

# Modeling Correlated Main-chain Motions in Proteins for Flexible Molecular Recognition

Maria I. Zavodszky<sup>1</sup>, Ming Lei<sup>2#</sup>, M. F. Thorpe<sup>4</sup>, Anthony R. Day<sup>3§</sup>, and Leslie A. Kuhn<sup>1,3\*</sup>

<sup>1</sup>*Department of Biochemistry and Molecular Biology*, <sup>2</sup>*Department of Physics and Astronomy*, and <sup>3</sup>*Center for Biological Modeling, Michigan State University, East Lansing, Michigan*, and <sup>4</sup>*Department of Physics and Astronomy, Arizona State University, Tempe, Arizona*

Key words: conformational sampling, flexibility, induced fit, conformational ensemble, macrocyclic ligands, cyclophilin A, cyclosporin, estrogen receptor, zearalenol, FIRST, ROCK, SLIDE.

Grant sponsors: American Heart Association, Grant number: 9940091N; National Institutes of Health, Grant numbers: GM067249 and AI053877.

\*Correspondence to: Leslie A. Kuhn, Protein Structural Analysis and Design Lab, Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824-1319. E-mail: [KuhnL@msu.edu](mailto:KuhnL@msu.edu)

# Present address: Department of Biochemistry, Brandeis University, Waltham, Massachusetts

§ Present address: Department of Physics and Engineering Physics, John Carroll University, University Heights, Ohio

## ABSTRACT

We describe a new method for modeling protein and ligand main-chain flexibility and show its ability to model flexible molecular recognition. The goal is to sample the full conformational space, including large-scale motions that typically cannot be reached in molecular dynamics simulations due to the computational intensity, as well as conformations that have not been observed yet by crystallography or NMR. A secondary goal is to assess the degree of flexibility consistent with protein-ligand recognition. Flexibility analysis of the target protein is performed using the graph-theoretic algorithm FIRST, which also identifies coupled networks of covalent and non-covalent bonds within the protein. The available conformations of the flexible regions are then explored with ROCK by random-walk sampling of the rotatable bonds. ROCK explores correlated motions by only sampling dihedral angles that preserve the coupled bond networks in the protein and generates conformers with good stereochemistry without using a computationally expensive potential function. A representative set of the conformational ensemble generated this way can be used as targets for docking with SLIDE, which handles the flexibility of protein and ligand side chains. The realism of this protein main-chain conformational sampling is assessed by comparison with time-resolved NMR studies of cyclophilin A motions. ROCK is also effective for modeling the flexibility of large cyclic and polycyclic ligands, as demonstrated for cyclosporin and zearalenol. The use of this combined approach to perform docking with main-chain flexibility is illustrated for the cyclophilin A–cyclosporin complex and the estrogen receptor in complex with zearalenol, while addressing the question of how much flexibility is allowed without hindering molecular recognition.

## INTRODUCTION

Forty-five years ago, Linderstrom-Lang and Schellman<sup>1</sup> noted that biomolecules can exist in a number of different conformations. Since then, experimental evidence has increasingly indicated that the crystal structure of a protein, although likely the lowest energy state under the crystallization conditions, is by no means the only conformational state existing in solution<sup>2-4</sup>. Surprisingly, the total configurational entropy can increase upon protein association with other proteins, DNA, or small molecules<sup>5-9</sup>, indicating an increase in the flexibility of the system. Thus, a realistic description of proteins and protein-ligand recognition should account for the molecules' inherent flexibility. Studies have been performed that point out the importance of modeling protein side chain flexibility in docking<sup>10,11</sup>, and many current docking tools include protein side chain flexibility<sup>12-18,37,55</sup>. Further analysis of the conformational changes upon complex formation, by comparing ligand-free and ligand-bound crystal structures of various proteins, showed that about 25-50% of the proteins also undergo substantial main-chain conformational changes when binding their ligands<sup>11,19</sup>. This means, in many instances, that the protein-ligand recognition process cannot be described accurately unless protein main-chain flexibility is taken into account. Excellent reviews have been published recently<sup>20,21</sup> summarizing the state of the art in flexible docking. Except for inter-domain hinge motions<sup>22</sup>, crystallographically determined alternative conformations<sup>23,24</sup>, or small to moderate motions accessible by molecular dynamics simulations<sup>25,26</sup>, main-chain flexibility has not been considered in docking.

Inducing changes in the protein main chain while performing docking is too computationally intensive to be practical at present, so efforts in this area are directed toward generating an ensemble of protein conformations to be used as docking targets. This approach is based on the idea that the ligand will recognize one of the low energy protein conformers in solution<sup>27-29</sup>, which is likely more realistic than the classical induced fit model<sup>30</sup> picturing a single protein conformation that is actively modified by the ligand. Ligands of various shapes can bind to any low-energy conformation of the unbound protein, not only to the one with the lowest free energy. A ligand with good shape and chemical complementarity to a less populated conformational state of the receptor can be a stronger binder than one that fits less tightly into the lowest energy conformation of the protein. Nevertheless, this tight-binding ligand would be missed if only the lowest energy conformer of the receptor or the average of several low energy structures was considered.

Alternative protein conformers for docking usually come from NMR studies, x-ray structures of the protein in complex with various ligands, or MD simulations. In their groundbreaking work, Kuntz and co-workers<sup>15</sup> have used ensembles of NMR and x-ray protein structures as docking targets for DOCK. Intermolecular potentials are calculated at grid points in the binding site. Variations among different observable conformations are taken into account by calculating the average of the force field values at each grid point, and two types of averaging are used: energy weighted, and geometry weighted. The first method involves calculating the contribution of each atom from each structure to the potential energy, then calculating a weighted potential by averaging over all structures. Geometry weighted averaging means that the averaging is performed at the structural level by calculating a mean position for every atom of the protein over the ensemble of conformers. Although this approach does not include receptor flexibility in a dynamic sense, the composite grid representing the interaction energies of the docked ligands with the different protein conformers is shown to outperform many of the

grids derived from individual structures in identifying known inhibitors for the cases studied. In a similar approach, Osterberg and coworkers also show encouraging results using weighted average methods for docking 21 peptidomimetic inhibitors to HIV-1 protease with AutoDock<sup>31</sup>. Claussen and co-workers use FlexE, an extension of FlexX<sup>32</sup>, to dock ligands into a set of alternative local conformations generated from the superimposed structures of the x-ray structure ensemble of the target protein<sup>23</sup>. While averaging the similar backbone and side-chain positions, the regions with larger variations are retained in the form of conformational libraries. New conformations of the receptor are created by combining compatible conformations of the various flexible regions of the binding site. The method can handle several side-chain conformations and smaller loop (up to three or four amino acid) movements, but not the motions of larger backbone segments.

The use of multiple experimental structures limits the conformational sampling to already observed and existing conformations. Some proteins do not have multiple x-ray structures or are too large or flexible for NMR structure determination. MD simulations can provide novel protein conformers to be used as targets for docking, however, they generate smaller scale movements than may be observed in nature due to the computational complexity of sampling the motions of thousands of atoms whose motions are governed by a full force field for tens of millions of femtosecond time steps. The development of a dynamic pharmacophore model for HIV-1 integrase is described by Carlson et al. by using snapshots of MD simulations and the multi-copy minimization method MUSIC to determine probe positions that are relatively static in the dynamic binding site<sup>33</sup>.

In this paper, a new and relatively efficient approach for modeling main-chain flexibility in docking and screening is described. Flexibility analysis from a single conformation of the target protein can be performed using the graph-theoretic algorithm FIRST<sup>34</sup>, followed by the generation of alternative conformations for the predicted flexible regions using ROCK<sup>35,36</sup>, a random-walk conformational sampling algorithm. This algorithm is unique in its ability to handle coupled ring systems and guarantees conformers with good stereochemistry without requiring expensive energy minimizations. By directly identifying and maintaining collective motions throughout the simulation, ROCK avoids the pitfalls of combining potentially incompatible conformations of various parts of the protein. A representative and diverse set of the conformational ensemble generated by ROCK was used as a series of targets for docking with SLIDE<sup>24,37</sup>, which handles protein and ligand side chains flexibly. ROCK is well suited for sampling conformations of coupled ring structures, including those formed by non-covalent bond networks, and can also be used to model the flexibility of macrocyclic ligands, as demonstrated for cyclosporin and zearalenol. The use of this combined method to perform flexible docking is illustrated on cyclophilin A (CypA) – cyclosporin and estrogen receptor (ER) – zearalenol complexes, while addressing the question of how much flexibility is allowed while maintaining specific recognition between the molecular partners.

CypA is a ubiquitous cytosolic protein composed of 165 amino acids catalyzing *cis-trans* isomerization in peptides and proteins<sup>38</sup>. CypA is also the target for the immunosuppressive drug cyclosporin A, a cyclic undecapeptide, in which 7 of the 11 amide nitrogens are methylated (Figure 1). The CypA-cyclosporin complex binds and inhibits the Ser-Thr phosphatase calcineurin, as well as blocking the activation of JNK and p38 signaling pathways, inhibiting T lymphocyte activation<sup>39</sup>. The role of CypA in immunosuppression seems to be unrelated to its *cis-trans* isomerase function but depends on ligand binding in the active site<sup>40</sup>. Another interesting aspect of CypA is that it binds to the HIV-1 Gag protein and is incorporated into the HIV-1 virion as a necessary element for HIV infection<sup>41</sup>. The human CypA – cyclosporin

complex serves as a good model system for analyzing protein-ligand flexibility for the following reasons: (1) as a peptidyl *cis-trans* isomerase, CypA has to be flexible enough to accommodate both the *cis* and the *trans* conformations of its substrate; (2) since the ligand, cyclosporin, is a cyclic peptide with many bond-rotational degrees of freedom, the same method used to model main-chain flexibility of the target protein can be applied to model the flexibility of this ligand; and (3) the flexibility of CypA and cyclosporin has been studied extensively, with a number of high resolution x-ray and NMR structures available for both the unliganded CypA and complexes of CypA with various peptidyl ligands, as well as dynamics data during catalysis.

Zearalenone is a common contaminant of the major cereal grains worldwide and is produced by plant pathogenic molds.  $\alpha$ - and  $\beta$ -zearalenol are metabolites usually found in mixture with zearalenone (Figure 2). The biological activity of these mycotoxins is mainly attributed to their estrogenic activity, which modulates or disrupts endocrine function in animals and possibly humans<sup>42,43</sup>.  $\alpha$ -Zearalenol (henceforth referred to as zearalenol) was shown to have the most prominent biological effect among the three compounds. It has similar binding affinity to the natural ligand, 17 $\beta$ -estradiol<sup>44</sup>, for the human estrogen receptor, despite having quite different structure and chemistry. Zearalenol is a highly flexible molecule, so it cannot be assumed it binds in the same conformation as it assumes in the crystal structure (CSD entry BEGDAS). Using the ability of ROCK to model large ring flexibility, the conformational space of zearalenol was explored. These conformers were docked into the human ER using SLIDE, to model the flexible interaction between zearalenol and ER and assess the molecular basis for the estrogenic activity of zearalenol.

## METHODS

### FIRST

FIRST (Floppy Inclusion and Rigid Substructure Topography) is a graph theoretical approach to identify rigid and flexible regions based on the protein bond network consisting of covalent bonds, hydrogen bonds and hydrophobic interactions<sup>34</sup>. Hydrophobic tethers represent hydrophobic interactions between carbon and/or sulfur atoms for which:

$$D = R_A + R_B + R$$

where  $D$  is the distance between the centers of the two hydrophobic atoms,  $R_A$  and  $R_B$  are their van der Waals radii, and  $R$ , the distance between their van der Waals surfaces, was empirically optimized to be 0.5 Å. This modeling of hydrophobic interactions by FIRST is a modified version of that described previously<sup>45</sup>, in that only those carbon or sulfur atoms covalently bonded only to other carbon, sulfur, and/or hydrogen atoms were considered to participate in hydrophobic interactions, whereas previously any carbon or sulfur atoms with van der Waals surfaces within 0.25 Å of each other were considered engaged in hydrophobic interactions. This change was implemented to only include hydrophobic interactions between significantly hydrophobic groups in the protein.

In much the same way that constraint counting approaches in engineering can identify underconstrained, just-stable, and overconstrained regions in the trusswork of a bridge or building, FIRST places and counts constraints in a protein structure according to its network of covalent and non-covalent bonds. The result is the identification of which bonds are constrained and which bonds remain free to rotate. Rigid regions may be just rigid (isostatic) or overconstrained (more than one bond must break to allow flexibility). Flexible regions are those with remaining degrees of freedom or not enough constraints to be rigid. The number of extra constraints or the number of remaining degrees of freedom is used to calculate the relative rigidity or flexibility index of the region. This computational approach is very fast and is able to

reliably predict the conformational flexibility of a protein from a single, static three-dimensional structure<sup>34,46</sup>.

## ROCK

ROCK (Rigidity Optimized Conformational Kinetics) uses restricted random walk sampling to search the conformational space available to proteins<sup>35,36,47</sup>. The conformational space is divided into allowed and disallowed regions, and sampling is only performed within the allowed regions. As a consequence, no energy function or Metropolis sampling is required. A conformation is either accepted or rejected, depending upon whether it meets a set of pre-defined bond length, angle and excluded volume constraints. A key aspect of ROCK is that it allows the sampling of motions that are consistent not only with covalent bonds, but also with maintaining a given set of non-covalent interactions. Proteins are viewed as complex networks of atoms connected by covalent bonds, hydrogen bonds and hydrophobic interactions (Figure 4E). We use FIRST<sup>33</sup> to define the network of covalent and non-covalent constraints to be preserved, including bond lengths and coordination angles, hydrophobic interactions, hydrogen bonds, and salt bridges. FIRST is also used to identify rotatable bonds to be sampled by ROCK as well as non-rotatable bonds that are completely constrained by the network. In addition, good main-chain stereochemistry for proteins is ensured by sampling only favored main-chain  $\phi, \psi$  angles from the distribution<sup>48</sup> used in the program PROCHECK<sup>49,50</sup>. A tolerance value can be set to allow small van der Waals overlaps during conformer generation, e.g., up to 30% of the sum of the atoms' van der Waals radii (which are expanded to reflect the presence of hydrogen atoms that are not included in the crystal structure's atomic coordinates from the Protein Data Bank<sup>51</sup> (PDB)). Allowing small van der Waals overlaps during conformer generation permits more efficient sampling of conformational space. After sampling, any remaining overlaps in the conformers can be annealed, for instance, by quick energy minimization or by a set of ROCK steps in which the tolerance for overlaps is reduced to 0.

New ROCK conformations are generated by successive, random dihedral rotations (typically a maximum of 5 or 10 degrees), ensuring that all the original bond constraints are obeyed, significant van der Waals overlaps between atoms are avoided, and all the rings formed by covalent and non-covalent constraints are closed. Rotating one bond in an interlocking ring system typically will result in opening several other rings. To obtain a valid new conformer, all the rings must be closed exactly and simultaneously after each angular step. The algorithm of Go and Scheraga<sup>52</sup> can solve the ring closure equations for a single ring, but this is insufficient to concurrently close all the interlocked rings in a complex bond network. Rather than identifying a set of dihedral angles that will close a single ring (i.e., solving for the roots of the ring closure equation  $F_{angle1, angle2, \dots} = 0$ ), ROCK makes random dihedral angle changes in the flexible rings of the network. A set of values for the remaining angles is then found that allows the rings to close. This is done by minimizing the ring closure potentials with respect to the angular variables for all rings simultaneously, until a set of angles is found that reduces the potentials to zero<sup>35,36</sup>. Some rotatable bonds are shared by interlocked rings, and these bond angles will appear in more than one ring closure function. However, the simultaneous minimization ensures that the resulting set of angles will close all the rings.

The computational cost of minimizing the total ring closure potential for a cluster of interlocking rings rises with the number of bonds in the network. ROCK maintains efficiency by sampling only those regions in the protein that are flexible, using the results of FIRST flexibility analysis. FIRST analysis of a number of protein structures has shown that a high density of hydrogen bonds and hydrophobic interactions reduces the number of internal degrees of freedom

to zero in some regions of a protein, resulting in local rigidity<sup>45,47,53,54</sup>. In fact, proteins typically have a large, virtually rigid core and several smaller rigid or flexible regions. Flexible regions of the protein that are remote in sequence can be correlated due to coupling through the intervening network of covalent and non-covalent bonds (Figure 4F). These coupled regions are identified by FIRST and sampled in ROCK as a ring cluster.

A non-linear constrained optimization algorithm is used for repositioning those side chains not participating in non-covalent interactions into positions consistent with the new main chain and non-covalently networked side chains<sup>36</sup>. The most distinct conformers are selected by a utility program of the ROCK software package as follows. The pairwise main-chain RMSD (root mean square atomic positional deviation) values for the flexible regions of the ROCK conformers are calculated relative to the initial structure (e.g., the crystal structure of CypA). The initial structure is the first conformer selected. The next ROCK conformer selected is the one with the largest RMSD value relative to the initial structure. For the remaining  $n-1$  conformers, we compare their RMSD values relative to the first two structures and select the smaller of the two values to characterize how similar each of the  $n-1$  structures is from the first two. From the set of  $n-1$  RMSD values obtained this way, the largest one is selected, and the corresponding, least similar, conformer is added to the most distinct set. This procedure is repeated until all conformers have been analyzed.

## SLIDE

The docking and screening software SLIDE<sup>24,37</sup> (Screening for Ligands by Induced-fit Docking, Efficiently) is used to perform flexible docking experiments using the protein and ligand conformations generated by ROCK. SLIDE models protein-ligand interactions based on steric complementarity combined with hydrophobic and hydrogen-bonding interactions, and can include water-mediated hydrogen bonds. Due to its ability to balance protein and ligand side-chain flexibility, SLIDE can identify and correctly dock diverse, known ligands into the ligand-free conformation of the binding site, as has been shown for a variety of proteins, e.g., subtilisin, cyclodextrin glycosyltransferase, uracil DNA glycosylase, rhizopuspepsin, HIV protease, estrogen receptor, Asn tRNA synthetase, thrombin, and glutathione S-transferase<sup>24,37,55,56</sup>. Scoring of the docked protein-ligand complex by SLIDE is based on the number of intermolecular hydrogen bonds and the hydrophobic complementarity between the ligand and its protein environment. The knowledge-based scoring function DRUGSCORE<sup>57</sup> is used in addition to the built-in scoring function of SLIDE to score and rank the dockings.

## Preprocessing of Structures

The CypA structure used as input to FIRST was taken from the 1.8 Å resolution CypA-cyclosporin complex<sup>58</sup>, Protein Data Bank<sup>51</sup> (PDB) entry 1bck. The protein was first analyzed with the ligand Thr2-cyclosporin removed, then the analysis was repeated with the ligand in the binding site. To allow direct comparison of the FIRST flexibility prediction with protein dynamics data provided by NMR measurements<sup>59</sup>, the structure of CypA in complex with the peptide<sup>58</sup>, Suc-Ala-Phe-Pro-Phe-4-NA (PDB entry 1rmh), was also analyzed.

There is no ligand-free crystal structure of the estrogen receptor available, so as a ligand-free target, the structure of the complex with 17β-estradiol<sup>60</sup> was used with estradiol removed (PDB entry 1ere). The 3D structure of the starting zearalenol conformation was taken from the Cambridge Structural Database<sup>61</sup> (CSD entry BEGDAS).

To prepare the crystallographic structures of the proteins for analysis by FIRST, WHAT IF<sup>62</sup> was used to add polar hydrogen atom positions. This program has been shown to reliably reproduce ~95% of the hydrogen atom positions observed in protein structures determined by neutron diffraction<sup>17</sup>. PROCHECK<sup>49,50</sup> was then used to assess whether the structures had good stereochemistry, since the flexibility analysis of the non-covalent bond network by FIRST is sensitive to defects in packing and main-chain dihedral angles. Since WHAT IF adds hydrogen positions only to proteins, the ligand hydrogen positions were added using the program InsightII (Accelrys). Surface-bound water molecules were not included in either the CypA or estrogen receptor ligand dockings.

## Energy Calculations

The potential energies of the CypA conformers were calculated using the CHARMM22 force field<sup>63</sup> with a Generalized Born solvent model using the MMTSB tool set<sup>64</sup>. Because small van der Waals overlaps are tolerated by ROCK to increase the probability of traversing small energy barriers during conformer generation, a short energy minimization of 100 steps was performed before the energy calculations for each protein conformer to resolve any remaining overlaps and result in meaningful conformational energies. A distance dependent dielectric function with a dielectric constant of 4.0 was used during energy minimization with CHARMM22. Backbone Ca atoms were restrained to their original positions with a force constant of 5 kcal/mol to maintain the main-chain conformation generated by ROCK.

## RESULTS

### CypA – Cyclosporin

#### *Flexibility Analysis of CypA*

FIRST identified 121 hydrophobic tethers and 213 hydrogen bonds satisfying the bond length and angle criteria<sup>34</sup> in the ligand-free structure of CypA (PDB entry 1bck with the ligand removed). Hydrogen bond energies were calculated using a modified version of the Mayo potential<sup>53,65</sup> and were ranked in decreasing order, with a maximum (least favorable) H-bond energy of  $E = -0.1$  kcal/mol.

One way to visualize the results of the flexibility analysis is by generating a hydrogen-bond dilution plot (Figure 3). The numbering along the top, from left to right, represents the amino acid sequence of the protein (residues 1-165). Each line below is colored to indicate the rigid and flexible parts present in the structure at the energy value listed on the left. Thin black lines represent flexible regions, while thick bars represent rigid regions, with identical colors for mutually rigid regions that belong to the same rigid cluster (which may or may not be contiguous in sequence). The consecutive lines illustrate the changes in flexibility of the protein as hydrogen bonds are removed in order of their increasing energy, a process analogous to thermal denaturation<sup>53</sup>. These energy values should be considered a reasonable ranking of relative energies, rather than measuring energies on an absolute scale. When the crystal structure (PDB entry 1bck) was analyzed (energy of  $-0.1$  kcal/mol, top of the plot in Figure 3), all 213 hydrogen bonds were present and the structure formed one rigid cluster (represented by the red bar), with only a few residues at the N terminus being flexible (represented by the thin black line segment). Moving down the plot, as the energy increased and hydrogen bonds incrementally broke, certain regions became flexible, as represented by black lines intervening between colored blocks. Hydrophobic interactions were kept intact during the thermal dilution, because hydrophobic

interactions apparently become stronger over moderate increases in temperature<sup>66</sup>. Eventually, proteins typically fragment into two or more independent rigid regions (which are internally stable but could move as rigid bodies relative to one another) represented by segments of different colors connected by flexible regions. The first and second columns in Figure 3 list the index number (rank, from strongest to weakest, where the strongest H-bond is given index 1, etc.) and energy values (in kcal/mol) of the hydrogen bonds whose breakage induced a rigid to flexible change in the structure, with the resulting structure shown on that line. The last two columns specify the residue numbers of the hydrogen donor (blue) and acceptor (red) of the respective hydrogen bond (denoted S for side chain, M for main chain, and W for buried water molecules). The donor and acceptor positions are also shown by carets beneath the bars in the plot. The third column lists the mean coordination number at the current H-bond energy. The mean coordination number  $\langle r \rangle$  is the average number of covalent and non-covalent bonds for the atoms in the protein, providing an overall description of the protein bond network that is useful when comparing rigid to flexible transitions in different proteins<sup>45</sup>. Moreover, at  $\langle r \rangle = 2.40$ , a rigid-to-flexible phase transition occurs in 26 proteins previously analyzed, and this transition point is shared with other molecular networks such as amorphous glasses<sup>45</sup>.

An energy cutoff value of -2.3 kcal/mol was selected from the hydrogen bond dilution plot as corresponding to the flexibility observed in the native state of the CypA protein, where the protein has one rigid core, but the outer loops are flexible (Figure 4A). These regions are found to be flexible in the well-determined NMR structure of ligand-free CypA<sup>67</sup>. When the bond network at this particular energy cutoff was analyzed by FIRST, the following regions were identified to be flexible: residues 12-15, 24-29, 43-47, 54-60, 65-76, 79-82, 87-94, 101-107, 116-127, 133-135, and 143-155. Three strands of the  $\beta$ -sheet forming the bottom of the binding site (Figure 4A) were rigid, while the loops surrounding the ligand binding site were flexible. To study the effect of ligand binding on the flexibility of CypA, FIRST analysis was also performed on the CypA-cyclosporin complex (Figure 4B). Cyclosporin rigidified part of the CypA binding site, especially the strand formed by residues 87-94 and loops formed by residues 24-29, 87-94, and 116-127. Except for residues 116-127, these regions are not in direct contact with the ligand but are coupled via the bond network to the binding site residues. This propagation of flexibility changes from the binding interface to distal regions of the protein also has been observed for other systems<sup>9, 68, 69</sup>.

### ***Main Chain Conformational Exploration***

The results of FIRST analysis for CypA included lists of the covalent bonds, hydrogen bonds stronger than -2.3 kcal/mol (using the hydrogen-bond energy cutoff found to be appropriate for CypA), and hydrophobic tethers, as well as a list of which bonds remained rotatable, the flexibility index evaluated for each bond, and lists of which bonds and atoms were coupled through rings of covalent and non-covalent interactions. This information and the crystal structure 1bck were used as the input for ROCK to generate alternative conformations for the flexible regions. The bond network in CypA identified by FIRST is shown in Figure 4E, with covalent bonds rendered as blue tubes, hydrogen bonds as magenta tubes, and hydrophobic interactions as triplets of white spheres. The flexible ring clusters identified by ROCK in this network are shown in the next panel (Figure 4F), with blue representing the rigid framework of CypA and each other color representing a region of collective motion<sup>46</sup>. While the purple, pink and orange regions are simple loops consisting of a single ring of covalent and non-covalent bonds, the regions colored in red, yellow, and green are complex ring clusters made up of 7 to 14

interlocked rings with multiple, coupled bond rotational degrees of freedom allowing conformational change.

Two ROCK runs were performed: one with main-chain dihedral angle rotation steps of up to 5 degrees, and the second with steps of up to 10 degrees. ROCK actively rotates a fixed percentage (10 or 20%) of the randomly chosen dihedrals. Other, “passively rotated” dihedrals may be changed by greater or lesser angles in order to close the ring systems generated by the covalent and non-covalent bond network surrounding the actively rotated bonds. Given the random walk nature of the conformational sampling with ROCK, independent runs, especially with different angle step sizes, have the potential of sampling different regions of the conformational space. Depending on the step size and the percentage of actively rotated dihedrals, it took an average of 3 to 6 minutes of CPU time to generate a new CypA conformation (on an AMD 1.9 GHz processor). The conformer generation speed is mainly determined by the size of the flexible regions of the protein, in particular by the number and complexity of the interlocking ring systems. Each ROCK run generated 600 conformers, out of which the most distinct 20 conformations, representing the largest conformational changes, were selected. The 20 most distinct conformations from each of the two runs were combined, and the 12 most distinct conformers (Figure 4C) of these 40 structures were identified and used as targets for docking. Accepting only conformers with favored main chain  $\phi$  and  $\psi$  values guaranteed good stereochemistry of these structures, as indicated by the Ramachandran plots of the x-ray structure and the most distinct conformer (Figure 5); the other conformers were of similar quality.

The set of 12 most distinct conformers was identified based on the backbone RMSD values of residues 42-46, 67-75, 79-81, 120-124, and 148-149 relative to the crystal structure of CypA (PDB entry 1bck). The movements of these regions were monitored because these are the regions in the binding site predicted to be the most flexible by FIRST. They also are the regions with the most significant backbone differences in NMR structures<sup>67</sup>. The conformers generated by ROCK (Figure 4C) sample approximately the same conformational space as the 20 lowest energy NMR structures of CypA (Figure 4D). The regions with the largest movements modeled by ROCK are those with the most variations among the individual NMR structures, with the ROCK conformers showing somewhat larger backbone deviations of up to 4.3 Å. To illustrate the range of motions captured by ROCK, the average of the pairwise main chain dihedral angle differences of the 12 most distinct conformers was plotted for each residue (Figure 6 panels A and B). As a comparison, the deuterium exchange rates of the backbone amide protons of free CypA are shown in Figure 6C (data generously provided by Marcel Ottiger and Kurt Wüthrich<sup>67</sup>). Significant  $\phi$  and  $\psi$  angle changes were found in regions surrounding the binding site (residues 65-76, 116-126, 143-155), but also in those residues not in direct contact with the ligand (residues 43-47, 79-82, 87-94, 134-136). These are also the backbone areas with very fast HD exchange rates observed with NMR (Figure 6C), indicative of the flexible regions of CypA in solution.

Conformers for the cyclic ligand cyclosporin were also generated with ROCK. Figure 7A shows the *cis* peptide (free in organic solvent) and *trans* peptide (protein-bound) conformations of cyclosporin. The protein-bound conformation of cyclosporin was used as a starting structure, since peptide bonds are locked in the present implementation of ROCK, and only the covalent bond lengths and coordination angles were used as constraints. This was considered the most appropriate approach, given that only one inter-residue hydrogen bond and no intramolecular hydrophobic tethers were identified for cyclosporin by FIRST. A total of 3000 cyclosporin conformers were generated in three separate runs using small angular steps at a speed of 0.5 to

0.8 seconds CPU time per conformer (on an AMD 1.9 GHz processor). The runs differed in the maximal step sizes of the dihedral angle rotations (up to 2.0 and 5.0 degrees, respectively) and the maximum percentage of bonds that could be rotated at each step (10% and 20%, respectively). As in the case of protein main-chain conformer generation, the active rotation of these randomly chosen bonds could result in greater rotation of other, passively rotated bonds. Passive rotation of bonds is typically required to close the rings of covalent and non-covalent interactions involving actively rotated bonds. Since cyclosporin is a relatively large and flexible ligand, and docking thousands of conformers to 13 CypA targets (12 conformers plus the original x-ray structure) would be very time consuming, the most distinct 395 conformers out of the 3000 cyclosporin structures were selected for docking. Following main chain superposition with the x-ray structure of the protein-bound cyclosporin (PDB code 1bck), the root-mean-square positional deviations (RMSD's) of these conformers were in the range of 0.5 – 4.9 Å, with Ca positional deviations of up to 4.5 Å. This indicates a diverse set of conformers and reasonable sampling of conformational space (Figure 7B).

### **Docking**

CypA and cyclosporin conformers were used for docking with SLIDE to probe the range of flexibility consistent with molecular recognition. Templates representing the binding site in CypA were created for the x-ray structure of CypA (PDB entry 1bck) and for each of the 12 ROCK-generated conformers. These templates included 77 to 108 points each. All the hydrogen bonding template points were assigned as key points to assure that only those dockings with at least one hydrogen bond between protein and the ligand were generated. Using a template of 77 points, SLIDE took an average of 6 minutes (elapsed time) to dock one conformer of cyclosporin, incorporating side-chain flexibility modeling for the protein and ligand (on a PentiumIII 800MHz processor). The docking time depends on several factors: size of the template describing the binding site, size and number of rotatable bonds of the ligands, parameter settings referring to how closely the template points are to be matched by the ligand interaction points, and how many iterations are allowed during the process of resolving van der Waals collisions, among others. The CypA-cyclosporin system was very time consuming due to the large size of the cyclosporin molecule (resulting in a very large number of potential matches with the CypA template, all of which were tested). Docked ligands maintaining interactions with the rigid base of the binding site (Table I) were considered to be correct dockings. A docking was classified as correct, or close to correct, if the contacts listed in Table I were maintained with a maximum distance of 4 Å (5 Å for the hydrophobic contact). This approach of using known recognition determinants<sup>58</sup> to identify good dockings was employed since the goal here is to assess how much flexibility in the protein and ligand is consistent with maintaining recognition.

Ca deviations of up to 4.3 Å were found in the CypA binding site. A wide range of cyclosporin ligand conformations could be accommodated, with backbone RMSD values of up to 4.3 Å compared to the x-ray conformation. Upon docking, key interactions with the more rigid portions of the binding site were maintained by residues 1, 2 and 11 of cyclosporin, while the effector loop of cyclosporin protruding from the binding site could flex considerably, reaching up to 8 Å in Ca deviations (Figure 8A and B), consistent with the NMR ensemble (Figure 8C). These results are also in good agreement with the x-ray results, showing high temperature factors for the effector loop of cyclosporin and loops 79-81, 65-76, and 143-155 of CypA (Figure 8D), indicating movement and/or disorder of these regions in the crystal structure of the CypA-cyclosporin complex (PDB entry 1bck).

The distribution of SLIDE and DRUGSCORE scores of correctly docked cyclosporin conformers (Figure 9) in CypA conformers indicates that the lower energy conformers generally

led to more favorable cyclosporin interactions. The CypA conformers with no symbols next to their names (conformers 164, 389, 431, and 548) could not recognize and accommodate ligand conformers in the correct binding mode. The energy values next to the protein conformer names are the potential energies of the ligand free conformers calculated with the CHARMM22 force field after a short energy minimization. The crystallographic structure of CypA from PDB entry 1bck (1bck\_000) was energy minimized the same way as the ROCK conformers, and was found to be the lowest energy conformation. It was expected that this structure would be among the lowest-energy conformations due to the energy refinement steps performed during crystallographic structure determination. We also know that this is the bioactive conformation (taken from the complex with cyclosporin) and thus is the lowest energy conformation of CypA in complex with cyclosporin (neglecting any energetic perturbations from crystal lattice formation). The potential energies of the ROCK generated conformers are of similar magnitude, affirming these are energetically reasonable conformations. Protein conformers with the lowest energies could accommodate the largest number of cyclosporin conformers, whereas those with higher energies could recognize only a few ligand conformers or none at all. The score distributions of the protein-ligand complexes are similar for the low energy CypA conformers and the 1bck x-ray structure. Low-energy conformers of CypA had similarly favorable interactions with cyclosporin in comparison with the x-ray structure, as judged by the SLIDE scoring function (where higher values are better). According to SLIDE scores, the best dockings correlate almost directly with the energy of the CypA conformation. DRUGSCORE was more sensitive and gave a considerably more favorable CypA-cyclosporin interaction score to the x-ray structure, relative to the lowest-energy conformers of CypA in complex with cyclosporin (where lower DRUGSCORE values are better). DRUGSCORE also clearly distinguished high-energy conformers as giving less favorable interactions with cyclosporin.

## **Estrogen receptor – Zearalenol**

### ***Flexibility Analysis of Estrogen Receptor***

FIRST flexibility analysis showed that the estrogen receptor structure selected for this study (PDB entry 1ere) with the ligand removed had a largely rigid binding site (Figure 10). For this reason, the zearalenol conformers were docked into the target structure 1ere with only the protein side chains but not the main chain handled flexibly; side-chain flexibility was incorporated by SLIDE to reflect that active-site side chains would likely need to accommodate zearalenol in different conformations than they adopted for 17 $\beta$ -estradiol (the ligand in the 1ere structure).

### ***Conformer Generation***

As in the case of cyclosporin, only the covalent bond lengths and coordination angles were used as constraints for the zearalenol conformational search. Two thousand zearalenol conformers were generated with ROCK with no more than 40% of the rotatable bonds rotated a maximum of 10 degrees at each step. This provided a diverse set of zearalenol conformers (Figure 11A) with a maximum RMSD of 3.3 Å compared to the starting conformation. The selection of the parameters (percent of bonds to rotate and maximum angular step size) depends on the size and complexity of the molecule for which the conformational space is explored. Details of ROCK applied to sample flexibility in ring systems can be found elsewhere<sup>35,36,47</sup>. More bonds should be rotated with larger maximum angles in the case of small molecules like zearalenol, while smaller parameter values more often result in collision-free conformers (hence, more efficient sampling) for larger molecules like cyclosporin or proteins with coupled ring systems.

## Docking

The 2000 ROCK conformers and the original x-ray structure of zearalenol (CSD entry BEGDAS) were used for docking. The zearalenol conformers were docked with an average speed of 7 seconds/ligand (elapsed time) using a template consisting of 39 points (on a PentiumIII 800MHz processor). Based on the number of dockings and their scores, two main zearalenol orientations were produced by SLIDE (Figure 11 panels C and D), one rotated by 180° relative to the other about the long axis of the zearalenol molecule. To illustrate the differences between the best docked and free zearalenol conformations, top scoring representatives of the two types of dockings are shown in Figure 11B, with heavy atom RMSD values relative to the free structure of 1.1 Å (magenta) and 1.6 Å (white) following superposition of the aromatic ring only. Both docked orientations showed a good steric match with the bound 17β-estradiol. The distribution of the scores of these dockings shows that SLIDE scored the first type of docking higher (Figure 12A), while DRUGSCORE preferred the second docking (Figure 12B; lower DRUGSCORES indicate more favorable dockings).

The interactions of the two zearalenol orientations with ER binding site residues are shown in the LIGPLOT<sup>70</sup> diagrams (Figure 13AB). Both orientations made a number of good interactions with the protein, except for the O1 oxygen of zearalenol, with its hydrogen bonding potential remaining unsatisfied in both cases. The hydrogen bonds formed by the native ligand 17β-estradiol, as well as the two types of zearalenol dockings with the binding site residues of estrogen receptor, are summarized in Table II. Both docking types made, on average, two out of the three hydrogen bonds of the native ligand, plus one or more additional hydrogen bonds. The protein-ligand hydrophobic contacts show a similar picture: some native contacts were made by the type 1 docking, some by the other, and both types made a number of additional contacts not observed between 17β-estradiol and the ER.

## DISCUSSION

The goal of this work is to develop and apply a method that can be used to sample bioactive conformations and conformational changes in proteins and their ligands, while incorporating the correlated motions that result from networks of covalent and noncovalent interactions. Both CypA and its ligand cyclosporin apparently have significant flexibility. In addition to evaluating the flexibility of CypA starting from the cyclosporin-bound structure (PDB entry 1bck), FIRST analysis was also performed on the ligand-free crystal structure<sup>71</sup> (PDB entry 2cpl), with a resolution of 1.63 Å. The results obtained were in excellent agreement with those for 1bck, indicating that FIRST results are not very dependent on the details of the individual x-ray structures, if they have good stereochemistry. The good agreement between the flexibility predictions of FIRST and the NMR and crystallographic data<sup>58,59,67</sup> indicating which regions are the most flexible in CypA (Figures 4, 6, and 8), as well as previous results on other proteins<sup>34,45,54</sup>, suggest that FIRST is a reliable method to predict the flexible regions of a protein from a single x-ray structure.

Comparing the FIRST flexibility predictions for CypA in complex with the *cis* form of one of its substrates, SIN-Ala-Ala-Pro-Phe-NIT (PDB code 1rmh), with the experimental NMR data measuring the dynamics of the enzyme during catalysis<sup>59</sup>, good agreement was found between the regions of CypA predicted to be flexible by FIRST (Figure 8E) and those found to show microsecond time scale dynamics by NMR (Figure 8F; data and figure generously provided by Dorothee Kern and Elan Eisenmesser<sup>59</sup>). Regarding the peptidyl ligand,

flexibility analysis predicted the N-terminal part to be flexible and surrounded by flexible loops of the protein, while the C-terminal half remained rigid (Figure 8E), attached to the rigid core of the protein by hydrophobic interactions. Based on these results, it seems most feasible for the N-terminal end of the peptide to rotate during the prolyl isomerization catalyzed by CypA to produce the *trans* conformation of the peptide. Kern and coworkers envision a different trajectory<sup>59</sup>, suggesting that the C-terminal end of the substrate is likely to swing around when the peptide bond is rotated by 180°. Further experiments are needed to elucidate the details of this process.

ROCK is unique in being designed to sample flexibility in multiple interlocked ring systems formed by covalent and non-covalent interactions and thus can model significant correlated motions, using a non-forcefield approach. The protein main-chain conformers generated for CypA are of good stereochemical quality and span a conformational space similar to that found in NMR solution structures. Comparisons between MD and the FIRST flexibility analysis used as input to ROCK show very good agreement for Ras, Raf, and the Ras-Raf complex<sup>9</sup>. Shortcomings of this method are the lack of a timescale associated with the modeled motions, as well as the lack of an energy function that allows assessment of the relative likelihood of the generated conformers. This is something our future work will address. However, the incorporation of hydrogen-bond and hydrophobic interactions for most side chains in the current implementation already ensures energetically favorable positioning of side chains in most cases. A major strength in the combined FIRST/ROCK approach is that it automatically identifies groups within the molecules that undergo coupled motion, without the need for expensive normal modes or essential dynamics calculations. Also, an efficient way of sampling the side chains on the modified main chains will be implemented to find not only a feasible side chain conformation for each residue, but also to identify the most favorable orientations.

When there is only one crystal or NMR structure available for a protein of interest, it can be difficult to identify which regions are flexible, and decide how much flexibility sampling is required for realistically representing the behavior of the protein. FIRST flexibility analysis can help provide this information. We used FIRST to identify that all but two side chains (Arg 394 and Phe 404) that are in direct interaction with 17 $\beta$ -estradiol in ER are attached to mutually rigid main chain segments (Figure 10). Phe 404 was shown to be anchored to neighboring hydrophobic side chains by several hydrophobic interactions. The only side chain free to move was Arg 394 at one end of the binding pocket. Since SLIDE explores side chain flexibility, no further conformational sampling by ROCK was necessary for ER. These findings are in good agreement with the results of Oostenbrink et al.<sup>72</sup>, who performed MD simulations on several ER – agonist complexes. They observed that the hydrogen bond between one of the ligand hydroxyl groups and the imidazole group of His 524 is formed 97% of the time during the simulation, while the other hydroxyl group of the ligand interacts with Glu 353 and Arg 394 only 40-60% of the time. This is consistent with greater flexibility of Arg 394 relative to the rest of the binding site.

The detailed analysis of the interactions between the zearalenol dockings and ER revealed that, on average, the first type of zearalenol docking made one hydrogen bond less than the second type (2.7 versus 3.6), while being positioned slightly off-center (closer to one side of the binding pocket) and engaging in extensive van der Waals contacts with the protein. These interactions increase the hydrophobic complementarity term of the SLIDE scores, resulting in higher average SLIDE scores for type 1 dockings. DRUGSCORE preferred the second docking type because the more central location of the zearalenol molecule in the binding site resulted in more favorable protein-ligand distance pairs over all ligand atoms. One might expect the new

ligand, in this case zearalenol, to make similar interactions with the protein as the native 17 $\beta$ -estradiol. In spite of their high scores, type 2 dockings fail to form a hydrogen bond with ND1 of His 524 in the majority of the cases, while this hydrogen bond was shown to be formed 97% of the time by 17 $\beta$ -estradiol during the MD simulation (Oostenbrink et al., 2000). The type 1 dockings do satisfy this hydrogen bond more than 50% of the time. On the other hand, NH2 of Arg 394 is unavailable for hydrogen bonding to the type 1 dockings due to steric constraints. These dockings form a hydrogen bond with the backbone oxygen of Leu 346 instead. Thus, based on the similarity of interactions of the zearalenol dockings with those made by the native ligand, it is not possible to clearly decide which docking type is more favorable, as some native interactions are lost but other favorable interactions are gained.

We suggest a single ER mutation to test which is the most favored binding mode of zearalenol, based on the interactions shown in Figure 13AB. Mutating Arg 394 to methionine would affect the hydrogen bonding network of the second, but not the first docking. To estimate the effects of this mutation, the two top scoring zearalenol dockings (Figure 13) were rescored using the protein structure after modeling the Arg394Met residue replacement using InsightII (Accelrys), followed by a short energy minimization of the replaced side chain and neighboring side chains within 5 Å. As expected, the Arg394Met replacement resulted in worse scores for the type 2 docking compared to the original (a change of 13138 relative to the -510452 DRUGSCORE of the wild type and a change of -2.2 relative to the 28.9 SLIDE score of the wild type), with minor or no effect on type 1 (a change of 4215 relative to the -469580 DRUGSCORE of the wild type ER and no change relative to the SLIDE score of 71.4 for wild type ER).

Comparison of the interactions made by 17 $\beta$ -estradiol and zearalenol with the ER can shed light on the molecular basis of zearalenol mimicry of 17 $\beta$ -estradiol. The main determinants of estradiol binding to ER are the two sets of hydrogen bonds formed at the ends of the ligand, the interactions with its aromatic ring, and additional nonpolar interactions with the central part of the steroid ring system<sup>73</sup>. The transcriptional activator property of 17 $\beta$ -estradiol is determined by its ability to maintain helix 12 of ER in a position where it seals the binding site, an alignment typically prevented by ER antagonists<sup>60</sup>. The dockings provided by SLIDE (Figure 11 panels C and D) indicate not only that the agonist zearalenol mimics the shape of 17 $\beta$ -estradiol, but also that it makes most of the same critical interactions with the ER binding site: hydrogen bonds at one or both ends, its aromatic ring in the same position as that of 17 $\beta$ -estradiol, and extensive nonpolar interactions involving the central part of the molecule. Zearalenol also fits into the binding pocket without having to disturb the position of the transactivator helix 12 of the ER, and as such, acts as an agonist<sup>44</sup>.

Out of the 2000 zearalenol conformers generated with ROCK, 1750 could be docked into the binding site of estrogen receptor sterically, though with different apparent affinities (docking scores). The large number of possible binding conformations of this ligand may indicate an unusually large degree of freedom maintained by the ligand after binding.

## CONCLUSIONS

In this paper, we have presented a new methodology to generate an ensemble of low energy conformations for a protein and two macrocyclic ligands, showing that they indeed can recognize each other in a range of conformations surrounding the crystal structure. Protein conformers with good stereochemistry are generated from a single starting structure, and these conformers can be used to model protein main chain flexibility and its influence on ligand recognition. Comparing these conformers with experimentally determined NMR conformations

for CypA showed that this method works well for modeling main-chain flexibility. The conformers correspond to snapshots of the molecule during moderate to large-scale correlated motions typically required for biological function. Another useful feature of this technique is its applicability to both proteins and ligands. The docking of cyclosporin conformers to CypA conformers using SLIDE confirmed the hypothesis that there is a considerable amount of flexibility tolerated in CypA-cyclosporin recognition, reflected by the fact that multiple target structures accommodate a wide range of ligand conformers while maintaining key interactions. This flexible docking approach also provided a testable prediction of the binding orientation of the natural product zearalenol and how it mimics the physiological hormone 17 $\beta$ -estradiol in activating the estrogen receptor. This approach to modeling biologically significant protein main-chain and macrocyclic ligand conformations is expected to enable a range of future applications, including probing the effects of individual hydrogen bonds and hydrophobic contacts on the available conformational space, and providing a flexible representation of proteins for structure based ligand design.

FIRST, ROCK, and SLIDE are available to the modeling community. A webserver for the Arizona State University version of FIRST is accessible at <http://firstweb.asu.edu> for academic users. Source code for FIRST, ROCK, and SLIDE is distributed through Michigan State University (contact [KuhnL@msu.edu](mailto:KuhnL@msu.edu)). More information about the software and licensing can be found at <http://www.bch.msu.edu/labs/kuhn> under Software.

## ACKNOWLEDGMENTS

The authors thank Marcel Ottiger and Kurt Wüthrich (ETH, Zurich) for providing the HD exchange data shown in Figure 6C and Elan Eisenmesser and Dorothee Kern (Brandeis University) for their dynamics data on CypA shown in Figure 8F. We also thank Judith Guenther and Gerhard Klebe (University of Marburg) for providing DRUGSCORE for our use, Tim Zacharewski and Jason Matthews (Michigan State University) for their insights on estrogen receptor-ligand interactions, and Michael Feig (Michigan State University) for his expertise with the MMTSB toolset.

## REFERENCES

1. Linderstrom-Lang, K. U. and Schellman, J. Protein structure and enzyme activity. Boyer, Lardy, and Myrbäck. The Enzymes, 1959,
2. Volkman BF, Lipson D, Wemmer DE, Kern D. Two-state allosteric behavior in a single-domain signaling protein. *Science* 291:2429-2433, 2001.
3. Frauenfelder H, Parak F, Young RD. Conformational substates in proteins. *Annu. Rev. Biophys. Chem.* 17:451-479, 1988.
4. Rader SD, Agard DA. Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 Å resolution. *Protein Sci.* 6:1375-1386, 1997.
5. Yu L, Zhu CX, Tse-Dinh YC, Fesik SW. Backbone dynamics of the C-terminal domain of Escherichia coli topoisomerase I in the absence and presence of single-stranded DNA. *Biochemistry* 35:9661-9666, 1996.
6. Olejniczak ET, Zhou MM, Fesik SW. Changes in the NMR-derived motional parameters of the insulin receptor substrate 1 phosphotyrosine binding domain upon binding to an interleukin 4 receptor phosphopeptide. *Biochemistry* 36:4118-4124, 1997.
7. Zidek L, Novotny MV, Stone MJ. Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat. Struct. Biol.* 6:1118-1121, 1999.
8. Arumugam S, Gao G, Patton BL, Semenchenko V, Brew K, Van Doren SR. Increased backbone mobility in beta-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. *J. Mol. Biol.* 327:719-734, 2003.
9. Gohlke, H., Kuhn, L. A., and Case, D. A. Change in protein flexibility upon complex formation: Analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins*, in press, 2004.
10. 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.* 13:547-562, 1999.
11. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. *Proteins* 39:261-268, 2000.
12. Althaus E, Kohlbacher O, Lenhof HP, Muller P. A combinatorial approach to protein docking with flexible side chains. *J. Comput. Biol.* 9:597-612, 2002.
13. Broughton HB. A method for including protein flexibility in protein-ligand docking: Improving tools for database mining and virtual screening. *J. Mol. Graph. Model.* 18:247-4, 2000.
14. Kallblad P, Dean PM. Efficient conformational sampling of local side-chain flexibility. *J. Mol. Biol.* 326:1651-1665, 2003.

15. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* 266:424-440, 1997.
16. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* 235:345-356, 1994.
17. Leach AR, Lemon AP. Exploring the conformational space of protein side chains using dead-end elimination and the A\* algorithm. *Proteins* 33:227-239, 1998.
18. Schaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins* 33:295-310, 1998.
19. Betts MJ, Sternberg MJ. An analysis of conformational changes on protein-protein association: Implications for predictive docking. *Protein Eng.* 12:271-283, 1999.
20. Carlson HA. Protein flexibility and drug design: How to hit a moving target. *Curr. Opin. Chem. Biol.* 6:447-452, 2002.
21. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47:409-443, 2002.
22. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: Insights from an open to closed conformational isomers. *Proteins* 32:159-174, 1998.
23. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 308:377-395, 2001.
24. Schnecke V, Kuhn LA. Database screening for HIV protease ligands: The influence of binding-site conformation and representation on ligand selectivity. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 242-251, 1999.
25. Carlson HA, Masukawa KM, McCammon JA. Method for including the dynamic fluctuations of a protein in computer-aided drug design. *J. Phys. Chem. A* 103:10213-10219, 1999.
26. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J. Am. Chem. Soc.* 124:5632-5633, 2002.
27. Bosshard HR. Molecular recognition by induced fit: How fit is the concept? *News Physiol. Sci.* 16:171-173, 2001.
28. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57:213-218, 2000.
29. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: A matter of pre-existing populations. *Protein Sci.* 11:184-197, 2002.

30. Koshland DEJ. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA* 44:98-104, 1958.
31. 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* 46:34-40, 2002.
32. Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* 37:228-241, 1999.
33. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* 43:2100-2114, 2000.
34. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. *Proteins* 44:150-165, 2001.
35. Lei, M., Kuhn, L. A., Zavodszky, M. I., and Thorpe, M. F. Sampling Protein Conformations and Pathways. *J. Comp. Chem.* in press, 2004.
36. Thorpe, M. F. and Lei, M. Macromolecular flexibility. *Philosophical Magazine Special Edition*, in press, 2004.
37. Schnecke V, Kuhn LA. Virtual screening with solvation and ligand-induced complementarity. *Perspectives in Drug Discovery and Design* 20:171-190, 2000.
38. Hacker J, Fischer G. Immunophilins: Structure-function relationship and possible role in microbial pathogenicity. *Mol. Microbiol.* 10:445-456, 1993.
39. Matsuda S, Koyasu S. Mechanisms of action of cyclosporin. *Immunopharmacology* 47:119-125, 2000.
40. Ke H, Mayrose D, Belshaw PJ, Alberg DG, Schreiber SL, Chang ZY, Etkorn FA, Ho S, Walsh CT. Crystal structures of cyclophilin A complexed with cyclosporin A and N-methyl-4-[(E)-2-butenyl]-4,4-dimethylthreonine cyclosporin A. *Structure* 2:33-44, 1994.
41. Sapphire AC, Bobardt MD, Gallay PA. Human immunodeficiency virus type 1 hijacks host cyclophilin A for its attachment to target cells. *Immunol. Res.* 21:211-217, 2000.
42. Miksicek RJ. Interaction of naturally occurring nonsteroidal estrogens with expressed recombinant human estrogen receptor. *J. Steroid Biochem. Mol. Biol.* 49:153-160, 1994.
43. Kuiper-Goodman T, Scott PM, Watanabe H. Risk assessment of the mycotoxin zearalenone. *Regul. Toxicol. Pharmacol.* 7:253-306, 1987.
44. Matthews J, Celius T, Halgren R, Zacharewski T. Differential estrogen receptor binding of estrogenic substances: a species comparison. *J. Steroid Biochem. Mol. Biol.* 74:223-234, 2000.

45. Rader AJ, Hespeneide BM, Kuhn LA, Thorpe MF. Protein unfolding: Rigidity lost. *Proc. Natl. Acad. Sci. USA* 99:3540-3545, 2002.
46. Jacobs DJ, Kuhn LA, Thorpe MF. Flexible and rigid regions in proteins. In: Thorpe MF, Duxbury PM, eds., *Rigidity theory and applications*. New York: Kluwer Academic/Plenum Publishers, 1999.
47. Thorpe MF, Lei M, Rader AJ, Jacobs DJ, Kuhn LA. Protein flexibility and dynamics using constraint theory. *J. Mol. Graph. Model.* 19:60-69, 2001.
48. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* 23:283-438, 1968.
49. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins* 12:345-364, 1992.
50. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. Procheck - A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26:283-291, 1993.
51. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 28:235-242, 2000.
52. Go N, Scheraga HA. Ring closure and local conformational deformations of chain molecules. *Macromolecules* 3:178-187, 1970.
53. Hespeneide BM, Rader AJ, Thorpe MF, Kuhn LA. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Graph. Model.* 21:195-207, 2002.
54. Thorpe MF, Hespeneide BM, Yang Y, Kuhn LA. Flexibility and critical hydrogen bonds in cytochrome c. *Pac. Symp. Biocomput.* 191-202, 2000.
55. Schnecke V, Swanson CA, Getzoff ED, Tainer JA, Kuhn LA. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* 33:74-87, 1998.
56. Zavodszky MI, Sanschagrín PC, Korde RS, Kuhn LA. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. *J. Comput. Aided Mol. Des.* 16:883-902, 2002.
57. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 295:337-356, 2000.
58. Kallen J, Mikol V, Taylor P, Walkinshaw MD. X-ray structures and analysis of 11 cyclosporin derivatives complexed with cyclophilin A. *J. Mol. Biol.* 283:435-449, 1998.
59. Eisenmesser EZ, Bosco DA, Akke M, Kern D. Enzyme dynamics during catalysis. *Science* 295:1520-1523, 2002.

60. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M. Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389:753-758, 1997.
61. Allen FH. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallogr. B* 58:380-388, 2002.
62. Vriend G. What If - A Molecular Modeling and Drug Design Program. *Journal of Molecular Graphics* 8:52-56, 1990.
63. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4:187-217, 1983.
64. Feig, M., Karanicolas, J., and Brooks III, C. L. MMTSB Tool Set, MMTSB NIH Research Resource, The Scripps Research Institute. 2001,
65. Dahiyat BI, Gordon DB, Mayo SL. Automated design of the surface positions of protein helices. *Protein Sci.* 6:1333-1337, 1997.
66. Baldwin RL. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U. S. A* 83:8069-8072, 1986.
67. Ottiger M, Zerbe O, Guntert P, Wuthrich K. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* 272:64-81, 1997.
68. Freire E. Can allosteric regulation be predicted from structure? *Proc. Natl. Acad. Sci. U. S. A* 97:11680-11682, 2000.
69. Pan H, Lee JC, Hilser VJ. Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. U. S. A* 97:12020-12025, 2000.
70. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 8:127-134, 1995.
71. Ke H. Similarities and differences between human cyclophilin A and other beta-barrel structures. Structural refinement at 1.63 Å resolution. *J. Mol. Biol.* 228:539-550, 1992.
72. Oostenbrink BC, Pitera JW, van Lipzig MM, Meerman JH, Van Gunsteren WF. Simulations of the estrogen receptor ligand-binding domain: affinity of natural ligands and xenoestrogens. *J. Med. Chem.* 43:4594-4605, 2000.
73. Anstead GM, Carlson KE, Katzenellenbogen JA. The estradiol pharmacophore: ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids* 62:268-303, 1997.

**Table I.** CypA-cyclosporin interactions (PDB structure 1bck) that were monitored to identify correct cyclosporin dockings.

		<b>CypA</b>	<b>Cyclosporin</b>	<b>Distance in Crystal Structure</b>
1	Hydrogen bond	Gln63: NE2	BMT1: O	3.18 Å
2	Hydrogen bond	Asn102: O	Thr2: N	2.97 Å
3	Hydrogen bond	Asn102: N	MVA11: O	3.54 Å
4	Hydrophobic interaction	Phe113: CD1	MVA11: CG1	3.44 Å

**Table II.** Hydrogen bonds formed by the native ligand 17 $\beta$ -estradiol and the two orientations of zearalenol with binding site residues in the estrogen receptor.

	<b>17<math>\beta</math>-estradiol</b>	<b>Zearalenol</b>	
		<b>type 1 docking</b>	<b>type 2 docking</b>
<b># H-bonds</b> (average values for zearalenol)	3	2.7	3.6
<b>H-bonding protein residue</b>	<b># H-bonds</b>	<b>% with H-bond</b>	<b>% with H-bond</b>
GLU 353	1	83.1%	92.1%
ARG 394	1	0.0%	92.7%
HIS 524	1	51.1%	5.7%
GLY 521	0	28.7%	5.5%
LEU 346	0	94.9%	60.8%
LEU 387	0	13.0%	94.6%
THR 347	0	0.0%	0.5%

## Figure Legends

*Figure 1.* The cyclic undecapeptide Thr2-cyclosporin. The difference between Thr2-cyclosporin and the immunosuppressive drug cyclosporin A is that the second residue of the latter does not have an OH group. The two cyclosporin molecules have comparable biological activity. The complex of human CypA–Thr2-cyclosporin was selected as the starting point for this study because of its higher resolution crystal structure (PDB code 1bck, resolution 1.8 Å) compared to the human CypA–cyclosporin A complex (PDB code 1cwa, resolution 2.1 Å).

*Figure 2.* Structures of zearalenone and its metabolites, alpha- and beta-zearalenol, which are estrogen receptor agonists.

*Figure 3.* Relative stability of regions in the unliganded human CypA as a function of hydrogen-bond dilution, simulating thermal denaturation<sup>53</sup>. The numbering along the top, from left to right, represents the amino acid sequence of the protein (residues 1-165). Each line below is colored to indicate the rigid and flexible parts present in the structure at the energy value listed at the left of the line. Thin black lines represent the flexible parts, while thick bars represent rigid regions, with identical colors for mutually rigid regions that belong to the same rigid cluster (which may or may not be contiguous in sequence). The consecutive lines illustrate the changes in flexibility of the protein as hydrogen bonds are removed in order of their increasing energy. The first and second columns list the index number (rank, from strongest to weakest, where the strongest H-bond is given index 1, etc.) and energy of those hydrogen bonds whose breakage induces a rigid to flexible change in the structure. The last two columns specify the residue numbers of the hydrogen donor (blue) and acceptor (red) of the respective hydrogen bond, and the donor and acceptor positions are also shown by carets beneath the rigid cluster decomposition plot. The third column lists the mean coordination number at the current H-bond energy.

*Figure 4.* Ribbon diagram of the ligand-free (A) and ligand-bound (B) CypA structures (PDB code 1bck) showing the flexible and rigid regions of the protein colored by flexibility index. Grey regions are isostatic or just rigid, blue regions are overconstrained, having more than enough bonds to make them rigid, while yellow to red regions are flexible. The binding site is occupied by the cyclosporin ligand (colored green) in (B). Comparison of the FIRST flexibility analysis of CypA with (A) and without (B) cyclosporin shows that the ligand rigidifies parts of the protein even at regions distant from the binding site (for example loops 24-29 and 87-94). (C) The 12 most distinct conformers of CypA generated by ROCK. Atoms of the ligand rendered in magenta spheres indicate the location of the binding site. (D) The ribbon diagram of the lowest energy NMR structure of free CypA (PDB code 1oca). The thickness of the tube is proportional to the maximum deviation of the backbone Ca atoms from the average Ca position of the 20 NMR structures from PDB entry 1oca. (E) The network of covalent bonds (blue tubes), hydrogen bonds (magenta tubes) and hydrophobic interactions (white spheres) of CypA identified by FIRST. (F) The flexible ring clusters, or collective motions, sampled by ROCK, where each independent ring cluster is colored by a different color. The blue region is rigid due to its overconstrained bond network. While the purple, pink and orange regions are simple loops consisting of a single ring, the regions colored in red, yellow, and green are complex ring clusters made up of 7 to 14 interlocked rings with multiple bond rotational degrees of freedom.

*Figure 5.* Ramachandran plots (generated with PROCHECK<sup>49,50</sup>) of the crystal structure of CypA (PDB entry 1bck) (A) and of the most distinct ROCK conformer (B). The dark grey areas correspond to most favored core regions, medium grey to allowed regions, and light grey to generously allowed regions in  $\phi, \psi$  main chain dihedral angle space. For the crystal structure, 86% of residues fall in the most favored regions of  $\phi, \psi$  space, and for the conformer (from the 175th iteration of ROCK), 80% are in these regions.  $\psi$  represents Gly residues, while  $\phi$  represents all other residue types.

*Figure 6.* The average of the pairwise  $\phi$  (A) and  $\psi$  (B) differences of the 12 most distinct CypA conformers generated with ROCK. To facilitate visual analysis of positive correlation with (A) and (B), the hydrogen-deuterium (HD) exchange rates of the backbone amide protons of free CypA are shown as white bars on a black background in (C). (Data generously provided by Marcel Ottiger and Kurt Wüthrich<sup>67</sup>.) Exchange rates of  $-1$  indicate that the exchange was too fast to be observed. Exchange rates of  $-5$  indicate very slow exchange. Note the high degree of correlation between regions with slow HD exchange and rigid regions from FIRST/ROCK ( $\phi, \psi = 0$ ). Complementarily, the flexible regions in the protein, appearing as peaks in (A) and (B), are found to correspond to regions of fast HD exchange (black regions in (C)). The locations of the regular secondary structure elements are shown at the bottom of (C), where straight lines indicate  $\beta$ -sheets and wavy lines correspond to  $\alpha$ -helices.

*Figure 7.* (A) Comparing the crystal structures of free (CSD code KEPNAU) and protein-bound conformations (PDB entry 1bck) of cyclosporin. The arrow indicates the location of the peptide bond between residues 9 and 10, which is *cis* in the free structures but *trans* in the protein-bound one, resulting in major conformational change of the cyclic compound, as indicated by the superposition of *cis* and *trans* conformers based on the backbone atoms of residues 8 and 9. CypA catalyzes the interconversion between *cis* and *trans* peptide bonds. (B) A subset of cyclosporin conformers (only backbones are shown) generated with ROCK from the protein-bound conformation (taken from PDB entry 1bck). The x-ray structure of cyclosporin is shown in thicker black bars.

*Figure 8.* (A) Cyclosporin conformers (showing only the main chains as tubes) docked into the binding sites of the ROCK-generated CypA conformers (shown as ribbons corresponding to the protein main chain). Each docked cyclosporin conformer is colored the same as the CypA conformer to which it docked best. (B) Ribbon diagrams of the 12 most distinct CypA conformers colored by flexibility index with the docked cyclosporin conformers (backbones shown as green tubes), illustrating the range of conformational changes during the protein-ligand recognition process as modeled by the FIRST-ROCK-SLIDE combined method. The magenta tube is the x-ray conformation of cyclosporin docked into the x-ray conformation of CypA. (C) Backbone diagrams of the 22 NMR structures of the CypA-cyclosporin complex (PDB code 3cys) colored by the average RMSD values of the individual atoms relative to the corresponding atoms from the other 21 structures. (D) The ribbon diagram of the CypA-cyclosporin complex colored by the crystallographic temperature factor (PDB code 1bck). (E) FIRST flexibility analysis of CypA in complex with the *cis* form of SIN-Ala-Ala-Pro-Phe-NIT (PDB code 1rmh) shows that the N-terminal part of the substrate is flexible while its C-terminus is part of the same rigid cluster as the protein's rigid core. This view of the protein can be obtained from the position seen in panels A-D by a 90 degree rotation to the right around the vertical axis. (F) Residues in CypA exhibiting microsecond time scale dynamics during catalysis (Data and figure

generously provided by Dorothee Kern and Elan Eisenmesser<sup>59</sup>). Structure of the *cis* conformation of the substrate Suc-Ala-Phe-Pro-Phe-4-NA (green) bound to CypA, based on the x-ray structure of CypA complexed with the *cis* form of Suc-Ala-Ala-Pro-Phe-4-NA (PDB entry 1rmh). CypA residues with chemical exchange in both the presence and absence of substrate are color coded in blue (F67, N71, G74, S77, and S110). Residues in red exhibit chemical exchange only during turnover (R55, K82, L98, S99, A101, N102, A103, and G109). Residues shown in magenta exhibit chemical exchange in the absence of substrate, but increase in its presence (T68 and G72). Some additional residues were found to be dynamic in the microsecond time regime using the more sensitive CPMG method (Eisenmesser et al., to be published).

*Figure 9.* The distribution of ligand-protein interaction scores from (A) SLIDE and (B) DRUGSCORE for cyclosporin conformers correctly docked into the CypA conformers. The CypA conformers with no symbols next to their names (conformers 164, 548, 431, and 389) could not recognize and accommodate ligand conformers in the correct binding mode. The energy values next to the protein conformer names are the potential energies of the ligand free CypA conformers calculated with the CHARMM22 force field<sup>63</sup>.

*Figure 10.* The ligand binding site of estrogen receptor with bound 17 $\beta$ -estradiol (shown as van der Waals spheres). The light gray ribbons represent the main chain portion of the estrogen receptor predicted to be flexible by FIRST. Among the residues lining the binding site, only Phe 404 and Arg 394 were part of this flexible region of the main chain.

*Figure 11.* (A) Alpha-zearalenol conformers generated with ROCK. The thick tube shows the starting conformation (CSD structure BEGDAS). (B) Comparing two of the top scoring zearalenol conformers to the CSD structure. The conformer shown in magenta was docked in the orientation shown on (C), while the one shown in white was docked in the orientation shown in (D). There were two dominant binding modes for the conformers, shown in (C) and (D) in comparison with the crystal structure orientation of 17 $\beta$ -estradiol (PDB structure 1ere). Docked zearalenol conformers are shown in tubes colored by atom type, where green is used for carbon and red for oxygen atoms. 17 $\beta$ -estradiol is shown in yellow tubes, and the protein side chains are shown as purple tubes, from crystal structure 1ere.

*Figure 12.* Score distributions of all zearalenol conformations according to (A) SLIDE and (B) DRUGSCORE. Higher SLIDE scores are more favorable, whereas lower values are more favorable in DRUGSCORE.

*Figure 13.* Ligplot<sup>70</sup> diagrams showing the interactions of the two main types of zearalenol dockings with binding site residues of estrogen receptor (PDB entry 1ere): (A) the top scoring orientation by SLIDE score (type 1), and (B) the top scoring orientation by DRUGSCORE<sup>57</sup> (type 2).



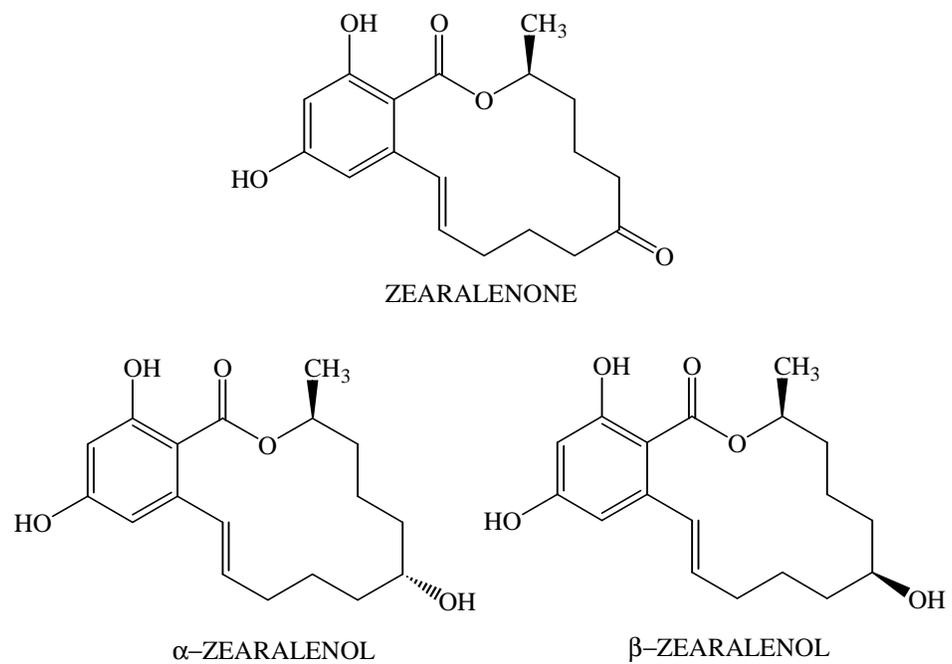


Figure 2.

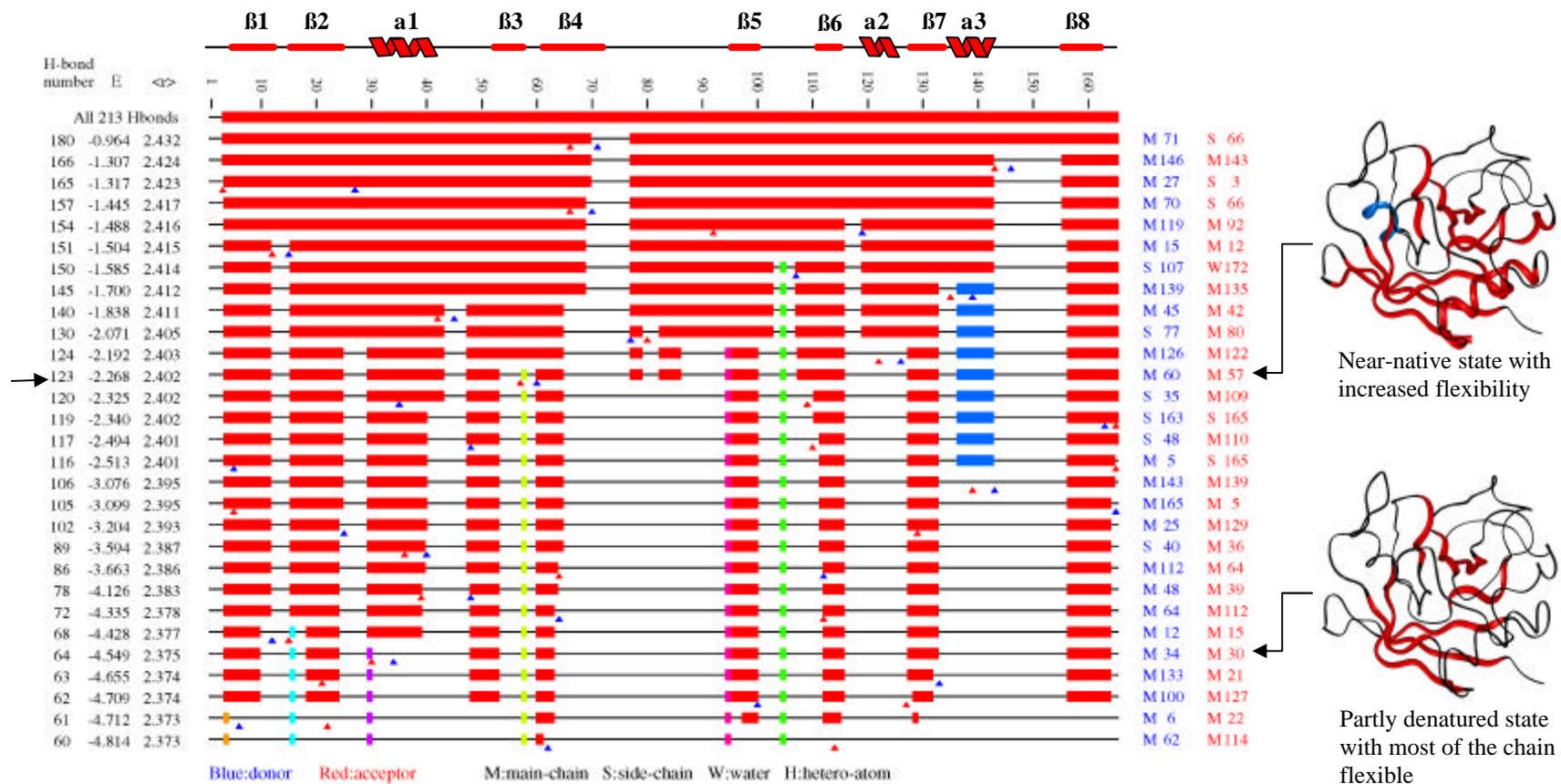


Figure 3.

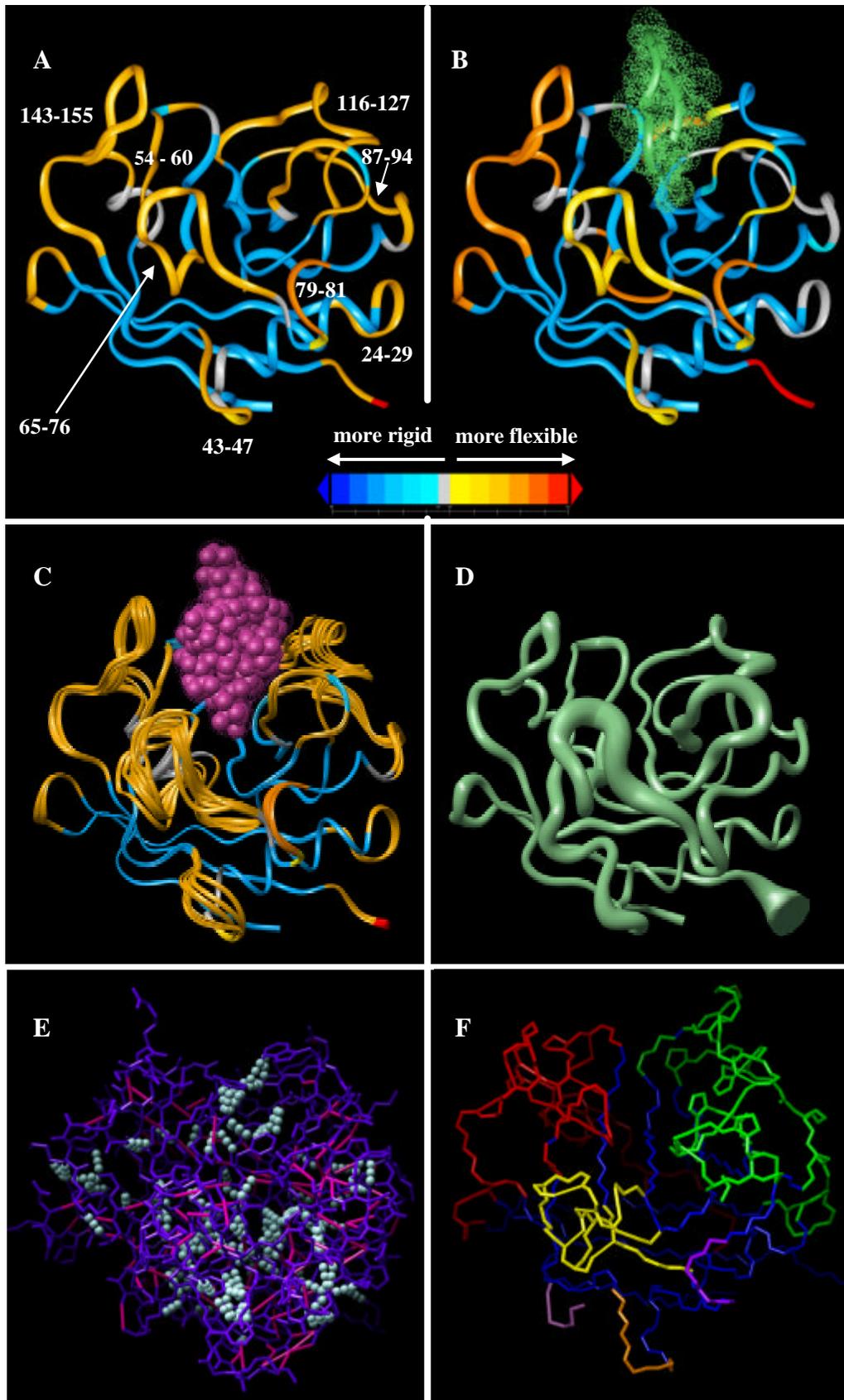


Figure 4.

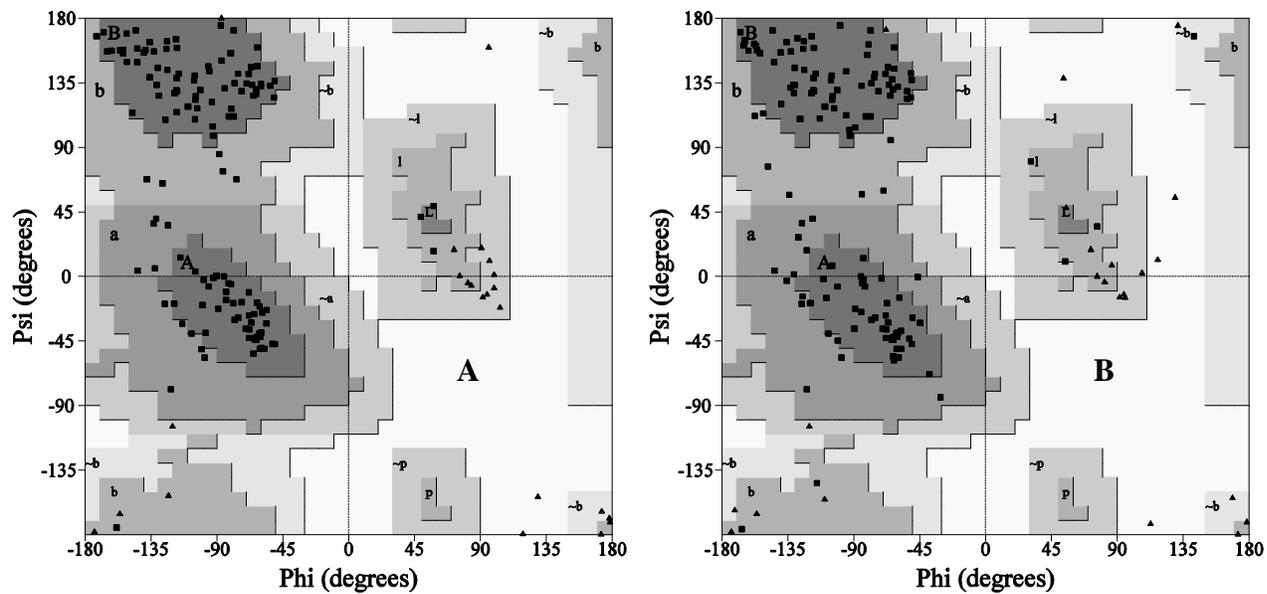


Figure 5.

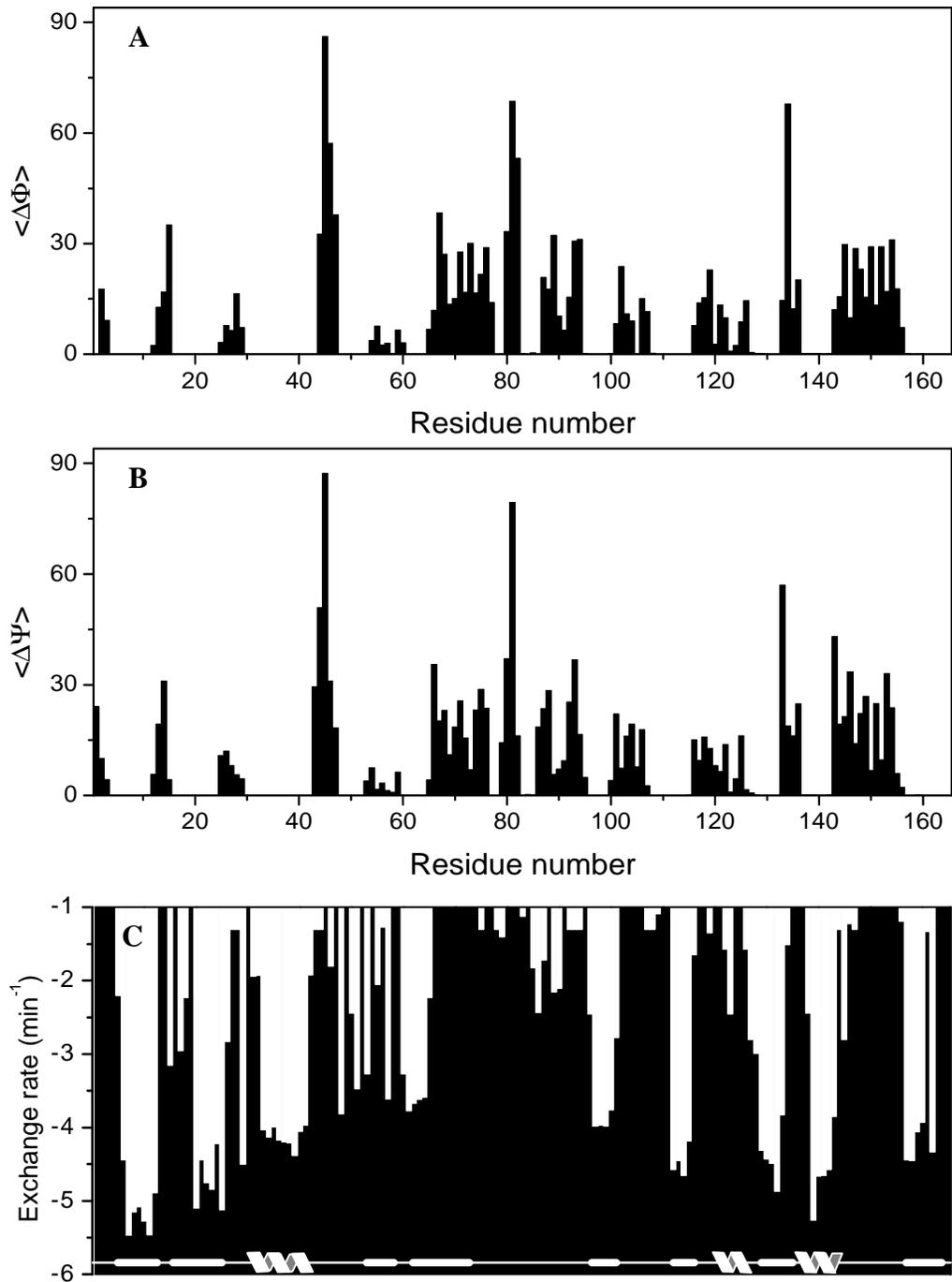
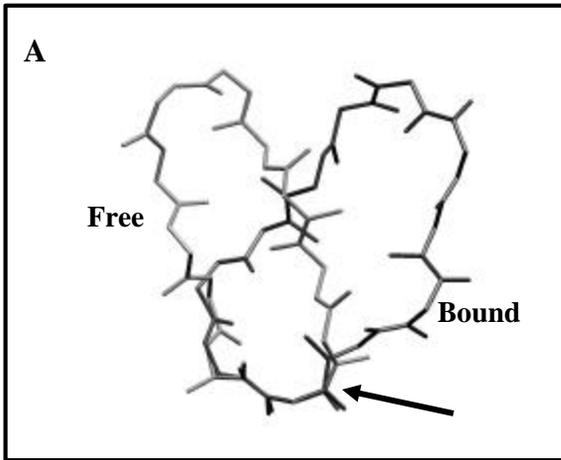


Figure 6.



*Figure 7.*

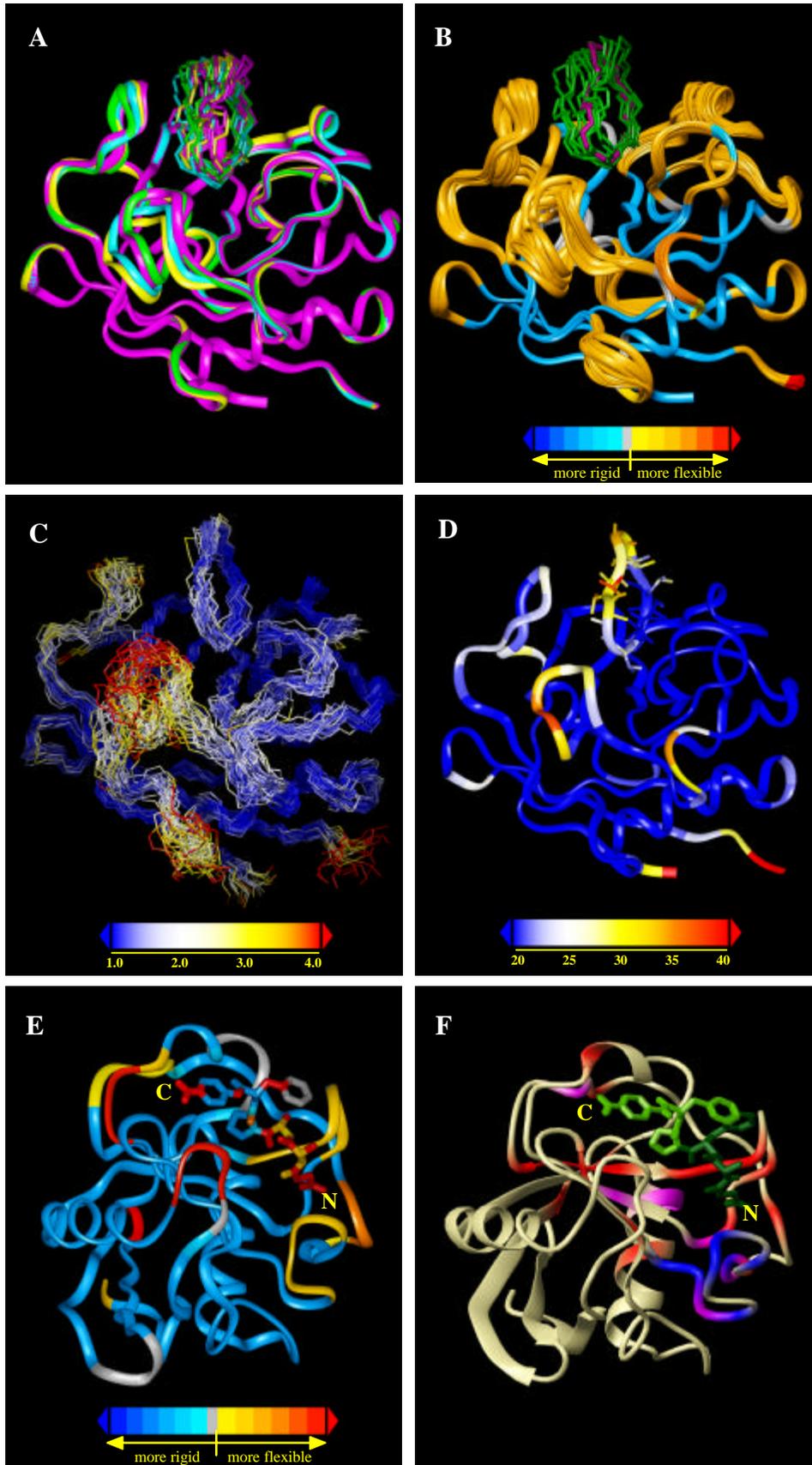


Figure 8.

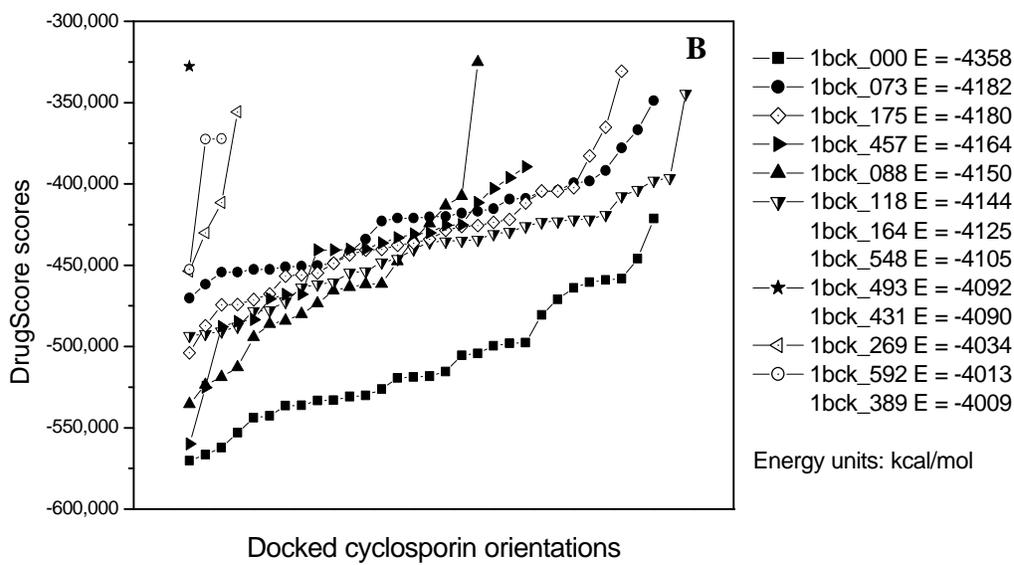
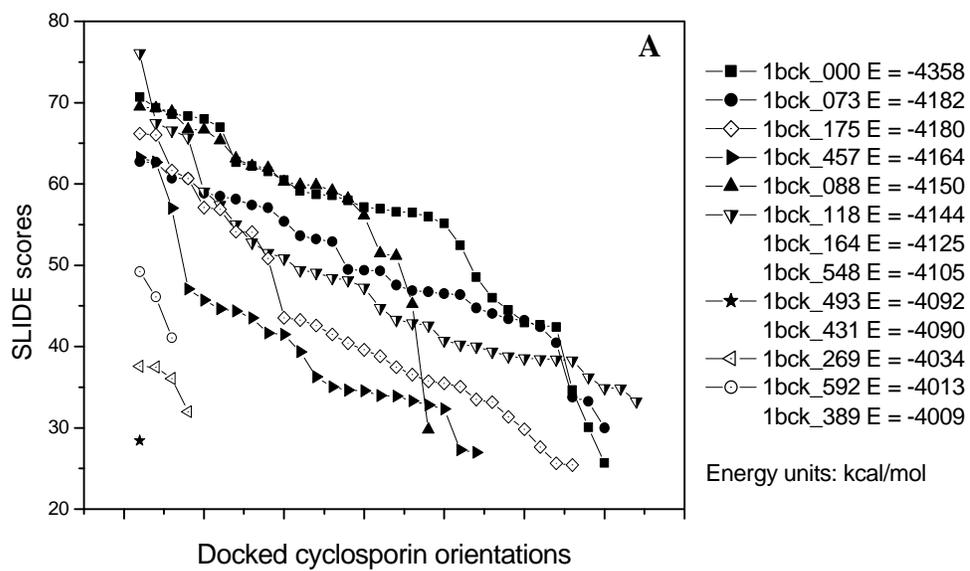
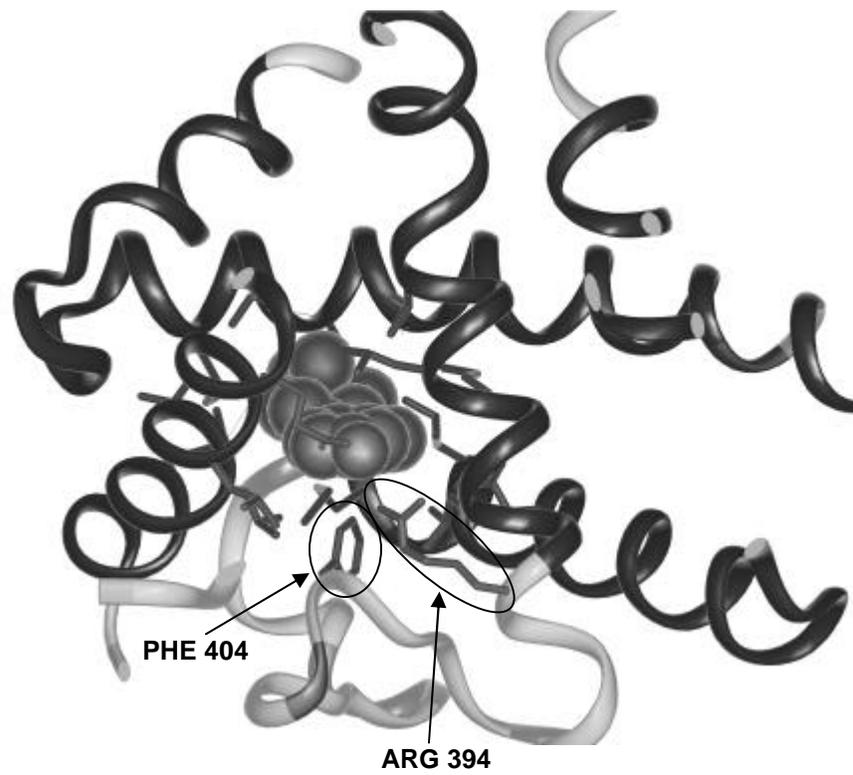


Figure 9.



*Figure 10.*

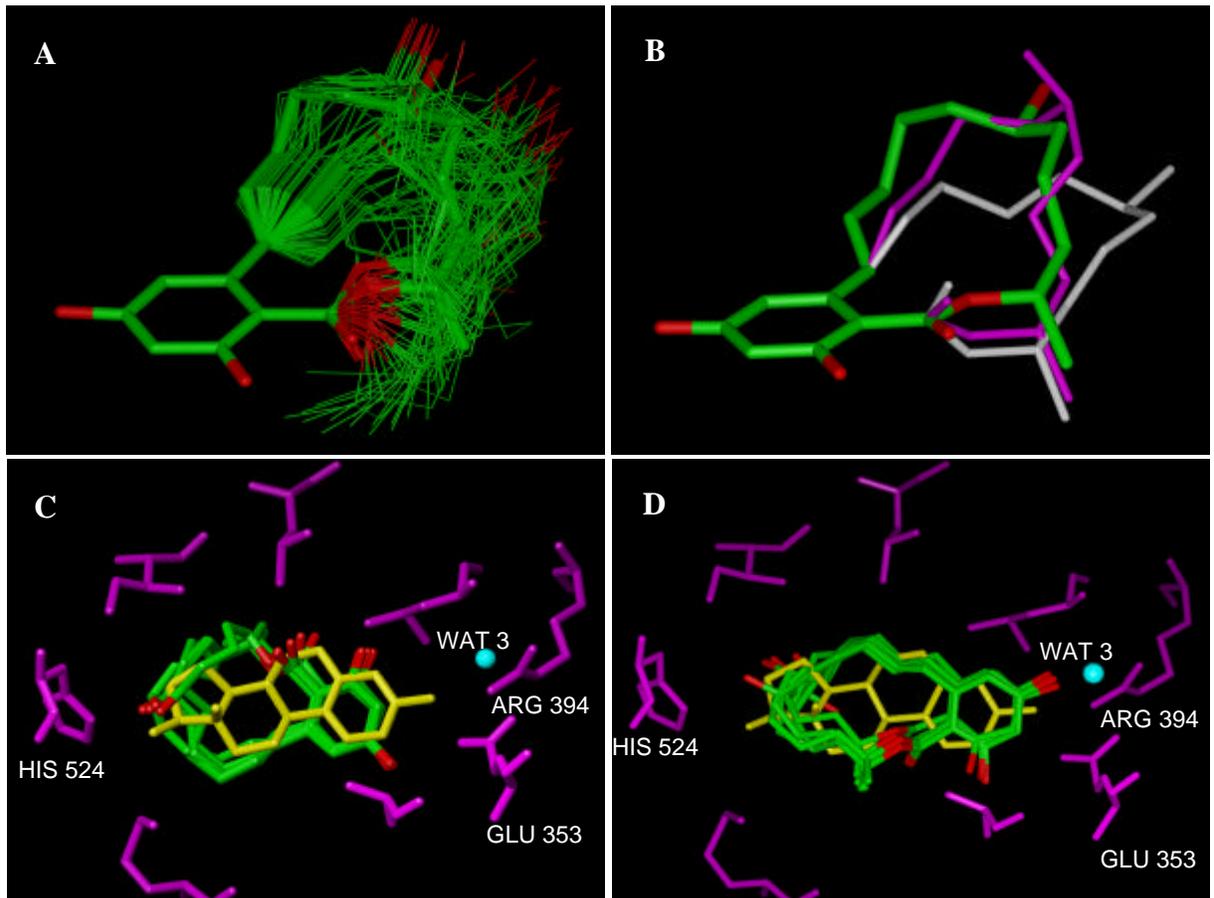


Figure 11.

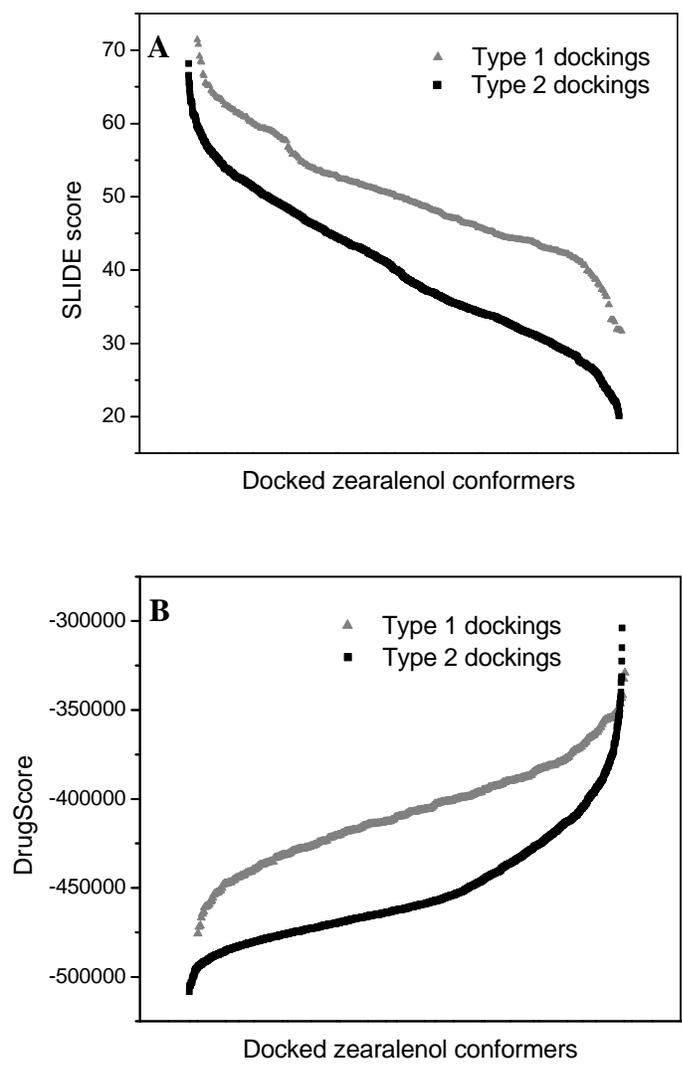


Figure 12.

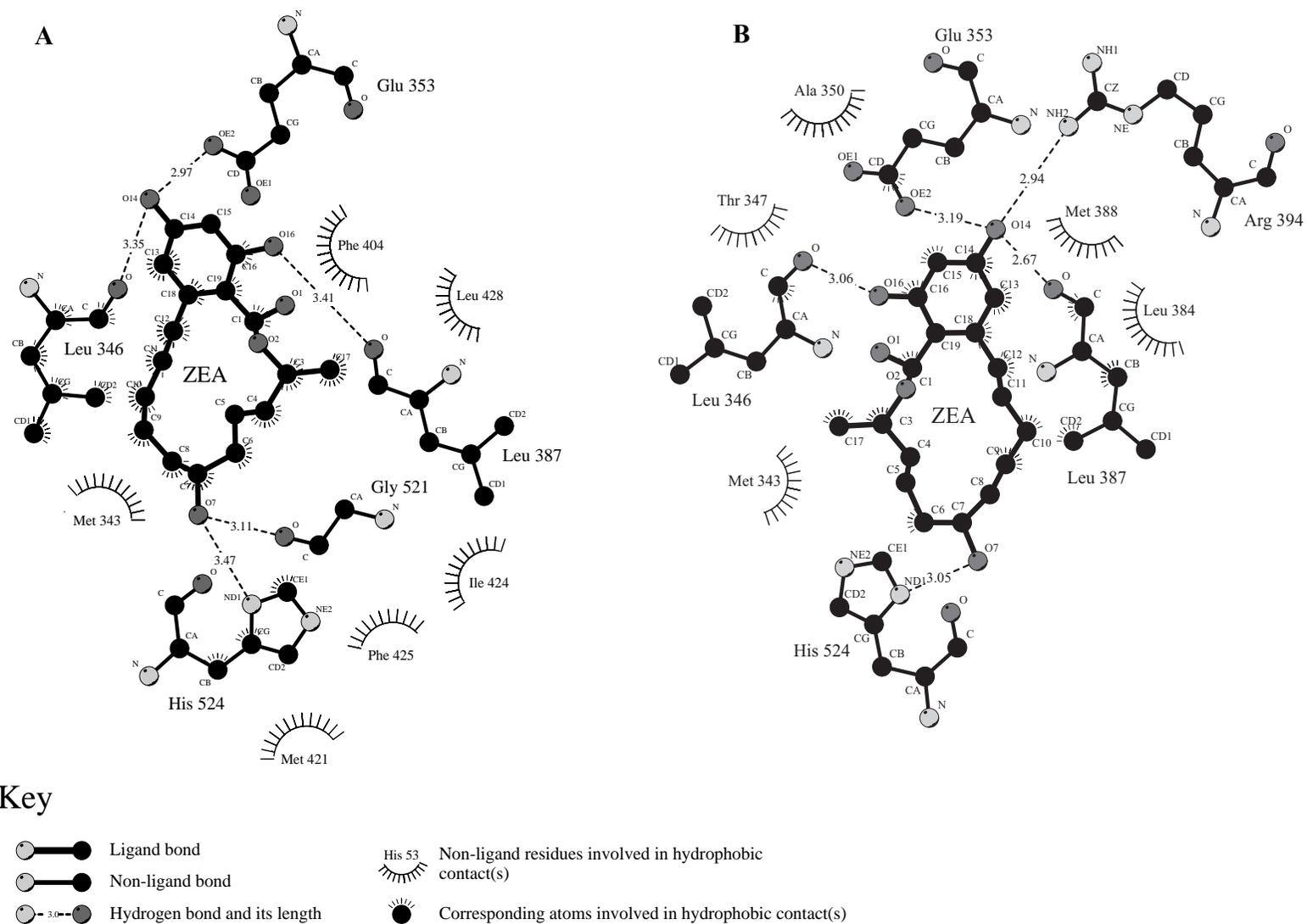


Figure 13.