Published in final edited form as:

Arch Biochem Biophys. 2011 May 15; 509(2): 127–132. doi:10.1016/j.abb.2011.02.020.

Active-site residues move independently from the rest of the protein in a 200 ns molecular dynamics simulation of cytochrome P450 CYP119

Relly Brandman[†], Jed N. Lampe^{††}, Yigal Brandman, and Paul R. Ortiz de Montellano^{†,*}

[†] Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA 94158-2517, USA

^{††} Department of Pharmacology, Toxicology and Therapeutics, The University of Kansas Medical Center, 3901 Rainbow Boulevard, MS-1018, Kansas City, KS 66160

Abstract

The conformational dynamics of cytochrome P450 enzymes are critical to their catalytic activity. In this study, the correlated motion between residues in a 200ns molecular dynamics trajectory of the thermophilic CYP119 was analyzed to parse out conformational relationships. Residues that are structurally related, for example residues within a helix, generally have highly correlated motion. In addition, clusters of non-adjacent residues that show correlated motion (“hot spots”) are seen in various regions, including at the base of the F and G helices that make up the most dynamic region of the enzyme. A modified k-means algorithm that clusters residues based on their correlated motion indicates that functionally related residues are in the same cluster (e.g., the catalytic threonines and the heme). Tightly coupled clusters form a solvent-exposed “shell” around the enzyme, whereas less coupling between clusters is seen in regions that are critical to ligand interactions, redox partner interactions, and catalysis. Most notably, we find that residues in the active site move independently from the rest of the enzyme, effectively insulating the catalytic machinery from other regions of the protein.

Keywords

molecular dynamics simulations; mutual information; cytochrome P450; CYP119; protein dynamics

1. Introduction

The cytochrome P450 enzymes (CYP P450s) are versatile catalysts that have been referred to as Nature’s “molecular blowtorch” [1, 2]. This designation derives from their ability to oxidize a wide variety of chemically diverse substrates with exquisite regio and stereo selectivity [3, 4]. Due to this unique capability, P450 enzymes have an exciting potential in biotechnological applications [5], including the production of biofuels [6], bioremediation

© 2011 Elsevier Inc. All rights reserved.

*Corresponding author: Paul R. Ortiz de Montellano, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158-2517, Tel: 415 476-2903, Fax: 415 502-4728, ortiz@cgl.ucsf.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

[7], and the generation of authentic metabolic standards for drug discovery and development [8–11]. Although much progress has been made in understanding the detailed mechanism of the P450 catalytic cycle, efforts to replicate the efficiency and selectivity of P450 mediated oxidation in biomimetic catalysts have met with only limited success [12, 13]. This testifies to the importance of the apoprotein structure in the catalytic process, which is emphasized by an extremely conserved protein fold [3, 12]. Furthermore, the important role of protein dynamics in the P450 catalytic process has been highlighted in a number of recent publications [14–17]. Consequently, a precise understanding of the role that protein structure and dynamics play in ligand recognition and catalysis is needed for the further development of cytochrome P450 enzymes as practical catalysts. Long molecular dynamics simulations [18], which take place on the hundreds of nanosecond to microsecond timescales, provide an approach to model the major conformational changes that occur during catalysis and have proven to be useful in understanding enzymological phenomena [19–21]. However, the abundance of data obtained from such simulations can be difficult to evaluate. Analyzing the correlated motion of residues in the protein can provide insight into which regions of the enzyme are directly involved in the catalytic process. This type of analysis has been successfully applied to identify domain-domain correlations in both HIV protease and cytochrome P450 BM3 [22, 23]. Here we introduce a quantitative analysis of molecular dynamics correlation matrices applied to a long-timescale simulation of cytochrome P450 CYP119. We establish that residues that are functionally related have coupled dynamics. Furthermore, we are able to designate clusters of residues with coupled dynamics, and examine how much coupling occurs between clusters. Finally, the resulting model predicts that residues in the active-site move relatively independently from the rest of the protein, effectively insulating the catalytic machinery from external dynamics.

2. Materials and Methods

2.1 Molecular dynamics simulations

The molecular dynamics starting structure was chain A of a 2.05 Å ligand-free crystal structure of CYP119, a cytochrome P450 from the archaeon *Sulfolobus acidocaldarius* (PDB 1IO9) [24]. Missing residues at the C-terminus were added using PyMOL [25]. Molecular dynamics were carried out using the Gromacs 4.0 simulation software package [26] with the OPLS force field [27] and explicit TIP3P water [28]. The water box was created 15 Å from the protein boundaries, and periodic boundary conditions were used. Simulations were run at 300K with 4 fs timesteps, with constant temperature maintained by the V-rescale algorithm [29]. Hydrogen bonds were constrained using the LINCS algorithm [30]. Seven Na⁺ ions were added to achieve a neutral charge. Long range electrostatic interactions were treated by the Particle-Mesh Ewald method [31]. Previously published heme atom parameters such as partial charges were used [32], with the heme Fe³⁺ coordinated to O₂. The simulation contained 77,588 atoms, including: 6,664 atoms of the protein, 23,639 waters, and 7 Na⁺ ions. Energy minimization was followed by a 200 ps equilibration under constant pressure maintained by coupling the system to a pressure bath at 1.0 atm [33]. Following equilibration, the simulations were maintained at constant volume. Data are based on one 200 ns trajectory collected on 8 processors in a linux cluster.

2.2 Mutual information

Mutual information (MI) between two random variables is calculated as:

$$MI(X;Y)=\sum_{y\in Y}\sum_{x\in X}p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right). \quad [34]$$

To calculate the MI in the dynamics between pairs of residues, we represent the trajectory of each residue as a distance from a moving average over time (i.e. distance over time), a compressed version of the real trajectory (i.e. (x,y,z) coordinates over time). Specifically, the average position of the CA atom is calculated in a 500 ps window centered on the current timestep, and the distance between the current position and moving average position is determined. MI was calculated for all pairs of residues in the protein. Mutual information between the heme and residue X is calculated as the average MI between all heme atoms and the CA of residue X. MI is normalized by joint entropy so that MI between lower entropy residues is weighted equally with MI between higher entropy residues:

$$MI_{norm}(X, Y) = \frac{MI(X; Y)}{H(X, Y)}.$$

Normalized MI (MI_{norm}) is reported throughout the manuscript.

Mutual information calculations were implemented as a Gromacs accessory tool written in the C programming language. An external-memory algorithm was used to reduce the memory requirement. This tool will be made available as free software under the GNU General Public License.

2.3 Clustering

A modified k-means algorithm is used to cluster the residues based on MI. The distance metric is defined as

$$d(X, Y) = H(X, Y) - MI(X; Y)_{[34]}$$

Instead of starting with randomly assigned cluster centers as per the k-means algorithm, we assign each CA position as a cluster center and iterate through all residues 100 times.

3. Results

Correlated motion in a 200 ns MD trajectory of CYP119 was calculated using mutual information (MI), a method that has been used previously in conjunction with molecular dynamics to reveal functionally important correlations [35–37]. MI is a statistical measure of correlation that includes both linear and non-linear correlations and as such captures more biologically relevant correlations than other methods [36].

3.1 Global protein motion

The overall structure of CYP119 stayed stable over the entire 200ns trajectory, with an average C_{α} root mean squared deviation (RMSD) of 2.5Å. As published previously, the region with the greatest RMSD was the F/G loop. The F/G loop begins in the open conformation, samples the closed conformation, and returns to the open conformation [38].

3.2 Highest MI_{norm} pairs are along peptide chains, within secondary structure elements, and in “hot spots” connecting secondary structure elements

MI_{norm} between all pairs of residues was calculated (see *Materials and Methods* for more details) and the resulting 368×368 matrix is shown in Figure 1A. High MI_{norm} (i.e. more tightly coupled motion) between residues is indicated in red, and low MI_{norm} (i.e. residues that move more independently) is indicated in blue. The definition of correlated motion calculated by MI imparts the following two properties on the matrix: (1) the matrix is

symmetric because $MI(X,Y) = MI(Y,X)$, and (2) the diagonal line ($MI(X,X)$, the correlated motion between a residue and itself) is the highest MI value for each residue.

Pairs of residues with the highest correlated motion are shown in Figure 1B–D. Figure 1B highlights just the residue pairs with the highest (red, $MI_{\text{norm}} > 0.35$) and second highest correlated motion (green, $0.3 < MI_{\text{norm}} \leq 0.35$). All other residue pairs are colored blue. As fits with intuition, the highest correlated motion is between neighboring residues along the peptide chain (near the diagonal) and within secondary structure elements.

Helices appear as a thicker diagonal in the MI matrix (i.e. MI between primary sequence neighbors is higher within a helix than in non-helical regions) and residues within a helix are amongst the most correlated residue pairs (Figure 1A,B). The exception to this typical helical pattern is the I-helix, the longest helix which traverses the entire protein and includes residues that participate in catalysis [3]. MI in the I-helix reveals residues that move in three independent groups, rather than the typical one coupled group per helix (see Supplemental Figure 1 for a close up figure of this region). The regions at the N- and C-terminal ends of the I-helix show coupled dynamics that are consistent with other helices, although there is little coupling between these two regions. Residues in the middle region of the I-helix, a conserved region which includes all catalytic residues, move relatively independently from the rest of the residues in the I-helix as well as the rest of the enzyme (rather than the more typical coupled motion seen between residues in the same helix). The I-helix of cytochrome P450's often contain a kink in the alpha helical structure in this middle region [3].

Beta strands appear as a line in the MI matrix (for example with antiparallel beta strands, residue X is coupled with residue Y, residue X+1 is coupled with residue Y-1, etc.) and residues between beta strands are amongst the most correlated residue pairs (Figure 1A,B). The three off-diagonal regions with high MI_{norm} pairs are circled in Figure 1B and colored red on the ligand-free crystal structure (1IO9 [24]) in Figure 1C. These regions mainly surround the beta strand secondary structure.

The second highest correlated motion is seen with “hot spot” residue pairs that are dispersed throughout the structure and connect between different secondary structure elements (green in Figure 1B, orange in Figure 1C). “Hot spots” are defined as residue pairs that have high MI_{norm} and cluster in 3D space, but are not adjacent in the primary sequence. The largest cluster of hot spots occurs at the base of the F and G helices. The loop between these two helices, the FG-loop, is the most dynamic region of the enzyme [3, 39] and indeed this region is the most dynamic region in our simulation. Motion of this loop in and out of the active site renders the enzyme in either a “closed” or “open” state [39, 40]. In our simulation the FG loop moves independently from the rest of the enzyme (low MI_{norm}), providing an explanation for previous crystallographic evidence that shows little movement in the rest of the enzyme between the open (ligand-free [24]) and closed (ligand-bound [41]) states. Loop regions are also in tightly coupled MI hot spots, including loops near the active site. Previous studies provide data that is consistent with hypotheses generated from coupled motion in the dynamics. For example, directed evolution towards a more active form of the cytochrome P450 enzyme vitamin D₃ hydroxylase [42] results in four mutations that are far from the active site, shifting the conformational equilibrium of the enzyme towards the closed state. One mutation is located in the hot spot region we identified in CYP119 at the base of the F and G helices, and the other three mutations are within a few residues of other hot spots we identified. One hypothesis consistent with this data is that mutation of hot spot residues results in structural changes, and further studies in which residues both near and far from hot spots are mutated would be of great interest.

3.3 Residues can be clustered into 23 coupled units

The 368×368 residue MI_{norm} matrix (Figure 1A) can be simplified by clustering residues with coupled dynamics together. K-means clustering results in 23 coupled units of the enzyme that move together (see *Materials and Methods* for details). Intra-cluster residues are more coupled to each other than to residues outside the cluster. Figure 2A shows the 23×23 matrix where each element is the average MI_{norm} between all residues in each cluster. Figure 2B shows each coupled unit on the structure.

The K-means algorithm clusters the heme cofactor together with the catalytic residues in the I-helix. Residues in the F/G loop are also clustered together in a separate cluster from the catalytic residues. The fact that functionally related residues are clustered together supports the relevance of clustering based on dynamics. The K-means algorithm also clusters the I-helix clusters into the same three regions previously described by the I-helix dynamics (see Section 3.1).

3.3 Inter-cluster coupling reveals a tightly coupled “outer shell” and more loosely coupled active-site and regions involved in ligand and redox partner interactions

The K-means algorithm clusters residues that are most coupled to each other (coupled units of residues that move together), and the average MI_{norm} between these clusters is also calculated (off-diagonal elements in Figure 2A). All coupled units with average $MI_{\text{norm}} > 0.3$ between them are shown with surface representation in Figure 3. Tight coupling between coupled units is seen around the solvent-exposed surface of the enzyme, whereas coupled units in the active site move more independently from their neighboring coupled units. These more loosely coupled residues include catalytic residues, the heme cofactor, the dynamic F and G helices, the FG loop implicated in ligand binding, and the B, B', C helices implicated in ligand egress. Previously identified egress channels are all located in the more loosely coupled regions [43]. The proximal surface of CYP119 has been implicated in interactions with redox partners [44]. All of the proximal surface is part of the tightly coupled “outer shell” except for one cluster that contains the proximal cysteine ligand, Cys117. Thus, surfaces that are involved in interacting with ligand or ligand binding are more loosely coupled than the rest of the enzyme.

3.4 The active site moves relatively independently from the rest of the enzyme

The active site of CYP119 is a deep, hydrophobic cavity in the center of the enzyme which contains the heme cofactor. The iron-coordinated cysteine thiolate ligand (Cys317) is located on the proximal side of the heme, whereas the catalytic residues of the I-helix and substrate binding residues in the B, B', and C helices are located on the distal side of the heme. MI correlation analysis suggests that active site residues move relatively independently from the rest of the enzyme. Figure 4 shows the average MI_{norm} within 5 Å for each residue, with residues that are highly coupled to their neighboring residues colored red and residues that move relatively independently from their neighbors colored blue. Residues in the active site, including the heme cofactor, catalytic residues, and F/G loop are the least coupled to their neighboring residues. Figure 5 shows the mean MI_{norm} between each residue and the heme cofactor (high MI_{norm} colored orange, low MI_{norm} colored blue). The catalytic residues in the I-helix, the proximal cysteine thiolate ligand, and a few residues in active site loops are the most coupled to the heme cofactor.

4. Discussion

Correlated motion between protein residues is essential for functions such as interaction with ligands, catalysis, and allostery [14, 36, 45]. Here, we have demonstrated the utility of the mutual information technique (MI) to analyze correlated motion in a 200 ns molecular

dynamics simulation of P450CYP119, a thermostable cytochrome P450 from *Sulfolobus acidocaldarius*. We have previously established that a 200 ns molecular dynamics simulation of the same enzyme captures transitions between known conformations in the absence of ligand [38], suggesting that functionally relevant dynamics occur on this timescale.

P450 enzymes exhibit profound conformational plasticity [3]. In particular, the F/G loop (residues 152–163) is a dynamic region on the distal side of the catalytic site that has been implicated in ligand recognition and binding. Prior crystallographic and NMR studies have demonstrated that this region undergoes dramatic conformational rearrangement in and out of the active site to create an open [24] and closed [41] state. These structural rearrangements adjust the volume and the electrostatic surface potential of the active site so that it may accommodate different ligands [3, 39]. Indeed, in our simulation this region exhibits the greatest flexibility [38]. MI reveals that the F/G loop has low MI compared to other residues in the enzyme (the loop moves independently rather than in a coupled fashion), and thus the loop is able to move in and out of the active site without affecting the positions of other residues. F/G loop dynamics thus provide an explanation for previous crystallographic evidence that shows little movement in the rest of the enzyme between the open and closed states.

MI patterns in other regions of the enzyme can also be linked to structural and functional properties of the enzyme. For example, residues that are structurally related through covalent bonds (e.g. adjacent in the peptide chain) or side-chain interactions (e.g. between beta strands) have high MI (coupled motion). P450 secondary structure is predominantly helical and residues within the same helix generally have coupled dynamics, with the I-helix being a notable exception. We observe several regions in which there is a cluster of coupled motion between residues that are not adjacent in a peptide chain and we term these regions “hot spots”. While determining the significance of these hot spots requires additional study, it is intriguing to note experimental evidence that mutations near these hot spots affect catalysis in a related P450 CYP, despite the fact that they are far from the active site [42].

We reduce the complexity of the 368×368 MI matrix (Figure 1) by clustering together residues with coupled dynamics using a modified k-means algorithm. A total of 23 clusters emerge from this analysis (Figure 2). The functional significance of the clusters is suggested by the fact that the catalytic residues in the I-helix are in the same cluster as the heme co-factor. Coupling between the clusters (off diagonal elements in the 23×23 matrix) reveals how much coupling there is between different regions of the enzyme. MI patterns reveal that clusters with the highest amount of coupling between them form a surface-exposed “shell” around the enzyme (Figure 3). Clusters that are more loosely coupled to other clusters occur in the inner regions of the enzyme (catalytic site), regions involved in ligand interactions (distal side), and regions involving redox partner interactions (proximal side) (Figure 3). On the distal side, ligand binding loops (the F/G loop) [39] and all the ligand egress channels (near the B, B' and C loops) [43] occur in regions where the clusters are more loosely coupled to one another. The cluster that is loosely bound to other clusters on the proximal side contains Cys317, the cysteine thiolate ligand that is bound to the heme co-factor. While the significance of these coupling patterns will need further study, one hypothesis that emerges is that the enzyme is able to bind to ligands and protein partners without affecting the dynamics in other regions of the enzyme.

A striking feature of the dynamics is that the enzyme active site moves independently from the rest of the protein (Figures 4 and 5). This implies that many of the functionally significant conformational changes that take place during the catalytic process are limited to a subset of residues clustered within the active site itself. It has previously been proposed

that, because of the reactive nature of the perferryl species, any substrate that enters the active site is likely to undergo oxidation to some degree [3]. In this context, much of the apoprotein simply serves as a scaffold to maintain a productive reactive environment around the heme prosthetic group. Dynamics in the I-helix support the ability of active site residues to move independently. Whereas typically residues within the same helix have high MI to one another, residues in the I-helix do not. Residues in the center region of the I-helix, including the catalytic threonine residues Thr213,215 move relatively independently from one another and from the other residues in the I-helix. P450 I-helices contain structurally conserved “kinks” (bond angles that are not typical of helices) in the center region of the I-helix. It is possible that this conserved structural mechanism is related to the ability of these catalytic residues to move without influencing the positions of non-catalytic residues.

This study provides evidence that conformational dynamics in CYP119 are linked to function. Functionally related residues move together. An active site that moves independently from the rest of the protein indicates that catalytic activity will be less sensitive to non-catalytic events such as interactions with other proteins in the crowded cellular environment. Segregation of the catalytic core from the rest of the protein is attractive in functional terms and is likely to occur in other proteins. Only further research will determine whether it is a property of thermophilic P450 enzymes, a broad range of P450 enzymes, heme proteins in general, or even a general characteristic of enzyme function.

5. Conclusion

Enzyme dynamics are intricately related to function and residues with correlated motion are likely to be functionally related. By building computer simulations of enzyme dynamics in all-atom detail, we can begin to build hypotheses for the role of dynamics in function. Correlated motions in our simulation suggest that active-site dynamics are decoupled from those of the rest of the protein, effectively insulating the catalytic machinery from interactions at the protein surface.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Jonathan B. Johnston for helpful feedback on the manuscript. This work was supported by grant GM25515 from the National Institutes of Health.

Abbreviations used

CYP	cytochrome P450
MI	mutual information
MD	molecular dynamics simulations

References

1. Guengerich FP. Chem Biol. 2009; 16:1215–1216. [PubMed: 20064429]
2. Schlichting I, Berendzen J, Chu K, Stock AM, Maves SA, Benson DE, Sweet RM, Ringe D, Petsko GA, Sligar SG. Science. 2000; 287:1615–1622. [PubMed: 10698731]
3. Ortiz de Montellano, PR. Cytochrome P450: Structure, Mechanism, and Biochemistry. Kluwer Academic/Plenum Publishers; New York: 2005.
4. Guengerich FP. Aaps J. 2006; 8:E101–111. [PubMed: 16584116]

5. Guengerich FP. *Nat Rev Drug Discov.* 2002; 1:359–366. [PubMed: 12120411]
6. Chemier JA, Fowler ZL, Koffas MA, Leonard E. *Adv Enzymol Relat Areas Mol Biol.* 2009; 76:151–217. [PubMed: 18990830]
7. Li S, Wackett LP. *Biochemistry.* 1993; 32:9355–9361. [PubMed: 8369306]
8. Serizawa N, Nakagawa K, Hamano K, Tsujita Y, Terahara A, Kuwano H. *J Antibiot (Tokyo).* 1983; 36:604–607. [PubMed: 6874577]
9. Andersen JF, Hutchinson CR. *J Bacteriol.* 1992; 174:725–735. [PubMed: 1732208]
10. Schmidt-Dannert C, Umeno D, Arnold FH. *Nat Biotechnol.* 2000; 18:750–753. [PubMed: 10888843]
11. Jennewein S, Croteau R. *Appl Microbiol Biotechnol.* 2001; 57:13–19. [PubMed: 11693909]
12. Lohmann W, Karst U. *Anal Bioanal Chem.* 2008; 391:79–96. [PubMed: 18163163]
13. Gotardo MC, Moraes LA, Assis MD. *J Agric Food Chem.* 2006; 54:10011–10018. [PubMed: 17177535]
14. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D. *Nature.* 2005; 438:117–121. [PubMed: 16267559]
15. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. *Nature.* 2007; 450:913–916. [PubMed: 18026087]
16. Bakan A, Bahar I. *Proc Natl Acad Sci U S A.* 2009; 106:14349–14354. [PubMed: 19706521]
17. Benkovic SJ, Hammes GG, Hammes-Schiffer S. *Biochemistry.* 2008; 47:3317–3321. [PubMed: 18298083]
18. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE. *Curr Opin Struct Biol.* 2009; 19:120–127. [PubMed: 19361980]
19. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KF, Becker S, Meiler J, Grubmuller H, Griesinger C, de Groot BL. *Science.* 2008; 320:1471–1475. [PubMed: 18556554]
20. Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson MA, Petsko GA, Karplus M, Hubner CG, Kern D. *Nature.* 2007; 450:838–844. [PubMed: 18026086]
21. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, Jensen MO, Dror RO, Kuriyan J, Shaw DE. *Proc Natl Acad Sci U S A.* 2009; 106:139–144. [PubMed: 19109437]
22. Swaminathan S, Harte WE Jr, Beveridge DL. *Journal of the American Chemical Society.* 1991; 113:2717–2721.
23. Arnold GE, Ornstein RL. *Biophys J.* 1997; 73:1147–1159. [PubMed: 9284282]
24. Park SY, Yamane K, Adachi S, Shiro Y, Weiss KE, Maves SA, Sligar SG. *J Inorg Biochem.* 2002; 91:491–501. [PubMed: 12237217]
25. The PyMOL Molecular Graphics System, Version 1.2r3pre. Schrödinger, LLC;
26. Hess B, Kutzner C, van der Spoel D, Lindahl E. *J Chem Theory Comput.* 2008; 4:435–447.
27. Jorgensen WL, Tirado-Rives J. *J Am Chem Soc.* 1988; 110:1657–1666.
28. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79
29. Bussi G, Donadio D, Parrinello M. *J Chem Phys.* 2007; 126:014101. [PubMed: 17212484]
30. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. *Journal of Computational Chemistry.* 1998; 18:1463–1472.
31. Darden T, York D, Pedersen L. *J Chem Phys.* 1993; 98:10089.
32. Gogonea V, Shy JM 2nd, Biswas PK. *J Phys Chem B.* 2006; 110:22861–22871. [PubMed: 17092038]
33. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR. *J Chem Phys.* 1984; 81:3684–3690.
34. Cover, TM.; Thomas, JA. *Elements of Information Theory.* Wiley-Interscience; 1991.
35. Meitzler JL, Brandman R, Ortiz de Montellano PR. *J Biol Chem.* 285:40991–41000. [PubMed: 20947510]
36. Lange OF, Grubmuller H. *Proteins.* 2008; 70:1294–1312. [PubMed: 17876828]
37. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP. *J Chem Theory Comput.* 2009; 5:2486–2502. [PubMed: 20161451]

38. Lampe JN, Brandman R, Sivaramakrishnan S, de Montellano PR. *J Biol Chem*. 2010; 285:9594–9603. [PubMed: 20097757]
39. Nishida CR, Ortiz de Montellano PR. *Biochem Biophys Res Commun*. 2005; 338:437–445. [PubMed: 16139791]
40. Savino C, Montemiglio LC, Sciara G, Miele AE, Kendrew SG, Jemth P, Gianni S, Vallone B. *J Biol Chem*. 2009; 284:29170–29179. [PubMed: 19625248]
41. Yano JK, Koo LS, Schuller DJ, Li H, Ortiz de Montellano PR, Poulos TL. *Journal of Biological Chemistry*. 2000; 275:31086–31092. [PubMed: 10859321]
42. Yasutake Y, Fujii Y, Nishioka T, Cheon WK, Arisawa A, Tamura T. *J Biol Chem*. 285:31193–31201. [PubMed: 20667833]
43. Cojocaru V, Winn PJ, Wade RC. *Biochim Biophys Acta*. 2007; 1770:390–401. [PubMed: 16920266]
44. Chang YT, Loew G. *Biochemistry*. 2000; 39:2484–2498. [PubMed: 10704198]
45. Berendsen HJ, Hayward S. *Curr Opin Struct Biol*. 2000; 10:165–169. [PubMed: 10753809]

Figure 1.

Correlated motion between residue pairs as measured by normalized mutual information (MI_{norm}). (A) MI_{norm} between all residues pairs. The axes are labeled by the position of the helices. MI_{norm} between a residue and itself (the diagonal line, by definition the highest MI_{norm} of 1) has been removed from the color spectrum to facilitate viewing of the off-diagonal pairs. (B) Only the highest (red, $MI_{\text{norm}} > 0.35$) and second highest MI_{norm} pairs (green, $0.3 < MI_{\text{norm}} \leq 0.35$) are highlighted. Circled in yellow are the most correlated off-diagonal residues, generally regions with beta strand secondary structure. (C) Residue pairs that are not neighbors in the primary sequence and have the highest (red) and second highest (orange) MI_{norm} are highlighted on the ligand-free CYP119 structure used as the starting structure for molecular dynamics (11O9 [24]). View is from the distal side of the enzyme, and rotated to highlight the base of the F and G helices in (D).

Figure 2.

A modified k-means algorithm was used to cluster residues based on correlated motion. (A) 23 clusters emerge, and the simplified 23×23 matrix of the MI_{norm} within and between clusters is shown. (B) The residues in each of the 23 clusters are colored to illustrate their positions in the structure. Functionally related residues, for example the catalytic threonines and the heme co-factor, are in the same cluster. Cluster numbers corresponding to Panel A are labeled on the structure.

Figure 3.

Coupling between clusters (the off-diagonal elements in the 23×23 matrix in Figure 2) is indicated in a view of the structure from the distal (A) and proximal (B) sides of the enzyme. Each of the 23 clusters has a different color, as in Figure 2. Clusters that are highly coupled to neighboring cluster are shown in a surface representation, whereas clusters that are more loosely coupled to neighboring clusters are shown in cartoon representation. Tightly coupled clusters form a surface-exposed “shell” around the enzyme. Regions that have more loosely coupled clusters are involved in ligand interactions, redox partner interactions, and catalysis.

Figure 4.

The mean MI_{norm} between each residue and all neighboring residues within 5 Å (high MI_{norm} is red, low MI_{norm} is blue/green). The heme cofactor and residues in the catalytic site have the lowest MI_{norm} , indicating that these residues move relatively independently from the rest of the enzyme.

Figure 5.

The mean MI_{norm} between each residue and the heme cofactor in the active site (high MI_{norm} is orange, low MI_{norm} is blue). Residues most coupled to the heme cofactor are in the I-helix (distal side), including the catalytic residues, and in the loop containing the 317Cys thiolate ligand (proximal side).