Structure-Based CoMFA As a Predictive Model - CYP2C9 Inhibitors As a Test Case

Kazuya Yasuo,*,† Noriyuki Yamaotsu,‡ Hiroaki Gouda,‡ Hideki Tsujishita,† and Shuichi Hirono‡

Discovery Research Laboratories, Shionogi & Co., Ltd. 12-4, Sagisu 5-Chome, Fukushima-ku, Osaka 553-0002, Japan, and School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8641, Japan

Received September 2, 2008

In this study, we tried to establish a general scheme to create a model that could predict the affinity of small compounds to their target proteins. This scheme consists of a search for ligand-binding sites on a protein, a generation of bound conformations (poses) of ligands in each of the sites by docking, identifications of the correct poses of each ligand by consensus scoring and MM-PBSA analysis, and a construction of a CoMFA model with the obtained poses to predict the affinity of the ligands. By using a crystal structure of CYP 2C9 and the twenty known CYP inhibitors as a test case, we obtained a CoMFA model with a good statistics, which suggested that the classification of the binding sites as well as the predicted bound poses of the ligands should be reasonable enough. The scheme described here would give a method to predict the affinity of small compounds with a reasonable accuracy, which is expected to heighten the value of computational chemistry in the drug design process.

INTRODUCTION

Computational chemistry has been used widely in pharmaceutical research, in the hope of facilitating the drug design process. For further success of computational chemistry in this field, one of the critical factors would be an accurate prediction of affinities of small molecules (ligands) to their target proteins. In this view, many efforts have been made for decades to develop predictive models that could accurately estimate the affinities of small molecules.

As a part of such efforts, the free energy perturbation (FEP) method[1] and the thermodynamic integration (TI) method[2] were actively studied in 1980s. In theory, these methods are able to estimate the affinity of ligands (difference of binding free energy) precisely. However, those methods require a long computation time to complete a sufficient search of the configuration space, which makes it difficult to use those methods in practical drug design processes at pharmaceutical industries. In addition, due to some inaccuracies residing in the force field, those methods do not always estimate the binding affinities correctly. For the above reasons, those methods have not come to use broadly in the field of drug design.

As an alternative of FEP or TI methods, MM-PBSA (molecular mechanics with Poisson−Boltzmann surface area) analysis has been proposed recently. This method was originally introduced by Kuhn et al.,[3] to estimate the binding free energy in a more efficient manner than FEP or TI. In their paper, this MM-PBSA method was used in an attempt to select the most energetically stable pose among docking solutions. In the formulation of MM-PBSA, the net change in the binding free energy accompanying the formation of a protein−ligand complex is estimated by the following equation

$$\Delta G_{\mathrm{bind}} = \Delta E_{\mathrm{MM}} + \Delta G_{\mathrm{solv}} - T\Delta S_{\mathrm{solute}} \quad (1)$$

where ΔE_{MM} is the change in molecular mechanics (MM) energy expressed as the sum of the internal, electrostatics, and van der Waals contributions to binding in vacuo, ΔG_{solv} is the net change in solvation free energy upon binding expressed as the sum of polar and nonpolar solvation free energies ($\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{psolv}} + \Delta G_{\mathrm{npsolv}}$, respectively) (eq 2), T is the absolute temperature of the system, and $\Delta S_{\mathrm{solute}}$ is the internal entropy change of the solute upon binding. In MM-PBSA analysis, however, the estimation of the entropy difference in solute ($\Delta S_{\mathrm{solute}}$) often has a doubtful accuracy because it is generally estimated with a normal-mode analysis of harmonic frequencies calculated at the molecular mechanics level, which sometimes makes it difficult to evaluate ΔG_{bind} of structurally different ligands.[63]

On the other hand, so-called "scoring functions" have been also proposed, as more simple and rapid ways for the estimation of the ligand affinity. Most of the scoring functions are designed to calculate the binding affinity from a single three-dimensional structure of a protein−ligand complex. Some of the scoring functions are force-field based, where the binding affinity is calculated from the interactions between a protein and a ligand by using molecular mechanics.[4−7] In other scoring functions, coefficients of function terms are determined empirically so as to reproduce the actual binding affinities of a set of ligands.[8−13] There are so-called "knowledge-based" scoring functions as well, such as PMF (potential mean force)[14] or DrugScore,[15] which are calculated from the probability of the existence of specific atom types in the ligand at the specific positions in the protein cavity. Those probabilities are determined by a statistics analysis of existing protein−ligand complexes in the Protein Data Bank. Those scoring functions are now used widely for compound database screening (virtual screening), because those scores can be computed in a relatively short time. However, while those scoring functions are useful for

* Corresponding author e-mail: kazuya.yasuo@shionogi.co.jp.
† Shionogi & Co., Ltd.
‡ Kitasato University.

10.1021/ci800313h CCC: $40.75
Published on Web 03/10/2009

identifying potential binders to a target protein among many compounds, their accuracy in predicting the binding affinity is usually very limited.

As briefly summarized above, a method for the accurate prediction of the binding affinity of small ligands with a marginal time and cost still needs to be developed. In addition to the methods mentioned above, which are based on the structures of protein−ligand complexes, another choice for predicting the ligand affinity would be to utilize a QSAR (quantitative structure−activity relationship) approach. In this method, a model to describe the affinities of a set of compounds (in many cases, their biological activities such as IC50 or *K*i) is built by a multivariable regression between the affinities and physicochemical properties of existing compounds. Then the resulting model can be used to predict the affinity (biological activity) of other compounds that are not included in the original model building. 3D-QSAR is an extension of QSAR, in which some properties derived from three-dimensional structures of compounds are utilized in the model building. For example, in the CoMFA (comparative molecular field analysis) method,[16] the three-dimensional structures of the compounds in a training set (i.e., compounds to be used for the model building) are aligned to each other in three-dimensional space so as to reflect their relative orientations in the ligand binding site on a target protein. Then, the interaction energies between each of the compounds and the surrounding space are calculated, by placing atomic probes on a grid around the aligned compounds. The quantitative relationships between those interaction energies and the affinities of the compounds can be derived by a multivariable regression such as PLS (partial least-squares),[17] by which one can identify the important interactions between the compounds and the surrounding grids. The derived model can be used to predict the affinity of a new compound as well. In a CoMFA model, those interaction energies with the surrounding probes ("molecular field") can be considered as a mimicry of the ligand binding site of a target protein. If the compounds are aligned so as to reflect their bound conformations and orientations in the ligand binding site, the molecular field could represent the environment of the binding site correctly. With such a molecular field, the CoMFA model would become a reliable tool for the prediction of the affinities of any ligands that bind to the target binding site. This means that the determination of appropriate alignments of ligand molecules would be critical for an accurate CoMFA model building, and once we could establish a scheme for obtaining the correct molecular alignment, we could then use the CoMFA method as a predictive model for the binding affinities of ligands. In this paper, we will describe our attempts to develop such a scheme to produce an appropriate molecular alignment, by using cytochrome P450s (CYPs) inhibition as a test case.

CYPs are a large group of monooxygenase enzymes, which have broad substrate specificities for metabolizing generally small, lipophilic molecules such as drugs into more water-soluble metabolites.[18] A drug that has CYP inhibitory potential may cause problems when it is administered together with other drugs, namely potentially dangerous drug-drug interactions can occur.[18] Enzyme assays of CYP inhibition have been become possible after various CYPs were isolated in 1990s. To eliminate potential CYP inhibitors early in the development stages, various computational methods based on experimental data have been attempted. It has become increasingly important to predict CYP inhibition activities of candidates in drug discovery processes.

As a part of such efforts, some CYP binding models for 3D-QSAR CoMFA/CoMSIA (comparative molecular similarity indices analysis)[64] have been reported. An overview of recently derived models for CYPs can be found in several reviews.[19,20] One QSAR model for CYP2C9 inhibition utilized 27 compounds. In this paper, the molecular alignment of the compounds was done by using the ligand structures only, where the 27 compounds were aligned with their common structures of 9(*S*),11(*R*)-cyclocoumarol and not by the sites of oxidation, for model building. The obtained model predicted the presence of an aromatic binding region and a sterically forbidden region.[21]

Another CoMFA/CoMSIA analysis also utilized the ligand-based molecular alignment, where the minimized structures of 17 benzbromarone-related compounds were superimposed by aligning the benzylic ring of (*S*)-warfarin, phenyl rings of the compounds, and the common benzofuran rings between the C-7 and C-6 positions of the warfarin fused ring. The model suggested the presence of a hydrophobicity-favorable region around the phenyl rings of analogues and a sterically forbidden region around the alkyl residues of the benzofuran ring.[22]

In most 3D-QSAR models, the molecular alignments have been done based on ligand structures. One limitation of such a "ligand-based" model is that only structurally similar molecules can be used in the model building, because common substructures in the molecules are used as a guide to make up the molecular alignment in most cases. Moreover, a model derived from structurally similar molecules would have poor ability to predict the activity of a molecule that is not similar to those used in the model building. This would cause a significant drawback in modeling CYP binding and any other proteins that have broad substrate/ligand specificities. For such proteins, another approach for the molecular alignment would be more suitable for obtaining a predictive 3D-QSAR model that would be applicable to structurally different ligands.

Recently, the X-ray structures of human CYP 2C9 were solved by Wester et al. (PDB code: 1R9O)[23] and by Williams et al. (PDB code: 1OG2 and 1OG5).[24] These X-ray structures enable us to more generally discuss the bound conformations of various substrates and inhibitors, regardless of the similarity in their chemical structures. Namely, if the conformations and orientations of ligands in the binding pockets in CYP 2C9 are determined correctly by computational techniques with those X-ray structures, an appropriate set of the aligned molecules could be obtained, which would lead to a predictive 3D-QSAR model for CYP inhibition.

In this study, we developed a computational scheme to determine the bound conformations of various CYP 2C9 inhibitors in the X-ray structure of the protein. This scheme consists of the following: (1) ligand-binding pocket exploration on the target protein, (2) generation of the probable binding conformations (poses) by docking the ligands into each of the detected pockets, and (3) selection of the best one from the obtained poses of each ligand molecule by docking scores and MM-PBSA analysis. Finally, CoMFA

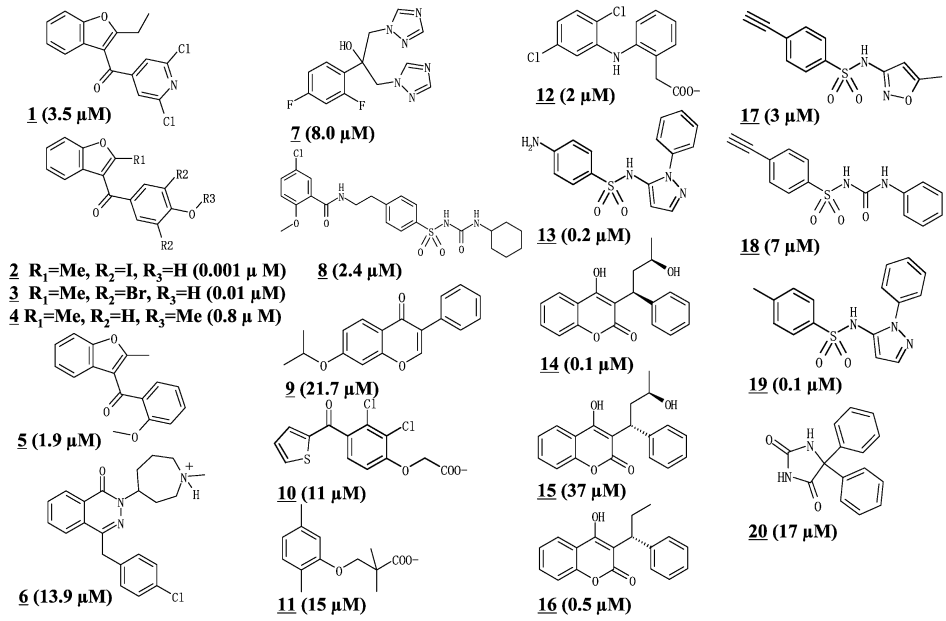


Figure 1. Chemical structures of the twenty CYP 2C9 inhibitors used in this study. The values in the parentheses indicate *K*i of the inhibitors.

models were built by using the best poses of the ligands, to see whether the expected quality of models could be achieved or not.

As mentioned in the beginning of this section, the computational affinity prediction models based on the protein−ligand interactions have had a rather limited success. This fact probably indicates that the direct computation of the ligand affinity from the protein−ligand interactions would be a very difficult problem. The intention of our work is to use CoMFA or other 3D-QSAR techniques, rather than a direct computation, for the affinity prediction, and to focus on the development of a computational scheme to produce the correct alignment of ligands by using a protein structure. This is because the correct molecular alignment is mandatory to obtain a predictive CoMFA model. Once this scheme is established successfully, we could come to obtain predictive models for ligand affinities, which would dramatically heighten the value of computational chemistry in drug design process.

METHODS

Overall Procedure. We first collected 20 CYP 2C9 inhibitors with their inhibitory activities on (*S*)-warfarin hydroxylation. To determine the correct binding conformation of each molecule, we adopted the following procedure. First, the CYP 2C9 X-ray structure was investigated with a pocket-searching program (HBOP and HBSITE, see below) since CYP 2C9 might have multiple binding pockets. We then tried to dock the ligands into each of the identified pockets with the protein structure kept fixed but could not obtain docking solutions for many of the ligands, which suggested that there would be conformational changes in the protein upon ligand binding. Therefore, an MD simulation with data for explicit water molecules was carried out in order to sample plausible conformations of the pockets on ligand binding. Next, a docking study for each compound was performed for all of the binding pockets of the sampled CYP 2C9 structures, and the most energetically stable pose determined with the MM-PBSA analyses was employed as the actual binding conformation. Finally, a CoMFA model was constructed with the alignment composed of the overlaid poses. More details are described below.

Ligand Data. The ligand structures used in this study are shown in Figure 1. These compounds do not clearly have common structures over all, and it is expected to be difficult to do a usual ligand-based alignment for those compounds. In this work, we checked how the structure-based alignment we developed should work for this kind of data set. The data set was used to explore variation in activity over 4 orders of magnitude[22,25−28] and is therefore appropriate for establishing statistically significant relationships between experimental and predicted *K*i values.

CYP 2C9 Structure Preparation. There are two crystal structures[23,24] of human CYP2C9 in the Protein Data Bank. We selected the crystal structure of 1R9O because it contains no mutated residues in its amino-acid sequence. There are four loop regions not assigned in the X-ray structure. We modeled two of those regions using the SYBYL[29] Biopolymer module. Namely, residues 38−42 and 214−220 were built using the residues 20−24 and 196−202 of the human CYP 2C8 X-ray structure (PDB code 1PQ2) which have a homology with the corresponding regions of CYP2C9 (Figure 2). The other two regions were not modeled and were capped with an acetyl moiety for the N-terminal nitrogen atom and an *N*-methyl moiety for C-terminal carbonyl carbon atom because these atoms were originally members of peptide bonds and not charged.

Binding Pocket Search. The structure prepared above was explored with HBOP and HBSITE programs[30] developed by our laboratory for potential ligand-binding sites on the protein. These programs predict the ligand-binding pockets according to the hydrophobic-energy function proposed by Israelachvili and Pashley, eq 3.[31]

$$\Delta G_H = -2.0R_{ij}\exp(-D_{ij}/10) \qquad (3a)$$

$$R_{ij} = R_i R_j/(R_i + R_j) \quad (3b)$$

$$D_{ij} = d_{ij} - (R_i + R_j) \quad (3c)$$

R_i and R_j are the radii of the carbon atom of the protein and the probe on the grid point, respectively, which of a lattice is generated around the protein surface and the thickness of the lattice was 10 Å, and the grid spacing was 1 Å. The probe was the sp3 carbon atom, and R_j was equal to 1.52 Å. d_{ij} was the distance between the carbon atom of the protein and the grid point.

Only the carbon atoms, except for the amide carbons included in the hydrophobic residues (Gly, Ala, Val, Leu, Ile, Met, Trp, Phe, and Pro), were used to calculate the hydrophobic potential. The carbon atoms that did not belong to the protein (i.e., small molecules such as ligands and cofactors) were not used for the calculation. A radius of each atom was taken from that of corresponding Tripos atom type implemented in SYBYL 6.8.1.[65] HBOP outputs the coordinates of the grid points as well as the hydrophobic potentials on the grid points, which are described as 20 hydrophobic levels. Those values are used in the subsequent analysis by HBSITE, a utility program for grid clustering.

HBSITE uses three parameters: the lowest hydrophobic level to be used (low_{HB}), the radius for grid clustering (r_{grid}), and the minimum volume (min_{vol}; that is, the minimum number of grid points to be included in a cluster). In this study, low_{HB}, r_{grid}, and min_{vol} were set to "HB5" (the fifth lowest level), 1.1 Å (slightly larger than the value of grid spacing), and 10 grid points, respectively. Before the clustering, the grid points with hydrophobic levels less than low_{HB} were rejected. If a distance between two grid points was within r_{grid}, those two points were assigned to the same cluster. All of the grid points were grouped into clusters in this way. The grid points that had not been assigned to any cluster were eliminated, and only the clusters that had grid points more than or equal to the min_{vol} were kept. HBSITE then made a new grid file that contains the grid points that belong to those clusters. In our laboratory, it has been shown that HBOP and HBSITE programs were able to predict the ligand-binding pockets for various kinds of proteins.[30]

With this analysis, we found two pockets (called Site A and Site B, see below) on the CYP 2C9 crystal structure, and these pockets were subjected to the following steps.

MD Simulation. Initial attempts of docking of the CYP 2C9 inhibitors to the binding pockets identified with HBOP and HBSITE had failed for a half of the ligands, when using the protein X-ray structure as is (see below). It was thought that there might be a possibility of conformational changes associated with ligand binding about the pockets, considering the fact that CYP2C9 could adopt various types of substrates and inhibitors.[32] Hence, to obtain possible conformations of the binding pockets of CYP2C9 in solution, we performed an MD simulation using explicit water molecules with SANDER in the AMBER (v8.0) simulation package.[33] The force field was taken from ff03.[34] We removed all crystallographic water molecules and added a 10 Å layer of surrounding water molecules using the TIP3P water model.[35] Several monovalent counterions were used to neutralize the system. The system was then subjected to a steepest-descent minimization for 100 steps to remove unfavorable steric contacts. We then performed an MD simulation only for the water molecules and the counterions, while the protein part was kept fixed. Next, the whole system was subjected to a steepest-descent minimization for 100 steps, followed by a conjugate gradient minimization until the root-mean-square of gradient became less than 0.05 kcal/mol Å. The protein structure finally obtained was used for the following equilibration and production dynamics.

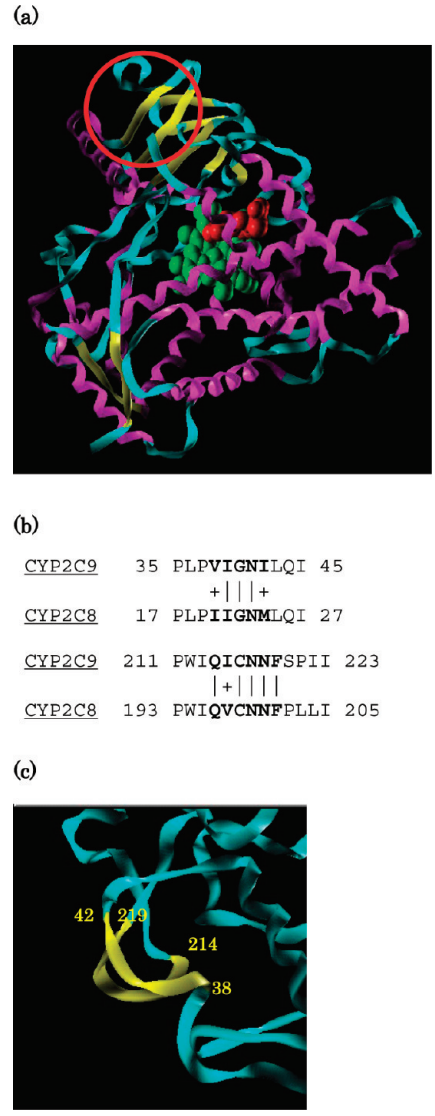


Figure 2. (a) The crystal structure of CYP 2C9 (PDB ID: 1R9O). Only the main chain is shown as ribbon representation. The color scheme is as follows: yellow for β-sheets, magenta for α-helixes, and cyan for loops. Heme and the bound substrate flurbiprophen in the catalytic site are shown in red and green spheres, respectively. The red circle indicates the two regions which lack in the X-ray structure. The structures of these regions were modeled in this study (see text). (b) The sequence alignment of the modeled regions in CYP 2C9 with homological sequences of CYP 2C8 X-ray structure (1PQ2). (c) The structure of the modeled regions of CYP2C9. The modeled regions are shown in yellow ribbon, and the numbers indicate the amino acid numbers at both ends of the regions.

During the equilibration and production dynamics, all of the bond lengths between hydrogen atoms and other heavy atoms were constrained by SHAKE[36] which was required when using a water model such as TIP3P. The numerical integrator used was the leapfrog Verlet algorithm.[37] Periodic boundary conditions were enforced with a time step of 1 fs. The cutoff distance was 10.0 Å for the nonbonded interactions with Particle Mesh Ewald method[38,39] for long-range

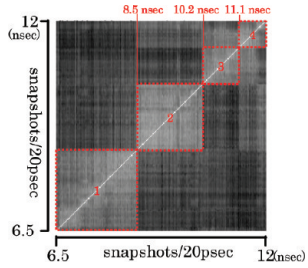
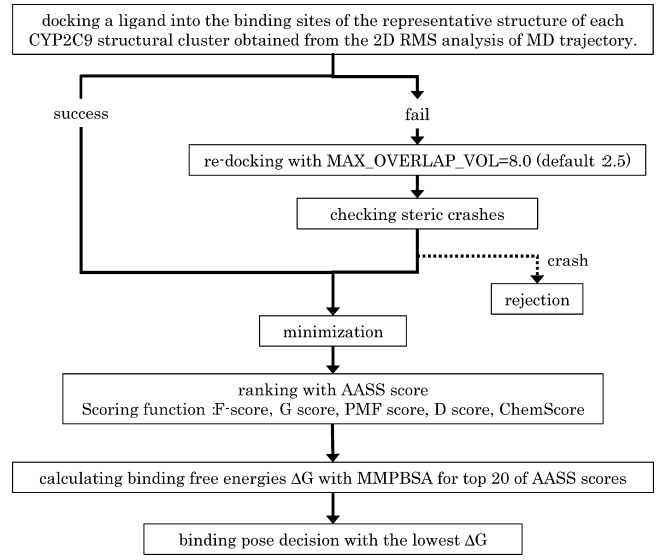


Figure 3. The result of 2D rmsd analysis with the trajectory (6.5 to 12 ns) of the MD simulation. The vertical and horizontal axes indicate the MD snapshots, and each point on the graph indicates the similarity (rmsd) of two snapshots: structurally similar snapshots with lower rmsd values are indicated by a brighter color in the matrix. Squares enclosed by dotted, red lines express the clustered protein structures of the MD simulation, and the number in each of the squares indicates the cluster number. The simulation time-points are also shown. There are four clusters as shown above.

electrostatic interactions. The equilibrate dynamics was started from the minimized configuration and performed by gradually heating from 0 to 300 K for 20 ps under constant-pressure (1 bar) conditions using an Andersen piston.[40] The final structure from the equilibration was passed on to a production MD that was under constant-pressure condition and maintained at the target temperature of 300 K for 12 ns using the Berendsen weak-coupling scheme.[41] During the production MD, we saved snapshots of the system at every 1000 steps along the trajectory and used them for the postproduction analysis described below. The MD part of the calculation took approximately 10.6 min per ps CPU time, when running parallel on eight PowerPC G5 2.5 GHz processors.

MD Trajectory Clustering. We analyzed the MD trajectory with PTRAJ from AMBER to investigate the amino-acid residue conformations included in the two binding pockets (Site A and Site B) found with HBOP and HBSITE. For the first step, we checked the trajectory and observed that trajectories between 6.5 to 12 ns were equilibrated for the energetic fluctuation. Equilibration of the total energy means that the protein structure in a system moves within a fixed range of energy in the simulated solution, namely, solution structures could be obtained from the trajectories of 6.5 to 12 ns. Hence, we examined the conformations of the binding pocket residues included in these trajectories with 2D-rms analysis using PTRAJ. This function allows us to calculate the root-mean-square of the coordinate file with respect to another coordinate file in the same trajectory file as a function of time. In this analysis, the rms values are first calculated between each snapshot of a trajectory in a 'round-robin' system. In this step, atoms or residues used for the calculation can be defined; the residues used were those within the binding pockets predicted with HBOP and HBSITE. Next is the creation of the matrix of rms sorted by the simulation time (namely, *x*-axis and *y*-axis are the simulation time and each element in the matrix has the value of rms); it is expressed visually in a PostScript file. In this form, rms is expressed in a gradationally colored form around the value: lower rms values are whiter and higher values are darker. Hence, the elements between the same snapshots on a diagonal line in the matrix are white in color (that is, they have the lowest rms value) (Figure 3). The snapshots were clustered based on the colored matrix, and the representative structure of each area was extracted from the MD trajectory as a snapshot whose element (rms value) was the closest to the average of rms values included in the area. These representative structures were minimized using SANDER. The first 100 steps of the minimization were conducted by the steepest-descent method, and the rest of the calculation was carried out using the conjugate-gradient method, until the energy gradient became less than 0.05 kcal/mol Å. The refined structures were used for the following computational docking studies.

Scheme 1. Schematic Representation of the Docking Poses Determination Procedure



Computational Docking. The protocol to obtain the poses is shown in Scheme 1. We used the FlexX program (Tripos Inc.)[42] in these docking studies for each predicted cavity of the classified structures and obtained up to 100 poses. The amino acid residues within 4 Å of Site A and Site B were used for the FlexX docking. In a case where no result could be obtained with the docking study described above, we employed the MAX_OVERLAP_VOL parameter multiplied by 8.0, which controls the maximum overlap van der Waals volume between the protein atoms and the docked ligand atoms to soften the van der Waals contacts between the protein and the ligand. The poses obtained were used for the next pose determination step.

Pose Determination. For the structure-based alignment for CoMFA, plausible ligand binding conformations are needed. Usually, several poses are obtained in a docking calculation for one ligand, and when there are a sufficient number of poses, it is expected that the correct pose should be included in those docking poses. We next tried to determine the actual pose of each inhibitor from the results

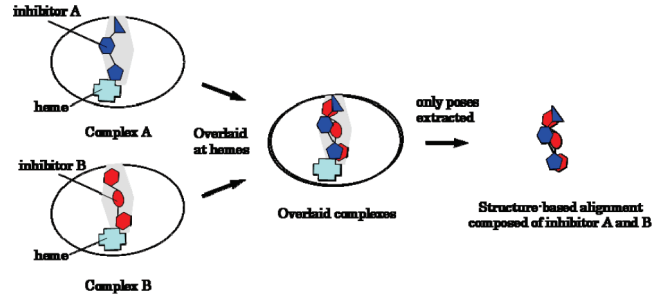


Figure 4. A schematic representation of the structure-based alignment construction.

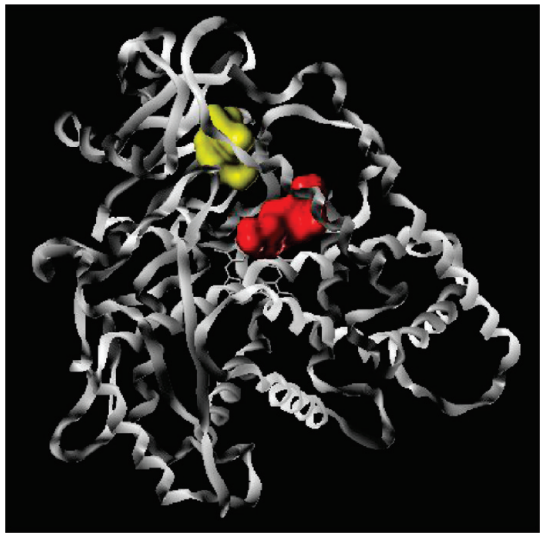


Figure 5. The two binding sites identified by HBOP. The resulted grids are shown in solid surfaces: Site A (red) and Site B (yellow). The main chain of CYP2C9 is shown in ribbon. The heme molecule is shown in gray capped-stick.

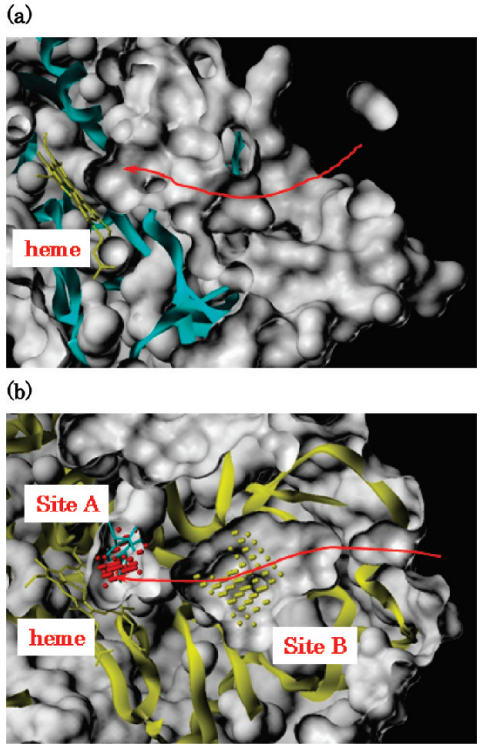


Figure 6. (a) The X-ray structure of mammalian CYP 2C5 (PDB code: 1DT6). The plausible substrate access route (see text) is indicated by a red arrow. The protein is depicted as a sliced gray surface, and the main chain is shown in ribbon. (b) The CYP2C9 structure used in this study. The red and yellow dots indicate Sites A and B, respectively. From the comparison with the CYP 2C5 structure, Site B is suggested as a part of the substrate access channel (red arrow).

of docking studies (that is, the candidates of the correct pose) for Site A and Site B. For the determination of the correct poses, we employed MM-PBSA analysis, but this method generally consumes computational cost. Therefore, we first attempted to limit the probable candidates for MM-PBSA analysis using scoring functions. Although many scoring functions have been introduced as mentioned above, each type has different strengths and weaknesses, often making the selection of the optimal scoring function a challenging task. To overcome these shortcoming, a "consensus scoring" method has been introduced for evaluating docking models.[42,43] Consensus scoring methods are promising as they can potentially improve performance by merging results from different functions (namely, a product set). Consequently, we utilized both a consensus scoring method, AASS (Average Auto-Scaled Score),[44] and MM-PBSA analysis for the determination of poses. The details are described below.

AASS Scoring. We used AASS (Average Auto-Scaled Score) scoring[44] for the first selection of the pose candidates about each classified pocket. AASS includes five scoring functions: FlexX score,[10] Gold score,[4] PMF score,[14] Dock score,[15] and ChemScore.[11] These five scores were independently calculated for the docking models and awarded marks of one to five for the validities of the models. AASS scoring was estimated with the following equation

$$x^{\mathrm{AASS}} = \left(\sum_{\mathrm{i\_Score}}^{n} \frac{x^{\mathrm{i\_Score}} - x_{\mathrm{min}}^{\mathrm{i\_Score}}}{x_{\mathrm{max}}^{\mathrm{i\_Score}} - x_{\mathrm{min}}^{\mathrm{i\_Score}}} \right) / n \qquad (4)$$

where $x^{\mathrm{i\_Score}}$ is the value of each score, $x^{\mathrm{i\_Score}}{}_{\mathrm{max}}$ and $x^{\mathrm{i\_Score}}{}_{\mathrm{min}}$ are the maximum and minimum scores, respectively, and n is the number of scores used. In AASS scoring, each score was normalized between 0 and 1 for its deviation, and the sum of five normalized scores was averaged. The smaller the AASS score, the more reasonable is the ligand-docking mode. We employed the top 20 poses for the subsequent MM-PBSA analysis. If the total number of poses was less than 20, we used all poses for the analysis.

MM-PBSA Analysis. As the second evaluation of the poses selected above, we employed an MM-PBSA analysis to estimate the binding free energy and select the most energetically stable pose among the docking solutions.

As described above, when we try to estimate accurate ΔG_{bind} values of each protein−ligand complex, the calculation of $\Delta S_{\mathrm{solute}}$ is required, but this step generally takes a high computational cost and a long calculation time. Recently, some reports have described the possibility of estimating protein−ligand binding affinities with a reasonable accuracy even if the $\Delta S_{\mathrm{solute}}$ term of the original MM-PBSA methodology is ignored.[45,46]

In this study, we used this modified formulation of MM-PBSA. In our study, MM-PBSA was used to compare the different conformations of the same molecules bound to the protein, and in such a case, $\Delta S_{\mathrm{solute}}$ could be almost the same among the conformation because $\Delta S_{\mathrm{solute}}$ does not depend much on conformations of molecules. Thus the contribution of $\Delta S_{\mathrm{solute}}$ to ΔG_{bind} would be negligible if the different conformations of the same molecule are compared.

The MM-PBSA analysis was applied to the energy-minimized complexes of the 20 docked inhibitors. For the minimization of complexes, we used the standard AMBER charges for the protein and the AM1-BCC[47] charges for the heme and the ligands. AM1-BCC is very efficient for calculating partial charges, which makes it ideal for use for large libraries of compounds. The force field parameters for the small molecules are taken from the generalized AMBER force field (GAFF).[48] The assignments for AM1-BCC partial charges and atomic force field parameters were done using ANTECHAMBER (v1.0).[49] The polar solvation free energies

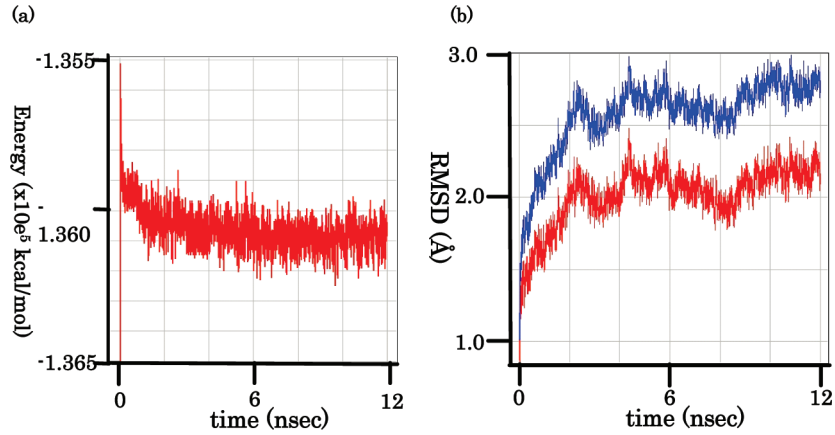


Figure 7. (a) The total energy changes of the system in the MD simulation. (b) Atomic rmsd changes from the initial structure in the MD simulation. rmsd calculated with all protein atoms is indicated in red, and the value calculated with main chain is in blue.

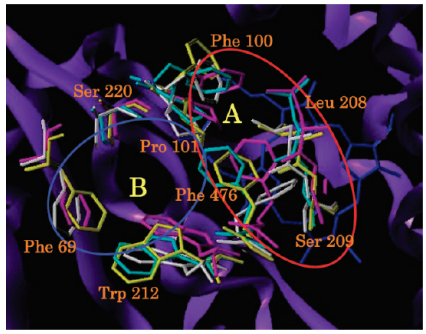


Figure 8. The residues which have different conformations in four representative CYP 2C9 structures: white, capped-stick expresses the first representative structure of Figure 3; yellow, the second structure; magenta, the third structure; and cyan, the fourth structure. The red and blue circles represent Site A and B, predicted with HBOP. CYP 2C9 and the heme molecules are expressed in ribbon and blue capped-stick, respectively.

were calculated by solving the Poisson−Boltzmann (PB) equation with the AMBER PBSA module.[50] We used the Poison-Boltzmann solver of AMBER with atomic cavity radii and atomic charges taken from the AMBER topology files (AMBER ff03 force field). The grid spacing was set at 0.5 Å, and dielectric constants of 1 and 80 were used for the interior and exterior of the protein, respectively.

The nonpolar contribution (ΔG_{npsolv} in eq 4) to the solvation free energy was calculated with methods dependent on the solvent accessible surface area (SASA)[51] using the equation $\Delta G_{npsolv} = \gamma SASA + b$. We used Molsurf[51] with $\gamma = 0.005$ kcal mol^{-1} Å^{-2} and $b = 0.0$ kcal mol^{-1} in combination with AMBER PB polar solvation energies. These parameters are well documented in the literature.[52−59] In this step, the pose that had the smallest binding free energy was finally accepted as the correct pose.

Comparative Molecular Field Analysis (CoMFA) Modeling. In this study, we employed the alignment for CoMFA as the overlaid poses which were determined with the procedures described above (so-called "structure-based CoMFA"). The alignment was obtained as follows (Figure 4): (1) the protein−ligand complexes were overlaid at the heme molecules and (2) the proteins and the hemes were removed from the overlaid structures and the remaining structures, namely, the overlaid poses of the ligands, were employed as the alignment.

In this study, we constructed a CoMFA predictive model over the two different pockets (namely, Site A and Site B). That is, the ligands that were predicted to bind to the different pockets were included in one model. The aim of this study is to predict the binding affinities of compounds that bind to any of binding pockets on a protein. Therefore, we thought that it would be better if one model could explain the affinities of all ligands even when they bind to different binding pockets. As we employed a protein structure-based alignment in this study, it was expected to be able to obtain such a CoMFA predictive model if the poses of the ligands in appropriate pockets were identified correctly.

All CoMFA analyses in this paper used the standard defaults. The analysis was performed using the CoMFA module in the SYBYL program package. The QSAR model was determined by partial least-squares (PLS) analysis[17] of the inverse natural logarithm of biological activity (an inhibitor *K*i value) versus CoMFA electrostatic and steric fields. The obtained CoMFA models were validated by cross-validation to exclude models with poor prediction ability.

RESULTS AND DISCUSSION

Active Site Characteristics. We first investigated the X-ray structure with HBOP and HBSITE to search for binding pockets because proteins such as enzymes and receptors do not always have only one ligand binding site. As a result, two pockets (Site A and Site B) were found (Figure 5). HBOP and HBSITE predict pockets according to the hydrophobic-interaction function only: this is based on the hypothesis that ligand-binding sites of a protein have a tendency to be more hydrophobic than other regions because hydrophobic ligands are more energetically stable in a protein−ligand complex. One of those two hydrophobic regions, designated as Site A, was located near the heme molecule in the catalytic site, and Site B was located near Phe69, Pro367, and Phe476. Actually, another reported X-ray structure of CYP 2C9 (PDB code: 1OG5) has a warfarin molecule bound in this site. With respect to Site B, the existence of substrate access channels around this site in various CYP structures has been suggested through use of random expulsion molecular dynamics (REMD) by Wade et al.[60] They carried out a REMD simulation of progesterone-mammalian CYP 2C5 complex and observed the ligand leaving the active site via the channel between the G and I helices and the BC loop. This channel is observed in an X-ray structure of mammalian CYP 2C5 (PDB code: 1DT6). CYP 2C5 has structural homology with CYP 2C9, and the overlaid

Table 1. Pose Numbers Obtained from the Docking Calculation for the Two Detected Binding Sites of the Four Representative CYP 2C9 Structures

	structure 1		structure 2		structure 3		structure 4			structure 1		structure 2		structure 3		structure 4	
ligand no.	Site A	Site B	Site A	Site B	Site A	Site B	Site A	Site B	ligand no.	Site A	Site B	Site A	Site B	Site A	Site B	Site A	Site B
1	3	4	1	0	0	1	5	4	11	8	45	57	2	41	17	21	0
2[a]	1	51	4	71	26	1	6	77	12	100	100	43	14	100	37	100	100
3[a]	66	67	7	36	40	6	13	95	13	3	25	6	40	9	4	19	4
4[a]	16	35	12	4	92	16	96	92	14	2	4	0	0	2	3	12	8
5	12	39	1	77	63	66	38	20	15	10	11	3	0	5	13	10	11
6	30	0	4	9	14	0	19	3	16	50	23	0	0	5	13	23	26
7	34	100	28	42	100	52	65	17	17	40	28	42	11	38	0	100	9
8[a]	0	0	7	15	88	24	100	0	18	23	22	35	2	41	8	26	40
9	8	0	0	0	7	0	6	0	19	5	10	3	0	10	0	11	4
10	47	43	0	1	17	27	16	37	20	34	100	28	42	100	52	65	17

[a] Docked with MAX_OVERLAP_VOL=8.0.

Table 2. Minimum ΔG Values Calculated with MM-PBSA among the Docking Poses of Each Ligand[a]

ligand no.	structure no.	sites	minimum ΔG (kcal/mol)	ligand no.	structure no.	sites	minimum ΔG (kcal/mol)
1	4	A	−62.13	11	4	A	−545.57
2	2	A	−54.12	12	1	B	−71.60
3	2	A	−55.87	13	2	A	−60.43
4	4	A	−61.82	14	4	A	−712.65
5	2	B	−56.86	15	1	B	−54.06
6	1	A	−55.97	16	4	A	−107.95
7	2	A	−64.10	17	3	A	−55.26
8	4	A	−141.58	18	1	B	−54.29
9	1	A	−59.26	19	2	A	−61.35
10	1	A	−91.83	20	3	A	−52.29

[a] The representative CYP 2C9 structure number and the binding site that gave the minimum ΔG of each ligand are also shown.

structure of CYP 2C5 and 2C9 shows that Site B might be a part of the active site access channel for CYP 2C9 substrates (Figure 6).

We first attempted to dock the twenty CYP2C9 inhibitors into the identified pockets of the crystal structure, but half of the inhibitors (**2**, **3**, **6**, **8**, **10**, **11**, **14**, **16**, **17**, **19**, and **20** in Figure 1) were failed to dock into them. We thought that some conformational changes in the pockets should occur upon binding of those inhibitors. Therefore, we next tried to do a sampling of possible conformations of the protein pockets with an MD simulation in explicit water.

The rmsd values from the initial structure and the total energies of the system during the MD simulation are shown in Figure 7. As indicated in this figure, the rmsd and the total energy became stable and equilibrated from about 6.5 ns. This means that the protein structure in the system became to move within a fixed range of energy in the simulated solution from that time. Therefore, we employed the trajectory from 6.5 to 12 ns for the next 2D rmsd analysis. Here, we focused on the movement of the amino acid residues included in Site A and Site B found with HBOP. The result of the analysis is shown in Figure 3. We clustered the MD trajectory into four areas based on this result. The representative structure of each of the four areas was extracted from the MD trajectory as a snapshot with its rms value closest to the average of the rms values of snapshots existing in that area. Among the four representative structures of CYP 2C9, the binding pockets (Site A and B) had similar shapes of the main chain (rmsd = 0.78 Å), but different conformations were observed in the side chains of several amino acids

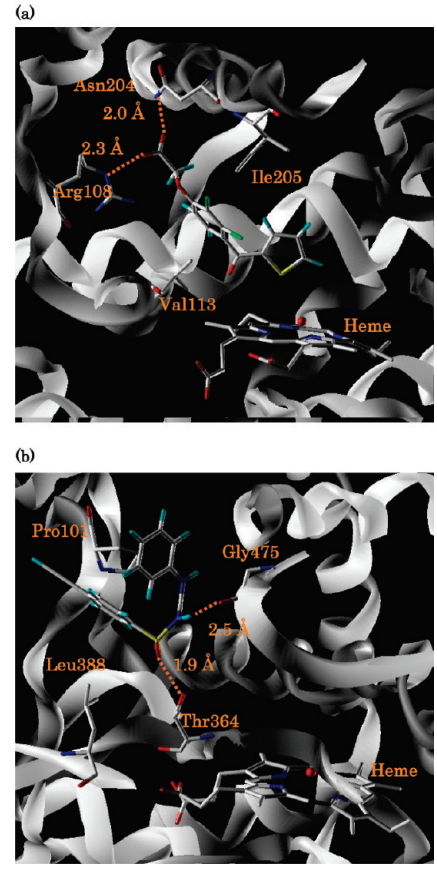


Figure 9. The examples of the docking poses determined with AASS and MM-PBSA. CYP 2C9 is expressed in ribbon and the heme molecule, and the interacting residues and ligands are expressed in capped-stick, respectively. (a) The pose of **10**. The compound interacts with Arg108 and Asn204 with the distances of 2.3 Å and 2.0 Å, respectively. (b) The pose of **18**. The compound interacts with Thr364 and Gly475 with the distances of 1.9 Å and 2.5 Å, respectively.

such as Phe69, Phe100, Trp212, and Phe476; in these residues, the aromatic rings in the side chain were flipped among the four protein structures (Figure 8). This observation would mean that those residues play an important role in the accommodation of various substrates, by changing their side chain conformations. As shown in Figure 8, the ligand-binding pocket was mainly composed of hydrophobic amino acids such as Phe, Trp, Pro, and Leu. The hydrophobic feature of the pocket was consistent with the known structure of other CYPs and might be related to the CYP function of

Table 3. Result of CoMFA

cross-validated			conventional			contribution (%)	
q^{2a}	$s_{press}{}^{b}$	no. of components	r^{2c}	F[d]	s[e]	ST[f]	EL[g]
0.728	0.679	3	0.941	84.707	0.317	60.4	39.6

[a] Predicted correlation coefficient. [b] Standard deviation for the sum of square predicted errors. [c] The conventional values of correlation coefficient. [d] F value. [e] Standard deviation. [f] The contribution values of steric contribution. [g] The contribution values of electrostatic contribution.

binding with hydrophobic substrates and their oxidative metabolization.

Pose Determination. Computational docking trials were carried out for the twenty ligands using FlexX. The two binding sites (Site A and Site B) of the four representative CYP 2C9 structures were used for the docking so that the effect of the conformational changes of the sites upon ligand-binding could be considered properly. As a result, eight docking calculations were performed per ligand. As shown in Table 1, the numbers of poses obtained for each ligand are very different among the representative protein structures or the binding sites (Site A and Site B), suggesting that there might be various ligand-binding modes for every ligand. The poses were evaluated by AASS scoring first, because the number of the poses was too great to allow a calculation of the binding free energy with MM-PBSA analysis. This consensus scoring method was developed in our laboratory, and it has been shown that the evaluation of the top twenty poses according to score-ranking was sufficient to capture the actual pose of a ligand.[44] We then estimated the binding free energies for the twenty poses with MM-PBSA, and the most energetically stable pose was assumed as the actual pose used for the following CoMFA.

Table 2 summarizes the representative protein structure and the binding pocket that gave the minimum MM-PBSA ΔG_{bind} for each ligand. It should be noted that the MM-PBSA calculations in this study did not include the entropy term ($-T\Delta S_{solute}$) and that the calculated values of ΔG_{bind} are not always necessary to correlate with the actual *K*i values of the inhibitors. There are some cases in which the difference of the entropy upon ligand binding makes a significant contribution to the ligand affinity. For example, a structurally flexible ligand experiences a bigger loss of entropy upon binding than a rigid ligand. Therefore, it is sometimes essential to consider the $-T\Delta S$ solute term for discussing the difference of the affinities of multiple ligands. In this study, however, the MM-PBSA analysis was used for the evaluation of different conformations of the same ligand. In such a case, the contribution of the entropy term would be almost the same among the conformations, and this term could be neglected in the calculation.

From the results shown in Table 2, sixteen of the twenty ligands were judged to bind at Site A near the heme molecule, and the four at Site B according to the MM-PBSA evaluation. These results suggest that there should be at least two types of inhibition: binding to Site A, which is a competitive inhibition, and binding to Site B, which is an unknown type of CYP inhibition, but as mentioned above, Site B is suggested to be a part of an access route to the catalytic site (Site A) of substrates and might result in inhibition of the substrate access to the heme.

The examples of the docking poses are shown in parts a (Site A) and b (Site B) of Figure 9. For the pose of **10** (Figure 9a), the carboxylic moiety of the ligand interacts with Arg108 and Asn204, and the phenyl ring of **10** is supported by Val113 and Ile205. Around Site B, compound **18** binds to the cavity and interacts with Thr364 and Gly475 (Figure 9b), where the position of the phenyl ring of the compound is kept by Pro101 and Leu388. Arg108 is suggested to play an important role in binding the substrates,[61,62] and lipophilic interactions such as Val113, Ile205, Pro101, and Leu388 are supposed to be important in binding substrates.

CoMFA Analysis. First, a structure-based alignment was made up of the most energetically stable poses obtained above (Figure 4). Interestingly, some of structurally similar inhibitors, for example, **14** and **15**, were judged to bind to the different sites respectively, while such compounds would be superposed on each other in a usual ligand-based alignment. Then a CoMFA model was constructed with that alignment. The statistics of the obtained model are shown in Table 3. As shown in Table 3, a statistically significant model, as judged by leave-one-out cross-validation, was obtained. The standard r^2 value for the nonvalidated model using three components was 0.941. The standard error of the estimate (s) was 0.317 and the F value, 84.697. A cross-validated r^2 (q^2) value of 0.728 indicated that this model has good predictive ability.[16,17] The contributions of the electrostatic (EL) and steric (ST) fields were 39.6% and 60.4%, respectively, which might explain the substrate preference of CYPs for hydrophobic compounds. The predicted values (*K*i) are shown in Table 4 and plotted against experimental values in Figure 10, which indicates that the predicted *K*i values have a good correlation to the experimental values.

CoMFA contour plots around Site A are shown in Figure 11. A yellow area (unfavorable steric feature) was observed near Leu208 and Leu233, and a green area (favorable steric feature) was observed around a glycerol molecule in the 1R9O crystal structure. There was sufficient space to accommodate a bulky group of inhibitors. A red area (a region preferring a negative charge) was observed around the positively charged ferrous atom position of the heme. In addition, a large blue area (a region preferring a positive charge) was observed near Phe90 and Phe114 where there was a region of relatively high electron density. The observed CoMFA contours in Site A well reflected the environment of the binding site, which suggests the obtained poses and alignment of the inhibitors would be reasonable.

On the other hand, no significant electrostatic or steric contours were observed around Site B. This might be ascribed to the fact that the number of inhibitors bound to Site B was too small to derive a significant contribution. One of our intensions was to make a single equation to predict the affinity of any compounds that bind to the same protein. Therefore, we had tried to build a model with compounds bound to different binding pockets: we thought it should give a reasonable result if the prediction of the bound conformations were done appropriately. Unfortunately, the obtained model did not give any significant contours around compounds in Site B. We have tried to build a model only with Site A compounds, which gave a comparable result (data not shown). These results indicate that the contribution from the Site B compounds was not significant, probably due to

Table 4. Calculated −log *K*i (calc) versus Experimentally Measured −log *K*i (expt) Values for the Twenty Ligands

		−log *K*i					−log *K*i		
no.	sites	expt	calc	error	no.	sites	expt	calc	error
1	A	5.46	5.75	−0.29	11	A	4.82	4.68	+0.14
2	A	9.00	8.30	+0.70	12	B	5.70	5.60	+0.10
3	A	8.02	8.19	−0.17	13	A	6.70	6.77	−0.07
4	A	6.10	6.06	+0.04	14	A	7.00	7.37	−0.37
5	B	5.73	5.58	+0.15	15	B	4.43	4.74	−0.31
6	A	4.86	4.42	+0.44	16	A	6.30	6.63	−0.33
7	A	5.10	5.05	+0.05	17	A	5.52	5.72	−0.20
8	A	5.62	5.50	+0.12	18	B	5.15	5.34	−0.19
9	A	4.66	4.48	+0.18	19	A	7.00	6.77	+0.23
10	A	4.96	4.74	+0.22	20	A	4.77	5.20	−0.43

an insufficient number of the ligands in this site. Further study should be needed to clarify this point.

In this study, we tried to establish a general scheme to create a model that could predict the affinity of small compounds to their target proteins. This scheme consists of a search for ligand-binding sites on a protein, a generation of bound conformations (poses) of ligands in each of the sites by docking, identifications of the correct poses of each ligand by consensus scoring (AASS) and MM-PBSA analysis, and a construction of a CoMFA model with the obtained poses to predict the affinity of the ligands. By using a crystal structure of CYP 2C9 and the twenty known CYP inhibitors as a test case, we suggested that this protein would have two different sites (Site A and B) for ligand binding. Furthermore, we classified the twenty inhibitors into Site A and Site B binders and predicted their bound conformations in those sites. In the final CoMFA model building, we used the bound poses of the ligands obtained above and included both the Site A and B binders into one model. As a result, we obtained a CoMFA model with a good statistics, which suggested that the classification of the binding sites as well as the predicted bound poses of the ligands should be reasonable enough. Strictly speaking, a good CoMFA model does not always ensure that the poses and the alignment of the ligands used in the analysis are correct: some experimental analysis such as X-ray crystallography would be needed to prove it. Nevertheless, in the obtained model, the properties of the binding pockets were well reflected in the CoMFA contour maps, which suggests that our scheme described here would have a reasonable accuracy.

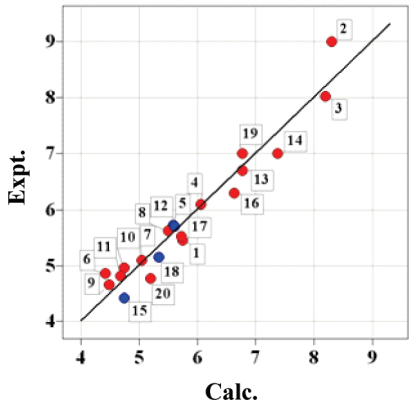


Figure 10. Plot between the predicted and experimental *K*i values (−log *K*i). The ligands bound to Sites A and B are indicated by red and blue, respectively. The compounds' numbers correspond to those in Figure 1.

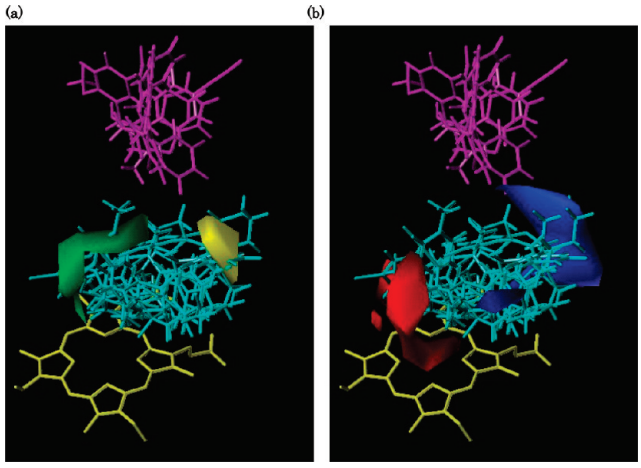


Figure 11. The contour maps of the obtained CoMFA model. (a) Steric contours: the yellow area indicates where steric interactions hinder binding, and the green area indicates where steric interactions enhance binding. (b) Electrostatic contours; the red areas indicate where more negative charge will enhance binding, and the blue areas indicate where positive charge will enhance binding. The molecules are expressed in capped-stick models, the heme is visualized in yellow for reference (not included in CoMFA), the ligands bound to Site A are expressed in blue, and the ones bound to Site B are expressed in magenta.

There are some cases in which some conformational changes in binding pockets of a protein occur upon ligand binding. Actually, half of the twenty inhibitors used in this study failed to dock into the crystal structure of CYP 2C9. We showed that a sampling of possible conformations of the protein pockets with an MD simulation in explicit water and a subsequent 2D rmsd analysis would be useful for such a case. This MD simulation would be a time-consuming step, but it needs to be done only once for one target protein: once representative protein structures are obtained through the MD analysis, one can then use these representative structures for docking of any ligands.

In conclusion, we established a scheme for the prediction of the affinities of small molecules, by combining the prediction of binding sites and bound conformations of ligands and 3D-QSAR. With the test case of CYP 2C9 inhibition, this scheme gave a reasonable accuracy in predicting the affinity of the inhibitors. In order to prove the further generality and robustness of the method, attempts to predict the affinities of other CYP 2C9 inhibitors with the model in this paper as well as applications of the scheme to other target proteins would be needed. Such studies are ongoing in our laboratory, and results will be published elsewhere.

REFERENCES AND NOTES

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

(2) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Inversion of receptor binding preferences by mutagenesis: free energy thermodynamic integration studies on sugar binding to L-arabinose binding proteins. *Biochemistry* **1993**, *32*, 7428–7434.

(3) Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, *43*, 3786–3791.

(4) Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

(5) Morris, G. M.; Goodsell, D. S.; et al. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(6) Ewing, T. J. A.; Makino, S.; et al. DOCK 4.0: Search Strategies For Automated Molecular Docking of Flexible Molecule Databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.

(7) Yin, S.; Biedermannova, L.; Vondrasek, J.; Dokholyan, N. V. MedusaScore: An Accurate Force Field-Based Scoring Function for Virtual Drug Screening. *J. Chem. Inf. Model.* **2008**, *48*, 1656–1662.

(8) Gehlhaar, D. K.; Verkhivker, G. M.; et al. Molecular Recognition of the Inhibitor Ag-1343 by Hiv-1 Protease - Conformationally Flexible Docking by Evolutionary Programming. *Chem. Biol.* **1995**, *2*, 17–324.

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

(10) Rarey, M.; Kramer, B.; et al. A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(11) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(12) Wang, R. X.; Lai, L. H.; et al. Further Development and Validation of Empirical Scoring Functions for Structure-Based Binding Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11–26.

(13) Wang, R. X.; Liu, L.; et al. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. *J. Mol. Model.* **1998**, *4*, 379–394.

(14) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(15) Gohlke, H.; Hendlich, M.; et al. Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.

(16) Clark, M.; Cramer, R. D., III.; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative molecular field analysis (CoMFA). 2. Toward its use with 3D-structural databases. *Tetrahedron Comp. Meth.* **1990**, *3*, 47–59.

(17) Clark, M.; Cramer, R. D. III The Probability of chance correlation using partial least squares (PLS). *Quant. Struct.-Act. Relat.* **1993**, *12*, 137–145.

(18) Miners, J. O.; Birkett, D. J. Cytochrome P4502C9: an enzyme of major importance in human drug metabolism. *Br. J. Clin. Pharmacol.* **1998**, *45*, 525–538.

(19) de Groot, M. J.; Vermeulen, N. P. E. Modeling the active site of cytochrome P450s and glutathione S-transferases, two of the most important biotransformation enzymes. *Drug Metab. Rev.* **1997**, *29*, 747–799.

(20) Ekins, S.; de Groot, M. J.; Jones, J. P. Pharmacophore and three-demensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. *Drug Metab. Dispos.* **2001**, *29*, 936–944.

(21) Jones, J. P.; He, M.; Trager, W. F.; Rettie, A. E. Three-demensional quantitative structure-activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab. Dispos.* **1996**, *24*, 1–6.

(22) Sreedhara, R.; Ron, A.; Michael, S.; William, F. T.; Allan, R.; Jeffrey, P. J. A refined 3-demensional QSAR of cytochrome P450 2C9: Computational predictions of drug interactions. *J. Med. Chem.* **2000**, *43*, 2789–2796.

(23) Michael, R. W.; Jason, K. Y.; Guillaume, A. S.; Christine, Y.; Keith, J. G.; C.; David, S.; Eric, F. J. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-Å resolution. *J. Biol. Chem.* **2004**, *279*, 35630–35637.

(24) Pamera, A. W.; Jose, C.; Alison, W.; Hayley, C. A.; Dijana, M. V.; Harren, J. Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **2003**, *424*, 464–468.

(25) Nakajima, M.; Ohyama, K.; Nakamura., S.; Shimada, N.; Yamazaki, H.; Yokoi, T. Inhibitory effects of azelastine and its metabolites on drug oxidation catalyzed by human cytochrome P-450 enzymes. *Drug Metab. Dispos.* **1999**, *27*, 792–797.

(26) Locuson, C. W., II.; Rock, D. A.; Jones, J. P. Quantitative binding models for CYP2C9 based on benzbromarone analogues. *Biochemistry* **2004**, *43*, 6948–6958.

(27) Katalin, M.; Laszlo, V.; Ferenc, L.; Istvan, S. Ipriflavone as an inhibitor of human cytochrome P450 enzymes. *Br. J. Pharmacol.* **1998**, *123*, 605–610.

(28) Kyoung-Ah, K.; Ji-Young, P. Inhibitory effect of glyburide on human cytochrome P450 isoforms in human liver microsomes. *Drug Metab. Dispos.* **2003**, *31*, 1090–1092.

(29) *SYBYL, 7.1*; Tripos, Inc.: St. Louis, 2005.

(30) Oda, N.; Yamaotsu, A.; Hirono, S. Determination of ligand-binding sites on proteins using long-range hydrophobic potential. *Biol. Pharm. Bull.* **2008**, *31*, 1552–1558.

(31) Israelachvili, J. N.; Pashley, R. M. The hydrophobic interaction is long range, decaying exponentially with distance. *Nature* **1982**, *300*, 341–342.

(32) Park, H.; Lee, S.; Suh, J. Structural and dynamical basis of broad substrate specificity, catalytic mechanism, and inhibition of cytochrome P450 3A4. *J. Am. Chem. Soc.* **2005**, *127*, 13634–13642.

(33) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, I.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal-mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41.

(34) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(36) Ryckaert, J. P. ; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamica of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(37) Hockney, R. W. The potential calculation and some applications. *Methods Comput. Phys.* **1970**, *9*, 136–211.

(38) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald-an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(39) Petersen, H. G. Accuracy and efficiency of the particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 3668–3679.

(40) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.

(41) Brendsen, H. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(42) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.

(43) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Modell.* **2002**, *20*, 281–295.

(44) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380–391.

(45) Brown, S. P.; Muchmore, S. M., W. M. High throughput calulation of protein-ligand binding affinities: Modification and adaptation of the MM-PBSA protocol to enterprise grid computing. *J. Chem. Inf. Model.* **2006**, *46*, 999–1005.

(46) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15*, 7865–7877.

(47) Jakarian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC Model I: Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(48) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(49) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(50) Case, D. A.; Cheatham, T. E., III.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *16*, 1668–1688.

(51) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.

(52) Pearlman, D. A. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796–7807.

(53) Rafi, S. B.; Cui, G.; Song, K.; Cheng, X.; Tonge, P. J.; Simmerling, C. Insight through molecular mechanics Poisson-Boltzmann surface area calculations into the binding affinity of triclosan and three analogues for fabI, the E. coli enoyl reductase. *J. Med. Chem.* **2006**, *49*, 4574–4580.

(54) Adekoya, O. A.; Wilassen, N. P.; Sylte, I. Molecular insight into pseudolysin inhibition using the MM-PBSA and LIE methods. *J. Struct. Biol.* **2006**, *153*, 129–144.

(55) Kuhn, B.; Gerber, P.; Schultz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.

(56) Steinbrecher, T.; Case, D. A.; Labahn, A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J. Med. Chem.* **2006**, *49*, 1837–1844.

(57) Page, C. S.; Bates, P. A. Can MM-PBSA calculations predict the specificities of protein kinase inhibitors. *J. Comput. Chem.* **2006**, *27*, 1990–2007.

(58) Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–6606.

(59) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **2006**, *49*, 4805–4808.

(60) Wade, R. C.; Winn, P. J.; Schlichting, I. Sudarko A survey of active site access channels in cytochrome P450. *J. Inorg. Chem.* **2004**, *98*, 1175–1182.

(61) Dickmann, L. J.; Locuson, C. W.; Jones, J. P.; Rettie, A. E. Differential roles of Arg97, Asp293, and Arg108 in enzyme stability and substrate specificity of CYP2C9. *Mol. Pharmacol.* **2004**, *65*, 842–850.

(62) Ridderström, M.; Masimirembwa, C.; Trump-Kallmeyer, S.; Ahlefelt, M.; Otter, C.; Andersson, T. B. Arginines 97 and 108 in CYP2C9 are important determinants of the catalytic function. *Biochem. Biophys. Res. Commun.* **2000**, *270*, 983–987.

(63) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49*, 6596–606.

(64) 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*, 4130–4146.

(65) *SYBYL, 6.8.1*; Tripos, Inc.: St. Louis, 2002.

CI800313H