Molecular Dynamics Study of Time-Correlated Protein Domain Motions and Molecular Flexibility: Cytochrome P450BM-3

Gregory E. Arnold and Rick L. Ornstein
Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352 USA

ABSTRACT Time-correlated atomic motions were used to characterize protein domain boundaries from atomic coordinates generated by molecular dynamics simulations. A novel application of the dynamical cross-correlation matrix (DCCM) analysis tool was used to help identify putative protein domains. In implementing this new approach, several DCCM maps were calculated, each using a different coordinate reference frame from which protein domain boundaries and protein domain residue constituents could be identified. Cytochrome P450BM-3, from *Bacillus megaterium*, was used as the model protein in this study. The analyses indicated that the simulated protein comprises three distinct domain regions; in contrast, only two protein domains were identified in the original crystal structure report. Specifically, the DCCM analyses showed that the F-G helix region was a separate domain entity and not a part of the α domain, as previously designated. The simulations demonstrated that the domain motions of the F-G helix region effected both the size and shape of the enzyme active site, and that the dynamics of the F-G helix domain could possibly control access of substrate to the binding pocket.

INTRODUCTION

Protein structure is classified into a hierarchy of fundamental elements. These elements interact to form assemblies, and the assemblies coalesce to generate a cooperatively folded, three-dimensional structure. The domain element is the highest level of substructural classification in the hierarchical organization of a protein folded from a single polypeptide chain. The protein molecule may consist of either a single or multiple domain element/s.

Large-scale conformational transitions in proteins can occur as a result of relative movement between domains. This type of motion could conceivably produce a wide range of protein conformations with only minor expenditures in energy. Flexibility in short segments of the protein backbone allow corresponding domain motions to occur, with only minor structural perturbations in the domains themselves. Thus each domain is in essence a structurally rigid entity composed of amino acid residues that move both cooperatively and coherently.

Much of what is known about protein domain motion and how it relates to enzyme function has been learned by comparing the x-ray crystal structures of protein allomorphs (Gerstein et al., 1994). Computer simulation of protein atomic motions is a relatively new method that can be used to study the underlying cooperative atomic motions that lead to rigid body-like domain motions. The collective motions of a simulated protein can be identified by analyzing both the covariance and time correlation in the positional fluctuation of its atoms (Ichiye and Karplus, 1991). A dynamic cross-correlation matrix (DCCM) is a three-dimensional matrix representation that graphically displays time-correlated information among the residues of the protein; the DCCM residue-based time-correlation data can be quickly and comprehensively analyzed by visual pattern recognition (Swaminathan et al., 1991).

In this study, the DCCM analysis tool was applied in a novel way to identify protein domains. This was accomplished by examining the time-correlated atomic motions from a series of different coordinate reference frames (see below). This approach is different from past DCCM analysis in which only a single coordinate reference frame was used, that being a global superpositioning of all molecular dynamics (MD) snapshot structures on either a time-averaged MD structure or the incipient x-ray structure. This type of global superpositioning provides little or no information with regard to time-correlated protein interdomain motions. The new use of multiple spatial reference frames provides additional information regarding the dynamics of the protein that has not previously been demonstrated. The utility, in using multiple coordinate reference frames, is in the additional data gleaned, which allows for the characterization of time-correlated interdomain motions and the delineation of their boundaries.

METHODS

Molecular dynamics simulations were calculated separately for each of the two allomorphs in the asymmetrical unit of the cytochrome P450BM-3 crystal structure (Ravichandran et al., 1993). Each simulation was 200 ps in length, and an explicit 10-Å solvent layer was used to hydrate the protein. This brought the aggregate number of atoms in the systems to 19,135 and 19,168 for P450BM-3 molecules 1 and 2, respectively. The model system and molecular dynamics simulation data used in the present work are presented elsewhere (Paulsen and Ornstein, 1995). The authors have expended a substantial effort in developing a protocol for computing protein simulations (of modest duration in a water droplet model, such as those used here) and have demonstrated that these types of MD simulations

Received for publication 3 March 1997 and in final form 30 May 1997.

Address correspondence to Dr, Gregory E. Arnold, Biological Information Technologies, Inc., P.O. Box 1403, Richland, WA 99352. Tel.: 509-376-4423; Fax: 509-376-6767; E-mail: ge_arnold@pnl.gov.

Address reprint requests to Rick L. Ornstein, Fax: 509-375-6904; E-mail: rl_ornstein@pnl.gov.

© 1997 by the Biophysical Society

0006-3495/97/09/1147/13 $2.00

are reliable and reproducible and are carried out with suitable precautions (Arnold and Ornstein, 1994).

Solvent-accessible surface area calculations, molecular volumes, and active-site volumes were all calculated as implemented by the program GRASP (Nicholls et al., 1991). Solvent-accessible surface areas were calculated using a probe size of 1.4 Å. Active-site volumes were calculated as follows: 1) backbone atoms of residues 17, 21, 44, 47, 73, 76, 77, 188, and 189 were used to approximate an annulus, which served as the outer boundary of the active site; 2) atoms below this annulus projecting into the binding pocket were used to calculate the surface of the active site; and 3) the resulting surfaces, invaginating into the protein from this annulus, were used to define the active-site surface.

For the dynamic cross-correlation analyses (discussed below), time-averaged MD structures were calculated over the final 125 ps of the respective simulation. For all other analyses, time-averaged MD structures were calculated over the final 50 ps of the respective trajectory.

RESULTS AND DISCUSSION

MD simulations

The general stability and the convergence characteristics of the cytochrome P450BM-3 molecular dynamics trajectories are described elsewhere (Paulsen and Ornstein, 1995). Some of the properties monitored over the time course of the simulation included protein radius of gyration, atomic rms deviations away from the crystal structure coordinates, solvent-accessible surface area, and rms fluctuations in atomic position averaged on a per-residue basis. Based on these analyses, the trajectories were determined to be both stable and representative of the average experimental structures. The authors will provide, upon request, additional information with regard to these data that pertains to the quality of the MD simulations. The differences in molecular flexibility between the two trajectories and the delineation of P450BM-3's domain boundaries are described herein, as ascertained from the MD simulations. (In the ensuing discussion, we will refer to P450BM-3 M1 and M2 crystal structures as M1C and M2C, respectively, and P450BM-3 molecules 1 and 2 time-averaged MD structures as M1A and M2A, respectively. The following abbreviations will also be used: M1, molecule 1; M2, molecule 2; RF, reference frame.)

Time-correlated atomic motions

Dynamic cross-correlation maps (Ichiye and Karplus, 1991; McCammon and Harvey, 1987) were used to detect time-correlated motions in the protein as implemented by Swaminathan et al. (1991). Dynamic cross-correlation matrices were calculated by the expression

$$C_{ij} = \langle \Delta r_i \cdot \Delta r_j \rangle / (\langle \Delta r_i^2 \langle \Delta r_j^2 \rangle)^{1/2} \quad (1)$$

where r_i and r_j are the spatial backbone atom positions of the respective ith and jth amino acids. The spatial coordinate of an individual amino acid is calculated from the geometric mean of its respective N, C_α, and C backbone atom coordinates. In Eq. 1, Δr_i corresponds to the displacement of the ith residue from its mean position over the time interval.

There is a time scale associated with each C_{ij} element, and this time scale corresponds to a subset of contiguous snapshot structures taken from the temporal series of snapshot structures saved from the MD trajectory. This time scale (subset of structures) determines the time interval over which the C_{ij} elements are calculated (Swaminathan et al., 1991).

The C_{ij} elements of the matrix are symmetrical about the diagonal; however, for clarity, only negative correlation values are displayed in the upper triangle and only positive correlation values are displayed in the lower triangle (Figs. 2–5). Because Eq. 1 is normalized, the calculated C_{ij} values indicate only the direction of the time-correlated atomic motions, and not the magnitude of their fluctuations. Positive C_{ij} values result from backbone atom motions between residues i and j that are in the same direction along a given spatial coordinate; negative C_{ij} values result from backbone atom motions between residues i and j that are opposite in direction along a given spatial coordinate. Furthermore, because of the scalar product in the numerator of Eq. 1, pairwise backbone atomic motions whose directions are orthogonal with respect to each other will result in $C_{ij} = 0$. As a result, these types of time-correlated motions will not be detected and are a limitation of the DCCM method. However, it is unlikely that relative domain movements within a protein will consist only of orthogonal motions; hence this limitation should not obscure the detection of time-correlated domain motions by the DCCM method.

Spatial reference frame

There are translational and rotational components of motion associated with each snapshot structure saved from the MD trajectory. This, in effect, places each structure in a different coordinate frame of reference. To obtain meaningful results from the DCCM analysis, a common spatial reference for all structures must first be chosen.

In previous protein dynamics studies using DCCM analysis, a spatial reference frame was typically selected by using the backbone atoms of either x-ray crystal structure or the time-averaged MD structure as a template on which all MD snapshot structure were superimposed (Brunger et al., 1985; Harte et al., 1990; Ichiye and Karplus, 1991; Swaminathan et al., 1991; Venable et al., 1993). This type of global superpositioning can potentially obscure time-correlated motions of a protein, particularly those motions that involve many residues over large segments of the protein backbone (in Fig. 1, a simple schematic is used to illustrate this point). However, the use of a global reference frame does allow for the identification of higher order localized motions within the protein, albeit at the expense of potentially obscuring time-correlated domain motions of the protein.

The inherent limitation of using a single global reference frame can be overcome by calculating several DCCM maps using a series different coordinate reference frames. In this method, a specific spatial reference frame is selected by

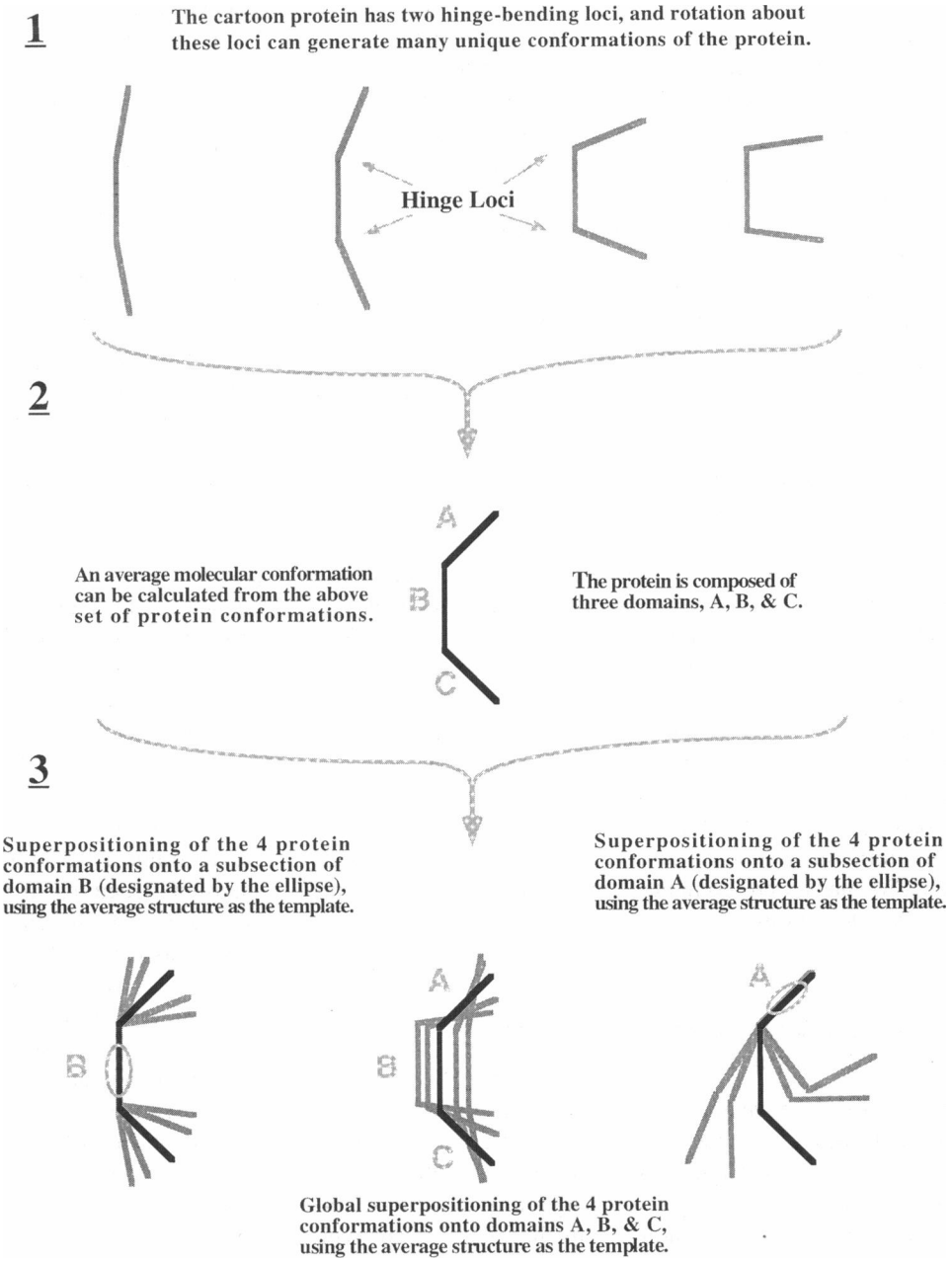

FIGURE 1 A schematic representation illustrating three different spatial reference frames for a set of protein conformations. The protein is depicted in cartoon form, using a stick figure representation. The protein is composed of three different stick segments, each of which represents a unique protein domain (domains A, B, and C). (*1*) Different protein conformations (*gray*) are generated by rotating the domains at the vertices of the sticks (the hinge-bending loci). (*2*) An average protein conformation (*black*) is determined from the four gray conformations. (*3*) A spatial reference frame is selected by superimposing the gray colored conformations on the average protein structure. Different spatial reference frames can be selected by superimposing the gray protein conformations on different regions of the average structure. Global superimposition of the gray conformations on the average structure could potentially obscure time-correlated domain motions. Superimposition of the gray conformations on a subregion of a domain (using the average structure as a template) will mask the time-correlated motions within that domain; however, the time-correlated motions of other domains could potentially be delineated. The use of multiple reference frames aids in the identification of domain boundaries.

superimposing MD snapshot structures on a particular subsection of the protein (Fig. 1). Different spatial reference frames are then selected by superimposing MD snapshot structures on different subsections of the protein.

In general, for a residue to be part of a domain, it must exhibit concerted atomic motions between itself and all other residues within the given domain. By superimposing MD snapshot structures on a subset of residues within a domain, the domain itself becomes a stationary point of reference (Fig. 1). This, in effect, masks the coherent motions of the residues within the reference domain, and as a result, no time-correlated motions will be detected among the residues comprising the reference domain. Moreover, all time-correlated motions that are detected in the DCCM map are relative to the stationary reference domain. Any strong positive cross-correlated motions that are observed will correspond to those domain motions other than the domain used as a stationary reference.

Given a particular subset of residues as the stationary point of reference, either 1) all other regions of the protein that exhibit no cross-correlated motions with the reference residues are members of the same protein domain as the reference residues, or 2) the motional properties of these residue are not coherent and/or are not time-correlated with any other part of the protein. Next, by selecting a subset of residues from a different coordinate reference frame from which a new DCCM map is calculated, the resulting data can be used to distinguish between these two possibilities.

Those regions of the protein that previously exhibited no cross-correlated motions, but now do, are members of the original domain; those regions of the protein that again show no cross-correlated motions are not members of the original reference domain. Furthermore, as we compute several DCCM maps using multiple reference frames, the resulting data can then be used to characterize putative domain boundaries and identify the residue constituents of a particular domain.

DCCM maps were calculated from four different reference frames as described below. Furthermore, there is a time scale associated with each C_{ij} element, and a time series of DCCM maps was calculated using 5-, 15-, 25-, 40-, 80-, and 125-ps time blocks. For each 5-, 15-, 25-, and 40-ps time block, two different DCCM maps were calculated from MD structures at 75 ps and 150 ps as starting points for the respective time series elements.

DCCM maps were calculated from trajectory time points greater than 75 ps to allow the simulated proteins as much time as possible to thermally equilibrate and enter into coherent modes of motion. Given the 200-ps length of the trajectories, only single DCCM maps could be computed for the 80- and 125-ps time block calculations because of the large number of MD structures needed for these two calculations. Had two starting points been used, the resulting matrices would have had to have been calculated from overlapping MD structures and/or MD structures from the equilibration portion of the trajectories. Thus only single DCCM maps were computed for the 80- and 125-ps time block calculations.

Time-correlated protein secondary structure motions

Interactions between different portions of the protein through the time course of the simulation are evident from the DCCM maps. High-density positive correlations along the diagonal result from nearest-neighbor interactions. Perpendicular plumes emanating from the diagonal are typically diagnostic of correlated motions between antiparallel β-strands. For example, in Fig. 3 the intense black plume emanating at and around residues 330–355 indicates time-correlated motions between strands 3 and 4 from β-sheet 1. Also evident are strong off-diagonal peaks that exhibit positive correlations between these residues and residues distally located at positions 38–44, 47–53, and 66–70. These distal residues correspond to strands 1, 2, and 5 from β-sheet 1. Triangular shapes that run along the diagonal are generally indicative of correlated motions from helices. For example, the large positive triangular block extending from residues ~265–325 corresponds to correlated atomic motions among the atoms in the contiguous series of helices I, J, J′, and K. These types of patterns are expected, because regular units of secondary structure move in concert. Several positive cross-correlation peaks between sequentially noncontiguous regions of the protein are also evident, and in general are found between spatially proximate regions within a particular domain. For example, the intense blocks of positive cross-correlation found at and around residues ~283–300 and 410–420 correspond to motions between spatially proximate portions of the J and L helices, both of which are located within the α domain.

Time-correlated substrate-binding pocket motions

Anticorrelated motions between residues 45 and 191 (shown in *green*, Fig. 6) on opposite sides of the mouth of the binding pocket have previously been identified (Paulsen and Ornstein, 1995). This behavior was observed only for the M2 trajectory and not for the M1 trajectory. This particular area of the binding pocket (the loop between the F and G helices and the turn between strands 1 and 2 from β-sheet 1) has the potential to act as mobile flaps covering the active site. To further investigate these specific motions and other time-correlated motions involving the binding pocket residues, DCCM maps were calculated for both the M1 and M2 trajectories. The reference frame was chosen such that residues that are part of the binding pocket region were free to move relative to the rest of the protein. For these analyses, the time-averaged MD structure was used as the superpositioning template, and the reference frame (reference frame 1, RF1) was selected by RMS fitting all MD snapshot structures to the core secondary structure backbone atoms of the entire protein, except for those atoms defined as part of the binding pocket.

The DCCM map calculated for M2 using RF1 and a 25-ps time block is shown in Fig. 2. Of particular interest are the negative cross-correlation peaks. A periodic change in displacement across the mouth of the binding pocket was previously measured between residues 45 and 191 (*green residues* in Fig. 6) (Paulsen and Ornstein, 1995). This distance varied between ~11 Å and ~18 Å in the M2 trajectory; however, only subtle changes in distance were observed for these same residues in the M1 trajectory. A strong negative cross-correlation peak is evident between the loop extending from the F and G helices (residues 191–206) and strands 1 and 2 from β-sheet 1 (residues 38–53). A weaker cross-correlation peak is observed between the F and G helix loop region and strand 5 from β-sheet 1 (residues 66–70). A strong negative cross-correlation peak is also observed between the F and G helix loop region and strand 2 from β-sheet 2 (residues 344–348). There is a time scale associated with these peaks, and these peaks were most intense for the 25-ps time block calculation. These same peaks were significantly weaker for the 15-ps and 40-ps time block calculations, and were not detectable in any other time scales used. These peaks had the same general shapes regardless of the starting structure used; however, there was some variation in peak intensity, depending on the starting structure used in the calculation. The variation in intensity is indicative of anharmonicities

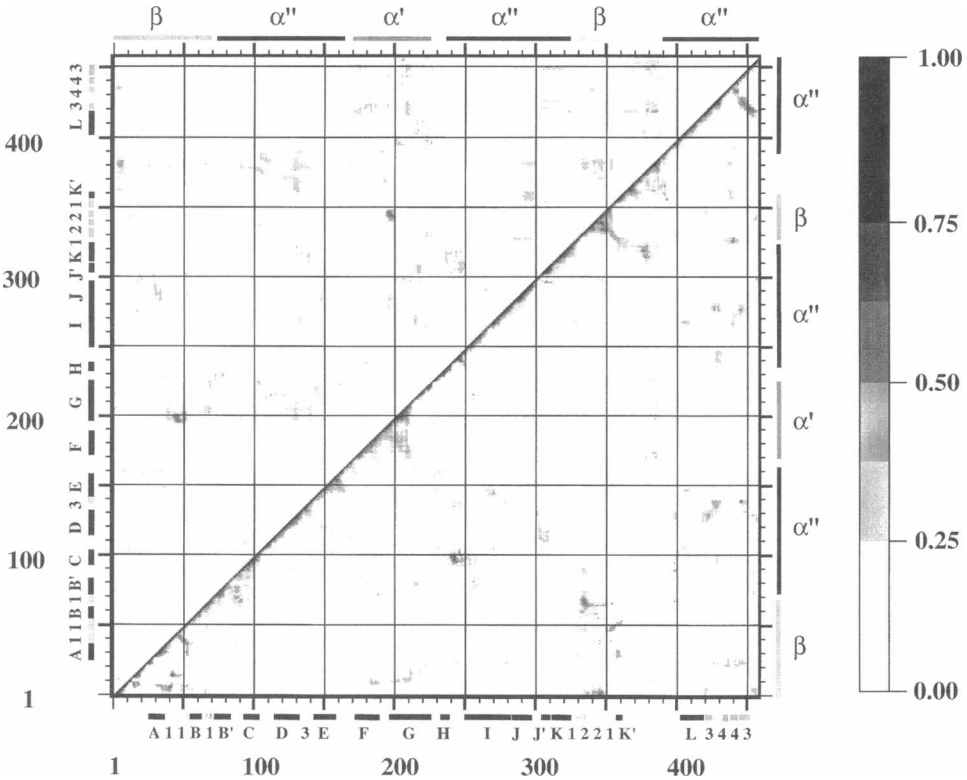

FIGURE 2 Dynamic cross-correlation map of the time-correlated backbone atom motions of trajectory M2 calculated with MD structures from the 150–175-ps time interval. RF1 is the coordinate reference frame. Only cross-correlations greater than 0.25 are shown; the size of the cross-correlation is proportional to the shading, as indicated by the scale at right. Negative correlations are shown in the upper triangle, and positive correlations are shown in the lower triangle. The bars along the bottom and left borders indicate positions of secondary structure elements. Black bars correspond to α-helices and are labeled with a capital letter; gray bars correspond to β-sheets and are labeled with their numeric designation. The bars along the top and right borders indicate positions of the putative protein domains.

associated with the particular atomic motions. These peaks were found only in the DCCM maps calculated from the M2 trajectory and were not observed in the DCCM maps calculated from the M1 trajectory. All other negative cross-correlation peaks varied in position, depending on which of the two starting structures was used, and as such are either not harmonic or not persistent motions with respect to the time series used in the DCCM calculations.

Time-correlated domain motions

In the initial report describing the x-ray structure of P450BM-3, two domains (α and β, Table 1) were identified (Ravichandran et al., 1993), and the F and G helices were classified as parts of the α domain. The authors also observed two allomorphs (structures M1 and M2) in the unit cell of the crystal. The conformational difference between structures M1 and M2 was described as rigid body rotations of ~5.0° for the β domain (*yellow backbone*, Fig. 6) and ~4.6° for the F-G helix region (*white backbone*, Fig. 6), relative to the rest of the protein. These data suggest that the F-G helix region could potentially move in concert and independently from the rest of the protein, in effect, behaving as an independent domain. To explore this possibility, DCCM maps were computed from the following additional three reference frames: the F and G helices (RF2), the β domain (RF3), and the α domain except for the F and G helices (RF4). For clarity, we will refer to the F and G helix region as the α′ domain (*white backbone*, Fig. 6) and refer to the previously defined α domain, minus the F and G helix region, as the α″ domain (*steel blue backbone*, Fig. 6). For these analyses, MD snapshot structures were superimposed on the average dynamics structure, using the backbone atoms within the respective reference frame.

TABLE 1 Domain definitions

	From crystal structure data (residue position)	From MD simulation data (residue position)
α Domain	72–325, 390–457	—
β Domain	1–70, 329–361	1–70, 84–92, 329–361, 387–401
α′ Domain	—	172–226 (F and G helix region), 233–239 (H helix), 250–264 (15 N-terminal I helix residues)
α″ Domain	—	138–141, 266–325, 402–457
Not classified	362–389	362–389
Transition regions	—	72–83, 94–104, 114–132, 142–158, 163–167 (α helices B′, C, D, E, and a-3_{10})

All DCCM maps analyzed and discussed below were calculated using the M2 trajectory (Figs. 3–5, RF2, RF3, and RF4); DCCM maps calculated using the M1 trajectory were very similar to those calculated from the M2 trajectory. For the analyses described below, the results and discussion of the M2 maps would in general apply equally well to the M1 DCCM maps.

In general, relatively strong and large cross-correlation peaks were found throughout the entire time series calculations (except for the 5-ps calculations, where a plethora of relatively weak peaks were found scattered throughout the

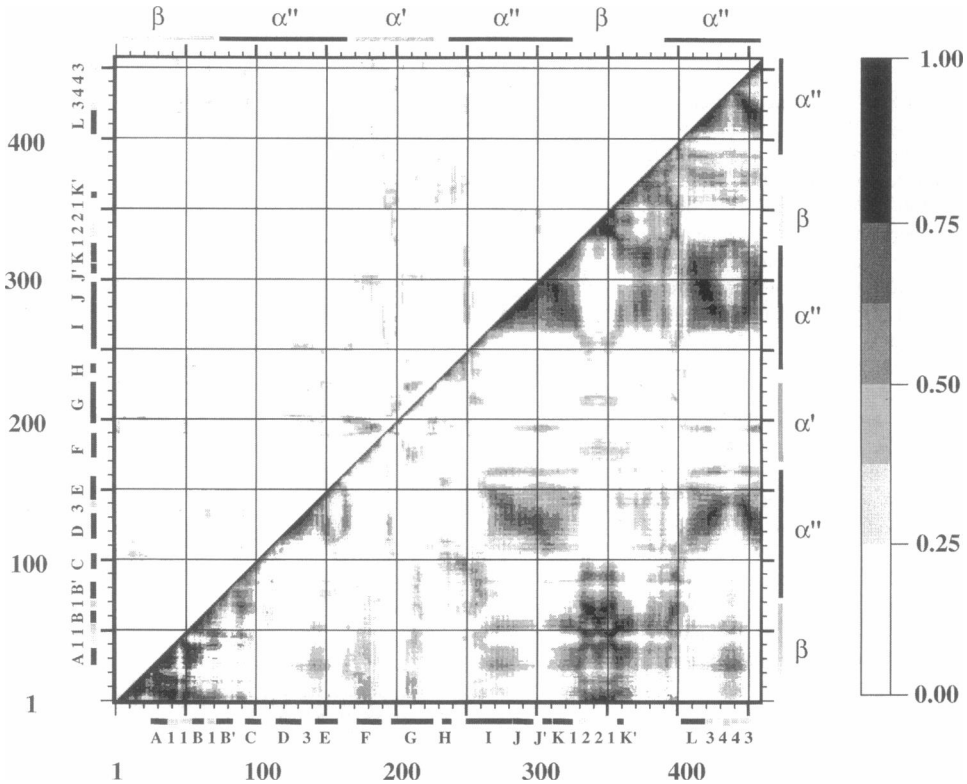

FIGURE 3 Dynamic cross-correlation map of time-correlated backbone atom motions calculated over the 150–190-ps time interval from the M2 trajectory, using coordinate reference frame RF2. Only cross-correlations greater than 0.25 are shown; the size of the cross-correlation is proportional to the shading, as indicated by the scale at right. Negative correlations are shown in the upper triangle and positive correlations are shown in the lower triangle. The bars along the bottom and left borders indicate positions of secondary structure elements. Black bars correspond to α-helices and are labeled with a capital letter; gray bars correspond to β-sheets and are labeled with their numeric designation. The bars along the top and right borders indicate the positions of the putative protein domains.

maps). The intensities of these values became more pronounced in progressing from the 5- through 40-ps time series calculations; only subtle changes in positions and intensities were observed in calculations using time blocks of 40 ps and larger. Some of the small and weak cross-correlation peaks exhibited variations in position through the DCCM times series calculations. These particular atomic motions are independent of the atomic motions associated with time-correlated domain motions that must persist through the time series.

The DCCM maps (Figs. 3–5) generally exhibited strong positive cross-correlation values among those residues previously defined as members of the α domain and among those residues defined as members of the β domain. However, for several residue segments that had previously been classified as members of either domain α or β, no time-correlated motions were observed in the DCCM maps as described below.

The DCCM maps calculated using RF2 (Fig. 3) generally exhibited strong positive cross-correlation values between intra- and interdomain regions among the respective residue segments that comprise the α″ and β domains. Unexpectedly, several residue segments that had previously been defined as domain components exhibited little or no time-correlated motions with their respective domains. The B′ and C helices exhibited positive cross-correlation peaks among themselves, but no strong positive cross-correlation peaks were found between these two helices and the majority of residues comprising the α″ domain. Weak to moderate positive cross-correlation values were observed between the B′ and C helices and residues of the β domain, H helix, and ~15 N-terminal residues of the I helix. DCCM maps calculated using RF3 (Fig. 4) show virtually no time-correlated atomic motions between the B′ and C helices and the β domain. However, weak to moderate positive cross-correlation values were found between the B′ and C helices and helices D, E, F, and G; moderate positive cross-correlation values were also found between the B′ and C helices and helix H and ~15 N-terminal residues of the I helix. Consistent with these findings are the data from DCCM maps calculated using RF4 (Fig. 5). As expected, virtually no positive cross-correlation peaks were observed among the residues comprising the α″ domain. However, again, weak to moderate cross-correlation peaks were found between helices B′ and C and a preponderance of β domain residues, helix H, and the ~15 N-terminal residues of the I helix. These data suggest that the motional properties of the B′ and C helices are generally independent of those of the majority of α″ domain residues. However, some positive time-correlated atomic motions were found between the B′ and C helices and those parts of the α′ and α″ domain that are spatially proximate.

The two extended residue segments, 84–92 and 387–401, have both been classified as part of the α domain, but only negligible positive cross-correlation was found between these residues and the rest of the α domain, regardless of the reference frame used (Figs. 3–5). However, moderate to strong positive cross-correlation was observed between these extended residue segments and the β domain, and helices B′ and C. These data strongly indicate that these

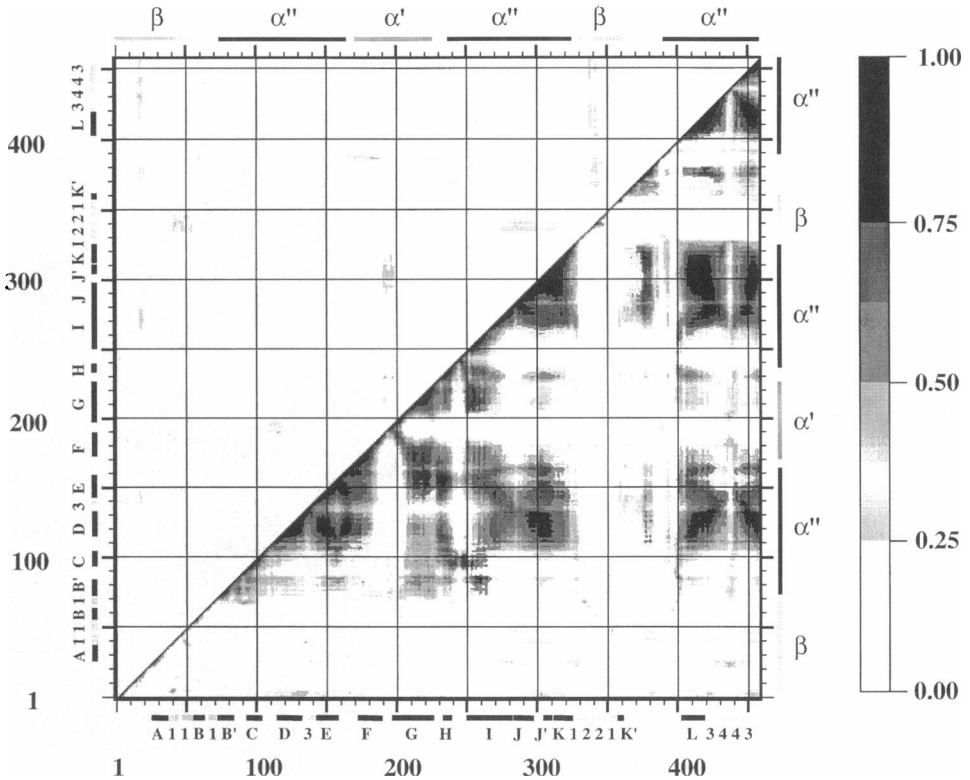

FIGURE 4 Dynamic cross-correlation map of time-correlated backbone atom motions calculated over the 150–190-ps time interval from the M2 trajectory, using coordinate reference frame RF3. Only cross-correlations greater than 0.25 are shown; the size of the cross-correlation is proportional to the shading, as indicated by the scale at right. Negative correlations are shown in the upper triangle, and positive correlations are shown in the lower triangle. The bars along the bottom and left borders indicate positions of secondary structure elements. Black bars correspond to α-helices and are labeled with a capital letter; gray bars correspond to β-sheets and are labeled with their numeric designation. The bars along the top and right borders indicate the positions of the putative protein domains.

extended residues are part of the β domain, and not the α domain.

From the perspective of RF2, weak to moderate positive cross-correlation values were found between the ~15 N-terminal residues of the I helix and helix H. In contrast, virtually no positive cross-correlation values were observed between these residue segments and the majority α'' or β domain residue constituents (Fig. 3). Weak to moderate

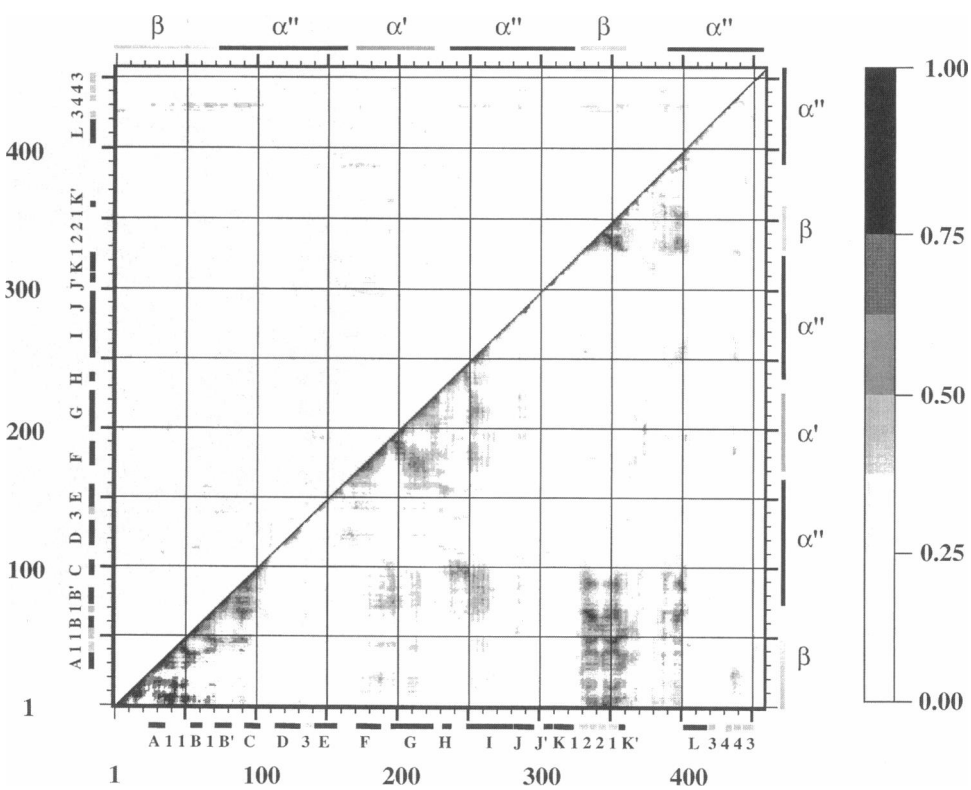

FIGURE 5 Dynamic cross-correlation map of time-correlated backbone atom motions calculated over the 150–190-ps time interval from the M2 trajectory, using coordinate reference frame RF4. Only cross-correlations greater than 0.25 are shown; the size of the cross-correlation is proportional to the shading, as indicated by the scale at right. Negative correlations are shown in the upper triangle, and positive correlations are shown in the lower triangle. The bars along the bottom and left borders indicate positions of secondary structure elements. Black bars correspond to α-helices and are labeled with a capital letter; gray bars correspond to β-sheets and are labeled with their numeric designation. The bars along the top and right borders indicate the positions of the putative protein domains.

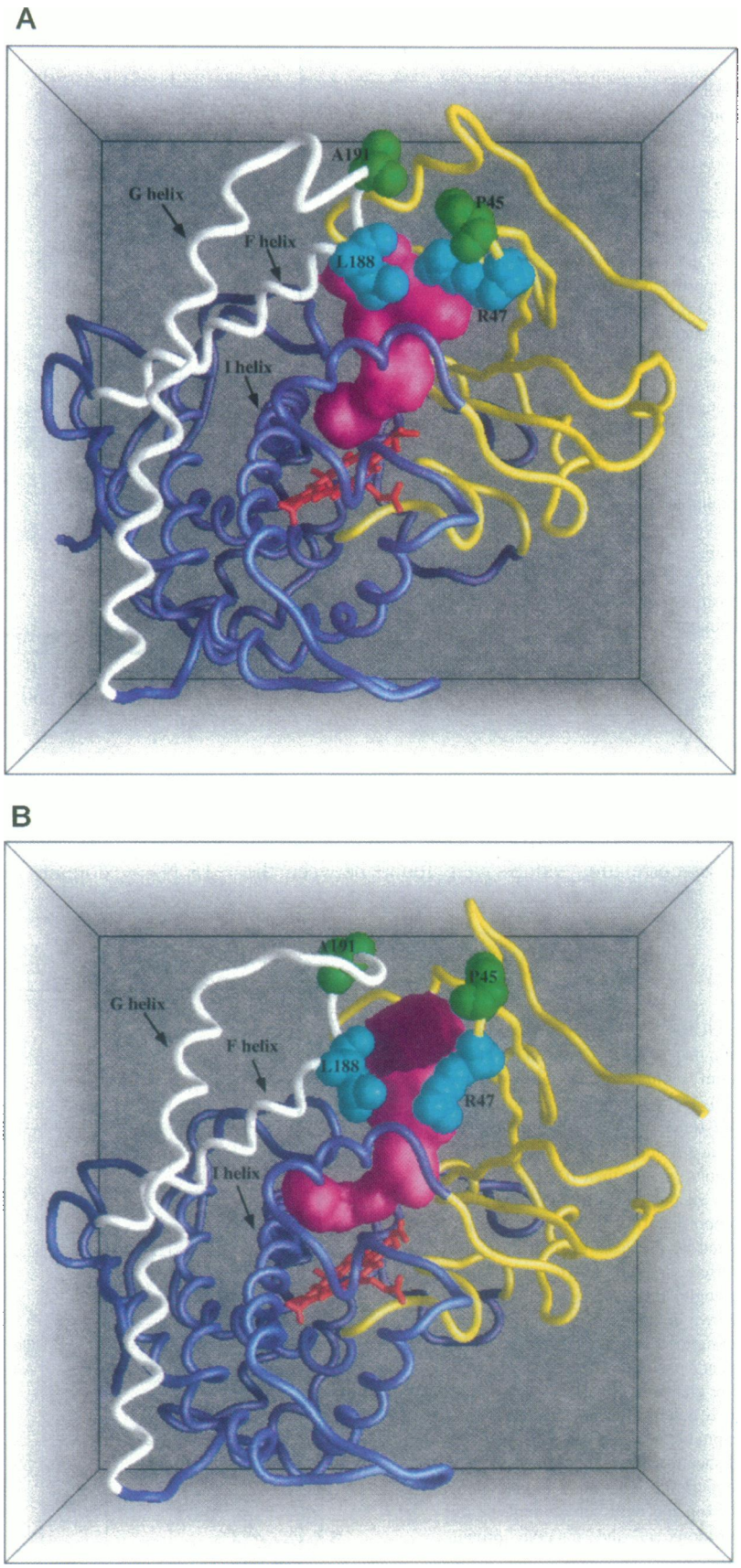

FIGURE 6 α-Carbon worm representation of structure M1A (*A*) and structure M2A (*B*). The F-G helix region is displayed in white; the β domain is drawn in yellow. The remaining portions of the protein are colored steel blue. The heme group is depicted by a stick representation and is displayed in red; the active-site cavity is represented by the magenta surface. Residues 45 and 191 are displayed in a green CPK representation. Two of the residues used in calculating the active-site annulus are depicted by a light blue CPK representation.

positive time correlations were found between these two residue segments and helices B′ and C. In general, similar data were observed for the ~15 N-terminal helix I residues and helix H, regardless of reference frame used (Figs. 4 and 5), with the exception of RF3, where moderate to moderately strong positive cross-correlation values were observed between these residue segments and helices D, E, F, and G (Fig. 4). Thus the atomic motions of the ~15 N-terminal residues of helix I and helix H are generally independent of the majority of α'' domain residues. But some time-correlated motional behavior is observed between these two residue segments and the adjacent helices B′, C, D, E, F, and G.

DCCM analyses of the F and G helix region indicate that the α' domain does exhibit motional properties largely independent of a significant portion of the α'' domain (except for the cases described above). Using RF2, strong cross-correlation values were found throughout the majority of α'' and β domain constituents, and virtually no cross-correlation peaks were found within the α' domain. Using RF3, strong positive cross-correlation values were observed between the F and G helices, indicating that the atomic motions of these two helices were concerted. Furthermore, moderate to strong positive cross-correlation values were also seen between the α' domain and helices D, E, H, and ~15 N-terminal residues helix I. Using RF4, positive cross-correlated motions between the F and G helices themselves and between the F and G helices and ~15 N-terminal residues of the I helix were again found. Also evident in these DCCM maps was the lack of positive cross-correlated atomic motions among the extended loop region, between the F and G helices, and the rest of the protein. These data suggest that the atomic motions of the F and G helices are concerted and are generally different from the α'' domain. However, the time-correlated atomic motions found between helices F and G and helices D, E, H, and ~15 N-terminal residues of helix I do indicate that cross-communication exists between the F and G helices and those parts of the α'' domain that are spatially proximate.

Interestingly, positive time-correlated atomic motions between the first 15 residues of helix I and helices B′, C, D, E, F, G, and H, and virtually no positive cross-correlated motions between these helices and the rest of the α'' domain are observed. Reference frames RF2 and RF3 both indicate that the off-diagonal I helix triangle does not include the first 15 N-terminal residues. The 15th I helix residue is a glycine (Gly^{265}), and inspection of structures M1A and M2A revealed a disruption of the I helix hydrogen-bonding network at and around this residue. We observed that the I helix had separated into two discrete helices (at residue position 265), and the DCCM data indicate that each helix segment exhibits distinct time-correlated atomic properties.

Domain classification based on DCCM analysis

Several differences were found in comparing the static crystal structure domain definitions against those determined from DCCM analyses of the MD trajectories. Specifically, the DCCM maps showed that positive cross-correlation values for the extended residues 387–401 were found only with β domain residues, regardless of the reference frame used. These data clearly indicate that this residue segment is a component of the β domain rather than the α domain. Furthermore, because of the strong positive cross-correlation values found between residues 84–92 and β domain residues, this residue segment has been reclassified as part of the β domain. The DCCM data also indicate that the time-correlated motions of the F-G helix region are synchronous and are, by and large, independent of those of domains α and β. For this reason the F-G helix region has been classified as a new protein domain, designated α'.

The specific DCCM domain characterization of helices B′, C, D, and E is less clear. The time-correlated atomic motions of helices B′ and C were tightly coupled, as were the time-correlated atomic motions between helices D and E. Because the time-correlated atomic motions of helices B′ and C are coupled with both the α' and β domains, helices B′ and C were classified as transitional residue segments intervening between the two domains. Similarly, because of the cross-correlation of the atomic motions of helices D and E with both domains α'' and β, helices D and E have been classified as transitional residue segments.

In summary, the DCCM data indicate that three distinct protein domains exist (consisting of α', α'', and β), with helices B′ and C, and D and E, serving as transition regions between domains α' and β and α' and α'', respectively. This information is summarized in Table 1 and depicted in Fig. 7.

Molecular flexibility

Differences in protein conformations can be directly analyzed by computing a difference distance matrix (DDM) (Nishikawa et al., 1972). The functional form of the equation is

$$\Delta\Delta r_{ij} = (\Delta r_{ij})_1 - (\Delta r_{ij})_2 \quad (2)$$

where Δr_{ij} are the α-carbon displacements within a given structure, and the differences in displacements ($\Delta\Delta r_{ij}$) are calculated between two distinct protein conformers. The individual $\Delta\Delta r_{ij}$ elements are then cast into matrix form, and the resulting matrix is symmetrical about the diagonal. Because the matrix is symmetrical, data in the upper and lower triangles are redundant. Different DDM calculations, with respect to the upper and lower triangles of each matrix, are shown, thus eliminating the redundant information. A negative $\Delta\Delta r_{ij}$ value indicates that the displacement of the residue pair has increased; a positive $\Delta\Delta r_{ij}$ value indicates that the diplacement of the residue pair has decreased.

DDM calculations for M1C–M2C are shown in the upper triangle of Fig. 8. The two crystal structures are virtually identical, except for the perturbations found at the entrance of the substrate-binding channel (Ravichandran et al., 1993). The negative difference distance values indicate that

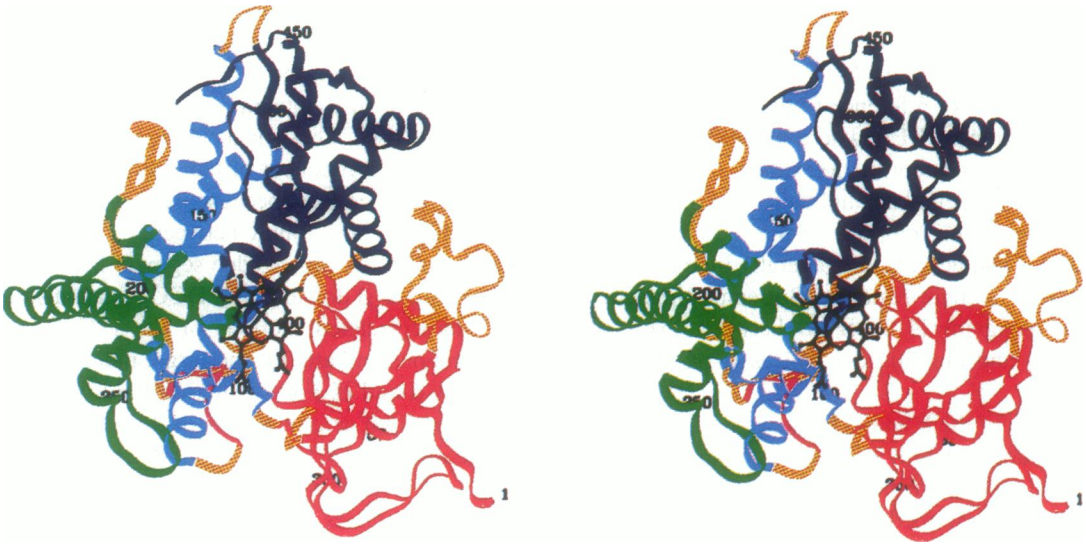

FIGURE 7 Regions of cytochrome P450BM3 identified from the MD simulations. Individual domains are color coded as follows: α' domain, dark blue; α'' domain, light blue; β domain, red. The transition regions between domains are colored magenta. The region of the protein that exhibited little or no coherent motions with any other portion of the protein (residues 362–389) is colored green. The heme group is depicted by a black stick representation.

this region in the M2C structure is more open relative to the M1C structure. A greater number of perturbations are found in the DDM calculation between structures M1C–M1A (*lower triangle*, Fig. 8). These data indicate that the two crystal structures are more similar to each other than are M1A and M1C. Of particular interest are the DDM values observed in the binding pocket region. Note that the negative values previously found in this region are gone and some positive difference distance values are now observed. These data indicate that this region of M1A has "closed" relative to either crystal structure. Also observed is a band of positive values between residues 9–13 and various segments throughout the rest of the protein. These residues are located at the entrance of the binding pocket and have decreased in displacement relative to the rest of the protein, as compared to these same displacement values in structure

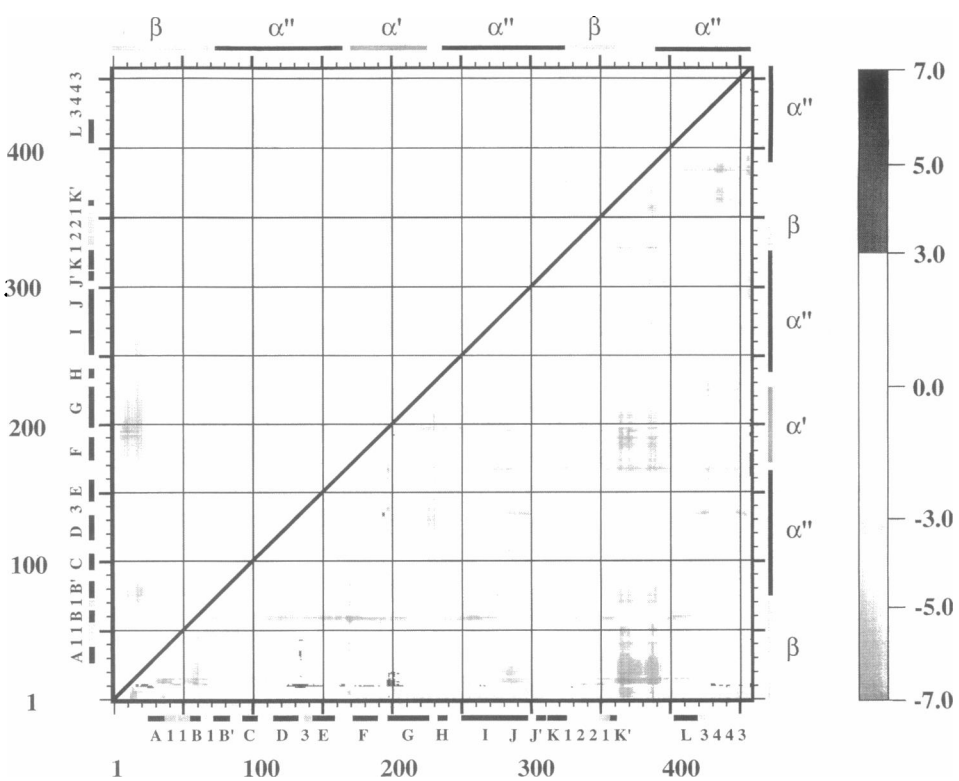

FIGURE 8 Difference distance matrix maps depicting differences in α-carbon displacements (measured in Å) between structures M1C and M2C (*upper triangle*) and M1C and M1A (*lower triangle*). Positive differences are plotted as dark gray contours, and negative differences are plotted as light gray contours. The magnitudes of the differences are proportional to the shading, as indicated by the scale at right.

M1C. Also evident is another band of negative difference distance values found for residues ~55–61. These data indicate that the B helix has extended away slightly from the rest of the protein. A similar increase in displacement between the extended residues 362–389 and the β domain is also observed.

DDM values calculated using structures M1C–M2A are shown in the upper triangle of Fig. 9. Of particular interest are the negative DDM values observed between residues at and around 3_{10}-helix −a and α-helices F and G and the N-terminal residues of helix I. The 3_{10}-helix −a is located at the mouth of the binding pocket opposite α-helices F and G and the N-terminal residues of helix I. The negative DDM values associated with these residues indicate that this region has opened relative to structure M1C. There is also a negative band of values associated with the F-G helix loop relative to the G helix, indicating that the loop has extended away from the G helix (except for residues 191–193, which have inverted and are now closer to the β domain). This is explicitly evident from comparing Fig. 6 *A* with Fig. 6 *B*. There are several other bands of negative DDM values that are visible. Again, an increase in displacement is found between the extended residues 362–387 and much of the protein; also observed are negative displacement values between 3_{10}-helices −b and −d and various segments of the protein. These stretches of residues are on the surface of the protein away from the active site, and the motions of these residues are most likely associated with the surface dynamics of the protein rather than active-site dynamics.

Displayed in the lower triangle of Fig. 8 are DDM values calculated using structures M1A–M2A. A general increase in distances between residues associated with the binding pocket is evident. In particular, greater displacement values are found between β domain residues ~5–45 and the F and G helix region, compared to all other DDM maps. Also observed is a significant increase in distance between helix G relative to helix K, and helix G relative to the first 15 residues of the I helix. There is also a band of negative values associated with active-site residues ~436–438 and the F and G helix region. Taken together, these data indicate that the active-site volume of structure M2A is larger than that of M1A. Positive DDM values found between the extended region on the back side of the protein (residues 361–390) β domain and the F helix indicate a decrease in displacement between these regions, and probably have little effect on the shape of the active site. Some stretches of positive DDM values are also observed between the C-terminal residues of the F helix and helix J, β sheet 4, and the C-terminal residues of the helix I, indicating some closure and deformation in this portion at the mouth of the binding pocket (compare Fig. 6 *A* with 6 *B*).

Structures M1C, M2C, M1A, and M2A were also analyzed in terms of rigid body rotation and translation, as described by Faber and Matthews (1990), of the F and G helix region relative to the β domain. In these analyses, the M1C structure was used as the template. Initially, structures M2C, M1A, and M2A were each superimposed on the backbone atoms of M1C in the F and G helix region. Subsequently, a transformation was calculated for each structure (consisting of a rotation and translation about an axis) such that the β domain of each of the three structures was superimposed on the β domain backbone atoms of the

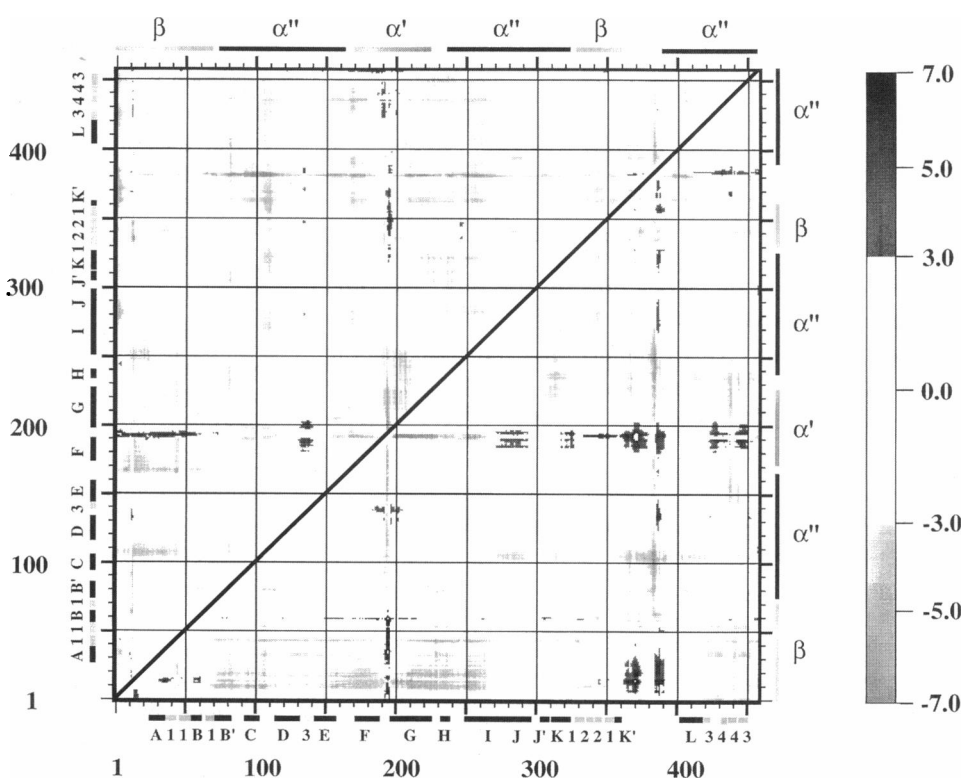

FIGURE 9 Difference distance matrix maps depicting differences in α-carbon displacements between structures M1C and M2A (*upper triangle*) and M1A and M2A (*lower triangle*). Positive differences are plotted as dark gray contours, and negative differences are plotted as light gray contours. The magnitudes of the differences are proportional to the shading.

M1C template structure. In addition, an identical set of calculations were made, using structure M1A as the template. For these calculations, a range of rotation angles was observed, from 4.8° to 12.6° (see Table 2). Structure M2A had the largest calculated rotation angle value regardless of the template used, and the M1C–M1A rotation angle was the smallest. Note also that all of the rotation angles increase in size when structure M1A is used as the template. This indicates that the binding pocket region of structure M1A is more closed relative to structure M1C; the rotation data also indicate that the binding pocket region of structure M2A is the most open relative to the other three structures.

Consistent with the rotation data are the measured volumes of the active-site pocket (Table 3 and Fig. 6). These values range from 640 Å to 980 Å^3, and there is an increasing progression in size for structures M1A, M1C, M2C, and M2A. In comparing Fig. 6 *A* with Fig. 6 *B*, one sees that the active-site surface is closed in structure M1A and open in structure M2A. Furthermore, an additional volume located above the heme, juxtaposed to the I helix, is found in structure M2A and is not observed in structure M1A. Because the active-site volume of the M1A structure is smaller than that of either of the two crystal structures, one might suspect an aberrant M1A trajectory resulting in general compaction and collapse of the simulated protein. Molecular volume and solvent-accessible surface area calculations indicate that this is not the case, because these values are both larger for structure M1A as compared to either of the crystal structures, and these values are quite similar between the two simulated proteins (Table 3). Thus the compaction of the active site in the M1A trajectory is a localized phenomenon in the M1A structure and agrees with the compaction observed in the x-ray crystal structure of cytochrome P450BM-3 complexed with a long-chain fatty acid (Li and Poulos, 1996).

CONCLUSIONS

Both dynamical cross-correlation maps and difference distance matrices were used to study domain structure and the dynamic properties of cytochrome P450BM-3. The MD trajectories indicated that the enzyme comprises three principle domains—only two domains were reported for the crystal structure. The simulation data suggest that the dynamics of the newly identified F-G helix domain region could potentially be important in allowing access of substrate to the enzyme-binding pocket.

TABLE 2 Rigid-body rotation angles of the F and G helix region relative to the β domain

BM3 structure	Angle of rotation with M1C template (degrees)	Angle of rotation with M1A template (degrees)
M1C	—	4.8
M2C	7.1	10.3
M1A	4.8	—
M2A	9.7	12.6

TABLE 3 Molecular surface measurements

BM3 structure	Active-site volume (Å^3)	Molecular volume (Å^3)	Solvent-accessible surface area (Å^2)
M1C	640	87,230	18,540
M2C	770	87,300	18,740
M1A	550	91,100	19,950
M2A	980	91,550	20,470

All measurements with regard to the size of the enzyme's active site, including both the DDM and hinge-bending angle calculations (Table 2), clearly indicate that the active site of structure M1A had contracted and that of structure M2A had expanded relative to the crystal structures. These differences are depicted in Fig. 6, illustrating the changes in both the size and shape of the binding pocket. These variations demonstrate that the substrate binding pocket is highly dynamic and can adapt a range of conformations in solution. P450BM-3 hydroxylates both saturated and unsaturated fatty acids ranging in length from 14 to 20 carbons; thus the active site must be highly flexible for the enzyme to be able to bind such a wide range of substrates. The MD simulations demonstrate that the F and G helix domain region can act as an independent hinge, allowing the binding pocket to adapt various sizes and shapes as necessary for the enzyme to accommodate a host of different substrates, which is consistant with the recent results of Li and Poulos.

The above data indicate that the M1 and M2 simulations have taken different trajectories and are exploring different portions of conformational space. Measured properties such as protein radius of gyration, atomic RMS deviations away from the crystal structure coordinates, solvent-accessible surface area, and thermal B factors all suggest that the two trajectories have thermally equilibrated (Paulsen and Ornstein, 1995). The 200-ps length of the two MD simulations is respectable by most standards for a protein system of this size, but clearly the length was not large enough to sample the entire conformational spectrum available to the protein. Nonetheless, the different trajectory paths taken by the two simulations demonstrate the inherent structural flexibility of this enzyme, which is consistent with the recent results of Li and Poulos (1996).

Amadei and co-workers have developed a two-principle-component analysis or essential dynamics analysis method to study the dynamic properties of protein MD trajectories (Amadei et al., 1993). Here, too, the use of a single globally positioned reference frame can potentially be inadequate to identify relevant interdomain motions. The essential dynamics method (like the DCCM method) relies on atomic displacements, and the spatial orientation among the different snapshot structures in the trajectory will clearly affect the positional fluctuations in these analyses. The use of multiple reference frames, in the essential dynamics

method, would provide additional information with regard to the dynamics of interdomain motions.

The primary objective of this study was to demonstrate a new implementation of the DCCM analysis method and its utility. The study emphasizes that when applying the DCCM analysis tool, the user must be cognizant of the coordinate reference frame in which the analysis is being carried out. Clearly, the use of only a single spatial reference (as in past DCCM studies) has the potential to obscure time-correlated atomic motions, particularly within protein domains. The use of multiple spatial reference frames can overcome this potential limitation and help identify concerted atomic motions that extend throughout the protein of interest.

We thank Dr. Mark Paulsen for the many discussions regarding BM-3. Special thanks go to Dr. Xue-jun Zhang (University of Oregon, Eugene) for making his EDPDB computer program available for calculating rotation angles in BM-3.

Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by the Battelle Memorial Institute under contract DE-AC06-76RLO 1830.

REFERENCES

Amadei, A., A. B. M. Linssen and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Protein Struct. Funct. Genet.* 17:412–425.

Arnold, G. E., and R. L. Ornstein. 1994. An evaluation of implicit and explicit solvent model system for the molecular dynamics simulation of bacteriophage T4 lysozyme. *Protein Struct. Funct. Genet.* 18:19–33.

Brunger, A. T., C. L. Brooks III, and M. Karplus. 1985. Active site dynamics of ribonuclease. *Proc. Natl. Acad. Sci. USA.* 82:8458–8462.

Faber, H. R., and B. W. Matthews. 1990. A mutant T4 lysozyme displays five different crystal conformations. *Nature.* 348:263–266.

Gerstein, M., A. M. Lesk, and C. Chothia. 1994. Structural mechanism of domain movements in proteins. *Biochemistry.* 33:6739–6749.

Harte, W. E. J., S. Swaminathan, M. M. Mansuri, J. C. Martin, I. E. Rosenberg, and D. L. Beveridge. 1990. Domain communication in the dynamical structure of human immunodeficiency virus 1 protease. *Proc. Natl. Acad. Sci. USA.* 85:4686–4690.

Ichiye, T., and M. Karplus. 1991. Collective motions in proteins—a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Protein Struct. Funct. Genet.* 11:205–217.

Li, H., and T. L. Poulos. 1996. Conformational dynamics in cytochrome P450-substrate interactions. *Biochimie.* 78:695–699.

McCammon, J. A., and S. C. Harvey 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge.

Nicholls, A., K. A. Sharp, and B. Honing. 1991. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* 11:281–290.

Nishikawa, K., T. Ooi, Y. Isogai, and N. Saito. 1972. Tertiary structure of proteins. I. Representation and computation of the conformations. *J. Phys. Soc. Jpn.* 32:1331–1337.

Paulsen, M. D., and R. L. Ornstein. 1995. Dramatic differences in the motions of the mouth of open and closed cytochrome P450BM-3 by molecular dynamics simulations. *Proteins Struct. Funct. Genet.* 27: 237–243.

Ravichandran, K. G., S. S. Boddupalli, C. A. Hasemann, J. A. Peterson, and J. Deisenhofer. 1993. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. *Science.* 261:731–736.

Swaminathan, S., W. E. Harte, and D. L. Beveridge. 1991. Investigation of domain structure in proteins via molecular dynamics simulation: application to HIV-1 protease dimer. *J. Am. Chem. Soc.* 113:2717–2721.

Venable, R. M., B. R. Brooks, and F. W. Carson. 1993. Theoretical studies of relaxation of a monomeric subunit of HIV-1 protease in water using molecular dynamics. *Proteins Struct. Funct. Genet.* 15:374–384.