Articles

Efficient Method for High-Throughput Virtual Screening Based on Flexible Docking: Discovery of Novel Acetylcholinesterase Inhibitors

Miho Yamada Mizutani and Akiko Itai*

Institute of Medicinal Molecular Design, Inc., 5-24-5 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received December 5, 2003

A method of easily finding ligands, with a variety of core structures, for a given target macromolecule would greatly contribute to the rapid identification of novel lead compounds for drug development. We have developed an efficient method for discovering ligand candidates from a number of flexible compounds included in databases, when the three-dimensional (3D) structure of the drug target is available. The method, named ADAM&EVE, makes use of our automated docking method ADAM, which has already been reported. Like ADAM, ADAM&EVE takes account of the flexibility of each molecule in databases, by exploring the conformational space fully and continuously. Database screening has been made much faster than with ADAM through the tuning of parameters, so that computational screening of several hundred thousand compounds is possible in a practical time. Promising ligand candidates can be selected according to various criteria based on the docking results and characteristics of compounds. Furthermore, we have developed a new tool, EVE-MAKE, for automatically preparing the additional compound data necessary for flexible docking calculation, prior to 3D database screening. Among several successful cases of lead discovery by ADAM&EVE, the finding of novel acetylcholinesterase (AChE) inhibitors is presented here. We performed a virtual screening of about 160 000 commercially available compounds against the X-ray crystallographic structure of AChE. Among 114 compounds that could be purchased and assayed, 35 molecules with various core structures showed inhibitory activities with IC_{50} values less than 100 μM. Thirteen compounds had IC_{50} values between 0.5 and 10 μM, and almost all their core structures are very different from those of known inhibitors. The results demonstrate the effectiveness and validity of the ADAM&EVE approach and provide a starting point for development of novel drugs to treat Alzheimer's disease.

Introduction

Drugs in general exhibit pharmacological activity by binding to a target protein, such as an enzyme or receptor. For recognition between the protein and ligand, it is important that the two molecules form a stable complex, rather than that the ligand has the same core structure as the natural or other synthetic ligands. The factors contributing to the stabilization of the complex structure include complementarity of shape, hydrogen bonding (H-bonding), and electrostatic and hydrophobic properties, as well as desolvation costs and internal strain when the complex is formed. Therefore, many compounds with a wide variety of core structures could act as ligands of a particular target. Experimental high-throughput screening (HTS) is generally used at present for discovering novel lead structures. In recent years, the techniques of isolation, purification, and structure determination of biomolecules have greatly advanced, and the functions and three-dimensional (3D) structures of increasing numbers of macromolecules are becoming available. Accordingly, a computer-aided approach to lead generation is currently of great interest to many researchers.

For computer-aided lead generation, two kinds of approaches, i.e., virtual screening of 3D databases and de novo structure construction, are possible when the 3D structure of the target is known or amenable to modeling. The latter approach has merit in terms of the variety of constructed candidate structures, while the former also the great advantage that, if available compounds (e.g., commercial or in-house ones) are searched, identified candidates can be tested for activity without the effort of synthesis. Moreover, virtual screening is superior to experimental HTS in several respects. For example, in virtual screening, compounds that have not been synthesized, and even "virtual ligands", can be tested, and it is also possible to apply the technique to target macromolecules for which assays are difficult or expensive in experimental HTS. Here, we focus on the virtual screening approach.

In structure-based virtual screening, ligand candidates should be selected based on intermolecular interaction with the ligand-binding region of the target protein in the most stable complexed form of the two molecules. For the correct prediction of the most stable

* To whom correspondence should be addressed. Tel: +81-3-5689-4051. Fax: +81-3-5689-4054. E-mail: itai@immd.co.jp.

10.1021/jm030605g CCC: $27.50
Published on Web 08/28/2004

complex structure in various systems, an efficient automated docking method, which satisfies the following requirements, is indispensable. First, the conformational space of compound should be fully explored, since the conformation in the isolated state (e.g., in crystal or solution, or in any energy-minimum state) is not necessarily similar to that of the bound form. The protein structure backbone should be more rigid than the structures of small compounds, but in both docking studies and virtual screening, one cannot avoid the "induced-fit" problem, that is, the conformational changes of side (and sometimes main) chains of a protein when various ligands bind to it. Second, an efficient and robust score should be used for estimating the feasibility of each protein−ligand mode throughout the docking procedure. If an inappropriate score is adopted, promising modes might be abandoned during docking or incorrect modes might be given higher ranking in the final step.

Many automated docking algorithms have been developed,[1−8] but none of them fully meets the above requirements. However, successful examples of application to virtual screening have been reported for DOCK, the first automated docking method developed by Kuntz and co-workers.[9−13] In the first successful application of DOCK to drug discovery, haloperidol was identified as a novel inhibitor of HIV-1 protease with a K_i of 100 μM.[9] Thereafter, successful drug discoveries using DOCK have been reported for many target macromolecules, e.g. thymidylate synthase,[11] protein tyrosine phosphatase-1B,[12] and protein kinase CK2.[13] Recently, active compounds have also been discovered by using other docking methods.[14−18] For example, novel FKBP inhibitors have been discovered by the program SANDOCK, which has a modified algorithm from that of DOCK,[14] and new thyroid hormone receptor antagonists have been identified by the program ICM.[16] Pang and co-workers have discovered farnesyl transferase inhibitor leads, one of which was also active in ex vivo assay, by using their program EUDOC, and they have validated their computational screening by means of a control study on randomly selected compounds.[15] These successes have been noteworthy, but we consider that a method and strategy that can discover, with higher probability, active compounds with a wide range of structural frameworks are still needed.

We have already developed a unique automated docking method for proteins and flexible ligands.[19−21] The method, named ADAM, is able to construct energetically favorable docking models, considering conformational flexibility of the ligand and intermolecular interaction between ligand and protein. ADAM starts the docking calculation by using the character of the H-bonding pattern, which provides specificity of direction and distance. The position, orientation, and conformation of H-bonding part of ligand are searched effectively and rapidly, by comparing the distances of ligand H-bonding heteroatoms with those of protein H-bonding sites. Then, the conformational space of the ligand is fully and continuously explored by combining systematic generation of conformers with appropriate structural optimization techniques. In most other docking methods, only discrete points can be sampled in the ligand conformational space, or part of the conformational space is explored by means of stochastic approaches (e.g., Monte Carlo methods, genetic algorithms, etc.). Using such methods, it is difficult to always reach correct docking modes, except for the straightforward case that the ligand in a crystallographic complex structure is docked to the protein structure in the same complex. As for ADAM, the merits of the method are high accuracy of the results, fully automatic generation of docking modes and short computational time. The efficiency and accuracy of ADAM docking have been confirmed in several protein systems.[21]

Using the algorithm of ADAM, we have developed a new virtual screening procedure for discovering flexible ligands from 3D databases. It is called ADAM&EVE. We have modified the ADAM algorithm to perform the docking calculation successively on compounds included in 3D databases, and to select promising ligand candidates based on various criteria involving inter- and intramolecular energy values, the numbers of H-bonds of docking models and characteristics of compounds. Several parameters were tuned up to accelerate the docking calculation. Furthermore, we have developed a new tool for automatically preparing the additional data necessary for flexible docking (e.g., various atom types, atomic charges, settings of bond rotation, etc.) for each compound in the 3D databases, prior to the database screening calculation.

ADAM&EVE has been applied to a dozen drug targets, including enzymes and receptors. One successful example is the discovery of novel aldose reductase inhibitors; this has already been reported, but the details of our method were not presented in that paper.[22] Here, we describe our ADAM&EVE procedure, together with another successful application to finding novel potent acetylcholinesterase (AChE) inhibitors with various core structures.

Methods

ADAM&EVE Procedure. Prior to the description of our virtual screening procedure, we will briefly summarize the docking algorithm of ADAM, on which the major search routine of ADAM&EVE is based. The details of this algorithm have already been reported, together with confirmation of its usefulness in several test cases.[19−21]

In the first step, the ADAM algorithm makes efficient use of the H-bonding pattern to rapidly obtain probable binding modes (i.e., positions and orientations) together with probable ligand conformations. In predicting the H-bonding schemes, we use H-bonding dummy atoms whose coordinates are automatically determined. The meaning of an H-bonding dummy atom is that if a heteroatom in the ligand is placed near a dummy atom, an H-bond is expected to be formed between the ligand and the protein. To determine the positions of dummy atoms, the H-bonding regions, i.e., the positions of ligand heteroatoms that can form H-bonds with protein functional groups, are calculated first, and then dummy atoms are placed at positions (e.g., the centers) inside the regions. Note that the dummy atoms are not placed at the positions of H-bonding functional groups or heteroatoms of the protein.

For each combination set of correspondences between dummy atoms and ligand heteroatoms, the possibility

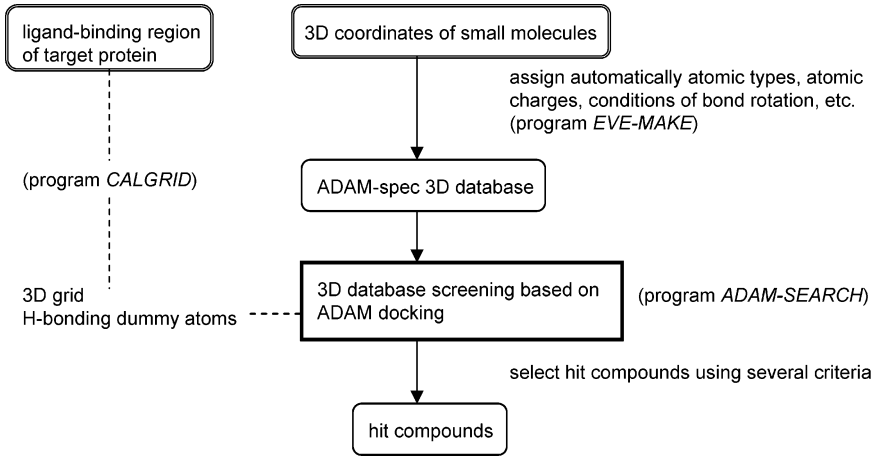


Figure 1. Flowchart of our virtual screening procedure ADAM&EVE.

of simultaneous formation of a given number of H-bonds is successively examined, by comparing the distance relation of dummy atoms and ligand heteroatoms. At this stage, all rotatable bonds in the ligand H-bonding part, i.e., the partial structure including all heteroatoms in each H-bonding combination set, are systematically rotated. All possible binding modes can be covered by examining all possible combination of H-bonding pairs. The conformers such that distance relations between dummy atoms and the corresponding ligand heteroatoms match well are selected, and the distance relation is optimized by changing the conformation of the ligand H-bonding part.

The selected conformers of the ligand H-bonding part with probable H-bonding schemes are placed in the protein cavity so as to fit the positions of ligand heteroatoms to those of the corresponding dummy atoms. Then, the interaction energy of the H-bonding part for each conformer is estimated by using a 3D grid that is calculated in advance inside the user-defined region of the protein, and the intramolecular energy of the H-bonding part is also calculated. The conformers with lower total energies are subjected to energy minimization.

For each likely structure of the H-bonding part, conformations in the remaining part are systematically explored. Energy minimizations of low-energy models are carried out, and several to several dozen stable docking models are output from the ADAM program.

In recent years, we have modified the ADAM algorithm to perform high-throughput virtual screening, and the method is called ADAM&EVE. The flowchart of ADAM&EVE is shown in Figure 1.

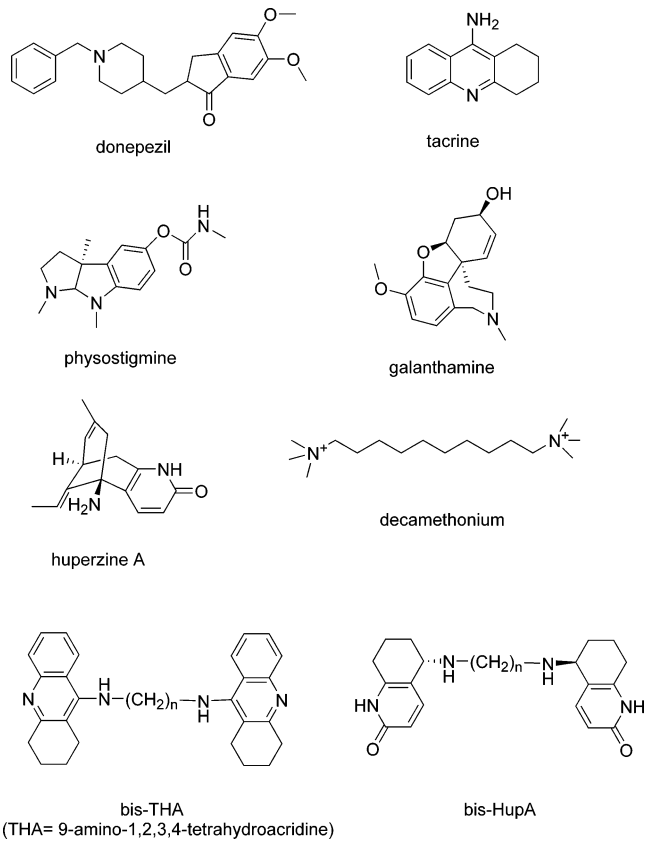
First, the allowed region for ligand-binding is indicated by the user. Inside this region, a 3D grid is generated, and various potential values, including vdW and electrostatic potentials, are stored at each grid point. Potential values are estimated by using the Cornell et al. AMBER force field.[23] At the same time, the H-bonding dummy atoms are placed in the ligand-binding region. Generation of the grid and dummy atoms is performed automatically by the program CALGRID.[24,25]

For the compounds to be screened, the 3D coordinates and connection tables of non-hydrogen atoms are taken from the various databases including the structures of commercial or corporate compounds, or crystal structures, etc. For efficient use of the ADAM algorithm, the following data need to be added for each compound: the 3D coordinates of hydrogen atoms; H-bonding type of each atom for predicting H-bonding schemes; atomic charge and AMBER force field atomic type of each atom for estimating intra- and intermolecular energies and minimizing them; the condition of bond rotation for considering the conformational flexibility of the molecules. We have developed a fast method called EVE-MAKE that automatically adds the above information for each compound in a rule-based fashion. The atomic charge calculation in the method is pursued based on the empirical method of Gasteiger and Marsili.[26] The definition of H-bonding types is common to the program CALGRID.[24,25] In setting the condition of bond rotation, the angle range of systematic rotation for each bond is determined taking account of 2-fold or 3-fold rotation symmetry. For a bond whose dihedral angle should be planar, such as a double bond or ester, only 0° and 180° are allowed for conformer generation. The torsion angle interval for general bond rotation can be defined by the user. Thus, an "ADAM-spec 3D database" for ADAM&EVE can be easily constructed.

Then, using the modified ADAM algorithm, named ADAM-SEARCH, the 3D database screening calculation is performed. In the docking calculation, only one ligand and one protein are managed, but in the 3D database screening, dockings of tens or hundreds of thousands of molecules to one target protein should be performed successively. So, for development of ADAM-SEARCH, the algorithm of ADAM was modified to execute docking of many molecules successively, and several parameters were tuned up to accelerate the calculation. The most time-consuming steps in our docking algorithm are several optimization processes, especially those for fitting the compound structure to the 3D grid in the protein cavity. To accelerate this optimization step, we decided to adopt the Powell method, instead of the simplex method used so far.[27] The former algorithm is much faster than the latter, and significantly speeds up the ADAM method (unpublished data). Furthermore, a looser condition for convergence is used in each optimization process of ADAM-SEARCH, and this also shortens the calculation time.

In addition, the user can determine the hit criteria for interaction energy, the number of H-bonds, etc., of the most stable docking model obtained for each com-

Chart 1. Chemical Structures of Known Noncovalent AChE Inhibitors



pound. Molecular weight, the number of atoms, heteroatoms, and ring structures, and the sorts of functional groups included in the compound, etc., may also be used as hit criteria. On the basis of these criteria, the promising ligand candidates are selected, and the 3D coordinates of docking models are output at the same time.

Application to Acetylcholinesterase. ADAM&EVE has been applied to AChE, which is considered to be an important target for treatment of Alzheimer's disease.[28−30] AChE inhibitors such as donepezil and tacrine have been used clinically.[31,32] In Chart 1, well-known noncovalent inhibitors of AChE are shown. We hoped to discover novel noncovalent inhibitors with different core structures from those of the known ligands, by searching 3D databases of commercially available compounds.

Preparation of Target Protein Structure. The X-ray crystal structure of electric eel AChE, which was used in the inhibition assays of this study, had not been solved, and even the sequence had not been reported when our study started. So, we decided to use the crystal structure of *Torpedo californica* AChE, even though the homology of the two sequences was unknown. Several crystal structures of *T. californica* AChE complexed with various ligands have been solved, and a major conformational difference between them is seen in the orientation of the side chain of Phe330.[33−35] The complex structure with the inhibitor decamethonium[33] has a wider ligand-binding cavity than other crystal structures (the complex structure with the strong inhibitor donepezil[35] was unavailable at the start of this study), owing to the orientation of the phenyl ring of Phe330,

Table 1. Dummy Atoms Used in the 3D Database Screening of AChE

no.	H-bonding atom of AChE[a]	H-bonding character[b]
1	Ser81 amide O	A
2	Trp84 amide O	A
3	Asn85 carboxamide Oδ1	A
4	Glu199 carboxylate Oϵ1	A
5	Ser200 hydroxyl Oγ	B
6	His440 amide O	A

[a] From the H-bonding heteroatoms in this column, H-bonding dummy atoms were generated. [b] The H-bonding characters assigned to the dummy atoms are hydrogen donor (D), hydrogen acceptor (A) or both (B).

and we expected that a larger number of new active compounds would be discovered by using the decamethonium complex structure. So, the 3D atomic coordinates of *T. californica* AChE were taken from the crystal structure of the complex with decamethonium, which has been deposited in the Protein Data Bank (PDB)[36] as 1ACL. The crystallographic resolution is 2.8 Å.

The inhibitor and all water molecules were removed from the structure. Atomic charges were assigned to the protein atoms according to the default values of the program AMBER 5.0, by using our program PDBFIL that performs preprocessing of the protein coordinate file.[24,25] The protonation state of His in the catalytic triad was properly determined. AChE has a deep cavity containing the catalytic triad at the bottom (it is called the "active site gorge"),[37] and we indicated the whole range of the active site gorge as the allowed region for ligand-binding.

Inside the ligand-binding region, a 3D grid with a regular interval of 0.4 Å and 42 H-bonding dummy atoms were generated by the program CALGRID. Among those dummy atoms, six located at the bottom of cavity were selected (Table 1). In the 3D database screening, ligand candidates that could bind to these essential sites were sought.

For these database screening, another 3D grid was prepared. It is possible that active compounds of large size or quite different shape cannot be docked computationally to the crystal structure of protein, owing to the "induced-fit" problem. Because the protein crystal conformation should adapt to the bound ligand structure, it is desirable to handle the conformational flexibility of protein explicitly in the docking calculation. But our docking algorithm assumes that the protein structure is rigid, so an alternative strategy is required. We developed the "vdW-offset" grid as an approach to solve the problem, at least in part. In the "vdW-offset" grid, the vdW energy curve for each atom pair is shifted to an extent that the user sets (Figure 2). The same offset value is applied to all protein atoms inside the ligand-binding region. As a result, the vdW potential surface is moved to widen the protein cavity, and bumping of the protein cavity and compound atoms may be avoided to some degree. In the virtual screening of AChE, a "vdW-offset" grid with a 0.5 Å offset value was prepared. The same dummy atoms as in the normal grid were adopted in the virtual screening using the "vdW-offset" grid.

Preparation of 3D Databases for ADAM&EVE. The basic data for small molecules to be searched were taken from the Available Chemicals Directory (ACD; MDL Information Systems) and MAYBRIDGE catalog

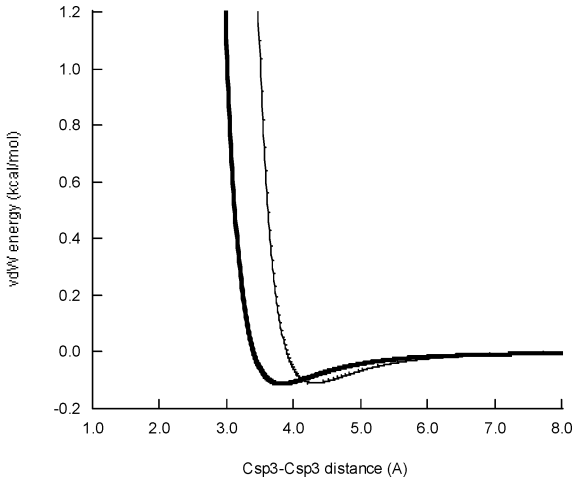


Figure 2. The vdW energy curves for Csp3−Csp3 atom pair. Normal (thin line) and 0.5 Å "vdW-offset" (bold line) conditions are compared.

database. ACD contained about 110 000 commercially available compounds and MAYBRIDGE had about 47 000 compounds. The original ACD database has only information on the 2D chemical structure of each compound, so the ACD-3D database that includes the 3D atomic coordinates was used. The 3D coordinates for MAYBRIDGE compounds were prepared by use of the program CONVERTER (Accelrys, Inc.).

The 3D coordinates and connection tables of compounds were taken from the ACD and MAYBRIDGE databases, and various data necessary for the effective use of ADAM were added automatically by the program EVE-MAKE, to construct the ADAM-spec 3D databases. The atomic charges were calculated by applying the Gasteiger−Marsili method, AMBER force field atomic types, and H-bonding types were automatically assigned, and information on functional groups was stored in the databases. As for the interval of torsion angle of each rotatable bond, the value of 120° was used for the step of generating conformers of H-bonding moieties. This rather large value (twice as large as the regular angle step in automated docking) was necessary to accelerate the docking calculation of each compound, but the accuracy of the resultant docking modes should not be greatly impaired, since the convergence range of the optimization used in this step is large. The torsion angle intervals in the non-H-bonding part were set to the same values as the defaults of ADAM docking, i.e., first the large angle interval 120° is used and then the much smaller value of 15°.

In generating the ADAM-spec 3D database, compounds including elements other than H, C, N, O, S, P, and halogens were excluded, because the parameters for such molecules are not included in the AMBER program.

Parameters of Database Search and Hit Criteria. The user-adjustable parameters in the ADAM-SEARCH calculation are common to those in the docking program ADAM. The definitions of docking parameters were described in our paper on ADAM.[21] The parameters used in this virtual screening were as follows. The threshold of function F, which is a matching score of the distance relation between dummy atoms and compound heteroatoms (F_{thres}), was 0.8. The threshold of the total potential energy used for selection of intermediate models in the docking process (E_{thres}) was 3000 kcal/mol. The intermolecular energy E_{inter} was used as a score function in the docking process and in selecting hit compounds. The E_{inter} is given as follows:

$$E_{\text{inter}} = E_{\text{vdw}} + E_{\text{elc}} - 2.5N_{\text{hb}}$$

The E_{vdw} and E_{elc} are intermolecular vdW and electrostatic energies calculated by using the 3D grid, and N_{hb} is the number of H-bonds.

Table 2. Hit Criteria for This Database Screening Calculation

criteria for characteristics of each compound
molecular weight ≥ 300
no. of atoms ≥ 40
no. of heteroatoms ≥ 2, ≤ 10
no. of ring structures ≥ 2
functional groups without carboxylate, phosphate, sulfate
criteria for most stable docking model obtained for each compound
intermolecular energy < −20.0 kcal/mol
intramolecular energy of compound < 50.0 kcal/mol
intermolecular H-bonds ≥1

The hit criteria of the screening are shown in Table 2. Small, very hydrophilic, or very hydrophobic compounds were excluded. To focus on druglike structures, a criterion for the number of ring structures was applied. Furthermore, compounds with anionic functional groups were excluded, because it has been suggested that the active site gorge is electrostatically negative, based on experimental and theoretical studies.[38,39] As regards the energy values of the most stable docking model of each compound, criteria for the inter- and intramolecular energies and the number of H-bonds were set.

Results and Discussion

3D Database Screening Calculation. First, database screening calculation using the normal 3D grid was performed, and a total of 640 hit compounds was obtained, together with 3D coordinates of docking models for each. Next, the hit compounds obtained with the normal grid were excluded from the 3D databases, and a second calculation using the "vdW-offset" grid was executed. This provided 911 hits. In this two-step screening, a total of 1551 compounds was obtained, so about 1% of the databases was hit.

The average computational time for docking of each molecule was about 6 s on the Linux machine (Pentium III, 1 GHz) that we currently use (at the time of this study, only computers with much slower CPU speeds were available). The speed of our docking calculation is considered satisfactory for application to 3D databases including a number of compounds, although the needed computational time will change according to the conditions of calculation and characteristics of the protein cavity.

The core structures of hit compounds were various, and most of them had quite different structures from known inhibitors. The docking modes of the hit compounds were observed using our in-house 3D graphics tool GREEN,[24,25] and the fitness of shape between each hit compound and the protein, H-bonding schemes,

Table 3. AChE-Inhibitory Activities of Hit Compounds from ADAM&EVE Virtual Screening

	no. of compounds	
IC_{50}[a] (μM)	normal grid	vdW-offset grid
0.5−1.0	0	3
1.0 − 5.0	0	6
5.0−10.0	1	3
10.0−50.0	4	13
50.0−100.0	3	2

[a] IC_{50} value of physostigmine (positive control) was 0.36 μM.

hydrophobic interaction, and compound conformations were checked. The hit compounds that could bind conformably to the target protein with rather extended conformation along the active site gorge were selected, but when there were several analogous compounds, only one or two representatives were selected. Among these selected compounds, a total of 114 available compounds (55 from the normal grid calculation and 59 from the "vdW-offset" grid calculation) could be purchased.

Inhibitory Activities and Docking Modes of Hit Compounds. The inhibitory activities of the 114 purchased compounds toward electric eel AChE were measured. A total of 35 compounds showed inhibitory activities with IC_{50} values of less than 100 μM (Table 3). Thus, among the compounds hit by our computational screening against AChE, more than 30% of those assayed were actually active, and this hit rate is quite high compared with that of experimental high-throughput screening or most other virtual screening approaches.[9−18,40] Furthermore, several potent inhibitors were included in the active compounds obtained from this study. A total of 13 compounds had IC_{50} values better than 9 μM, and among them, three compounds had IC_{50} values below 1 μM. The most potent compound had an IC_{50} of 0.59 μM. The IC_{50} of the well-known inhibitor physostigmine used as positive control was 0.36 μM, so some of the novel inhibitors discovered by our method have activities comparable to that of a clinically used inhibitor. We think these results are very promising, because there have been very few cases where so many active compounds, including ones active at submicromolar concentration, have been discovered by using only structure-based virtual screening.

The chemical structures of hit compounds that had IC_{50} values below 30 μM are shown in Figure 3. The novel inhibitors discovered by ADAM&EVE had various chemical frameworks, quite different from those of known ligands, except that compound **3** has an analogous core structure to donepezil. Compound **1** should be a dication when it binds to AChE, and its characteristic of bivalency is common to the known dimeric inhibitors, such as decamethonium, bis-THA,[41] and bis-HupA,[42] although its structural framework is different from them. These compounds are sufficiently active to be useful as lead compounds for drug development, so our results have provided very useful clues to a variety of frameworks for candidate new drugs to treat Alzheimer's disease.

As can be seen in Figure 3, all of the novel inhibitors discovered here are flexible compounds with several rotatable bonds. In many of them, the ligand conformations in the docking models were quite different from the input ones. We show a striking example in Figure 4 and Table 4. In this example, we performed the docking calculation again, starting from the 3D structure generated by our program KEY3D, which can reproduce bond lengths and angles very similar to those in the crystal structure in Cambridge Structural Database.[43] KEY3D is able to construct an extended and stable conformation for each compound, and such structures are suitable for demonstrating the features of our docking method. It can be seen that the input conformation of compound **4** (Figure 4a) is very different from that in the docking model (Figure 4b). As can be seen in Table 4, all of the torsion angles were greatly changed through the automated docking, and compound **4** could not have been properly docked to the protein cavity if it had been kept in the input conformation. We think that most of the new inhibitors discovered in this study would not have been hit, if we had used a virtual screening technique based on rigid docking. Furthermore, even if some active compounds had been hit by rigid-docking-based screening, the docking modes with the input ligand conformations might be incorrect, so that the binding energies would not be estimated properly. We considered that the flexible-docking-based algorithm of ADAM&EVE contributed greatly to our success in discovering many AChE inhibitors.

Several inhibitors (i.e., compounds **2**, **14**, and **15**) have only a few H-bonding heteroatoms. Our docking algorithm uses the geometrical relation of H-bonds as an initial clue to search intermediate docking modes at the early stage of the procedure, but matching of H-bonding pattern is not the only factor that contributes to the determination of final docking modes. In the latter part of the docking procedure, the total energy, including vdW and electrostatic terms, is also estimated to construct highly stable docking models, so compounds with only a few H-bonding heteroatoms can be hit, as in the application to AChE.

From the docking models constructed during the ADAM&EVE virtual screening process, it was suggested that the binding modes of active compounds were as varied as their chemical frameworks. In the docking models, the novel inhibitors bound conformably to the active cavity, with a variety of H-bonding schemes and interaction sites. As an example, we present the docking modes of three novel inhibitors, **1**, **4**, and **8,** in Figure 5. The most potent inhibitor **1** was stored as its neutral form in the database and identified by the virtual screening procedure in an improper protonation state. We repeated the docking of compound **1** in its protonated state and confirmed that it also satisfies the hit criteria in its correct protonation state. The resulting docking mode is shown in Figure 5a. The problem of the protonation state will be discussed later. Compound **1** occupies a similar region to decamethonium, which was included in the complex crystal structure used as the basis of this virtual screening. Both compound **1** and decamethonium have 2-fold rotation symmetry, and the positions of the two protonated imidazoline rings of **1** are almost the same as those of the two quaternary amines of decamethonium. In the docking mode of compound **4**, the aminobenzene ring is placed at the position of the aromatic ring of donepezil, the complex crystal structure of which is available from PDB as 1EVE.[35] The aminobenzene ring is stacked against

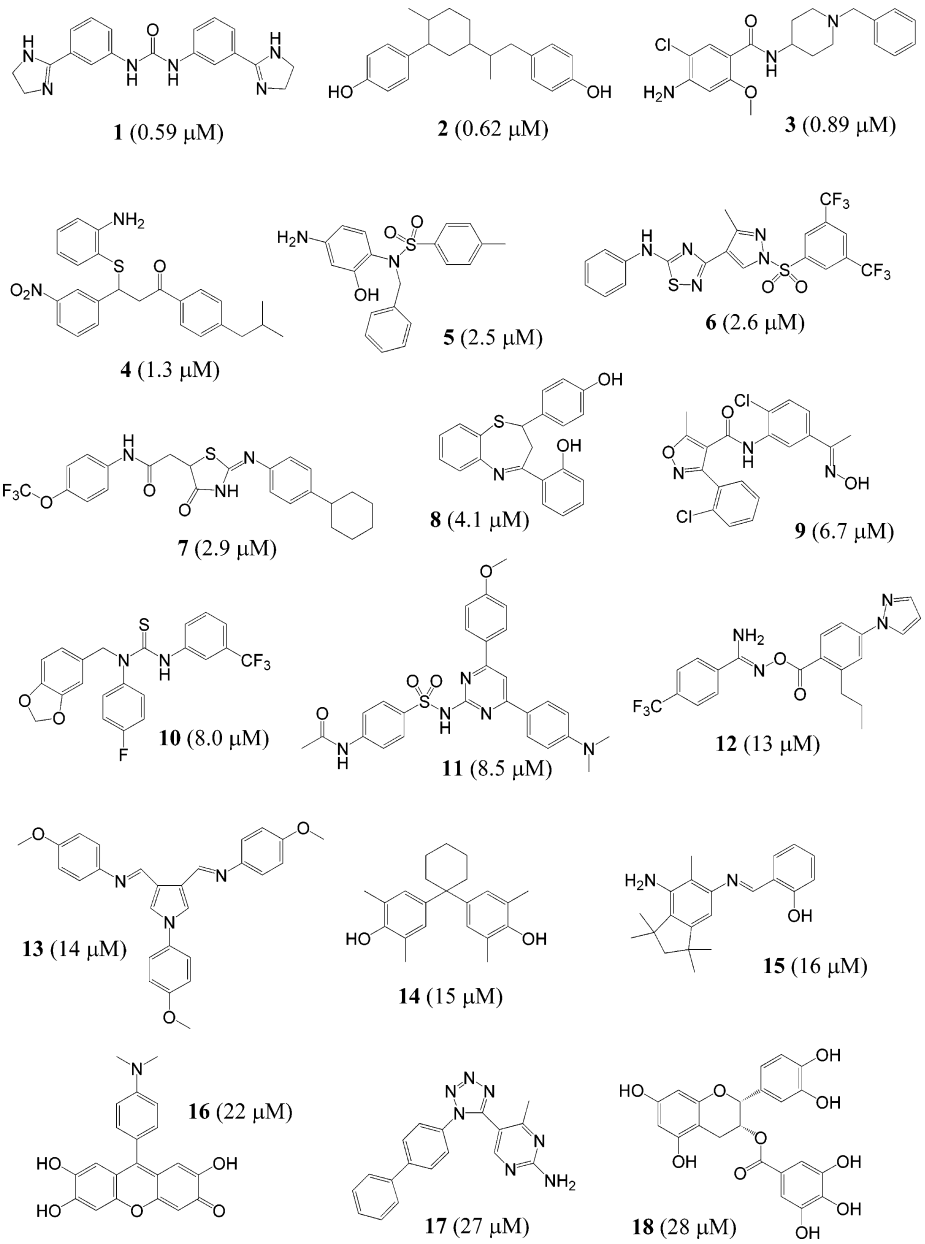


Figure 3. Chemical structures of hit compounds that show inhibitory activity toward AChE (IC_{50} < 30 μM) and their IC_{50} values.

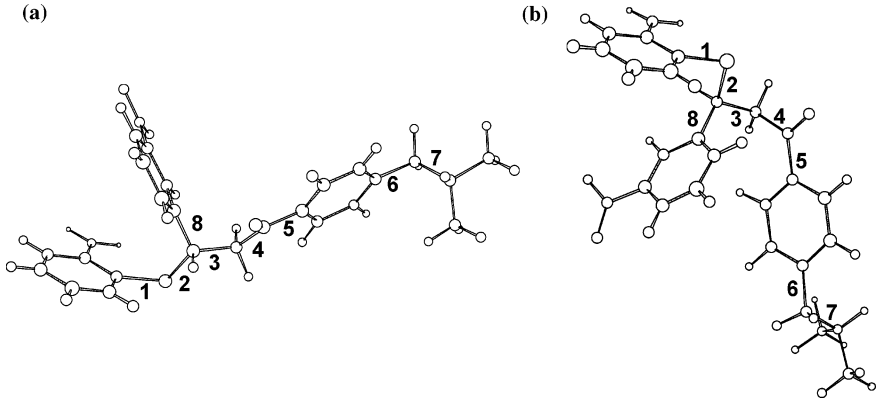


Figure 4. Comparison of input and docking conformations of compound **4**. (a) Unbound structure used in docking as an input structure. (b) Bound structure in the docking model. The directions of aminobenzene rings coincide in the two structures. The numbers on the rotatable bonds correspond to those of the torsion angles in Table 4.

Trp84, and an H-bond is formed between the amino group of **4** and amide O of His440 of AChE. Compound **4** occupies a part of the binding region of donepezil and a part of the binding site of the substrate, acetylcholine. As for compound **8**, one of the benzene rings is stacked against Trp84, and the other is placed at the benzene

Table 4. Comparison of the Input and Docking Conformations of Compound **4**

	torsion angles[a] (deg)							
	1	2	3	4	5	6	7	8
input structure	118.0	−173.7	−172.3	−162.7	−15.7	71.7	−172.5	64.6
docking model	−73.4	179.1	70.1	101.2	173.8	102.7	−76.7	−107.1

[a] The numbers of the torsion angles correspond to those of the rotatable bonds in Figure 4.

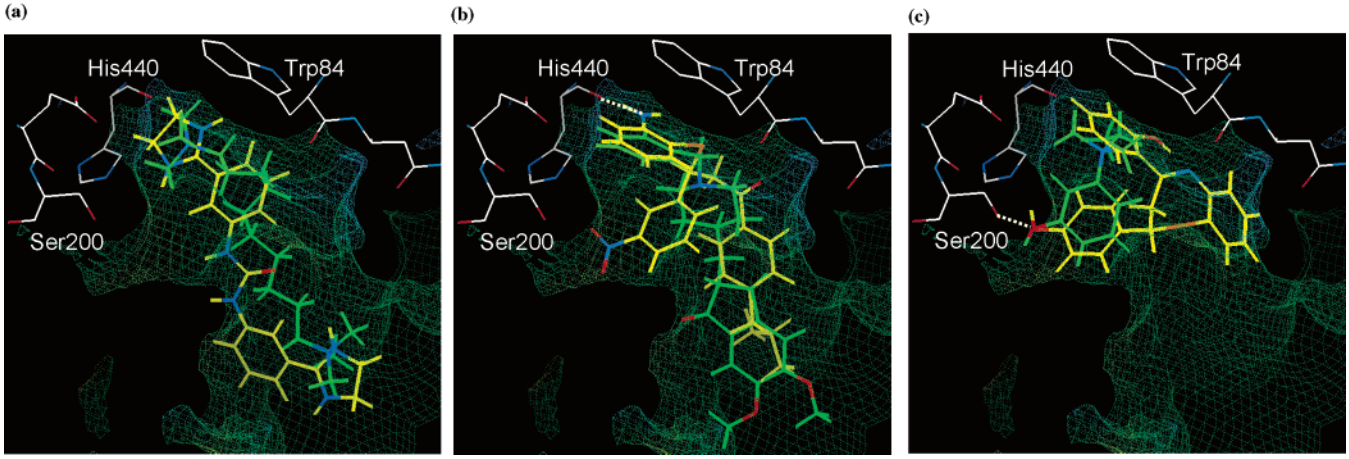


Figure 5. Docking models of **1**, **4**, and **8**, compared with complex crystal structures of known inhibitors: (a) **1** (carbons and hydrogens are yellow) and decamethonium (green), (b) **4** (yellow) and donepezil (green), (c) **8** (yellow) and edrophonium (green). The shape of ligand-binding region is displayed as a bird-cage model. Dotted lines represent intermolecular H-bonds. Hydrogens in the protein are omitted for clarity.

ring position of the known inhibitor edrophonium (PDB code is 2ACK).[33] An H-bond is formed between the hydroxyl group of **8** and Oγ of catalytic Ser200. Different from **1** and **4**, compound **8** occupies only the bottom region of the protein cavity, but shows good steric complementarity to the site. These docking modes seem reasonable, considering the binding modes in complex crystal structures of known inhibitors, but it should be noted that several different docking modes with closely similar energy values are also obtained by ADAM&EVE. We expect that, by modifying the structures based on these docking models and synthesizing the modified compounds, more potent inhibitors will be obtained.

It would be desirable to use an enzyme from the same source for both experimental assay and computational virtual screening, but the electric ray AChE used in virtual screening is not commercially available, so the electric eel enzyme was used instead. The 3D structure of eel AChE and even its sequence were unavailable at the start of this study, as noted above. The sequence of electric eel AChE is now known and has about 75% homology with that of *T. californica* AChE. All the residues exposed in the ligand-binding cavity are conserved in the two sequences, except that Phe330 in *T. californica* AChE is substituted by Tyr in electric eel AChE. We have performed another screen in which Phe330 was computationally changed to Tyr, while maintaining the conformation of the side chain as it was. The resulting hits mostly coincided those obtained with the native structure (data not shown). In the changed AChE structure, only a small H-bonding site was generated from the hydroxyl group of the substituted Tyr330, owing to the steric hindrance caused by the parallel alignment of the Tyr330 ring to the gorge axis. So, only a few compounds could form H-bonds with Tyr330 in their docking modes. However, it has been reported that the inhibitory potencies of some known ligands are significantly different between *T. californica* AChE and mammalian AChE with the Tyr substitution (Tyr337 in mouse and human) in place of Phe330 of *T. californica* AChE.[44] What caused this inconsistency? In the crystallographic study of mammalian AChE, it has been shown that the lower part of the gorge is much narrower than in *T. californica* AChE, owing to the shifted positions of not only Tyr337, but also Phe338 and Tyr341 (Phe331 and Tyr334 in *T. californica* AChE).[45,46] Of course, the H-bonding site for Tyr337 is quite different from that for Tyr330 in the *T. californica* AChE model. So, we consider that more detailed modeling that can reproduce the shape and H-bonding sites of the gorge of eel AChE is necessary, although the structural similarity of eel AChE and mammalian AChE has not been reported. We think it likely that, if assay was performed using *T. californica* AChE, more novel inhibitors might be discovered.

The 3D database search calculation using the "vdW-offset" grid discovered many more inhibitors than that using the normal grid, as shown in Table 3. Among the compounds hit using the "vdW-offset" grid, 45.8% of the assayed compounds showed inhibitory activities with $IC_{50} < 100\ \mu M$, while among those hit using the normal grid only 14.5% were actually active. Moreover, all of the highly potent inhibitors discovered in this study were from the calculation using the "vdW-offset" grid. We think this is because flexibility of the protein structure is implicitly included in the "vdW-offset" grid. By using the "vdW-offset" grid, it should become possible to dock properly compounds with larger volumes to the ligand-binding site; such large compounds frequently occupy a substantial region of the protein cavity.

However, there is an apparent limitation of the "vdW-offset" grid. When a large ligand binds to the protein cavity, not the whole structure around the ligand-binding site of the protein, but only a part of the side (and main) chains will be movable. So, it is possible that false positives (nonactive compounds) would be increased by use of the "vdW-offset" grid, although false negatives should be reduced. Furthermore, if there is drastic induced-fit change of protein structure, some active compounds might not be docked properly even using the "vdW-offset" grid. To solve these problems in part, we have been developing a procedure in which the docking modes of hit compounds are subjected to energy minimization, including local movement of the ligand-binding site of the protein, followed by ranking by our in-house scoring system. Both false positives and false negatives should be reduced by means of this procedure, and test calculations performed on several target proteins have been very promising (the results will be reported elsewhere).

Ranking of hit compounds was not done directly on the basis of binding free energy. Instead, several criteria including AMBER force field energy and certain characteristics of compounds (e.g., size, hydrophobicity, functional groups, etc.) were used for selection of promising ligand candidates. The force field energy values calculated in this case were crude, owing to the approximation of the 3D grid and lack of explicit consideration of protein induced-fit. The empirical method of Gasteiger and Marsili for calculating the atomic charge is very fast and efficient, but the semiempirical method should be more suitable for AMBER-based energy calculation, as reported by Pang et al.[7] As for the problem of the protonation state of compounds, each compound was handled in the protonation state given in the original 2D database, because it was difficult to predict the correct protonation state in the bound form to the target protein. In this study, compounds **1** and **3** were supposed to be in the neutral state, which seems to be improper, although they were identified by ADAM&EVE since they satisfied all the hit criteria despite underestimated electrostatic energies. In addition to the above problems, many other important factors were ignored in this energy calculation, for example, hydrophobic interaction, desolvation costs, and the changes of compound intramolecular energies in going from the unbound form and to the bound form to the target protein. In view of the difficulty of estimating precisely the stability of each docking model, the strategy adopted in this virtual screening, where hit compounds that met all of our criteria were treated equally in the step of further visual investigation, is considered be practical and effective. However, we cannot rule out the possibility that some active molecules might have been incorrectly eliminated in this study, owing to such crude estimation.

Our goal is to discover as many active compounds as possible from the 3D databases, with minimum false negatives and false positives, to provide medicinal chemists with the maximum number of promising leads, which could be developed into effective medicines. Although there is a great distance to that goal, we are continuing our efforts to overcome the difficulties. Recently, we have been developing a new, efficient score for more properly selecting promising ligand candidates. The score includes terms taking account of the desolvation costs, lost degree-of-freedom of the ligand, and change of intramolecular energy caused by complex formation. The docking modes of each compound in the 3D databases are estimated and ranked according to this score, after energy minimization, including a consideration of the local motion of the ligand-binding site. Furthermore, we have been improving our docking method to predict the protonation state of each compound and the protein side chain during the process of constructing the docking models. In addition, our new method KEY3D, which has been developed for generating 3D structures of compounds from their 2D chemical structures, can provide suitable structures for our docking procedure. KEY3D is also able to assign new force field atomic types (KMF types), which can be applied to all general compound structures, and to calculate MOPAC atomic charges for given molecules. Currently, we use the KEY3D program for generating 3D structural databases, and KMF types and MOPAC charges are utilized in the ADAM&EVE docking process and score estimation. These improvements will be reported elsewhere. If the above-mentioned developments are fully adopted in our virtual screening procedure, larger numbers of novel ligands with higher activities should be obtainable with less effort.

Conclusions

We have developed an efficient virtual screening procedure, ADAM&EVE, for discovering promising ligand candidates from 3D databases, using hit criteria based on the interaction with the target protein. One of advantages of our method is that docking is performed by covering all possible positions, orientations, and conformations of each compound effectively and rapidly. Our method can screen a large number of flexible compounds in a practical time and also provides the docking modes.

As one of our successful applications, high-throughput computational screening for discovering potent noncovalent AChE inhibitors is presented in this report. In this work, 35 novel inhibitors that have quite different core structures from known inhibitors were identified from 3D databases including commercially available compounds. Among them, 13 compounds showed significant inhibitory activity, and the most potent one had an IC_{50} value of 0.59 μM. These results confirm the efficiency and validity of our method. We have been developing a new score system for ranking hit compounds, and adoption of this in our virtual screening procedure should allow the discovery of novel, highly active ligands with less effort.

Experimental Section

Acetylcholinesterase Inhibition. The spectrophotometric method described in refs 47 and 48 was followed. The assay solution consisted of 20 mM phosphate buffer, pH = 8.0. First, 100 μL of 0.057 unit/mL AChE from electric eel (Sigma Chemical Co.) was added to each well of 96-well plates, and 10 μL of each sample solution was added. The solutions were incubated at 37 °C for 15 min. Then, 100 μL of 909 μM 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB) solution and 10 μL of

455 μM acetylthiocholine iodide solution were added, and the final solutions were incubated at 37 °C for 30 min. The absorbance at 405 nm was detected by a spectrophotometer, and the percent inhibition due to the presence of sample compound was calculated. Several concentrations of the compounds were assayed, and IC_{50} values were determined. Each assay was carried out in duplicate and the average value was calculated.

Acknowledgment. We wish to thank Dr. Kazuhiko Otoguro and Dr. Satoshi Omura of The Kitasato Institute for measuring the AChE-inhibitory activities.

References

(1) 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.
(2) Leach, A. R.; Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* **1992**, *13*, 730−748.
(3) Goodsell, D. S.; Olson, A. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195−202.
(4) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM − a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488−506.
(5) 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.
(6) 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.
(7) Pang, Y.-P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: A computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem*., **2001**, *22*, 1750−1771.
(8) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375−382.
(9) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz de Montellano, P. R.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644−6648.
(10) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078−1082.
(11) Shoichet, B. K.; Stroud, R. M.; Santi, D. V.; Kuntz, I. D.; Perry, K. M. Structure-based discovery of inhibitors of thymidylate synthase. *Science* **1993**, *259*, 1445−1450.
(12) Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213−2221.
(13) Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* **2003**, *46*, 2656−2662.
(14) Burkhard, P.; Hommel, U.; Sanner, M.; Walkinshaw, M. D. The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J. Mol. Biol.* **1999**, *287*, 853−858.
(15) Perola, E.; Xu, K.; Kollmeyer, T. M.; Kaufmann, S. H.; Prendergast, F. G.; Pang, Y.-P. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.* **2000**, *43* 401−408.
(16) Schapira, M.; Raaka, B. M.; Das, S.; Fan, L.; Totrov, M.; Zhou, Z.; Wilson, S. R.; Abagyan, R.; Samuels, H. H. Discovery of diverse thyroid hormone receptor anatagonists by high-throughput docking. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7354−7359.
(17) Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J. Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, *6*, 439−446.
(18) Schneider, G.; Böhm, H.-J. Virtual screening and fast automated docking methods. *Drug Discovery Today* **2002**, *7*, 64−70.
(19) Yamada, M.; Itai, A. Development of an efficient automated docking method. *Chem. Pharm. Bull.* **1993**, *41*, 1200−1202.
(20) Yamada, M.; Itai, A. Application and evaluation of the automated docking method. *Chem. Pharm. Bull.* **1993**, *41*, 1203−1205.
(21) Yamada Mizutani, M.; Tomioka, N.; Itai, A. Rational automatic search method for stable docking models of protein and ligand. *J. Mol. Biol.* **1994**, *243*, 310−326.
(22) Iwata, Y.; Arisawa, M.; Hamada, R.; Kita, Y.; Mizutani, M. Y.; Tomioka, N.; Itai, A.; Miyamoto, S. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J. Med. Chem.* **2001**, *44*, 1718−1728.
(23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(24) Tomioka, N.; Itai, A.; Iitaka, Y. A method for fast energy estimation and visualization of protein−ligand interaction. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 197−210.
(25) Tomioka, N.; Itai, A. GREEN: a program package for docking studies in rational drug design. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 347−366.
(26) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(27) Nelder, J. A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308−313.
(28) Bartus, R. T.; Dean III, R. L.; Beer, B.; Lippa, A. S. The cholinergic hypothesis of geriatric memory dysfunction. *Science* **1982**, *217*, 408−417.
(29) Coyle, J. T.; Price, D. L.; Delong, M. R. Alzheimer's disease: a disorder of cortical cholinergic innervation. *Science* **1983**, *219*, 1184−1190.
(30) Benzi, G.; Moretti, A. Is there a rationale for the use of acetylcholinesterase inhibitors in the therapy of Alzheimer's disease? *Eur. J. Pharmacol.* **1998**, *346*, 1−13.
(31) Eagger, S. A.; Levy, R.; Sahakian, B. J. Tacrine in Alzheimer's disease. *Lancet* **1991**, *337*, 989−992.
(32) Kawakami, Y.; Inoue, A.; Kawai, T.; Wakita, M.; Sugimoto, H.; Hopfinger, A. J. The rationale for E2020 as a potent acetylcholinesterase inhibitor. *Bioorg. Med. Chem.* **1996**, *4*, 1429−1446.
(33) Harel M.; Schalk I.; Ehret-Sabatier L.; Bouet F.; Goeldner M.; Hirth C.; Axelsen P. H.; Silman I.; Sussman J. L. Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9031−9035.
(34) Raves, M. L.; Harel, M.; Pang, Y.-P.; Silman, I.; Kozikowski, A. P.; Sussman, J. L. Structure of acetylcholinesterase complexed with the nootropic alkaloid, (−)-Huperzine A. *Nat. Struct. Biol.* **1997**, *4*, 57−63.
(35) Kryger, G.; Silman, I.; Sussman, J. L. Structure of acetylcholinesterase complexed with E2020 (Aricept): implications for the design of new anti-Alzheimer drugs. *Structure* **1999**, *7*, 297−307.
(36) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(37) Sussman, J. L; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman. Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein. *Science* **1991**, *253*, 872−879.
(38) Ripoll, D. R.; Faerman, C. H.; Axelsen, P. H.; Silman, I.; Sussman, J. L. An electric mechanism for substrate guidance down the aromatic gorge of acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5128−5132.
(39) Nolte, H.-J.; Rosenberry, T. L.; Neumann, E. Effective charge on acetylcholinesterase active sites determined from the ionic strength dependence of association rate constants with cationic ligands. *Biochemistry* **1980**, *19*, 3705−3711.
(40) Paiva, A. M.; Vanderwall, D. E.; Blanchard, J. S.; Kozarich, J. W.; Williamson, J. M.; Kelly, T. M. Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. *Biochim. Biophys. Acta* **2001**, *1545*, 67−77.
(41) Pang, Y.-P.; Quiram, P.; Jelacic, T.; Hong, F.; Brimijoin, S. Highly potent, selective, and low cost bis-tetrahydroaminacrine inhibitors of acetylcholinesterase. *J. Biol. Chem.* **1996**, *271*, 23646−23649.
(42) Carlier, P. R.; Du, D.-M.; Han, Y.-F.; Liu, J.; Perola, E.; Williams, I. D.; Pang, Y.-P. Dimerization of an inactive fragment of Huperzine A produces a drug with twice the potency of the natural product. *Angew. Chem., Int. Ed.* **2000**, *39*, 1775−1777.
(43) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday: A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystallographic Data Center: computer-based search, retrieval, analysis and display of information. *Acta Crystallogr.* **1979**, *B35*, 2331−2339.
(44) Saxena, A.; Redman, A. M. G.; Jiang, X.; Lockridge, O.; Doctor, B. P. Differences in active site gorge dimensions of cholinesterases revealed by binding of inhibitors to human butyrylcholinesterase. *Biochemistry* **1997**, *36*, 14642−14651.

(45) Bourne, Y.; Taylor, P.; Bougis, P. E.; Marchot, P. Crystal structure of mouse acetylcholinesterase. *J. Biol. Chem*. **1999**, *274*, 2963−2970.
(46) Wong, D. M.; Greenblatt, H. M.; Dvir, H.; Carlier, P. R.; Han, Y.-F.; Pang, Y.-P.; Silman, I.; Sussman, J. L. Acetylcholinesterase complexed with bivalent ligands related to Huperzine A: Experimental evidence for species-dependent protein−ligand complementarity. *J. Am. Chem. Soc.* **2003**, *125*, 363−373.
(47) Ellman G. L.; Courtney K. D.; Andres V., Jr; Featherstone R. M. A new and rapid colorimetric determination of acetylcholinesterase activity. *Biochem. Pharmacol.* **1961**, *7*, 88−95.
(48) Doctor, B. P.; Toker, L.; Roth, E.; Silman, I. Microtiter assay for acetylcholinesterase. *Anal. Biochem*. **1987**, *166*, 399−403.

JM030605G