Published in final edited form as:

J Comput Chem. 2014 September 15; 35(24): 1748–1756. doi:10.1002/jcc.23680.

Including Ligand Induced Protein Flexibility into Protein Tunnel Prediction

Laura J. Kingsley¹ and Markus A. Lill¹

¹ Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, 575 Stadium Mall Dr. West Lafayette, IN 47907

Abstract

In proteins with buried active sites, understanding how ligands migrate through the tunnels that connect the exterior of the protein to the active site can shed light on substrate specificity and enzyme function. A growing body of evidence highlights the importance of protein flexibility in the binding site upon ligand binding; however, the influence of protein flexibility throughout the body of the protein during ligand entry and egress is much less characterized. We have developed a novel tunnel prediction and evaluation method named IterTunnel, which includes the influence of ligand-induced protein flexibility, guarantees ligand egress, and provides detailed free energy information as the ligand proceeds along the egress route. IterTunnel combines geometric tunnel prediction with steered MD in an iterative process to identify tunnels that open as a result of ligand migration and calculates the potential of mean force (PMF) of ligand egress through a given tunnel. Applying this new method to cytochrome P450 2B6 (CYP2B6), we demonstrate the influence of protein flexibility on the shape and accessibility of tunnels. More importantly, we demonstrate that the ligand itself, while traversing through a tunnel, can reshape tunnels due to its interaction with the protein. This process results in the exposure of new tunnels and the closure of pre-existing tunnels as the ligand migrates from the active site.

Keywords

Tunnel prediction; cytochrome P450 2B6; steered molecular dynamics; potential of mean force; umbrella sampling; conformational ensemble; protein flexibility; induced fit

Introduction

Understanding how ligands migrate through tunnels that connect the exterior of the protein to the active site can shed light on substrate specificity and enzyme function. For instance, the enzyme acetylcholinesterase has a 20Å long entrance tunnel leading to the active site which is gated by five aromatic residues that reorient to open or close the tunnel, effectively preventing larger substrates from entering the active site while favoring the entrance of acetylcholine¹. Tunnel structure has also been shown to influence enzyme efficiency. Pavlova *et al.* produced a haloalkane dehalogenase mutant with a 32-fold increase in activity

Correspondence to: Markus A. Lill (mlill@purdue.edu).

Additional Supporting Information may be found in the online version of this article.

towards the compound trichloropropane by mutating two tunnel-lining residues in the enzyme². By altering the two residues, a previously accessible tunnel that otherwise allowed water to access the binding site was closed, resulting in the large increase in catalytic efficiency².

Cytochrome P450 (CYP) enzymes represent a large class of enzymes that catalyze the majority of drug metabolism by the human body³ and are therefore a key target in drug design. The entrance and exit tunnels of these enzymes are especially interesting because, despite structural similarities, they exhibit diverse, yet isozyme-specific substrate specificity⁴. The degree of substrate diversity or specificity of each isozyme is thought to be related to the unique structural flexibility of that isozyme⁵. There is evidence to suggest that the tunnels of CYP enzymes are dynamic and that structural changes within tunnels may be critical for ligand entry and egress⁶. For example, several CYP isozymes have been crystallized in both “open” and “closed” forms⁷. In the open form, tunnels are readily observed between the active site and the surrounding environment. Whereas in the closed form, fewer and less pronounced tunnels are observed⁷. These X-ray structures represent only a snapshot of the protein and do not capture the multitude of conformational states that might occur throughout the ligand migration process.

Geometric tunnel prediction is one of the most widely used classes of methods to identify tunnels. These methods typically use Voronoi based maps to locate continuous voids in a static protein structure and rank the tunnels based on the length and width of the tunnel. Several programs including MolAxis⁸, CAVER⁹⁻¹⁰, and MOLE¹¹⁻¹² utilize this basic premise. Geometric methods have been shown to perform well on protein structures with a single and well-defined tunnel, such as the tunnels of transmembrane channels¹³⁻¹⁴. However, the application of these methods to highly flexible systems such as CYP enzymes can be more challenging. One approach has been to use an ensemble of structures in the tunnel prediction process^{10,15}. However, this approach still lacks the ability to explicitly include ligand migration in tunnel prediction. Furthermore, geometric tunnel prediction does not take into account potential physicochemical interactions between the protein and the ligand when ranking the tunnels. In other words, tunnel ranking using geometric methods is not ligand specific and may be misleading depending on the ligand of interest.

Alternatively, tunnel prediction methods based on molecular dynamics (MD) simulations are much better suited for tunnel prediction in CYPs and other flexible enzymes. These methods directly incorporate both protein flexibility and protein-ligand interactions during the tunnel prediction process, yielding ligand-specific information about a given tunnel. Compared to geometric methods, where tunnels are ranked based on the protein structure alone (*i.e.* length and width), MD based methods can provide tunnel rankings based on the preferences of a particular ligand for a given tunnel or set of tunnels. While they provide more detailed, system-specific information, these methods are significantly more computationally demanding. For instance, the most widely used MD-based method, Random Accelerated Molecular Dynamics (RAMD)¹⁶, mimics substrate migration by applying a randomly oriented force to an explicit ligand and pushing it through the protein until the ligand exits the protein or until the allotted simulation time expires¹⁷⁻¹⁸. This procedure is repeated

multiple times and tunnels are assessed based on the number of simulations that result in successful ligand egress through each tunnel.

While RAMD represents a significant improvement over geometric methods, it requires a considerable number of simulations and does not guarantee successful ligand egress in all simulations¹⁹. Furthermore it does not directly provide specific information about the barriers encountered by the ligand during exit or the conformational changes associated with ligand egress. A detailed understanding of the specific interactions that occur between the ligand and the protein at given points along the tunnel can offer valuable insight for applications in protein engineering and drug design.

We have developed a novel tunnel prediction and evaluation method named IterTunnel, which includes the influence of ligand-induced protein flexibility, guarantees ligand egress, and provides detailed free energy information as the ligand proceeds along the egress route. IterTunnel combines geometric tunnel prediction with steered MD in an iterative process to identify tunnels that open as a result of ligand migration and calculates the potential of mean force (PMF) of ligand egress through a given tunnel.

Our method uses the geometric tunnel prediction program, MolAxis²⁰, to initially identify tunnels leading out of the binding site (**Figure 1A**). Using these tunnels as a guide the ligand is then pulled from the binding site along each initial tunnel using steered MD (**Figure 1B**). After a pre-defined time, the steered MD simulation is stopped and tunnels are recalculated from the new position of the ligand within the tunnel (**Figure 1C**). This allows for the identification of any new tunnels that may open or close as a result of ligand migration. Steered MD is then resumed along the three highest-ranked tunnels as well as the original tunnel (**Figure 1D**). This process is repeated until the ligand exits the protein at which point the simulation is terminated. Using the steered MD trajectories, umbrella sampling is then used to calculate the PMF of ligand transit.

Methods

MD Simulations

All simulations in this study were based on the PDB structure 3IBD²¹. The protein was prepared using Reduce²² to identify the proper rotamer and protonation states of histidine, and the proper rotameric states of asparagine and glutamine. The heme parameters for a nonoxygenated state were extracted from the literature²³. Gromacs was used to solvate the system in an octagonal water box of SPC216 waters and 6 chlorine ions were added to neutralize the system. The box size was selected to guarantee a minimum distance of 15Å between solute and box edge.

MD simulations were performed using Gromacs-4.5.5²⁴⁻²⁵ and the Amber03 force field. Ligand charges and parameters were determined using the antechamber²⁶ package from Amber and the bcc charge method. After solvation of the protein, 1000 steps of energy minimization were performed using the steepest descent method and particle mesh Ewald (PME) summation with a grid size of 0.12nm and 4th order interpolation to compute potential and forces between the grid points. To compute van der Waals interactions a

switching function was applied between 1.0nm and the cut-off of 1.4nm. The LINCS algorithm²⁷ was used to constrain bonds containing hydrogen atoms. The integration time step was 2fs. Next the hydrogen bond network of the surrounding waters was established using a 200ps simulation in which all but the waters were restrained. Simulations were performed at 300K using PME, Berendsen thermostat, and Parrinello-Rahman pressure coupling. Finally a 400ps equilibration run was performed to equilibrate the system prior to the production run.

IterTunnel Algorithm

To initiate tunnel prediction by IterTunnel, we performed equilibration of the protein structure in triplicate to generate three starting structures for input into IterTunnel. MolAxis was run on each of these structures, and the resultant tunnels were then clustered using k-means clustering (5.0Å cutoff), resulting in a total of six initial tunnels. Steered MD was then performed along each identified tunnel by restraining a dummy atom at the endpoint of the tunnel using a 9999 kJ mol⁻¹ nm⁻² force and pulling the ligand slowly toward the dummy atom. We referred to the literature when selecting a pull rate. In similar studies pull rates ranging from 10Å ns⁻¹ 15 - 25Å ns⁻¹ 28-29 were used, but recently it has been suggested that even 10Å ns⁻¹ may be too fast to allow adequate structural relaxation³⁰. We therefore chose a slower pull rate of 6Å ns⁻¹ to allow for longer relaxation.

After 800ps of pulling using steered MD, the simulation was stopped and potential tunnels were recalculated from the current center of mass of the ligand using MolAxis. Pulling was then initiated along the top three identified tunnels and was continued along the original tunnel. This process was repeated three times (at 800ps, 1600ps and 2400ps), producing a total of 40 unique tunnels per initial tunnel. Once the ligand reached bulk solvent, the simulation was terminated and the tunnels from all simulations, 240 tunnels in total, were clustered using the k-means clustering algorithm where k was adjusted such that maximum RMSD between any cluster member and the cluster center was less than 5.0Å.

It is necessary to prevent rotation of the protein during the simulation because even slight rotations can cause misalignment of the pull path between the fixed dummy atom, the ligand and the intended tunnel. In order to prevent rotation of the CYP2B6 system, the alpha carbons of K, L, D, and J helices were restrained during all steered MD simulations using a 1000 kJ mol⁻¹ nm⁻² force constant. These helices are located on the underside of the heme, as opposed to the site of enzymatic reaction. This portion of the protein is not believed to be engaged in ligand entry and exit, but to prevent bias any tunnels which exited below the plane of the heme were not considered.

The PMF of ligand transit along all identified tunnels was calculated using umbrella sampling and the g_wham³¹ tool of Gromacs. The reaction coordinate for the steered MD simulations was defined as the distance between the dummy atom and the center of mass of the chlorophenylimidazole ligand. Frames were extracted every 0.8Å along this reaction coordinate. Additional frames were added manually after the initial selection as necessary to ensure complete coverage of the reaction coordinate. In total, about 45 frames, also called umbrella sampling windows, were selected per tunnel.

Each window was equilibrated for 100ps before starting the production run. Production runs of 500ps, 1ns, 2ns, and 3ns were tested (**SI Figure 1**). No noticeable difference was found between the 2ns and 3ns simulation length, thus a minimum of 3ns was used for all umbrella sampling windows, but on average about 4.5ns per window was completed resulting in ~200ns of sampling per tunnel. While, it is difficult to directly comment on convergence without increasing the simulation length by at least an order of magnitude, 3ns windows are in line with previous studies³² and we believe this to be an efficient choice given the number of windows required.

During both the equilibration and the production run the restraints on the alpha carbons of the K, L, D, and J helices as well as the dummy atom were maintained. In addition, a 1000 kJ mol⁻¹ nm⁻² harmonic spring constant was used to restrain the ligand along the reaction coordinate in each window.

The g_wham³¹ program from Gromacs was used to estimate the PMF from the sampled windows. The average distance pulled was around 30-35Å which was broken into 100 bins by the g_wham algorithm. Error estimates were made using the b-hist bootstrapping method from the g_wham program. In this method, all histograms are considered as independent data points and random weights are assigned to each histogram to estimate the error.

Ensemble Based Tunnel Prediction

Tunnels predicted by IterTunnel were compared to tunnels predicted in a structural ensemble of CYP2B6 structures. The ensemble was generated by extracting frames every 10ps from a 10ns MD trajectory of chlorophenylimidazole bound CYP2B6. Caver3.0²⁰ was used to calculate tunnels in the ensemble. The probe radius was chosen based on the ligand chlorophenylimidazole. The widest distance between two heavy atoms in the phenyl-ring group of the ligand is approximately 2.5Å, therefore a probe radius at half of that width (1.25Å) was selected. By selecting this probe radius, we assume that minor changes within the tunnel are feasible which would allow the hydrogenated ligand to pass. The starting point was chosen to be 3Å above the iron atom of the heme group. The standard clustering settings of Caver3.0 were used and any resultant tunnels that exited below the plane of the heme were removed.

The ligand was then pulled through each tunnel identified by Caver3.0. To pull precisely along the paths defined by Caver3.0, tunnels were broken down into linear sections that where the angle between two sections differed by no more than 30 degrees. The ligand was pulled along each linear section, using the parameters described above, and the PMF was calculated.

All tunnels identified were named based on the nomenclature set forth by Cojocaru *et. al.*³³⁻³⁴

Results

IterTunnel was designed to incorporate ligand migration into the tunnel prediction and evaluation process. Using CYP2B6 as a model system, we compared the tunnels predicted by IterTunnel to those predicted in a 100-member ensemble generated by MD simulation. A

total of eleven potential exit routes were found using IterTunnel and a total of five potential exit routes were found in the MD ensemble.

Tunnels Exiting Near the F/F'-G/G' Helices

In the ensemble two tunnels, 2a and 2e were found to exit near the F/F' and G/G' helices (**Figure 2**, grey tunnels). Using IterTunnel, these two tunnels were identified as well as three additional tunnels 2ac, 2d, and 2f (**Figure 2**, colored tunnels). Each of the tunnels that were unique to IterTunnel were found to cause conformational changes of greater than 2.5Å in the protein backbone. For instance, ligand egress via tunnel 2d, causes a >3.5Å shift in the AA' loop (**Figure 3a**). However, the nearby tunnel 2a is open in most structures of the ensemble and ligand migration is found to induce only minor (~1Å) changes in the F'-G' loop and the β_1 sheet (**Figure 3b**).

Solvent Tunnels

The solvent tunnel is ubiquitous in CYP enzymes and is thought to regulate water access to the binding site³⁵. Often, the solvent tunnel is split into two branches in CYP isozymes³³. In CYP2B6 one branch was identified in both the ensemble and with IterTunnel while the second branch was only identified using IterTunnel.

The first branch passes the β_5 sheet, the F helix, and the E helix causing a minor conformational change in the C-terminal loop (**Figure 4**, grey tunnel). The same minor conformational change is observed during ligand egress through the second branch, but in addition a ~4Å shift in the FF' loop/F helix (**SI Figure S2**) before proceeding to exit between the β_5 sheet and the F helix (**Figure 4**, cyan tunnel).

Exit between the F and G Helices

Tunnels that exit between the F and G helices are broadly classified as tunnel 3. Using IterTunnel, two geographically-distinct subclasses of pathway 3 were found which will be referred to as 3a and 3b. Tunnel 3a is broken into two branches, one which exits near the C terminal region of the F helix (**Figure 5A**, green tunnel) and a second branch which exits at the middle of the F helix (**Figure 5A**, yellow tunnel). Tunnel 3b exits between the EF loop and the G helix (**Figure 5A**, blue tunnel). No tunnels were found to exit via any tunnel 3 route in the ensemble, likely a consequence of the bottleneck formed by the F and G helices. Using IterTunnel, significant conformational changes were observed in this region that were necessary to accommodate ligand egress.

The conformational changes induced by ligand migration through tunnel 3 are shown in **Figures 5b, 5c, and 5d**. F202 plays a key role in exit via all three tunnels and may act as a gate to substrate exit. Ligand egress causes this residue to rotate by nearly 180 degrees causing the outward shift in the F helix which in turn causes a widening of this tunnel to accommodate ligand passage. The average diameter along each tunnel for both the ensemble and IterTunnel is shown in **Figures 5e, 5f, and 5g**. In the ensemble, large portions of all three of these tunnels are closed due to steric occlusion by residues such as F202. However, using IterTunnel, migration of the ligand and its interaction with surrounding residues causes these tunnels to open allowing ligand passage.

Tunnel 1 and 5

Tunnel 5, was identified in two branches. One branch was observed only in the ensemble (**Figure 6**, orange tunnel) and one was found only using IterTunnel (**Figure 6**, green tunnel). The initial portion of this tunnel overlaps with the solvent tunnel, but it then turns and proceeds along the C-terminal loop before exiting near the K helix. Tunnel 1 exits on the opposite side of the CYP enzyme, near the H and C helices, and was uniquely identified in the ensemble. Although this was the only tunnel that was identified in the ensemble but not identified by IterTunnel, it was only found in a single member of the entire 100-member ensemble and was the worst-ranked by the Caver3.0¹⁰ scoring algorithm.

PMF to Estimate Tunnel Favorability

Given the large number and geographic diversity of the tunnels identified by both IterTunnel and in the ensemble, a critical consideration is the relative preference of the ligand for each of these tunnels. To identify the most biologically relevant tunnels for a given ligand, protein-ligand interactions and potential energetic barriers encountered by the ligand as it exits the protein must be taken into account. One way that this can be achieved is by computing the PMF along each pathway. This provides both an estimate of the total free energy of ligand egress as well as a detailed energetic account of ligand migration along each pathway.

The PMF of each tunnel was calculated using umbrella sampling and the weighted histogram analysis method (WHAM)³¹. Table 1 shows the PMF of the largest barrier encountered by the ligand as it progressed through each tunnel and which method (IterTunnel, ensemble or both) was used to identify the tunnel. Tunnels were classified as having either small or large barriers. If the ligand encountered any barrier exceeding 65KJ mol⁻¹, the tunnel was considered to have a large barrier, otherwise the tunnel was considered to have small barriers. The PMF profiles of all tunnels have been provided in the **Supporting Information** (Figures S3-S7).

Of the eight tunnels identified with small barriers, all were identified by IterTunnel and over half were unique to IterTunnel. Tunnels 1, 3 and 5, were found to have large barriers. Tunnel 1 was found only in the ensemble, tunnel 5 was found using both methods, and tunnel 3 was only found using IterTunnel.

In theory, the overall change in PMF between the bound state and the unbound state should be equivalent, regardless of the egress tunnel. We found that, in general, this held true for our simulations and that the difference between the bound and unbound state was on average around 57kJ mol⁻¹. However, some variations were observed between tunnels. In addition to standard error, small variations may be the result residual interactions between the ligand and protein. For instance, if the ligand is pulled into bulk solvent, but some contacts with surface exposed residues are still present, the final energy may reflect these residual interactions. We believe these two sources may account for fluctuations up to about 5-10kJ/mol. Alternatively, large fluctuations, such as those observed in tunnel 3 are likely caused by the structural changes in the protein that fail to return to an equilibrium state (see “Discussion”).

Discussion

Tunnels with Small Barriers

In most cases, binding was observed to be a downhill process, in other words, no significant barriers to ligand egress were found. This is phenomenon has been previously observed in CYP enzymes and is suggestive of a structure-function relationship^{19,36}. The biological role of CYP enzymes in the body is critically dependant on the ability to rapidly metabolize a diverse array of substrates. Significant barriers in the tunnels may prevent certain crucial substrates from binding or prevent or slow product egress. This, in turn, could result in slowed or altered metabolic ability and lead to increased vulnerability to potential side effects of exogenous compounds. Therefore, the tunnels of CYP enzymes must be able to efficiently accommodate the entry and egress of many diverse substrates. This trend may be even more apparent in smaller substrates, such as chlorophenylimidazole, used in this study.

Tunnel 2—All sub-tunnels of tunnel 2 were found to be favorable for chlorophenylimidazole egress in CYP2B6. With the exception of 2e, which passes through the BC loop, each subclass of tunnel 2 exits near the F and G helices, a structural unit known as the FG block³⁴. Due to its location directly above the active site and structural heterogeneity between different CYP isozymes the FG block is thought to be a key determinant of ligand entry and egress in CYP enzymes^{6,34,37}. For instance, studies of CYP P450cam binding to a large ruthenium-linked substrate have shown that the FG loop can shift by up to 7.5Å to accommodate ligand passage³⁷.

In comparison to other CYP enzymes like CYP P450cam, CYP2B6 has an especially helix-dense FG block, containing both the F' and G' helices in addition to the parent F and G helices connected by particularly short loops. In all three currently available crystal structures of CYP2B6, the F'G' region is in close contact with surrounding secondary structures. In other words, the only available CYP2B6 structures are in the closed form³⁴ and potential exit routes in this area are obscured.

The unique composition of the FG block in CYP2B6 is likely the underlying reason for differences in identification of tunnel 2 subclasses between the ensemble, which identifies only 2a and 2e, and IterTunnel, which identifies 2a and 2e as well as 2ac, 2d, and 2f. Although modest in comparison to the changes observed in P450cam³⁷, the use of IterTunnel led to observed shifts of up to 4Å in the FG loop region and surrounding structures that were not observed in the ensemble. Without ligand passage to influence the F'G' region and surrounding regions, the 2ac, 2d, and 2f tunnels remain closed and therefore could not be identified in the ensemble. These tunnels were found to be favorable for ligand egress, suggesting that ligand induced conformational changes in both the loops and helices in this region may be important for tunnel identification in closed CYP structures, such as CYP2B6.

Solvent Tunnel—Based on the PMF, both branches of the solvent tunnel were found to be energetically favorable for ligand egress in CYP2B6. The solvent tunnel is thought to be an important ligand egress tunnel in other CYP isozymes. It has been suggested as the main

ligand exit pathway in CYP2D6³⁸ and for specific ligands, including temazepam and testosterone, in CYP3A4²⁸.

Using IterTunnel, both branches of the solvent tunnel were identified. However, only one branch of the solvent tunnel was identified in the ensemble, and it was the second poorest ranked according to Caver3.0. One possible reason that the first branch was so poorly ranked and the second branch was not found at all in the ensemble is the method of ensemble generation. It has been suggested that the binding of high affinity ligands can cause the solvent tunnel to close in CYP enzymes³⁹. The ensemble was generated from a ligand-bound form of CYP2B6, wherein both solvent tunnels remained in a relatively closed conformation for the duration of the simulation leading to a poor ranking or complete lack of identification of the branches respectively. However, using IterTunnel, ligand egress caused the widening and opening of these branches, respectively. Therefore, both of these tunnels were able to be identified using IterTunnel while only one was identified in the ensemble.

Rare Tunnels and Tunnels with large barriers

Rare Tunnels—Both tunnel 1 and tunnel 5 are classified as rare because they are only observed in a small number of CYP enzymes and are generally thought to be unfavorable for ligand egress⁷. Our findings are in agreement with previous results and indicate that passage of chlorophenylimidazole through either of these tunnels is energetically unfavorable and that the ligand encounters significant energetic barriers upon egress (**SI, Figures S5 and S6**).

Tunnel 1 was the only tunnel that was not identified by IterTunnel. The most likely reason for this is that ligand migration through nearby tunnels causes a narrowing of this tunnel and prevents it from being identified by IterTunnel. Compared to tunnel 1, tunnels 2e and 2ac are located on the opposite side of the BC loop. Ligand migration through either of these tunnels causes a downward shift in the BC loop constricting tunnel 1. Therefore, if the ligand started to migrate down either tunnel 2e or 2ac, tunnel 1 may have become constricted, ultimately preventing its identification by IterTunnel.

Tunnel 5 was identified as two sub-tunnels, one of which was found in the ensemble while the other was found using IterTunnel. Interestingly, after accounting for ligand migration, the sub-tunnel found by IterTunnel was wider, but had many more barriers in PMF profile compared to the sub-tunnel found in the ensemble (**SI, Figure S6**). One possible explanation for this is that during egress the ligand interacts with different hydrophobic regions of the I helix. In the sub-tunnel identified by IterTunnel the largest PMF peak (65 kJ mol^{-1}) is caused when the ligand passes through a hydrophobic patch formed by the C-terminal region of the I helix (**Figure 7**). In the sub-tunnel found in the ensemble, however, the ligand passes by the center of the I helix and interacts with another hydrophobic patch near the middle of the I helix causing an earlier PMF peak (46 kJ mol^{-1}).

The PMF differences between these sub-tunnels emphasize two important considerations in tunnel identification and ranking. First, that the geometric features of length and width, commonly used to rank tunnels, may not be as accurate as methods which incorporate physicochemical features. Notably, a recently released geometric tunnel prediction

algorithm, Mole2.0 has attempted to incorporate physicochemical features in tunnel ranking⁴⁰. Second, this finding also highlights the importance of tunnel clustering, which is routinely used in tunnel prediction algorithms. Our findings suggest that tunnels which are physically very close to one another (i.e. would likely be considered as members of the same cluster) can have significantly different PMF profiles.

Tunnel 3—Ligand egress between the F and G helices was only observed using IterTunnel. It is thought that this region acts as a clamp that holds the ligand in the active site³⁷. IterTunnel identifies two tunnels that exit via this region, both of which cause significant rearrangement of this clamp. It is likely that egress through these tunnels significantly perturbs the hydrogen bonding and hydrophobic packing of these helices. Although tunnel 3a branch 2 was found to be favorable, there is a significant peak in the PMF as it passes through these helices (Supporting Information Figure S7, panel B). This was the only tunnel with a small barrier that demonstrated such a peak. The remaining tunnels, 3a branch 1 and 3b were both found to have large PMF values for ligand egress (72 and 77 kJ/mol respectively). The difference in the final PMF values of these two tunnels and the other tunnels tested suggests that even after the ligand exits the protein, the packing cannot be re-established in the simulation time allotted.

While our results suggest that these two tunnels are not a likely for ligand egress, the identification and characterization of such tunnels could be important for protein engineering and the study of naturally occurring mutations. Recently, the effect of ligand egress in this region of CYP17A1, which has naturally occurring mutations in both the F and G helices, has been investigated¹⁵. These mutants have a decreased ligand binding affinity and it has been proposed that the reason for this is that the mutations perturb the packing of the F and G helices and allow the ligand to prematurely escape the binding site¹⁵. The identification of egress routes like tunnel 3 suggest that IterTunnel could be used to highlight areas that are susceptible to ligand egress/entrance. These data could then be used in protein engineering studies, for example, to finely tune access to binding site(s).

Conclusions

We have developed a novel tunnel prediction and characterization method called IterTunnel. We compared tunnels found in a 100-member ensemble of CYP2B6 to tunnels found using our method and found significant differences in the tunnels identified by IterTunnel as compared to tunnels identified in the ensemble. By including ligand migration during the tunnel identification process, tunnel opening was observed as a result of ligand-induced conformational changes in the protein. In the ensemble, these tunnels were not identified because such conformational changes could not be achieved without the inclusion of the ligand at various points along the egress pathway. Furthermore, by computing the PMF of each tunnel, we found that tunnels identified by only IterTunnel were energetically favorable for ligand egress.

Our study further demonstrates that tunnel identification in CYP enzymes is complicated by both the inherent flexibility of the enzyme and its interaction with the substrate. It is well known that a protein can adapt its conformational state in accordance with the ligand bound

in the binding site, as theorized by the induced fit or conformational selection model. Such conformational adaptation to the bound ligand is observed for a large class of protein-ligand systems^{41–43}. This study suggests that protein conformational adaptation is also important as the ligand traverses from the binding site to bulk solvent. Although our study was focused on one particular protein system, our conclusions about the importance of ligand-induced protein adaptation for tunnel identification are likely to be broadly applicable in other systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank Hamed Tabatabaei Ghomi, William McGee, and Gregory Wilson for valuable suggestions during the preparation process. The authors gratefully acknowledge grants from the NIH (GM092855) and the U.S. Department of Education (GAANN) for supporting this research.

References and Notes

1. Bui J, Tai K, McCammon J. *J Am Chem Soc.* 2004; 126:7198–7205. [PubMed: 15186156]
2. Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, Wade RC, Tsuda M, Nagata Y, Damborsky J. *Nat Chem Biol.* 2009; 5:727–733. [PubMed: 19701186]
3. Wienkers LC, Heath TG. *Nat Rev Drug Discov.* 2005; 4:825–833. [PubMed: 16224454]
4. Lindberg RLP, Negishi M. *Nature.* 1989; 339:632–634. [PubMed: 2733794]
5. Otyepka M, Berka K, Anzenbacher P. *Curr Drug Metab.* 2012; 13:130–142. [PubMed: 22208528]
6. Schleinkofer K, Sudarko, Winn PJ, Ludemann SK, Wade RC. *EMBO Rep.* 2005; 6:584–589. [PubMed: 16028306]
7. Cojocaru V, Winn PJ, Wade RC. *BBA-Gen Subjects.* 2007; 1770:390–401.
8. Yaffe E, Fishelovitch D, Wolfson H, Halperin D, Nussinov R. *Proteins.* 2008; 73:72–86. [PubMed: 18393395]
9. Petrek M, Otyepka M, Banas P, Kosinova P, Koca J, Damborsky J. *BMC Bioinformatics.* 2006; 7:316. [PubMed: 16792811]
10. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, Gora A, Sustr V, Klvana M, Medek P, Biedermannova L, Sochor J, Damborsky J. *PLoS Comput Biol.* 2012; 8:e1002708. [PubMed: 23093919]
11. Petrek M, Koainová P, Ko a J, Otyepka M. *Structure.* 2007; 15:1357–1363. [PubMed: 17997961]
12. Sehnal D, Svobodova Varekova R, Berka K, Pravda L, Navratilova V, Banas P, Ionescu C-M, Otyepka M, Koca J. *Journal of Cheminformatics.* 2013; 5:39. [PubMed: 23953065]
13. Howard RJ, Murail S, Ondricek KE, Corringier P-J, Lindahl E, Trudell JR, Harris RA. *Proc Natl Acad Sci USA.* 2011; 108:12149–12154. [PubMed: 21730162]
14. Payandeh J, Scheuer T, Zheng N, Catterall WA. *Nature.* 2011; 475:353–358. [PubMed: 21743477]
15. Cui Y-L, Zheng Q-C, Zhang J-L, Xue Q, Wang Y, Zhang H-X. *Journal of Chemical Information and Modeling.* 2013
16. Ludemann S, Lounnas V, Wade R. *J Mol Biol.* 2000; 303:797–811. [PubMed: 11061976]
17. Winn PJ, Lüdemann SK, Gauges R, Lounnas V, Wade RC. *Proc Natl Acad Sci USA.* 2002; 99:5361–5366. [PubMed: 11959989]
18. Lüdemann SK, Lounnas V, Wade RC. *J Mol Biol.* 2000; 303:813–830. [PubMed: 11061977]
19. Shen Z, Cheng F, Xu Y, Fu J, Xiao W, Shen J, Liu G, Li W, Tang Y. *PLoS ONE.* 2012; 7:e33500. [PubMed: 22442693]

20. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R. *Proteins*. 2008; 73:72–86. [PubMed: 18393395]
21. Gay SC, Shah MB, Talakad JC, Maekawa K, Roberts AG, Wilderman PR, Sun L, Yang JY, Huelga SC, Hong W-X, Zhang Q, Stout CD, Halpert JR. *Mol Pharmacol*. 2010; 77:529–538. [PubMed: 20061448]
22. Word JM, Lovell SC, Richardson JS, Richardson DC. *J Mol Biol*. 1999; 285:1735–1747. [PubMed: 9917408]
23. Oda A, Yamaotsu N, Hirono S. *Journal of Computational Chemistry*. 2005; 26:818–826. [PubMed: 15812779]
24. Hess B, Kutzner C, van der Spoel D, Lindahl E. *J Chem Theory Comput*. 2008; 4:435–447.
25. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. *J Comput Chem*. 2005; 26:1701–1718. [PubMed: 16211538]
26. Wang J, Wang W, Kollman PA, Case DA. *J Mol Graph Model*. 2006; 25:247–260. [PubMed: 16458552]
27. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. *J Comput Chem*. 1997; 18:1463–1472.
28. Fishelovitch D, Shaik S, Wolfson HJ, Nussinov R. *The Journal of Physical Chemistry B*. 2009; 113:13018–13025. [PubMed: 19728720]
29. Sgrignani J, Magistrato A. *Journal of Chemical Information and Modeling*. 2012; 52:1595–1606. [PubMed: 22621202]
30. Otyepka M, Berka K, Anzenbacher P. *Curr Drug Metab*. 2012; 13:130–142. [PubMed: 22208528]
31. Hub JS, d. G. B. van der Spoel D. *J Chem Theory Comput*. 2010; 6:3713–3720.
32. Pathak AK, Bandyopadhyay T. *Proteins: Structure, Function, and Bioinformatics*. 2014 n/a-n/a.
33. Cojocaru V, Winn P, Wade R. *Biochim Biophys Acta*. 2007; 1770:390–401. [PubMed: 16920266]
34. Yu X, Cojocaru V, Wade RC. *Biotechnology and Applied Biochemistry*. 2013; 60:134–145. [PubMed: 23587001]
35. Cojocaru V, Balali-Mood K, Sansom MSP, Wade RC. *PLoS Comput Biol*. 2011; 7:e1002152. [PubMed: 21852944]
36. Cui Y-L, Zheng Q-C, Zhang J-L, Xue Q, Wang Y, Zhang H-X. *Journal of Chem Inf Model*. 2013
37. Dunn AR, Dmochowski IJ, Bilwes AM, Gray HB, Crane BR. *Proceedings of the National Academy of Sciences*. 2001; 98:12420–12425.
38. Rowland P, Blaney FE, Smyth MG, Jones JJ, Leydon VR, Oxbrow AK, Lewis CJ, Tennant MG, Modi S, Eggleston DS, Chenery RJ, Bridges AM. *Journal of Biological Chemistry*. 2006; 281:7614–7622. [PubMed: 16352597]
39. Haines DC, Tomchick DR, Machius M, Peterson JA. *Biochemistry-US*. 2001; 40:13456–13465.
40. Berka K, Hanak O, Sehnal D, Banas P, Navratilova V, Jaiswal D, Ionescu C-M, Svobodova Varekova R, Koca J, Otyepka M. *Nucleic Acids Res*. 2012; 40:W222–W227. [PubMed: 22553366]
41. Teague SJ. *Nat Rev Drug Discov*. 2003; 2:527–541. [PubMed: 12838268]
42. Cozzini P, Kellogg GE, Spyraakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA. *J Med Chem*. 2008; 51:6237–6255. [PubMed: 18785728]
43. Lill MA. *Biochemistry-US*. 2011; 50:6157–6169.

Figure 1.

IterTunnel method. This method begins with (A) the prediction of original tunnels using the geometric method, MolAxis⁸. Next, (B) steered MD is used to pull the ligand along each identified tunnel. As the ligand migrates through a tunnel, it can induce conformational changes in the protein which may open new tunnels or close pre-existing tunnels(C). In order to account for these changes, the steered MD simulation is halted after a set period of time and tunnels are recalculated. D) The ligand is then pulled through the original tunnel as well as any newly identified tunnels. E) The PMF of the original and any newly identified tunnels is calculated.

Figure 2. Location of the tunnel 2 subclasses in CYP2B6. Tunnels 2a and 2e were found by both IterTunnel and in the ensemble, while tunnels 2ac, 2d and 2f were only found by IterTunnel. These tunnels were exposed, in part, by ligand-induced conformational changes in the F'G' helix region

Figure 3.

Comparison of ligand induced conformational changes in tunnel 2d (A), which was only using IterTunnel and nearby tunnel 2a (B) which was identified by both IterTunnel and the ensemble method. The conformational state of the protein is shown in relation to the location of the ligand, both the chlorophenylimidazole ligand (shown in sticks) and the protein (shown in cartoon) transition from light to dark as the ligand migrates through the tunnel. Ligand migration through tunnel 2d (A) causes a $\sim 3.5\text{\AA}$ shift in the AA' loop shown in blue, but as it migrates through tunnel 2a (B) only causes minor conformational changes in the β_1 sheet and G' helix.

Figure 4.

Location of the solvent tunnels identified in CYP2B6. Branch 1 (grey) of the solvent tunnel was identified in both the ensemble and in IterTunnel, while branch 2 (cyan) was identified only using IterTunnel.

Figure 5.

Tunnel egress through tunnel 3, identified only by IterTunnel. Location of the sub-tunnels 3a and 3b are shown in A. Conformational changes of selected residues as the ligand migrates from the binding site (bound ligand conformation shown in white) through the tunnel (shown in increasingly darker shades) are shown in B, C, and D. Although these tunnels were not identified in the ensemble, the diameter along these tunnels in the ensemble was calculated using Caver3.0 (top portions of E, F, and G). The diameter of the tunnel along its length from the binding site to bulk solvent (left to right) is shown. Red represents a smaller diameter, *i.e.* a bottleneck, and the green represents a wider bottleneck. On average, the tunnels are found to be much wider using IterTunnel (bottom portions of E, F, and G) due to the conformational changes induced by ligand migration.

Figure 6. Locations of tunnel 5 and tunnel 1 in CYP2B6. The ensemble method identifies the orange subclass of tunnel 5 and tunnel 1 (red), whereas IterTunnel identifies the green subclass of tunnel 5.

Figure 7.

Hydrophobic patches encountered by the ligand in the two subclasses of tunnel 5. Hydrophobic residues are colored in red. The subclass found using IterTunnel encounters two hydrophobic patches, one near the β_5 sheet, and another near C-terminus of the I helix. The sub-tunnel identified by the ensemble passes by a hydrophobic patch near the center of the I-helix.

Table 1

Tunnel PMF. The largest PMF peak observed for each identified tunnel and the method by which each tunnel was identified. The distance from the binding site at which the largest barrier occurred has also been provided.

	Tunnel Name	Largest PMF barrier	Location of Barrier (distance from active site)	Identification Method
Small Barriers	2f	48 kJ mol ⁻¹	3.8nm	IterTunnel
	2ac	55 kJ mol ⁻¹	4.3nm	IterTunnel
	Solvent Tunnel Branch 2	55 kJ mol ⁻¹	3.5nm	IterTunnel
	3a Branch 2	58 kJ mol ⁻¹	1.3nm [†]	IterTunnel
	2a	56 kJ mol ⁻¹	3.7nm	Both
	2d	61 kJ mol ⁻¹	3.2nm	IterTunnel
	2e	62 kJ mol ⁻¹	3.4nm	Both
	Solvent Tunnel Branch 1	63 kJ mol ⁻¹	3.0nm	Both
Large Barriers	1	66 kJ mol ⁻¹	2.2nm	Ensemble
	5	65 kJ mol ⁻¹	3.4nm	Ensemble
		65 kJ mol ⁻¹	2.7nm [†]	IterTunnel
	3a Branch 1	72 kJ mol ⁻¹	3.7nm	IterTunnel
	3b	77 kJ mol ⁻¹	2.5nm	IterTunnel

[†] Denotes barriers that occur within the tunnel, before the ligand exits into bulk solvent.