potential energy at r = 0 to a finite value (Figure 9):

$$V(r) = c_{m,n}\, \varepsilon\, [(\sigma/r)^m - (\sigma/r)^n]$$

m and n are positive integer number and m > n (usually 12 and 6, respectively), ε is the depth of the potential well and σ is the finite distance at which the interparticle potential is zero. $(1/r)^m$ and $(1/r)^n$ describe repulsion and attraction terms, respectively. $c_{m,n}$ is chosen to get the minimum value of V(r) being $V_{min} = -\varepsilon$.

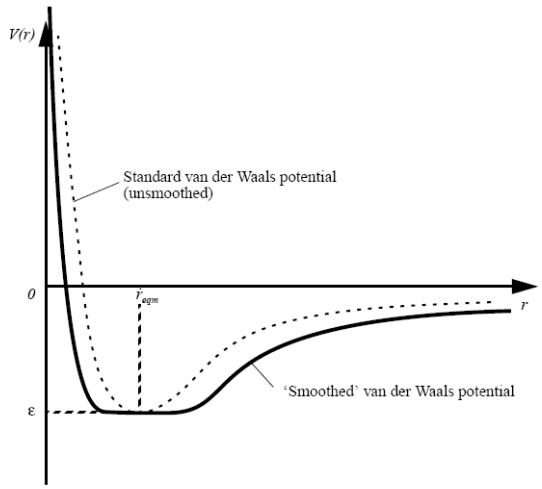


Figure 9. Comparison between the smoothed van der Waals potential used by AutoDock program and the classic unsmoothed van der Waals potential.

As a result, an affinity grid is obtained for each atom type of the ligand as well as a grid of the electrostatic potential is calculated through Coulombic interactions between the macromolecule and a point charge of $+1.60219 \times 10^{-19}$ C used as a probe (Figure 10), by means of the following equation:

$$F = 1/4\pi\varepsilon_{(r)}\ q_1 q_2 / r^2$$

where $\varepsilon_{(r)}$ is a sigmoidal distance-dependent dielectric function used to model solvent screening, q_1 and q_2 are the charge values of the probe and the corresponding atom of the macromolecule and r is the distance between q_1 and q_2.

The energetic profile of a particular substrate configuration is eventually found by counting both the affinity values and the electrostatic interactions of the grid points surrounding each atom of the given ligand.

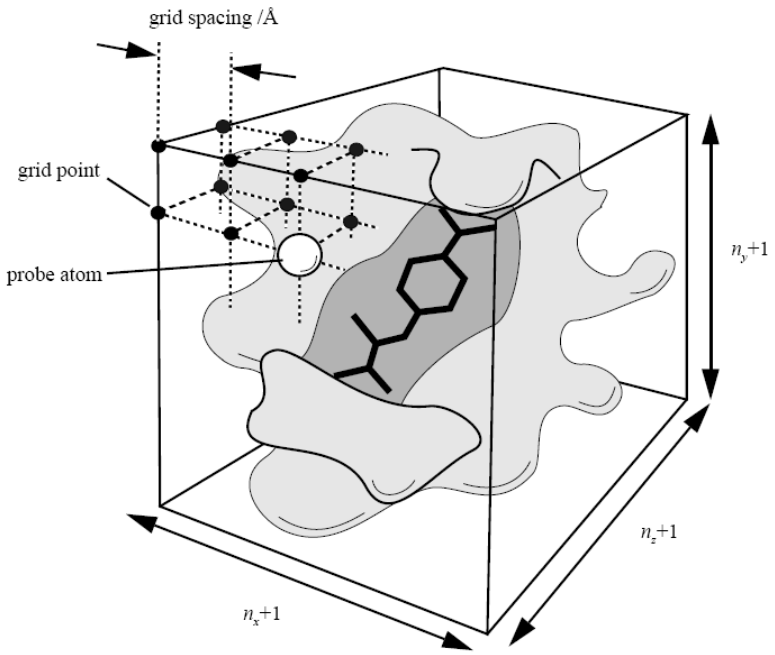


Figure 10. Graphical exemplification of the grid points used by AutoGrid to calculate protein maps.

The scoring function of AutoDock was parameterized using a series of protein-inhibitor complexes for which both the structures and the inhibition constant, K_i, are known. Looking at the theory, AutoDock scoring function is based on the Hess's law of heat summation. This law states that since the free energy is a state function, the change in the free energy between two states will be the same regardless of what pathway is taken to achieve the

products. So the calculation of the free energy of binding in solvent can be expressed by the following equation:

$$\Delta G_{binding\ solution} = \Delta G_{binding\ vacuo} + \Delta G_{solvation(EI)} - \Delta G_{solvation(E+I)}$$

since $\Delta G_{binding\ vacuo}$ can be calculated by the docking simulation, and we can also estimate the solvation free energy for the complex (EI) and for the separate inhibitor (I) and protein (E), we eventually are able to calculate the free energy change upon binding of the ligand to the protein in solution, $\Delta G_{binding\ solution}$, and the corresponding inhibition constant, K_i.

Ligand setup.

The protonation state of ligands **1**-**6** was calculated using MarvinSketch tools (available at http://www.chemaxon.com/marvin/doc/dev/example-sketch1.1.html) at the pH value of the corresponding biological essay. The absolute stereochemistry of each ligand was considered as reported in literature. Due to the undetermined stereochemistry on the N-terminal hydroxyl group of compound **2**, both possible isomers were taken into account for the docking calculations. For **2**, **3** and **4** the Cambridge

Structural Database (CSD)[38] was searched for the conformational preference of the cyclohexyl and the piperazinone moieties. Energy minimizations of the obtained structures were achieved with the TRIPOS force field using the SYBYL/MAXIMIN2 minimizer by applying the BFGS (Broyden, Fletcher, Goldfarb and Shannon) algorithm[39] with a convergence criterion of 0.001 kcal/mol. Partial atomic charges were assigned by using the Gasteiger-Marsili formalism.[40] All the relevant torsion angles were treated as rotatable during the docking process, allowing thus a search of the conformational space.

Protein setup.

All the X-ray structures of BACE (PDB entry codes = 1FKN, 1W51, 1TQF, 1XN3 and 2G94)[13,26-29] were set up for docking as follows: polar hydrogens were added using the BIOPOLYMERS module of the SYBYL program (the side chain of Asp32 was taken as protonated[24,41] while all other Asp, Glu, Lys and Arg side chains were considered ionized and all His were considered neutral by default), Kollman united-atom partial charges were assigned and all waters were removed. In order to optimize the side chains and the hydrogen positions, the protein structures were minimized

using both steepest descent and conjugate gradient, keeping the backbone atoms constrained, employing the DISCOVER program with the CVFF force field.[42] ADDSOL utility of AutoDock program was used to add salvation parameters to the protein structures and the grid maps representing the proteins in the docking process were calculated using AutoGrid. The grids, one for each atom type in the ligand, plus one for electrostatic interactions, were chosen to be large enough to include not only the catalytic site, but also a significant part of the protein around it. As a consequence, for all docking calculations, the dimensions of grids map was 60 x 60 x 60 • with a grid -point spacing of 0.375 •. The centre of the grid was set to be coincident with one of the two oxygens of Asp228.

Docking simulation.

For each ligand, 100 separate docking calculations were performed. Each docking calculation consisted of 1×10^6 energy evaluations using the Lamarckian genetic algorithm local search (GALS) method. A low-frequency local search according to the method of Solis and Wets is applied to docking trials to ensure that the final solution represents a local minimum. Each docking run was performed with a population size of 150,

and 300 rounds of Solis and Wets local search were applied with a probability of 0.06. A mutation rate of 0.02, a crossover rate of 0.8 and an elitism value of 1 were used to generate new docking trials for subsequent generations. The GALS method evaluates a population of possible docking solutions and propagates the most successful individuals from each generation into the next one. The docking results from each of the 100 calculations were clustered on the basis of root-mean square deviation (rmsd = 2 Å) between the Cartesian coordinates of the ligand atoms and were ranked on the basis of the free energy of binding.

The result with the lowest free energy of binding was taken as the representative of each cluster.

Energy refinement of the BACE-1/ligand complexes.

Energy optimizations of the obtained complexes were carried out using 3000 steps of steepest descent followed by 2000 steps of conjugated gradient algorithm with employing the CVFF force field as implemented in the DISCOVER program.[42] A convergence criterion on the gradient of 0.001 kcal mol^{-1} $Å^{-1}$ was set. Only the ligand and the side chains of all residues within a radius of 8 Å around the ligand were allowed to relax.

Connolly surfaces of BACE and cathepsin D were calculated using PyMOL software.[43]

REFERENCES

[1] G. G. Glenner, C. W. Wong, Biochem. Biophys. Res. Commun. 1984, 120, 885-890.

[2] R. C. Petersen, R. G. Thomas, M. Grundman, D. Bennett, R. Doody, S. Ferris, D. Galasko, S. Jin, J. Kaye, A. Levey, E. Pfeiffer, M. Sano, C. H. van Dyck, L. Thal, J. New England J. Med. 2005, 352, 2379-2388.

[3] a) D. K. Lahiri, M. R. Farlow, K. Sambamurti, N. H. Greig, E. Giacobini, L. S. A. Schneider, Curr. Drug Targets 2003, 4, 97-112; b) M. Citron, Neurobiol. Aging 2002, 23, 1017-1022.

[4] R. Vassar, Adv. Drug Delivery Rev. 2002, 54, 1589-1602.

[5] a) S. Roggo, S. Curr. Top. Med. Chem. 2002, 2, 359-370.
b) M. S. Wolfe, J. Med. Chem. 2001, 44, 2039-60.

[6] a) R. Vassar, B. D. Bennett, S. Babu-Khan, S. Kahn, E. A. Mendiaz, P. Denis, D. B. Teplow, S. Ross, P. Amarante, R. Loeloff, Y. Luo, S. Fisher, J. Fuller, S. Edenson, J. Lile, M. A. Jarosinski, A. L. Biere, E. Curran, T. Burgess, J. C. Louis, F. Collins, J. Treanor, G. Rogers, M. Citron, Science 1999, 286, 735-741; b) X. Lin, G. Koelsch, S. Wu, D. Downs, A. Dashti, J. Tang, Proc. Natl. Acad.

Sci. U.S.A. 2000, 97, 1456-1460; c) S. Sinha, I. Lieberburg, Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 11049-11053.

[7] a) D. J. Selkoe, Nature 1999, 399 (67388 Suppl), A23-A31; b) D. J. Selkoe, Physiol. Rev. 2001, 81, 741-766.

[8] a) H. Josien, Curr. Opin. Drug Discov. Devel. 2002, 5, 513-25; b) V. John, J. P. Beck, M. J. Bienkowski, S. Sinha, R. L. Heinrikson, J. Med. Chem. 2003, 46, 4625-30.

[9] D. H. Small, C. A. McLean, J. Neurochem. 1999, 73, 443-449.

[10] Y. Luo, B. Bolon, S. Kahn, B. D. Bennett, S. Babu-Khan, P. Denis, W. Fan, H. Kha, J. Zhang, Y. Gong, L. Martin, J. Louis, Q. Yan, W. G. Richards, M. Citron, R. Vassar, Nat. Neurosci. 2001, 4, 231-232.

[11] L. A. Thompson, J. J. Brown, F. C. Zusi, Curr. Pharm. Des. 2005, 11, 3383-3404.

[12] M. Willem, A. N. Garratt, B. Novak, M. Citron, S. Kaufmann, A. Rittger, B. DeStrooper, P. Saftig, C. Birchmeier, C. Haass, Science 2006, 314, 664-666.

[13] a) M. Maillard, C. Hom, A. Gailunas, B. Jagodzinska, L. Y. Fang, J. Varghese, J. N. Freskos, S. R. Pulley, J. P. Beck, R. E. Tenbrink, Patent WO 02/02512, 200;. b) L. Vuillard, S. Patel, J. Yon, A.

Cleasby, B. Hamilton, A. Shah, Patent WO 04/011641, 2004; c) S. Patel, L. Vuillard, A. Cleasby, C. W. Murray, J. Yon, J .Mol. Biol. 2004, 343, 407-416.

[14] a) G. R. Bhisetti, J. O. Saunders, M. A. Murcko, C. A. Lepre, S. D. Britt, J. H. Come, D. D. Deninger, T. Wang, Patent WO 02/088101, 2002; b) D. C. Cole, E. S. Manas, J. R. Stock, J. S. Condon, L. D. Jennings, A. Aulabaugh, R. Chopra, R. Cowling, J. W. Ellingboe, K. Y. Fan, B. L. Harrison, Y. Hu, S. Jacobsen, G. Jin, L. Lin, F. E. Lovering, M. S. Malamas, M. L. Stahl, J. Strand, M. N. Sukhdeo, K. Svenson, M. J. Turner, E. Wagner, J. Wu, P. Zhou, J. Bard, J. Med. Chem. 2006, 49, 6158-6161.

[15] R. Turner, G. Koelsch, L. Hong, P. Castenheira, A. K. Ghosh, J. Tang, Biochemistry 2001, 40, 10001-10006.

[16] a) B. Schmidt, H. A. Braun, R. Narlawar, Curr. Med. Chem. 2005, 12, 1677-1695; b) B. Schmidt, S. Baumann, H. A. Braun, G. Larbig, Curr. Top. in Med. Chem. 2006, 6, 377-392.

[17] V. John, J. P. Beck, M. J. Bienkowski, S. Sinha, R. L. Heinrikson, J. Med. Chem. 2003, 46, 4625-4630.

[18] a) A. K. Ghosh, L. Hong, J. Tang, Curr. Med. Chem. 2002, 9, 1135-1144; b) A. K. Ghosh, D. Shin, D. Downs, G. Koelsch, X. Lin, J. Ermolieff, J. Tang, J. Am. Chem. Soc. 2000, 122, 3522-3523.

[19] R. K. Hom, L. Y. Fang, S. Mamo, J. S. Tung, A. C. Guinn, D. E. Walker, D. L. Davis, A. F. Gailunas, E. D. Thorsett, S. Sinha, J. E. Knops, N. E. Jewett, J. P. Anderson, V. John, J. Med. Chem. 2003, 46, 1799-1802.

[20] J. N. Cumming, Y. Huang, G. Li, U. Iserloh, A. Stamford, C. Strickland, J. H. Voigt, Y. Wu, J. Pan, T. Guo, D. W. Hobbs, T. X. H. Le, J. F. Lowrie, Patent WO 05/014540, 2005.

[21] S. J. Stachel, C. A. Coburn, T. G. Steele, M. Crouthamel, B. L. Pietrak, M. Lai, M. K. Holloway, S. K. Munshi, S. L. Graham, J. P. Vacca, Bioorg. Med. Chem. Lett. 2006, 16, 641-644.

[22] C. A. Coburn, S. J. Stachel, K. G. Jones, T. G. Steele, D. M. Rush, J. DiMuzio, B. L. Pietrak, M. Lai, Q. Huang, J. Lineberger, L. Jin, S. Munshi, M. K. Holloway, A. Espeseth, A. Simon, D. Hazuda, S. L. Graham, J. P. Vacca Bioorg. Med. Chem. Lett. 2006, 16, 3635-3638.

[23] V. John, Curr. Top. Med. Chem. 2006, 6, 569-578.

[24] a) T. Polgàr, G. M. Keserü, J. Med. Chem. 2005, 48, 3749-3755; b) T. Polgàr, G. M. Keserü, J. Chem. Inf. Model. 2006, 46, 1795-1805.

[25] G. B. McGaughey, D. Colussi, S. L. Graham, M. T. Lai, S. K. Munshi, P. G. Nantermet, B. Pietrak, H. A. Rajapakse, H. G. Selnick, S. R. Stauffer, M. K. Holloway, Bioorg. Med. Chem. Lett. 2006 Nov 6; [Epub ahead of print].

[26] L. Hong, G. Koelsch, X. Lin, S. Wu, S. Terzyan, A. K. Ghosh, X. C. Zhang, J. Tang, Science 2000, 290, 150-153.

[27] C. A. Coburn, S. J. Stachel, Y. M. Li, D. M. Rush, T. G. Steele, E. Chen-Dodson, M. K. Holloway, M. Xu, Q. Huang, M. T. Lai, J. DiMuzio, M. C. Crouthamel, X. P. Shi, V. Sardana, Z. Chen, S. Munshi, L. Kuo, G. M. Makara, D. A. Annis, P. K. Tadikonda, H. M. Nash, J. P. Vacca, T. Wang, J. Med. Chem. 2004, 47, 6117-6119.

[28] R. T. Turner III, L. Hong, G. Koelsch, A. K. Ghosh, J. Tang, Biochemistry 2005, 44, 105-112.

[29] A. K. Ghosh, N. Kumaragurubaran, L. Hong, H. Lei, K. A. Hussain, C. Liu, T. Devasamudram, V. Weerasena, R. Turner, G. Koelsch, G. Bilcer, J. Tang, J. Am. Chem. Soc. 2006, 128, 5310-5311.

[30] S. F. Sousa, P. A. Fernandes, M. J. Ramos, Proteins 2006, 65,15-26.

[31] a) J. N. Freskos, Y. M. Fobian, T. E. Benson, M. J. Bienkowski, D. L. Brown, T. L. Emmons, R. Heintz, A. Laborde, J. J. McDonald, B. V. Mischke, J. M. Molyneaux, J. B. Moon, P. B. Mullins, P. D. Bryan, D. J. Paddock, A. G. Tomasselli, G. Winterrowd, Bioorg. Med. Chem. Lett. 2006 Oct 4; [Epub ahead of print]; b) J. N. Freskos, Y. M. Fobian, T. E. Benson, J. B. Moon, M. J. Bienkowski, D. L. Brown, T. L. Emmons, R. Heintz, A. Laborde, J. J. McDonald, B. V. Mischke, J. M. Molyneaux, P. B. Mullins, P. D. Bryan, D. J. Paddock, A. G. Tomasselli, G. Winterrowd, Bioorg Med Chem Lett. 2006 Oct 4; [Epub ahead of print].

[32] H. A. Rajapakse, P. G. Nantermet, H. G. Selnick, S. Munshi, G. B. McGaughey, S. R. Lindsley, M. B. Young, M. T. Lai, A. S. Espeseth, X. P. Shi, D. Colussi, B. Pietrak, M. C. Crouthamel, K. Tugusheva, Q. Huang, M. Xu, A. J. Simon, L. Kuo, D. J. Hazuda, S. Graham, J. P. Vacca, J. Med. Chem.; 2006, Nov 10; [Ahead of print].

[33] S. F. Brady, S. Singh, M. C. Crouthamel, M. K. Holloway, C. A. Coburn, V. M. Garsky, M. Bogusky, M. W. Pennington, J. P. Vacca, D. Hazuda, M. T. Lai, Bioorg Med Chem Lett. 2004, 14, 601-604.

[34] J. Rahuel, V. Rasetti, J. Maibaum, H. Rueger, R. Goschke, N. C. Cohen, S. Stutz, F. Cumin, W. Fuhrer, J. M. Wood, M. G. Grutter, Chem. Biol. 2000, 7, 493-504.

[35] S. Diment, M. S. Leech, P. D. Stahl, J. Biol. Chem. 1988, 263, 6901-6907.

[36] SYBYL Molecular Modelling System, version 7.2; Tripos Inc.: St. Louis, MO, 2003.

[37] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, J. Comput. Chem. 1998, 19, 1639-1662.

[38] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, Acta Crystallogr. 1979, B35, 2331-2339.

[39] J. Head, M. C. Zerner, Chem. Phys. Lett. 1985, 122, 264-274.

[40] J. Gasteiger, M. Marsili, Tetrahedron 1980, 36, 3219-3228.

[41] a) N. Moitessier, E. Therrien, S. A. Hanessian, J. Med. Chem. 2006, 49, 5885-5894; b) H. Park, S. Lee, J. Am. Chem. Soc. 2003, 125, 16416-16422.

[42] a) DISCOVER, Version 95.0; BIOSYM Technologies, 10065 Barnes Canyon Rd, San Diego, CA 92121; b) A. F. Hagler, S. Lifson, P. Dauber, J. Am. Chem. Soc. 1979, 101, 5122-5130.

[43] DeLano Scientific LLC, The PyMOL Molecular Graphics System, 2002, http:// www.pymol.org.