

Structure-Based and Multiple Potential Three-Dimensional Quantitative Structure–Activity Relationship (SB-MP-3D-QSAR) for Inhibitor Design

Qi-Shi Du,^{†,‡} Jing Gao,[‡] Yu-Tuo Wei,[§] Li-Qin Du,[§] Shu-Qing Wang,^{*,||} and Ri-Bo Huang^{†,§}

[†]State Key Laboratory of Non-food Biomass Energy and Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Academy of Sciences, Nanning, Guangxi 530007, China

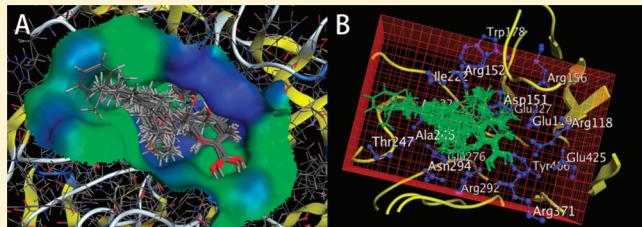
[‡]Department of Anesthesiology, The Second Hospital of Tianjin Medical University, Tianjin 300211, China

[§]State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Life Science and Biotechnology College, Guangxi University, Nanning, Guangxi, 530004, China

^{||}Tianjin Key Laboratory on Technologies Enabling Development of Clinical Therapeutics and Diagnostics (Theranostics), School of Pharmacy, Tianjin Medical University, Tianjin 300070, China

[‡]Gordon Life Science Institute, San Diego, California, United States

ABSTRACT: The inhibitions of enzymes (proteins) are determined by the binding interactions between ligands and targeting proteins. However, traditional QSAR (quantitative structure–activity relationship) is a one-side technique, only considering the structures and physicochemical properties of inhibitors. In this study, the structure-based and multiple potential three-dimensional quantitative structure–activity relationship (SB-MP-3D-QSAR) is presented, in which the structural information of host protein is involved in the QSAR calculations. The SB-MP-3D-QSAR actually is a combinational method of docking approach and QSAR technique. Multiple docking calculations are performed first between the host protein and ligand molecules in a training set. In the targeting protein, the functional residues are selected, which make the major contribution to the binding free energy. The binding free energy between ligand and targeting protein is the summation of multiple potential energies, including van der Waals energy, electrostatic energy, hydrophobic energy, and hydrogen-bond energy, and may include nonthermodynamic factors. In the foundational QSAR equation, two sets of weighting coefficients $\{a_j\}$ and $\{b_p\}$ are assigned to the potential energy terms and to the functional residues, respectively. The two coefficient sets are solved by using iterative double least-squares (IDLS) technique in the training set. Then, the two sets of weighting coefficients are used to predict the bioactivities of inquired ligands. In an application example, the new developed method obtained much better results than that of docking calculations.



INTRODUCTION

The quantitative structure–activity relationship (QSAR) may be the most frequently used technique in computer-aided drug design.^{1–5} The QSAR module in the famous software package SYBYL⁶ is the best selling point and widely used by researchers in drug design and in other research fields of life science. The most remarkable advantage of QSAR over other drug design techniques is that it is a target structure-independent approach. In other words, QSAR can be applied in the cases that the structures of drug targets are not available.⁷

The inhibitions of enzymes (proteins) are two-side interactions, which mainly determined by the binding free energies between ligands and the host proteins. However, traditional QSAR is a one-side technique, only considering the structures and physicochemical properties of inhibitors. The three-dimensional structures of drug targeting proteins are the basis of structure-based drug design approaches, such as docking^{8–10} scoring, and screening. The structure-independent advantage of the traditional QSAR technique, on the other hand, may be a shortcoming.

The theoretical basis of QSAR is the relationship between Gibbs free energy ΔG_i° and reaction equilibrium constant K_i of ligand–receptor interaction,

$$\log K_i = \frac{-1}{2.303RT} \Delta G_i^\circ \quad (1)$$

In traditional 2D-QSAR, the Gibbs free energy, ΔG_i° , is estimated using a linear free energy equation,¹⁰

$$\begin{aligned} \log K_i &= a_1 f_{H-b}(X_h) + a_2 f_{elec}(X_e) + a_3 f_{st}(X_s) + \dots \\ &= -\Delta G_i^\circ \end{aligned} \quad (2)$$

In eq 2, f_{H-b} refers to the hydrogen bond term, f_{elec} to the electrostatic term, f_{st} to the steric term of the drug candidate i , and $\{a_1, a_2, a_3, \dots\}$ are the weighting coefficients of the interaction terms. The interaction terms can affect either the ligand–receptor

Received: February 3, 2012

Published: April 5, 2012



recognition or activation. The interaction free energy, ΔG_i° , is a linear or quadratic function of the corresponding physicochemical parameter X_i . Commonly parameters are the following: (i) the *n*-octanol–water partition coefficient for a hydrophobic term, (ii) Taft E_s parameter for steric effect, (iii) Hammett constants to describe electronic effects, and (iv) the molar refractivity to account for dispersion forces. Lately, some nonthermodynamic properties were included in the list, such as topology indices.^{4–6,10}

In the 3D-QSAR approach, represented by CoMFA (comparative molecular field analysis),^{11,12} the ligands are aligned in a cubic grid, established by a set of space points $\{r_j\}$. The interaction Gibbs free energy, ΔG_i° of ligand i is evaluated by the potential energies $u_i(r_j)$ at the grid points,

$$-\Delta G_i^\circ = \sum_{j=1}^N b_j u_i(r_j) = pK_i \quad (3)$$

where pK_i refers to the logarithm bioactivity of sample i , which is proportional to the binding free energy between ligand and receptor. In eq 3, the $\{b_j\}$ is the weighting coefficients of space points. Usually the interaction energy contribution $u_i(r_j)$ of sample i at point r_j is the sum of electrostatic interaction and van der Waals interaction,^{11,12}

$$\begin{aligned} u_i(r_j) &= f_{i,j}^{\text{elec}} + f_{i,j}^{\text{vdW}} \\ &= \sum_{k=1}^{M_i} \frac{q_k}{r_{j,k}} + \sum_{k=1}^{M_i} \varepsilon_k \left[\left(\frac{\sigma_k}{r_{j,k}} \right)^{12} - 2 \left(\frac{\sigma_k}{r_{j,k}} \right)^6 \right] \end{aligned} \quad (4)$$

where $f_{i,j}^{\text{elec}}$ are the electrostatic field terms, $f_{i,j}^{\text{vdW}}$ are the van der Waals field terms, $r_{j,k}$ is the distance between point j and atom k , q_k and σ_k is the charge and diameter of atom k , respectively.

Actually, eqs 2 and 3 of both 2D- and 3D-QSAR cannot yield the ligand–receptor interaction free energy ΔG° , because in the traditional QSAR the structures of host proteins are unknown. We cannot evaluate the interaction energy only based on the structures of ligands without knowing the structures of receptors.

With the fast development of protein structure test techniques in both hardware and software, such as X-ray and NMR, more and more three-dimensional structures of drug targeting proteins are available.¹³ Involving the structural information of drug targeting proteins in the QSAR technique is the new direction of QSAR development. In this article we explore the theoretical foundations and computational techniques of the structure-based QSAR. A newly developed structure-based and multiple potential three-dimensional QSAR (SB-MP-3D-QSAR) is introduced.

THEORY AND METHOD

In this section, we introduce the detailed calculation procedure and mathematical techniques of SB-MP-3D-QSAR. The theoretical considerations and implications in the formulation are illustrated by using equations and graphics.

Multiple Docking Calculations of Protein and Ligands. The SB-MP-3D-QSAR, actually, is a combinatorial method of dock approach and QSAR technique. The first step of SB-MP-3D-QSAR is the multiple docking calculations between ligands and targeting protein.^{7–9,14,15} In the docking calculations the structure of host protein is rigid and the structures of ligands are flexible. After multiple docking calculations we have the three-dimensional coordinate system of host protein and ligands, as shown in Figure 1A and B.

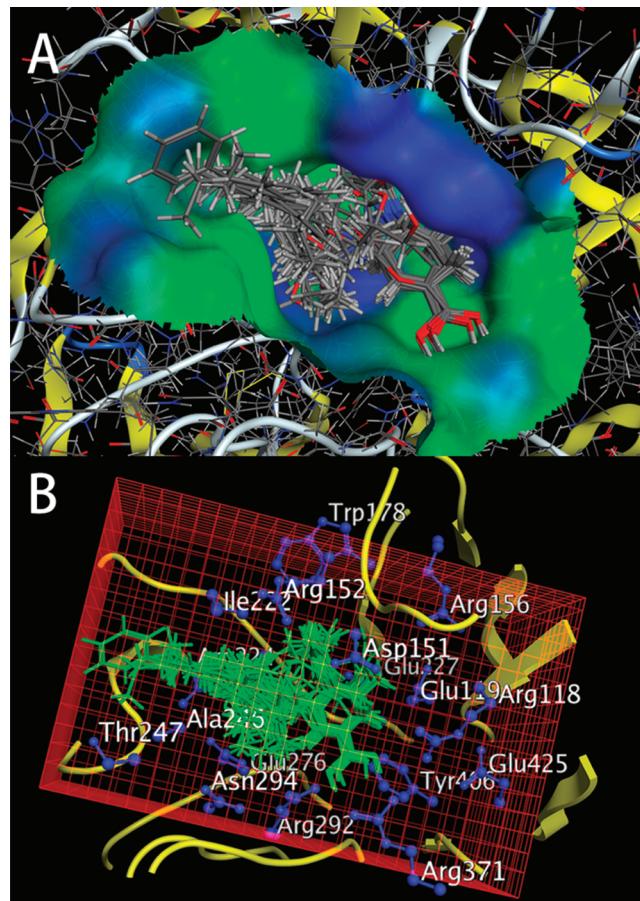


Figure 1. Protein–ligand docking structures and the interaction coordinate system between host protein and ligands for SB-MP-3D-QSAR. (A) Multiple docking structures between host protein (pdb code: 1f8b) and ligands. In the docking calculations the structure of host protein is rigid and the structures of ligands are flexible. (B) Protein–ligand interaction coordinate system and the positions of functional residues in the host protein. On the basis of the docking structural system, the functional residues of protein are selected, which play the key roles in the binding interactions and make the main contributions to the binding free energies.

Unlike common docking calculations, in the SB-MP-3D-QSAR, we only compute the binding energies between ligands and the functional residues^{16–19} of the host protein, which play the key roles in the binding interactions and make the main contributions to the binding free energies. The selection of functional residues will be introduced in next section. The total binding energy between ligand i and host protein is organized as the summation of the interaction energies $u_{i,j}$ between ligand i and functional residue j ,

$$E_i^{\text{total}} = \sum_{j=1}^M u_{i,j} \quad (5)$$

where M is the number of functional residues in host protein. The interaction energy $u_{i,j}$ between ligand i and residue j is the summation of atom pairs,

$$u_{i,j} = \sum_{k=1}^{K_i} \sum_{l=1}^{L_j} e_{k,l}(r_{k,l}) \quad (6)$$

where K_i is the number of atoms in ligand i , L_j is the number of atoms in functional residues j , the index k is for the atoms in the ligand, l is for the atoms in residue, and the $r_{k,l}$ is the distance between atom k and atom l . In eq 6, the atomic pairwise energy

term $e_{k,l}(r_{k,l})$ between atom k of ligand i and atom l of residue j is evaluated using the Coulomb equation and Lennard-Jones equation,

$$\begin{aligned} e_{k,l}(r_{k,l}) &= t_{k,l}^{\text{elec}} + t_{k,l}^{\text{vdW}} + \dots \\ &= \frac{q_k q_l}{r_{k,l}} + \varepsilon_{k,l} \left[\left(\frac{\sigma_{k,l}}{r_{k,l}} \right)^{12} - 2 \left(\frac{\sigma_{k,l}}{r_{k,l}} \right)^6 \right] + \dots \end{aligned} \quad (7)$$

where $t_{k,l}^{\text{elec}}$ are the electrostatic energy terms, $t_{k,l}^{\text{vdW}}$ are the van der Waals energy terms, q_k and q_l are the atomic charges, and $\sigma_{k,l}$ is the mean diameter of atom k and l . In eq 7, more potential energy terms can be included (e.g., lipophilic interaction term $t_{k,l}^{\text{lip}}$ and hydrogen bond term $t_{k,l}^{\text{H-b}}$). In the multiple potential QSAR, eq 7 is rewritten as the general form,

$$e_{k,l}(r_{k,l}) = t_{k,l}^{\text{elec}} + t_{k,l}^{\text{vdW}} + t_{k,l}^{\text{lip}} + t_{k,l}^{\text{H-b}} \dots = \sum_p t_{k,l}^p \quad (8)$$

where index p is for the energy types, P is the number of potential energy types, and $t_{k,l}^p$ is the p th potential energy type between atom k and atom l .

Selection of Functional Residues in Targeting Protein.

On the basis of the multiple docking structural system between host protein and ligands, the functional residues^{16–19} of targeting protein are selected. The functional residues are defined as the residues that make the main contributions to the binding free energies, or play the key roles in host-ligand recognition. The functional residues were frequently studied and discussed in the literature.^{16–19} In SB-MP-3D-QSAR, the functional residues can be found in the ligand–residue pairwise docking calculations. If all ligand–residue pairwise interaction energies $\{u_{i,j}\}$ are listed in order, the functional residues should be ranked in the first 20 positions. In general, the functional residues are in 5 Å range surrounding the ligands, and binding with the ligands through hydrogen bond, salt bridge, cation-π interaction,^{20,21} hydrophobic interactions, and other main molecular interactions. In the selection of functional residues the previous works of other authors and experimental data have to be taken into consideration.^{18,19} The functional residues of neuraminidase 1F8B^{22,23} are shown in Figure 1B.

Foundational Equation of SB-MP-3D-QSAR. Inserting eqs 6 and 8 into the eq 5, we get eq 9, the total interaction energy between ligand and functional residues of targeting protein,

$$E_i^{\text{total}} = \sum_{j=1}^M \sum_{k=1}^{K_j} \sum_{l=1}^{L_j} \sum_{p=1}^P t_{k,l}^p \quad (9)$$

When the summations to the index k (for the atoms of ligand) and l (for the atoms of residue j) are over,

$$g_{i,j,p} = \sum_{k=1}^{K_j} \sum_{l=1}^{L_j} t_{k,l}^p \quad (10)$$

eq 9 is simplified to the following form,

$$E_i^{\text{total}} = \sum_{j=1}^M \sum_{p=1}^P g_{i,j,p} \quad (11)$$

where $g_{i,j,p}$ is the interaction energy contribution of ligand i and residue j in the p th potential energy type.

On the basis of the total docking energy between ligand and functional residues of targeting protein (eq 11), the foundational equation of SB-MP-3D-QSAR is derived as follows. Two sets of weighting coefficient sets are inserted in eq 11: $\{a_j\}$ is

the weighting coefficients of the functional residues and $\{b_p\}$ is the weighting coefficients of the potential energy types

$$-\Delta G_i^\circ = \sum_{j=1}^M a_j \sum_{p=1}^P b_p g_{i,j,p} = pK_i \quad (i = 1, 2, \dots, N) \quad (12)$$

where index i is for the ligands and N is the total number of ligands. The energy elements $g_{i,j,p}$ form a three-dimensional matrix $G_{N \times M \times P}$. In eq 12, the total interaction energy E_i^{total} of eq 11 is replaced by free energy ΔG° , and correlated with bioactivity (pK_i) of ligands.

Iterative Double Least-Squares (IDLS) Solution. The foundational equation eq 12 of SB-MP-3D-QSAR is a three-dimensional simultaneous linear equation, in which there are two sets of coefficients $\{a_j\}$ and $\{b_p\}$ to be solved. The iterative double least-squares (IDLS) technique is used to determine the values of the coefficient sets $\{a_j\}$ and $\{b_p\}$ alternately and iteratively.^{4,5,24,25} By using a set of initial values of coefficients $\{a_j^{(0)}\}$, the three-dimensional data matrix $G_{N \times M \times P}$ is reduced to a two-dimensional data matrix $D_{N \times P}^{(1)}$ with the elements given by

$$d_{i,p}^{(1)} = \sum_{j=1}^M a_j^{(0)} g_{i,j,p} \quad (13)$$

Thus, the three-dimensional simultaneous linear equation set (eq 12) is reduced to a set of two-dimensional equations, i.e.,

$$\sum_{p=1}^P b_p^{(1)} d_{i,p}^{(1)} = pK_i \quad (i = 1, 2, \dots, N) \quad (14)$$

The above equation set can be solved by using the least-squares approach, yielding the first solutions of the coefficients $\{b_p^{(1)}\}$. Then the values of $\{b_p^{(1)}\}$ are used to reduce the three-dimensional data matrix $G_{N \times M \times P}$ to a two-dimensional data matrix $V_{N \times M}^{(1)}$ with the elements given by

$$v_{i,j}^{(1)} = \sum_{p=1}^P b_p^{(1)} g_{i,j,p} \quad (15)$$

Similarly, the three-dimensional simultaneous linear equation set (eq 12) is reduced to a two-dimensional equation set, as given by

$$\sum_{j=1}^M a_j^{(1)} v_{i,j}^{(1)} = pK_i \quad (16)$$

The above equation set can be solved by using the least-squares approach, leading to the solution of coefficients $\{a_j^{(1)}\}$. Then the values of $\{a_j^{(1)}\}$ are used for the new solutions of the weighting coefficients $\{b_p^{(2)}\}$ of potential energy types. The above procedure is performed iteratively for n steps, i.e., until reaching the converged solutions as denoted by $\{a_j^{(n)}\}$ and $\{b_p^{(n)}\}$. The convergence criterion for the iterative procedure is given by the following equation,

$$\begin{aligned} |Q^{(n+1)} - Q^{(n)}| &= \sqrt{\frac{1}{N} \sum_{i=1}^N (pK_i - pK_i^{(n+1)})^2} \\ &\quad - \sqrt{\frac{1}{N} \sum_{i=1}^N (pK_i - pK_i^{(n)})^2} \\ &\leq \epsilon (10)^{-6} \end{aligned} \quad (17)$$

where $Q^{(n)}$ represents the square root of the summation of squared differences between the experimental bioactivities and the predicted bioactivities in the n th step, and $Q^{(n+1)}$ that in the $(n+1)$ th step. Now, the values of $\{a_j^{(n)}\}$ and $\{b_p^{(n)}\}$ can be used to predict the bioactivity pK_i^{pred} of the i -th query ligand through the following equation,

$$\sum_{j=1}^M a_j \sum_{p=1}^P b_p g_{i,j,p} = pK_i^{\text{pred}} \quad (18)$$

where the $g_{i,j,p}$ are the interaction energy elements of potential energy type p between inquiry ligand i and residue j of protein. The energy elements $g_{i,j,p}$ are calculated using eqs 7, 8, and 10 based on the docking structure of host protein and inquiry ligand. The mathematical procedure of IDLS is schematically illustrated in flow Figure 2.

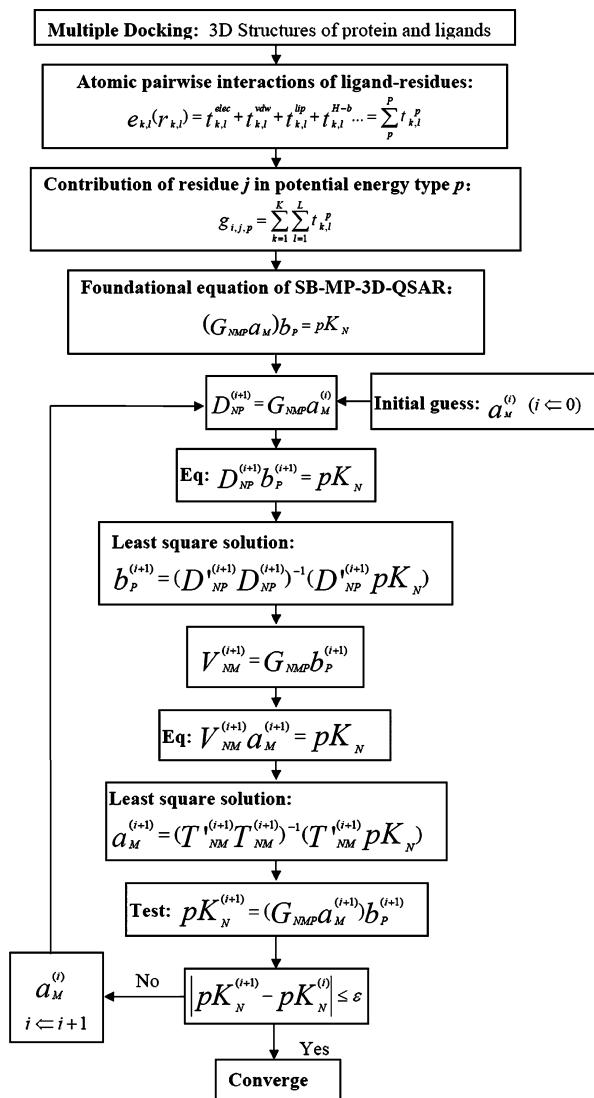


Figure 2. Flow of structure-based and multiple potential three-dimensional quantitative structure–activity relationship (SB-MP-3D-QSAR) and mathematical procedure of iterative double least-squares (IDLS).

■ APPLICATION EXAMPLE

As an example, the SB-MP-3D-QSAR is applied in the study of the neuraminidase (NA) inhibitor design against influenza A

viruses.^{26,27} The demonstrative example is helpful for the illustration of the application procedure of SB-MP-3D-QSAR. Total 49 ligand samples are used in the training set. The molecular structures and bioactivities pIC_{50} of the 49 ligand molecules are selected from refs 22, 23, 28, and 29 and listed in Table 1.

In the standard 3D-QSAR methods, such as CoMFA^{11,12} and CoMSIA,^{30,31} a cubic grid is set up surrounding the aligned ligand molecules, then thousands and millions potential values at these grid points are calculated. However, the potential energy differences of grid points inside and outside ligand molecules may be up to 10^6 , and huge number of grid points may cause over correlation problem. In the SB-MP-3D-QSAR calculations, we only focus on the interactions between ligands and functional residues of host protein. The X-ray crystal structure of N9 neuraminidase (PDB code: 1f8b)³² is used as the host protein. Several commercial available drugs against influenza A viruses are designed based on this structure.^{27,32} The 49 ligands are docked in the host pocket of NA, as shown in Figure 1A. On the basis of the multiple docking complex system of host protein and ligands, 17 functional residues (Arg118, Glu119, Asp151, Arg152, Arg156, Trp178, Ile222, Arg224, Glu227, Ala246, Thr247, Glu276, Arg292, Asn294, Arg371, Tyr406, and Glu425) are selected from the NA protein, and shown in Figure 2, where the locations of functional residues are indicated. Some functional residues interact directly with the inhibitors and make main contributions to the binding free energies, and some functional residues hold the substrate and create a suitable environment for catalytic interaction.^{16–19,33}

In SB-MP-3D-QSAR the atomic pairwise interaction energies between ligands and residues are calculated using eqs 6 and 8. Five potential energy terms are included in the QSAR calculations: electrostatic interaction $t_{k,l}^{\text{elec}}$, van der Waals interaction $t_{k,l}^{\text{vdW}}$, hydrogen-bond donor interaction $t_{k,l}^{\text{Hb-d}}$, hydrogen-bond acceptor interaction $t_{k,l}^{\text{Hb-a}}$, and lipophilic interaction $t_{k,l}^{\text{lip}}$. The electrostatic interaction terms are computed using classical Coulomb equation and atomic partial charges, the van der Waals interaction terms using 6–12 Lennard-Jones equation, the hydrogen bond donor and acceptor interaction terms using 4–8 Lennard-Jones equation,³⁴ and the lipophilic interaction terms using the equation suggested in our previous studies,^{35–37}

$$t_{\text{lip}}^{k,l}(r_{k,l}) = -\frac{l_k l_l}{r_3^{k,l}} \quad (19)$$

where the l_k and l_l are the lipophilic indices of atom k and l .^{36,38–41} According to eq 19, if the lipophilic indices of atom k and l take the same sign (both are lipophilic or hydrophilic), the lipophilic interaction makes negative contribution to the binding free energy ΔG° ; otherwise, the lipophilic contribution of atom k and l is positive.

After the calculations of interaction energy terms between functional residues and ligands are over, we get the three-dimensional matrix $G_{N \times M \times P}$ and foundational equation (eq 12) of SB-MP-3D-QSAR, where $N = 49$ is the number of ligands, $M = 17$ is the number of functional residues, and $P = 5$ is the number of potential energy types. The three-dimensional simultaneous linear equation (eq 12) is solved using the iterative double least-squares (IDLS) technique described in the Theory and Method section. The initial values of weighting coefficients of five potential energy types $\{a_p^{(0)}\}$ are assigned to be 1. This is a

reasonable initial guess for $\{a_p^{(0)}\}$, meaning that the roles of all potential energy types are equally important.

In the IDSL procedure the correlation coefficient R between experimental pIC_{50} and calculated pIC_{50} increases and the

Table 1. Structures and Experimental pIC_{50} of 49 Molecules Used in the Training Set^a

M01 pIC_{50} 6.65		M02 pIC_{50} 7.22	
M03 pIC_{50} 9.52		M04 pIC_{50} 7.92	
M05 pIC_{50} 7.66		M06 pIC_{50} 9.00	
M07 pIC_{50} 9.00		M08 pIC_{50} 6.21	
M09 pIC_{50} 6.28		M10 pIC_{50} 8.00	
M11 pIC_{50} 9.00		M12 pIC_{50} 9.00	
M13 pIC_{50} 8.52		M14 pIC_{50} 6.22	
M15 pIC_{50} 6.74		M16 pIC_{50} 6.68	
M17 pIC_{50} 6.70		M18 pIC_{50} 6.57	
M19 pIC_{50} 5.20		M20 pIC_{50} 8.05	
M21 pIC_{50} 9.00		M22 pIC_{50} 5.43	
M23 pIC_{50} 7.22		M24 pIC_{50} 5.70	
M25 pIC_{50} 6.89		M26 pIC_{50} 6.74	
M27 pIC_{50} 6.52		M28 pIC_{50} 6.70	
M29 pIC_{50} 7.05		M30 pIC_{50} 6.28	
M31 pIC_{50} 5.66		M32 pIC_{50} 5.70	
M33 pIC_{50} 3.60		M34 pIC_{50} 7.32	
M35 pIC_{50} 7.19		M36 pIC_{50} 6.74	

Table 1. continued

M37 8.22		M38 7.00	
M39 6.70		M40 7.05	
M41 7.07		M42 7.92	
M43 6.70		M44 7.96	
M45 5.57		M46 5.19	
M47 5.40		M48 5.04	
M49 8.70			

^aExperimental pIC₅₀ data are from refs 22, 23, 28, and 29.

prediction residue Q decreases with the iterations. The curves of correlation coefficients R vs iterations are shown in Figure 3A, where R_a is for the weighting coefficients $\{a_k^{(n)}\}$ and R_b is for the weighting coefficients $\{b_j^{(n)}\}$. The prediction residue Q between the calculated bioactivities and the experimental bioactivities of ligands are shown in Figure 3B, where Q_a is for $\{a_k^{(n)}\}$ iteration and Q_b is for $\{b_j^{(n)}\}$ iteration. It has been observed in Figure 3 that after 300 iterations, the R and Q converged smoothly and constantly. For the training data set, the correlation coefficient is $R_{\text{trai}} = 0.9319$ and the average residue of calculated bioactivity is $Q_{\text{trai}} = \pm 0.0646$. The experimental and calculated pIC₅₀ values of 49 ligands are listed in Table 2.

In order to check the prediction ability of SB-MP-3D-QSAR to the inquired ligands, a jackknife test^{42–45} is used in this study. In statistical prediction, the independent data set test, subsampling test, and jackknife test are three cross-validation tests often used in the literature. Of these three, the jackknife test is deemed the most rigorous and objective one.⁴² Therefore, the jackknife test has been adopted by more and more investigators for testing the power of a predictor. The predicted pIC₅₀ values of 49 ligands in the jackknife test are shown in Table 2. For comparison the pIC₅₀ values of docking calculations are also listed in Table 2, which are calculated using AutoDock4 software package.⁴⁶

DISCUSSION AND CONCLUSION

The structure-based and multiple potential three-dimensional QSAR (SB-MP-3D-QSAR), actually, is the combination of common 3D-QSAR and docking calculations, which possesses five advantages over other inhibitor design approaches. (i) The first advantage of SB-MP-3D-QSAR over traditional QSAR methods is that the structure information of host proteins is

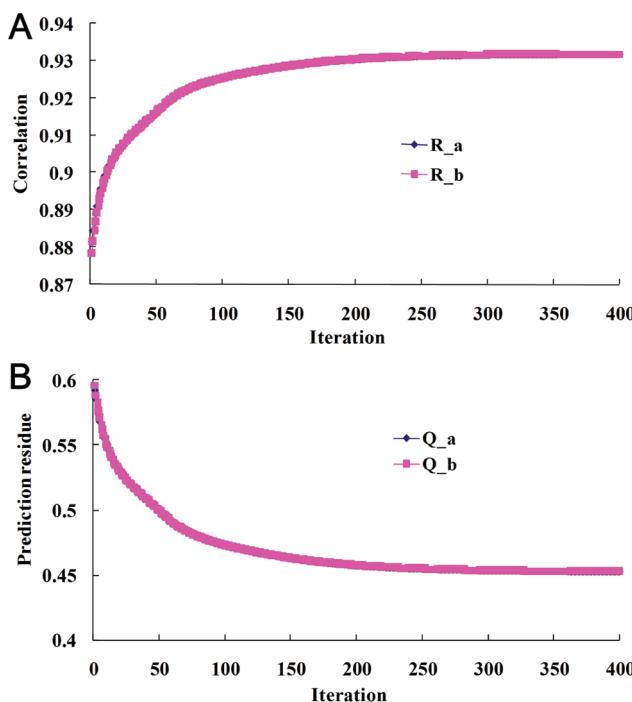


Figure 3. (A) Correlation coefficients of the calculated pIC₅₀ in training set. R_a is for $\{a_k^{(n)}\}$ iterations, and R_b is for $\{b_j^{(n)}\}$ iteration. The correlation coefficient R increases from 0.8765 to 0.9321. (B) Residue Q between predicted bioactivities and experimental bioactivities of molecules in training set. Q_a is for $\{a_k^{(n)}\}$ iteration, and Q_b is for $\{b_j^{(n)}\}$ iteration. The residue Q decreases from 0.5963 to 0.4626 with the iterations. Promising improvements are from the better theoretical model and the advanced mathematical technique IDLS (iterative double least-squares) of SB-MP-3D-QSAR.

Table 2. pIC_{50} Data of 49 1F8B Neuraminidase Inhibitors from Experiments, SB-MP-3D-QSAR Training Calculations, Jackknife Tests, and Docking Calculations, Respectively

ligand	exp^a	SB-MP-3D-QSAR ^b		dock^c
	pIC_{50}	$\text{pIC}_{50}^{\text{trai}}$	$\text{pIC}_{50}^{\text{test}}$	pIC_{50}
M01	6.65	6.401	5.769	8.013
M02	7.22	7.497	7.928	8.713
M03	9.52	9.059	8.863	9.546
M04	7.92	8.888	10.442	9.276
M05	7.66	7.370	6.522	8.733
M06	9.00	8.380	9.103	9.239
M07	9.00	8.846	8.805	8.852
M08	6.21	6.084	5.940	8.320
M09	6.28	6.200	6.161	9.039
M10	8.00	7.926	7.899	8.662
M11	9.00	8.492	8.332	8.922
M12	9.00	8.397	7.862	9.078
M13	8.52	7.466	7.101	8.808
M14	6.22	5.978	2.958	8.492
M15	6.74	6.895	7.119	9.523
M16	6.68	7.173	7.315	9.036
M17	6.70	6.738	6.743	8.975
M18	6.57	6.585	6.594	9.266
M19	5.20	4.941	4.837	8.458
M20	8.05	7.851	7.722	8.577
M21	9.00	8.845	8.360	9.191
M22	5.43	6.236	6.427	8.903
M23	7.22	6.87	6.714	8.283
M24	5.70	6.421	7.282	8.530
M25	6.89	6.505	6.322	8.702
M26	6.74	8.121	8.864	8.615
M27	6.52	6.560	6.563	8.892
M28	6.70	6.734	6.738	8.948
M29	7.05	7.175	7.903	9.522
M30	6.28	6.646	7.020	9.414
M31	5.66	6.244	6.455	8.174
M32	5.70	5.598	5.568	8.469
M33	3.60	3.507	1.615	9.028
M34	7.32	7.376	10.678	8.953
M35	7.19	6.567	6.371	9.026
M36	6.74	6.558	6.499	9.116
M37	8.22	8.422	8.547	8.717
M38	7.00	6.867	6.777	9.356
M39	6.70	6.678	6.665	9.272
M40	7.05	6.576	6.306	8.587
M41	7.07	7.901	8.323	9.173
M42	7.92	7.928	7.933	8.669
M43	6.70	6.709	6.278	9.009
M44	7.96	7.919	8.926	8.943
M45	5.57	4.942	4.720	8.677
M46	5.19	5.666	5.817	8.680
M47	5.40	5.762	6.021	9.110
M48	5.04	4.960	3.096	7.570
M49	8.70	8.780	10.038	8.959
SB-QSAR training		$R_{\text{trai}} = 0.9319$	$Q_{\text{trai}} = \pm 0.0646$	
SB-QSAR test ^b		$R_{\text{test}} = 0.7602$	$Q_{\text{test}} = \pm 0.1621$	
dock ^c		$R_{\text{dock}} = 0.3227$	$Q_{\text{dock}} = \pm 0.3158$	

^aExperimental pIC_{50} data from refs 22, 23, 28, and 29. ^bPredicted in jackknife test. ^cDocking calculations using AutoDock4⁴⁶ software package.

taken into consideration in the inhibitor design. The inhibition of proteins (enzymes) is a two-side interaction: one side is the inhibitor and the other is the host protein. The structural information of host protein is certainly helpful for inhibitor design. (ii) The second advantage of SB-MP-3D-QSAR over common 3D-QSAR is that only functional residues of host protein are used in the calculations of interaction free energy. The number of functional residues is much less than the cubic grid points used in CoMFA^{11,12} and CoMSIA.^{30,31} In this way the overcorrelation problem and the cutoff threshold problem are avoided. (iii) The third advantage of SB-MP-3D-QSAR over common 3D-QSAR is that there is no limitation to the number of potential energy terms. More potential energy terms can be used in the calculations of interaction free energy, including nonthermodynamic terms. (iv) The fourth advantage of SB-MP-3D-QSAR over common docking type methods is that it does not directly use the binding free energy of docking calculations. Two sets of weighting coefficients $\{a_k\}$ and $\{b_j\}$ are used in the prediction equation for potential energy types and for functional residues. The binding free energy between ligand and host protein is the main contribution to the inhibition interaction; however, it may be not the only factor. The two weighting coefficient sets contain the information of other interaction factors, which are not included in the binding free energy. In the application example the correlation coefficient R_{dock} between experimental pIC_{50} and the pIC_{50} of docking calculations is merely 0.3227. In contrast, the correlation coefficient R_{test} between experimental pIC_{50} and the jackknife test pIC_{50} of SB-MP-3D-QSAR is 0.7602. (v) The fifth advantage is that more advanced mathematical technique IDSL is used in SB-MP-3D-QSAR calculations, which improves the prediction ability promisingly. In the IDSL technique, two least-squares procedures are performed iteratively and alternately for the solutions of two weighting coefficient sets $\{a_k\}$ and $\{b_j\}$. In the iterative process the correlation coefficient R increases and the prediction residue Q decreases smoothly and constantly.

As illustrated in the Theory and Method section, SB-MP-3D-QSAR is the combination of docking approach and QSAR technique, in which the QSAR calculation is performed at the basis of docking calculations. In principle, SB-MP-3D-QSAR can be combined with any type of docking methods and can improve the prediction ability of docking calculations. The bioactivities of drugs are affected by many factors, including thermodynamic and nonthermodynamic factors. In this study, five potential energy types (electrostatic term, van der Waals term, hydrogen-bond donor term, hydrogen-bond acceptor term, and lipophilic term) are used in the SB-MP-3D-QSAR calculations. For further improvement of SB-MP-3D-QSAR, more potential terms can be included, especially the nonthermodynamic terms, such as topological term^{47,48} and chirality term.^{49–52} The interaction coordinate system of SB-MP-3D-QSAR depends on the molecular docking calculations, which determine the locations and orientations of ligands in the host protein. For better prediction results of SB-MP-3D-QSAR, reliable docking calculations are vital. The foundational data matrix $G_{N \times M \times P}$ of SB-MP-3D-QSAR is the interaction energies between ligands and functional residues, which are calculated using force field parameters and empirical equations. For the further improvement of SB-MP-3D-QSAR, we need better force field parameters⁵³ and empirical formulations.

AUTHOR INFORMATION

Corresponding Author

*Phone: 086-22-2354-2621/086-0-13820660330. E-mail: wsq1975@yahoo.com.cn.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants 30970562 and 31160032 from the National Science Foundation of China (NSFC), grant 2010GXNSFD013030 from the Key Natural Science Foundation of Guangxi (GXNSFD), and grant 2011KY43 from the Tianjin medical university. We appreciate the constructive suggestions of the anonymous reviewers and the editor in improving the manuscript.

REFERENCES

- (1) Paula, S.; Tabet, M. R.; Farr, C. D.; Norman, A. B.; Ball, W. J., Jr. Three-dimensional quantitative structure-activity relationship modeling of cocaine binding by a novel human monoclonal antibody. *J. Med. Chem.* **2004**, *47* (1), 133–142.
- (2) Sanders, J. M.; Ghosh, S.; Chan, J. M.; Meints, G.; Wang, H.; Raker, A. M.; Song, Y.; Colantino, A.; Burzynska, A.; Kafarski, P.; Morita, C. T.; Oldfield, E. Quantitative structure-activity relationships for gammadelta T cell activation by bisphosphonates. *J. Med. Chem.* **2004**, *47* (2), 375–384.
- (3) Akula, N.; Lecanu, L.; Greeson, J.; Papadopoulos, V. 3D QSAR studies of AChE inhibitors based on molecular docking scores and CoMFA. *Bioorg. Med. Chem. Lett.* **2006**, *16* (24), 6277–6280.
- (4) Du, Q. S.; Huang, R. B.; Wei, Y. T.; Pang, Z. W.; Du, L. Q.; Chou, K. C. Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design. *J. Comput. Chem.* **2009**, *30* (2), 295–304.
- (5) Du, Q. S.; Huang, R. B.; Wei, Y. T.; Du, L. Q.; Chou, K. C. Multiple field three dimensional quantitative structure-activity relationship (MF-3D-QSAR). *J. Comput. Chem.* **2008**, *29* (2), 211–219.
- (6) Homer, R. W.; Swanson, J.; Jilek, R. J.; Hurst, T.; Clark, R. D. SYBYL line notation (SLN): a single notation to represent chemical structures, queries, reactions, and virtual libraries. *J. Chem. Inf. Model.* **2008**, *48* (12), 2294–2307.
- (7) García-Sosa, A. T.; Sild, S.; Takkis, K.; Maran, U. Combined approach using ligand efficiency, cross-docking, and antitarget hits for wild-type and drug-resistant Y181C HIV-1 reverse transcriptase. *J. Chem. Inf. Model.* **2011**, *51* (10), 2595–2611.
- (8) Huang, S.-Y.; Zou, X.-Q. Construction and test of ligand decoy sets using MDock: community structure-activity resource benchmarks for binding mode prediction. *J. Chem. Inf. Model.* **2011**, *51* (9), 2107–2114.
- (9) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.-B. Christine humblet comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49* (6), 1455–1474.
- (10) Hansch, C. L., A. J. *Substitute constants for correlation analysis in chemistry and biology*; John Wiley: New York, 1979.
- (11) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967.
- (12) Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Recent advances in comparative molecular field analysis (CoMFA). *Prog. Clin. Biol. Res.* **1989**, *291*, 161–165.
- (13) Protein Data Bank. <http://www.rcsb.org/pdb/home/> (accessed March 3, 2012).
- (14) Moitessier, N.; Therrien, E.; Hanessian, S. A method for induced-fit docking, scoring, and ranking of flexible ligands. Application to peptidic and pseudopeptidic beta-secretase (BACE 1) inhibitors. *J. Med. Chem.* **2006**, *49* (20), 5885–5894.
- (15) Thomsen, R.; Christensen, M. H. MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* **2006**, *49* (11), 3315–3321.
- (16) Colman, P. M.; Hoyne, P. A.; Lawrence, M. C. Sequence and structure alignment of paramyxovirus hemagglutinin-neuraminidase with influenza virus neuraminidase. *J. Virol.* **1993**, *67* (6), 2972–2980.
- (17) Shim, J. Y. Understanding functional residues of the cannabinoid CB1. *Curr. Top. Med. Chem.* **2010**, *10* (8), 779–798.
- (18) Benitez, P. A.; Cardenas, B. S. Dissection of functional residues in receptor activity-modifying proteins through phylogenetic and statistical analyses. *Evol. Bioinform. Online* **2008**, *4*, 153–169.
- (19) Xin, F.; Radivojac, P. Computational methods for identification of functional residues in protein structures. *Curr. Protein. Pept. Sci.* **2011**, *12* (6), 456–469.
- (20) Du, Q. S.; Long, S. Y.; Meng, J. Z.; Huang, R. B. Empirical formulation and parameterization of cation-π interactions for protein modeling. *J. Comput. Chem.* **2012**, *33*, 153–162.
- (21) Du, Q. S.; Liao, S. M.; Meng, J. Z.; Huang, R. B. Energies and physicochemical properties of cation-π interactions in biological structures. *J. Mol. Graph. Model.* **2012**, *34*, 38–45.
- (22) Kim, C. U.; Lew, W.; Williams, M. A.; Liu, H.; Zhang, L.; Swaminathan, S.; Bischofberger, N.; Chen, M. S.; Mendel, D. B.; Tai, C. Y.; Laver, W. G.; Stevens, R. C. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **1997**, *119* (4), 681–690.
- (23) Kim, C. U.; Lew, W.; Williams, M. A.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. Structure-activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J. Med. Chem.* **1998**, *41* (14), 2451–2460.
- (24) Du, Q. S.; Huang, R. B.; Chou, K. C. Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Curr. Protein. Pept. Sci.* **2008**, *9* (3), 248–260.
- (25) Du, Q. S.; Huang, R. B.; Wei, Y. T.; Wang, C. H.; Chou, K. C. Peptide reagent design based on physical and chemical properties of amino acid residues. *J. Comput. Chem.* **2007**, *28* (12), 2043–2050.
- (26) Lew, W.; Chen, X.; Kim, C. U. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. *Curr. Med. Chem.* **2000**, *7* (6), 663–672.
- (27) Dunn, C. J.; Goa, K. L. Zanamivir: a review of its use in influenza. *Drugs* **1999**, *58* (4), 761–784.
- (28) Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Ali, S. M.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E. A.; Finley, J.; Luo, M.; Air, G. M.; Laver, G. W. Potent inhibition of influenza sialidase by a benzoic acid containing a 2-pyrrolidinone substituent. *J. Med. Chem.* **1999**, *42* (13), 2332–2343.
- (29) Lew, W.; Wu, H.; Mendel, D. B.; Escarpe, P. A.; Chen, X.; Laver, W. G.; Graves, B. J.; Kim, C. U. A new series of C3-aza carbocyclic influenza neuraminidase inhibitors: synthesis and inhibitory activity. *Bioorg. Med. Chem. Lett.* **1998**, *8* (23), 3321–3324.
- (30) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37* (24), 4130–4146.
- (31) Tekiner, G. B.; Temiz, A. O.; Yıldız, I.; Aki, S. E.; Yalcın, I. 3D-QSAR study on heterocyclic topoisomerase II inhibitors using CoMSIA. *SAR. QSAR. Environ. Res.* **2006**, *17* (2), 121–132.
- (32) Smith, B. J.; Colman, P. M.; Von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein Sci.* **2001**, *10* (4), 689–696.
- (33) Liu, H.; Yao, X.; Wang, C.; Han, J. In silico identification of the potential drug resistance sites over 2009 influenza A (H1N1) virus neuraminidase. *Mol. Pharm.* **2010**, *7* (3), 894–904.
- (34) Bohm, M.; Klebe, G. Development of new hydrogen-bond descriptors and their application to comparative molecular field analyses. *J. Med. Chem.* **2002**, *45* (8), 1585–1597.

- (35) Du, Q.; Liu, P. J.; Mezey, P. G. Theoretical derivation of heuristic molecular lipophilicity potential: a quantum chemical description for molecular solvation. *J. Chem. Inf. Model.* **2005**, *45* (2), 347–353.
- (36) Du, Q. S.; Li, D. P.; He, W. Z.; Chou, K. C. Heuristic molecular lipophilicity potential (HMLP): lipophilicity and hydrophilicity of amino acid side chains. *J. Comput. Chem.* **2006**, *27* (6), 685–692.
- (37) Du, Q. S.; Mezey, P. G.; Chou, K. C. Heuristic molecular lipophilicity potential (HMLP): a 2D-QSAR study to LADH of molecular family pyrazole and derivatives. *J. Comput. Chem.* **2005**, *26* (5), 461–470.
- (38) Viswanadhan, V. N.; Ghose, A. K.; Hanna, N. B.; Matsumoto, S. S.; Avery, T. L.; Revankar, G. R.; Robins, R. K. Analysis of the in vitro antitumor activity of novel purine-6-sulfenamide, -sulfinamide, and -sulfonamide nucleosides and certain related compounds using a computer-aided receptor modeling procedure. *J. Med. Chem.* **1991**, *34* (2), 526–532.
- (39) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **1996**, *3* (10), 842–848.
- (40) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation energies of amino acid side chains and backbone in a family of host–guest pentapeptides. *Biochemistry* **1996**, *35* (16), 5109–5124.
- (41) White, S. H.; Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
- (42) Zhang, C. T.; Chou, K. C. An analysis of protein folding type prediction by seed-propagated sampling and jackknife test. *J. Protein Chem.* **1995**, *14*, 583–593.
- (43) Jiang, Y.; Iglini, P.; Kurgan, L. Prediction of protein folding rates from primary sequences using hybrid sequence representation. *J. Comput. Chem.* **2009**, *30* (5), 772–783.
- (44) Chou, K. C. Prediction of G-protein-coupled receptor classes. *J. Proteome Res.* **2005**, *4*, 1413–1418.
- (45) Liu, H.; Wang, M.; Chou, K. C. Low-frequency Fourier spectrum for predicting membrane protein types. *Biochem. Biophys. Res. Commun.* **2005**, *336*, 737–739.
- (46) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
- (47) Gineityte, V. A. Simple topological factor determining the allowance of pericyclic reactions. *Int. J. Quantum Chem.* **2008**, *108* (6), 1141–1154.
- (48) Estrada, E.; Uriarte, E. Recent advances on the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8* (13), 1573–88.
- (49) Brooks, W. H.; Guida, W. C.; Daniel, K. G. The significance of chirality in drug design and development. *Curr. Top. Med. Chem.* **2011**, *11* (7), 760–70.
- (50) Cintas, P. Tracing the origins and evolution of chirality and handedness in chemical language. *Angew. Chem., Int. Ed.* **2007**, *46* (22), 4016–4024.
- (51) Barone, G.; Duca, D.; Silvestri, A.; Gomez-Paloma, L.; Riccio, R.; Bifulco, G. Determination of the relative stereochemistry of flexible organic compounds by Ab initio methods: conformational analysis and Boltzmann-averaged GIAO ^{13}C NMR chemical shifts. *Chemistry* **2002**, *8* (14), 3240–3245.
- (52) Tsuzuki, Y.; Tanabe, K.; Akagi, M.; Tejima, S. The Rotatory Dispersion and Stereochemistry of Organic Compounds. VII. Monosulfides and Disulfides Derived from Glucose and Glucosamine. *Bull. Chem. Soc. Jpn.* **1965**, *38*, 270–274.
- (53) Pham, T. A.; Jain, A. N. Parameter estimation for scoring protein-ligand interactions using negative training data. *J. Med. Chem.* **2006**, *49* (20), 5856–5868.