The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme

4 AUTHORS, INCLUDING:

Jung-Hsin Lin
Academia Sinica

52 PUBLICATIONS **1,517** CITATIONS

SEE PROFILE

Jung-Hsin Lin
Alexander L. Perryman
Julie R. Schames
J. Andrew McCammon
Howard Hughes Medical
Institute,
Department of Chemistry and
Biochemistry,
and Department of
Pharmacology,
University of California at
San Diego,
9500 Gilman Dr.
La Jolla, CA 92093-0365

The Relaxed Complex Method: Accommodating Receptor Flexibility for Drug Design with an Improved Scoring Scheme

Abstract: *An extension of the new computational methodology for drug design, the "relaxed complex" method (J.-H. Lin, A. L. Perryman, J. R. Schames, and J. A. McCammon,* Journal of the American Chemical Society, *2002, vol. 24, pp. 5632–5633), which accommodates receptor flexibility, is described. This relaxed complex method recognizes that ligand may bind to conformations that occur only rarely in the dynamics of the receptor. We have shown that the ligand–enzyme binding modes are very sensitive to the enzyme conformations, and our approach is capable of finding the best ligand enzyme complexes. Rapid docking serves as an efficient initial filtering method to screen a myriad of docking modes to a limited set, and it is then followed by more accurate scoring with the MM/PBSA (Molecular Mechanics/Poisson Boltzmann Surface Area) approach to find the best ligand–receptor complexes. The MM/PBSA scorings consistently indicate that the calculated binding modes that are most similar to those observed in the x-ray crystallographic complexes are the ones with the lowest free energies.* © 2002 Wiley Periodicals, Inc. Biopolymers 68: 47–62, 2003

Keywords: *computational drug design; receptor flexibility; molecular dynamics; docking; scoring*

INTRODUCTION

The "Trinity" in the field of computational drug design has been an efficient searching algorithm, a very accurate scoring function, and a sufficiently large

configuration space. The dimensions of the configuration space should include a good library of compounds, locations, orientations, and conformations of ligands and receptors. There have been some docking programs that can locally replace a few atoms or part
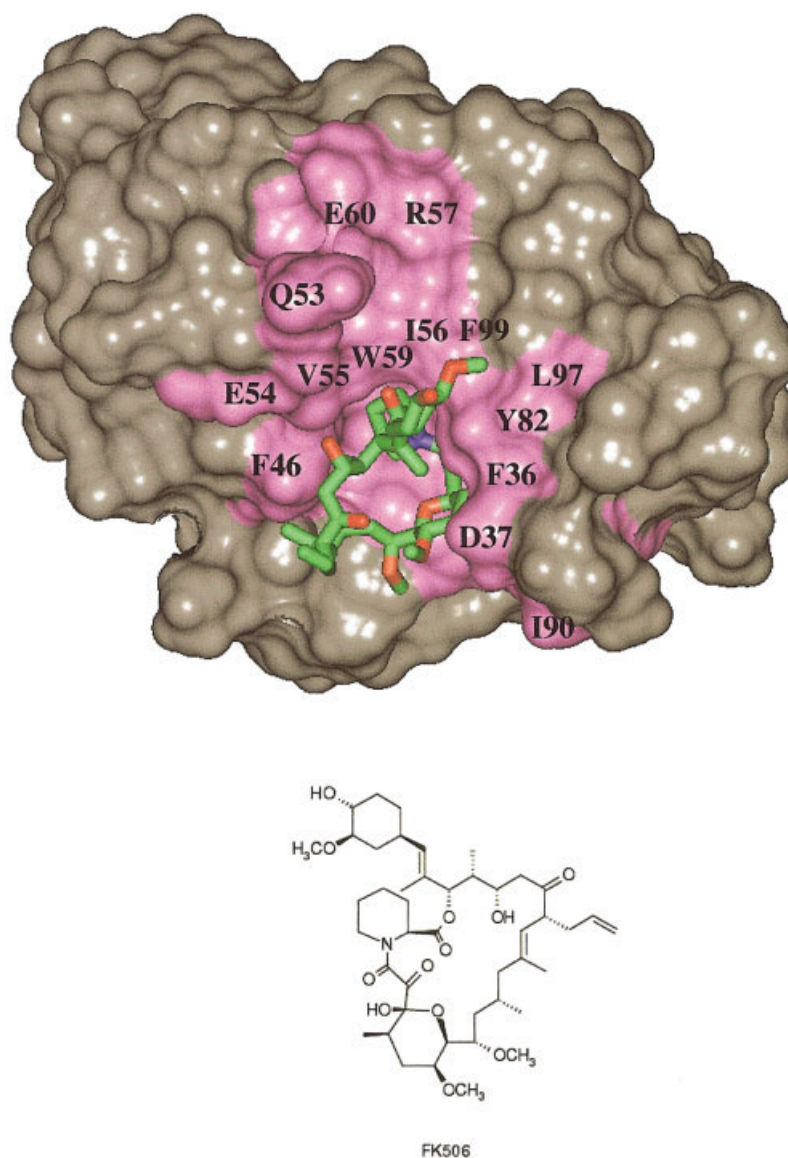
FIGURE 1   FK506 binds to FKBP-12, which is represented by the Connolly solvent-accessible surface. Structure adopted from Protein Data Bank (PDB code: 1FKF). Active site residues are color in magenta. The chemical structure of FK506 is included.
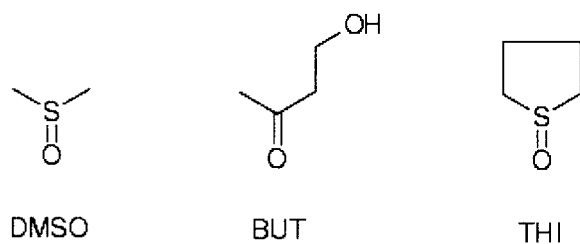

FIGURE 2   Chemical structural of DMSO, BUT, and THI.

of a molecular moiety of a ligand in order to optimize a lead compound.[1–3] While ligand flexibility has been incorporated in many docking schemes, most programs still treat the receptors as rigid bodies.[3–5] In general, ligands that are designed from combinatorial approaches may bind to conformations of the receptor that occur infrequently in the unliganded receptor; consequently, this rigid body assumption may fail to find correct ligand–receptor binding modes. Due to its apparent significance, some attempts have been made to account for the receptor flexibility.[6,7] These issues have also been reviewed recently.[8] Inspired by two recent successful experimental methods for the rapid

Table I Weighting Factors for Each Energetic Contribution Defined in Eqs. (2)–(6)

W_{vdw}	0.1485
W_{hbond}	0.0656
W_{elec}	0.1146
W_{tor}	0.3113
W_{sol}	0.1711

discovery of ligands that bind strongly to a receptor—namely the "SAR by NMR" method[9] and the "tether" method,[10] here we elaborate a novel computational approach, called the "relaxed complex" method, which incorporates receptor flexibility.

This method recognizes that a ligand may bind optimally to conformations that occur only rarely in the dynamics of the receptor and that strong binding often reflects multivalent attachment of the ligand to the receptor, as shown in our recent work.[11] Similar to the "dynamic pharmacophore" method,[12] in the "relaxed complex" scheme, a long molecular dynamics (MD) simulation of the unliganded receptor will first be conducted to extensively sample the receptor's conformations. The second phase of the relaxed complex method involves the rapid docking of minilibraries of candidate ligands to a large ensemble of the receptor's MD conformations.

With the advent of new docking algorithms and an enhanced empirical free energy function, AutoDock
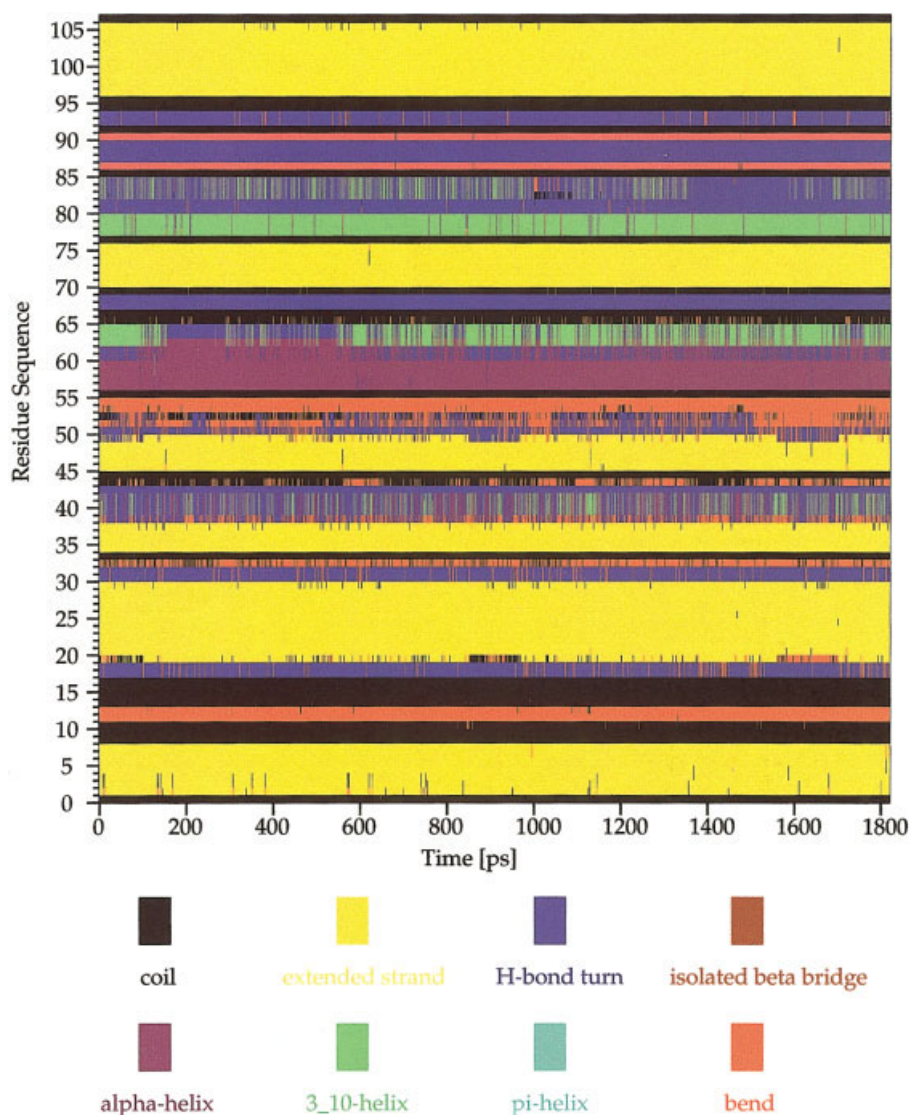
FIGURE 3 Stability of the secondary structures of FKBP-12 in the MD simulation.
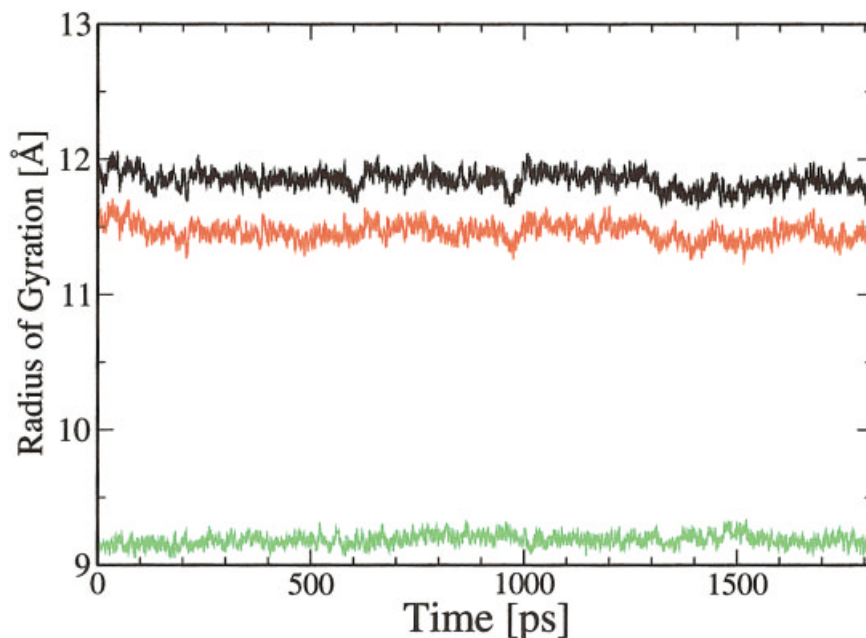
FIGURE 4   Radii of gyration of FKBP in the MD simulation.

$3.0.5^4$ is capable of performing very efficient docking of large flexible ligands and was adopted in the relaxed complex scheme. The Lamarckian genetic algorithm (LGA) is a hybrid of the original genetic algorithm with the adaptive local search (LS) method. The so-called genome in the genetic algorithm consists of floating point "genes," each of which encodes

one state variable describing the molecule's position, orientation, or conformation. The local searcher modifies the phenotype, which is then allowed to update the genotype. The initial position, orientation, and conformation of the ligand are randomly assigned to a region outside the receptor. The ligand then samples many different orientations, conformations, and trans-
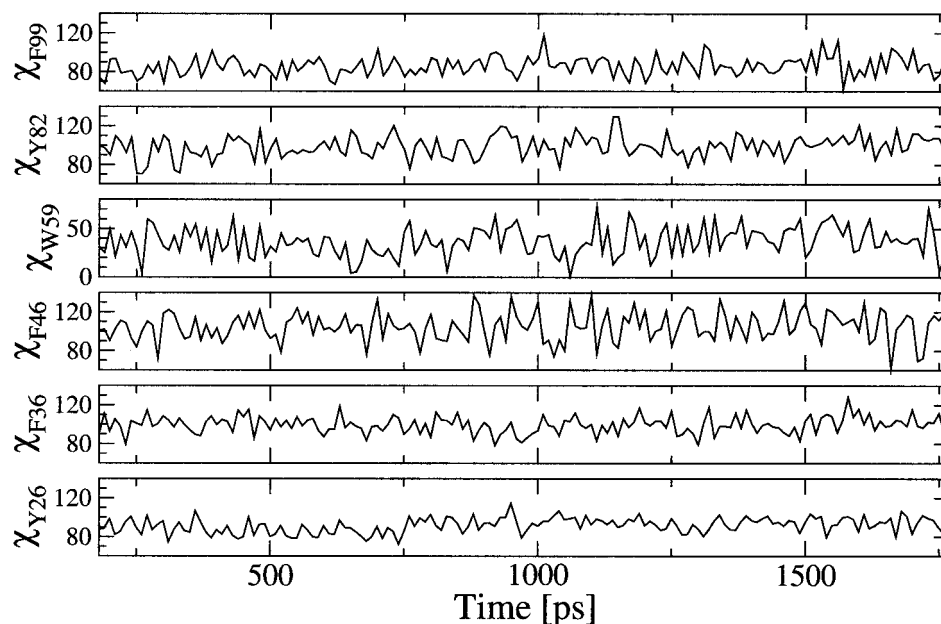

FIGURE 5   Librational motion of the aromatic side chains in the FKBP-12 active site during the MD simulation. The χ is the dihedral angle of C_{α}–C_{β}–C_{γ}–$C_{\delta 1}$ of an aromatic residue.
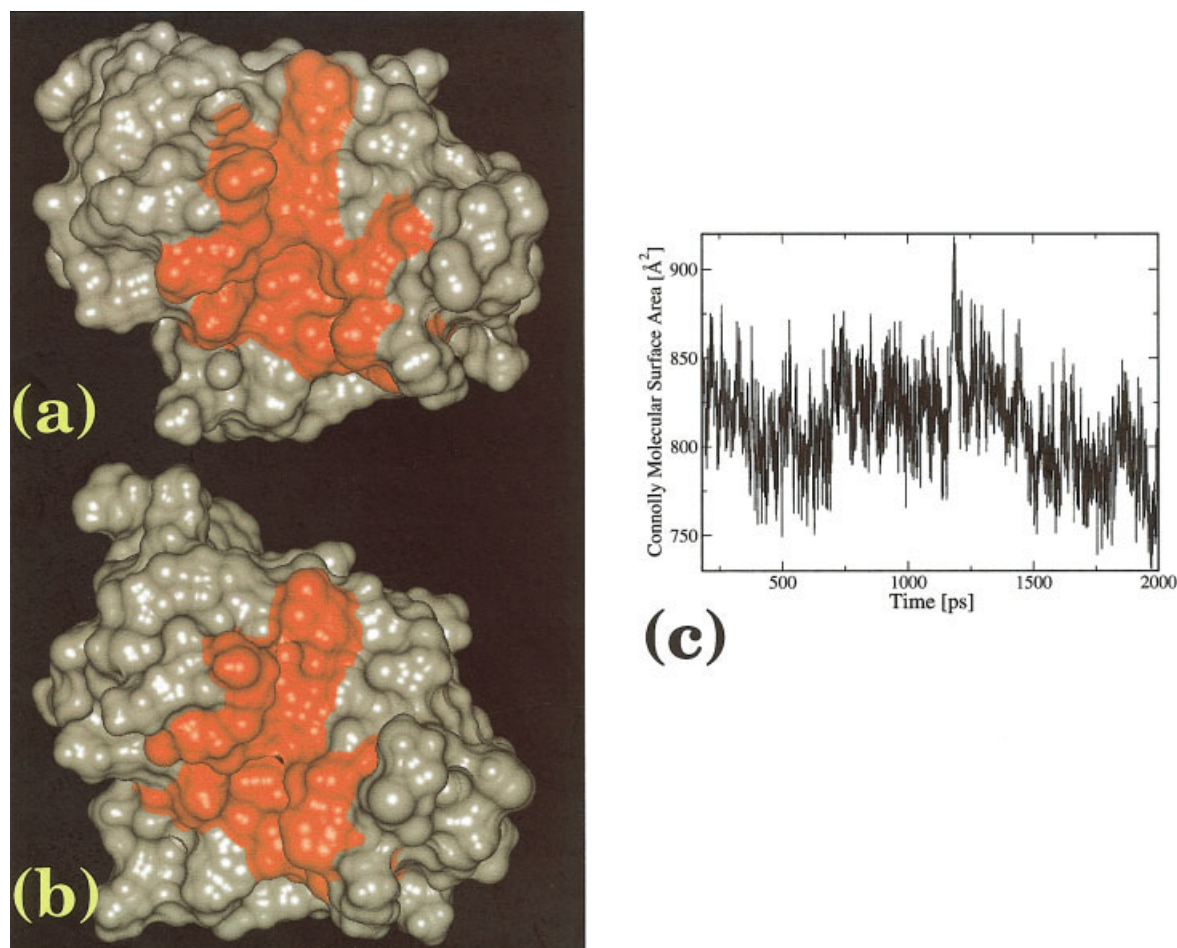
FIGURE 6 (a) Snapshot with the largest surface area of the active site. ($t = 1180$ ps). (b) Snapshot with the smallest surface area in the active site ($t = 1710$ ps). Active site surfaces are colored in red. Both (a) and (b) display the same orientation of FKBP. (c) Changes in molecular surface area of the active site of FKBP-12 in the MD simulation.

lations until an ideal complex has formed. However, due to the current accuracy of the AutoDock scoring function, such ideal sites from each run do not necessarily represent the correct binding mode of the ligand–receptor complex. In this work we employ the MM/PBSA (Molecular Mechanics/Poisson Boltzmann Surface Area) approach to re-score the docking results, which allows us to find the correct binding modes (as judged by the crystallographic complexes).

The experimentally well-characterized system FK506 binding protein, FKBP-12, was considered. The immunophilin FKBP-12 is the soluble receptor for the natural immunosuppressant drug, FK506 (tacrolimus). Immunophilins, when complexed to immunosuppressive ligands, appear to inhibit signal transduction cascades that would normally result in exocytosis and transcription. FK506 belongs to a special category of compounds, called "molecular glue" inhibitors or "chemical inducers of dimerization,"

which induce dimerization of two protein partners that normally have little or no affinity for each other.[13] The active site of FKBP-12 is formed by Y26, F36, D37, F46, Q53, E54, V55, I56, R57, W59, E60, Y82, H87, I90, L97, and F99. The binding of FK506 to FKBP-12 is illustrated in Figure 1.

Alternatives to the natural FK506 have been much sought after in this research field.[14–16] In this work three ligands, dimethylsulfoxide (DMSO), 4-hydroxy-2-butanone (BUT), and tetrahydrothiophene-1-oxide (THI) (see Figure 2), were selected for the docking studies, and their binding modes were compared with x-ray crystallographic data.[17]

COMPUTATIONAL DETAILS

Protein Structure

The FKBP-12 structure was taken from the Protein Data Bank (http://www.rcsb.org/pdb/). The first pub-
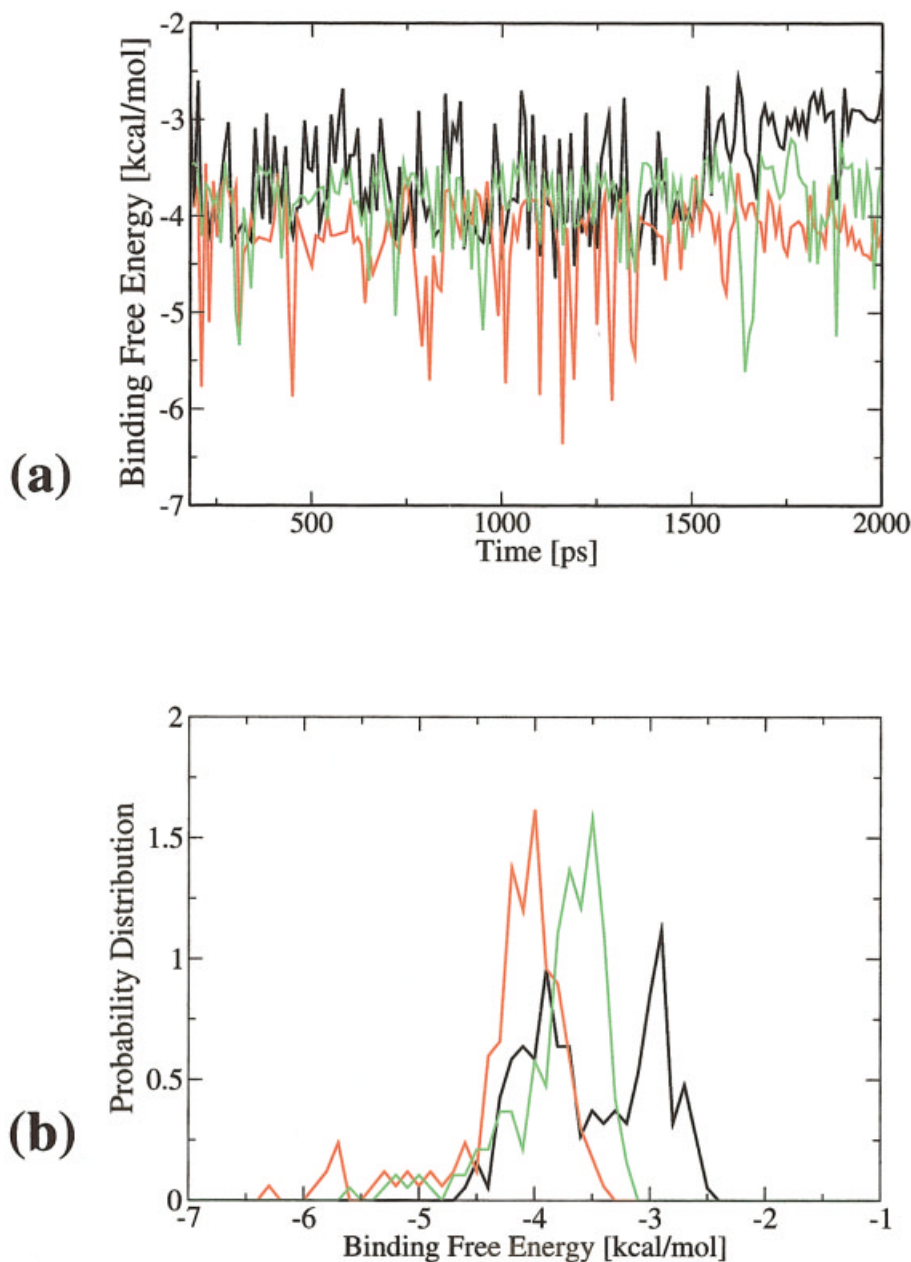
FIGURE 7   (a) The best binding free energies of each compound to the time series of MD conformations. (b) The probability distribution of binding free energies. Black: DMSO, red: BUT, green: THI.

lished crystal structure of FKBP-12 in the unliganded form[17] (PDB code: 1D6O) was used, which has a resolution of 1.85 Å. The protonation states of the ionizable residues, including the states the histidine residues, adopt (δ, ε, or protonated), were determined according to the NMR structure (1FKT).[18] The structure was carefully examined by WhatCheck.[19] The warnings reported by WhatCheck (e.g., some abnormal bond lengths and angles) disappeared after an initial energy minimization. The water molecules

Table II   Best Binding Free Energies (BFEs) of DMSO, BUT, and THI and Their Corresponding MD Snapshots[a]

Compound	BFE	MD Time (ps)
DMSO	−4.64	1140
BUT	−6.36	1160
THI	−5.29	1640

[a] Energetics are in the unit of kcal/mol for all the tables.

FIGURE 8 Binding modes of DMSO-FKBP complexes from AutoDock. The FKBP configuration is taken from $t = 230$ ps MD snapshot.

FIGURE 9 Binding modes of BUT-FKBP complexes from AutoDock. The FKBP configuration is taken from $t = 240$ ps MD snapshot.

FIGURE 10   Binding modes of THI-FKBP complexes from AutoDock. The FKBP configuration is taken from $t = 470$ ps MD snapshot.

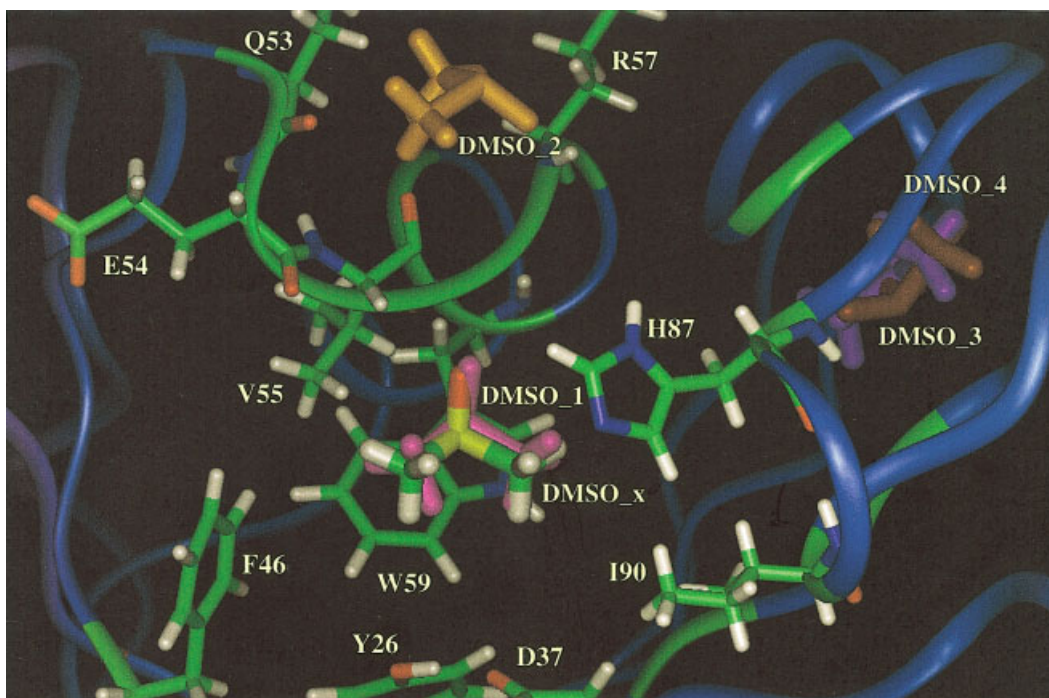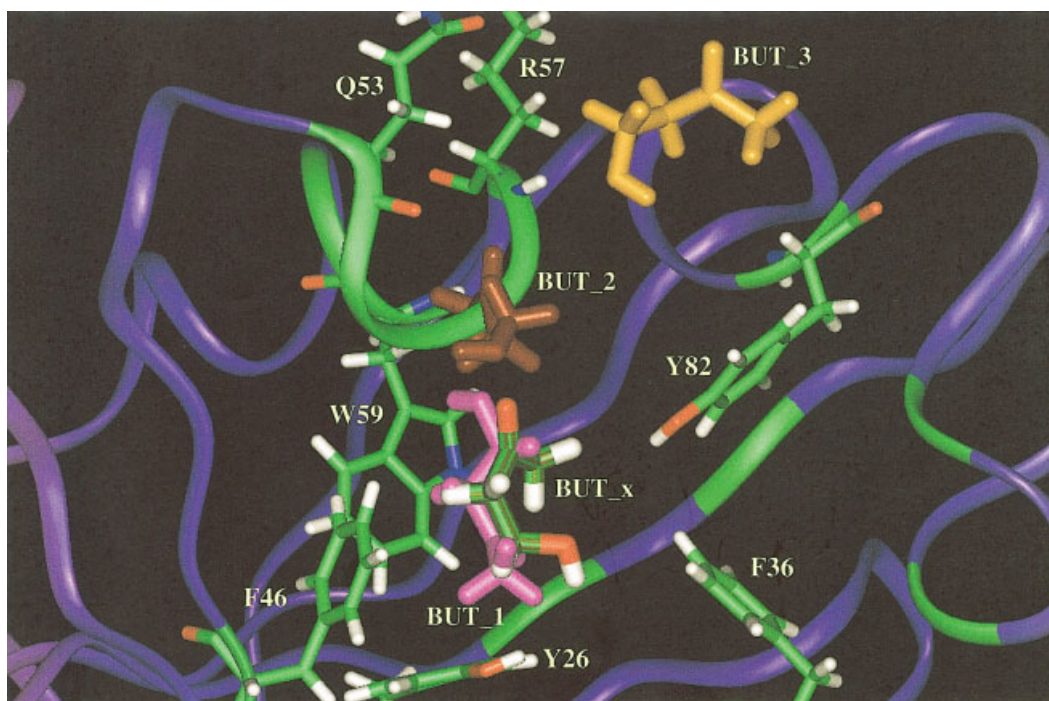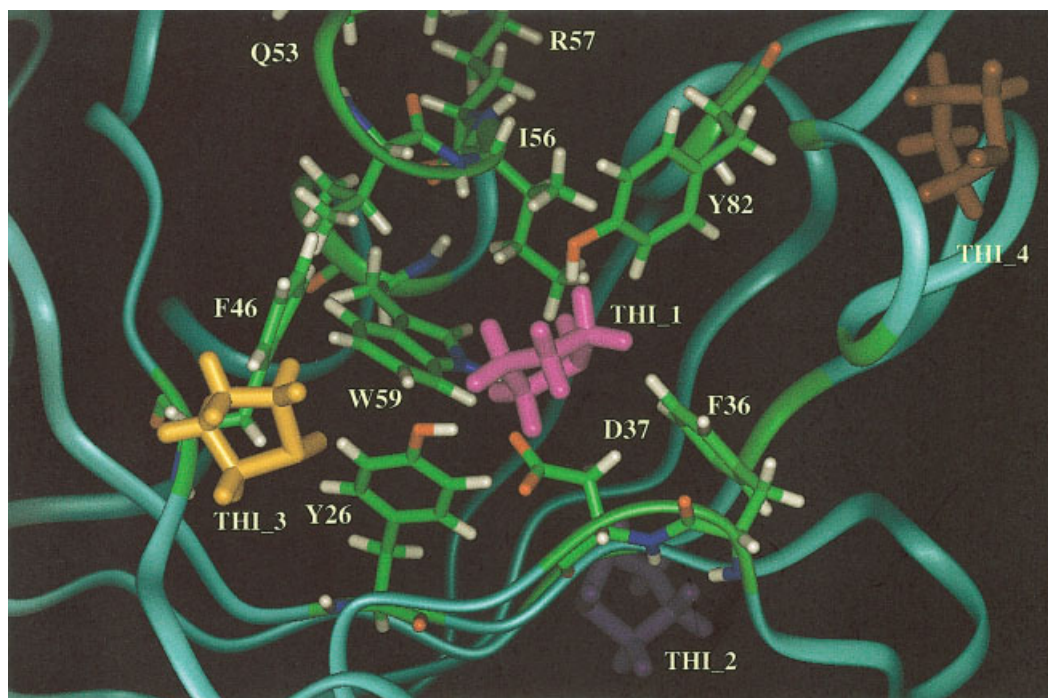solved by x-ray crystallography were kept in our model for MD simulations.

Simulation Protocol

The molecular dynamics simulations were conducted using the SANDER module of AMBER 7.0.[20] The preparation of the topology files and parameter files was done by using LEaP[21] and our own programs. By using the SHAKE algorithm,[22] the bond stretching degrees of freedom for the hydrogen atoms are prohibited, which permits the use of the longer time step of 2 fs. The isotropic scaling for pressure regulation was used. The reference pressure in each dimension was set to 1 Bar, and the Berendsen barostat[23] was used with the pressure coupling constant, τ_P, of 0.1 ps. The temperature was kept at 300 K, and the Berendsen thermostat was used with the coupling constant, τ_T, of 0.2 ps. Solute and solvent atoms were coupled to a single thermostat, and the center of mass motion was removed at each pico second. The particle mesh Ewald method[24] with cubic spline approximation was used to treat the long-range Coulombic interactions, and the direct sum tolerance was set to 0.00001.

The protein was soaked with water molecules in a cubic box of 79.0 Å in length. The TIP3P model of water[25] was used. A 30 ps NPT (isothermal-isobaric)

simulation was conducted to relax the whole system and to reach the equilibrated density, which is about 1.0188 g/cc. The final box size is about 70 Å. It should be noted that if a subset of atoms is very strongly restrained or fixed, the box size will not be adjusted even though the NPT condition is used. The condition was then switched to NVT (constant volume, constant temperature), and restrained potentials were applied to all the protein's atoms and to all the oxygen atoms of the crystallographic water molecules. The original minimized structure was taken as the reference structure, which allows the protein to be brought back to the original structure and to have the correct density of the solvent environment with that initial configuration. The restraints were gradually reduced from 30 kcal/mol to zero. In the subsequent production run, the NVT condition was still used to

Table III   DMSO/FKBP Complexes Ranked by AutoDock Binding Free Energies (see Figure 8)

ID	Rank	BFE	Site
1	1	−4.18	DMSO_4
2	1	−4.18	DMSO_4
3	2	−3.74	DMSO_3
4	3	−2.49	DMSO_2
5	4	−2.32	DMSO_1

Table IV   MM Decomposition of DMSO/FKBP Complexes

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	283.690	756.735	1017.756	−373.213	−355.275	1329.693
2	283.664	756.740	1017.756	−363.940	−355.297	1338.923
3	283.703	756.739	1017.756	−158.152	−352.751	1547.295
4	283.703	756.736	1017.756	−408.763	−354.574	1294.862
5	283.690	756.731	1017.756	−431.564	−354.539	1272.074

increase performance and to reduce perturbation to the system. It produced trajectories of about 50 ps per day on a single Compaq Alpha EV67 CPU or about 245 ps per day on the SDSC Meteor cluster on 16 processors.

DOCKING DETAILS

Ligand Preparation

Sybyl[26] was used to construct the molecular files for the ligands. All hydrogen atoms are present for all of the constructed ligands. The **autotors** module of AutoDock was used to assign the active torsions. The ligands were optimized using Gaussian 98[27] at the Hartree–Fock level with the 6-31 G* basis set. The RESP model,[28] an electrostatic potential-based method using charge restraints, was used to derive the partial charges on the atoms of the ligands.

Docking Protocol

AutoDock 3.0.5[4] was used for the docking study, and its implementation of the Lamarckian genetic algorithm (LGA) was activated to search for the globally optimized complexes. All the partial charges on the atoms of the enzyme were taken from the AMBER force field parameters. The **addsol** module of AutoDock was then used to generate the enzyme's PDBQS files, which include the solvation parameters. Affinity and electrostatic grids were generated by the

autogrid module. The grid spacing was 0.375 Å in each dimension, and each grid map consisted of 151 × 151 × 151 grid points. The centers of the grid maps were assigned to the geometric center of the enzyme. Each LGA job consists of 50 runs, and the number of generations in each run was 10,000. The maximum number of iterations of the pseudo-Solis and Wets local search[4] was 300. The maximum number of energy evaluations was set to 100,000,000 so that the actual number of energy evaluations used in each run was below this limit. The rest of the parameters were set to the default values. In this study, the compounds were docked to the conformations that occurred at every 10 ps throughout the entire 2 ns MD simulation (i.e., 200 different conformations were targeted).

Free Energy Evaluation

The free energy function in AutoDock 3.0.5 is defined as

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{elec} + \Delta G_{tor} + \Delta G_{sol} \tag{1}$$

where the formula for each term is as follows:

$$\Delta G_{vdw} = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{2}$$

$$\Delta G_{hbond} = W_{hbond} \times \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} - E_{hbond} \right) \tag{3}$$

Table V   MM/PBSA Analysis of DMSO/FKBP Complexes [$\gamma = 0.025$ kcal/mol/Å^2, $T = 298.15$ K, ε(solute) = 2, ε(solvent) = 78.54 for all MM/PBSA Analyses]

ID	E_{MM}	$-TS_{MM}$	G_{PB}	γA	G_{total}	Rank
1	1329.603	−1206.150	−893.708	153.827	−616.338	3
2	1338.923	−1206.154	−894.107	153.827	−607.511	4
3	1547.295	−1200.879	−894.141	153.827	−393.898	5
4	1294.862	−1206.216	−895.518	153.417	−653.455	2
5	1272.074	−1207.096	−897.793	153.062	−679.753	1

Table VI   MM Decomposition of DMSO/FKBP Minimized Complexes

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	31.871	183.585	815.590	−632.528	−331.804	66.714
2	31.856	183.585	815.585	−632.524	−331.803	66.714
3	31.856	190.146	827.095	−648.626	−324.233	76.238
4	31.531	182.960	815.725	−634.767	−332.614	62.835
5	31.613	182.940	816.521	−635.246	−332.879	62.949

$$\Delta G_{elec} = W_{elec} \times \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \quad (4)$$

$$\Delta G_{tor} = W_{tor} \times N_{tor} \quad (5)$$

$$\Delta G_{sol} = W_{sol} \times \sum_{i,j} (S_i V_j + S_j V_i) e^{-r_{ij}^2/2\sigma^2} \quad (6)$$

where $E(t)$ is the directional weight based on the angle, t, between the probe and the target atom, E_{hbond} is the empirically − estimated average energy of the hydrogen bonding of water with a polar atom, and N_{tor} is the number of sp^3 bonds in the ligand, which is a measure of the unfavorable entropy of ligand binding due to the restriction of its conformational degrees of freedom. S_i and V_i are the solvation parameter and the fragmental volume of atom i,[29] respectively. The weighting factors for different energetic contributions are listed in Table I. These weighting factors were obtained by fitting a large set of energetics analyses of ligand–receptor complexes.[4]

In order to enhance efficiency, AutoDock takes a grid-based approach. The affinity and the electrostatic potential of the regions between the grid points are evaluated by a trilinear interpolation scheme.

One advantage of the AutoDock scoring function is that the "observed free energy," ΔG_{obs}, from the docking results can be readily related to the experimental dissociation constant, K_D, using the following equation:

$$\Delta G_{obs} = RT \ln K_D \quad (7)$$

where R is the gas constant, 1.987 cal/K/mol, and T is the absolute temperature.

RESULTS

Secondary Structures of FKBP

FKBP has an antiparallel β-sheet folding topology that results in a novel loop crossing and produces a large cavity lined by a conserved array of aromatic residues; this cavity serves as the rotamase active site and drug-binding pocket. There are other significant structural features (such as a protruding positively charged loop and an apparently flexible loop) that may be involved in the biological activity of FKBP.

Figure 3 shows the changes in the secondary structures of FKBP, as defined by DSSP,[30] in the MD simulation. Most of the secondary structures of the residues at the active site (e.g., Y26, F36, D37, F46, E54, V55, I56, W59, H87, I90, and F99) do not fluctuate much during the simulation. The β-strand structures remain intact. However, rapid transitions of secondary structures were observed in the regions of residues 38–42, 49–53, 60–66, and 82–85, which contain some active site residues.

Radii of the Gyration of FKBP

Given the Cartesian coordinates and masses of the atoms, the moment of inertia tensor, $\mathbb{I}$, is

Table VII   MM/PBSA Analysis of DMSO/FKBP Minimized Complexes

ID	E_{MM}	$-TS_{MM}$	G_{PB}	γA	G_{total}	Rank
1	66.714	−1206.150	−789.681	144.150	−1784.968	4
2	66.714	−1206.154	−789.811	144.149	−1785.103	3
3	76.238	−1200.879	−795.701	140.753	−1779.589	5
4	62.835	−1206.216	−797.958	143.306	−1798.034	2
5	62.949	−1207.096	−798.194	143.120	−1799.221	1

Table VIII   BUT/FKBP Complexes Ranked by AutoDock Binding Free Energies (See Figure 9)

ID	Rank	BFE	Site
1	1	-133×10^5	BUT_2
2	2	-3.64	BUT_1
3	2	-3.63	BUT_1
4	2	-3.63	BUT_1
5	3	-3.37	BUT_3

$$
\mathbb{I} = \begin{pmatrix}
\sum_i m_i(y_i^2 + z_i^2) & -\sum_i m_i x_i y_i & -\sum_i m_i x_i z_i \\
-\sum_i m_i y_i x_i & \sum_i m_i(x_i^2 + z_i^2) & -\sum_i m_i y_i z_i \\
-\sum_i m_i z_i x_i & -\sum_i m_i z_i y_i & \sum_i m_i(x_i^2 + y_i^2)
\end{pmatrix}
\tag{8}
$$

The radii of gyration of a given structure can be calculated from the three principal moments of inertia,

$$
R_i = \sqrt{\frac{I_i}{M}} \quad i = 1, 2, 3
\tag{9}
$$

where M is the total mass of the molecule.

For a rigid body the rotational motion can be fully described by its principal moments. Figure 4 shows the changes of the radii of gyration of FKBP in the 2 ns simulation, and it appears that the overall shape of FKBP is quite stable.

The above two subsections support the stability of the gross FKBP conformation during the MD simulation.

Librational Motions of Aromatic Side Chains at the Active Site

Although the overall shape and backbone structure of FKBP were stable in the simulation, it is conceivable that the side chains in the active site still underwent significant conformational changes. Correspondingly, the steric and affinity interactions at the active site could be altered by changes in the side chain orientations of, for example, the aromatic residues. The active site of FKBP-12 is rich in hydrophobic (V55, I56, I90, L97) and aromatic residues (Y26, F36, F46, W59, Y82, F99). Figure 5 depicts the librational motion of the side chains of the aromatic residues in the FKBP-12 active site during the MD simulation. It is clear that all of these side chains have wobbled substantially in the simulation; therefore, conformations of the active site have been sampled extensively. For compounds that are very specific to a particular subsite of the receptor, e.g., between two adjacent aromatic side chains, such librational motion could provide an "on/off" gating or switching mechanism. On the other hand, such rapid librational motions might be a characteristic of only the unliganded receptors. For a liganded receptor, it is conceivable that the librational motions of the residues close to the bound ligand would be largely suppressed.

Changes of the Molecular Surface Area of the Active Site

To further assess the correlation of the local motions of the residues at the active site to other physical properties, the solvent-accessible molecular surface of the FKBP-12 active site was calculated by MSD, which uses an analytical algorithm by Connolly.[31] Because this surface is defined by rolling a probe atom on the receptor's molecular surface, it also represents the accessibility of ligand atoms to the active site. Larger surface areas will correspond to less steric hindrance of ligands to the active site, and vice versa. As shown in Figure 6(c), the difference between the largest and smallest areas in this 2 ns MD simulation is about 187 Å^2, which is approximately the area of a small ligand (DMSO: 153 Å^2, BUT: 245 Å^2, THI: 165 Å^2). For large ligands that bind to the receptor with very specific steric arrangements, the influence of the changes in the surface area on the preferred binding modes is also apparent.

Table IX   MM Decomposition of BUT/FKBP Complexes (Nan = Not a Number)

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	277.416	778.221	1032.382	Nan	Nan	Nan
2	277.407	778.196	1031.838	-380.869	-347.481	1359.091
3	277.410	778.205	1032.732	-389.861	-347.460	1351.026
4	277.399	778.228	1032.741	-390.385	-347.452	1350.531
5	277.402	778.206	1033.639	-414.967	-347.899	1326.381

Table X   MM/PBSA Analysis of BUT/FKBP Complexes

ID	E_{MM}	$-TS_{MM}$	G_{PB}	γA	G_{total}	Rank
1	Nan	Nan	−876.049	153.810	Nan	5
2	1359.091	−1206.256	−874.298	153.874	−568.589	4
3	1351.026	−1205.546	−873.959	152.834	−575.645	3
4	1350.531	−1206.257	−874.128	152.846	−577.008	2
5	1326.381	−1208.331	−871.143	154.033	−599.060	1

Table XI   MM Decomposition of BUT/FKBP Minimized Complexes

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	Nan	Nan	Nan	Nan	Nan	Nan
2	31.677	184.730	812.954	−644.909	−328.360	56.092
3	31.528	184.235	812.684	−645.522	−326.847	56.078
4	31.677	184.730	812.950	−644.909	−328.360	56.088
5	31.617	184.853	813.635	−642.005	−329.530	58.570

Sensitivity of the Docking Free Energy to MD Conformations

As anticipated previously, the local motions of the active site residues can alter the binding of ligands in a drastic manner. The lowest binding free energies of ligand–receptor complexes can also vary according to the receptor's conformational changes. Indeed, this is shown very clearly in Figure 7(a), where the best binding free energies of a ligand change markedly even for neighboring MD conformations. The lowest binding free energy for DMSO, BUT, and THI and their corresponding MD snapshots are listed in Table II. As shown in Figure 7(b), the distributions of these binding free energies are very wide (at least 2 kcal/mol). The binding modes of the best docked complexes vary in different MD snapshots. Furthermore, some designed ligands may not bind to the active site for certain receptor conformations. It should be noted that the accuracy of the scoring function plays a very important role. In some cases the binding modes that correspond to the largest calculated binding free energy are not necessarily the actual binding modes found in the crystallographic complexes. It is usually required to verify the docking results with other scoring schemes.

Rescoring with the MM/PBSA Scheme

There are many computational methods that can be used to calculate the standard and relative free energies of complex formation,[32–34] but most of them require substantial computational efforts. The MM/PBSA scheme[35–39] is a postprocessing method to evaluate the standard free energies of molecules or the binding free energies of molecular complexes in a less computationally demanding manner.

The MM/PBSA method combines molecular mechanical (MM) energies, continuum solvation energies, and estimates of solute entropy as follows:

$$\langle G_{mol} \rangle = \langle E_{MM} \rangle + \langle G_{PBSA} \rangle - T \langle S_{MM} \rangle \qquad (10)$$

where $\langle G_{mol} \rangle$ is the average "standard" free energy[32] of the molecule of interest, which can be the ligand, the receptor, or their complex. The average molecular mechanical energy, $\langle E_{MM} \rangle$, is typically defined as

Table XII   MM/PBSA Analysis of BUT/FKBP Minimized Complexes

ID	E_{MM}	$-TS_{MM}$	G_{PB}	γA	G_{total}	Rank
1	Nan	Nan	Nan	Nan	Nan	5
2	56.093	−1206.256	−800.895	141.444	−1809.558	3
3	56.078	−1205.546	−813.242	141.681	−1816.383	1
4	56.093	−1206.257	−800.299	141.441	−1808.965	4
5	58.569	−1208.331	−806.328	142.750	−1812.109	2

Table XIII   THI/FKBP Complexes Ranked by AutoDock Binding Free Energies (See Figure 10)

ID	Rank	BFE	Site
1	1	−3.81	THI_1
2	1	−3.81	THI_1
3	2	−3.69	THI_3
4	3	−3.66	THI_2
5	4	−3.45	THI_4

$$\langle E_{\text{MM}} \rangle = \langle E_{\text{bond}} \rangle + \langle E_{\text{angle}} \rangle + \langle E_{\text{torsion}} \rangle + \langle E_{\text{vdw}} \rangle + \langle E_{\text{elec}} \rangle$$

$$(11)$$

where E_{bond}, E_{angle}, E_{torsion}, E_{vdw}, E_{elec} are the bond, angle, torsion, van der Waals, and electrostatics terms of intramolecular energy, respectively.

The molecular solvation free energy can be further decomposed to

$$\langle G_{\text{PBSA}} \rangle = \langle G_{\text{PB}} \rangle + \gamma \cdot \langle A \rangle \qquad (12)$$

where $\langle G_{\text{PBSA}} \rangle$ denotes the average electrostatic contribution of molecular solvation, γ is the surface tension of water, and A is the solvent-accessible surface area (SASA). The electrostatic solvation free energy, G_{PB}, is calculated by

$$G_{\text{PB}} = \sum_{i=1}^{N} \left[q_i(\phi_i^{aq} - \phi_i^{g}) \right] \qquad (13)$$

where N is the number of atoms in the molecule, q_i is the electrostatic charge of atom i, and ϕ_i^{aq} and ϕ_i^{g} are the electrostatic potentials of atom i in the aqueous and gas phase, respectively, which are usually obtained by solving the Poisson–Boltzmann equation[40,41]:

$$-\nabla \cdot [\varepsilon(r)\nabla\phi(r)] + \kappa^2(r)\sinh\left[\frac{e_c\phi(r)}{k_{\text{B}}T}\right] = \rho_{\text{mol}}(r)$$

$$(14)$$

where $\varepsilon(r)$ is the position-dependent dielectric constant, k_{B} is the Boltzmann constant, T is the absolute temperature, e_c is the electron charge, κ is the inverse Debye–Hückel screening length,

$$\kappa^2(r) = \frac{2Ie_c^2}{k_{\text{B}}T\varepsilon(r)} \qquad (15)$$

where I is the ionic strength, and $\rho_{\text{mol}}(r)$ is the charge distribution of the molecule, given by a sum of δ functions:

$$\rho_{\text{mol}}(r) = \sum_{i=1}^{N} q_i\delta(r - r_i) \qquad (16)$$

The Poisson–Boltzmann equation was solved using the parallel focusing methods available in the Adaptive Poisson–Boltzmann Solver (APBS) software package,[42] described in more detail at http://mccammon.ucsd.edu/apbs.

The entropy of a polyatomic molecule, S_{MM}, can be decomposed into three parts:

$$S_{\text{MM}} = S_{\text{trans}} + S_{\text{rot}} + S_{\text{vib}} \qquad (17)$$

where S_{trans}, S_{rot}, and S_{vib} represent the entropies from the translational, rotational, and vibrational degrees of freedom, respectively. It should be noted that the conformational entropy is not taken into account in the current MM/PBSA scheme. The translational entropy can be calculated by[42]

$$S_{\text{trans}} = k_{\text{B}} \ln\left[\left(\frac{2\pi M k_{\text{B}}T}{h^2}\right) e^{5/2} \right] \qquad (18)$$

where h is the Planck constant, M is the molecular mass, and e is the natural exponent. The rotational entropy can be calculated using the classical approximation[43]:

Table XIV   MM Decomposition of THI/FKBP Complexes

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	295.295	805.675	999.948	−324.180	−347.075	1429.663
2	295.354	805.659	999.951	−324.063	−347.075	1429.826
3	232421.603	1393.251	1002.536	−297.367	−343.341	234176.682
4	261681.558	1620.446	1002.471	−321.140	−343.274	263640.060
5	28807.596	1618.684	1002.795	−201.082	−338.425	1427.179

Table XV   MM/PBSA Analysis of THI/FKBP Complexes

ID	E_{MM}	$-TS_{\mathrm{MM}}$	G_{PB}	γA	G_{total}	Rank
1	1429.663	−1214.322	−916.905	155.699	−545.866	2
2	1429.826	−1214.318	−916.906	155.699	−545.699	3
3	234176.682	−1211.660	−917.131	156.270	232204.161	4
4	263640.060	−1211.689	−920.977	156.470	261663.864	5
5	1427.179	−1212.426	−925.409	156.127	−554.529	1

$$S_{\mathrm{rot}} = k_{\mathrm{B}} \ln \left[\frac{\pi^{1/2} e^{3/2}}{\sigma} \left(\prod_{i=1}^{3} \frac{h^2}{8\pi^2 k_{\mathrm{B}} I_i} \right) \right] \qquad (19)$$

where σ is the symmetry number and I_i are the three principal moments of inertia, which are obtained by solving the eigenvalues of the matrix defined in Eq. (9).

The vibrational entropy can be calculated by normal mode analysis[43] (assuming a nonlinear molecule),

$$S_{\mathrm{vib}} = k_{\mathrm{B}} \sum_{j=1}^{3N-6} \left[\frac{\Theta_{\omega_j/T}}{e^{\Theta_{\omega_j/T}} } - \ln(1 - e^{\Theta_{\omega_j/T}}) \right] \qquad (20)$$

where the Θ_{ω_j} are the normal mode vibrational temperatures,

$$\Theta_{\omega_j} \equiv \frac{\hbar \omega_j}{k_{\mathrm{B}}} \qquad (21)$$

where ω_j are normal mode angular frequencies.

The NMODE module[44] of the AMBER 7 suite was used for entropy calculations.

In order to demonstrate how the MM/PBSA scheme could improve the scorings of the binding modes, snapshots that have multiple ligand binding sites found by AutoDock were selected, as shown in Figures 8 (DMSO), 9 (BUT), and 10 (THI). The binding free energies and rankings by AutoDock are listed in Tables III, VIIII, and XIII. In Figures 8 and 9, LIGAND$_x$ denotes the locations of ligands in the

x-ray structures. Apparently, the best binding modes for DMSO and BUT found by AutoDock were not the same as those found in the x-ray studies. Since only snapshots of the binding modes are available, it is not possible to obtain the ensemble-averaged free energies by following the original MM/PBSA scheme described above, which requires long-time MD simulations for each of the binding modes. It should be noted that conformations that were taken directly from AutoDock may change when judged by a more sensitive and accurate energy evaluation scheme. Although AutoDock adopts parameters similar to those in the MM/PBSA scheme for the nonbond interactions, it uses a grid-based approach, and the interactions were smoothed at the grid points to avoid singularity problems. A potential problem of such approaches is clearly demonstrated in Figure 9. The best binding mode of the calculated BUT/FKBP complex (see Table VIII) is in fact unrealistic, because BUT is inserted partially into the backbone of the receptor. This is due to the numerical artifact from the cancellation of large electrostatic attraction and van der Waals repulsion. This erroneous binding mode was easily detected by the MM energetic analysis, shown in Table IX.

As described previously, the conformations taken from AutoDock are not necessarily the most suitable representatives of the corresponding binding modes. The energy-minimized conformations, which sit in the basin of an ensemble of neighboring conformations, may be better representatives. On the other hand, a large number of iterations in the genetic

Table XVI   MM Decomposition of THI/FKBP Minimized Complexes

ID	E_{bond}	E_{angle}	E_{dihe}	E_{vdw}	E_{elec}	E_{MM}
1	30.767	184.870	811.027	−630.345	−323.511	72.808
2	30.767	184.871	811.029	−630.343	−323.515	72.808
3	31.052	188.435	818.784	−636.952	−325.325	75.994
4	31.442	187.097	810.855	−631.099	−327.473	70.822
5	30.792	185.717	813.281	−630.677	−324.532	74.580

Table XVII MM/PBSA Analysis of THI/FKBP Minimized Complexes

ID	E_{MM}	$-TS_{MM}$	G_{PB}	γA	G_{total}	Rank
1	72.808	−1214.322	−853.898	144.327	−1851.085	1
2	72.808	−1214.318	−850.392	144.305	−1847.597	2
3	75.994	−1211.660	−818.373	141.360	−1812.679	3
4	70.822	−1211.689	−810.503	143.669	−1807.701	5
5	74.580	−1212.426	−816.016	144.895	−1808.967	4

algorithm may also alter the ligand conformations, e.g., bond lengths, angles, and dihedrals, and introduce large energetic contributions in the MM/PBSA analyses (e.g. Table XIV). To diagnose such problems, the original docked conformations and the minimized-complex conformations were both used to evaluate their MM/PBSA free energies. As a general trend, the energy minimizations "flatten" the rugged energy landscapes, which can be seen by comparing the original MM energetics and minimized MM energetics (Tables IV, VI, IX, XI, XIV, and XVI).

The energy minimization and the normal mode entropy calculation are the most computationally demanding parts of the MM/PBSA scheme; therefore, the solute entropies were sometimes neglected in the previous MM/PBSA analyses.[36,38] As seen from Table V, X, and XV, the entropic contributions to the free energy differences can be as large as 6–7 kcal/mol (Table V). Thus, in general, they should not be neglected.

From Tables VII, XII, and XVII, it is clearly shown that the best MM/PBSA-ranked minimized complexes are those that exhibit similar binding modes to the x-ray crystallographic complexes. Unminimized complexes do exhibit a similar tendency (Tables V and XV), but exceptions can be found in certain scenarios (Table X).

In principle, the binding free energy of a ligand with its receptor can be calculated as

$$\Delta G_{binding} = \Delta G_{complex} - \Delta G_{receptor} - \Delta G_{ligand}$$

where $\Delta G_{complex}$, $\Delta G_{receptor}$, and ΔG_{ligand} are the standard free energy of the receptor, ligand, and their complex, respectively. Ideally, it would be best to conduct long duration, explicitly solvated MD simulations for the ligand, its receptor, and their complex, and perform MM/PBSA analyses on these three trajectories. However, since many binding modes to many different MD conformations are sampled in the current relaxed complex scheme, the binding free energies of different ligands can likely also be accurately estimated by some statistical analyses.

CONCLUSIONS

In summary, we have demonstrated that docking results are very sensitive to the conformations of a receptor. The average distribution of the docking free energies is about 2–3 kcal/mol, which is sufficient to misrank a potent drug candidate as a weak binder, and vice versa. By using the MM/PBSA scheme and ranking different binding modes by the free energies of the minimized complexes, it has been shown that the binding modes that are in agreement with the x-ray studies are now consistently ranked as the most stable complexes. Future work must include more efficient methods for energy minimizations, for normal mode analyses, and a statistical approach for calculating the binding free energies of different ligands.

REFERENCES

1. Wang, R.; Gao, Y.; Lai, L. J Mol Model 2000, 6, 498–516.
2. Budin, N.; Majeux, N.; Tenette-Souaille, C.; Caflish, A. J Comp Chem 2001, 22, 1956–1970.
3. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. J Comput Aid Mol Des 2001, 15, 411–428.
4. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. J Comput Chem 1998, 19, 1639–1662.
5. Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. J Mol Biol 2001, 308, 377–395.
6. Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. J Comp Chem 1995, 16, 454–464.
7. Burkhardt, C.; Zacharias, M. Nucleic Acids Res 2001, 29, 3910–3918.

8. Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Curr Opin Struct Biol 2002, 12, 197–203.

9. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Science 1996, 274, 1531–1534.

10. Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R.-M. Proc Natl Acad Sci 2000, 97, 9367–9372.

11. Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. J Am Chem Soc 2002, 124, 5632–5633.

12. Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. J Med Chem 2000, 43, 2100–2114.

13. Babine, R. E.; Bender, S. L. Chem Rev 1997, 97, 1359–1472.

14. Muegge, I.; Martin, Y. C.; Hajduk, P. J.; Fesik, S. W. J Med Chem 1999, 42, 2498–2503.

15. Dubowchik, G. M.; Vrudhula, V. M.; Dasgupta, B.; Ditta, J.; Chen, T.; Sheriff, S.; Sipman, K.; Witmer, M.; Tredup, J.; Vyas, D. M.; Verdoorn, T. A.; Bollini, S.; Vinitsky, A. Org Lett 2001, 3, 3987–3990.

16. Sich, C.; Improta, S.; Cowley, D. J.; Guenet, C.; Merly, J. P.; Teufel, M.; Saudek, V. Eur J Biochem 2000, 267, 5342–5354.

17. Burkhard, P.; Taylor, P.; Walkinshaw, M. D. J Mol Biol 2000, 295, 953–962.

18. Michnick, S. W.; Rosen, M. K.; Wandless, T. J.; Karplus, M.; Schreiber, S. L. Science 1991, 252, 836–839.

19. Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. Nature 1996, 381, 272–272.

20. Case, D. A.; Pearlman, D. A.; Caldwell, J. W., Cheatham III, T. E.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 7, San Francisco, 2002.

21. Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V. LEaP, San Francisco, 1995.

22. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327–341.

23. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684–3690.

24. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. J Chem Phys 1995, 103, 8577–8593.

25. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. J Chem Phys 1983, 79, 926–935.

26. Sybyl Molecular Modeling Software: Tripos Inc., 1966 S. Hanley Road; St. Louis, MO 63144-2913, http://www.tripos.com

27. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; R. E. Stratmann, J. A. M.; Burant, J. C.; Dapprich, S.; Millam, J.-M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, Revision A.1, Pittsburgh, PA, 1998.

28. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. J Phys Chem 1993, 97, 10269–10280.

29. Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. Mol Simul 1993, 10, 97–120.

30. Kabsch, W.; Sander, C. Biopolymers 1983, 22, 2577–2637.

31. Connolly, M. L. J Appl Cryst 1983, 16, 548–558.

32. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. Biophys J 1997, 72, 1047–1069.

33. Beveridge, D. L.; DiCapua, F. M. Annu Rev Biphys Biophys Chem 1989, 18, 431–492.

34. Kollman, P. A. Chem Rev 1993, 93, 2395–2417.

35. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.-A.; Thomas E. Cheatham, I. Accts Chem Res 2000, 33, 889–897.

36. Srinivasan, J.; Thomas E. Cheatham, I.; Cieplak, P.; Kollman, P. A.; Case, D. A. J Am Chem Soc 1998, 120, 9401–9409.

37. Massova, I.; Kollman, P. A. J Am Chem Soc 1999, 121, 8133–8143.

38. Wang, W.; Kollman, P. A. J Mol Biol 1999, 303, 567–582.

39. Lee, M. R.; Duan, Y.; Kollman, P. A. Proteins Struct Funct Genet 2000, 39, 309–316.

40. Davis, M. E.; McCammon, J. A. Chem Rev 1990, 94, 7684–7692.

41. Honig, B.; Nicholls, A. Science 1995, 268, 1144–1149.

42. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Proc Natl Acad Sci 2001, 98, 10037–10041.

43. McQuarrie, D. A. *Statistical Mechanics;* Harper & Row: New York, 1976.

44. Case, D. A. Curr Opin Struct Biol 1994, 4, 285–290.