Pharmacophore-based virtual screening and density functional theory approach to identifying novel butyrylcholinesterase inhibitors

Sugunadevi SAKKIAH, Keun Woo LEE*

Division of Applied Life Science (BK21 Program), Systems and Synthetic Agrobiotech Center (SSAC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Research Institute of Natural Science (RINS), Gyeongsang National University (GNU), Jinju 660–701, Korea

Aim: To identify the critical chemical features, with reliable geometric constraints, that contributes to the inhibition of butyrylcholinesterase (BChE) function.

Methods: Ligand-based pharmacophore modeling was used to identify the critical chemical features of BChE inhibitors. The generated pharmacophore model was validated using various techniques, such as Fischer’s randomization method, test set, and decoy set. The best pharmacophore model was used as a query in virtual screening to identify novel scaffolds that inhibit BChE. Compounds selected by the best hypothesis in the virtual screening were tested for drug-like properties, and molecular docking study was applied to determine the optimal orientation of the hit compounds in the BChE active site. To find the reactivity of the hit compounds, frontier orbital analysis was carried out using density functional theory.

Results: Based on its correlation coefficient (0.96), root mean square (RMS) deviation (1.01), and total cost (105.72), the quantitative hypothesis Hypo1 consisting of 2 HBA, 1 Hy-Ali, and 1 Hy-Ar was selected as the best hypothesis. Thus, Hypo1 was used as a 3D query in virtual screening of the Maybridge and Chembridge databases. The hit compounds were filtered using ADME, Lipinski’s Rule of Five, and molecular docking to reduce the number of false positive results. Finally, 33 compounds were selected based on their critical interactions with the significant amino acids in BChE’s active site. To confirm the inhibitors’ potencies, the orbital energies, such as HOMO and LUMO, of the hit compounds and 7 training set compounds were calculated. Among the 33 hit compounds, 10 compounds with the highest HOMO values were selected, and this set was further culled to 5 compounds based on their energy gaps important for stability and energy transfer. From the overall results, 5 hit compounds were confirmed to be potential BChE inhibitors that satisfied all the pharmacophoric features in Hypo1.

Conclusion: This study pinpoints important chemical features with geometric constraints that contribute to the inhibition of BChE activity. Five compounds are selected as the best hit BchE-inhibitory compounds.

Keywords: cholinesterase; butyrylcholinesterase; Alzheimer’s disease; virtual screening; density functional theory; molecular docking; ligand-based pharmacophore modeling theory

Original Article

Introduction

Cholinesterases (ChEs) are involved in the degradation of choline and show similarity in protein sequence but differences in their kinetic properties. On the basis of their substrate and inhibitor specificities, cholinesterases are divided into two subfamilies: acetylcholinesterases (AChEs; EC 3.1.1.7) and butyrylcholinesterases (BChEs; EC 3.1.1.8). AChE is predominantly present in the central and peripheral nervous system, as well as in muscles. In muscles, AChE terminates impulse transmission by the rapid hydrolysis of acetylcholine to acetic acid and choline[1]. BChE is primarily synthesized in the liver and secreted into plasma, and it is responsible for the hydrolysis of a variety of choline (hydrophilic and hydrophobic) and non-choline esters[2]. BChE plays a key role in cholinergic synapses by terminating acetylcholine action, although the complete physiological function of BChE remains unclear[3]. Both cholinesterase enzymes belong to the super family of α/β-hydrolase fold proteins[4]. Both AChE and BChE exist as multimers of catalytic subunits in globular forms such as G1, G2, and G4 that contain one, two, or four subunits, respec-
tively. The hydrolysis of substrates by both enzymes proceeds through a transacylation step involving nucleophilic and general acid-base elements\textsuperscript{[3]}. \(\text{BChE}\) acts as a scavenger protein that protects the cholinergic system against anticholinesterase poisons. \(\text{BChE}\) is the sole carboxylesterase\textsuperscript{[6,7]} with recognized toxicological and pharmacological importance in scavenging and detoxification of numerous ester-containing drugs, prodrugs\textsuperscript{[8,9]}, and poisonous carbamyl- and phosphoryl-esters, including nerve agents\textsuperscript{[10,11]}.

Currently, \(\text{BChE}\) is emerging as an important pharmacological target in Alzheimer’s disease (AD) therapy\textsuperscript{[12]}. A 40\%–90\% increase in \(\text{BChE}\) expression and activity have been found in AD brain neuronal plaques\textsuperscript{[13]}. \(\text{BChE}\) is capable of compensating for reduced \(\text{AChE}\) catalytic functions in the synaptic cleft\textsuperscript{[14,15]} and shows significantly increased activity (30\%–60\%) during the time course of AD\textsuperscript{[16,17]}. Hence, in recent years, many scientists and researchers have shown keen interest in designing small molecules that can inhibit \(\text{BChE}\) activity\textsuperscript{[18–23]}.

However, there is also increasing evidence of \(\text{BChE}\)’s involvement in non-cholinergic functions such as cell differentiation\textsuperscript{[24]}, neurogenesis, and the formation of amyloid plaques in AD\textsuperscript{[25–27]}.

In this work, we used computer-aided drug design approaches to identify novel and potent inhibitors of \(\text{BChE}\). Pharmacophore studies are more cost-effective than experimental chemical screening of large databases. A 3D pharmacophore model was generated for \(\text{BChE}\) based on a series of well-known inhibitors. The best quantitative model was used as a 3D query for virtual screening of chemical databases to discover novel hit compounds. The virtual screening results revealed a small subset of database compounds that were promising potential hit compounds for \(\text{BChE}\) inhibition. The hits were subsequently filtered by Lipinski’s Rule of Five, ADME (absorption, distribution, metabolism, and excretion) properties, and molecular docking. Finally, density functional theory (DFT) was used to calculate the orbital energy value and energy gap for the molecules screened by docking.

**Computational methods**

Pharmacophore modeling is one of the most frequently used and valuable methods to discover novel scaffolds for various targets.

**Selection of compounds**

To construct the \(\text{BChE}\) data set, 71 compounds were collected with their corresponding reported inhibitory activity values (IC\textsubscript{50}) which were tested using the same bioassay technique from various publications\textsuperscript{[28–32]}. The \(\text{BChE}\) data set was divided into two sets: training and test sets that contained 26 and 45 compounds, respectively. The training set was prepared based on the following criteria: (i) a minimum of 16 diverse compounds were selected to avoid any chance correlation; (ii) the activity data should have a range of 4–5 orders of magnitude; (iii) the compounds should be selected to provide clear, concise information to avoid redundancy or bias in terms of both structural features and activity range; (iv) the most active compounds should be included so that they provide information on the most critical features required for a reliable/rational pharmacophore model; and (v) the inclusion of any compound known to be inactive due to steric hindrance must be avoided. The training set was used to build the quantitative hypothesis based on principles of structural diversity and IC\textsubscript{50} values that spanned a wide activity range, from 3.6 nmol/L to 11 000 nmol/L (Figure 1). The test set was used to evaluate the predictive ability of the generated pharmacophore model. Both the training and test set compounds were classified into three categories based on their activity values. The compounds with IC\textsubscript{50} values less than or equal to 100 nmol/L were considered to be highly active (+++), compounds with an activity range between 100 nmol/L and 10 000 nmol/L were considered to be moderately active (++), and compounds with IC\textsubscript{50} values greater than or equal to 10 000 nmol/L were set as low activity compounds (+). The 2D structures of the training and test set molecules were drawn using ChemSketch\textsuperscript{[24]} and the structures were converted into their corresponding 3D form using DS.

**Pharmacophore modeling**

Quantitative hypotheses were generated, and the best hypothesis was selected based on the models’ ability to predict the biological activity of novel compounds from various chemical databases using Discovery Studio v2.5.5 (DS, www.accelrys.com, San Diego, CA, USA). There are generally two methods to generate molecular conformation: FAST and BEST. The FAST algorithm only considers existing conformers and interrupts a search as soon as a pharmacophore matching conformation is found, whereas the BEST algorithm additionally “tweaks” bond distances, angles, and dihedral angles of pre-generated conformers on the fly to achieve the best matches. Herein, we used the BEST conformation method to generate multiple acceptable conformations for each compound present in the training and test sets with 20 kcal/mol as the energy cutoff\textsuperscript{[33]}. All default parameters were used to generate the pharmacophore, except the uncertainty default value (3.0) was changed to 2.0\textsuperscript{[34]}. The uncertainty is the ratio of the reported activity value relative to the minimum, and the maximum values must be greater than 1.0. The uncertainty value affects the categorization of ligands in the data set as either active or inactive compounds and is used during the constructive and subtractive phases. Here, an uncertainty value of 2.0 was more suitable for our data set because the compound activities spanned the requisite 4 orders of magnitude; this choice has been confirmed by evidence in the literature\textsuperscript{[35,36]}. The feature mapping/DS protocol was used to identify common features present in the active inhibitors of \(\text{BChE}\). This protocol computes a maximum of 1000 possible pharmacophore features mappings for the selected ligands. The selected features from the feature mapping were used as one of the key inputs for the 3D-QSAR pharmacophore generation module using a HypoGen algorithm. The HypoGen algorithm further estimates the activity of each training set compound by computing regres-
sion analysis using parameters such as the relationship of geometric fit value versus the negative logarithm of the activity. While generating the quantitative model, a minimum of 0 to a maximum of 5 features were selected to build a series of hypotheses. Ten quantitative pharmacophore models were generated with corresponding statistical parameters such as cost values, root mean square (RMS), and fit values. The best quality hypothesis was selected based on cost values as defined by Debnath’s methods [34].

Hypothesis validation
In general, pharmacophore models should be statistically significant, accurately predict the activity of molecules, and retrieve active compounds from databases. The best pharmacophore model was validated using various potent approaches such as Fischer’s randomization, test set, and decoy set [35].

The main purpose of validating a quantitative pharmacophore model is to determine its capacity to identify active compounds, as well as its predictive ability for corresponding molecules. Fischer’s randomization test was performed simultaneously during the original hypotheses generation and produced a number of random spreadsheets depending on the selected significance level (90%, 95%, 98%, and 99%) by shuffling the activity values present in the training set. Here, a 95% significance level was selected. Nineteen random spreadsheets were produced by randomly shuffling the activity value of the training set compounds, and the test generated hypotheses using the same chemical features and parameters used to develop the original hypothesis. Test and decoy sets were used to check whether the best hypothesis was able to select molecules with orders of magnitude of activity similar to that of the active training set and to determine how well the model hypothesis could differentiate potential BChE inhibitors from other compounds, respectively. The test set consisted of structurally diverse chemical compounds from the training set to ascertain the broadness of pharmacophore predictability. The decoy set was prepared by calculating the 1D property of 25 active inhibitors of BChE and 2075 inactive or unknown molecules. EF and GF were calculated using the following equations:

\[
EF = \frac{\text{number of correct hypothesized molecules}}{\text{total number of molecules}}
\]

\[
GF = \frac{\text{number of correct hypothesized molecules}}{\text{number of hypothesized molecules}}
\]
The main aim of the orbital energies calculation was to provide valuable information about the electrostatic properties of the BChE inhibitors. DFT is a successful and promising approach adopted by quantum chemists in the quantum mechanical simulation of periodic systems\[41\]. There is substantial evidence that DFT provides an accurate description of the electronic and structural properties of small molecules by computing the electronic structure of matter. The selected docked poses of the hit compounds from the molecular docking studies were used as input for the DFT instead of the compounds’ bioactive conformations. Because the docking results showed the suitable binding orientation of hit compounds, it was suitable for calculating the orbital energies such as HOMO and LUMO using DS. Calculating the orbital energy using B3LYP provided information regarding the capacity of the molecules to transfer their energies from a HOMO, which can act as an electron donor, to a LUMO, which can act as an electron acceptor. These electrostatic property calculations could provide useful information for designing novel BChE inhibitors.

Results and discussion

A ligand-based pharmacophore method was used to elucidate the spatial arrangement of chemical features that were crucial for the interaction of structurally diverse and potent BChE inhibitors with their target protein. Ligand-based approaches reveal the important and common chemical features of diverse ligands, and these features can then be used as 3D query in virtual screenings of large chemical databases to identify novel hit compounds.
Pharmacophore model

The HypoGen algorithm was used to construct quantitative hypotheses that correlated the experimental and the predicted activity values of the inhibitors. At the end of each run, the top ten hypotheses were generated based on a set of 26 chemically diverse inhibitors of BChE (Figure 1), and the statistical parameters values such as cost, correlation \( r \), and RMS for each hypothesis are shown in Table 1.

Among the ten hypotheses, nine hypotheses contained 1 hydrogen bond acceptor (HBA) and 1 hydrophobic aliphatic (Hy-Ali) group, which indicates that these chemical features are necessary for BChE inhibition. Out of the 10 hypotheses, only 3 hypotheses were selected for further processing based on the maximum fit value (greater than 9). Debnath’s analysis\(^{[42]}\), used to select the best hypothesis, states that the best pharmacophore model should have the highest cost difference, good correlation coefficient, least RMS, and lowest total cost values. Cost differences represent the difference between the null and total cost of hypothesis. A 40–60 bit difference leads to a predictive correlation probability of 75%–90%, and if the difference is greater than 60 bits, the hypothesis is assumed to have a correlation probability of greater than 90%\(^{[31]}\). Hypo1 showed the highest cost difference of 120.12 bits, compared with Hypo4 and Hypo5, indicating its significance. The correlation coefficient is based on linear regression derived from the geometric fit index; Hypo1 showed the highest correlation coefficient (0.96), demonstrating its high predictive ability. The RMS factor represents the deviation of the predicted activity value from the experimental value, and the RMS values were 1.02, 1.23, and 1.24 for Hypo1, Hypo4, and Hypo5, respectively. This result also supports the conclusion that Hypo1 was significant when compared with the two other hypotheses. The reliability of a pharmacophore model also depends on whether the total cost value is distant from the null cost and close to the fixed cost. The fixed cost represents a simple model that fits all data perfectly, while the null cost presumes that there is no relationship in the data and that the experimental activities are normally distributed around their average value. The fixed and total cost values of Hypo1 were 94.82 and 108.57, respectively. Thus, Hypo1, which consisted of 2 HBA, 1 Hy-Ali, and 1 hydrophobic aromatic (Hy-Ar), was selected as the best hypothesis and was employed for further analyses. The chemical features and 3D spatial arrangement of Hypo1 are depicted in Figure 2.

Hypo1 was used to estimate the inhibitory activities of 26 training set compounds to elucidate its predictive accuracy. Hypo1 was able to predict the inhibitory activity value of the 26 training set compounds in the same order of magnitude (Table 2). One moderately active and two inactive compounds

---

Table 1. Information of statistical significance and predictive power presented in cost values measured in bits for the top 10 hypotheses as a result of automated 3D QSAR pharmacophore generation.

| Hypo No | Total cost | Cost difference\(^a\) | RMS\(^b\) | Correlation | Features\(^a\) | Max fit |
|---------|------------|------------------------|----------|-------------|--------------|---------|
| Hypo1   | 105.72     | 120.12                 | 1.01     | 0.96        | 2 HBA, 1 Hy-Ali, 1 Hy-Ar | 8.58    |
| Hypo2   | 103.93     | 116.91                 | 1.10     | 0.95        | 2 HBA, 1 Hy-Ali, 1 RA   | 9.54    |
| Hypo3   | 109.97     | 115.87                 | 1.15     | 0.94        | 2 HBA, 1 Hy-Ali, 1 Hy-Ar| 8.85    |
| Hypo4   | 112.57     | 113.27                 | 1.23     | 0.93        | 2 HBA, 1 Hy-Ali, 1 RA   | 9.18    |
| Hypo5   | 112.98     | 112.86                 | 1.24     | 0.93        | 2 HBA, 1 Hy-Ali, 1 RA   | 9.48    |
| Hypo6   | 116.06     | 109.78                 | 1.34     | 0.92        | 2 HBA, 1 Hy-Ali, 1 RA   | 7.25    |
| Hypo7   | 117.41     | 108.43                 | 1.38     | 0.91        | 2 HBA, 1 Hy-Ali, 1 RA   | 7.28    |
| Hypo8   | 117.43     | 108.41                 | 1.39     | 0.91        | 2 HBA, 1 Hy-Ali, 1 Hy-Ar| 7.86    |
| Hypo9   | 118.78     | 107.06                 | 1.42     | 0.91        | 2 HBA, 1 Hy-Ali, 1 RA   | 7.50    |
| Hypo10  | 118.82     | 107.02                 | 1.43     | 0.91        | 2 HBA, 2 Hy-Ar          | 7.86    |

\(^a\) Cost difference between the null and the total cost. The null cost, the fixed cost and the configuration cost are 225.84, 92.399 and 15.79, respectively.

\(^b\) Abbreviation used for features: RMS, root mean square deviation; HBA, hydrogen bond acceptor; Hy-Ali, hydrophobic aliphatic, Hy-Ar, hydrophobic aromatic and RA, ring aromatic.
were underestimated and overestimated as inactive and moderately active, respectively. All of the active compounds were predicted in their own activity ranges, indicating the predictive ability of Hypo1. Hypo1 aligned with the most active compound 1 \((IC_{50} = 3.6 \text{ nmol/L})\) and least active compound 26 \((IC_{50} = 11 400 \text{ nmol/L})\) in the training set (Figure 3). From this analysis, we suggest that Hypo1 was able to estimate the activity of compounds to a high degree of accuracy relative to their experimental \(IC_{50}\) values (Table 2). The error value was defined as the ratio between experimental and predicted activity value, and Hypo1 demonstrated remarkable consistency. The best pharmacophore model, Hypo1, was validated by various methods such as Fisher’s randomization, a test set, and a decoy set to demonstrate its robustness and statistical significance.

### Table 2. Actual and estimated activity of the training set molecules based on the pharmacophore model Hypo1.

| Compound | Fit value\(^a\) | Exp \(IC_{50}\) nmol/L | Pred \(IC_{50}\) nmol/L | Error\(^b\) Exp scale\(^c\) | Pred scale\(^c\) |
|----------|-----------------|--------------------------|--------------------------|--------------------------|------------------|
| 1        | 6.89            | 3.6                      | 5.5                      | +1.5                     | +++              |
| 2        | 6.38            | 7.8                      | 18                       | +2.3                     | +++              |
| 3        | 6.28            | 10                       | 22                       | +2.2                     | +++              |
| 4        | 6.36            | 21                       | 19                       | -1.1                     | +++              |
| 5        | 6.22            | 22                       | 26                       | +1.2                     | +++              |
| 6        | 6.08            | 23                       | 36                       | +1.6                     | +++              |
| 7        | 6.11            | 25                       | 33                       | +1.3                     | +++              |
| 8        | 6.31            | 26                       | 21                       | -1.2                     | +++              |
| 9        | 6.39            | 29                       | 18                       | -1.7                     | +++              |
| 10       | 6.12            | 34                       | 33                       | -1.0                     | +++              |
| 11       | 6.16            | 56                       | 30                       | -1.9                     | +++              |
| 12       | 6.3             | 76                       | 21                       | -3.5                     | +++              |
| 13       | 4.85            | 250                      | 600                      | +2.4                     | ++               |
| 14       | 4.75            | 490                      | 760                      | +1.5                     | ++               |
| 15       | 4.25            | 610                      | 2 400                    | +3.9                     | ++               |
| 16       | 4.85            | 650                      | 600                      | -1.1                     | ++               |
| 17       | 4.62            | 800                      | 1 000                    | +1.3                     | ++               |
| 18       | 4.9             | 900                      | 540                      | -1.7                     | ++               |
| 19       | 4.55            | 1 200                    | 1 200                    | +1.0                     | +                |
| 20       | 4.01            | 1 600                    | 4 200                    | +2.6                     | +                |
| 21       | 4.72            | 1 800                    | 810                      | -2.2                     | +                |
| 22       | 3.94            | 1 800                    | 810                      | -2.2                     | +                |
| 23       | 3.94            | 3 000                    | 5 000                    | +1.7                     | +                |
| 24       | 4.29            | 3 900                    | 2 200                    | -1.8                     | +                |
| 25       | 4.01            | 5 700                    | 4 200                    | -1.3                     | +                |
| 26       | 4.09            | 7 100                    | 3 500                    | -2.1                     | +                |
| 27       | 4.28            | 11 000                   | 2 200                    | -5.1                     | +                |

\(^a\) Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule.

\(^b\) Division of higher value of experimental or predicted \(IC_{50}\) by lower predicted or experimental \(IC_{50}\) value. ‘+’ indicates that the predicted \(IC_{50}\) is higher than the experimental \(IC_{50}\) and ‘–’ indicates that the predicted \(IC_{50}\) is lower than the experimental \(IC_{50}\); a value of 1 indicates that the predicted \(IC_{50}\) is equal to the experimental \(IC_{50}\).

\(^c\) Activity scale: \(IC_{50}<100 \text{ nmol/L}=+++\) (highly active); 100 nmol/L \(\leq IC_{50}<1000 \text{ nmol/L}=++\) (moderately active); \(IC_{50}\geq10000 \text{ nmol/L}=+\) (low active).

**Figure 3.** The best pharmacophore model Hypo1 aligned to training set compounds: A) active compound 1 \((IC_{50} = 3.6 \text{ nmol/L})\) and B) low activity compound 26 \((IC_{50} = 11 400 \text{ nmol/L})\). The pharmacophore features are color coded (HBA, hydrogen bond acceptor: green; Hy-Ali, hydrophobic aliphatic: light cyan; Hy-Ar, hydrophobic aromatic: cyan).

**Validation of the pharmacophore model**

**Fisher’s randomization test**

Fisher’s test was applied to evaluate the significance of Hypo1 based on statistical validation. A confidence level of 95% was chosen, and a total of 19 random spreadsheets were generated to produce the hypothesis. The significance of the hypothesis was calculated using the formula \(S=[1-(1+X)/\ Y]\times100\), where \(X\) is the total number of hypotheses having a total cost lower than the original hypothesis, and \(Y\) is the total number of HypoGen runs (initial+random runs). Here, \(X=0\) and \(Y=(1+19)\), hence \(95\% =\ [1-\ (1+0)/(19+1)]\times100\). The total cost of 19 random pharmacophore models compared with Hypo1 showed that the original hypothesis was far superior to the 19 other hypotheses, which indicated that the Hypo1 was not generated by chance (Figure 4). This result provided confidence that the Hypo1 could be a best hypothesis that contains all the necessary chemical features to inhibit BChE activity.

**Test set validation**

The test set contains 45 structurally distinct compounds from training set molecules. The test set was used to examine the ability of Hypo1 to predict the activity of external compounds in the same activity range. Except for one active compound that was underestimated as moderately active, all of the remaining compounds are predicted on their own activity range by Hypo1 (Table 3). Hypo1 shows the strong correlation coefficient of 0.94 between experimental and predicted BChE inhibitory activity values for the test set (Figure 5). This result also showed that Hypo1 fit not only for the training set compounds but also for the external compounds; this result...
also demonstrated the predictive ability of Hypo1 to differentiate the active and inactive BChE inhibitors.

Decoy set validation
As a final validation, decoy set screening was performed using the Best Flexible searching module/DS. To determine the robustness of Hypo1, four parameters were calculated: false positives, false negatives, enrichment factor (EF), and goodness of fit score (GF). EF and GF were calculated using the following set of parameters: hit lists (Ht), number of active percent yields (%Y), percent ratio of actives in the hit lists (%A), false negatives, and false positives (Table 4). Hypo1 succeeded in the retrieval of 76% of the active compounds from the decoy set. It predicted 6 active compounds to be inactive compounds (false negatives). Hypo1 showed a GH score of 0.86, indicating that Hypo1 had a greater tendency to show true positives. On the basis of the overall validations, we were strongly assured that the Hypo1 demonstrated excellent prediction of BChE inhibitor activities.

Virtual screening
The validated hypothesis, Hypo1, was used as a 3D structural query for retrieving novel candidate molecules from the Maybridge (60 000) and Chembridge (50 000) chemical databases. A total of 45 496 hit compounds were obtained from the first screening; among these, 1619 compounds were selected for further analysis by applying maximum fit value of greater than 8. Subsequently, these molecules were tested for ADMET and Lipinski’s Rule of Five. ADMET properties calculated the values of blood-brain barrier (BBB) penetration, solubility, cytochrome P450 (CYP450) 2D6 inhibition, hepatotoxicity, HIA, plasma protein binding (PPB), and assessed a broad range of ligand toxicity. The drug should not cross the BBB, hence the level ‘3’ was selected to represent low penetration of BBB. The value of 0 in CYP26 and hepatotoxicity indicates that the molecules are of low toxicity. The cutoff values of solubility and HIA were 3 and 0, respectively. Out of the 1619 molecules, 202 molecules passed the BBB level, absorption, solubility, and toxicity criteria. These hit compounds were subjected to further filtering by applying Lipinski’s Rule of Five, which states that LogP should be less than 5, the molecular weight less than 500, the number of hydrogen bond donors less than 5, the number of hydrogen bond acceptors less than 10, and the number of rotatable bonds less than 10. The flexibility of the molecules and the total number of hydrogen bond acceptors and hydrogen bond donors are important predictors that a compound will have good oral bioavailability. Ultimately, 84 molecules were selected as hits based on drug-like properties. The hit compounds from the virtual screening process were subjected to molecular docking to reduce the false positive rate.

Molecular docking
A BChE complex with butanoic acid from PDB was chosen as the target protein for molecular docking. The establishment and reorganization of specific covalent or non-covalent

Figure 4. The difference in cost between 19 scrambled runs. The 95% confidence level was selected.

Figure 5. Plot of the correlation between the experimental activity and the activity predicted by Hypo1 for the test set molecules (in brown) and training set molecules (in blue).
Table 3. Experimental and predicted IC_{50} values of 45 test set molecules against Hypo1.

| Compound | Structure | Fit value (nmol/L) | Exp IC_{50} (nmol/L) | Pred IC_{50} (nmol/L) | Errora | Exp Scaleb | Pred Scaleb |
|----------|-----------|--------------------|----------------------|-----------------------|--------|------------|-------------|
| 1        | ![Structure 1](image1.png) | 6.90               | 3.0                  | 5.34                  | +1.78  | +++        | +++         |
| 2        | ![Structure 2](image2.png) | 6.36               | 3.3                  | 8.51                  | +2.58  | +++        | +++         |
| 3        | ![Structure 3](image3.png) | 6.36               | 3.9                  | 8.51                  | +2.18  | +++        | +++         |
| 4        | ![Structure 4](image4.png) | 6.07               | 4.9                  | 6.85                  | +1.40  | +++        | +++         |
| 5        | ![Structure 5](image5.png) | 6.07               | 5.0                  | 6.85                  | +1.37  | +++        | +++         |
| 6        | ![Structure 6](image6.png) | 5.93               | 13.9                 | 15.31                 | +1.10  | +++        | +++         |
| 7        | ![Structure 7](image7.png) | 5.35               | 22.1                 | 19.34                 | -1.14  | +++        | +++         |
| 8        | ![Structure 8](image8.png) | 6.05               | 24.0                 | 38.57                 | +1.61  | +++        | +++         |
| 9        | ![Structure 9](image9.png) | 5.73               | 24.5                 | 29.42                 | 1.20   | +++        | +++         |
| 10       | ![Structure 10](image10.png) | 5.77               | 24.8                 | 72.32                 | +2.92  | +++        | +++         |
| 11       | ![Structure 11](image11.png) | 6.69               | 25.2                 | 28.84                 | +1.14  | +++        | +++         |
| 12       | ![Structure 12](image12.png) | 6.69               | 28.0                 | 28.84                 | +1.03  | +++        | +++         |
| 13       | ![Structure 13](image13.png) | 5.17               | 31.0                 | 29.62                 | -1.05  | +++        | +++         |
| 14       | ![Structure 14](image14.png) | 5.17               | 31.6                 | 28.62                 | -1.10  | +++        | +++         |
| 15       | ![Structure 15](image15.png) | 7.08               | 34.1                 | 35.36                 | +1.04  | +++        | +++         |

(to be continued)
| Compound No | Structure | Fit value (nmol/L) | Exp IC$_{50}$ (nmol/L) | Pred IC$_{50}$ (nmol/L) | Error* | Exp Scale$^b$ | Pred Scale$^b$ |
|-------------|-----------|-------------------|------------------------|-------------------------|--------|--------------|--------------|
| 16          | ![Structure 16](image1) | 5.33              | 40.0                   | 22.05                   | -1.81  | +++          | +++          |
| 17          | ![Structure 17](image2) | 5.71              | 42.6                   | 82.64                   | +1.94  | +++          | +++          |
| 18          | ![Structure 18](image3) | 5.71              | 44.3                   | 82.64                   | +1.87  | +++          | +++          |
| 19          | ![Structure 19](image4) | 5.96              | 48.3                   | 47.38                   | -1.02  | +++          | +++          |
| 20          | ![Structure 20](image5) | 5.92              | 50.0                   | 51.57                   | +1.03  | +++          | +++          |
| 21          | ![Structure 21](image6) | 5.80              | 54.0                   | 67.56                   | +1.25  | +++          | +++          |
| 22          | ![Structure 22](image7) | 5.96              | 54.6                   | 47.38                   | -1.15  | +++          | +++          |
| 23          | ![Structure 23](image8) | 6.19              | 57.9                   | 27.49                   | -2.11  | +++          | +++          |
| 24          | ![Structure 24](image9) | 5.75              | 65.0                   | 76.99                   | -1.18  | +++          | +++          |
| 25          | ![Structure 25](image10) | 5.79              | 70.3                   | 68.81                   | -1.02  | +++          | +++          |
| 26          | ![Structure 26](image11) | 5.28              | 74.8                   | 25.39                   | -2.95  | +++          | +++          |
| 27          | ![Structure 27](image12) | 5.91              | 84.4                   | 52.36                   | -1.61  | +++          | +++          |

(to be continued)
| Compound | Structure | Fit value | Exp IC<sub>50</sub> (nmol/L) | Pred IC<sub>50</sub> (nmol/L) | Error<sup>a</sup> | Exp Scale<sup>b</sup> | Pred Scale<sup>b</sup> |
|----------|-----------|-----------|-----------------------------|-----------------------------|--------------|----------------|----------------|
| 28       | ![Structure28] | 5.75      | 90                          | 75.87                       | -1.19        | +++            | +++            |
| 29       | ![Structure29] | 5.30      | 98                          | 115.96                      | +1.18        | +++            | ++             |
| 30       | ![Structure30] | 5.22      | 120                         | 257.32                      | +2.14        | ++             | ++             |
| 31       | ![Structure31] | 4.93      | 130                         | 198.69                      | +1.53        | ++             | ++             |
| 32       | ![Structure32] | 5.10      | 136                         | 142.40                      | +1.05        | ++             | ++             |
| 33       | ![Structure33] | 5.58      | 153                         | 113.40                      | -1.35        | ++             | ++             |
| 34       | ![Structure34] | 5.74      | 155                         | 178.64                      | +1.15        | ++             | ++             |
| 35       | ![Structure35] | 5.62      | 200                         | 103.15                      | -1.94        | ++             | ++             |
| 36       | ![Structure36] | 5.70      | 205                         | 284.84                      | +1.39        | ++             | ++             |
| 37       | ![Structure37] | 5.16      | 220                         | 299.54                      | +1.36        | ++             | ++             |
| 38       | ![Structure38] | 5.70      | 237                         | 244.84                      | +1.03        | ++             | ++             |
| 39       | ![Structure39] | 5.11      | 660                         | 330.68                      | -2           | ++             | ++             |

(to be continued)
interactions between substrates or inhibitors play a crucial role in biological function. Three distinct domains in the active site confer selectivity of BChE inhibitors. The first domain is an acyl binding pocket that contains two residues (L286 and V288) responsible for the binding of larger substrates with acyl groups. A second domain is found near the lip of the active site cavity, and a third domain is defined as the choline binding site (or cation-pi site). The catalytic domain of BChE is composed of nucleophilic serine, histidine, and glutamate residues. The substrate was stabilized between the oxyanion hole and acyl binding pocket of the “catalytic triad” composed of S198, E197, and H438 of the active esteratic site. The mechanism of catalysis depends on the charge relay system, in which the imidazole ring of H438 relays electrons from E197 to S198 and causes the hydroxyl oxygen of S198 to become a nucleophile. A nucleophilic attack of this hydroxyl oxygen on the ester bond of the substrate leads to an acyl-enzyme intermediate and a free choline moiety. Then, the acyl group is hydrolyzed from S198 by the nucleophilic attack of a water molecule that is activated by taking a proton from H438 to form a catalytic triad.

Initially the co-crystal was docked in the active site of BChE to check whether the selected parameters are able to produce the most suitable binding orientation. The RMSD value of 0.79 Å was obtained when the best docked pose overlapped with the co-crystal, which revealed that the default parameters

| Compound | Structure | Fit value | Exp IC<sub>50</sub> (nmol/L) | Pred IC<sub>50</sub> (nmol/L) | Error<sup>a</sup> | Exp Scale<sup>b</sup> | Pred Scale<sup>b</sup> |
|----------|-----------|-----------|-----------------------------|-----------------------------|----------------|----------------|----------------|
| 40       | ![Structure](image1.png) | 4.95      | 711                         | 841                         | +1.18         | ++             | ++             |
| 41       | ![Structure](image2.png) | 4.35      | 761                         | 194.74                      | -3.91         | ++             | ++             |
| 42       | ![Structure](image3.png) | 5.07      | 1010                        | 1362.10                     | +1.35         | +              | +              |
| 43       | ![Structure](image4.png) | 4.74      | 4650                        | 4776.95                     | +1.03         | +              | +              |
| 44       | ![Structure](image5.png) | 3.62      | 5100                        | 1298.10                     | -3.93         | +              | +              |
| 45       | ![Structure](image6.png) | 3.59      | 11400                       | 10400                       | -1.09         | +              | +              |

<sup>a</sup> ‘+’ indicates that the predicted IC<sub>50</sub> is higher than the experimental IC<sub>50</sub>; ‘–’ indicates that the predicted IC<sub>50</sub> is lower than the experimental IC<sub>50</sub>; a value of 1 indicates that the predicted IC<sub>50</sub> is equal to the experimental IC<sub>50</sub>.

<sup>b</sup> Activity scale: IC<sub>50</sub> < 100 nmol/L = +++ (highly active); 100 nmol/L ≤ IC<sub>50</sub> < 1000 nmol/L = ++ (moderately active); IC<sub>50</sub> ≥ 1000 nmol/L = + (low active).

Table 4. Statistical parameter from screening test set molecules.

| No | Parameter | Values |
|----|-----------|--------|
| 1  | Total number of molecules in database (D) | 2100 |
| 2  | Total number of actives in database (A) | 25 |
| 3  | Total number of hit molecules from the database (Ht) | 21 |
| 4  | Total number of active molecules in hit list (Ha) | 19 |
| 5  | % Yield of active (Ht/HA) X 100 | 90.47 |
| 6  | % Ratio of actives (Ha/A) X 100 | 76 |
| 7  | Enrichment Factor (EF) | 76 |
| 8  | False negatives (A-Ha) | 6 |
| 9  | False Positives (H-Ha) | 2 |
| 10 | Goodness of fit score (GF) | 0.87 |
are valid to find the best orientation of BChE in the active site. The same parameters were therefore employed to dock the candidate compounds. The selected candidate molecules from the virtual screening were docked in the BChE active site. The top-ranked 84 compounds based on the docking score were selected as the best potential inhibitors and were manually validated for critical interaction with vital amino acids in the active site of BChE. Intermolecular hydrogen bonding was observed between active residues S198, E197, and H438 in the active site of BChE (Figure 6). From the 84 compounds, 33 candidate molecules showed hydrogen bond interactions with S198 and H438, as well as reliable hydrophobic interactions with Y323 and F329.

**Density functional theory**

The orbital energies such as HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) were calculated for 7 (5 active, one moderate and one inactive inhibitor) training set compounds and for the 33 hits from molecular docking. HOMO and LUMO are responsible for the charge transfer in a chemical reaction\[45\]. The calculated orbital energies of the 33 hits and known inhibitors were compared to analyze the energy transfer and stability of small molecules in protein active site. Comparing the HOMO energy with the activity value of known BChE inhibitors shows an inverse correlation, indicating that the HOMO energy of the inhibitor may transfer its electrons to some critical residues in the active site of BChE. The highest energy value of the HOMO in the hit compounds implied the greatest likelihood of strong inhibition of BChE. Hence, correlation of the HOMO values of hits and the training set molecules showed that 10 hit compounds possessed greater values than the reported inhibitors (training set) of BChE (Table 5). A smaller energy gap (between the LUMO and HOMO) of the hit molecules illustrates that the molecule are more reactive\[46\]. The wide energy gap in the hit molecules is unfavorable for the electron to be excited from the HOMO to the LUMO, which consequently leads to a weak affinity of the inhibitor for BChE. Among the 10 compounds, 5 hit compounds (Figure 7) were selected based on their lowest energy gaps that suggested the molecules would be reactive. Table 5 clearly showed that the moderate and inactive compounds had high energy gaps that were not suitable for the reactivity of the molecules. The atomic orbital composition of the frontier molecular orbital for compound 30080 is shown in Figure 8. On the basis of the results above, we suggest that the hit compounds may possess equivalent or greater electronic properties compared with most active compounds and could

| Name               | HOMO  | LUMO  | ΔE   | IC50  |
|--------------------|-------|-------|------|-------|
| SPB_07954          | -8.31 | -0.99 | 7.31 |       |
| Compound_Number_30238 | -8.32 | -0.23 | 8.08 |       |
| RJC_03502          | -8.55 | -0.44 | 8.11 |       |
| Compound_Number_14811 | -8.62 | -0.46 | 8.16 |       |
| BTB_07807          | -8.65 | -0.96 | 7.69 |       |
| Compound_Number_30080 | -8.65 | -1.19 | 7.46 |       |
| KM_03101           | -8.69 | -0.67 | 8.01 |       |
| KM_02281           | -8.70 | -0.94 | 7.76 |       |
| Compound_Number_15687 | -8.76 | -0.33 | 8.43 |       |
| Compound_Number_23227 | -8.76 | -0.97 | 7.79 |       |
| Training 1         | -8.77 | -0.89 | 7.89 | 3.6   |
| Training 4         | -8.80 | -0.70 | 8.09 | 21    |
| Training 5         | -8.80 | -0.10 | 8.70 | 22    |
| Training 7         | -8.92 | -0.22 | 8.69 | 24.8  |
| Training 8         | -8.98 | -0.78 | 8.21 | 26    |
| Training 17        | -9.18 | -0.34 | 8.84 | 800   |
| Training 26        | -9.12 | -0.28 | 8.83 | 11 400|

**Figure 6.** (A) Chembridge (compound 30080), (B) Maybridge (KM_02281), and (C) Maybridge (SPB_07954). The candidate compounds are represented as green sticks. Hydrogen bonds are shown in black. For the clarity of the docked view the Y323 was not shown.
be used to design novel classes of BChE inhibitors.

Conclusions
In this study, pharmacophore models were generated based on a series of known BChE inhibitors. The main purpose of quantitative pharmacophore generation is to predict or differentiate the active inhibitors from inactive compounds. The best pharmacophore model, Hypo1, consisting of 2 HBA, 1 Hy-Ali, and 1 Hy-Ar, was validated by Fischer’s randomization method, test set, and decoy set. Fischer’s method confirmed the 95% statistical confidence of Hypo1; the test set showed a fairly high correlation between experimental and predicted IC\textsubscript{50} values (correlation coefficient of 0.96), indicating satisfactory predictive ability; additionally, good EF (0.76) and GF (0.87) values for Hypo1 were calculated from the decoy set. The three validation methods confirmed that Hypo1 was the best hypothesis to differentiate the active inhibitors from inactive inhibitors of BChE. Thus, Hypo1 was used as a 3D query to screen molecular structural libraries, including the Maybridge and Chembridge databases. The hit compounds were filtered using ADMET, Lipinski’s Rule of Five, and molecular docking to reduce the number of false positive results. Finally, 33 compounds were selected based on their critical interactions with the significant amino acids in BChE’s active site. To confirm the inhibitors’ potencies, we calculated the orbital energies, such as HOMO and LUMO, for hit compounds and 7 training set compounds. From among the 33 hit compounds, 10 compounds with the highest HOMO values were selected, and this set was further culled to 5 compounds based on their energy gaps, which is important for stability and energy transfer. From the overall results, we confirmed that 5 hit compounds satisfied all the pharmacophoric features in Hypo1 and are potential BChE inhibitors.

Acknowledgements
This research was supported by the Basic Science Research Program (2009-0073267), the Pioneer Research Center Program (2009-0081539), and the Management of Climate Change Program (2010-0029084) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (MEST) of the Republic of Korea. This work was also supported by the Next-Generation BioGreen21 Program (PJ008038) from the Rural Development Administration (RDA) of the Republic of Korea.

Author contribution
Sugunadevi SAKKIAH contributed to all parts of the work, including experimental design, conduct, and analysis, and preparation of the manuscript. Prof Keun Woo LEE analyzed the results.

References
1 Massoulié J, Sussman J, Bon S, Silman I. Chapter 15: Structure and functions of acetylcholinesterase and butyrylcholinesterase. In: Cuello AC editor. Progress in Brain Research: Elsevier; 1993. p 139–46.
2 Sun H, El Yazal J, Lockridge O, Schopfer LM, Brimijoin S, Pang YP. Predicted michaelis-menten complexes of cocaine-butyrylcholinesterase. J Biol Chem 2001; 276: 9330–6.
3 Massoulié J, Pezzementi L, Bon S, Krejci E, Vallette FM. Molecular and cellular biology of cholinesterases. Prog Neurobiol 1993; 41: 31–91.
4 Valle AM, Radić Z, Rana BK, Whitfield JB, O’Connor DT, Martin NG, et al. The cholinesterases: Analysis by pharmacogenomics in man. Chem Biol Interact 2008; 175: 343–45.
5 Quinn DM. Acetylcholinesterase: enzyme structure, reaction dynamics, and virtual transition states. Chem Rev 1987; 87: 955–79.
6 Li B, Sedlacek M, Manoharan I, Boopathy R, Duyan EG, Masson P, et al. Butyrylcholinesterase, paraoxonase, and albumin esterase, but not carboxylesterase, are present in human plasma. Biochem Pharmacol 2005; 70: 1673–84.

Figure 7. A 2D representation of the final 5 hit compounds.

Figure 8. The orbital energy values and energy gap for compound 30080.
O, et al. Carbamates with differential mechanism of inhibition toward acetylcholinesterase and butyrylcholinesterase. J Med Chem 2008; 51: 4200–12.

45 Li Y, Wang Y, Li W, Dong P, Ge G, Yang L, et al. Investigation of binding features: effects on the interaction between CYP2A6 and inhibitors. J Comput Chem 2010; 31: 1822–31.

46 Queiroz AN, Gomes BAQ, Moraes WM Jr, Borges RS. A theoretical antioxidant pharmacophore for resveratrol. Eur J Med Chem 2009; 44: 1644–49.