# A Method for Induced-Fit Docking, Scoring, and Ranking of Flexible Ligands. Application to Peptidic and Pseudopeptidic β-secretase (BACE 1) Inhibitors

Nicolas Moitessier,*,† Eric Therrien,‡ and Stephen Hanessian‡

*Department of Chemistry, McGill University, 801 Sherbrooke Street W, Montréal, Québec H3A 2K6, Canada, and Department of Chemistry, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada*

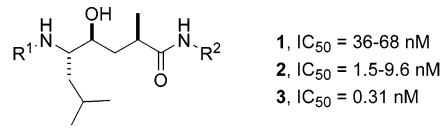
*Received February 12, 2005*

Inhibition of β-secretase (BACE 1) has recently been investigated as a promising therapeutic approach in the treatment of Alzheimer's disease, and a growing number of BACE 1 inhibitors and crystal structures of BACE 1/inhibitors complexes have been reported. We report herein a predictive computational method and its application to potential BACE 1 inhibitors. Using a training set of 50 known highly flexible inhibitors, we developed a docking method that accounts for the flexibility of both the protein and the inhibitors. Protein flexibility is accounted for using a specifically designed genetic algorithm. We next developed a scoring function consisting of force field evaluation of the inhibitor/protein interactions and two additional terms for hydrogen bonding and entropy change upon binding. Discarding three outliers from the training set, our protocol was found to perform well with an rmsd of 1.19 kcal/mol. Evaluation of the predictive power was next carried out by virtual screening of 80 synthetic compounds. The significant enrichment at the top of the ranking list in active compounds demonstrated the ability of the docking and scoring protocol to rank the compounds relative to their activities.

## Introduction

BACE 1 (β-secretase, memapsin-2) has recently been identified as one of the main enzymes involved in the cascade of physiological events that lead to Alzheimer's disease (AD).[1−4] This aspartic protease cleaves a β-amyloid precursor protein (APP) into poorly soluble β-amyloid, which further aggregates and deposits in the brain. The prevention of this proteolysis is therefore a promising therapeutic approach to control the onset and progression of AD. In a series of seminal papers, Tang and co-workers[5,6] reported on the synthesis of a series of peptidic inhibitors such as compounds **1**−**3**, containing a hydroxyethylene isostere (Figure 1). These compounds exhibited nanomolar and subnanomolar activities. The specific interactions of inhibitor **2** with BACE 1, shown in Figure 2, were deduced from the structure of a complex resolved within 1.9 Å.[5] This structural information has next been used in the design and synthesis of a series of potent inhibitors by Tang and Ghosh.[6] The preparation of a number of compounds covering a restricted range of structural classes and efforts directed at the identification of inhibitors of BACE 1 are now gaining momentum.[7−13]

BACE 1 is a structurally challenging protein target with multiple sites for effective binding. Furthermore, the high homology with other aspartic proteases including cathepsin D, pepsin, and renin requires the search and development of selective inhibitors.[9] Penetration of the blood−brain barrier is another consideration of importance for a therapeutically useful inhibitor. In this regard, many reported compounds with inhibitory activity in vitro and in cellular assays feature a highly polar surface, which makes them unsuitable for further development as drug candidates. A potential solution to this problem is to prepare large and diverse libraries of compounds by further exploring the available X-ray crystallographic data. Alternatively, computer-aided drug design can be a valuable tool for predicting potentially active compounds and offering a ranking based on a given set of parameters.




**1**, OM99-1, $R^1$ = H-Val-Asn-, $R^2$ = -Ala-Glu-Phe-OH
**2**, OM99-2, $R^1$ = H-Glu-Val-Asn-, $R^2$ = -Ala-Glu-Phe-OH
**3**, OM00-3, $R^1$ = H-Glu-Val-Asp-, $R^2$ = -Val-Glu-Phe-OH

**Figure 1.** Selected inhibitors.

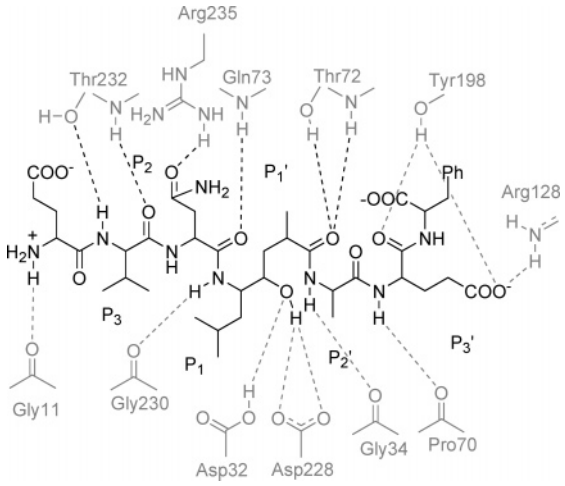



**Figure 2.** Bioactive conformation of OM99-2/BACE 1 complex (code PDB: 1FKN).

The linear interaction energy (LIE) method[14,15] was applied by Reynolds and co-workers[16,17] to a set of 12 compounds as BACE 1 inhibitors. This study revealed the difficulty of predicting the binding affinity of inhibitors to BACE 1 (root-mean-square deviation, rmsd, for predicted free energy of binding of 1.1 kcal/mol). The main difficulty stems from the protonation state of both the inhibitors and the protein (five

* To whom correspondence should be addressed. Phone: 514-398-8543. Fax: 514-398-3797. E-mail: nicolas.moitessier@mcgill.ca.
† McGill University.
‡ Université de Montréal.

10.1021/jm050138y CCC: $33.50 
Published on Web 10/14/2005

arginines and lysines and two aspartic acids in the binding site). As an example, the protonation state of the two aspartic acids in the catalytic site is not strictly defined. Recent molecular dynamics simulation studies favored the protonation of Asp32.[18]

To achieve potent enzyme inhibition, the inhibitor must benefit from optimal molecular recognition with the macromolecular biological target structure, a principle that has been the basis for the development of computational drug design methods.[19−22] Two aspects contribute to the success of structure-based drug design: the generation of reasonable binding modes and the highest scoring of those that correspond to the experimentally observed data. Predicting the correct binding mode of an inhibitor in an enzyme active site invokes the prior positioning of the ligand by a search engine that ensures an efficient and unbiased sampling (conformation/orientation/translation). This docking problem has been tackled using a panel of sampling methods that have been recently reviewed.[23] The binding mode of interest must next be identified.[19,24,25] A variety of scoring functions have been developed including regression-based empirical functions pioneered by Böhm (LUDI)[26] or knowledge-based approaches (e.g., DrugScore[27,28]). To date, the existing scoring functions rank (to some extent) compounds according to their biological activities, but their predictiveness still relies heavily on the target under study.[24,29]

Although the induced-fit and the conformational ensemble rationales are now well documented,[30] the *lock and key* concept describing the ligand−receptor binding has been the model largely exploited by the docking methods. However, this model does not account for conformational changes upon binding. When activity is correlated to conformational changes, usual approaches often generate misleading results.[31,32] In addition, although inhibitors are accurately docked back to their corresponding protein structure (self-docking), docking to other structures (cross-docking) is usually performed poorly.[33,34] To account for these side chain or backbone adjustments, strategies have been explored[35] that include single docking to conformational ensembles modeled as a single set of grids,[34,36,37] docking to a series of composite structures developed from predefined libraries of side chain rotamers and a single backbone conformation,[38−41] or sampling and clustering of the side chains conformations.[42] Construction of the protein structure on the fly using discrete receptor conformations[43,44] and protein adjustment upon binding using a set of template points to improve complementarity[45] have also been proposed. The more time-expensive molecular-dynamics-based methods[46−49] such as free energy perturbation were found to be highly accurate but cannot be applied to combinatorial problems. Other approaches that use a single conformation include the use of soft Lennard-Jones potentials[50] or pharmacophore-guided docking.[51]

Herein, we disclose our efforts toward the development of an oriented induced-fit docking method able to dock and predict the activity of highly flexible inhibitors of BACE 1. We have addressed this objective by first developing an accurate flexible ligand/flexible protein docking method and subsequently by the creation of a scoring function based on a force field.

## Theory and Implementation

**A New Genetic Algorithm-Based Docking Method.** Prior to the de novo design and synthesis of novel BACE 1 inhibitors, we wished to evaluate the reliability of existing automated docking programs and scoring functions. Initial attempts to dock the highly flexible inhibitors **2** and **3** using fully automated docking programs including AutoDock 3.0,[52,53] FlexX,[54,55] and DOCK 4.0[56−58] led to unrealistic binding modes when compared to the X-ray crystal structures. These failures were attributed to the large flexibility of the inhibitors (DOCK was recently successfully used in the docking of a rigid inhibitor)[18] and to the particular shape of the binding site. The $S_1$, $S_2$, and $S_3$ pockets in BACE 1 merge into a large cavity on one side, and $S_2'$ and $S_3'$ pockets merge into a slightly smaller volume. The $S_1'$ pocket and the catalytic site constitute a narrow channel, which links these two large and partly solvent-exposed pockets. In addition, the key hydrogen bond network between the crucial hydroxyl group of the hydroxyethylene subunit and the two catalytic aspartates is quite difficult to predict accurately. We therefore developed a new genetic algorithm that uses molecular mechanics to accommodate the docking of flexible inhibitors into large proteins. This genetic algorithm protocol, detailed in Methods, allowed us to dock the pseudopeptides **2** and **3** with binding modes close to those observed in the crystal structures of the same Tang inhibitors (Figure 3). To increase the accuracy and speed of the method, the key hydroxyl group in the hydroxyethylene isostere subunit was used as an anchor and was forced into the catalytic site while creating the initial population and was released for the following evolution. Directing the docking with the use of an anchor was critical to the success of this method. Initial attempts to dock compounds **2** and **3** without this bias led to inaccurate results or exceedingly long computation times. The main difference between the docked and experimentally observed binding modes of these two inhibitors comes from their terminal residues. The comparison of the available crystal structures reveals that these terminal residues do not make specific interactions with the binding site residues. Considering that this docking was performed in vacuo, good binding modes were found with rmsd values below 3 Å, which were below 2 Å when the two terminal residues of the inhibitors were not considered. With this

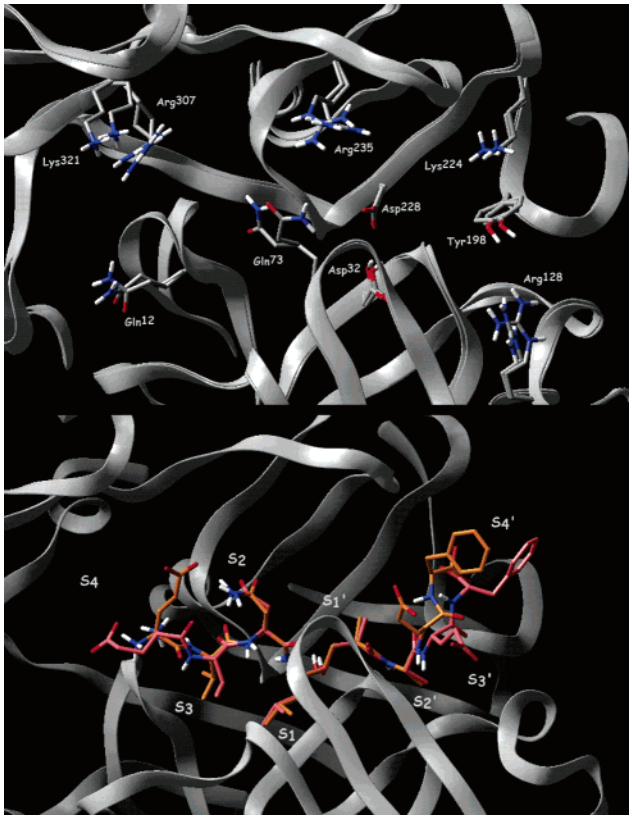



**Figure 3.** Top: superposition of two representative crystal structures. Bottom: binding mode of compound **2** with BACE 1 modeled using the developed genetic algorithm (pink) and experimental (orange).

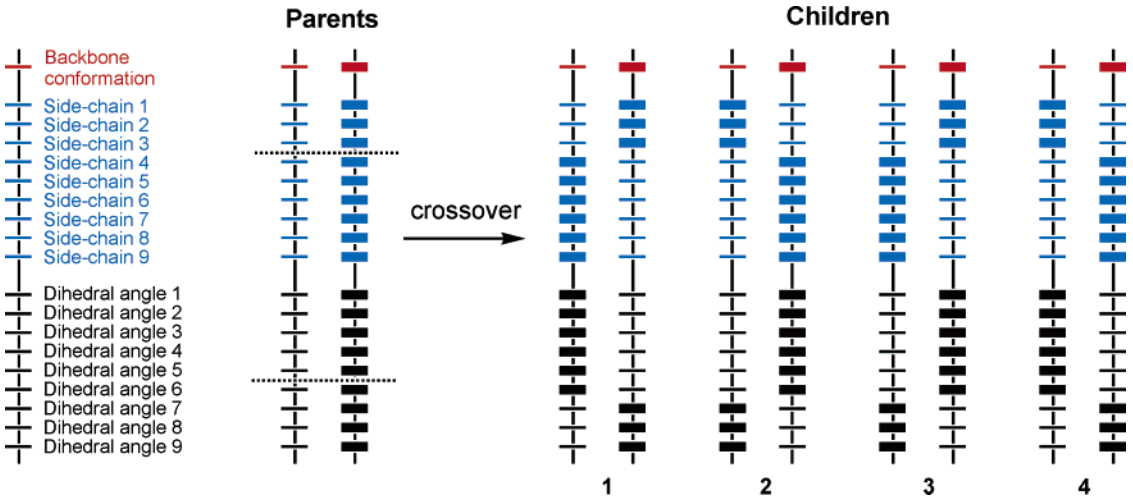



**Figure 4.** Chromosomes used for the docking of a flexible inhibitor (nine rotatable bonds) to a flexible protein (nine flexible side chains) and the two-point crossover operation used. The dotted lines designate the two points of crossover used for illustration.

algorithm in hand, we thought to develop a docking protocol that would also consider the flexibility of the protein.

**Docking to Flexible Proteins.** Critical motions of the binding site residues inducing additional interactions with the inhibitors were observed (Figure 3). The rmsd values of 0.35−1.0 Å between the available structures truncated in a manner detailed in Methods were measured. In fact, these rmsd values essentially reflect the flexibility of the arginine and lysine side chains found in the binding site. These motions revealed that an accurate protocol should account for the induced fit of the binding site.[31,32,35] To implement the flexibility of the protein, additional genes coding the protein binding site conformations were introduced into the genetic algorithm protocol. One gene codes the backbone conformation (red in Figure 4), 19 genes code 19 relevant side chains of the binding site residues (in blue in Figure 4), and other genes code each rotatable bonds of the inhibitor.

During the evolution, a two-point crossover approach was used (Figure 4). The first crossover operator was applied to the portion of the chromosome corresponding to the ligand (in black in Figure 4), while a second crossover operator was applied to the portion corresponding to the protein side chains (in blue in Figure 4). In practice, a regular two-point crossover operation would produce only pair 1 in Figure 4. To improve the efficiency of the evolution, a specific treatment of the junctions between the three sections (identified as black, blue, and red in Figure 4) of the chromosome was implemented in order to unlink these three portions. A switch (random number 0 or 1) has been added at the two junctions and virtually allows exchanges between the chromosomes at these points. This approach can be defined as a four-point crossover operation with a floating point on each section and two points fixed at the junctions. This implemented method could now potentially lead to the four possible pairs of children depicted in Figure 4 and to independent treatments of the protein and of the ligand albeit with a single chromosome. To further exploit this approach, a crossover rate of 80% was used. This rate allowed single-point crossover or no crossover operations to be applied to a fraction of the population while the other two switching operations were still effective. In these cases, the parents will simply exchange protein conformation and/or ligand conformation. Mutations were also applied to the chromosomes with a user-defined rate of 5%.

To accelerate the evolution, intermediate minimization steps applied to the ligands were added. This approach is similar to the Lamarckian genetic algorithm used in AutoDock.[53] Initial attempts using intermediate minimization of the side chain structures while docking the ligand failed to provide reasonable protein conformations. The genes coding the protein conformations were therefore restricted to experimentally observed side chain and main chain conformations. This approach can be related to the use of predefined libraries of side chain conformations and combinatorial construction of protein structures.[38−41] However, in the present case, the main-chain flexibility was also modeled. So far, the main-chain flexibility has only been modeled with more extensive methods such as molecular dynamics. More interestingly, the flexibility was modeled *while* docking the ligand. This technique differs from existing methods that successively dock the ligands to each composite conformation.[40,41,59,60] Furthermore, our approach is not restricted to a few side chains. One can probably also use predefined libraries of side chain rotamers in the developed algorithm instead of experimentally observed conformations, but this approach has not been tested herein. Obviously, a larger number of input structures would probably lead to improved accuracy. Of course, a very large number at the same time creates intractable computations with the increase of time therefore not recommended for screening libraries.

During the minimization stages, the hydroxyl hydrogens of the protein were also free to move while the rest of the protein was fixed. This procedure, also used in GOLD,[61] allowed the optimization of the interactions of the Thr and Tyr hydroxyl groups with the inhibitor via hydrogen bonds, adding further flexibility to the protein.

In the present study, six X-ray crystal structures (two monomeric structures and four structures from two dimers) were available[5,62,63] and were used to prepare the initial population. Thus, each residue side chain can independently adopt six different conformations, and the same treatment applies to the backbone. With these conformation sets, the genetic algorithm protocol now optimizes both the ligand binding mode and the protein binding site conformation. To increase the speed of the computations, AMBER united-atom was selected instead of all-atom force fields. With this approach the CPU time required for a single docking run ranged from $^1/_2$ to 20 h on a standard workstation, depending on the flexibility of the inhibitor (up to 33 rotatable torsions for compound **3**). Developments are in progress in our laboratory to reduce the time required by the whole protocol by at least 1 order of magnitude.

To assess the accuracy of this approach, inhibitors cocrystallized with the enzyme were first studied. The two highly flexible inhibitors **2** and **3** were docked three times. Each run

led to binding modes and protein conformations highly similar to the crystal structure, indicating a great convergence of this protocol. The rmsd between experimental and modeled structure of BACE 1 when bound to inhibitor **2** is 0.18 Å, while the six initial protein structures have rmsd of 0.35−1.0 Å when compared to each other. Similarly, when inhibitor **3** is docked to BACE 1, the modeled structure of the protein is very close to the experimentally observed structure (rmsd = 0.24 Å). The high convergence observed with such large peptidic inhibitors demonstrates the great potential of the developed induced-fit docking method.

**RankScore Scoring Function.** When docking compounds, the scoring function should discriminate between conformers of a single molecule. When ranking compounds, the function should discriminate between different molecules. While the desolvation and entropy contributions of two conformers are similar, the same contributions differ significantly between two molecules. We therefore hypothesized that a poor description of the contribution of water and entropy may explain the variable discrimination of active versus inactive compounds with various scoring functions.[64−66] A careful investigation of both the desolvation/solvation phenomena and of the entropy change is needed in order to develop an accurate scoring function. The proposed hybrid function (eq 1) uses a LUDI-type empirical evaluation of the entropy change of the ligand, a LUDI-type empirical evaluation of the hydrogen bonds, and the three terms used for the LIE method: an evaluation of the electrostatic and van der Waals interactions based on a force field, and the solvation contribution to the free energy of binding. This last term was required to account for the desolvation of highly charged compounds included in the training set.

$$\Delta G_{\text{binding}} = \Delta G_0 + 0.14N_{\text{ROT}} + \sum(\text{scale factor})[(0.26U_{\text{vdW}}^{\text{inh-prot}} + 0.035U_{\text{elec}}^{\text{inh-prot}} + 0.80f_{\text{hb}}(\Delta r,\Delta\alpha)] \quad (1)$$

In eq 1, the hydrogen bond term appears virtually twice because most of the force fields implicitly consider hydrogen bond interactions as electrostatic in nature. However, the docking to flexible BACE 1 was carried out with the AMBER94[67,68] force field, which treats electrostatic interactions and hydrogen bonds separately.

LUDI evaluates the hydrogen bond strength as varying linearly as a function of the α (e.g., O···H−N, ideally 180°) and $\beta$ (e.g., C═O···H, ideally 120°) angles and the length.[26] The hydrogen bond contribution in our scoring function was evaluated using a similar approach. Because the docking process relied on a force field evaluation of the binding mode, we thought that including the $\beta$ angle (which is not optimized during this process) would not be appropriate and therefore used only the α angle. LUDI scoring function accounts for the complementarity of shape by evaluating the contact surface, whereas force fields compute the van der Waals contribution for long-range interaction.[26] Our scoring function takes into account an intermediate treatment of this contribution by calculating van der Waals interactions only with binding site residues. The choice of such short-range interactions was dictated by the higher predictiveness of the force field scoring when only residues in proximity were considered (Table 1, entries 11 and 14). To the hydrogen bond and van der Waals terms were added the Coulombic interactions evaluated by the force field.

When one deals with compounds exhibiting a large range of lipophilicities, the contribution of desolvation to the binding is critical. In addition, to evaluate the whole binding process (as does LIE), the initial state in water and the final state (ligand bound to the protein) should be considered. Thus, the solvation difference between the final state (complex solvation energy) and the initial state (ligand and protein solvation energies) was computed. Six solvation models, which include methods based on atomic parameters[69] or on solvent accessible surface area or employ a more accurate calculation of the polar contribution using a finite difference of the nonlinear Poisson−Boltzmann equation, were evaluated. Surprisingly, none of these methods were found to significantly improve the accuracy of the function (see Results and Discussion). The solvation contribution was therefore removed for this work even though preliminary assessment of the transferability of this scoring function revealed that this contribution significantly improves the accuracy when looking at other enzymes.

**Table 1.** Spearman Coefficients for the Ranking of the Training Set

| entry | scoring function | rigid protein docking | flexible protein docking |
|---|---|---|---|
| 1 | PLP2 | 0.75 | 0.64 |
| 2 | PLP1 | 0.68 | 0.62 |
| 3 | LUDI | 0.47 | 0.37 |
| 4 | PMF | 0.21 | 0.29 |
| 5 | D-Score | 0.04 | 0.61 |
| 6 | G-Score | 0.29 | 0.20 |
| 7 | Chem-Score | 0.20 | 0.17 |
| 8 | LigScore2 | 0.71 | 0.47 |
| 9 | LigScore1 | 0.41 | 0.06 |
| 10 | force field[a] | 0.48[c] | 0.51[d] |
| 11 | force field, vdW[a] | 0.36[c] | 0.54[d] |
| 12 | force field, elec[a] | 0.39[c] | 0.41[d] |
| 13 | force field[b] | 0.55[c] | 0.57[d] |
| 14 | force field, vdW[b] | 0.45[c] | 0.69[d] |
| 15 | force field, elec[b] | 0.42[c] | 0.40[d] |
| 16 | RankScore (this work) | 0.68 | 0.80 |

[a] Intermolecular interaction; cutoff = 15 Å. [b] Intermolecular interaction with the binding site residues only (with at least one atom within 5 Å from the ligand). [c] CFF91. [d] AMBER94.

Although a number of empirical scoring functions explicitly include an entropy penalty for each frozen bond of the ligand, only a few account for the protein entropy, which is currently assumed to be constant regardless of the ligand.[70] The force-field-based scoring functions also lack a term for the entropy contribution to binding. The intrinsic consideration of the entropy in some empirical scoring functions (e.g., PLP2 and LigScore2) may explain their greater ability in ranking compounds (vide infra). The entropy term due to the freezing of torsions used here was similar to the one proposed in the LUDI-type scoring functions. The binding site of BACE 1 features five highly flexible residues (Arg128, Arg235, Arg307, Lys224, Lys321). Lysine and arginine side chains have been shown to be the two most flexible side chains in proteins.[71] In addition, the selected training set contains inhibitors of different lengths that are expected to interact with a different number of residues within the binding site. The mobility of these residues is therefore modulated by the ligands. As the binding enthalpy increases, the mobility of the protein decreases, resulting in a greater decrease in entropy.[72] Thus, the tighter a ligand is bound, the more frozen the side chains will be. This is indeed one of the more common explanations of the enthalpy−entropy compensation in proteins.[72] To account for this phenomenon, the interactions (van der Waals, electrostatic, and hydrogen bonding) with the side chains were scaled. The scaling pattern is related to the flexibility of the side chains. The interactions with the backbone and the catalytic aspartic acids were not scaled.

## Results and Discussion

**Induced-Fit Docking.** In this work, the enzyme/inhibitor complex is coded as a chromosome and optimized by means of a genetic algorithm. One part of the chromosome codes the binding mode of the ligand, and the second part codes the backbone and side chain conformations of the protein. Each torsion of the ligand is coded with one gene, and each entire side chain or entire backbone of the protein is coded with one gene. We considered a library of conformations restricted to the experimentally observed conformations of the protein. Initial work with more flexible side chains led to inconsistent protein conformations. Full optimization of the side chain conformations would in fact require the study of solvated complexes. Although in this case study only six starting structures were used, a more complete set of structures can be developed from routine molecular dynamics simulations or obtained from NMR studies. As revealed later, the accuracy of the docking protocol and of the developed scoring function was enhanced when the protein flexibility was considered. These results, which stem from the optimized shape complementarity of both partners when the protein is allowed to move, are consistent with observations from other groups.[31−45]

A training set of 50 highly diverse compounds exhibiting from subnanomolar to no activity was selected. Application of the induced-fit docking method led to 50 modeled complexes. A close look at these structures reveals that the 50 compounds are bound with similar orientations and conformations. However, a simple positioning by analogy with crystallized inhibitors and minimization would not be able to optimize the binding mode of compounds such as **42**, **45**, and **50** (Table A, Supporting Information). The similarity of the binding modes is additional proof of the high convergence of the method. It is worth mentioning that all these docked compounds are highly flexible and that flexibility has been, to date, an obstacle to accurate docking methods. In addition, 19 side chains and the backbone of the protein were also considered as flexible.

**Existing Scoring Functions.** Most of the existing scoring functions attempt to approximate the free energy of binding and rely on additive effects.[24] To evaluate the predictive power of some available scoring functions, a training set of 50 known inhibitors (22 charged and 28 neutral inhibitors described in the Supporting Information), which cover a large cross section of biological activities (subnanomolar to micromolar and even inactive), was used. This set was selected with an emphasis on chemical diversity, although the reported structures are built around similar scaffolds. The four compounds that were reported as inactive were arbitrarily attributed the value of 1 000 000 nM. This allowed us to calculate a binding free energy for these inactive compounds. Using other arbitrarily assigned values did not significantly affect the predictive power of the developed scoring function. Compounds in this training set were alternatively docked in the rigid and flexible protein and were scored using nine available scoring functions (LUDI,[26] PLP1,[73] PLP2,[74] PMF,[75] LigScore1,[76] LigScore2,[76] D-Score,[56] G-Score,[61] and ChemScore[77]) and two force fields (CFF91 and AMBER94[67,68]). During the course of our work, Wang et al.[64] reported the use of the Spearman coefficient as a tool for evaluating the ability of the scoring functions to rank the compounds. For the sake of comparison, this coefficient was also used in this work, and the results correlate well with the data provided by Wang and co-workers. A coefficient of 1 indicates a perfect ranking of the compounds, while a coefficient of 0 indicates a random ranking of compounds. Table 1 summarizes the ranking of the docked conformations of the training set by known scoring functions as well as by RankScore.

Thus, for each scoring function, the entire training set of 50 compounds was analyzed and ranked accordingly. For the rigid docking study, PLP2 and LigScore2 appeared as the most accurate scoring functions with Spearman coefficients above 0.7. A more in-depth study of each scoring function pointed out the importance of hydrogen bonds (as calculated by LUDI) and van der Waals interaction (as evaluated by force fields). These observations were further exploited in the development of a new scoring function (see below). It is not clear to us why the LigScore1 and LigScore2 performance was so affected by the incorporation of the flexibility and why D-Score benefited so greatly from the flexibility of the protein. Nevertheless, the performance of the other scoring functions was about the same with either the rigid or the flexible protein.

The great performance of AMBER scoring and PLP2 is in good agreement with the work from Charifson et al.[78] However, they have also found ChemScore to be accurate.[78] The need for a highly predictive scoring function prompted us to develop a more detailed protocol for ranking the relative binding affinity of sets of potential inhibitors.

**RankScore Scoring Function.** To date, most of the reported scoring functions accurately guide the docking of compounds by identifying the experimentally observed binding mode among a large number of different poses. However, they perform poorly when trying to discriminate between a series of active and inactive compounds.[29,64−66,79] A recent comparative study of 11 scoring functions confirmed this limitation.[64] When applied to the docking of inhibitors to BACE 1, the empirical scoring function PLP2 and LigScore2 performed well with a good ranking ability while other functions poorly ranked the compounds.

With the objective of securing a more discriminating protocol, we developed our own scoring function (RankScore) focusing on the loss of entropy and on the water effect upon binding as presented in Theory and Implementation. A multiple linear regression was carried out to assign the weights to each contribution, and rmsd values of 2.14 and 1.59 kcal/mol were obtained for the docking to rigid and flexible proteins, respectively. The mathematical basis for RankScore is shown in eq 1. Interestingly, the weights associated with the van der Waals (0.260) and electrostatic (0.035) terms are very similar to those developed by Reynolds and co-workers (LIE study, 0.204−0.239 and 0.014−0.060).[16,17] This correlation indicates that the scoring of the single docked conformation is a good approximation of the scoring of a larger sampling of structures obtained from molecular dynamics (MD) simulations. This also demonstrates that the docking protocol developed in this work is accurate given the time needed to dock a compound (hours) compared to the time needed for an LIE study (days). While this manuscript was in preparation, a similar approach was successfully applied by Huang and Caflish to BACE 1 and HIV 1 protease.[80] They combined energy minimization and accurate solvation estimation to develop predictive equations (rmsd of about 1.0 kcal/mol). In this case, the weight attributed to the van der Waals contribution (0.2737) to the free energy of binding was similar to ours, while the electrostatic was assigned a higher weight (0.1795) than in this work. Addition of an entropy contribution slightly increased the accuracy.

We next optimized the scaling factors to model the entropy cost associated with the side chain freezing. The values were 1.00 for the backbone, 0.80 for flexible side chains, or 0.60 for Arg and Lys side chains. This is an indication of the role of the

entropy associated with the side chain mobility. This may also partly account for the desolvation of the charged residues.

The weights attributed by linear regression analysis to each term of RankScore can be related to experimental values.[25] For instance, an energy of 0.80 kcal/mol was attributed to an ideal hydrogen bond. This value is in good agreement with the reported experimental values (1.1 kcal/mol).[81] Similarly, a contribution of 2 kcal/mol for a salt bridge is calculated by the developed function and found to be in good agreement with the literature (2−5 kcal/mol).[25] The freezing of a bond upon binding was found to increase the Gibbs energy by 0.14 kcal/mol, a value that is lower than that of LUDI and AutoDock scoring functions (0.33 and 0.31 kcal/mol, respectively). The remaining constant ($\Delta G_0$) represents the rigid-body entropy loss associated with the rotational and translational restrictions imposed on the bound ligand and enzyme desolvation. This latter contribution can be considered as roughly constant from one system to another.

The use of scaling factors to model the entropy cost greatly enhanced the accuracy of the RankScore scoring function. Thus, the developed protocol includes the protein flexibility and the evaluation of the entropy loss upon binding to an inhibitor. When these scaling factors were removed, linear regression led to weights with no physical meanings; the weight attributed to the electrostatic contribution was negative. An rmsd of 1.74 kcal/mol was calculated when no scaling was employed, and a weight of zero was assigned to the electrostatic contribution. These last results validated our assumption that the entropy loss of the macromolecule should be appropriately estimated. This also confirmed that the poor ability of the available scoring functions to rank compounds may be related to the oversimplification of this term.

A low weight was initially attributed to the solvation contribution; however, the role it played was not significant. This last result is quite unexpected when considering the large variety of polar and nonpolar compounds in the training set. A recent study on BACE 1 from Reynolds and co-workers[17] led to a similar conclusion. In the meantime, Brooks and co-workers[66] have investigated the role of the generalized Born and the Poisson continuum models in combination with a force field (CHARMm). This force-field-based scoring showed accuracy close to that obtained with available scoring functions. In the present work, scaling down the electrostatic and van der Waals contributions and scaling down the interactions with lysine and arginine side chains seem to account for the desolvation/solvation process.

As a result, the Spearman coefficient for RankScore (Table 2, entry 16) is now higher than those computed for other available scoring functions. The quality of the final scoring function was illustrated by an rmsd of 1.59 kcal/mol and an $r^2$ value of 0.624 (Figure 5). Removing the three major outliers ($\Delta G$(obsd) − $\Delta G$(pred) > 3 kcal/mol) led to an rmsd of 1.19 kcal/mol and an $r^2$ value of 0.789. The CPU-intensive but more accurate LIE procedure as applied by Reynolds and co-workers led to an optimized rmsd of 1.10 kcal/mol (12 compounds) and an optimized rmsd of 0.87 kcal/mol when the two charged compounds were removed. The more similar LIE approach used by Huang and Caflish demonstrated a better rmsd of 1.0 kcal/mol.[80] However, the large training set used in the present work gathers compounds from different sources and SAR data were collected using different biological assay conditions. The theoretical error (rmsd < 2 kcal/mol) is therefore believed to be within a reasonable range. The three outliers are **12** (pred, 61 $\mu$M; obsd, 0.04 $\mu$M), **27** (pred, 0.9 $\mu$M; obsd, 1400 $\mu$m),

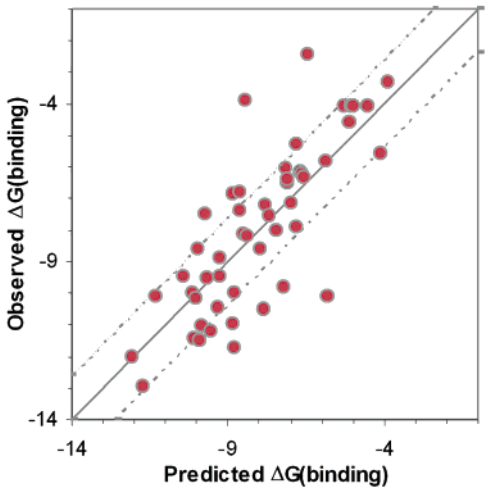



**Figure 5.** Predicted binding energies vs observed binding energies for the training set. Dotted lines delineate errors of 1 order of magnitude for the $K_i$ values.

and **32** (pred, 15 $\mu$M; obsd, 15 000 $\mu$M) listed in the Supporting Information. Visual inspection of docked compound **32** revealed that the phenyl ring was not well tolerated (not flat) by the enzyme. The lack of a term evaluating the ligand strain in the developed scoring function may explain this failure. Compounds **12** and **27** feature nitrogen-containing heterocycles. The protonation state of these groups is strongly influenced by the local p$K_a$ of the binding site, and local proton exchange between the protein and the inhibitor may occur upon binding. However, a standard docking procedure cannot account for such a phenomenon.

**Ranking of a Validation Set.** An ideal computational tool in virtual screening would extract the most active compounds from a large and diverse library, and the best protocol would assign high ranking to these active compounds. To evaluate the performance of the developed RankScore protocol, a validation set of 80 peptidomimetics and pseudopeptides developed in our laboratory was selected.[63,82,83] Most of these 80 compounds are similar to the compounds found in the training set, although they feature nonpeptidic moieties such as cyclopentanes.[63] One-eighth of the selected library exhibit $IC_{50}$ values below 500 nM (Figure 6). Thus, the selected compounds were docked, scored, and ranked using RankScore. For the sake of comparison, the same docked structures were ranked with PLP2 and LigScore2, which were found to be the two best scoring functions (Spearman coefficients for the training set of 0.64−0.75 and 0.47−0.71, respectively). Once the molecules were ranked, the performances of the different scoring functions were compared. Unfortunately, the available biological data were not sufficient to compute a correlation factor. These data included $IC_{50}$ values only for the most active compounds and not for compounds with inhibitory activity lower than 50% at 10 $\mu$M. In Figure 6, we compare histograms representing the composition of the initial library (random ranking) to ranked lists with RankScore, PLP2, and LigScore2. It is clear that the top of the ranking lists for RankScore and PLP2 is enriched in actives relative to the random ranking list. It also appears that LigScore2 poorly discriminates actives and inactives within this validation set. A focused library extracted from a LigScore2 ranking list would have nearly the same composition as a random library. This failure is striking when considering the good prediction of LigScore2 for the training set.

Several reasons can explain the weaker prediction of RankScore for two of the good binders in the validation set. First,

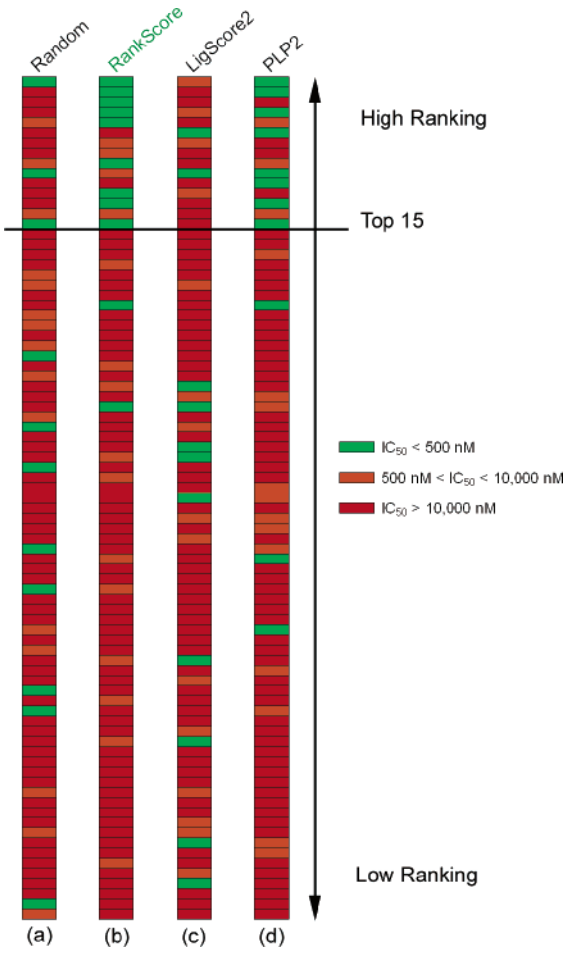



**Figure 6.** Distribution of (a) randomly generated library, (b) library scored/ranked with RankScore, (c) library scored/ranked with LigScore2, (d) library scored/ranked with PLP2.

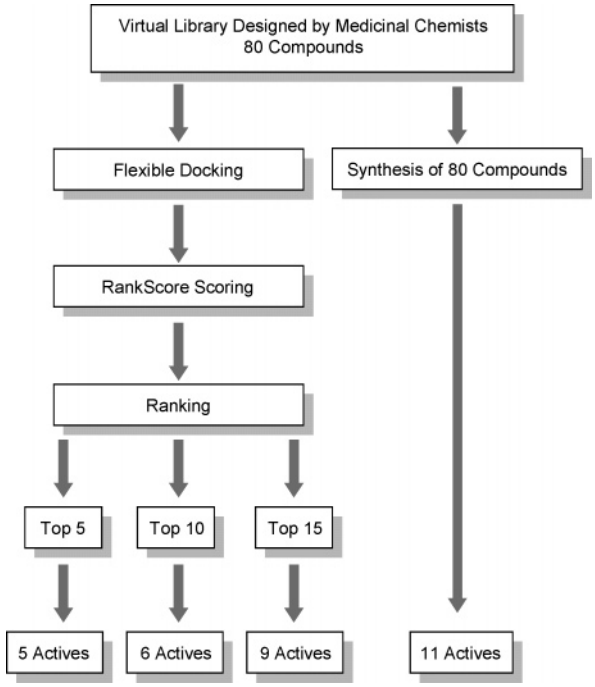
the docking was performed in a vacuum, while water molecules can have an active role to play in the binding. Second, the scoring might have been computed on poorly docked compounds. Third, the initial set of six structures of the protein may not cover the conformational space properly. This last issue can easily be addressed using molecular dynamics simulations or NMR structures. An additional experimental error can be attributed to the use of $IC_{50}$ values whereas the binding energies predicted for the training set were correlated to $K_i$ values. It is also worth mentioning that the RankScore scoring function was developed for this particular enzyme and that it was developed from docked and not experimentally observed structures.

From a practical point of view, if one had screened the validation set prior to synthesis and biological assays and had synthesized only the top 15 molecules, one would have made 8 (PLP2) or 9 (RankScore) of the 11 highly active compounds ($IC_{50}$ below 500 nM) (Chart 1). Overall, the timeline for the discovery of new active entities would have been significantly reduced. In addition, RankScore ranked the 11 highly active structures in the top 33, while two compounds ranked 48 and 55 with PLP2. In Chart 1, we show the protocol used and the number of actives predicted. When a desolvation/solvation term was included, the 11 highly active structures were found in the top 27.

## Conclusion

A docking method was first developed that properly docked the highly flexible inhibitors contained in the training set. Next,

**Chart 1.** Flow Chart of the Protocol Used for the Validation of the Developed Docking/Scoring Method




an extended version of this method that accounts for protein fit upon binding was developed and showed improved efficiency. The method is based on a genetic algorithm where chromosomes code the entire complex. Genes code the ligand conformation, while other genes code the protein conformation. The modeling of both the side chains and the main chain conformations while inhibitors are being docked is unique and easy to implement. The docking data clearly indicate that the oriented docking approach effectively accounts for the large flexibility of the inhibitors (up to 33 rotatable bonds for compound **3**) and for adjustments in the protein structure. The docking method has also been found to be highly convergent (three runs with the most flexible inhibitors led to similar binding modes and scores) although costly in time.

A force-field-based scoring function for BACE 1 (RankScore), which accounts for the protein entropy loss, ligand desolvation, and complex solvation, was next developed and demonstrated a great ability to discriminate between active and inactive compounds. Accounting for the entropy loss of the protein using a scaling pattern applying to flexible binding site side chains significantly enhanced the accuracy of the function. The predictive power of RankScore was compared to that of LigScore2 and PLP2. In fact, RankScore was able to discriminate compounds with similar structures but different biological activities, while the ranking of this validation set with PLP2 was less accurate. The ability to rank docked structures also demonstrates the appropriate binding modes proposed by the docking procedure.

The capacity of the approach described herein to discriminate actives and inactives has great potential for the future design of peptidic and pseudopeptidic BACE 1 inhibitors. The extension of this protocol to other enzymes and the evaluation of the transferability of RankScore are underway.

## Methods

**General.** The developed protocol was fully interfaced within InsightII from Accelrys Inc.[85] using BCL code, awk, and perl scripts.[86] Standard force field (AMBER94) atom partial charges

were assigned to the protein, while the ligand charges were computed using semiempirical calculations with MNDO. Missing parameters (e.g., sulfonamide) were obtained from the Amber* force field implemented in MacroModel software.[87] Graphical displays were printed from the Maestro molecular modeling system.[88] Scores with LigScore1, LigScore2, PLP1, and PLP2 were obtained using Cerius 2 software programs.[89] Scores with PMF, D_Score, G_Score, and Chem_Score were obtained using CScore standalone from Tripos Inc.[90] This protocol includes an additional minimization using the Tripos force field.

**Starting Enzyme Structure.** The X-ray structure of inhibitor/BACE 1 complexes were retrieved from the Brookhaven Protein Data Bank (PDB codes 1FKN, 1M4H, 1YM4) and from a patent[91] and used as a starting point. The hydrogen atoms were added, the energy was minimized, and the result was visually inspected. The inhibitors and water molecules were removed from the obtained complexes. To reduce the computational time, the enzyme structures were next truncated for the docking study. Truncation was achieved by keeping residues with at least one atom within 15 Å from the OM99-2 structure in the relaxed complex. Geometric considerations led us to protonate Asp32 and deprotonate Asp128. The protonation state of the enzyme has been investigated by different groups. However, the different studies led to opposite conclusions.[18,92,93]

**Genetic Algorithm-Based Docking Method.** The algorithm was written using BTCL code as implemented under Discover 3.0 and is as follows:[94] (1) Create the initial population (300 individuals in this work) by randomly rotating all the rotatable bonds of the inhibitor with the hydroxyl group positioned near the two catalytic aspartates followed by relaxation of the inhibitors by quick minimization; keep those with potential energy of less than a user-defined value (3000 kcal/mol). The initial population was constructed using six X-ray structures for the protein. (2) Begin a user-defined number of genetic operations (population usually converges within 100−200 generations). (2.1) Select two parents within the 300 individuals. (2.2) Produce two children by the developed two-point crossover (one point for the ligand and one point for the protein). (2.3) Apply mutation with a user-defined rate (0.05 in this work). (2.4) In a user-defined fraction of the population (0.25 in this study), optimize the solution (local search) by a user-defined number of steps of conjugate gradient energy minimization (250 in this work) with a convergence criterion of 0.001 kcal/mol. During this minimization stage, the Thr, Ser, and Tyr hydroxyl hydrogens are free to move while the rest of the protein is fixed. (2.5) Replace in a steady-state mode the least fit parent if the child's fit is better (potential energy is lower). (2.6) Reiterate the process until convergence is obtained.

The docking fitness function was a simple force field energy evaluation with a scaled (0.5) Coulombic term and a distance-dependent dielectric constant ($\epsilon = 4r$) to account for the desolvation and the solvent shielding, respectively. This electrostatic scaling was indeed found to significantly improve the efficiency of the docking method. The following side chains were considered as flexible: Gln12, Leu30, Asp32, Tyr71, Thr72, Gln73, Phe108, Ile110, Ile118, Ile126, Arg128, Tyr198, Lys224, Asp228, Thr231, Thr232, Arg235, Arg307, Lys321. Each of these side chains was coded as a gene, while the rest of the protein (backbone and other side chains) was coded in a single gene.

**Scoring Function.** The resulting docked structures were further optimized by conjugate gradient energy minimization with a convergence criterion of 0.001 kcal/mol and scored using the RankScore scoring function. This function includes (1) the van der Waals and Coulombic interactions of the complex measured using AMBER94, (2) the hydrogen bond profile of the complex using scripts that read the coordinate files, (3) an extra term that can also be included, which is the solvation free energy of binding calculated using DELPHI module (Insight II Users' Guide) by computation of the complex and the ligand solvation energies. AWK and perl scripts combine all these values into a final score.[85]

**Acknowledgment.** We thank the Canada Foundation for Innovation (N.M.) and NSERC through the NSERC/CRD Program with Novartis (Basel, Switzerland) (S.H.) for financial support.

**Supporting Information Available:** Table listing the compounds in the training set and their activities. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Lahiri, D. K.; Farlow, M. R.; Sambamurti, K.; Greig, N. H.; Giacobini, E.; Schneider, L. S. A Critical Analysis of New Molecular Targets and Strategies for Drug Developments in Alzheimer's Disease. *Curr. Drug Targets* **2003**, *4*, 97−112.
(2) Citron, M. Emerging Alzheimer's Disease Therapies: Inhibition of $\beta$-Secretase. *Neurobiol. Aging* **2002**, *23*, 1017−1022.
(3) Vassar, R. Beta-Secretase (BACE) as a Drug Target for Alzheimer's Disease. *Adv. Drug Delivery Rev.* **2002**, *54*, 1589−1602.
(4) Roggo, S. Inhibition of BACE, a Promising Approach to Alzheimer's Disease Therapy. *Curr. Top. Med. Chem.* **2002**, *2*, 359−370.
(5) Hong, L.; Koelsch, G.; Lin, X.; Wu, S.; Terzyan, S.; Ghosh, A. K.; Zhang, X. C.; Tang, J. Structure of the Protease Domain of Memapsin 2 (Beta-Secretase) Complexed with Inhibitor. *Science* **2000**, *290*, 150−153.
(6) Ghosh, A. K.; Hong, L.; Tang, J. Beta-Secretase as a Therapeutic Target for Inhibitor Drugs. *Curr. Med. Chem.* **2002**, *9*, 1135−1144.
(7) Cumming, J. N.; Iserloh, U.; Kennedy, M. E. Design and Development of BACE 1 Inhibitors. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 536−556.
(8) Hom, R. K.; Gailunas, A. F.; Mamo, S.; Fang, L. Y.; Tung, J. S.; Walker, D. E.; Davis, D.; Thorsett, E. D.; Jewett, N. E.; Moon, J. B.; John, V. Design and Synthesis of Hydroxyethylene-Based Peptidomimetic Inhibitors of Human $\beta$-Secretase. *J. Med. Chem.* **2004**, *47*, 158−164.
(9) Brady, S. F.; Singh, S.; Crouthamel, M.-C.; Holloway, M. K.; Coburn, C. A.; Garsky, V. M.; Bogusky, M.; Pennington, M. W.; Vacca, J. P.; Hazuda, D.; Lai, M.-T. Rational Design and Synthesis of Selective BACE 1 Inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 601−604.
(10) Lamar, J.; Hu, J.; Bueno, A. B.; Yang, H.-C.; Guo, D.; Coop, J. D.; McGee, J.; Gitter, B.; Timm, D.; May, P.; McCarthy, J.; Chen, S.-H. Phe*-Ala-Based Pentapeptide Mimetics Are BACE Inhibitors: P2 and P3 SAR. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 239−243.
(11) Chen, S.-H.; Lamar, J.; Guo, D.; Kohn, T.; Yang, H.-C.; McGee, J.; Timm, D.; Erickson, J.; Yip, Y.; May, P.; McCarthy, J. P3 Cap Modified Phe*-Ala Series BACE Inhibitors. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 245−250.
(12) Hu, B.; Fan, K. Y.; Bridges, K.; Chopra, R.; Lovering, F.; Cole, D.; Zhou, P.; Ellingboe, J.; Jin, G.; Cowling, R.; Bard, J. Synthesis and SAR of Bis-Statine Based Peptides as BACE 1 Inhibitors. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 3457−3460.
(13) Kimura, T.; Shuto, D.; Kasai, S.; Liu, P.; Hidaka, K.; Hamada, T.; Hayashi, Y.; Hattori, C.; Asai, M.; Kitazume, S.; Saido, T. C.; Ishiura, S.; Kiso, Y. KMI-358 and KMI370, Highly Potent and Small-Sized BACE1 Inhibitors Containing Phenylnorstatine. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 1526−1531.
(14) Åqvist, J. Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Çomput. Chem.* **1996**, *17*, 1587−1597.
(15) Åqvist, J. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res*. **2002**, *35*, 358−365.
(16) Tounge, B. A.; Reynolds, C. H. Calculation of the Binding Affinity of $\beta$-Secretase Inhibitors Using the Linear Interaction Energy Method. *J. Med. Chem.* **2003**, *46*, 2074−2082.
(17) Rajamani, R.; Reynolds, C. H. Modeling the Binding Affinities of $\beta$-Secretase Inhibitors: Application to Subsite Specificity. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 4843−4846.
(18) Park, H.; Lee, S. Determination of the Active Site Protonation State of $\beta$-Secretase from Molecular Dynamics Simulation and Docking Experiment: Implications for Structure-Based Inhibitor Design. *J. Am. Chem. Soc.* **2003**, *125*, 16416−16422.
(19) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. *Proteins: Struct., Funct., Genet*. **2002**, *47*, 409−443.
(20) Böhm, H. J.; Klebe, G. What Can We Learn from Molecular Recognition in Protein−Ligand Complexes for the Design of New Drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588−2614.
(21) Klebe, G. Recent Developments in Structure-Based Design. *J. Mol. Med*. **2000**, *78*, 269−281.
(22) *Structure-Based Ligand Design*; Gubernator, K., Böhm, H.-J., Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley: New York, 1998.

(23) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A Review of Protein−Small Molecule Docking Methods. *J. Comput.-Aided Mol. Des*. **2002**, *16*, 151−166.

(24) Tame, J. R. H. Scoring Functions: A View from the Bench. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 99−108.

(25) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed*. **2002**, *41*, 2644−2676.

(26) Böhm, H.-J. The Development of a Simple Empirical Scoring Function To Estimate the Binding Constant for a Protein−Ligand Complex of Known Three-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.

(27) Gohlke, H.; Hendlich, M.; Klebe, G. Predicting Binding Modes, Binding Affinities and "Hot Spots" for Protein−Ligand Complexes Using a Knowledge-Based Scoring Function. *Perspect. Drug Discovery Des.* **2000**, *20*, 115−144.

(28) Sotriffer, C. A.; Gohlke, H.; Klebe, G. Docking into Knowledge-Based Potential Fields: A Comparative Evaluation of DrugScore. *J. Med. Chem*. **2002**, *45*, 1967−1970.

(29) Bissantz, C.; Folkers, G.; Rognan, D. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem*. **2000**, *43*, 4759−4767.

(30) Bursavich, M.; G.; Rich, D. H. Designing Non-Peptide Peptidomimetics in the 21st Century: Inhibitors Targeting Conformational Ensembles. *J. Med. Chem*. **2002**, *45*, 541−558.

(31) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The Sensitivity of the Results of Molecular Docking to Induced Fit Effects: Application to Thrombin, Thermolysin and Neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547−562.

(32) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.

(33) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol*. **2004**, *337*, 209−225.

(34) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated Docking to Multiple Target Structures: Incorporation of Protein Mobility and Structural Water Heterogeneity in AutoDock. *Proteins: Struct., Funct., Genet*. **2002**, *46*, 34−40.

(35) Carlson, H. A. Protein Flexibility and Drug Design: How To Hit a Moving Target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.

(36) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular Docking to Ensembles of Protein Structures. *J. Mol. Biol*. **1997**, *266*, 424−440.

(37) Broughton, H. B. A Method for Including Protein Flexibility in Protein−Ligand Docking: Improving Tools for Database Mining and Virtual Screening. *J. Mol. Graphics Modell*. **2000**, *18*, 247−257.

(38) Leach, A. R. Ligand Docking to Protein with Discrete Side-Chain Flexibility. *J. Mol. Biol*. **1994**, *235*, 345−356.

(39) Leach, A.; Lemon, A. Exploring the Conformational Space of Protein Side-Chains Using Dead-End Elimination and the A* Algorithm. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 227−239.

(40) Frimurer, T. M.; Peters, G. G.; Iversen, L. F.; Andersen, H. S.; Moller, N. P.; Olsen, O. H. Ligand-Induced Conformational Changes: Improved Predictions of Ligand Binding Conformations and Affinities. *Biophys. J.* **2003**, *84*, 2273−2281.

(41) Anderson, A. C.; O'Neil, R. H.; Surti, T. S.; Stroud, R. M. Approaches to Solving the Rigid Receptor Problem by Identifying a Minimal Set of Flexible Residues during Ligand Docking. *Chem. Biol.* **2001**, *8*, 445−457.

(42) Källblad, P.; Dean, P. M. Efficient Conformational Sampling of Local Side-Chain Flexibility. *J. Mol. Biol.* **2003**, *326*, 1651−1665.

(43) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol*. **2001**, *308*, 377−395.

(44) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a Flexible-Receptor Docking Algorithm in a Model Binding Site. *J. Mol. Biol*. **2004**, *337*, 1161−1182.

(45) Schnecke, V.; Kuhn, L. A. Virtual Screening with Solvation and Ligand-Induced Complementarity. *Perspect. Drug Discovery Des*. **2000**, *20*, 171−190.

(46) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(47) Kidera, A.; Nakajima, N.; Higo, J.; Nakamura, H. Flexible Docking of a Ligand Peptide to a Receptor Protein by Multicanonical Molecular Dynamics Simulation. *Chem. Phys. Lett*. **1997**, *278*, 297−301.

(48) Kumar, S.; Payne, P. W.; Vasquez, M. Method for Free-Energy Calculations Using Iterative Techniques. *J. Comput. Chem*. **1996**, *17*, 1269−1275.

(49) Mangoni, M.; Roccatano, D.; Di Nola, A. Docking of Flexible Ligands to Flexible Receptors in Solution by Molecular Dynamics Simulation. *Proteins: Struct., Funct., Genet*. **1999**, *35*, 153−162.

(50) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem*. **2004**, *47*, 5076−5084.

(51) Moitessier, N.; Henry, C.; Maigret, B.; Chapleur, Y. Combining Pharmacophore Search, Automated Docking, and Molecular Dynamics Simulations as a Novel Strategy for Flexible Docking. Proof of Concept: Docking of Arginine-Glycine-Aspartic Acid-like Compounds into the $\alpha_v\beta_3$ Binding Site. *J. Med. Chem*. **2004**, *47*, 4178−4187.

(52) Goodsell, D. S.; Olson, A. J. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Struct., Funct., Genet*. **1990**, *8*, 195−202.

(53) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639−1662.

(54) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol*. **1996**, *261*, 470−489.

(55) Rarey, M.; Kramer, B.; Lengauer, T. Multiple Automatic Base Selection: Protein−Ligand Docking Based on Incremental Construction without Manual Intervention. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 369−384.

(56) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule−Ligand Interactions. *J. Mol. Biol*. **1982**, *161*, 269−288.

(57) DesJarlais, R.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure. *J. Med. Chem.* **1988**, *31*, 722−729.

(58) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078−1082.

(59) Cavasotto, C. N.; Abagyan, R. A. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol*. **2004**, *337*, 209−225.

(60) Källblad, P.; Todorov, N. P.; Willems, H. M. G.; Alberts, I. L. Receptor Flexibility in the in Silico Screening of Reagents in the S1′ Pocket of Human Collagenase. *J. Med. Chem.* **2004**, *47*, 2761−2767.

(61) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol*. **1997**, *267*, 727−748.

(62) Hong, L.; Turner, R. T., III; Koelsch, G.; Shin, D.; Ghosh, A. K.; Tang, J. Crystal Structure of Memapsin 2 ($\beta$-Secretase) in Complex with an Inhibitor OM00-3. *Biochemistry* **2002**, *41*, 10963−10967.

(63) Hanessian, S.; Yun, H.; Hou, Y.; Yang, G.; Bayrakdarian, M.; Therrien, E.; Moitessier, N.; Roggo, S.; Veenstra, S.; Tintelnot-Blomley, M.; Rondeau, J.-M.; Ostermeier, C.; Strauss, A.; Ramage, P.; Paganetti, P.; Neumann, U.; Betschart, C. Structure-Based Design, Synthesis, and BACE Inhibitory Activity of Carbocyclic and Heterocyclic Peptidomimetics. *J. Med. Chem*. **2005**, 5175−5190.

(64) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.

(65) Stahl, M.; Rarey, M. Detailed Analysis of Scoring Functions for Virtual Screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.

(66) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. Assessing Scoring Functions for Protein−Ligand Interactions. *J. Med. Chem.* **2004**, *47*, 3032−3047.

(67) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. A. New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(68) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

(69) *Delphi*, release 4.1.1; Accelrys Inc.: San Diego, CA, 2002.

(70) Giodanetto, F.; Cotesta, S.; Catana, C.; Trosset, J.-Y.; Vulpetti, A.; Stouten, P. F. W.; Kroemer, R. T. Novel Scoring Functions Comprising QXP, SASA, and Protein Side-Chain Entropy Terms. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 882−893.

(71) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-Chain Flexibility in Proteins Upon Ligand Binding. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 261−268.

(72) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Binding Affinities of Host−Guest, Protein−Ligand, and Protein−Transition-State Complexes. *Angew. Chem., Int. Ed*. **2003**, *42*, 4872−4897.

(73) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 317−324.
(74) Gehlhaar, D. K.; Bouzida, D.; Rejto, P. A. *Rational Drug Design: Novel Methodology and Practical Applications*; Parrill, L., Rami Reddy, M., Eds.; ACS Symposium Series 719; American Chemical Society: Washington, DC, 1999; pp292−311..
(75) Muegge, I.; Martin, Y. C. A General and Fast Scoring Function for Protein−Ligand Interactions: A Simplified Potential Approach. *J. Med. Chem*. **1999**, *42*, 791−804.
(76) Krammer, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.* **2005**, *23*, 395−407.
(77) Eldrege, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 425−445.
(78) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109
(79) Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jørgensen, F. S. A New Concept for Multidimensional Selection of Ligand Conformations (MultiSelect) and Multidimensional Scoring (MultiScore) of Protein−Ligand Binding Affinities. *J. Med. Chem.* **2001**, *44*, 2333−2343.
(80) Huang, D.; Caflisch, A. Efficient Evaluation of Binding Free Energy Using Continuum Electrostatics Solvation. *J. Med. Chem.* **2004**, *47*, 5791−5797.
(81) Williams, D. H.; Westwell, M. S. Aspects of Weak Interactions. *Chem. Soc. Rev*. **1998**, *27*, 57−63.
(82) Hanessian, S.; Hou, Y.; Bayrakdarian, M.; Tintelnot-Blomley, M. Stereoselective Synthesis of Constrained Oxacyclic Hydroxyethylene Isosteres of Aspartic Protease Inhibitors: Aldol and Mukaiyama Aldol Methodologies for Branched Tetrahydrofuran 2-Carboxylic Acids. *J. Org. Chem.* **2005**, *70*, 6735−6745.
(83) Hanessian, S.; Yun, H.; Hou, Y.; Tintelnot-Blomley, M. Stereoselective Synthesis of Constrained Azacyclic Hydroxyethylene Isosteres as Aspartic Protease Inhibitors: Dipolar Cycloaddition and Related Methodologies toward Branched Pyrrolidine and Pyrrolidinone Carboxylic Acids. *J. Org. Chem.* **2005**, *70*, 6746−6756.
(84) Coburn, C. A.; Stachel, S. J.; Li, Y.-M.; Rush, D. M.; Steele, T. G.; Chen-Dodson, E.; Holloway, M. K.; Xu, M.; Huang, Q.; Lai, M.-T.; DiMuzio, J.; Crouthamel, M.-C.; Shi, X.-P.; Sardana, V.; Chen, Z.; Munshi, S.; Kuo, L.; Makara, G. M.; Annis, D. A.; Tadikonda, P. K.; Nash, H. M.; Vacca, J. P.; Wang, T. Identification of a Small Molecule Nonpeptide Active Site $\beta$-Secretase Inhibitor That Displays a Nontraditional Binding Mode for Aspartyl Proteases. *J. Med. Chem.* **2004**, *47*, 6117−6119.
(85) *InsightII Modeling Environment*, release 2000.1; Accelrys Inc.: San Diego, CA, 2002.
(86) The scripts written specifically for docking to flexible BACE 1 are available on request from the authors. More general software is being coded and will be available in due course.
(87) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. MacroModel, an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.
(88) *Maestro*, version 6.5; Schrödinger, LLC: Portland, OR, 1999−2004.
(89) *Cerius2 Modeling Environment*, release 4.8.1; San Diego: CA, 2002.
(90) *SYBYL*, version 6.9.1; Tripos Inc. (1699 South Hanley Road, St. Louis, MO, 63144).
(91) Choppa, R.; Svenson, K.; Annis, B.; Akopian, T. N.; Bard, J.; Stahl, M. L.; Somers, W. S. Crystal structure of beta-site APP-cleaving enzyme (BACE) and uses thereof. PCT Int. Appl. WO 2002025276, March 28, 2002.
(92) Polgar, T.; Keseru, G. M. Virtual Screening for $\beta$-Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States at Asp32 and Asp228. *J. Med. Chem.* **2005**, *48*, 3749−3755.
(93) Rajamani, R.; Reynolds, C. H. Modeling the Protonation States of the Catalytic Aspartates in $\beta$-Secretase. *J. Med. Chem.* **2004**, *47*, 5159−5166.
(94) A first version of this algorithm has been previously reported: Moitessier, N.; Henry, C.; Len, C.; Chapleur, Y. Toward a Computational Tool Predicting the Stereochemical Outcome of Asymmetric Reactions. 1. Application to Sharpless Asymmetric Dihydroxylation. *J. Org. Chem.* **2002**, *67*, 7275−7282.

JM050138Y