DMD Fast Forward. Published on April 8, 2005 as doi:10.1124/dmd.105.004200

Possible Pathway(s) of Testosterone Egress from the Active Site of Cytochrome P450 2B1: A Steered Molecular Dynamics Simulation

Weihua Li, Hong Liu, Emily E. Scott, Frauke Gräter , James R. Halpert,

Xiaoming Luo, Jianhua Shen, and Hualiang Jiang

Center for Drug Discovery and Design, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Graduate School of the Chinese Academy of Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P. R. China (W.L., H.L., F.G., X.L., J.S., H.J.)

Department of Pharmacology and Toxicology, University of Texas Medical Branch, 301 University Blvd., Galveston, Texas 77555-1031(J.R.H)

Department of Medicinal Chemistry, University of Kansas, 1251 Wescoe Hall Dr., Lawrence, KS 66045-7582 (E.E.S.)

School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China (H.J.)

Running Title: SMD simulation of testosterone egress from P450 2B1

Corresponding author:

Dr. Hong Liu, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chong Zhi Road, Zhangjiang Hi-Tech Park, Shanghai 201203, China. Tel: +86-21-50806600 ext. 5416; Fax: +86-21-50807088; E-mail: hliu@mail.shcnc.ac.cn

Number of Text pages: 32

Number of Figures: 12

Number of References: 39

Number of words in *Abstract*: 228

Number of words in *Introduction*: 748

Number of words in *Discussion*: 1495

Abbreviations: P450s, cytochromes P450; SMD, steered molecular dynamics; 3D, three dimensional; MD, molecular dynamics; RMS, root-mean-square; S.D., standard deviation; REMD, random repulsion molecular dynamics.

Abstract

To probe possible substrate exit channel(s) in P450 2B1 and to clarify the role of residues previously identified by site-directed mutagenesis, a homology model was constructed based on the x-ray crystal structure of a P450 2B4-inhibitor complex. Testosterone was docked into the active site of P450 2B1 and was then pulled out through three putative channels using steered molecular dynamics simulations. The results indicated that of the three channels, the “solvent channel”, lined by helices E, F and I and the β3 hairpin, required the largest rupture force and backbone motion, which rendered it unlikely as an exit route. The relatively small rupture forces and backbone motions for the other two channels suggested them as possible candidates for testosterone passage. The opening of channel 1, located between helices G and I and the B'-C loop, is characterized by rotation of the aromatic ring of Phe297 together with a bending of the B'-C loop. The opening of channel 2, penetrating through the B'-C loop/ B' helix, is achieved by an expansion of this region and a small displacement of the backbone. Interestingly, during the egress of testosterone along channel 1, Phe297 and Phe108 appear to act as two clamps to stabilize testosterone binding and prevent it from leaving the active site. Phe115 acts as a gatekeeper for channel 2. These results are in agreement with previous site-directed mutagenesis experiments.

Introduction

An intriguing question about cytochromes P450 is how substrates enter and leave the active site, which is deeply buried in the center of the P450 fold (Wade et al., 2004). This issue is important because experimental data indicate that the regio- and stereospecificities of P450s may be influenced not only by residues in the active site but also by residues far from the binding site (Domanski and Halpert, 2001). A few structures of bacterial P450s suggest several potential routes for substrate access to and exit from the active site. The structures of substrate-free P450 102 (Ravichandran et al., 1993) and inhibitor-bound P450 101 (Raag et al., 1993) indicate a possible access channel located between the F-G loop and β1 sheet. This channel has also been suggested as a route for substrate/product egress in several other bacterial P450s (Wade et al., 2004). A very different channel has been described in the structure of P450 51 (Podust et al., 2001). This channel threads through the B-C loop and is almost perpendicular to the channel found in P450 102.

To date, the structures of five mammalian P450s, including rabbit 2C5 (Williams et al., 2000) and 2B4 (Scott et al., 2003) and human 2C8 (Schoch et al., 2004), 2C9 (Williams et al., 2003) and 3A4 (Williams et al., 2004; Yano et al., 2004), have been determined. Most of these structures adopt a closed conformation, in which no obvious channels are present for substrate entry or product egress. Thus, an intriguing problem arises: which parts open up to allow substrate/product passage in the closed form? Specific P450 regions have been shown to alter their conformation in response to ligand. The most dramatic differences in protein conformation are observed for 2B4. The substrate-free structure (Scott et al., 2003) reveals a large open cleft

that extends from the protein surface directly to the heme iron, while an inhibitor-bound structure (Scott et al., 2004b) adopts a closed conformation similar to that observed in the mammalian 2C enzymes. The differences between the open and closed structures of 2B4 are primarily limited to helices F through G, helices B' through C, the N-terminus of helix I, and the β4 region. The conformational change upon ligand binding implies that these specific flexible regions may be involved in substrate access or egress channels in the closed P450 form, but the residues involved and the mechanisms of channel opening and closing are unknown. An understanding of these questions is important to explain the broad substrate specificity and regio- and stereospecificities of P450s. In the present study, we use steered molecular dynamics (SMD) simulation to probe the possible substrate exit pathway(s) of P450 2B1. P450 2B1 has been chosen because of the wealth of site-directed mutagenesis information available.

Since the crystal structure of P450 2B1 is not available, the model used in the present simulations was constructed based on the structure of a ligand-bound P450 2B4 complex (Scott et al., 2004b). Testosterone, a typical substrate of P450 2B1, was docked into the active site of P450 2B1 in a reactive binding orientation, and then pulled out along three putative exit channels using SMD (Figure 1), chosen based on available P450 structures. Channel 1 is located between helices I and G and the B'-C loop and was selected based on the large cleft located in this region in P450 2B4 and an open channel found in P450 2C5 (Williams et al., 2000). Channel 2 is directed through helix B' and the B'-C loop and is consistent with an opening of this region in the unliganded 2B4 structure and the channels found in P450 51 (Podust et al., 2001), P450 152A1 (Lee et al., 2003), and P450 3A4 (Williams et al., 2004). Channel 3 corresponds to a “solvent

channel" between the active site and the protein surface found in several P450 crystal structures (Haines et al., 2001; Wester et al., 2003). This channel is lined by helices E, F and I and β3 hairpin, and has been suggested to be important for controlling proton access to the active site (Haines et al., 2001). However, it is not yet clear whether this solvent channel can permit the passage of large ligands. The channel between the F-G loop and β1 sheet in P450 101 and P450 102 was not investigated, because this region is occupied by two short helices F' and G' in the 2B1 model. Moreover, site-directed mutagenesis data from 2B1 do not support this region as a channel for substrate passage (Scott et al., 2002).

Methods

Modeling the Structure of P450 2B1. The sequence of rat P450 2B1 was obtained from the SwissProt database (accession number P00176). The three-dimensional (3D) model of P450 2B1 was constructed based on the crystal structure of P450 2B4 with an inhibitor bound (PDB entry 1SUO) (Scott et al., 2004b) using the InsightII software package (InsightII, 2000). The coordinates of the conserved residues were assigned based on the corresponding residues of P450 2B4 using the Homology module of InsightII. The heme coordinates were copied from P450 2B4 into the corresponding site of P450 2B1 model.

After the coordinate assignment, the initial model of P450 2B1 was refined using the Gromacs 3.1 software package (Lindahl et al., 2001) with an extended version of the GROMOS87 force field (van Gunsteren et al., 1996). Energy minimization was performed with 100 steps of steepest descent followed by 300 steps conjugate gradient to release conflicting contacts among residues. The protein was then solvated with water in a rectangular periodic box.

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

The simple point charge (SPC) model (Berendsen et al., 1981) was used to describe water molecules. Since the protein-water system has a total charge of –5e, the system was neutralized by Na^+ ions, which replaced water molecules at the positions of lowest Coulomb potential in the system. The solvent was relaxed by energy minimization while restraining the protein atomic positions with a harmonic potential. The system was then energy minimized without restraints for 2000 steps using a combination of steepest descent and conjugate gradient. Afterwards, molecular dynamics (MD) equilibration at constant temperature was performed in order to provide further structural relaxation. Equilibration was conducted at 300 K with decreasing harmonic restraints over a 15 ps time interval followed by 1 ns of equilibration without restraints. The model obtained was energy minimized again for 1000 steps using conjugate gradient method. The optimized structure was used as the model for subsequent substrate docking and SMD simulations. During the MD simulation, the LINCS algorithm (Hess et al., 1997) was used to constrain all bonds. A dielectric constant of 1, and a time step of 2 fs were used. The electrostatic interactions were calculated using the Particle-Mesh Ewald method (Essmann et al., 1995) with a 0.9 nm cutoff. The temperature was kept constant by coupling solute and solvent separately to a thermal bath at 300 K with a coupling constant τ_T = 0.1 ps. Pressure was kept constant by coupling to a pressure bath at 1.0 bar, using a coupling constant τ_P = 0.5 ps.

The overall quality of the final 3D model of P450 2B1 was assessed with respect to its geometry and energy. The Procheck (Laskowski et al., 1993) and Prostat modules of InsightII (InsightII, 2000) were employed for geometric evaluation. Prosa2003 (Sippl, 1993) was used to evaluate the quality of consistency between the native fold and the sequence and to examine the

energy of residue-residue interactions. The energy is transformed to a z-score by

$$z_{S,C} = (E_{S,C} - \hat{E}_{S,C})/\sigma_S, \quad (1)$$

where $E_{S,C}$ and $\hat{E}_{S,C}$ are respectively the residue-residue interaction energy and average residue-residue interaction energy of sequence S in conformation space C, and σ_S is the associated standard deviation.

Docking Simulation. The structure of testosterone was constructed using the Builder module of the Insight II package (InsightII, 2000). Placement of testosterone into the active site of the 3D model of P450 2B1 was accomplished by using the Affinity program encode in the Insight II (InsightII, 2000) as previously described (Scott et al., 2004a). Testosterone was automatically docked in a reactive orientation leading to 16α hydroxylation.

Steered Molecular Dynamics (SMD) simulation. Prior to the SMD simulations of the P450 2B1-testosterone complex, a 200 ps MD simulation was performed to evaluate the stability of testosterone in the reactive binding site of the 3D model, yielding an equilibrated starting structure for SMD simulation. The force field parameters for testosterone were generated using the Dundee Prodrug Server (van Aalten et al., 1996). Atomic charges of testosterone were calculated employing the ChelpG method, an electrostatic potential fitting approach (Breneman and Wiberg, 1990), at the HF/6-31G* level. This *ab initio* calculation was carried out using the Gaussian98 program (Frisch et al., 1998).

SMD is an extended molecular dynamics simulation method mimicking the principle of the atomic force microscopy (AFM) (Grubmuller et al., 1996). SMD has been widely used to explore the binding and unbinding properties of biomolecules and their responses to external

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

mechanical manipulations at the atomic level (Isralewitz et al., 2001). SMD has also been successfully applied to identify ligand pathways and to explore the elastic properties of several proteins (Isralewitz et al., 2001; Shen et al., 2003; Xu et al., 2003). In SMD simulation, time-dependent external forces are applied to a ligand in order to accelerate binding or unbinding processes. From the accelerated dissociation process of the ligand, the SMD simulation can reveal information about the protein's flexibility and its response to the dissociation of ligand. Analyses of interactions between the ligand and the protein and the relationship between the applied forces and the ligand pathway can yield important information about the structure-function relationships of the protein-ligand complex, the binding and unbinding pathway(s), and possible mechanisms of ligand recognition and inhibition.

In the current SMD simulations, testosterone was pulled out from the binding pocket through the putative exit channels (Figure 1) by an external force. The movement of testosterone along the channels was determined using the criterion of minimal collision with amino acid residues. The center of mass of testosterone was harmonically restrained to a point moving with a constant velocity along the desired directions. Steered molecular dynamics were performed using the Gromacs 3.1 software package (Lindahl et al., 2001) and an extended version of GROMOS87 force field (van Gunsteren et al., 1996). During the SMD simulations, the pulling velocity was set to 0.005 $nm \bullet ps^{-1}$ and the spring force constant was assigned as 500 $kJ \bullet mol^{-1} \bullet nm^{-2}$. The forces of testosterone with P450 2B1 were monitored throughout the SMD simulations and the interactions between the protein residues and testosterone were analyzed along the SMD trajectory. At each time point the presence of direct hydrogen bonds, water

bridges, and hydrophobic interactions between testosterone and 2B1 was analyzed using the LIGPLOT program (Wallace et al., 1995).

Results

3D model of P450 2B1. A Blastp search (Madden et al., 1996) confirmed that the sequence identity (79%) is the highest between 2B1 and 2B4 among all the known P450s structures in the Protein Data Bank. The high sequence identity between 2B1 and 2B4 ensures the accuracy of the modeled structure of 2B1 based on the crystal structure of 2B4. A comparison between the ligand-free and ligand-bound structures of 2B4 reveals that the latter adopts a closed conformation that is similar to 2C5 (Wester et al., 2003; Schoch et al., 2004). The substrate-bound structure of 2B1 is likely similar to the ligand-bound structure of 2B4. Accordingly, the crystal structure of a 4-(4-chlorophenyl)imidazole-2B4 complex (PDB entry 1SUO) (Scott et al., 2004b) was selected as the template to construct the 3D model for 2B1.

In the crystal structure of the P450 2B4 complex, the coordinates for the N-terminal residues 1-27 were missing. As the N-terminus does not affect substrate binding, the 3D model of P450 2B1 was constructed only for residues 28-491. Figure 2 displays the sequence alignment between P450 2B1 and P450 2B4. Because of the lack of gaps in the alignment, the entire structures were considered as conserved regions.

Total energy and heavy atom root-mean-square deviation (RMSD) from the energy-minimized model structure are two important criteria for the convergence of the free MD simulation. These properties are shown in Figure 3 for the 1 ns MD simulation of P450 2B1. The results indicate that the total energy is stabilized after 1 ns equilibration. The RMSD of the heavy

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

atoms from the energy-minimized model increased slowly in the first 600 ps time period, and then reached a plateau in the sequent simulation time. All these properties converged after 1 ns MD simulation, indicating that the model is stable and can be used for subsequent SMD simulations.

After 1ns MD simulation, the overall quality of the P450 2B1 homology model was judged from energetic and geometric criteria. Figure 4 shows the total Prosa energy in terms of the z-score (Eq. (1)) versus the protein residue position. A high z-score corresponds to stressed or strained sections of the chain and may point to problematic parts of a fold. The z-score in this model is negative at all residue positions, which indicates reasonable side chain interactions. The overall score of Procheck geometric assessment is –0.15 for the homology model. An overall score of –0.5 or greater is considered as an indicator for a high quality structure. The percentage of Φ-Ψ angles in core Ramchandran regions was 84.9% in the P450 2B1 model. The geometric assessment of P450 2B1 was also performed using Prostat/InsightII (InsightII, 2000). The cutoff used, which represents the significant difference for bond length, bond angle, and torsion from the reference value obtained from known protein crystal structures, is 5 S.D. (standard deviation). For the P450 2B1 model, none of the bond distances, none of the bond angles, and only three dihedral angles were found to have more than 5 S.D. Thus, the P450 2B1 model is of reasonable quality compared to the crystal structure of the P450 2B4 complex.

3D Model of Testosterone-P450 2B1 Complex. In order to investigate the exit pathways using the SMD simulation method, testosterone was docked into the active site of 2B1 in the 16α-hydroxylation orientation leading to the formation of the major product. Contacts between

protein and substrate are predominately hydrophobic. The side chains of Arg98, Gly99, Thr100, Ile101, Ile104, Ile114, Phe115, Phe206, Phe297, Ala298, Glu301, Thr302, Val363, Ile365, Val367, Pro368, and Ile477 lie within 5 Å of testosterone. Most of these residues, including Ile114, Phe115, Phe206, Phe297, Ala298, Thr302, Val363, Ile365, Val367, and Ile477, correspond to key residues responsible for regio- and stereoselectivities revealed by previous mutagenesis studies on family 2 enzymes (Domanski and Halpert, 2001). Additional residues, including Thr100, Ile101 and Ile104, are located in the first of six substrate recognition sites (SRS-1), as suggested by Gotoh (Gotoh, 1992) based on a comparison between family 2 enzymes and P450 101, and have been shown to influence the kinetic constants of some substrates in very recent site-directed mutagenesis experiments (Honma et al., 2005).

In order to evaluate the stability of testosterone in the active site of the P450 2B1 model, a 200 ps MD simulation was performed. Figure 5 shows the RMSD of testosterone and the distance between the iron atom and oxidation site C16 of testosterone as a function of simulation time. The RMSD fluctuates around 0.1 nm during the simulation, which indicates that the testosterone structure does not deviate significantly from the initial docked pose. The iron-C16 distance in the complex fluctuates around 0.45 nm, the typical length found in the crystal structures of other P450s for carbon oxidation. Thus, during the free MD simulation, the testosterone remained in the active site and maintained an orientation allowing C16 hydroxylation.

Egress of Testosterone from P450 2B1. The force profile of testosterone egress along channel 1 is shown in Figure 6A. During the first 230 ps a steady increase of the applied force was

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

observed. At the beginning of the simulation, the substrate-protein complex is stabilized by a water bridge to Glu301 and hydrophobic contacts with Ile104, Ile114, Phe297, Thr302, Val363, Ile365, Pro368 and Ile477, as shown in Figure 7A. At 230 ps testosterone forms a new hydrogen bond to Gly99 and hydrophobic interactions with Arg98, Ile101, Ile104, Ile114, Ala298, Val367, Pro368, and Ile477. These residues are located in the active site, demonstrating that the substrate has not left the vicinity of the heme. Breaking these hydrogen bond and hydrophobic interactions produces the highest peak in the force profile, corresponding to a rupture force of 860 pN. At 260 ps, testosterone has moved out of the active site and reached the entrance of channel 1, and a local minimum appears in the force profile. At this time, the direct hydrogen bond between substrate and Gly99 is broken, but testosterone forms a water bridge with Ser294 and several hydrophobic contacts with Ile104, Phe108, and Tyr111 in helix B' and the B'-C loop, residue Leu293 in helix I, and Leu238 and Leu242 in helix G, as shown in Figure 7B.

At 305 ps, the substrate forms two new water bridges with Gln239 and Ser294 and several hydrophobic contacts. Breaking these interactions produces the second peak of the force profile (Figure 6A). At 350 ps, the substrate forms two water bridges to Glu286 and Ser294 and hydrophobic contacts with residues Phe108, Tyr111, Leu242, Met289, and Leu293. At this time, the substrate adjusts its orientation to facilitate egress. At 450 ps, the force reaches the third peak and the substrate forms a hydrophobic interaction with residue Ile290. At 480 ps, the force falls to a local minimum. At this time, testosterone forms hydrophobic interactions with His285 and Met289 and a water bridge to Glu286, as shown in Figure 7C. Afterwards, testosterone exits channel 1 and moves into solvent. The force fluctuation after 650 ps probably is attributed to the

interaction of testosterone with solvent molecules. From 260 to 480 ps, the force profile is relatively flat, indicating that once testosterone overcomes the first peak at 230ps, it can exit the binding pocket smoothly through channel 1.

Figure 6B shows the force profile of testosterone exiting the P450 2B1 active site via channel 2. Initially, the exit of testosterone proceeded slowly, since the substrate remained tightly bound while the applied force increased steadily. At 240 ps of the simulation time, testosterone forms water bridges with Arg98, Gly99, Phe115, Asn117, Ser294, Glu301, and Arg370 and hydrophobic interactions with Ile114, Ile362, Val367, Pro368, and Ile477. Breaking these interactions produces the highest force peak corresponding to a rupture force of 830 pN. At 275 ps, the substrate forms two water bridges with Asn117 and Arg370 and six hydrophobic contacts with residues Gly99, Thr100, Ile101, Ala102, Glu105, and Phe115, as shown in Figure 7E. All of these residues except Arg370 are located in helix B' or the B'-C loop. These interactions demonstrate that the substrate has moved out of the active site pocket and reaches the entrance of channel 2. At 310 ps, the force decreases down to a local minimum, there is no water bridge between testosterone and 2B1, and the substrate only forms a hydrophobic interaction with Gly99. At 395 ps, a moiety of the substrate has penetrated the B'-C loop and testosterone adjusts its orientation to interact with other residues, which results in a small increase in the rupture force. A force of 400 pN is necessary to overcome the hydrophobic interactions of the substrate with residues Ile104, Glu105, Phe108, and Phe115 (Figure 7F). After 500 ps, the force has reached a global valley and the substrate has been completely pulled out of the protein.

Figure 6C shows the force profile of testosterone leaving the P450 2B1 active site via

channel 3. The rupture force for the substrate leaving this channel is up to 1400 pN, which is more than 1.5 times higher than channels 1 and 2. The highest force peak is produced at 400 ps. At this point, testosterone forms two hydrogen bonds to Lys479 and Ile480 and hydrophobic interactions with Phe206, Leu209, Phe297, Glu301, Pro364, Ile477, and Gly478. Several residues, including Leu209, Gly478, Lys479, and Ile480, are located at the entrance of channel 3, indicating that at this time testosterone has moved out of the active site and is penetrating through the channel. During the exit of testosterone via this channel, a hydrogen bond between $O^{\varepsilon 2}$ in Glu301 and N in Lys479 is noted, as shown in Figure 8. During the first 250 ps, the hydrogen bond between residue Glu301 and residues Lys479 is relatively stable. From 250 ps to 400 ps, the hydrogen bond is formed intermittently. After 400 ps, the hydrogen bonding is entirely broken, which enlarges the space enough to allow testosterone to pass through. Therefore, the hydrogen bonding between Glu301 and Lys479 may play an important role in preventing substrates from passing through this channel.

In order to compare the protein motions for testosterone exit via these three putative channels, the non-hydrogen atom RMS deviations (RMSD) and Cα RMS fluctuations (RMSF) from the initial structure were monitored during SMD simulations. The global RMSD after testosterone expulsion from channel 1, 2 and 3 fluctuated around 0.21 nm, 0.19 nm, and 0.18 nm, respectively, as shown in Figure 9. Figure 10 shows the Cα RMSF from the initial structure during SMD simulations along these three putative channels. The maximal Cα RMSF for testosterone exit via channel 3 is 0.27 nm, which is contributed by the flexible β3 hairpin region (the third shaded region in Figure 10C). The maximal Cα RMSF for testosterone exit via channel

1 and channel 2 is 0.22 nm and 0.21 nm, respectively, which is contributed by the H-I loop region. Since this loop is completely exposed to the surface of protein and thus freely mobile, it is not surprising that the region has a large RMSF for all the putative channels. In addition, the B' helix/ B'-C loop region displays a relatively large RMSF in testosterone exit via channel 1 and channel 2 (the first shaded regions in Figure 10A and 10B), which indicates that testosterone exit via both channel 1 and channel 2 involves the flexibility of this region.

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

Discussion

In this study, the possible substrate exit channel(s) in P450 2B1 and the interactions between testosterone and protein residues were investigated. A 3D model of P450 2B1 was first constructed, followed by docking testosterone into the active site in a reactive binding orientation. Afterwards, testosterone was pulled out of the binding pocket using SMD simulations. The SMD simulations show that of the three possible exit channels, channel 3 is unlikely to serve as the exit channel. This is supported by the following observations: 1) the maximum rupture force via channel 3 was 1.5 times higher than that via channel 1 and 2; 2) global RMSD and Cα RMSF via channel 3 is higher than those via channel 1 and channel 2; 3) the hydrogen bond formed between Glu301 and Lys479, which stabilizes the protein structure, has to be broken in order for testosterone to exit through channel 3. Since our simulations yield a similar force profile and rupture force for exit along channel 1 and channel 2, our results do not allow discrimination between these two channels for substrate egress. However, opening mechanisms of these two channels differ, as discussed below.

Although there is no obvious opening between helices I and G and loop B'-C in the closed, inhibitor-bound 2B4 structure, the ligand-free 2B4 structure (Scott et al., 2003) reveals a large cleft in this region that extends directly from protein surface to the heme iron. Site-directed mutagenesis revealed that substitution of residues in the N-terminus of helix I can alter the activity and stereoselectivity of P450 2B1 (Scott et al., 2004a). Moreover, REMD simulations have indicated a similar substrate egress pathway in P450 cam (Ludemann et al., 2000). In the current SMD simulation the secondary structures of helices I and G are well maintained, and the

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

rupture force is relatively small during testosterone exit along channel 1. All these findings indicate that a channel may exist between helices I and G and the B'-C loop. The simulation results suggest some residues that may play important roles in the egress of testosterone along channel 1. After entering channel 1, testosterone forms a water bridge to Ser294, which persisted until 450 ps. During this period, the hydrogen-bonding network formed between Ser294 and testosterone discourages the egress of testosterone. It can be speculated that abolishing the hydrogen-bonding network will alter the enzyme activity. This is consistent with the recent mutagenesis data, indicating that S294A mutant exhibited a 3-fold decrease in K_{cat}/K_m for the 16α product (Scott et al., 2004a). A similar function is attributed to Glu286, which also formed a water bridge to testosterone. The replacement of Glu286 by Ala led to a 2-fold decrease in K_{cat}/K_m for the 16α product (Scott et al., 2004a).

An analysis of the trajectory of SMD along channel 1 reveals that two residues located at the entrance of the putative channel 1, Phe108 and Phe297, appear to act as two flexible clamps during substrate egress. The hydrophobic interactions Phe108 and Phe297 with testosterone not only stabilize substrate binding and guide the orientation of the substrates, but also prevent the substrate from leaving the active site. The B'-C loop must rearrange to open wide enough for testosterone passage, and at the same time, the side chain torsion of Phe297 rotates to enlarge the space, as shown in Figure 11. To quantify this opening and rotation, the distance between Phe297 and Phe108 and the side-chain torsions χ of Phe297 have been monitored, as shown in Figure 12. In the first 230 ps, the distance remains around 0.9 nm. In the following 50 ps, displacement of the B'-C loop occurred, and the distance increased significantly to 1.4 nm in order for

testosterone to pass through. At the same time, the torsion of Phe297 increased from 22º to a maximal 120º. After testosterone crossed over the bottleneck, the χ torsion returned to its initial value.

The simulation result implies that the substitution of Phe297 and Phe108 with other residues may affect the orientation of the substrate in the active site, and thereby reduce substrate binding. This simulation result is in agreement with the mutagenesis result that F297A, F297I, and F297 exhibited a 5- to 10-fold decrease in catalytic activity (Domanski et al., 2001). Mutagenesis data on Phe108 show that substitution of this residue can also alter the kinetic constants of 2B1(Honma et al., 2005). The SMD simulation result together with the site-directed mutagenesis results support the inference that channel 1 is a possible route for substrate egress in 2B1.

The structure of P450 51 (Podust et al., 2001) reveals an access channel from protein surface to the heme, which penetrates through the open B-C loop. In P450 152A1 (Lee et al., 2003), a similar channel is proposed to permit water to escape from the active site and to allow access of hydrogen peroxide to substrate-bound enzyme. An opening between helices B' and C is also apparent in the ligand-free P450 2B4 structure (Scott et al., 2003) and 3A4 structures (Williams et al., 2004; Yano et al., 2004). Based on these observations, this region was selected as a possible exit channel in the current study. The SMD simulation results show that a relatively small rupture force and only slight backbone motions are required for testosterone exit along channel 2, as shown in Figure 6B and Figure 11B. Our results in conjunction with findings from crystal structures of several mammalian P450s suggest that channel 2 may serve as a common channel for ligand passage in some bacterial and mammalian P450s.

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

A detailed analysis of the interactions between testosterone and the protein revealed that Phe115 acts as a gatekeeper, not only playing an important role in stabilizing the orientation of testosterone in the active site, but also preventing ligands from escaping from the active site. The corresponding Phe114 in 2C9 was found to have a strong hydrophobic contact with the phenyl group of *S*-warfarin and thereby to stabilize its orientation in the 2C9 enzyme (Williams et al., 2003). An analysis of snapshots of the exit process through channel 2 shows that after testosterone leaves the active site, the phenyl ring of Phe115 flips to occupy the space left by testosterone and consequently closes the channel. This implies that a mutation of Phe115 to other residues might change the orientation of substrate in the active site and influence substrate egress from the active site. This suggestion is in agreement with the site-directed mutagenesis data, which indicate that F115A exhibits a 3-fold decrease in the testosterone hydroxylase activity (Domanski et al., 2001).

Although the opening of both channels is due to the flexibility of helix region B'-C loop/ B', the opening mechanism of channel 2 is different from that of channel 1. The opening of channel 1 is characterized by a rotation of Phe297 in conjunction with the B'-C loop rearrangement, involving a relatively large displacement of the B'-C loop (the first shadow region in Figure 10A and Figure 11A). In contrast, testosterone exit through channel 2 is achieved by the expansion of B'-C loop (Figure 11B) together with a relatively small displacement of the backbone (Figure 10B). It has been noted that in family 2 P450s two highly conserved GlyXGly motifs flank the region between helix B' and the B'-C loop. Glycines are known for their torsional flexibility due to the lack of a side chain, and the peptide backbone can readily adopt a wide range of

conformations. The GlyXGly motif thus can be assumed to constitute the hinge for the B'-C loop conformational changes.

The solvent channel, our channel 3, has been suggested to play an important role in controlling proton access to active site (Haines et al., 2000; Wester et al., 2003). However, no evidence at present directly indicates a ligand pathway. Our simulation shows that substrate egress via this route requires large rupture forces, large backbone motion, and the rupture of a hydrogen bonding network involving Glu301. This indicates a low probability of channel 3 to serve as an egress channel for testosterone in P450 2B1. In light of apparent opening of this channel in the structures of 2C5, 102, and inhibitor-bound 2B4, and the smaller ligand size in the 2B4 complex, access/egress through channel 3 cannot be ruled out. This remains to be tested by additional experimental and computational studies.

In conclusion, we used SMD simulations to investigate three possible ligand channels of P450 2B1 and to clarify the role of some residues previously identified by site-directed mutagenesis experiments. Our results demonstrate that of the three possible channels, the “solvent channel” is unlikely to serve as exit channel, whereas both other channels are equally good candidates for the egress channels. However, opening mechanisms of these two channels differ. The opening of channel 1 is characterized by a rotation of Phe297 in conjunction with a relatively large displacement of the B'-C loop. In contrast, testosterone exit through channel 2 is achieved by the expansion of B'-C loop together with a relatively small displacement of the backbone.

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

Acknowledgments

We thank the molecular dynamics group at Department of Biophysics Chemistry of Groningen University for their kindness in offering us the Gromacs program (www.gromacs.org).

References

Berendsen HJC, Postma JPM, van Gunsteren WF and Hermans J (1981) Interaction models for water in relation to protein hydration. In: Intermolecular Forces. (Pullman,B. ed.) pp331-342, D.Reidel Publishing Company Dordrecht.

Breneman CM and Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J.Comp.Chem.* **11:**361-397.

Delano WL (2004) *"The PyMOL Molecular Graphics System". Delano Scientific LLC, San Carlos, CA, USA.* http://www.pymol.org.

Domanski TL and Halpert JR (2001) Analysis of mammalian cytochrome P450 structure and function by site-directed mutagenesis. *Curr Drug Metab* **2:**117-137.

Domanski TL, He YQ, Scott EE, Wang Q and Halpert JR (2001) The role of cytochrome 2B1 substrate recognition site residues 115, 294, 297, 298, and 362 in the oxidation of steroids and 7-alkoxycoumarins. *Arch Biochem Biophys* **394:**21-28.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H and Pedersen LG (1995) A smooth particle mesh ewald potential. *J.Chem.Phys.* **103:**8577-8592.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Cli_ord S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Stefanov

BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES and Pople JA (1998) Gaussian98,Revision A.7. *Gaussian,Inc.,Pittsburgh,PA.*

Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267:**83-90.

Grubmuller H, Heymann B and Tavan P (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271:**997-999.

Haines DC, Sevrioukova IF and Peterson JA (2000) The FMN-binding domain of cytochrome P450BM-3: resolution, reconstitution, and flavin analogue substitution. *Biochemistry* **39:**9419-9429.

Haines DC, Tomchick DR, Machius M and Peterson JA (2001) Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* **40:**13456-13465.

Hess B, Bekker H, Fraaije J and Berendsen HJ (1997) A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18:**1463-1472.

Honma W, Li W, Liu H, Scott EE and and Halpert JR (2005) Functional role of residues in the helix B' region of cytochrome P450 2B1. *Arch Biochem Biophys* **435:**157-165.

InsightII V, Accerlrys Molecular Simulations Inc., San Diego,CA (2000).

Israleewitz B, Gao M and Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* **11:**224-230.

Laskowski RA, MacArthur MW, Moss DS and Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26:**283-291.

Lee DS, Yamada A, Sugimoto H, Matsunaga I, Ogura H, Ichihara K, Adachi S, Park SY and Shiro Y (2003) Substrate recognition and molecular mechanism of fatty acid hydroxylation by cytochrome P450 from Bacillus subtilis. Crystallographic, spectroscopic, and mutational studies. *J Biol Chem* **278:**9761-9767.

Lindahl E, Hess B and van der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **7:**306-317.

Ludemann SK, Lounnas V and Wade RC (2000) How do substrates enter and products exit the buried active site of cytochrome P450cam? 2. Steered molecular dynamics and adiabatic mapping of substrate pathways. *J Mol Biol* **303:**813-830.

Madden TL, Tatusov RL and Zhang J (1996) Applications of network BLAST server. *Methods Enzymol* **266:**131-141.

Podust LM, Poulos TL and Waterman MR (2001) Crystal structure of cytochrome P450 14alpha -sterol demethylase (CYP51) from Mycobacterium tuberculosis in complex with azole inhibitors. *Proc Natl Acad Sci U S A* **98:**3068-3073.

Raag R, Li H, Jones BC and Poulos TL (1993) Inhibitor-induced conformational change in cytochrome P-450CAM. *Biochemistry* **32:**4571-4578.

Ravichandran KG, Boddupalli SS, Hasermann CA, Peterson JA and Deisenhofer J (1993) Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science* **261:**731-736.

Schoch GA, Yano JK, Wester MR, Griffin KJ, Stout CD and Johnson EF (2004) Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J Biol Chem* **279:**9497-9503.

Scott EE, He YA, Wester MR, White MA, Chin CC, Halpert JR, Johnson EF and Stout CD (2003) An open conformation of mammalian cytochrome P450 2B4 at 1.6- Å resolution. *Proc Natl Acad Sci U S A* **100:**13196-13201.

Scott EE, He YQ and Halpert JR (2002) Substrate routes to the buried active site may vary among cytochromes P450: mutagenesis of the F-G region in P450 2B1. *Chem Res Toxicol* **15:**1407-1413.

Scott EE, Liu H, Qun He Y, Li W and Halpert JR (2004a) Mutagenesis and molecular dynamics suggest structural and functional roles for residues in the N-terminal portion of the cytochrome P450 2B1 I helix. *Arch Biochem Biophys* **423:**266-276.

Scott EE, White MA, He YA, Johnson EF, Stout CD and Halpert JR (2004b) Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-Chlorophenyl)imidazole at 1.9-Å resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J Biol Chem* **279:**27294-27301.

Shen L, Shen J, Luo X, Cheng F, Xu Y, Chen K, Arnold E, Ding J and Jiang H (2003) Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. *Biophys J* **84:**3547-3563.

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* **17:**355-362.

van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW and Vriend G (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J Comput Aided Mol Des* **10:**255-262.

van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Kuger P, Mark AE, Scott WRP and Tironi IG (1996) The GROMOS96 Manual and User Guide, Biomos, Zurich, Switzerland.

Wade RC, Winn PJ, Schlichting I and Sudarko (2004) A survey of active site access channels in cytochromes P450. *J Inorg Biochem* **98:**1175-1182.

Wallace AC, Laskowski RA and Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8:**127-134.

Wester MR, Johnson EF, Marques-Soares C, Dansette PM, Mansuy D and Stout CD (2003) Structure of a substrate complex of mammalian cytochrome P450 2C5 at 2.3 Å resolution: evidence for multiple substrate binding modes. *Biochemistry* **42:**6370-6379.

Williams PA, Cosme J, Sridhar V, Johnson EF and McRee DE (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol Cell* **5:**121-131.

Williams PA, Cosme J, Vinkovic DM, Ward A, Angove HC, Day PJ, Vonrhein C, Tickle IJ and Jhoti H (2004) Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **305:**683-686.

Williams PA, Cosme J, Ward A, Angove HC, Matak Vinkovic D and Jhoti H (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* **424:**464-468.

Xu Y, Shen J, Luo X, Silman I, Sussman JL, Chen K and Jiang H (2003) How Does Huperzine A Enter and Leave the Binding Gorge of Acetylcholinesterase? Steered Molecular Dynamics Simulations. *J Am Chem Soc* **125:**11340-11349.

Yano JK, Wester MR, Schoch GA, Griffin KJ, Stout CD and Johnson EF (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J Biol Chem* **279:**38091-38094.

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

Footnotes

A) Financial Support

The authors gratefully acknowledge financial support from the State Key Program of Basic Research of China (Grant 2002CB512802), the National Natural Science Foundation of China (Grants 20372069, 29725203, 20472094 and 20102007), the Basic Research Project for Talent Research Group from the Shanghai Science and Technology Commission, the Key Project from the Shanghai Science and Technology Commission (Grant 02DJ14006), the Key Project for New Drug Research from CAS, the Qi Ming Xing Foundation of Shanghai Ministry of Science and Technology (Grant 03QD14065),and the 863 Hi-Tech Programm (Grants 2002AA233061, 2002AA104270, 2002AA233011, and 2003AA235030) These studies were also supported by grant ES03619 and Center Grant ES06676 from the National Institutes of Health of the United States.

B) Name and Address for reprint Requests to:

James R. Halpert, Ph.D., Department of Pharmacology and Toxicology, The University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555-1031. E-mail: jhalpert@utmb.edu

Hualiang Jiang, Ph. D., Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chong Zhi Road, Zhangjiang Hi-Tech Park, Shanghai 201203. Phone: +86-21-50806600 ext. 1210; Fax: +86-21-50807088; E-mail: hljiang@mail.shcnc.ac.cn

FIGURE LEGENDS

Figure 1. Ribbon schematic representations of a homology model of P450 2B1 and the chosen channels examined by SMD. The heme is represented by a *red stick*; testosterone is represented by a *salmon CPK* model. Major helices and sheets are labeled. During the SMD simulation, the bound testosterone is pulled away from the active site using a harmonic potential symbolized by an artificial spring that is connected to the center of mass of testosterone. This pulling potential moves with a constant velocity of 0.005 $nm \bullet ps^{-1}$ in the direction indicated by the arrow. Images generated using Pymol (www.pymol.org) (Delano, 2004) unless otherwise noted.

Figure 2. The sequence alignment between P450 2B1 and P450 2B4 (PDB entry 1SUO) from residues 28 to 491. The asterisk (*) indicates an identical or conserved residue; a colon (:) indicates a conserved substitution; a dot (•) indicates a semi-conserved substitution. *Boxes* and *underlines* represent helices and β sheets, respectively.

Figure 3. Total energy and RMSD with respect to simulation time for 1 ns free molecular dynamics simulation on the P450 2B1 model.

Figure 4. z-score versus protein residues position for the P450 2B1 model. The model is from a 1

ns molecular dynamics simulation followed by energy minimization for 1000 steps using conjugate gradient method.

Figure 5. (A) Distance between iron and C16 with respect to simulation time during 200 ps MD simulation on a testosterone-2B1 complex prior to SMD simulations. (B) RMSD of testosterone in the active site of 2B1 during 200 ps MD simulation on a testosterone-2B1 complex prior to SMD simulations.

Figure 6. Force profiles for SMD simulations along three different putative channels. (A) Channel 1; (B) Channel 2; (C) Channel 3.

Figure 7. Snapshots of the relative positions of testosterone and the protein throughout the simulation along channel 1((A) 0 ps, (B) 260 ps, and (C) 480 ps) and Channel 2 ((D) 0 ps, (E) 275 ps, and (F) 395 ps). The residues within 5 Å from testosterone at the individual time are shown in *green sticks*. The heme is shown in *red stick*. Testosterone atoms are represented as a *yellow stick*.

Figure 8. Distance between $O^{\varepsilon 2}$ in Glu301 and N in Lys479 with respect to simulation time during the SMD simulation along channel 3.

Figure 9. RMSD of non-hydrogen atom with respect to simulation time for testosterone exit

along three putative channels during SMD simulations. *Dotted* line, *thin* line, and *bold* line represents testosterone egress along channel 1, channel 2, and channel 3, respectively.

Figure 10. Cα RMSF with respect to residues position during SMD simulations along channel 1 (A), channel 2 (B), and channel 3 (C). The shadow region from left to right represents the B' helix/ B'-C loop region, H-I loop, and β3 hairpin region.

Figure 11. (A) Superposition of the starting structure (*yellow*) and the structure at 260 ps (*green*) during the SMD simulation along channel 1. Helix I and the B'-C loop are shown as *cartoons*, F297 and F108 as *sticks*. (B) Superposition of the starting structure (*yellow*) and the structure at 280 ps (*green*) during the SMD simulation along channel 2. Helix I and the B'-C loop are shown as *cartoons*, F115 and T100 as *sticks*.

Figure 12. (A) The variation of distance between residue F197 and F108 versus simulation time. (B) The variation of the torsion χ of F297 versus simulation time.

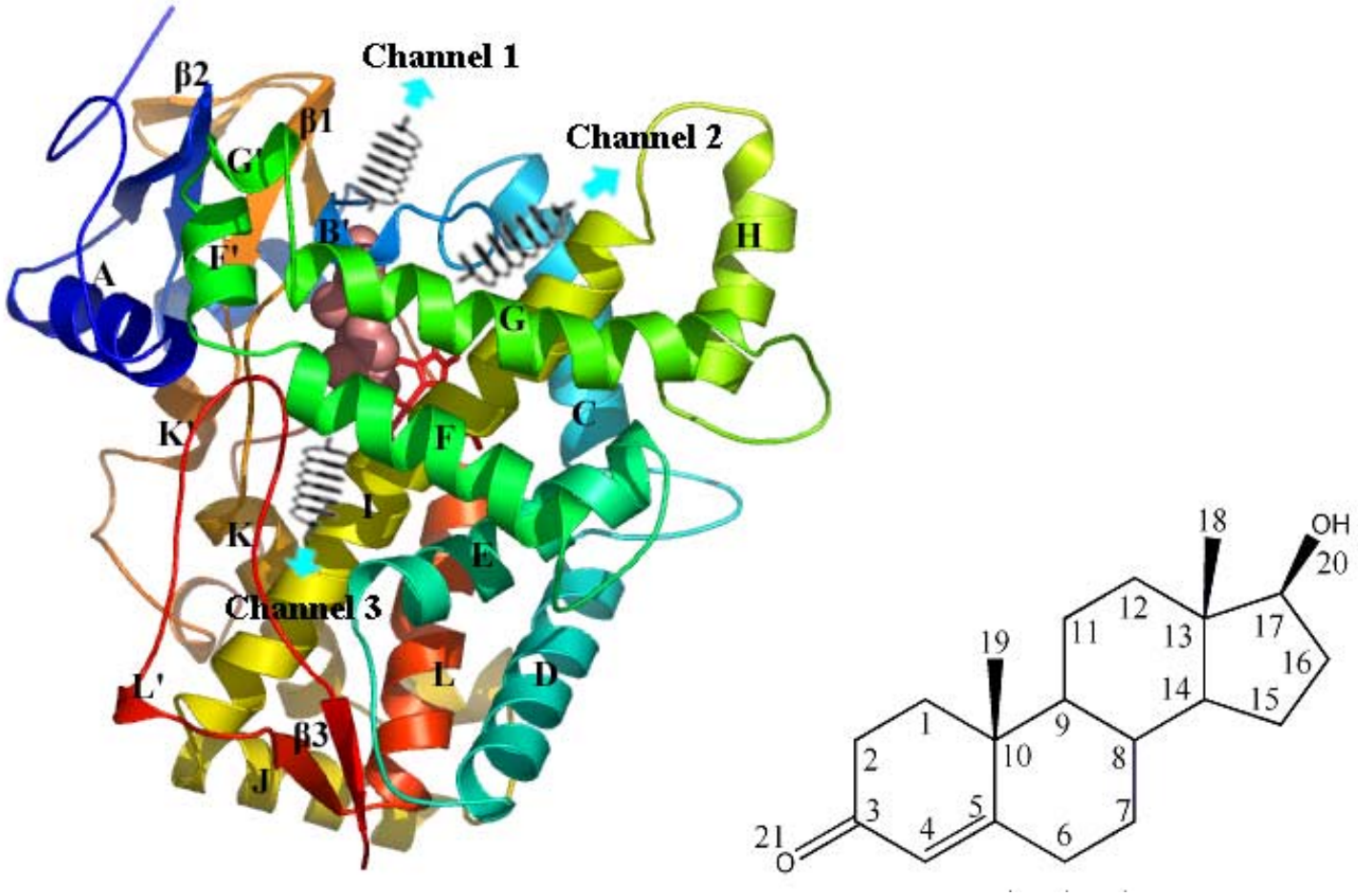

Figure 1

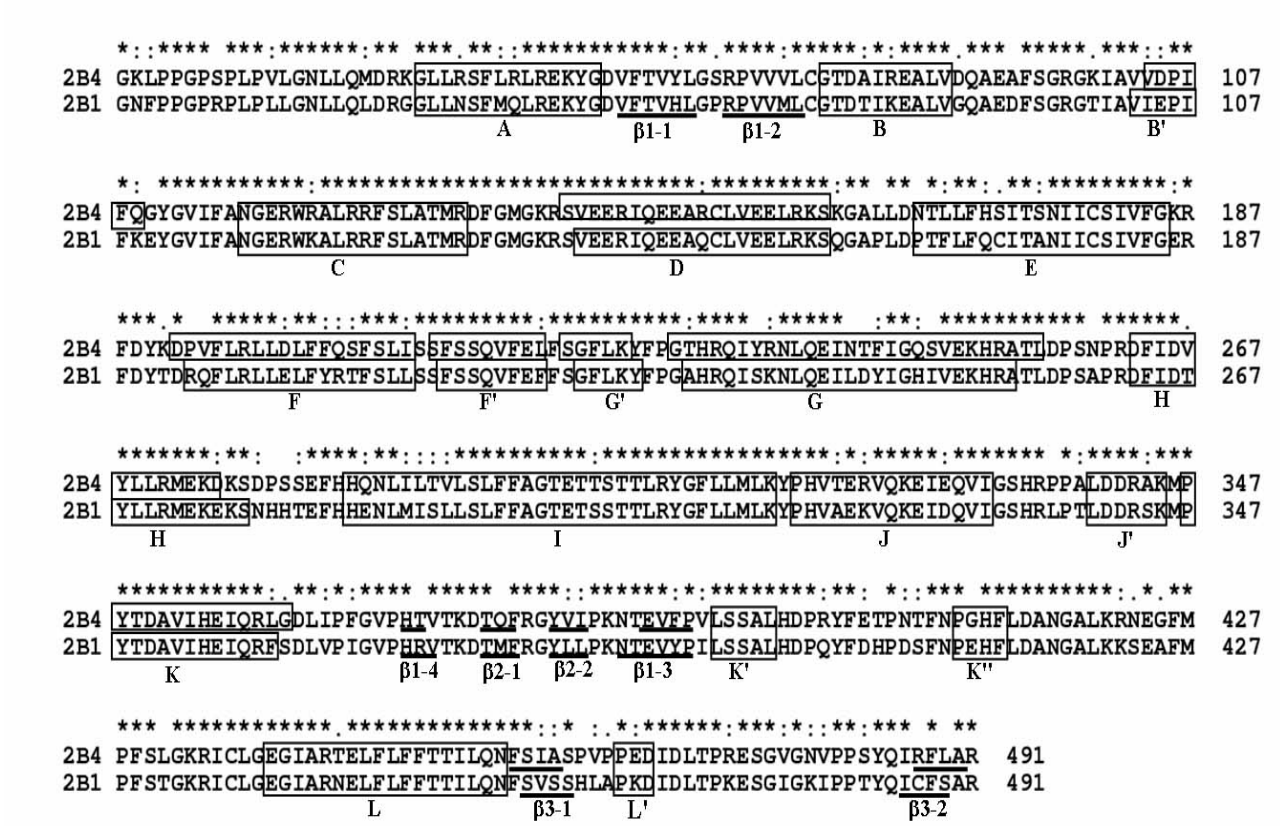

Figure 2

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200

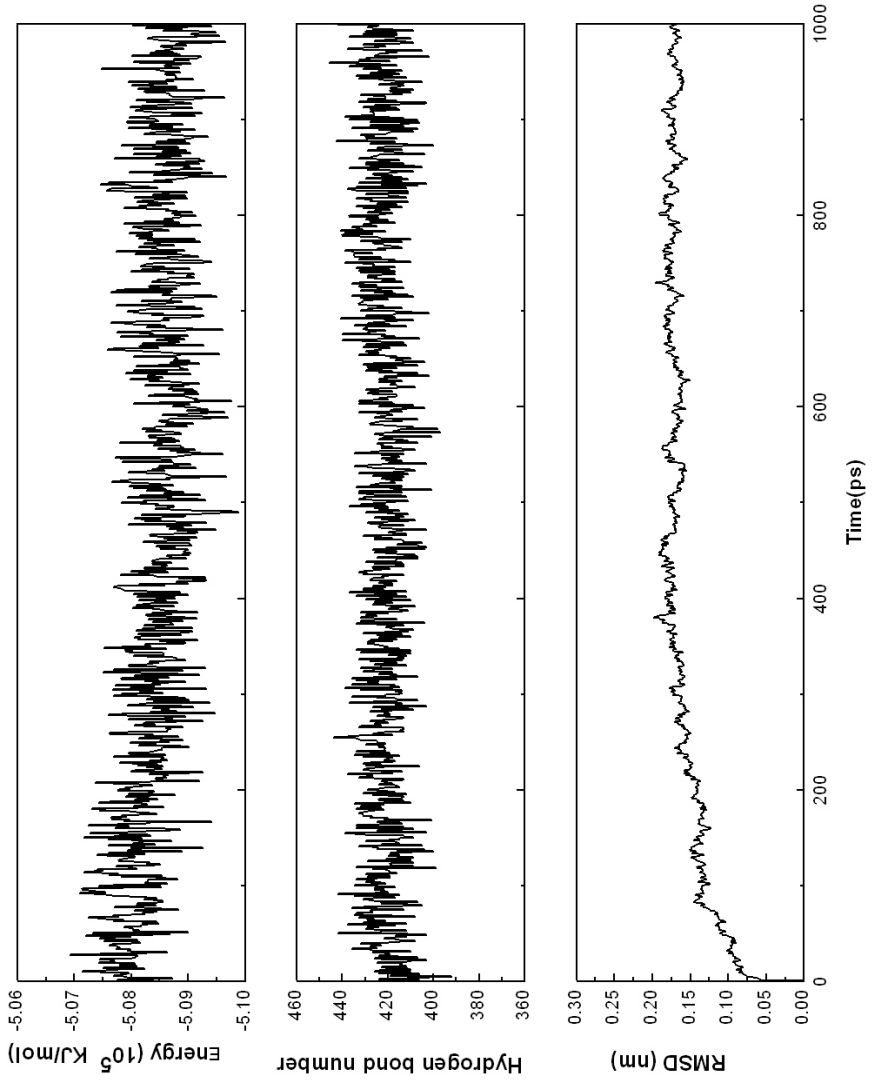

Figure 3

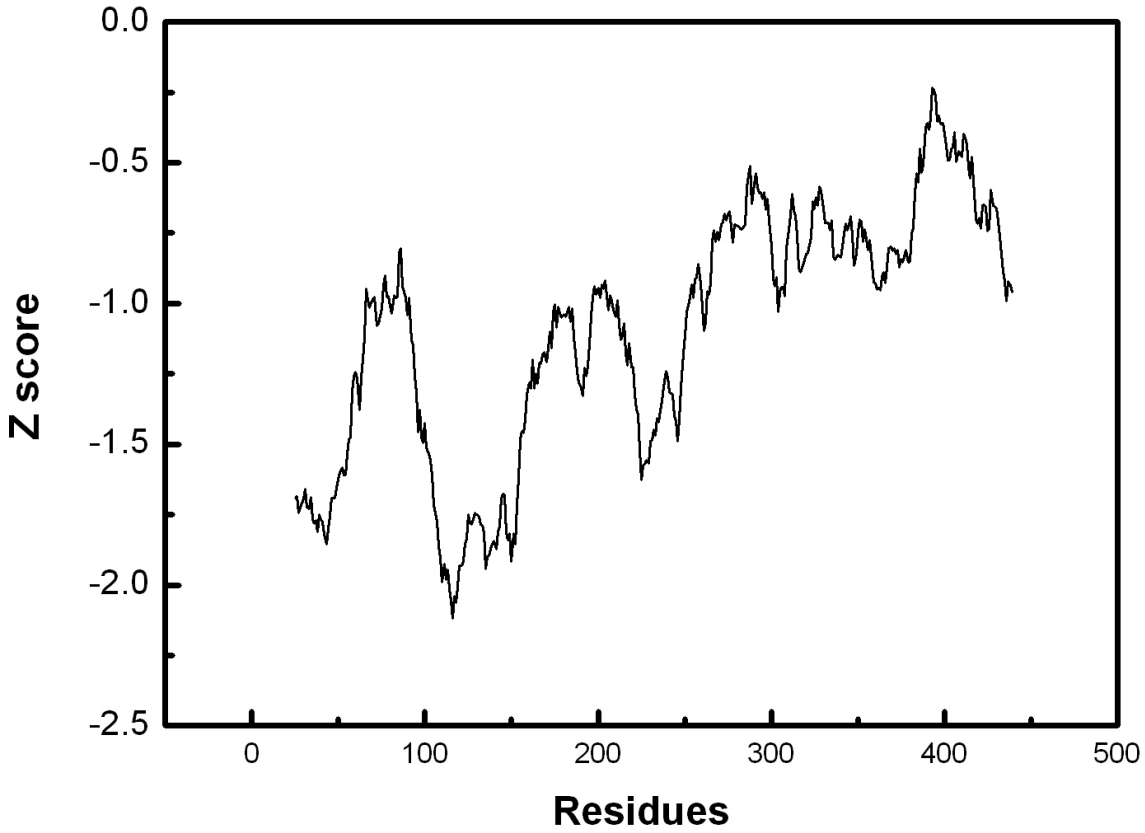

Figure 4

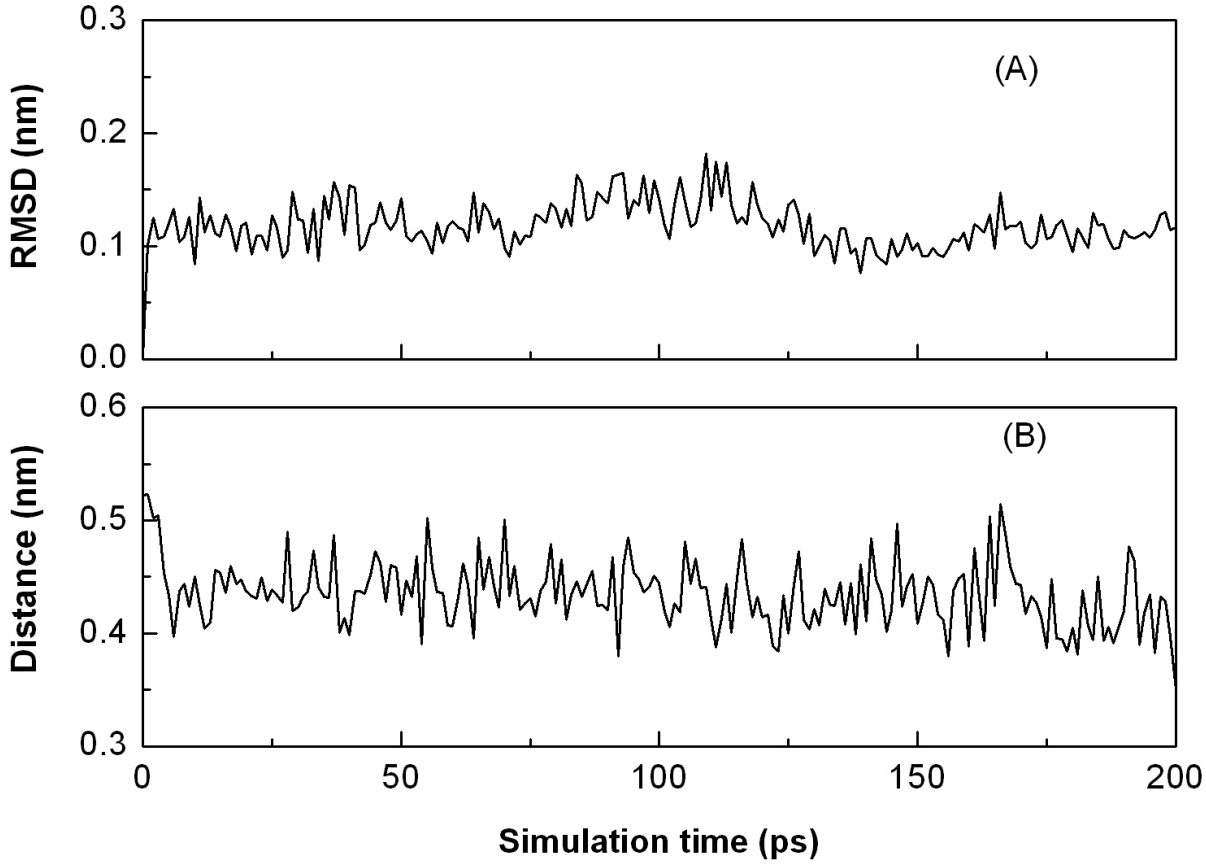

Figure 5

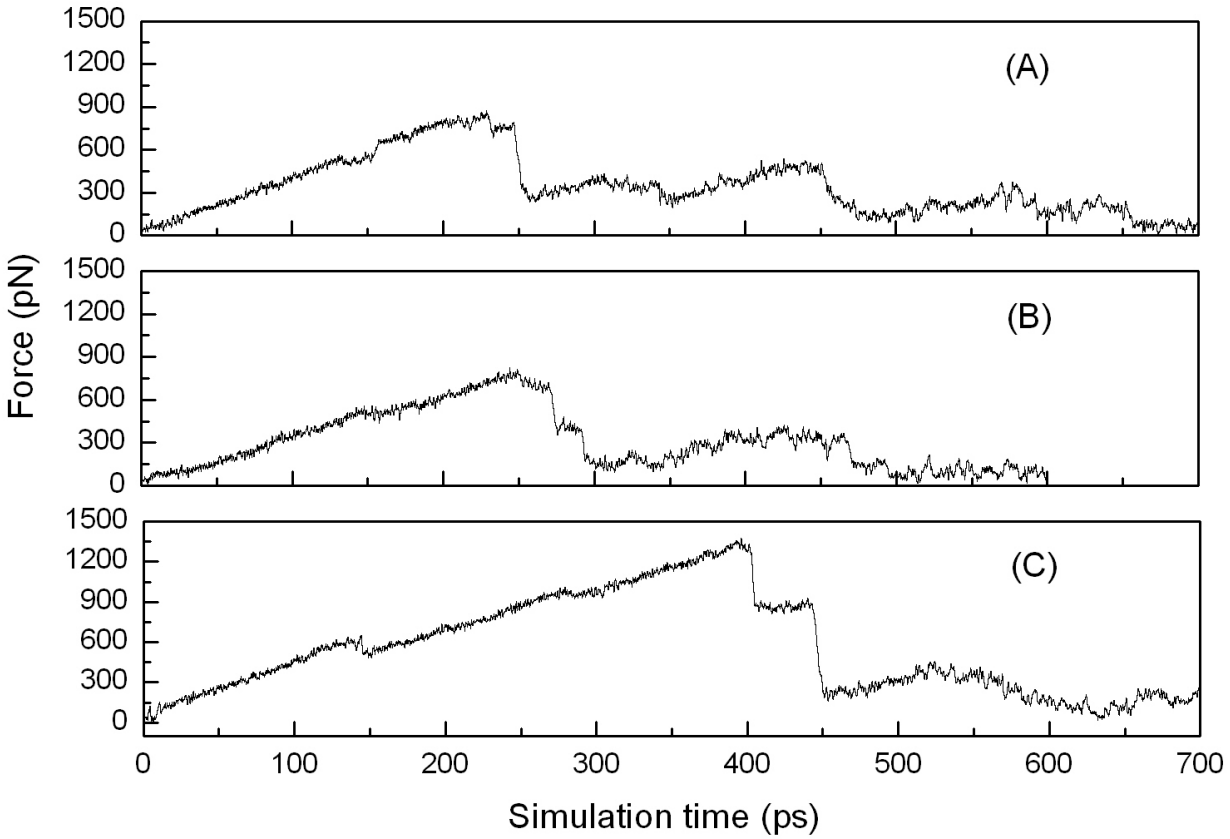

Figure 6

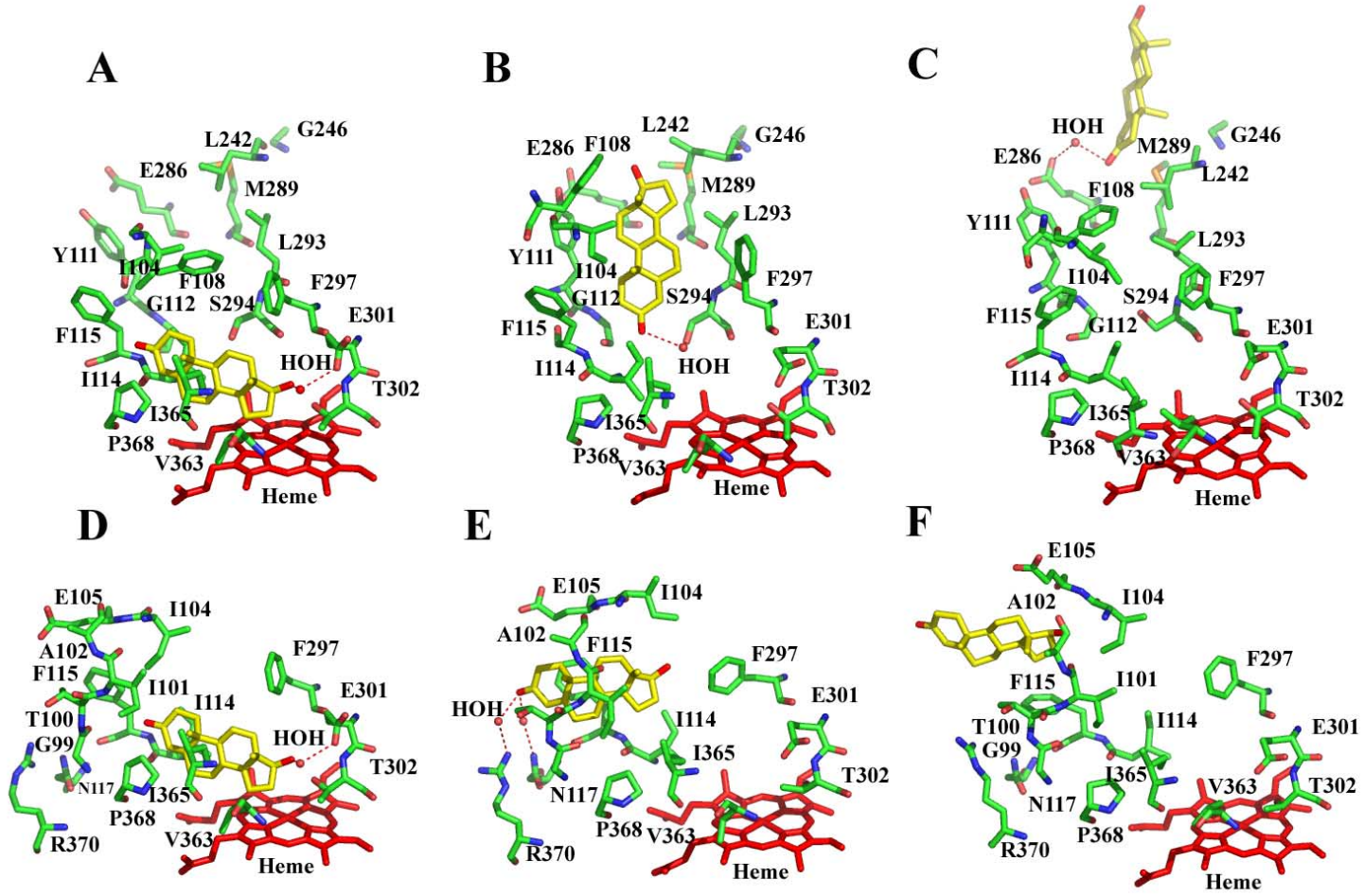

Figure 7

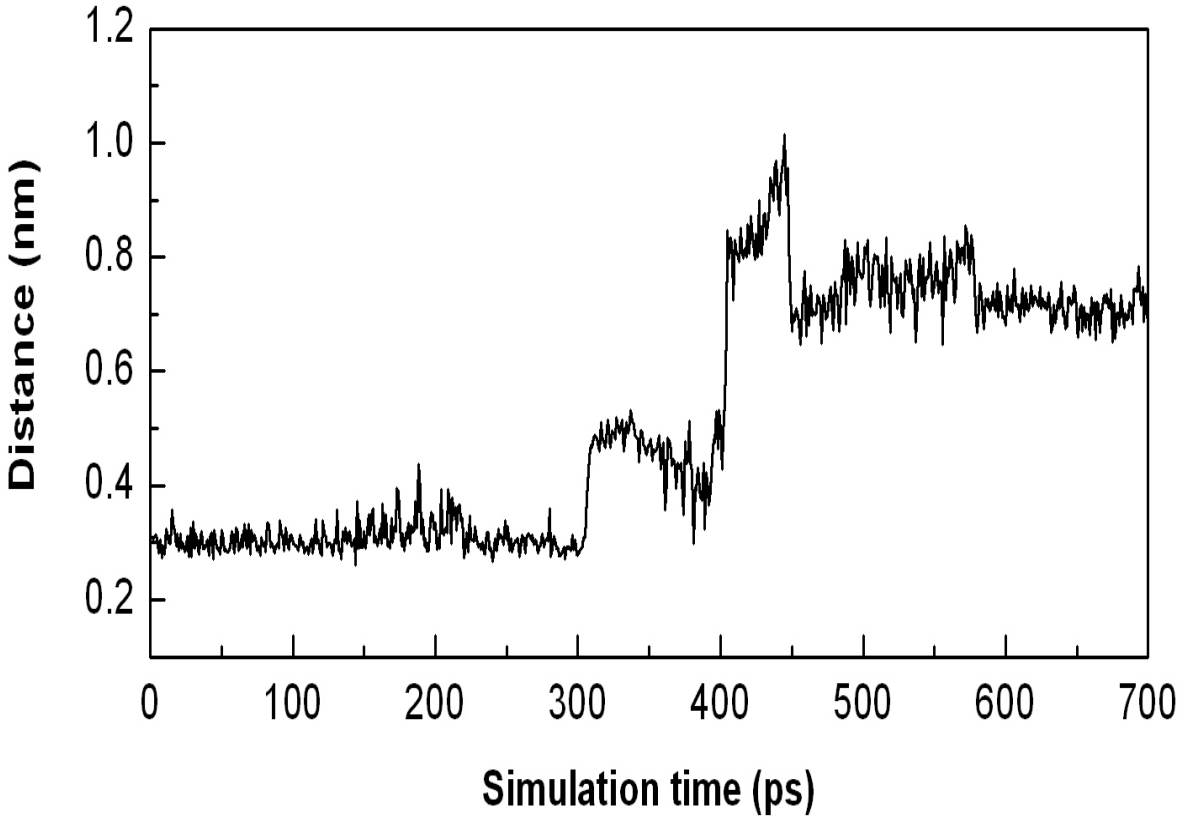

Figure 8

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200

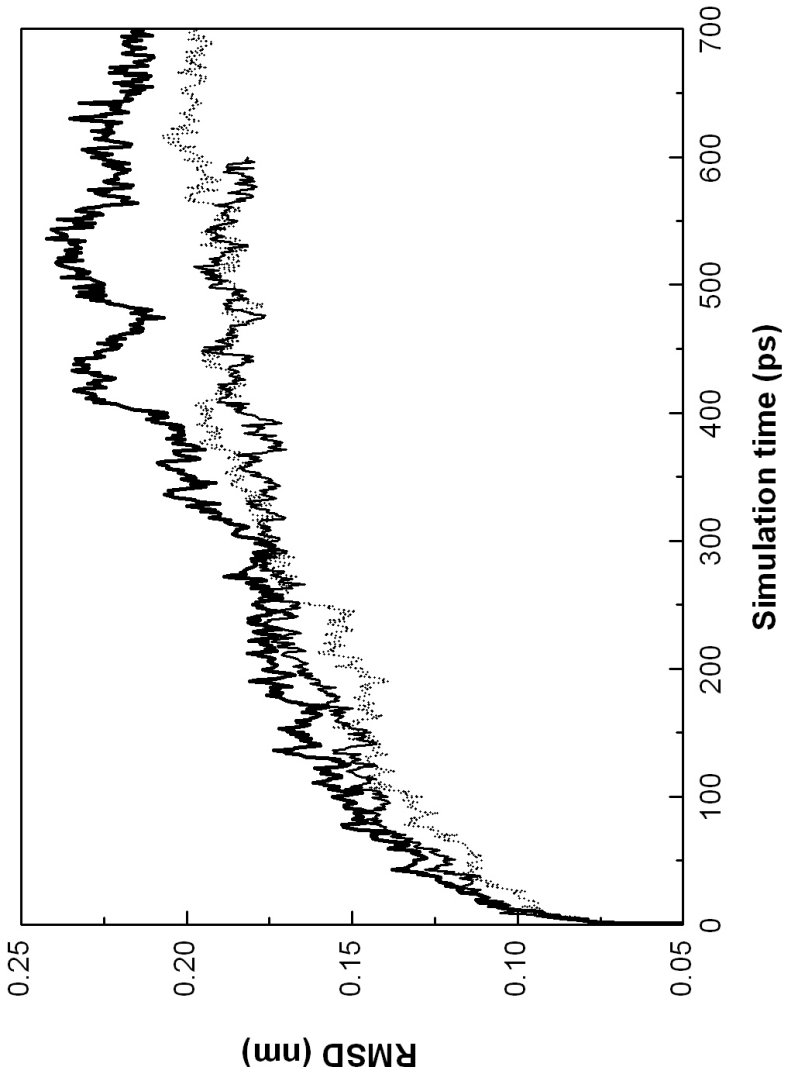

Figure 9

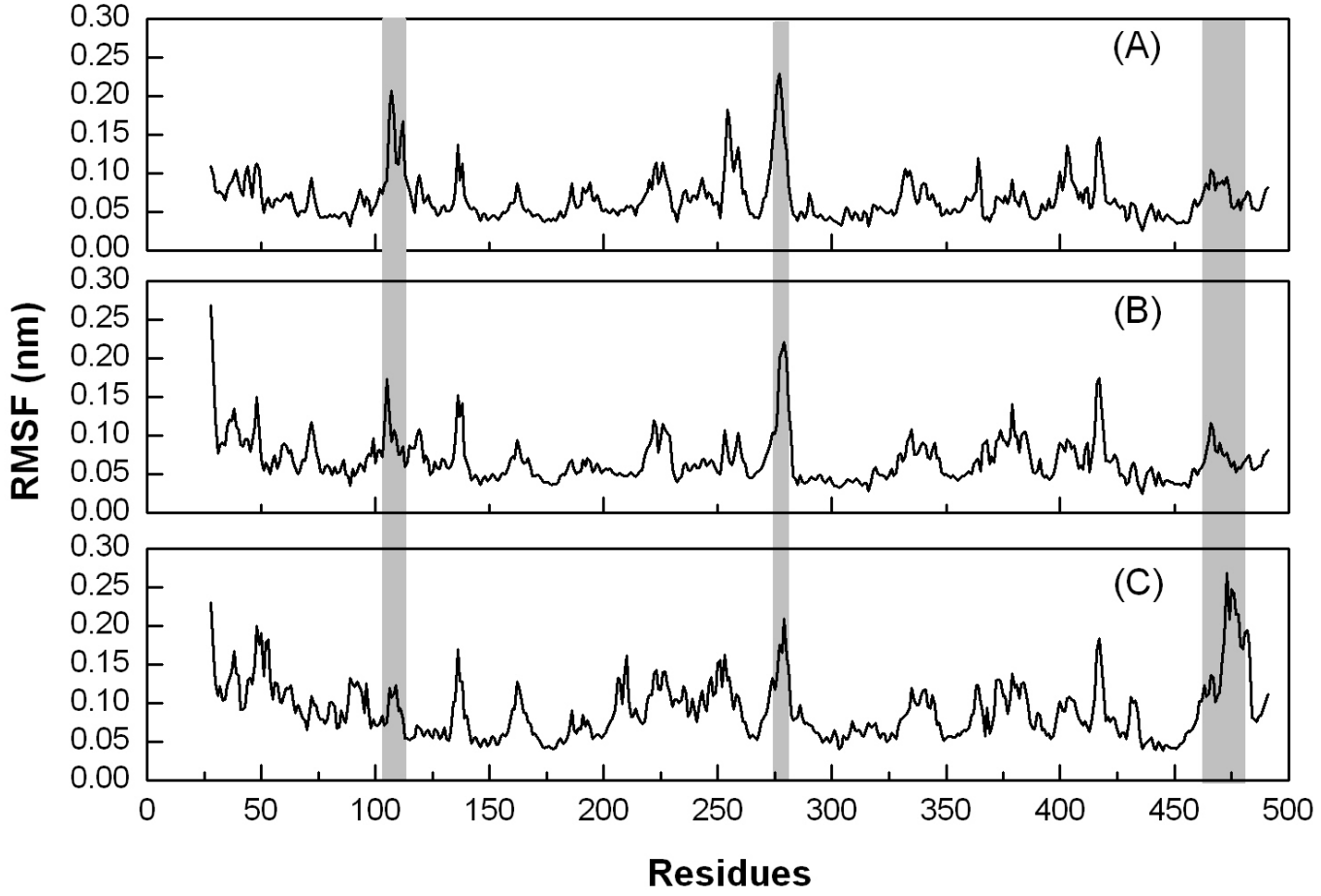

Figure 10

DMD Fast Forward. Published on April 8, 2005 as DOI: 10.1124/dmd.105.004200
This article has not been copyedited and formatted. The final version may differ from this version.

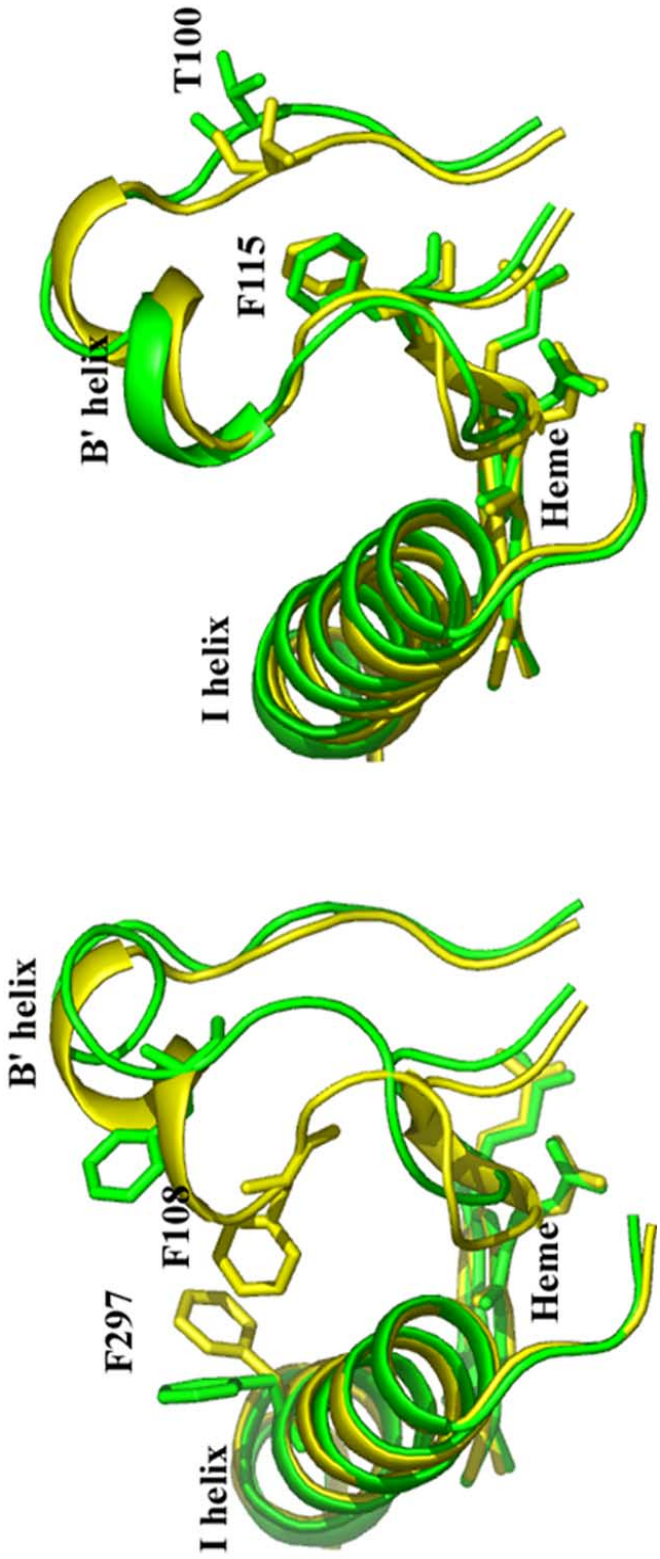

Figure 11

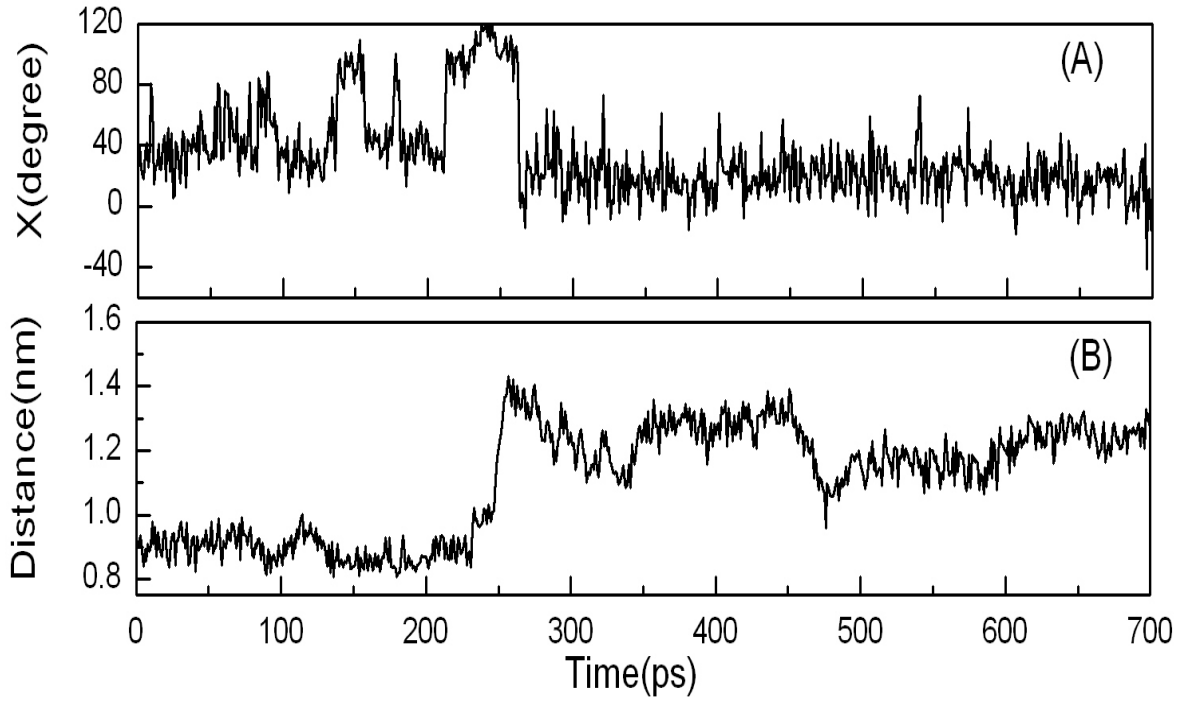

Figure 12