Protein−Ligand Docking Accounting for Receptor Side Chain and Global Flexibility in Normal Modes: Evaluation on Kinase Inhibitor Cross Docking

Andreas May and Martin Zacharias*

School of Engineering and Science, Jacobs University Bremen, Campus Ring 6, D-28759 Bremen, Germany

Received January 24, 2008

Efficient treatment of conformational changes during docking of drug-like ligands to receptor molecules is a major computational challenge. A new docking methodology has been developed that includes ligand flexibility and both global backbone flexibility and side chain flexibility of the protein receptor. Whereas side chain flexibility is based on a discrete rotamer approach, global backbone conformational changes are modeled by relaxation in a few precalculated soft collective degrees of freedom of the receptor. The method was applied to docking of several known cyclin dependent kinase 2 inhibitors to the unbound kinase structure and to cross-docking of inhibitors to several bound kinase structures. Significant improvement of ranking and deviation of predicted binding geometries from experiment was obtained compared to docking to a rigid receptor. The inclusion of only the soft collective degrees of freedom during docking resulted in improved docking performance at a very modest increase (doubling) of the computational demand.

Introduction

In recent years, considerable progress has been achieved in modeling the conformational flexibility of ligands during computational ligand−receptor docking.[1–4] Available approaches include geometric hashing in combination with fragment based docking[5] and Monte Carlo (MC)[6–9] or genetic algorithm[10] searches or conformational ensembles.[11] Although of similar importance, the problem of appropriate treatment of receptor flexibility has not been solved satisfactorily so far. Methods developed to tackle this problem are often computationally very demanding, including for example molecular dynamics (MD)[a] or MC simulations or methods based on docking to an ensemble of different receptor conformations. Especially during virtual screening of large drug-like compound libraries, the target protein structure is typically kept rigid or flexibility is allowed only for a few selected amino acid side chains. However, often ligand binding to a protein is accompanied by a variety of conformational changes on a local structural level (side chain or loop motions) as well as on a global level affecting the protein backbone geometry upon complex formation. It is desirable to include such conformational changes during receptor−ligand docking simulations. The increased availability of protein model structures generated based on homology to a known structure holds the potential to use such structures also in ligand−receptor docking studies. Depending on the degree of sequence similarity to the known structure, a structural model can contain errors of both side chain and backbone conformation that requires the inclusion of backbone and side chain flexibility during docking.

Several available docking programs can include side chain conformational changes at a proposed ligand binding site at the level of searching for optimal discrete side chain dihedral angle combinations (rotamers). However, this is computationally already much more demanding than docking to a rigid receptor. Therefore, explicit inclusion of side chain and backbone coordinates has been considered even less frequently because of the significant further increase of the computational demands.[10] Part of the protein backbone can be treated by discrete sets of (rigid) backbone structures compatible with the three-dimensional fold of the protein.[12] Other docking methods approximately account for flexibility by representing the target receptor as an ensemble of (rigid) structures.[11,13] Although fully flexible methods (for example based on MD simulations) have also been used to evaluate ligands docked into receptor binding sites[14] the applicability is restricted to one or a few ligands and possible binding sites because of the large computational demand.[15]

In addition to appropriately accounting for conformational changes, realistic scoring of predicted solutions is another important issue during docking. The difficulties of flexible docking and scoring are often considered separately. However, because scoring functions are often exclusively parametrized using crystal structures of protein−ligand complexes, it is important to generate high-quality structural models already in the sampling phase. Appropriate treatment of conformational flexibility during docking is tightly connected to the improvement of scoring of a docked ligand−receptor geometry. Highly accurate scoring of a ligand placement is only possible if the complex geometry has also been predicted with high precision. A scoring function that tolerates many errors in a given complex can only provide limited specificity to distinguish between incorrect placements of the ligand and/or to rank realistically high affinity vs low affinity ligands. Consequently, there is a direct relation between the robustness or softness of a scoring function and the number of false positives obtained in a virtual screen. The design of more rigorous and more specific scoring functions requires at the same time an improvement of the prediction accuracy of binding modes in terms of deviation from the experimental binding geometry.

The experimental analysis of protein motions and results of MD simulations in recent years indicate that in many cases the conformational fluctuations near an equilibrium state can be described by a few collective degrees of freedom.[16–19] These

* To whom correspondence should be addressed. Phone: ++49-421-200-3541. Fax: ++49-421-200-3249. E-mail: m.zacharias@jacobs-university.de. Address: School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany.

[a] Abbreviations: ANM: anisotropic network model; ATP: adenosine triphosphate; CDK2: cyclin dependent kinase 2; GNM: Gaussian network model; ENM: elastic network model; MD: molecular dynamics; MC: Monte Carlo; RT: rotamer trial.

10.1021/jm800071v CCC: $40.75
Published on Web 06/03/2008

degrees of freedom can be obtained by the principle component analysis of the covariance of atomic fluctuations during MD simulations (also termed essential dynamics analysis).[16] Alternatively, normal-mode analysis, that is, the analysis of the curvature of the potential energy function around an energy minimum, can be used to extract collective degrees of large mobility in proteins. Zacharias and Sklenar suggested the use of such softest modes from a normal-mode analysis as additional variables during flexible docking and applied it to ligand binding to the DNA minor groove.[20] Instead of normal modes, it is also possible to employ essential modes from the MD simulations as has been explored for flexible docking of an immunosuppressant to an "unbound" protein receptor structure using the program PCRelax.[21] Use of soft modes as additional variables allows for rapid relaxation of the receptor structure during docking (induced fit) and for an estimation of the receptor deformation energy. The amplitude of motions in soft modes during docking is much larger than during typical force field based energy minimization in independent Cartesian coordinates of protein atoms. The approach can in principle lead to a dramatic reduction of the computational complexity to account for receptor flexibility during docking. This may form the basis for systematic docking, including global conformational changes in the receptor.

Gaussian network models (GNM) and related anisotropic network models (ANM) of protein motion[17,22] have become popular to investigate the conformational flexibility of proteins based on a simple spring model and on the assumption that the mobility of a protein region is mainly determined by the local density (or the locally available free space). Harmonic mode analysis of this simple energy function allows one to very rapidly identify possible flexible (soft) collective degrees of freedom (soft modes) of the protein within a few minutes of computer time.[22] Tama and Sanejouand[23] found that such approximate mode calculations resulted in predicted soft modes that show significant overlap with observed conformational changes in proteins determined experimentally under different conditions (e.g., different crystal forms or apo vs bound form of a protein molecule). In some cases, over 50% of the conformational difference between two structures of a protein, determined, for example, by X-ray analysis of two crystal forms or as ligand-free and bound forms, could be assigned to a single approximate soft mode of the protein.[23]

The application of ANM derived soft modes as flexible variables during protein−protein docking was systematically explored by May and Zacharias[24−26] by employing a reduced (coarse-grained) model for the protein partners. Inclusion of up to five softest modes for protein partners during docking improved the docking results at a very modest increase of computer time by a factor 2−3 compared to rigid docking. However, the systematic test on several protein−protein complexes also indicated that, in order to achieve realistic docking predictions, both side chain flexibility and global flexibility need to be accounted for simultaneously during docking.[25]

In the current study, such an approach is presented and applied to the problem of ligand docking to protein kinases. Protein kinases are important drug targets and the 3D structures of many protein kinases in complex with different inhibitors have been determined using X-ray crystallography. The unbound (apo) structure of protein kinases often differs significantly both in terms of side chain but also in terms of the backbone structure around the inhibitor binding site (typically near the ATP binding site) from the inhibitor bound form (by several Ångstroms). In addition, even the structures of inhibitor-bound forms show significant variability, especially of the protein backbone structure. Hence, application of one selected inhibitor bound form as a docking target to search for new putative ligands may miss putative ligands because of the neglect of backbone conformational adaptation during docking. The set of known structures of the protein kinase CDK2 (cell-cycle dependent kinase 2) in complex with different inhibitors offers to test the docking methodology by cross-docking the inhibitors to the apo form and to different bound receptor structures. Application of the precalculated soft modes as flexible variables both with and without inclusion of side chain flexibility resulted in improved ranking as well as ligand placement during docking. Especially, the application of conformational relaxation in normal modes alone gave already a significant improvement of the docking results compared to rigid docking at very modest increase in computer time. The method could therefore be applicable for virtual screening of large numbers of putative ligands of a given protein receptor molecule.

Materials and Methods

Receptor−Ligand Complexes. As test systems, several complex structures of CDK2 cocrystallized with different inhibitors were used (PDB entries: 1E1V,[27] 1E9H,[28] 1FVV,[29] 1H1S,[30] 1JSV,[31] 1KE5,[32] as well as the structure without ligand (apo structure, PDB entry: 1HCL[33]). The structures were superimposed using the STAMP structural alignment tool[34,35] as included in VMD.[36] Crystal waters and heteroatoms were discarded. To approximate unbound ligand conformations in solution, ligand structures were subjected to conjugate gradient minimization (in Cartesian coordinates) with the AMBER(parm99) force field[37] for 2000 steps using a generalized Born term and a cutoff of 10 Å. This resulted in ligand structures with near-optimal bond lengths, bond angles, and dihedral angles according to the AMBER force field.

Force Field and Parametrization. For the proteins standard Amber parm99 parameters[33] were used. For ligand parametrization, the Antechamber program[38] of the Amber package was employed, i.e., the Gaff Lennard-Jones parameters.[39] Partial charges were calculated using Gaussian03[40] and a restricted Hartree−Fock run using the 6−31G* basis set and RESP fitting.[41] Systematic docking searches were performed using a C++ version of the program PCRelax[21] at all-atom resolution. To calculate ligand−receptor interactions, a Lennard-Jones energy function and an electrostatic interaction term with a distance dependent dielectric constant was used:

$$V_{\text{rigid}} = \sum \left(\frac{R_{\text{AB}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\text{AB}}}{r_{ij}}\right)^{6} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \qquad (1)$$

It was found that during the docking searches an $\varepsilon(r) = 4r$ gave the best results in terms of positional deviation of the ligand from experiment, whereas for the evaluation (scoring) of the docking placements an $\varepsilon(r) = 20r$ gave results in better agreement with experiment. Dihedral torsion energies were calculated according to the parm99 force field, although torsion barriers were scaled down (factor 0.25).

The objective function governing the side chain selection procedure included the following terms: interaction energy between receptor and ligand, intramolecular energy of the ligand consisting of the nonbonded contributions and the scaled dihedral angle energy from the parm99 force field, and the intramolecular energy of the receptor consisting of the interaction of each flexible interface side chain with the rest of the receptor. A 10 Å cutoff was employed for nonbonded interactions.

To circumvent the singularity of the nonbonded interaction at zero distance, a soft core (separation-shifted) potential was employed during the sampling phase.[42]

$$V_{\text{soft}} = \sum \left(\frac{R_{\text{AB}}^2 + \delta}{r_{ij}^2 + \delta}\right)^{6} - \left(\frac{R_{\text{AB}}^2 + \delta}{r_{ij}^2 + \delta}\right)^{3} + \frac{q_i q_j}{\varepsilon^*(r_{ij}^2 + \delta)} \qquad (2)$$

A very small $\delta = 0.001$ Å^2 was used that has a negligible effect on the shape of the ligand binding site cavity. However, even such a very small δ resulted in much better convergence of the energy minimization and a larger number of favorable docking solutions. For scoring, the standard Amber force field function was used.

Modeling of Backbone Conformational Changes Using Elastic Network Normal Modes. Soft collective backbone degrees of freedom used during docking corresponded to eigenvectors of the protein calculated using an approximate normal-mode analysis related to Gaussian network models as described by Hinsen.[22] The normal modes were calculated with respect to the protein backbone (C_α atoms) based on a pairwise distance dependent energy function:

$$V(R_1, ..., R_N) = \sum_{C\alpha-\text{pairs}} V_{ij}(R_i - R_j) \quad (3)$$

with the pairwise term:

$$V_{ij} = k\left(R_{ij}^{(0)}\right)\left(|r| - R_{ij}^{(0)}\right)^2 \quad (4)$$

where $Rij^{(0)}$ is the pair's equilibrium distance. The force constant k is distance-dependent:

$$k(r) = \lambda \cdot \exp\left(-\frac{|r|^2}{r_0^2}\right) \quad (5)$$

such that small distances are strongly restrained and long distances are weakly controlled. The constant λ controls the overall flexibility. Harmonic modes with respect to the above energy function can be obtained by diagonalization of the second derivative matrix of the energy function as described by Hinsen.[22] In test calculations, it was found that setting $r_0 = 4$ Å gave the best overlap between the softest nontrivial modes and the conformational difference between apo and holo structures.[25]

Deformations in normal modes were employed exclusively for modeling global low-frequency backbone movements. Sampling in each normal mode was allowed for both possible deformation directions. Side chains followed the motion of the backbone (means the whole side chain moves as a rigid body attached to the backbone). This separation allows in principle an independent efficient treatment of side chain mobility (e.g., in terms of discrete rotameric states)[43] on top of the global backbone motion. During docking, including global flexibility, the protein structure was minimized in 10 softest modes obtained from the approximate normal mode calculations (in addition to the 6 translational + orientational degrees of freedom of the ligand protein). Following earlier experience, the deformability of the protein in the soft modes[20,21,25] was limited by a fourth-order function in the deformation along each mode. The total energy of the system is

$$V_{\text{flex}} = V_{\text{rigid}} + \lambda \sum_{m=1}^{M} \kappa_m (R^0 - R_m)^4 \quad (6)$$

The parameter λ is a scaling factor that was set to 6.0 kcal/mol based on previous experience on protein−protein docking.[25] Note, that the parameter λ was the same for all calculations and no adjustments for the individual docking cases were introduced. R^0 is the coordinate set describing the experimental structure of the protein, $\|R^0 - R_m\|$ is the conformational change of the protein in mode m. The force constant κ_m is the square of the eigenvalue corresponding to mode m. M is the total number of modes employed.

Modeling of Side Chain Conformational Changes and Ligand Flexibility. Side chain conformational changes were modeled using a rotamer trial (RT)[43] protocol employing a set of rotamers from the backbone independent Dunbrack rotamer library.[44,45] In addition to the apo or native side chain conformation from the starting structure, the 3−6 most common rotamer states for amino acids from the library were used that corresponded to the ones that exhibit the highest residence probability. Table 1 gives an overview on the number of rotamer states per amino acid. Ligand flexibility was accounted for on the level of dihedral torsion angles around single bonds. All bond lengths and bond angles were fixed during docking minimization. Dihedral angles within ring systems were left unaltered during docking runs. Full ligand dihedral angle flexibility was switched on during all docking runs.

Table 1. Number of Rotamers per Amino Acid[a]

amino acid	number of rotamer states	amino acid	number of rotamer states
arginine	4	methionine	4
asparagine	3	phenylalanine	3
cysteine	3	proline	2
glutamine	3	serine	3
glutamate	4	threonine	3
histidine	3	tryptophan	3
isoleucine	3	tyrosine	3
leucine	6	valine	3
lysine	5		

[a] For each amino acid, the rotamers with highest probability were considered according to the Dunbrack rotamer library.[44]

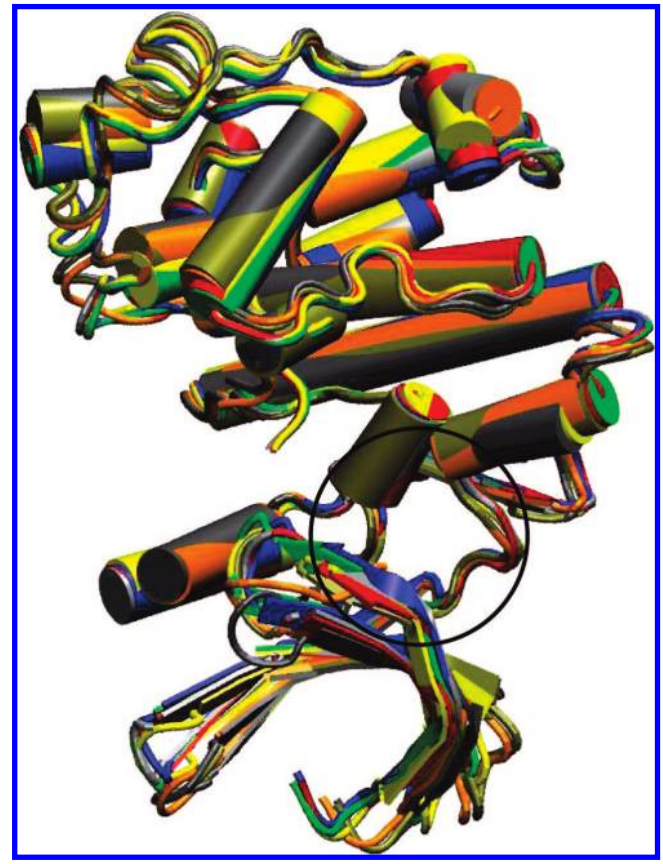

Figure 1. Superposition of all CDK2 kinase backbone structures used in this study to illustrate the conformational variability: 1E1V (blue), 1E9H (dark gray), 1FVV (orange), 1G5S (yellow), 1H1S (tan), 1JSV (light gray), 1KE5 (green), as well as the apo structure 1HCL (red). The backbone rmsd range of all bound structures with respect to the apo form was 0.7−2.2 Å. The ligand binding region is indicated as a circle.

Ligand−Receptor Docking Protocols. The individual docking runs involved several energy minimization runs. During each minimization, the local degrees of freedom (presumably with a shorter associated relaxation time) were optimized first, i.e., side chain conformations and ligand dihedral angles. Subsequently, the low-frequency degrees of freedom were energy minimized in a 6 + n-dimensional space (n is the number of modes + 6 rotational and translational degrees of freedom) as illustrated in Figure 1. For each docking run, 15 starting positions near the ligand binding region and 130 different orientations per start point (approximately 2000 starting structures) were generated. The starting positions were (approximately) evenly distributed at the binding site with a

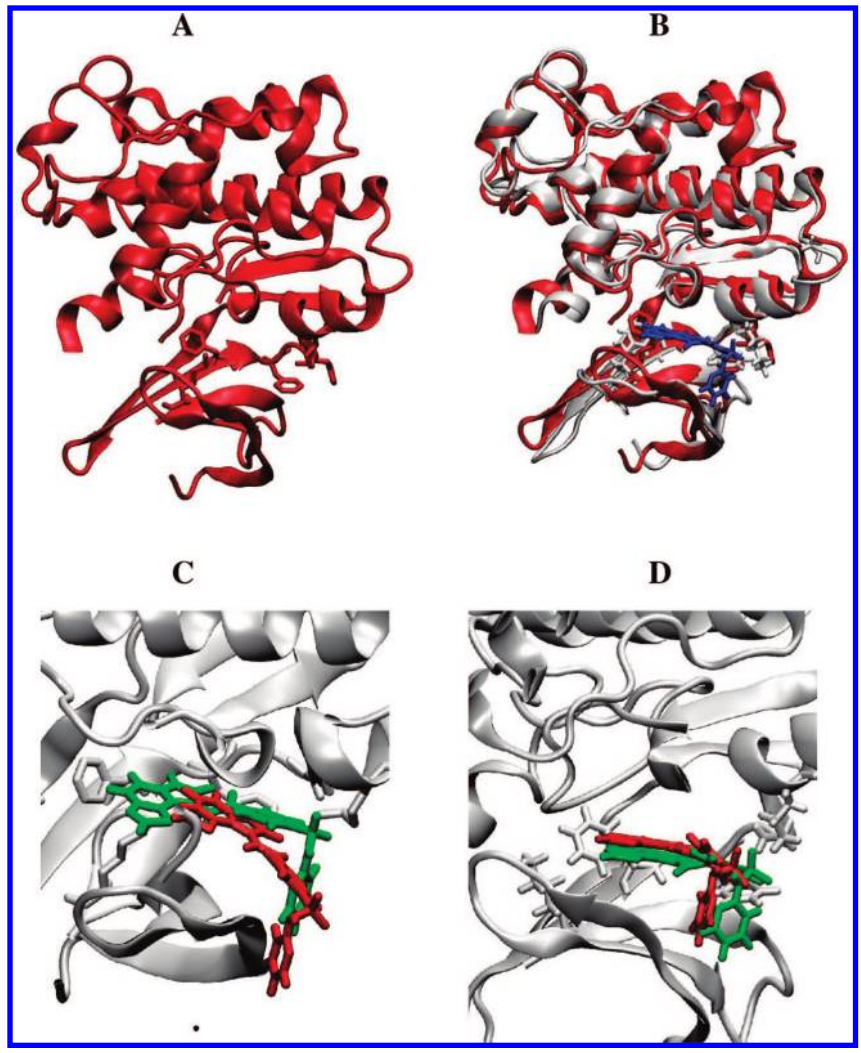

Figure 2. Illustration of induced fit in CDK2 upon ligand binding to the apo structure. (A) Cartoon representation of the apo structure (PDB 1HCL). (B) superposition of CDK2 apo structure (red) and an inhibitor bound structure (PDB 1E9H, gray cartoon) including the inhibitor (blue stick model). (C) Most accurate ligand placement (red) after docking of the inhibitor (from PDB 1E9H) to the rigid apo structure (gray cartoon, only the ligand was flexible). The ligand placement as observed in PDB 1E9H inhibitor bound form is shown in green. (D) Most accurate ligand placement (red) after docking the ligand into the flexible apo structure (gray, after flexible docking according to the RT/NM protocol: receptor backbone and side chains, as well as the ligand were flexible). Ligand placement as observed in PDB 1E9H is shown in green.

minimum distance of 2 Å with respect to any receptor atom and approximately the same distance between each start point (3 Å).

Four docking protocols differing in the degree of included receptor flexibility were tested (ligand flexibility was always included): (1) no receptor flexibility (rigid receptor protocol, run time for a single docking minimization: ~0.2 s); (2) inclusion side chain flexibility in the binding pocket (RT protocol, run time for a single minimization: ~11 s); (3) inclusion of global backbone flexibility only (NM protocol, run time for a single minimization: ~0.4 s); (4) inclusion of both side chain and backbone flexibility (RT/NM protocol, run time for a single minimization: ~8 s). All minimizations were performed employing the L-BFGS algorithm developed by Liu and Nocedal,[46] which is a quasi-Newton optimization routine. For the final scoring of the ligand receptor complexes, the complete ligand−receptor interaction energy was used (eq 1).

Results and Discussion

Flexible docking searches have been performed using the ligands from six different CDK2 complex structures. The receptor structure was either one of the six bound conformations or the apo (unbound) form of the protein. The apo as well as the bound CDK structure deviate considerably in the backbone and side chain structure near the ligand binding site (Figure 1). The overall deviation of each bound structure from other bound

Table 2. Ligand $rmsd_{heavy\ atoms}$ from Crystal Structure (and Ranking) for Docking Solution Closest to Experiment upon Docking to CDK2 apo Structure (PDB 1HCL)[a]

	1e1v	1e9h	1fvv	1h1s	1jsv	1ke5
Rigid receptor	0.8 (4)	2.2 (19)	4.3 (7)	2.2 (3)	2.0 (15)	4.2 (5)
RT	1.6 (1)	1.3 (1)	2.5 (2)	2.0 (1)	0.6 (1)	1.5 (2)
NM (10 modes)	1.5 (1)	1.4 (12)	0.9 (3)	2.0 (1)	1.5 (6)	1.9 (2)
RT+NM (10 modes)	1.2 (1)	0.8 (7)	1.6 (2)	1.8 (1)	1.7 (1)	1.4 (3)

0 – 1.0 Å | 1.0 – 2.0 Å | 2.0 – 3.0 Å | > 3.0 Å

[a] Ligand heavy atom rmsd for most accurate docking solution in terms of ligand rmsd (heavy atoms) and rank (in brackets) for the four different protocols as discussed in the Methods Section. Color coding according to legend.

Table 3. CDK2 Inhibitor Cross-Docking: Ligand $rmsd_{heavy\ atoms}$ from Crystal Structures (and Ranking) for Docking Solution Closest to Experiment[a]

A	1e1v	1e9h	1fvv	1h1s	1jsv	1ke5
1e1v	0.5 (1)	4.5 (6)	5.2 (2)	2.3 (4)	2.2 (14)	4.2 (8)
1e9h	1.8 (1)	0.8 (1)	3.7 (5)	2.1 (7)	2.3 (6)	1.8 (18)
1fvv	2.0 (2)	1.3 (1)	1.1 (1)	1.1 (1)	1.1 (8)	1.2 (1)
1h1s	1.8 (2)	1.2 (1)	1.0 (1)	1.0 (1)	0.9 (1)	1.2 (1)
1jsv	0.8 (1)	4.1 (12)	4.3 (18)	1.5 (1)	0.8 (2)	4.0 (8)
1ke5	1.5 (2)	1.4 (1)	1.6 (1)	2.1 (3)	0.4 (1)	0.5 (1)

C	1e1v	1e9h	1fvv	1h1s	1jsv	1ke5
1e1v	1.4 (1)	1.0 (8)	2.1 (2)	1.0 (9)	1.5 (4)	2.2 (9)
1e9h	1.8 (2)	0.7 (3)	1.4 (1)	1.4 (1)	1.4 (18)	1.6 (2)
1fvv	1.2 (2)	1.0 (2)	1.2 (1)	1.3 (1)	0.5 (5)	1.6 (1)
1h1s	0.9 (1)	1.4 (1)	0.8 (2)	1.3 (1)	1.0 (1)	1.2 (1)
1jsv	0.9 (3)	1.7 (7)	1.1 (15)	0.9 (2)	0.8 (1)	1.7 (2)
1ke5	1.8 (1)	1.0 (2)	1.3 (1)	2.0 (1)	0.9 (1)	0.8 (2)

B		1e9h	1fvv	1h1s	1jsv	1ke5
1e1v	1.3 (1)	1.8 (2)	2.8 (2)	2.2 (1)	0.5 (5)	1.2 (2)
1e9h	2.0 (1)	1.0 (2)	1.8 (1)	1.5 (1)	1.5 (2)	1.7 (2)
1fvv	1.9 (3)	1.2 (4)	0.6 (1)	1.1 (1)	1.2 (8)	1.2 (1)
1h1s	1.8 (2)	1.2 (1)	1.2 (1)	0.9 (2)	1.0 (1)	1.2 (2)
1jsv	0.9 (1)	1.5 (1)	2.1 (4)	2.0 (1)	0.5 (1)	0.9 (2)
1ke5	1.4 (1)	1.4 (1)	1.3 (1)	1.4 (1)	0.4 (1)	0.6 (1)

D	1e1v	1e9h	1fvv	1h1s	1jsv	1ke5
1e1v	1.1 (1)	1.0 (5)	1.8 (9)	1.4 (7)	1.5 (4)	1.4 (6)
1e9h	1.8 (1)	1.4 (2)	0.9 (7)	1.6 (2)	1.5 (2)	1.1 (8)
1fvv	1.6 (5)	0.8 (5)	1.0 (1)	1.0 (3)	0.9 (2)	1.2 (2)
1h1s	1.0 (2)	1.4 (1)	1.3 (1)	1.1 (1)	0.9 (1)	1.5 (1)
1jsv	1.5 (1)	0.6 (6)	1.4 (2)	1.4 (1)	1.4 (1)	2.0 (3)
1ke5	1.7 (2)	1.2 (2)	1.5 (1)	1.7 (1)	1.0 (1)	1.7 (1)

0 – 1.0 Å | 1.0 – 2.0 Å | 2.0 – 3.0 Å | > 3.0 Å

[a] Ligand heavy-atom rmsd and rank (in brackets) for most accurate solution in terms of ligand rmsd of individual cross-docking runs for different protocols: (A) docking to rigid receptor; (B) docking including side chain rotamer search; (C) inclusion of normal mode relaxation during docking; (D) inclusion of side chain rotamer search and normal mode relaxation. In each panel, column 1 corresponds to receptor structures, whereas row 1 corresponds to ligand structures.

forms or the apo form was in the range of 0.6−2.2 Å (protein backbone, Figure 1). However, near the ligand binding site, the pairwise deviation could reach more than 3 Å. It has already been shown in a number of studies that soft modes obtained from elastic network models (ENMs) can show a significant overlap with experimentally observed conformational changes.[17,19,23] The current flexible protein−ligand docking approach includes soft collective degrees of freedom of protein structures as additional variables during docking minimization (illustrated in Figure 2). Analysis of the current protein kinase example indicated that about 50% of the conformational difference between apo and bound forms can be described by deformations in 5−10 softest modes calculated for the unbound structure. A similar overlap was found when comparing two structures bound to different ligands (data not shown). However, often the overlap of observed conformational changes and calculated normal modes is distributed among several soft modes (not concentrated in just the first or second softest modes). As a compromise between computational demand and accurate representation of possible conformational changes during the docking simulations, the number of soft modes was limited to the first 10 softest modes of the ENM analysis. Inclusion of more soft modes did not change the results significantly but increased the computer time for energy minimization. During docking minimization, all interactions between ligand and receptor and between flexible side chains and other parts of the protein were treated explicitly based on the Amber parm99 force field. However, the energetic changes associated with global deformations of the protein along the soft collective modes were included based on a force constant derived from the ENM analysis (see Methods Section for details). This treatment makes the relaxation in the soft modes computationally highly efficient (computational demand only 2 times larger compared to docking to a rigid receptor).

Docking to the apo Structure of CDK2. As a first test, the performance of different docking protocols was compared for the systematic docking of the six ligands to the apo receptor. The results indicate that the rigid receptor hypothesis (even with full dihedral flexibility on the ligand) leads to unsatisfactory results for most of the ligand cases both in terms of the placement and orientation of the ligand (given as rmsd: root-mean-square deviation from the experimental placement) and the ranking of this solution (Table 2). In only one case, the rigid receptor approximation resulted in a ligand placement close to experiment and a reasonable ranking (first entry in the Table

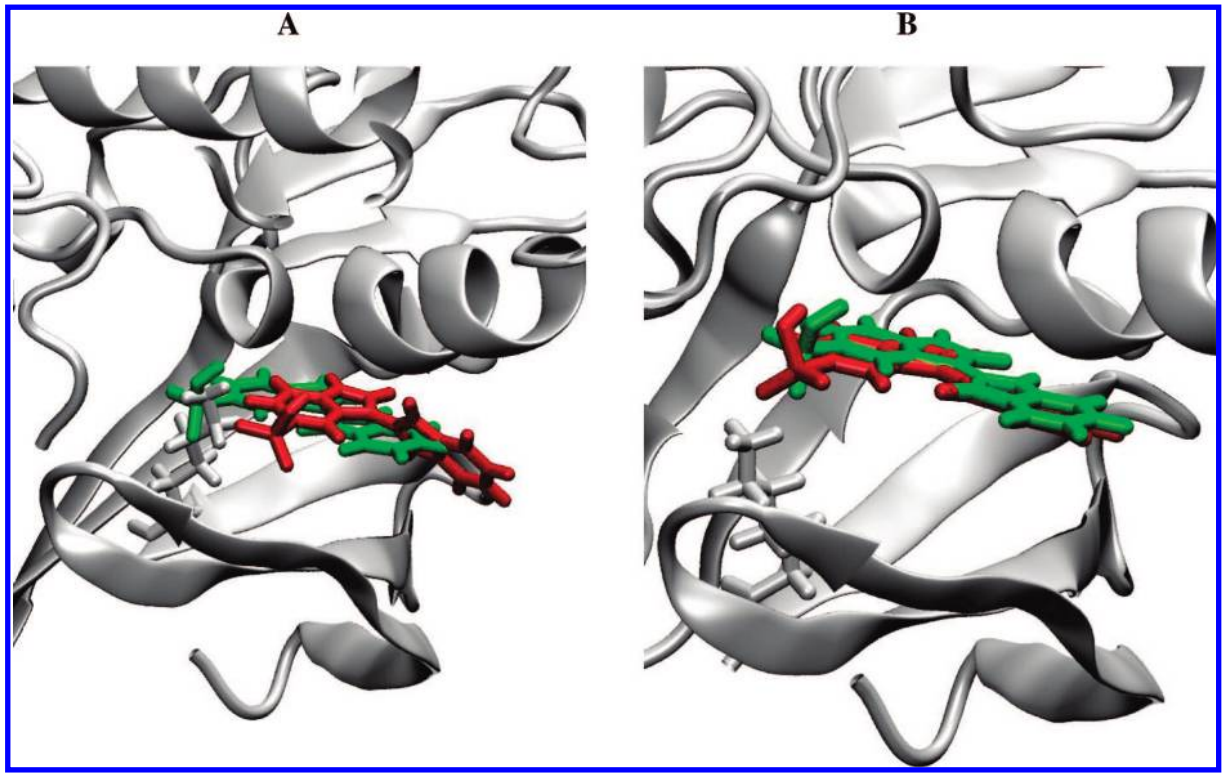

Figure 3. Cross-docking of ligand from PDB 1FVV to receptor from PDB 1E1V. (A) Most accurate ligand placement (red) after docking to the rigid receptor structure (gray: PDB 1E1V) (only the ligand was flexible), ligand placement as in PDB 1FVV shown in green. (B) Most accurate ligand placement (red) after docking to the flexible receptor structure (gray, PDB 1E1V, after flexible docking) (receptor backbone and side chains, as well as the ligand were flexible), ligand placement as in PDB 1FVV shown in green.

2). Allowing for side chain flexibility at the receptor binding site (RT protocol) leads to significant improvement of both the rmsd and the ranking of the docking solution closest to experiment (termed the near native docking solution). Interestingly, receptor relaxation in the 10 softest modes (no receptor side chain flexibility) also gave a very significant improvement of ranking and rmsd of the near-native docking solutions. The ranking was overall slightly better in case of including side chain flexibility at the rotamer level. One should keep in mind, however, that inclusion of normal mode relaxation only doubles the required computer time compared to docking to a rigid receptor. Inclusion of a search for the best rotameric side chain states during the optimization of ligand placement increased the computer time by a factor of ∼50. Applying the combined relaxation in soft modes and search for best side chain rotamers resulted on average in the best ligand rmsd, but the ranking of the best solution among alternative docking geometries was slightly better with the RT protocol. Interestingly, the combined RT/NM approach was compuationally faster than the RT approach due to a more rapid convergence of the energy minimization (the inclusion of flexibility in NMs may soften the energy landscape such that fewer energy minimization steps are sufficient for convergence).

Cross-Docking of Ligands to Different CDK2 Bound Structures. During virtual screening efforts to identify new putative ligands, a receptor structure bound to a ligand is usually preferred over the apo form of the receptor protein. The assumption is that structural differences between complexes with different ligands are smaller than the conformational difference with respect to the apo structure. However, especially in case of protein kinases, the conformational difference between bound structures complexed to different ligands can be quite considerable (Figure 1). Therefore, allowing for conformational changes during cross-docking of ligands to a bound receptor structure crystallized in the presence of another ligand might be as important as in the case of docking to the apo form. To further systematically compare the performance of the different docking protocols, we performed cross-docking searches (starting again from ∼2000 start geometries of the ligand near the known binding site).

In the case of docking of each flexible ligand to the kinase receptor in its corresponding rigid bound form, a placement with very small rmsd with respect to experiment was always found as the top-ranking solution (or second rank for one case, diagonal entries of Table 3A). This indicates that the force field scoring function is sufficiently accurate to reproduce the correct ligand placement and high ranking of the near-native solution as long as the receptor structure does not deviate from the bound structure. However, similar to docking to the apo structure employing a rigid receptor during cross-docking resulted in many cases in an rmsd of the ligand >3 Å from the experimental structure and also in unsatisfactory ranking of this solution (off-diagonal entries in Table 3A). Both the application of the RT protocol as well as the NM protocol greatly improved the results to on average similar degrees (Table 3B,C). The combined protocol resulted in the best ligand placements but a slightly worse overall ranking of the near-native solution compared to the RT protocol (Table 3D). A comparison of the cross-docking result using a rigid receptor and employing the RT/NM protocol is illustrated in Figure 3. Similar to the result on docking to the CDK2 apo structure, the application of the NM protocol (without considering side chain flexibility) improves the results considerably at a very modest increase of the computational demand (∼factor 2) compared to docking to a rigid receptor.

Conclusions

Many proteins, including some of the most prominent drug targets like protein kinases and HIV protease, undergo significant

conformational changes upon complex formation with substrates or inhibitors. Conformational changes include not only local side chain flips but also adjustment of loop structures and global changes of the protein backbone conformation. It has been recognized that computational approaches to identify putative drug-like ligand molecules for a given protein target structure need to appropriately account for such conformational changes.[17,19,23] As already mentioned, there is also a close relation between the accuracy of the scoring of a ligand receptor complex and the inclusion of conformational changes during docking. A soft scoring function that tolerates inaccurate placement of a ligand in a binding pocket or allows a large degree of overlap between ligand and receptor atoms can principally provide only limited specificity.

It was demonstrated that using normal modes to continuously model backbone flexibility together or without a rotamer library-based protocol for side chain flexibility can substantially improve docking results. Comparing average run times of the different protocols, the gain in accuracy, when going from a rigid receptor to a flexible receptor only in normal mode space is especially remarkable. In contrast to the frequently applied softening of the scoring function to allow for some overlap between ligand and receptor at the interface, the relaxation in soft normal modes can deform the interface in specific directions and can result in complete removal of atom overlap. This avoids an unspecific reduction of the repulsion due to atom overlap and can result in the formation of an altered cavity shape that exactly fits the ligand.

A possible extension of the approach could be the inclusion of side chains atoms and solvent molecules at the ligand binding site in the normal mode calculations. This would allow for small but rapid adjustments of side chains and water molecules upon ligand docking.

Homology modeled structures usually deviate from a realistic experimental structure to a similar degree as the different apo and bound structures used in the present study during cross docking. The successful application in cross docking may indicate that the approach could also be very useful for docking searches that employ homology modeled protein structures.

It should be emphasized that the approach was not yet optimized with respect to computational speed. For example, it could be possible to design smarter strategies for generating start structures that eliminate all configurations with too much overlap between ligand and receptor or to eliminate ligands from a library that are much larger than the cavity prior to docking. Also, it might be possible to speed up the docking process by combining it with docking approaches that are based on incremental construction of ligands based on a library of rigid fragments.[5]

Acknowledgment. This work was performed using the computational resources of the CLAMV (Computational Laboratories for Animation, Modeling and Visualization) at Jacobs University Bremen. We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support (grant Za-153/5).

References

(1) Muegge, I.; Rarey, M. Small Molecule Docking and Scoring. *Rev. Comput. Chem.* **2001**, *17*, 1–60.
(2) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935–949.
(3) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−ligand docking: current status and future challenges. *Proteins* **2006**, *65*, 15–26.
(4) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein−small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
(5) 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.
(6) Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.
(7) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375–382.
(8) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM, A New Method for Protein Modeling and Design: Applications to Docking and Structure Prediction from Distorted Native Conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
(9) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.
(10) Jones, G.; Willett, 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.
(11) Lorber, D. M.; Shoichet, B. K. Flexible Ligand Docking Using Conformational Ensembles. *Protein Sci.* **1998**, *7*, 938.
(12) Clausen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: efficient molecular docking considering protein structure validation. *J. Mol. Biol.* **2001**, *308*, 377–395.
(13) Carlson, H.,A.; Masukawa, K. M.; McCammon, J. A. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* **1999**, *103*, 10213–10219.
(14) 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.
(15) Wong, C. F.; McCammon, J. A. Protein flexibility and computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 31–45.
(16) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–25.
(17) Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* **1997**, *2*, 173–181.
(18) Tobi, D.; Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18908–18913.
(19) Tirion, M. M. Low-amplitude elastic motions in proteins from a single-parameter atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905.
(20) Zacharias, M.; Sklenar, H. Harmonic modes as variables to approximately account for receptor flexiblity in ligand−receptor docking simulations: applications to DNA minor groove ligand complex. *J. Comput. Chem.* **1999**, *20*, 287–300.
(21) Zacharias, M. Rapid protein−ligand docking using soft modes from molecular dynamics simulations to account for protein deformability: binding of FK506 to FKBP. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 759–767.
(22) Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins* **1998**, *33*, 417–429.
(23) Tama, F.; Sanejouand, Y. H. Conformational changes of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1–6.
(24) May, A.; Zacharias, M. Accounting for global protein deformability during protein−protein and protein−ligand docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225–231.
(25) May, A.; Zacharias, M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein−protein docking. *Proteins* **2007**, *70*, 794–809.
(26) May, A.; Zacharias, M. Protein−protein docking in CAPRI using ATTRACT to account for global and local flexibility. *Proteins* **2007**, *69*, 774–780.
(27) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* **2000**, *43*, 2797–804.
(28) Davies, T. G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J. A.; Noble, M. E. Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate. *Structure* **2001**, *9*, 389–397.
(29) Davis, S. T.; Benson, B. G.; Bramson, H. N.; Chapman, D. E.; Dickerson, S. H.; Dold, K. M.; Eberwein, D. J.; Edelstein, M.; Frye, S. V.; Gampe, R. T., Jr.; Griffin, R. J.; Harris, P. A.; Hassell, A. M.; Holmes, W. D.; Hunter, R. N.; Knick, V. B.; Lackey, K.; Lovejoy, B.; Luzzio, M. J.; Murray, D.; Parker, P.; Rocque, W. J.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Prevention of chemotherapy-induced alopecia in rats by CDK inhibitors. *Science* **2001**, *291*, 134–137.
(30) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.;

Noble, M. E.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745–749.

(31) Clare, P. M.; Poorman, R. A.; Kelley, L. C.; Watenpaugh, K. D.; Bannow, C. A.; Leach, K. L. The cyclin-dependent kinases cdk2 and cdk5 act by a random, anticooperative kinetic mechanism. *J. Biol. Chem.* **2001**, *276*, 48292–48299.

(32) Bramson, H. N.; Corona, J.; Davis, S. T.; Dickerson, S. H.; Edelstein, M.; Frye, S. V.; Gampe, R. T. , Jr.; Harris, P. A.; Hassell, A.; Holmes, W. D.; Hunter, R. N.; Lackey, K. E.; Lovejoy, B.; Luzzio, M. J.; Montana, V.; Rocque, W. J.; Rusnak, D.; Shewchuk, L.; Veal, J. M.; Walker, D. H.; Kuyper, L. F. Oxindole-based inhibitors of cyclin-dependent kinase 2 (CDK2): design, synthesis, enzymatic activities, and X-ray crystallographic analysis. *J. Med. Chem.* **2001**, *44*, 4339–4358.

(33) Schulze-Gahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S. H. Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins* **1995**, *22*, 378–391.

(34) Russell, R. B.; Barton, G. J. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **1992**, *14*, 309–323.

(35) Eargle, J.; Wright, D.; Luthey-Schulten, Z. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics* **2006**, *22*, 504–506.

(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38, 27−28.

(37) Case D, P. D.; Caldwell, J. W.; Cheatham , T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *Amber 8*; University of California: San Francisco, 2003.

(38) 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.

(39) 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.

(40) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(41) Bayly CI, C. P.; Cornell, W. D.; Kollman, P. A. The restrained electrostatic potential fit. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(42) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(43) Leach, A. R. Ligand docking to proteins with discrete side chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345–356.

(44) Dunbrack, R. L., Jr.; Cohen, F. E. Bayesian statistical analysis of protein side chain rotamer preferences. *Protein Sci.* **1997**, *6*, 1661–1681.

(45) Dunbrack, R. L., Jr.; Karplus, M. Backbone-dependent rotamer library for proteins. Application to side chain prediction. *J. Mol. Biol.* **1993**, *230*, 543–574.

(46) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large-scale optimization. *Math. Programming* **1989**, *45*, 503–528.

JM800071V