PROTEINS: Structure, Function, and Bioinformatics 54:759–767 (2004)

Rapid Protein–Ligand Docking Using Soft Modes From Molecular Dynamics Simulations to Account for Protein Deformability: Binding of FK506 to FKBP

Martin Zacharias
Computational Biology, School of Engineering and Science, International University Bremen, Bremen, Germany

ABSTRACT **Most current docking methods to identify possible ligands and putative binding sites on a receptor molecule assume a rigid receptor structure to allow virtual screening of large ligand databases. However, binding of a ligand can lead to changes in the receptor protein conformation that are sterically necessary to accommodate a bound ligand. An approach is presented that allows relaxation of the protein conformation in precalculated soft flexible degrees of freedom during ligand–receptor docking. For the immunosuppressant FK506-binding protein FKBP, the soft flexible modes are extracted as principal components of motion from a molecular dynamics simulation. A simple penalty function for deformations in the soft flexible mode is used to limit receptor protein deformations during docking that avoids a costly recalculation of the receptor energy by summing over all receptor atom pairs at each step. Rigid docking of the FK506 ligand binding to an unbound FKBP conformation failed to identify a geometry close to experiment as favorable binding site. In contrast, inclusion of the flexible soft modes during systematic docking runs selected a binding geometry close to experiment as lowest energy conformation. This has been achieved at a modest increase of computational cost compared to rigid docking. The approach could provide a computationally efficient way to approximately account for receptor flexibility during docking of large numbers of putative ligands and putative docking geometries. Proteins 2004;54:759–767.**

Key words: ligand–receptor interaction; principal components of motion; receptor flexibility; global deformations; molecular mechanics; docking minimization

INTRODUCTION

The identification of possible ligand-binding sites on protein receptor molecules is important for the prediction of ligand–receptor binding geometry and for specific ligand design. Although several current ligand–receptor docking approaches treat the ligand as at least partially flexible, most methods employ rigid receptor structures during the search for a putative ligand or ligand fragment.[1–7] Such approaches can be very successful if ligand docking to a receptor is performed in the so-called "bound" conformation that is a receptor structure determined in complex with a ligand. The structure of a protein crystallized in an "unbound" form can deviate from the conformation in the ligand bound form,[6,8] such that computational docking may result in unfavorable sterical overlaps between ligand and receptor even if the ligand has been placed correctly into the putative binding pocket.[4,8–11] This can either lead to complete failure of the docking attempt or to an unfavorable ranking of the putative binding site and geometry.[9–11] The application of comparative protein homology-building techniques allows generation of structural models of proteins with sequence similarity to a known protein structure.[12] Since many proteins of biological and pharmaceutical interest can be modeled based on sequence similarity to a known protein fold, a desirable long-term goal is to use such structures in computational ligand-docking studies.[13] However, depending on the sequence identity relative to the target structure and the details of the modeling procedure modeled, protein structures can contain inaccuracies that may interfere with ligand–receptor docking efforts for reasons similar to those outlined above for ligand docking to rigid "unbound" protein structures.

Complex formation can not only lead to local conformational changes in a receptor protein such as side-chain or protein loop rearrangements but can also cause global changes that correspond to adjustments of secondary structural elements or whole domains.[6] Given the need to obtain docking results within seconds of workstation time, a full inclusion of all Cartesian receptor degrees of freedom during docking is currently not computationally feasible. Even if possible, treating the receptor protein as fully flexible may lead to very unrealistic complex conformations due to force field limitations and neglect of solvent molecules. In some docking approaches, a selected set of protein side-chains close to the putative binding site is flexible (e.g., Schapira et al.[14]) or is represented as a discrete set of rotamers.[15,16] Part of the protein backbone can also be treated by discrete sets of backbone structures compatible with the protein three-dimensional (3D)-fold.[16]

Grant sponsor: Deutsche Forschungsgemeinschaft; Grant number: SFB 604.

Correspondence to: Martin Zacharias, Computational Biology, School of Engineering and Science, International University Bremen, Campus Ring 1, D-28759 Bremen, Germany. E-mail: m.zacharias@iu-bremen.de

Received 20 May 2003; Accepted 26 August 2003

Other methods approximately account for receptor flexibility by representing the receptor target as an ensemble of structures.[17,18] The structural ensemble can, for example, be a set of structures obtained from NMR spectroscopy. In these approaches, the receptor-binding pocket and its interaction with a ligand is represented by a composite grid that corresponds to a weighted average over grids, each representing one of the various receptor conformations. In an approach by Schafferhans and Klebe,[13] the receptor-binding site is represented by Gaussian functions that describe physical properties of the binding site and the ligand is docked onto the smooth binding-site descriptors. In these approaches, however, an atomic resolution sterical fit between ligand and receptor is not performed. In the so-called relaxed complex method, an ensemble of protein structures is generated using molecular dynamics (MD) simulations. Subsequently, docking approaches that assume a rigid receptor are applied to dock putative ligands to the individual conformational snapshots of the simulation.[19–21] This approach has shown promising results on test cases;[21] however, it is computationally expensive, since, depending on the size of the conformational ensemble, docking to many target receptor structures (possibly several hundred) needs to be performed.

Ligand binding to a receptor molecule can involve conformational changes of the receptor that are nonlocal; even small global conformational changes of less than 1 Å root-mean-square deviation (RMSD) between receptor bound and free structures can significantly affect ligand–receptor interaction.[6] However, approaches that include a fully flexible receptor structure during docking are computationally very demanding and limited to single (or very few) putative ligands and possible binding geometries.[22–24] In order to approximately account for global receptor flexibility in docking simulations, Zacharias and Sklenar[25] suggested use of soft harmonic modes as additional variables during docking. Such collective degrees of freedom can be calculated at an energy minimum of the receptor structure (by diagonalization of the second derivative matrix of the energy with respect to the atomic coordinates). The modes as additional variables allow rapid relaxation of the receptor structure during docking and estimation of receptor deformation energy. It also avoids the computationally costly calculation of the internal receptor energy at every docking minimization step. An additional advantage to docking methods that employ ensembles of discrete (rigid) receptor structures is that the receptor conformation can change continuously during docking in the precalculated soft degrees of freedom and has therefore a much greater capacity for induced fit adaptation. However, precalculation of harmonic modes of a large receptor molecule requires very extensive energy minimization and is usually performed in the absence of solvent. Energy minimization under these conditions can lead to large deviations from a realistic receptor geometry, and the calculated soft harmonic modes may not correspond to realistic soft degrees of freedom.

Calculation of flexible degrees of freedom of a protein molecule can be performed under more realistic conditions using a principal components analysis (PCA) of motions obtained during an MD simulation including surrounding waters and ions.[26]

In the present study, the possibility of using soft degrees of freedom calculated during an MD simulation as additional variables to dock a ligand molecule (immunosuppressant FK506) to an "unbound" conformation of a FK506-binding protein FKBP[27–29] has been explored. Accounting for relaxation in the precalculated soft modes of the receptor significantly improves the docking performance compared to docking to a rigid "unbound" FKBP receptor structure at a modest increase in computational demand. The approach might be useful for rapid preselection of putative ligand–receptor complexes that can be refined using a more accurate representation of receptor flexibility.

MATERIALS AND METHODS

Molecular Dynamics Simulations

The atomic coordinates for the crystal structures of the FKBP12 rotamase (a FK506-binding protein) in the ligand-free (unbound) form [Wilson et al.[29]; Protein Data Bank (PDB) entry:1fkk] were used as starting structures for MD simulations. The AMBER (Assisted Model Building with Energy Restraints) suite of programs,[30] with the Cornell et al. force field[31] was used for all simulations. The crystal structure was prepared for the simulations using the Leap module of AMBER. The protein was neutralized by adding counterions and solvated by adding ~2500 transferable intermolecular potentials 3 (TIP3) water molecules.[32] For the nonbonded short-range interactions, a 9.0 Å cutoff was used. The particle mesh Ewald summation technique with a grid size of 0.9 Å was employed to calculate long-range electrostatic interactions for distances greater than 9.0 Å.[33] The conformation of the solvated protein was first relaxed via energy minimization. Following minimization, the system was gradually heated from 50 to 300 K, with positional restraints on the protein atoms over a period of 0.1 ns. During another 0.1 ns simulation time at 300 K, the positional restraining force constant was gradually reduced from 50 kcal/mol^{-1} Å^{-2} to zero. The simulation system was further equilibrated without any restraints for 0.4 ns followed by a 1 ns data gathering period (2 fs time step). MD simulations were performed at constant pressure of 1 bar with relaxation time of 5 ps. Solute coordinates were stored each 0.4 ps simulation time.

During the data-gathering period, the simulated structure stayed close to the experimental structure (Fig. 1). A PCA of the trajectory (heavy atoms) was performed as described.[34–36] The positional covariance matrix,

$$C_{ij} = 1/N_k \cdot \sum^{k} [x_i(k) - \langle x_i \rangle] \cdot [x_j(k) - \langle x_j \rangle]$$

(x_i atomic positions, summation, k, is over all structures, N_k, in the trajectory) was diagonalized. The resulting eigenvectors with large eigenvalues (principal components, PCs) describe flexible orthogonal degrees of freedom of the protein molecule. The eigenvalues are a measure of the conformational fluctuation or flexibility in the corre-

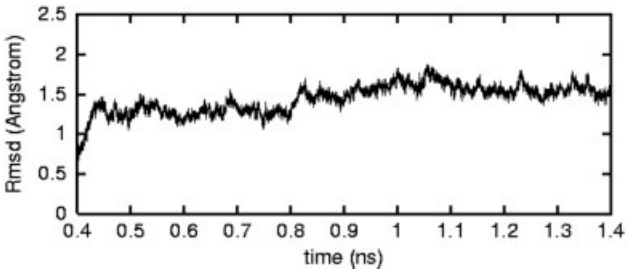


Fig. 1. The RMSD time course of the FKBP-binding protein simulation (after 0.4 ns equilibration) with respect to the first structure in the trajectory.

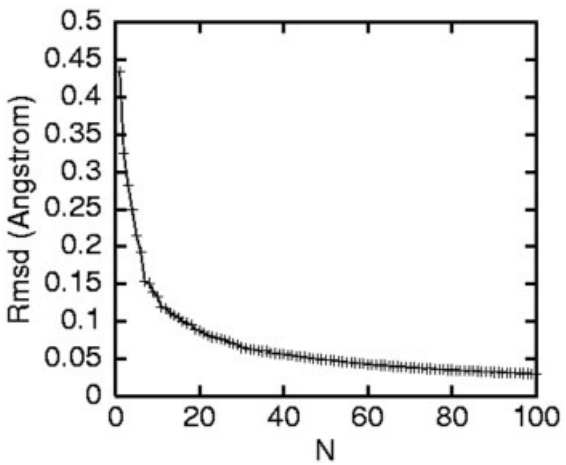


Fig. 2. Contribution of the first 100 principal components (with largest eigenvalues) to the FKBP-binding protein atomic fluctuations.

sponding eigenvector direction. The contribution of each calculated PC to the protein deformation observed during the simulation is shown in Figure 2. As has also been found in other MD simulation studies of proteins and nucleic acids, most of the observed conformational fluctuations (~80%) are due to motions in the first ~10–20 PCs with large eigenvalues.[26,34–37] These principal flexible directions are also called essential modes,[26] or soft collective degrees of freedom (soft modes).

Ligand–Receptor Docking Calculations Accounting for Receptor Flexibility

A docking program has been written (PCRELAX) that allows minimization of the position and orientation of a ligand molecule relative to a receptor molecule. During the minimization, all interligand–receptor interactions are calculated. For the present ligand–receptor system, only sterical interactions between heavy atoms have been considered using the Amber/optimized potentials for liquid simulations (OPLS) united atom parameters.[38] In addition to ligand orientation and translation with respect to the receptor, the PCRELAX approach allows simultaneous deformation or relaxation of the protein–receptor conformation in the direction of the precalculated PCs of motion (only for heavy atoms). The eigenvalues for the calculated PCs are a measure of the deformability in PC direction and can be used to derive an intrinsic energy function for protein deformations. A possible choice could be a quadratic energy function for a selected PC, with a force constant that limits deformations (within an energy change of RT: R, gas constant; T, temperature) up to the range observed during the simulation. Since the eigenvalue gives the square of the coordinate fluctuations in the corresponding eigenvector (PC) direction, an appropriate choice would be RT/eigenvalue. This would lead to a receptor energy increase of RT for conformational changes in the order of the value of the eigenvalue in the corresponding PC direction (compatible with the magnitude of the fluctuations observed during the simulation in that direction). However, test calculations indicated that a steeper energy increase for larger deformations leads to less severe and unrealistic protein deformations for unfavorable ligand placements. Therefore, an energy function $V(\mathbf{PC_i})$ proportional to the fourth power in the magnitude of the PC deformation was used:

$$V(\mathbf{PC_i}) = 1.5\ RT/\text{eigenvalue}^2((\mathbf{R_{rec}} - \mathbf{R_{rec0}}) \cdot \mathbf{PC_i})^4 \quad (1)$$

$\mathbf{R_{rec}}$ indicates the actual receptor coordinates and corresponds to the receptor coordinates for the fully relaxed protein structure (vectors are in bold, · indicates a scalar product). The term $(\mathbf{R_{rec}} - \mathbf{R_{rec0}}) \cdot \mathbf{PC_i}$ gives the scalar product between receptor deformation and the ith PC and corresponds to the contribution (projection) of the receptor deformation in the direction of the $\mathbf{PC_i}$. The force constant for a deformation in a selected PC was $1.5*RT$/eigenvalue^2. This choice is equivalent to the choice in case of using a quadratic energy function (described above) and results in an energy increase of $1.5*RT$ upon deformation of the receptor protein in PC direction by an amount equivalent to the corresponding eigenvalue (the eigenvalue^2 gives the fourth power of the equilibrium deformation in the PC direction observed during the MD simulation). The energy function proportional to the fourth power in magnitude of the PC deformation and the factor 1.5 allow considerable freedom for small deformations but more strongly limit larger receptor deformations. It creates a strong penalty for deformations in eigenvector direction that go beyond deformations observed during the MD simulation.

The target energy function that is minimized in the present docking simulations consists of Lennard–Jones terms that evaluate the sterical complementarity between ligand and receptor, and an intrinsic receptor deformation energy. The latter term can be calculated very rapidly avoiding calculation of the intrareceptor energy as a sum of intrareceptor atom pairs at every minimization step. In addition, the directions that are allowed for receptor deformation have been obtained as a set of flexible degrees of freedom under more realistic simulation conditions than vacuum conditions that are typically used during docking. A Quasi-Newton minimizer that requires only first derivatives of the energy function (no mixed second derivatives) was used for all energy minimizations. Minimization in the rotational and translational degrees of freedom was

TABLE I. Docking of FK506 to the FKB Receptor

	Rigid docking to experimental receptor structure	Rigid docking to average MD receptor structure	Docking to average MD receptor structure with PC[a] relaxation
$E_{interaction}$ (kcal/mol)	−58.5	128.2	−49.4 (−54.4)
$RMSD_{ligand}$ (Å)	0.4	1.45	1.2
$RMSD_{proteinCA}$ (Å)	0.0	1.5	1.45
$RMSD_{pocketRes}$ (Å)	0.0	1.8	1.45

$E_{interaction}$ (kcal/mol) is the calculated sterical ligand receptor interaction energy. It includes both the receptor deformation energy (see Methods section) and the ligand–receptor interaction energy (values in parenthesis give only the ligand–receptor interaction energy). $RMSD_{ligand}$ (Å) is the root-mean-square deviation of the ligand atoms from the position obtained in the experimental complex structure (PDB code: 1fkj) after superposition of the minimized complex on the experimental structure (with respect to protein atoms). $RMSD_{proteinCA}$ (Å) is the deviation of the receptor structure (after docking) from the experimental receptor structure in the complex (PDB code: 1fkj). $RMSD_{pocketRes}$ (Å) is the deviation of side-chains that contact the ligand (residues: 26, 36, 37, 46, 54, 55, 56, 59, 82, 87, 99) of the final receptor conformation from the corresponding position in the experimental structure (after superposition of the protein backbone).
[a]18 principal components obtained from a molecular dynamics simulation of the unbound FKBP protein have been included during docking.

possible by using gradients with respect to centers of geometry of the ligand and Euler angles that describe the angular orientation of the ligand. For the PCs, the forces consist of two contributions. The first contribution is due to ligand–protein receptor interactions (in the present case only Lennard–Jones interactions) that exert forces on receptor atoms. The receptor protein is only allowed to relax in the direction of the PCs. The force in the direction of a $\mathbf{PC_i}$ can be calculated as the projection of the receptor force vector $\mathbf{F_{rec}}$ onto the PC direction. Second, the penalty energy function [Eq. (1)] limits the deformation in the PC direction. The overall gradient $\mathbf{g}_i$ in the direction of a selected $\mathbf{PC_i}$ is given by

$$\mathbf{g}_i = 6.0\,RT/\text{eigenvalue}^2((\mathbf{R_{rec}} - \mathbf{R_{rec0}}) \cdot \mathbf{PC_i})^3 - \mathbf{F_{rec}} \cdot \mathbf{PC_i} \quad (2)$$

The first term in Eq. (2) is the derivative of the penalty energy function [Eq. (1)] and the second term corresponds to the forces due to ligand–receptor interactions (the negative sign is due to the fact that a force equals minus the gradient).

A complete minimization in ligand orientational and translational degrees of freedom and up to 18 soft receptor deformation modes can be performed in less than 0.5 s on a reasonable workstation (AMD Athlon, 1.8 GHz) down to very low residual energy changes per step ($<10^{-5}$ kcal/mol^{-1}). It should be noted that docking energy minimization (EM) of complex geometries allows a more reproducible evaluation of complexes than simple ligand placement into the binding pocket (small changes can greatly affect the evaluation). In this regard, a rapid convergence down to very small residual gradients and energy changes allows also a more reproducible evaluation than EM in Cartesian coordinates. In the case of Cartesian coordinates, achieving convergence is very demanding; therefore, final energies and docked conformations can be strongly dependent on the number of minimization steps.

RESULTS

Docking of FK506 in the Position And Orientation Observed in the Experimental Complex

Minimization of the FK506 ligand position and orientation in the experimental complex structure between FKBP and FK506 (Wilson et al.[29]; PDB code:1fkj) resulted in a placement with an RMSD of only 0.4 Å from the corresponding position in the experimental complex structure. It resulted in a sterical interaction energy of −58 kcal/mol^{-1}. To some degree this is expected, since the protein taken from the complex structure contains already a "preformed" cavity that optimally accommodates the ligand. The energy minimized average structure obtained from the MD simulation (starting from the free FKBP structure) served as a model for the free "unbound" receptor protein. The FK506 ligand was placed at the same position relative to the average receptor structure obtained from the MD simulation and in the same conformation as observed in the known crystal structure. This was achieved by superimposing the crystal structure of the complex onto the minimized average MD structure. Although the RMSD (protein backbone) between minimized average MD structure and crystal structure of the FKBP protein in the experimental complex was only 1.5 Å the ligand showed some sterical overlap with receptor atoms. Energy minimization in orientational and translational degrees of freedom led to a geometry with an RMSD of ~1.45 Å from the experimental ligand position (this means the RMSD of only the ligand with respect to the experimental placement) and still a positive interaction energy of ~128 kcal/mol^{-1} (see Table I). The sterical strain could not be completely relieved, since some receptor atoms hinder the ligand's escape from the binding pocket.

In contrast, using the same receptor structure and the same initial ligand placement but allowing for relaxation in the 18 softest precalculated PCs from the MD simulation of the protein resulted in a calculated binding energy of −49.4 kcal/mol^{-1} and a placement of the ligand with an RMSD of 1.2 Å with respect to the experimental ligand

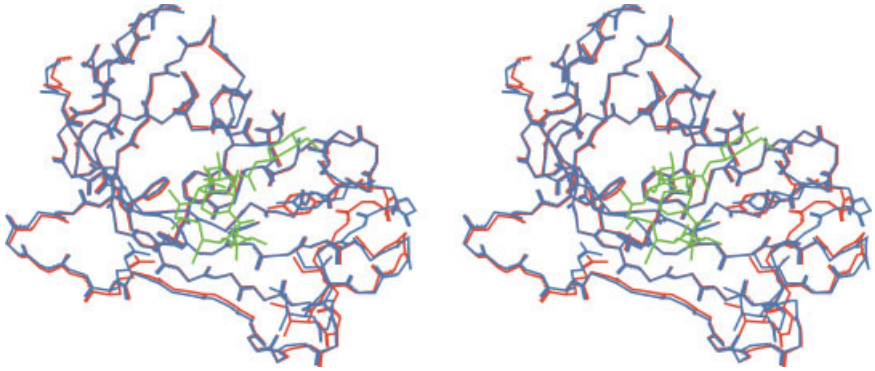
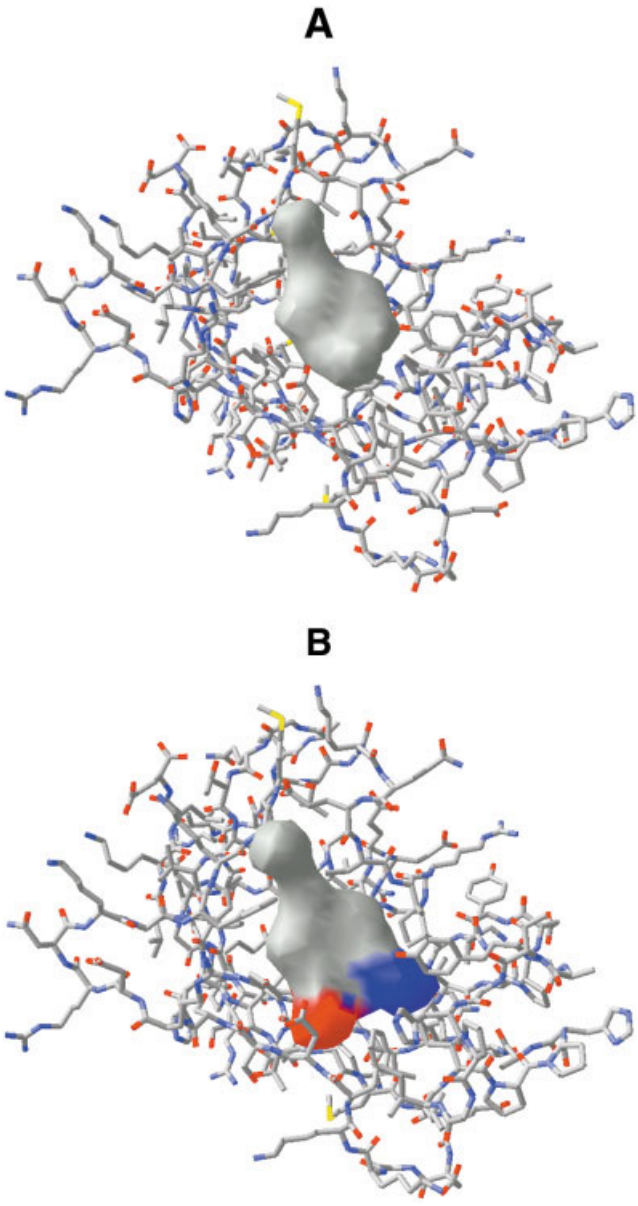


Fig. 3. Stereo view of the minimized average FKBP binding protein structure from MD before (blue) and after relaxation (red) in the 18 softest principal components (obtained from the MD simulation) upon docking of FK505 (green). For clarity, only the protein main-chain and side-chain heavy atoms of a few residues that participate in ligand contacts are included (Tyr26, Arg42, Phe46, Trp59, Tyr82, Ile90). All heavy atoms of the FK506 ligand are shown.



position. As indicated in Figure 3, most of the structural adaptation is due to small global adjustments of a protein loop region that relaxes the protein such that it can accommodate the ligand in the binding pocket. Although the RMSD of the relaxed protein structure from the average MD structure was only ~0.8 Å this small adaptation dramatically improves the sterical interaction between receptor and ligand from an unfavorable +128 kcal/mol^{-1} down to −54.4 kcal/mol^{-1} (the difference to the total complex energy given above is due to the implicit receptor deformation energy; see Table I). This is close to the value obtained for the experimental complex structure with a "preformed" ligand-binding pocket. It is also interesting to note that the dramatic increase in ligand–receptor complementarity is achieved by a relatively small readjustment of the ligand position and orientation, and a small cost in receptor deformation energy of ~5 kcal/mol^{-1} (this is the difference between complex energy of −49.4 kcal/mol^{-1} and the ligand–receptor interaction energy of −54.4 kcal/mol^{-1}), which corresponds to less than RT per PC (18 PCs have been included).

The relaxed protein structure is overall in slightly closer agreement with the receptor structure observed experimentally than the minimized average MD structure. However, if one compares only the side-chains involved in interactions with the ligand after backbone superposition of relaxed protein and experimental receptor, a more significant movement toward the experimental geometry was observed (Table I). This indicates that the relaxation of the protein in the soft modes leads to an improved rearrangement of the binding pocket in slightly closer agreement with the conformation in the experimental complex struc-

Fig. 4. Mapping the deformability of the FKBP ligand-binding pocket. A spherical probe was docked at various positions of the ligand-binding pocket. During docking, a weak harmonic restraint forced the probe toward the center of the protein (see text for details). The resulting accessible regions to the probe are shown as filled gray volumes in case of a rigid receptor (**A**) and as allowing for protein relaxation in the 18 most flexible principal components of motion obtained from the MD simulation (**B**). Two deformable regions are indicated in red (near residues Tyr26 and Phe36) and blue (near residues Tyr81 and Ile99). The FKBP protein structure is shown as a stick model (only heavy atoms).

TABLE II. Minimization of Displaced Ligands

Initial ligand placement		Rigid docking to experimental receptor structure		Rigid docking to average MD receptor structure		Docking to average MD receptor structure with PC relaxation	
N	$RMSD_{start}$ (Å)	E(kcal/mol)	RMSD (Å)	E(kcal/mol)	RMSD(Å)	E(kcal/mol)	RMSD(Å)
1	0.5	−58.6	0.4	128.2	1.5	−49.5 (−54.4)	1.2
2	0.8	−58.6	0.4	128.2	1.5	−49.5 (−54.4)	1.2
3	1.0	−58.6	0.4	−36.6	3.3	−39.5 (−41.4)	2.2
4	1.1	−58.6	0.4	−32.0	3.0	−37.9 (−38.6)	2.8
5	1.8	−58.6	0.4	128.2	1.5	−49.5 (−54.4)	1.2
6	2.1	−58.6	0.4	128.2	1.5	−49.5 (−54.4)	1.2
7	2.5	−31.8	3.7	−36.6	3.3	−43.4 (−45.4)	3.2

$RMSD_{start}$(Å) is the deviation of the ligand start placement from the experimentally found ligand placement. For each docking experiment (see legend of Table I) the ligand–receptor complex energy, E, including receptor deformation energy and ligand–receptor interaction energy (values in parenthesis give only the ligand–receptor interaction energy) and final RMSD of the ligand (from the experimental position) are given.

TABLE III. Systematic Docking Minimization Starting Outside of the FKBP Binding Pocket

	Lowest energy solution		2nd lowest energy solution		Solution closest to experiment		
Target receptor	E (kcal/mol)	RMSD (Å)	E (kcal/mol)	RMSD (Å)	Rank	E (kcal/mol)	RMSD (Å)
Exp. receptor	−58.5	0.4 (7)[a]	−43.8	8.3 (5)	1	−58.5	0.4 (7)
Mean MD structure − PCs	−40.4	7.7 (11)	−40.0	8.7 (4)	11	−36.8	3.3(2)
Mean MD structure + PCs	−49.6 (−54.5)	1.2 (4)	−47.4 (−49.4)	9.1 (2)	1	−49.6 (−54.5)	1.2 (4)

Energies include receptor deformation and ligand–receptor interaction (numbers in parenthesis are sterical interaction energies).
[a]Numbers in parenthesis indicate the number of times the docking minima were obtained.

ture than the average MD structure of the free receptor protein. One needs to keep in mind that the "free" receptor structure deviates in several regions (far away from the binding pocket) from the "bound" form, and it is unlikely that an improved receptor adaptation at the binding pocket leads to a simultaneous closer agreement between free and bound receptor structures in other regions.

Docking From Slightly Misplaced FK506 Positions and Orientations

In order to test whether there are alternative ligand-docking minima close to the experimentally observed placement, the initial position and orientation of the ligand was slightly and randomly distorted (7 start structures with increasing RMSD from the experimental position, see Table II). Docking minimization using the experimental bound receptor conformation as target yielded the same minimum structure as observed when starting from the experimental placement in 6 cases. Only the minimization starting from position 7 with the largest deviation, with respect to the experimental placement, yielded a new (less favorable) docking minimum with significant deviation from the experimental binding geometry (RMSD: 3.7 Å). For rigid docking to the average MD receptor structure (minimizing ligand in translational and orientational variables), the same minimum as observed starting from the experimental ligand placement (see above and Table I) was found in 4 cases resulting in reasonable agreement with the experimental binding geometry (RMSD of the ligand: 1.5 Å) but a very unfavorable interaction (128 $kcal/mol^{-1}$, see Table II). For three initial ligand placements, docking mimization yielded binding geometries with a negative interaction energy but significant deviation from the experimental binding placement (RMSD > 3 Å, the ligand escapes from the binding pocket). The case of docking including the 18 softest receptor modes gave a minimum close to experiment (RMSD of the ligand: 1.2 Å, same minimum as obtained when starting from experimental ligand placement) for 4 start geometries and a very favorable sterical ligand–receptor interaction. For 3 cases, docking minima outside the binding pocket were reached. However, in contrast to rigid docking, these alternative minima are predicted to be less favorable than the placement close to the experimental binding position.

Systematic Docking From Multiple Start Sites Outside of the Binding Pocket

For systematic docking studies, two random start positions outside of the ligand-binding pocket were selected such that the distance to any protein atom was slightly larger (by ~2.5 Å) than the radius of the FK506 ligand (the radius is here the maximum distance of any atom from the center of FK506). At these positions, ~150 different orientations of the ligand were generated (such that an approximately uniform density on the Euler angle sphere was achieved). The start geometries were minimized in two phases: In a first minimization, a harmonic distance constraint between the center of the protein and the ligand

atom that is closest to the protein center was applied, followed by a second minimization without constraints. In case of rigid docking (i.e., using translational and rotational degrees of freedom for the FK506 ligand) to the experimental receptor in the bound conformation, a docking geometry identical to the energy-minimized experimental complex was obtained 7 times. This geometry had the lowest energy (-58.5 kcal/mol^{-1}) and is well separated from the second-lowest energy docking geometry, with an RMSD of 8.3 Å from the experimental placement and an energy of -43.8 kcal/mol^{-1} (Table III). The result indicates that rigid docking, even if one uses only sterical complementarity as a scoring energy term, can be quite selective and accurate if one uses the receptor in the "bound" conformation.

However, in case of using the same procedure with exactly the same number of start structures and start geometries but using the "unbound" average MD receptor structure leads to a lowest energy docking geometry that strongly deviates from the experimental placement. In this case, the lowest energy ligand placement had a RMSD of 7.7 Å with respect to the experimental ligand placement. In addition, the ligand-binding geometry that comes closest to experiment ranks is at position 11 (out of ~70 docking minima) and this geometry still shows a large deviation from the experimental placement (RMSD: 3.3 Å, Table III). In contrast, using the mean MD receptor structure and including relaxation of soft receptor modes during systematic docking yielded a ligand-docking geometry close to experiment (RMSD: 1.2 Å, the same docking minimum as the one obtained by directly minimizing the experimental complex) as the lowest energy docking geometry. This geometry was obtained 4 times (among ~70 different docking minima).

The result indicates that rigid ligand docking with an "unbound" target conformation of the FKBP fails to identify a realistic ligand-docking geometry, since small but significant conformational adjustments in the receptor are necessary to accommodate the ligand in the experimental geometry. In contrast, the present docking approach that includes relaxation of precalculated soft modes of the receptor identifies a geometry close to experiment as lowest energy structure with a sterical interaction energy that comes close to the interaction energy in the case using the bound receptor conformation.

Mapping Deformability of the Ligand-Binding Pocket

Besides docking complete ligands to a binding pocket, the present approach can also be used to rapidly map out the flexibility of a putative receptor binding pocket. For this purpose, a number of spherical probes (with carbon Lennard–Jones parameters) were placed at the surface of the ligand-binding region of the FK506 binding protein [Fig. 4(a)]. A number of reference points (~40) at a distance of 10 Å from one of the spheres, which was located approximately in the middle of the ligand-binding site, were generated randomly. An additional constraint was that the reference points were located within 12 Å of the receptor center. In order to map out the deformability of the ligand-binding side, the positions of the spherical probes were minimized, applying harmonic restraints between the various reference points and the probe atoms (requiring one minimization for each probe and each reference point). The harmonic restrain between surface probe and the various reference points forced the probe to move toward the protein (in a number of slightly different directions depending on the position of the reference points). As a result, the probe atoms were forced to move toward cavities and holes in the binding region in order to reduce the distances to the reference points. As illustrated in Figure 4(a), the inclusion of soft mode receptor relaxation leads to larger movements of the probes and a significant increase of the binding pocket in certain directions compared to probe docking to a rigid receptor [Fig. 4(b)]. The relaxation of the receptor in the precalculated soft modes allows opening up cavities that are inaccessible in the case assuming a rigid receptor. In the case of soft mode receptor relaxation, these small conformational adjustments require only a small increase of the (implicitly) calculated receptor deformation energy. The allowed deformation energy during the probe docking was ~0.5 kcal/mol^{-1} (~RT, R, gas constant; T, temperature) for most final probe positions. Significant deformability of the binding pocket was found, for example, close to residues Ile99 and Tyr81 [indicated in blue in Fig. 4(b)]. These two residues are part of a long loop (residues 79–94) that participates also in contacts in case of a bound FK506 ligand. It is the ligand-binding region that undergoes the largest displacement upon FKBP-FK506 complex formation (upon docking including receptor relaxation). Another deformable region was identified near Tyr26 and Phe36 of FKBP [shown in red in Fig. 4(b)]. Identification of such deformable regions could help to suggest how to extend a given ligand, still allowing it to bind to the pocket.

DISCUSSION

Ligand–receptor docking approaches aim at solving three tasks: selection of putative ligands that may bind to a given receptor binding pocket; prediction of a realistic binding geometry; and selection of this geometry as the lowest energy or highest ranking docking solution. The fulfilment of these goals combined with the computational demand to allow rapid virtual screening of a large set of possible ligands requires introduction of a number of approximations during the docking process. One of the most common approximations that greatly increases the speed of computational docking methods is to assume a rigid receptor structure or to allow for only local adjustments (e.g., allowing limited flexibility of some side-chains close to the binding site) during docking in most current approaches. However, even in the case of using a rigid receptor and flexible ligands, the virtual screening of a database of ligand structures can take weeks or months on multiprocessor computers (e.g., Schapira et al.[14]). In particular, for docking to "unbound" receptor structures, average NMR structures, or model built protein structures, the assumption of a rigid receptor–protein structure

during docking may limit the usefulness of rigid receptor docking methods. In the present example, the minimized average FKBP protein structure obtained from MD simulations can serve as a typical example for an "unbound" structure or a modeled receptor structure. Although this structure shows only a small deviation from both the "unbound" FKBP structure and the "bound" form (RMSD to both structures ~1.5 Å), the ligand does not sterically fit into the known binding pocket even upon minimization in translational and orientational degrees of freedom. Systematic docking minimization starting from various positions and orientations near the known binding site or from start positions slightly shifted from the experimentally observed placement failed to identify any geometry close to experiment as low-energy binding placement. This result indicates that accounting for receptor flexibility during docking can be very critical even in the case of using a known high-affinity ligand. It is likely that docking attempts can fail for the same reason in the case of other receptor–protein structures in the unbound form or for modeled structures that are of limited precision.[9–11] In contrast to rigid docking, inclusion of precalculated soft modes that represent the flexible directions of motion for FKBP during docking resulted in complexes close to experiment as the top-ranking solution both for systematic docking and starting from slightly misplaced ligand positions. The sterical interaction energy was almost as low as the interaction energy in the case of rigid docking to the (preformed) "bound" receptor conformation. This result indicates that the present docking approach is a clear improvement compared to rigid docking in the case of using a receptor structure in the unbound conformation, at a modest additional computational cost of about a factor of ~3–5. It should be emphasized that energy minimization of the full receptor structure in Cartesian coordinates (or other internal coordinates) during docking is much more time-consuming than the present approach for two main reasons. First, such an approach requires recalculating the full force field energy at every minimization step; second, due to the larger number of variables, it shows also much slower convergence (requires many thousand minimization steps). In addition, the receptor minimization during docking needs to be performed in the absence of solvent, which may result in unrealistic conformations. In the present method, the flexible receptor degrees of freedom have been precalculated under relatively realistic conditions, including surrounding aqueous solvent and ions. Full minimization to residual energy changes of $<10^{-5}$ kcal/mol^{-1} require less than 100 minimization steps. Second, receptor energy changes are estimated from the intrinsic deformability of the protein in each soft mode (magnitude of conformational fluctuation in the selected mode observed during the MD simulation) and avoid calculation of receptor internal energy by evaluating all pairwise atom–atom interactions. Hence, the energy calculation at each docking minimization step is much faster than in the case of using all-atom receptor degrees of freedom.

A disadvantage of the present soft modes is, however, the fact that very mobile parts of the structure such as flexible surface side-chains are not very accurately represented by the precalculated modes. If such side-chains undergo extensive torsion angle flips during the MD simulation, the deformation of such side-chain atoms in the soft modes can result in a distortion of the bonded geometry. It needs to be emphasized that the primary focus of the present approach is not to provide an accurate ligand–receptor complex with a perfect stereochemistry but to account for receptor relaxation in a rapid but only approximate way in order to rapidly preselect putative ligands and binding geometries. In this regard, the approach appeared to be very successful. In a subsequent step, a greatly reduced set of selected low-energy ligand–receptor docking complexes can, for example, be further refined by energy minimization using all receptor and ligand degrees of freedom.

A possible extension of the method is to combine protein backbone motion represented by precalculated soft modes from an MD simulation with rapid side-chain rotamer building on the flexible protein backbone structure. Such approach could combine the description of global backbone protein motion by precalculated soft modes with a rotamer side-chain description more appropriate for describing possible states of mobile surface-exposed amino acids.

As shown in the Results section, the precalculated modes can also be used to map out the softness or flexibility of a ligand-binding pocket on the protein receptor. This was achieved by pushing a spherical probe (carbon atom) placed at the ligand-binding region toward the protein interior and allowing the protein to relax in the soft modes. The binding pocket can then be defined as the ensemble of spheres obtained from the probe docking approach. The resulting binding region is larger than the region defined during probe docking to a rigid receptor and might be useful to define regions of the ligand-binding pocket that are accessible upon small conformational changes in the receptor associated with only a small energy penalty. Such information might be useful for docking strategies based on a negative image representation of the binding pocket by spheres and for ligand design strategies that try to explore possible extensions of a ligand that can still be accommodated by structural receptor relaxation.

ACKNOWLEDGMENTS

My thanks to Drs. F. Pineda, J. Suehnel, and D. Roccatano for helpful discussions.

REFERENCES

1. Kuntz ID, Blaney J, Oatley S, Langridge R, Ferrin T. A geometric approach to macromolecule–ligand interactions. J Mol Biol 1982; 161:269–280.
2. Sobolev V, Wade R, Vriend G, Edelman M. Molecular docking using surface complementarity. Proteins 1996;25:120–129.
3. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
4. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol Pharmacol 2000;57:213–218.

5. Klebe G. Recent developments in structure-based drug design. J Mol Med 2000;78:269–281.
6. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
7. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struc 2003;32:335–373.
8. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
9. Murray CW, Baxter CA, Frenkel AD. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des 1999;13: 547–562.
10. Betts M, Sternberg M. An analysis of conformational changes on protein–protein association: implications for predictive docking. Protein Eng 1999;12:271–283.
11. Verkhivker, GM, Bouzida, D, Glhaar, DK, Rejto PA, Arthurs S, Colson AB, Freer ST, Larson V, Luty BA, Marrone T, Rose PW. Deciphering common failures in molecular docking of ligand–protein complexes. J Comput Aided Mol Des 2000;14:731–751.
12. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct 2000;29:291–325.
13. Schafferhans A, Klebe G. Docking ligands onto binding site representations derived from proteins built by homology modelling. J Mol Biol 2001;307:407–427.
14. Schapira M, Raaka BM, Samuels HH, Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. Proc Natl Acad Sci USA 2000;97:1008–1013.
15. Leach A. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 1995;235:345–356.
16. Clausen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
17. Knegtel RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.
18. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40.
19. Carlson HA, Masukawa KM, McCammon JA. Method for including the dynamic fluctuations of a protein in computer-aided drug design. J Phys Chem A 1999;103:10213–10219.
20. Lin JH, Perryman AL, Schames JR., McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc 2002;124:5632–5633.
21. Wong CF, McCammon, JA Protein flexibility and computer-aided drug design. Annu Rev Pharmacol Toxicol 2003;43:31–45.
22. Apostolakis J, Plueckthum A, Caflisch A. Docking small ligands in flexible binding sites. J Comput Chem 1998;19:21–37.
23. Mangoni M, Roccatano D, Di Nola A. Docking of flexible ligands to flexible receptors in solution using molecular dynamics simulations. Proteins 1999;35:153–162.
24. Pak Y, Wang S. Application of a molecular dynamics simulation method with a generalized effective potential to the flexible molecular docking problems. J Phys Chem B 2000;104:354–359.
25. Zacharias M, Sklenar H. Harmonic modes as variables to approximately account for receptor flexibility in ligand–receptor docking simulations: applications to DNA minor groove ligand complex. J Comput Chem 1999;20:287–300.
26. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
27. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. Science 1991;252:839–842.
28. Michnick SW, Rosen MK, Wandless TJ, Karplus M, Schreiber SL. Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. Science 1991;252;836–839.
29. Wilson KP, Yamashita MM, Sintchak MD, Rotstein SH, Murcko MA, Boger J, Thomson JA, Fitzgibbon MJ, Navia MA. Comparative X-ray structures of the major protein for the immunosuppressant FK506 (Tacrolimus) in unliganded form and in complex with FK506 and rapamycin. Acta Crystallogr D Biol Crystallogr 1995; 51:511–522.
30. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman PA. Comput Phys Commun 1995;91:1–41.
31. Cornell WD, Cieplak P, Bayley CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 1995;117:5179–5197.
32. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
33. Darden T, York D, Pedersen L. Particle mesh Ewald: an NlogN method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
34. Sherer, EC, Harris SA, Soliva R, Orozco M, Laughton CA. Molecular dynamics studies of DNA A-tract structure and flexibility. J Am Chem Soc 1999;121:5981–5991.
35. Wlodek ST, Clark TW, Scott LR, McCammon JA. Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. J Am Chem Soc 1997;119:9513–9522.
36. Zacharias, M. Comparison of molecular dynamics and harmonic mode calculations on RNA. 2000;54:547–560.
37. Levy RM, Srinivasan AR, Olson WK, McCammon JA. Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 1984;23:1099–1112.
38. Jorgensen WJ, Tirado-Rives J. The OPLS potential functions for proteins: Energy minimization for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;1110:1657–1666.