Homology Modeling, Molecular Dynamics Simulations, and Analysis of CYP119, a P450 Enzyme from Extreme Acidothermophilic Archaeon *Sulfolobus solfataricus*[†]

Yan-Tyng Chang* and Gilda Loew

Molecular Research Institute, 2495 Old Middlefield Way, Mountain View, California 94043-2316

Received August 20, 1999; Revised Manuscript Received December 16, 1999

ABSTRACT: The recent characterization of a thermophilic and barophilic CYP119 from *Sulfolobus solfataricus* offers a new opportunity to identify the origin of its stability by comparing it with mesophilic P450s with known structures. Since the three-dimensional structure of CYP119 is not yet available, homology modeling techniques were used to build model structures for this enzyme. The overall quality and stability of the models were assessed using three protein analysis programs and by monitoring structural stability during 1 ns of molecular dynamics simulations at 300 and 390 K. The results show the CYP119 models to be of good quality. Possible origins of the thermo- and barostability of CYP119 were then investigated by examining the amino acid compositions and the three-dimensional structure of CYP119 compared with the five mesophilic templates. Three possible factors were identified that could contribute to the enhanced stability of CYP119. The first was the higher relative population of salt bridges and the presence of a few unique salt bridges found in CYP119 that were absent in all five template CYP450s. The second factor was a decreased population of Ala and an increased population of Ile found in the interior of CYP119, which are likely to improve packing in CYP119. The third factor was a more extensive aromatic cluster seen in CYP119 which was not found in all five template P450s. In addition, the model CYP119 three-dimensional structures were also used to determine key properties related to its function. Specifically, binding site residues and surface residues for redox partner interactions were identified. These residues identified together with those residues found that might contribute to the increased stability are suggested for future mutagenesis studies. The results obtained from these experimental studies shall then provide further validation of the suggested origins of stability and the structure−function relationships derived from the model structures of this enzyme.

Life on earth shows an enormous capability in adapting to different extreme conditions, including temperatures ranging from −5 to 110 °C, pressures from 0.1 to 120 MPa (1200 bar), water activity from 1.0 to 0.6, and pH from 1 to 12 (*1*, *2*). Enzymes from organisms, mainly microorganisms, that adapt to the various extreme environments are called psychrophilic, thermophilic, barophilic, halophilic, acidophilic, and alkalophilic. Studies of these enzymes not only provide the opportunity to probe the origin of their intrinsic stability but also allow possible applications in biotechnology. For example, using thermostable enzymes in industrial applications offers the benefits of increased rates of reaction, higher substrate solubility, decreased media viscosity, longer enzyme self-lives at normal storage temperature, and lower risk of microbial contamination.

Most of the extremophilic enzymes found are thermophilic belonging to diverse families of bacteria (*2*). Some examples are citrate synthase from *Pyrococcus furiosus* with a half-life of 170.0 min at 100 °C, rubredoxin from *P. furiosus* with a melting temperature above 176 °C, and adenylate kinase from *Bacillus stearothermophilus* with a melting temperature of 74.5 °C. Many thermophilic enzymes have normal mesophilic counterparts that function at ambient temperatures and belong to the same protein family.

Recently, the first extremophilic enzyme in the metabolizing cytochrome P450 superfamily has been identified. The gene of this enzyme, CYP119, was cloned from *Sulfolobus solfataricus* (*Ss*[1]), an acidothermophilic archaebacteria found in sulfurous volcanic hot springs (*3*). The optimal growth conditions for *Ss* were found to be 87 °C and pH 4.5 (*4*). McLean et al. (*5*) have recently reported comparative studies of the stabilities of CYP119 and the mesophilic CYP101 (P450cam). It was found that the melting temperature (T_m) of CYP119 was 91 °C, compared to 54 °C for the T_m of CYP101. Moreover, CYP119 was stable at hydrostatic pressures of up to 2 kbar as evidenced by the amount of P450 present, while there was little or no native CYP101 left at 2 kbar. Similar thermostability and barostability of CYP119 have also been observed by Koo et al. (*6*).

The fundamental challenge posed by the existence of these proteins is the identification of differences in their properties that can account for the enhanced thermal stability of the thermophile. The search for these differences is currently an active area of investigation.

[†] Support for this work by National Institutes of Health Grant GM56125 is gratefully acknowledged.

* To whom correspondence should be addressed. E-mail: chang@purisima.molres.org. Telephone: (650) 210-0310. Fax: (650) 210-0318.

[1] Abbreviations: CYP, cytochrome P450 enzyme; *SS*, *Sulfolobus solfataricus*; T_m, melting temperature; MD, molecular dynamics; com, center of mass; rmsd, root-mean-square deviation; FMN, flavin mononucleotide.

10.1021/bi991966u CCC: $19.00
Published on Web 02/16/2000

One approach has been to examine the X-ray crystal structures of a variety of thermophilic proteins and compare them with mesophilic proteins in the same family. There is now a considerable library of such proteins in the public Protein Data Bank. Direct comparisons of these experimental structures have been made in the hope of identifying the factors leading to the difference in stability. Various factors such as improved packing, changes in loop flexibility, domain−domain interaction, and H-bond and ion pair salt bridge interactions have been suggested as origins for the enhanced stability. However, none of the differences described has been shown to be definitely linked to enhanced thermal stability in all thermophilic enzymes.

There are two major difficulties in the attempts to link directly observed differences in X-ray structures in mesophilic and thermophilic partners to differences in thermal stability. The first is that these structural differences need not be very large. Specifically, as pointed out by Jaenicke and co-workers (*1*, *2*), the difference in the stabilities of mesophilic and thermophilic proteins [$\Delta\Delta G$(natural − unfolded)] does not exceed 100 kJ/mol. This free energy difference corresponds to very subtle differences in the protein structures, for example, the presence or absence of just a few noncovalent interactions. The second difficulty in using X-ray structures alone is that such differences can be manifest mainly in the dynamic properties of the two proteins that cannot be readily deduced from crystal structure comparisons alone.

Due to the ubiquitous nature of the CYP450 enzymes (*7*) and the involvement of them in a wide spectrum of oxidation reactions, the discovery of the first acidothermophilic and barophilic CYP119 enzyme offers a new opportunity to identify the origin of its stability and to use the knowledge that is obtained to convert other mesophilic to thermophilc CYP450 for potential industrial and commercial applications. However, in the case of the CYP450s, there is an additional impediment. While the X-ray structures of five mesophilic CYP450s of bacterial or fungal origins are known, the X-ray structure of the thermophilic CYP119 is not. In the only study thus far reported addressing possible differences between the mesophilic CYP101 and the thermo- and barophilic CYP119 by McLean et al. (*5*), a model structure of CYP119 was used as obtained by the automated program Modeler (*8*). However, except for the observation of a clustering of aromatic residues centered around Trp281 of CYP119, no other properties were found in their model that could explain the thermo- and barostability of this enzyme.

In the work reported here, we continue to probe the origin of the stability of CYP119. The approach used here differs from the initial one in a number of important ways: (i) the strategies used to construct the three-dimensional (3D) models of CYP119, (ii) the assessment of these models, (iii) the probing of the dynamics of their structural behavior at different temperatures, and (iv) comparisons of the results with even-handed studies of mesophilic CYP450s.

Specifically, two homology models of CYP119 were constructed and assessed using strategies developed and validated in this laboratory (*9*−*11*). Molecular dynamics simulations with a length of 1 ns were then performed using these models. Analysis of amino acid compositions and structural features of the models was then made, and they were compared to those of other mesophilic P450 enzymes to provide possible explanations for its remarkable stability. In addition to probing the origin of its stability, we examined the properties relating to the enzymatic function of CYP119. Among them were the characterization of the topology of the active site of CYP119 and the protein−protein interactions with redox partners.

The residues that were identified that could (1) contribute to the enhanced stability, (2) modulate substrate and product specificity, and (3) affect the binding and/or electron transfer between CYP119 and redox partners are suggested for future mutagenesis studies. The results obtained from these experimental studies shall then provide further validation of the suggested origins of stability and the structure−function relationships derived from the model structures of this enzyme.

METHODS

Homology Modeling

The steps used to generate 3D models of the CYP119 enzyme are similar to those used in previous works to generate models for CYP105C1 (*9*) (P450choP), CYP2B4 (*10*), and CYP4A11 (*11*). The sequence of CYP119 with 368 amino acids was retrieved from the databank in the National Center for Biotechnology Information (www.ncbi.nih.gov). At present, the 3D crystal structures of five P450 enzymes, including the bacterial CYP101 (*12*) (P450cam), CYP102 (*13*, *14*) (P450BM3), CYP107A (*15*) (P450eryF), CYP108 (*16*) (P450terp), and the fungal CYP55A1 (*17*) (P450nor), are available for use as templates.

From pairwise sequence alignments, it was found that the level of sequence identity between CYP119 with each one of the templates is low, ranging from ∼18% with CYP102, 19% with CYP101, and 23% with CYP108 and CYP55A1 to ∼27% with CYP107A. It was thus more reliable to use all of them as templates. These five P450s will be referred to as the template P450s throughout the text.

Generation of an Initial Sequence Alignment and an Initial Model. The multiple-sequence alignment of four of the templates (CYP101, CYP102, CYP107A, and CYP108) based on structure alignment has been obtained previously (*9*). Addition of the fifth one, CYP55A1, also based on structure alignment was performed in this work. The sequence of the target protein, CYP119, was then aligned with the multiple-sequence alignment among the five templates. This was performed by using the pairwise sequence alignment result between CYP119 and CYP107A as a guide, and modified with manual adjustment. This initial sequence alignment was then used to generate an initial model (model 0) of CYP119. Procedures for the generation of backbone atoms in structurally conserved regions, nonstructurally conserved regions, and side chain conformations using Insight/Homology for the initial model are identical to those used in previous work (*9*−*11*).

Improvement of the Sequence Alignment and Backbone Structure. As described in detail in our previous work of homology modeling of CYP4A11 (*11*), to assess and iteratively improve the quality of sequence alignment and of the backbone structure of a model, the protein structure analysis program (Prosa) by Sippl (*18*) was used. Specifi-

cally, sequence alignments and backbone structures in all loops between the B and C, D and E, E and F, F and G, G and H, H and I, and J and K helices were varied and optimized using the Prosa energy plot and the Prosa *z* score as a guide. The optimum sequence alignment and loop structures in all except the region between the B and C helices were obtained. Due to the greater uncertainty in the B−C region, two sets of sequence alignments that differ only in the alignment in this region were kept. Using these two sequence alignments, two separate 3D models of CYP119 were generated. In the first model, the sequence of CYP119 in the B−C region was aligned mostly with that of CYP107A (P450eryF). In the second model, this region of CYP119 was aligned mainly with that of CYP108 (P450terp). The two models then differ only in the structure of this B−C loop: one that mimics CYP107A and the other that mimics CYP108.

Refinement of the Side Chain Conformations. The next step in the generation of the 3D models was improvement of the quality of the side chain conformations. In previous work in our laboratory involving a CYP4A11 model, we have compared the reliability of two different strategies for side chain generation, TORSO (*19*) and SCWRL (*20*). The SCWRL method was chosen in this work on the basis of a better agreement with experiments of the CYP4A11 model generated with it. As in the previous works, the side chain conformations generated by Insight/Homology for identical residues between CYP119 and the templates were kept unchanged, while those of nonidentical residues were generated by SCWRL. The two models, starting with a different backbone structure in the B−C loop region, generated by this step were labeled as model IA and model IIA, respectively.

Generation of Two Complete 3D Models of CYP119. For each of the two models, model IA whose B−C region mimics CYP107A and model IIA whose B−C region mimics CYP108, the following additional steps were carried out.

Constrained energy minimization of both the backbone and the side chains of nonstructurally conserved regions was performed using Amber 4.1 (*21*). The Amber db94 force field parameters were used for the amino acids. The parameters for the resting state heme were obtained from an in-house database generated previously.

Structural waters, including buried and surface waters, were then added to stabilize the two models before full protein minimization. This step was performed using the HYDRAS program developed in our laboratory that examines the local residue environment for hydrogen-bonding and steric interaction between residues and waters (*22*). Using this procedure, 320 and 328 structural waters were added to model IA and model IIA, respectively. Energy minimization of the water molecules and polar hydrogens in the amino acid residues was then performed to allow the best hydrogen-bonding network to form.

Finally, an unconstrained full protein minimization using a nonbonding cutoff of 8 Å and a distance-dependent radial dielectric $\epsilon = r$ was performed. All residues, the structural water molecules, and the resting state heme unit were allowed to move in the minimization using combined steepest descent and conjugate gradient methods to obtain the fully minimized models. The two fully optimized models were labeled as model IB and model IIB, respectively.

Assessment of the Overall Quality of the Models. For model IB and model IIB, evaluation for their overall quality was performed using available protein analysis programs. This includes the program Prosa (*18*) for examination of the quality of residue−residue interaction, Whatif (*23*) for packing and Procheck (*24*) for many stereochemical properties such as main chain and side chain conformations, etc.

Molecular Dynamics Simulations

Molecular dynamics (MD) simulations were performed for two purposes: (1) to examine the quality of the model structures by checking their stability over long simulations at room temperature and (2) to examine the effect of increased temperature on the unfolding of the model structures by performing simulations at a temperature above the melting temperature of CYP119.

All MD simulations were performed at a constant temperature and with a constant number of particles. At the beginning of the simulation, initial velocities were assigned from a Maxwellian distribution at the temperature that was used. The Berendsen coupling algorithm was used during the simulation for velocity rescaling to maintain a constant temperature. SHAKE was used to constrain bonds involving hydrogens. A time step of 1 fs (10^{-15} s), a long non-bonding cutoff distance of 15 Å, and a distance dependent dielectric $\epsilon = r$ were used. The non-bonding pair list was updated every 10 time steps. The same force field parameters described earlier for energy minimization were also used for MD simulations.

Simulations for the CYP119 models were performed at two different temperatures: 300 and 390 K. The 390 K temperature is above the reported melting temperature of 91 °C (364 K) for this enzyme (*5*).

For the simulations at 300 K, all atoms were constrained in position with a finite harmonic force constant *k* of 10 kcal/Å^2 for 1 ps (10^{-12} s) at the beginning of the simulation. The constraint was gradually reduced to 5, 1, 0.5, and 0.1 kcal/Å^2 for a total of an additional 5 ps. The constraint was then removed completely, and 1000 ps of unconstrained simulation was performed.

For the simulations at 390 K, a similar procedure was used except that the number and duration of segments of constrained MD simulations were increased. Specifically, constrained simulations with *k* values of 10, 5, 1, 0.5, 0.1, 0.05, 0.01, 0.005, and 0.001 kcal/Å^2 on all atoms were performed with a duration of 2 ps for each constraint. This was then followed by 1000 ps of simulation. In addition, during the 1000 ps simulations, a very small constraint *k* of 0.001 kcal/ Å^2 was applied to the oxygen atom of all water molecules to allow reasonable movement of water molecules yet to prevent evaporation which would cause disruption of the trajectory at this higher temperature using the AMBER program.

To monitor the stability of the model structures at two different temperatures, two properties were calculated during the MD simulations. The first one is the radius of gyration (without taking into account the contribution of structural water) of the models at every 20 ps interval. The second one is the root-mean-square deviation (rmsd) of all the Cα atoms between the fully minimized structure and each structure taken at 20 ps intervals from the MD simulations.

Use of the CYP119 Structure To Probe the Origin of Its Enhanced Stability and Structure−Function Relationships

Comparison between CYP119 and Mesophilic CYP450s with Known Structures. In this study, possible origins of thermo- and barostabilities of CYP119 were investigated by comparing the amino acid compositions and the 3D structures of CYP119 with the five template CYP450s. Specifically, amino acid types that were found to have significant differences in populations in CYP119 and the average of the five template P450s were further investigated for their structural consequences in CYP119.

Identification of Substrate Binding Site Residues. Since the natural substrates of CYP119 have not been discovered, the residues making up the active site in the two models were identified as those located on the distal side of the heme unit and within a certain distance cutoff from the resting state heme iron atom. In this work, those residues within 10 Å of the heme iron were reported to comprise the substrate binding pocket.

Identification of Surface Residues That May Be Involved in CYP119−Redox Partner Interaction. Since the natural physiological redox partner of CYP119 has not been identified, it was decided to use two proteins that are known redox partners of two of the mesophilic templates to probe the possible surface binding site on CYP119 for interaction with its redox partner(s). These are putidaredoxin, the physiological redox partner of CYP101, and a flavin mononucleotide (FMN)-binding protein which is the natural redox partner for the self-sufficient P450BM3 (CYP102) enzyme.

In the absence of a known crystal structure, a solution structure of putidaredoxin was extracted from the first of the 12 sets of NMR structures of PDB entry 1put (*25*). The crystal structure of the FMN-binding protein for P450BM3 was extracted from the bound FMN domain−heme domain complex in PDB entry 1bvy (*26*).

To probe all possible bound complexes starting from each of the two unbound CYP119 models (model IB and model IIB) and each of the two selected redox partners, the program GRAMM was used (*27*, *28*). An evaluation of the usefulness of this GRAMM program through the prediction of the association between the FMN-binding protein and P450BM3 was presented in another work (*29*). Briefly, the conclusions obtained from this previous evaluation of the GRAMM program are as follows. (1) Although GRAMM was able to find two complex structures very close to that found in the crystal structure, the scores obtained for these two complexes were ranked 57 and 69 out of the first 100 highest-score complexes. It was thus difficult to select the "completely correct" bound complex from the pool of largely "partially correct" or incorrect complexes. (2) The best use of the GRAMM program is to apply statistical analysis of the first 100 highest-score predicted complexes. This procedure allows the identification of the surface residues on the P450 enzymes most frequently found to have close contact with the redox partners. Thus, a similar procedure was used here to find surface residues on model IB and model IIB of CYP119 that have close contact with putidaredoxin and with the FMN-binding protein. The details of this procedure are given below.

(1) The separate structures of CYP119 and each of the redox partners were used as input to the GRAMM program. A grid step size of 3.4 Å was used for the search for possible structures of the CYP119 redox partner. The coordinates of CYP119 were kept fixed in the search process. The best 100 predicted coordinates of the redox partners, i.e., putidaredoxin and the FMN-binding protein, were then collected.

(2) These 100 predicted structures of the redox partner relative to the fixed coordinates of CYP119 model were then clustered on the basis of their center of mass (com) positions using a cutoff of 30.0 Å as a criterion for clustering.

(3) For each prediction in the largest cluster, the pairwise distances between each Cα atom in the redox partner and each Cα atom in the P450s was calculated.

(4) For each prediction in the largest cluster, a cutoff was applied to these Cα−Cα distances to compile a list of residues that comprise the region to which the redox partner binds.

(5) The lists of residues from each separate prediction from the largest cluster were combined, and a statistical analysis was performed on the combined results.

(6) Two criteria were used to select residues on CYP119 for suggestion of mutation studies. (a) The first set of residues selected for initial mutagenesis studies were those within an 8.0 Å Cα−Cα distance cutoff of the redox partner residues and occurred with a frequency of ≥40%. (b) The additional residues on CYP119 suggested for mutagenesis study were charged residues that fall within an 11.0 Å Cα−Cα distance cutoff and occurred with a frequency of >40%.

The residues from these two sets provide the map for the surface region on CYP119 where each redox partner could bind.

RESULTS

Homology Models

Table 1 shows the results of the final multiple-sequence alignment among the five templates and CYP119 obtained through iterative assessment and improvement of the initial sequence alignment and initial 3D model structure (model 0). The nomenclature used by Hasemann et al. (*30*) for the helices and β sheets was adopted here. In this table, two possible alignments of CYP119, labeled as 119(1) and 119-(2), and differing only in the region between the B and C helices with respect to the five templates were included.

The decision to keep two possible alignments was due to the uncertainty in the alignment of the B−C region. The B−C variable region is usually part of the substrate binding and oxidation site. The high degree of variability in both the sequences and the structures of this region reflects the characteristic property of enzyme-specific substrate and product specificities among the many CYP450 isozymes.

Among the five templates, the lengths of the B−C variable region in CYP107A (P450eryF) and CYP108 (P450terp) are closest to the corresponding length in CYP119. Both of these are only three residues longer than that in CYP119. Although the lengths of the B−C variable region are the same for CYP107A and CYP108, their structures in this region are dramatically different. Specifically, the B′ helical axis in CYP107A is oriented perpendicular to the heme plane in contrast to the direction of the B′ helical axis of CYP108 which lies in the same plane as the heme. Furthermore, in CYP107A, a 13-residue loop (PKKKYPGVEVEFP) just

Table 1: Multiple-Sequence Alignment of CYP119 with the Five Templates[a]

```
                                        A-HELIX          β1-1
102    :       TIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADEL--GEIFKFE   43
101    :   NLAPLPPHVPEHLVFDFDMYNPSNLSAGVQEAWAVLQESNV—PDLVWTR   57
55A1   :                      APSFPFSRASGPEPPAEFAKLRATNPVSQVK   35
108    :        MDARATIPEHIARTVILPQGYADDEVIYPAFKWLRDEQPLAMAH   44
107A   :                     ATVPDLESDSFHVDWYSTYAELRETAPVTPVR   33
119(1):                                MYDWFSEMRKKD---PVYYDG   18
119(2):                                MYDWFSEMRKKD---PVYYDG   18
              β1-2  B-HELIX     β1-5
102    : APGR-VTRYLSSQRLIKEACDESRFDKNLS----------------QALKF   77
101    : CNG--GHWIATRGQLIREAYEDYRHFSSE----------------CPFIPRE   91
55A1   : LFDGSLAWLVTKHKDVCFVATSE-KLSK-VRT----RQGFPELSASGKQA   79
108    : IEGYDPMWIATKHADVMQTGKQPGLFSNA----------------EGSEILYDQ   82
107A   : FLGQ-DAWLVTGYDEAKAALSDL-RLSSDPKKKYPGVEVEFPAYLGFPED   81
119(1): -----NIWQVFSYRYTKEVLNNFSKFSSDLTGYH--ERLEDLRNGKIRFD   61
119(2): -----NIWQVFSYRYTKEVLNNFSKFSSD----------------LTGYHE   48
                      +  ++  +
        B′-REGION                                C-HELIX C*-HELIX
102    : VRDFA----------GDGLFTSWTHEKNWKKAHNILLPSFSQQAMKG---  114
101    : AGEAY----------DFIPTSMDPPEQ-RQFRALANQVVGMPVVDK---  126
55A1   : AKAKP----------T--FVDMDPPEH-MHQRSMVEPTFTPEAVKN---  112
108    : NNEAFMRSISGGCPHVIDSLTSMDPPTH-TAYRGLTLNWFQPASIRK---  128
107A   : VRNYF----------ATNMGTSDPPTH-TRLRKLVSQEFTVRRVEA---  116
119(1): IPTRY----------T--MLTSDPPLH-DELRSMSADIFSPQKLQT---   94
119(2): RLEDLRNGKIRFDIPTRYTMLTSDPPLH-DELRSMSADIFSPQKLQT---   94
                                  + ++ ++  ++  ++
              D-HELIX                  β3-1 E*-HELIX  E-HELIX
102    : ------YHAMMVDIAVQLVQKWERL-NADEHIEVPEDM-TRLTLDTIGLC  156
101    : ------LENRIQELACSLIESLRPQ----GQCNFTEDYAEPFPIRIFMLL  166
55A1   : ------LQPYIQRTVDDLLEQMKQKGCANGPVDLVKEFALPVPSYIIYTL  156
108    : ------LEENIRRIAQASVQRLLDFD---GECDFMTDCALYYPLHVVMTA  169
107A   : ------MRPRVEQITAELLDEVGDS----GVVDIVDRFAHPLPIKVICEL  156
119(1): ------LETFIRETTRSLLDSI-DP----REDDIVKKLAVPLPIIVISKI  133
119(2): ------LETFIRETTRSLLDSI-DP----REDDIVKKLAVPLPIIVISKI  133
                     F-HELIX
102    : GFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQR-----ANPDDPAYDEN  201
101    : AGLPEE------DIPHLKYLTDQMTRPD----------------------G  189
55A1   : LGVPFN------DLEYLTQQNAIRTNGS--------------------ST  180
108    : LGVPED------DEPLMLKLTQDFFGVHEPDEQAVAAPRQSA-----DEA  208
107A   : LGVDEA------ARGAFGRWSSEILVMD--------------------PER  181
119(1): LGLPIE------DKEKFKEWSDLVAFRL--------------------GK  157
119(2): LGLPIE------DKEKFKEWSDLVAFRL--------------------GK  157
          G-HELIX                                H-HELIX β5-1
102    : KRQFQEDIKVMNDLVDKIIADRKASGE-----------QSDDLLTHMLNG  240
101    : SMTFAEAKEALYDYLIPIIEQRRQK-------------PGTDAISIVANG  226
55A1   : AREASAANQELLDYLAILVEQRLVE-------------PKDDIISKLCTE  217
108    : ARRFHETIATFYDYFNGFTVDRRSC-------------PKDDVMSLLANS  245
107A   : AEQRGQAAREVVNFILDLVERRRTE-------------PGDDLLSALISV  218
119(1): PGEIFELGKKYLELIGYVKDHLN---------------SGTEVVSRVVNS  192
119(2): PGEIFELGKKYLELIGYVKDHLN---------------SGTEVVSRVVNS  192
```

```
                  β5-2     I-HELIX                I-HELIX
102    : K-----DPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPH  285
101    : QV-------NGRPITSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLAKSPE  269
55A1   : QVK-------PGNIDKSDAVQIAFLLLVAGNATMVNMIALGVATLAQHPD  260
108    : KL-------DGNYIDDKYINAYYVAIATAGHDTTSSSSGGAIIGLSRNPE  288
107A   : QDD------DDGRLSADELTSIALVLLLAGFEASVSLIGIGTYLLLTHPD  262
119(1): NLS------------DIEKLGYIILLLIAGNETTTNLISNSVIDFTRF-N  229
119(2): NLS------------DIEKLGYIILLLIAGNETTTNLISNSVIDFTRF-N  229
                                                        #
        J-HELIX                                K-HELIX       β1-4
102    : VLQKAAEEAARVLVDPV-PSYKQVKQLKYVGMVLNEALRLWPTAP-AFSL  333
101    : HRQELIERP------------------ERIPAACEELLRRFSLV--ADGR  299
55A1   : QLAQLKANP------------------SLAPQFVEELCRYHTASALAIKR  292
108    : QLALAKSDP------------------ALIPRLVDEAVRWTAPVK-SFMR  319
107A   : QLALVRADP------------------SALPNAVEEILRYIAPPE-TTTR  293
119(1): LWQRIREE-------------------NLYLKAIEEALRYSPPVM-RTVR  259
119(2): LWQRIREE-------------------NLYLKAIEEALRYSPPVM-RTVR  259
                #                      #+
           β2-1   β2-2   β1-3   K′-HELIX       meander
102    : YAKEDTVLGGEYPLEKGDELMVLIPQLHRDKTIWGDDVEEFRPERFENPS  383
101    : ILTSDYEFHGV-QLKKGDQILLPQMLSGLDERENA-CPMHVDFSRQ----  343
55A1   : TAKEDVMIGDK-LVRANEGIIASNQSANRDEEVFE-NPDEFNMNRK---W  337
108    : TALADTEVRGQ-NIKRGDRIMLSYPSANRDEEVFS-NPDEFDITRF----  363
107A   : FAAEEVEIGGV-AIPQYSTVLVANGAANRDPSQFP-DPHRFDVTRD----  337
119(1): KTKERVKLGDQ-TIEEGEYVRVWIASANRDEEVFH-DGEKFIPDRN----  303
119(2): KTKERVKLGDQ-TIEEGEYVRVWIASANRDEEVFH-DGEKFIPDRN----  303
                                                #
                  β-BULGE       L-HELIX              β3-3
102    : --AIPQHA---FKPFGNGQRACIGQQFALHEATLVLGMMLKHFDFEDHT-  427
101    : -KVS-------HTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIA-  384
55A1   : PP-----QD-PLGFGFGDHRCIAEHLAKAELTTVFSTLYQKFPDLKVAV   380
108    : -PNR-------HLGFGWGAHMCLGQHLAKLEMKIFFEELLPKLKSVELS-  404
107A   : -TRG-------HLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLGI  379
119(1): -PNP-------HLSFGSGIHLCLGAPLARLEARIAIEEFSKRFRHIEI--  343
119(2): -PNP-------HLSFGSGIHLCLGAPLARLEARIAIEEFSKRFRHIEI--  343
          ++        +    +++++ +  +  +
            β4-1     β4-2   β3-2
102    : -NYELDIK-ETLTLKPEGFVVKAKSKKIPLGG                   457
101    : PGAQIQHK-SGIVSGVQALPLVWDPATTKAV                    414
55A1   : PLGKINYTPLNRDVGIVDLPVIF                            403
108    : --GPPRLVATNFVGGPKNVPIRFTKA                         428
107A   : DADDVVWRRSLLLRGIDHLPVRLDG                          404
119(1): --LDTEKVPNEVLNGYKRLVVRLKSNE                         368
119(2): --LDTEKVPNEVLNGYKRLVVRLKSNE                         368
                              #
```

[a] Two different sequence alignments for CYP119 (1 and 2) with respect to other P450s were given due to differences between the B helix and the C melix. Model I of CYP119 was based on alignment 1, and model II of CYP119 was based on alignment 2. Five positions where a conserved Pro was found among the five template P450s are marked with a # sign. Those residues predicted to be in close contact with redox partners were labeled with a + sign.

after the B′ helix extrudes into solvents. This is not seen in any other CYP450s with known structures. Since there is no *a prioro* criterion for determining which template is better for the modeling of this variable region for CYP119, two model structures were constructed. One (model I) is based on sequence alignment 119(1) and uses CYP107A as the template for this region. The other (model II) uses sequence alignment 119(2) and CYP108 as the template for this region.

Figure 1 shows the superposition of the Cα trace of these two final energy-minimized models, model IB and model IIB. As seen in this figure, the variable region between the B helix and the C helix adopts a different backbone conformation in model IB (yellow) and model IIB (magenta) while the backbone conformations for all other regions (in purple for both models) are almost identical.

As described in Methods, Prosa was used to assess and guide the improvement of all initial (model 0) and intermediate models leading to final fully energy-minimized model IB and model IIB. Figure 2a shows the comparison of the Prosa residue−residue interaction energy between the initial model (model 0, dotted line) and the fully energy-minimized model IB (solid line). As seen in this figure, many regions, including (1) residues 25−60 (helix B to region B′), (2) residues 100−200 (helix D to helix H), and (3) the C-terminal end, have been greatly improved as evidenced

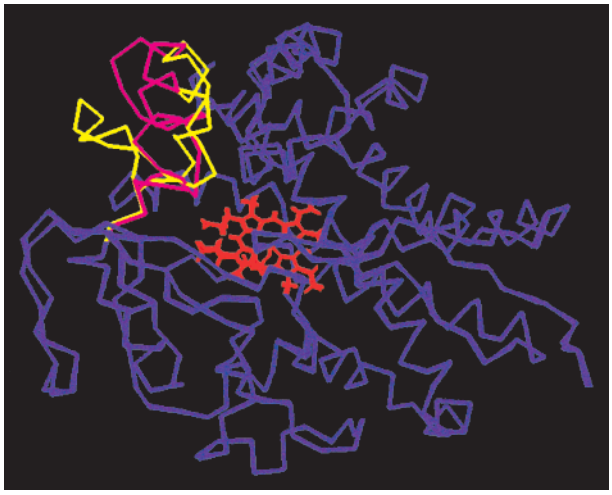

FIGURE 1: Superposition of the Cα trace of two final models, model IB and model IIB, of CYP119: red, heme; yellow, loop between the B helix and the C helix in model IB; magenta, loop between the B helix and the C helix in model IIB; and purple, the rest of the protein in both models.

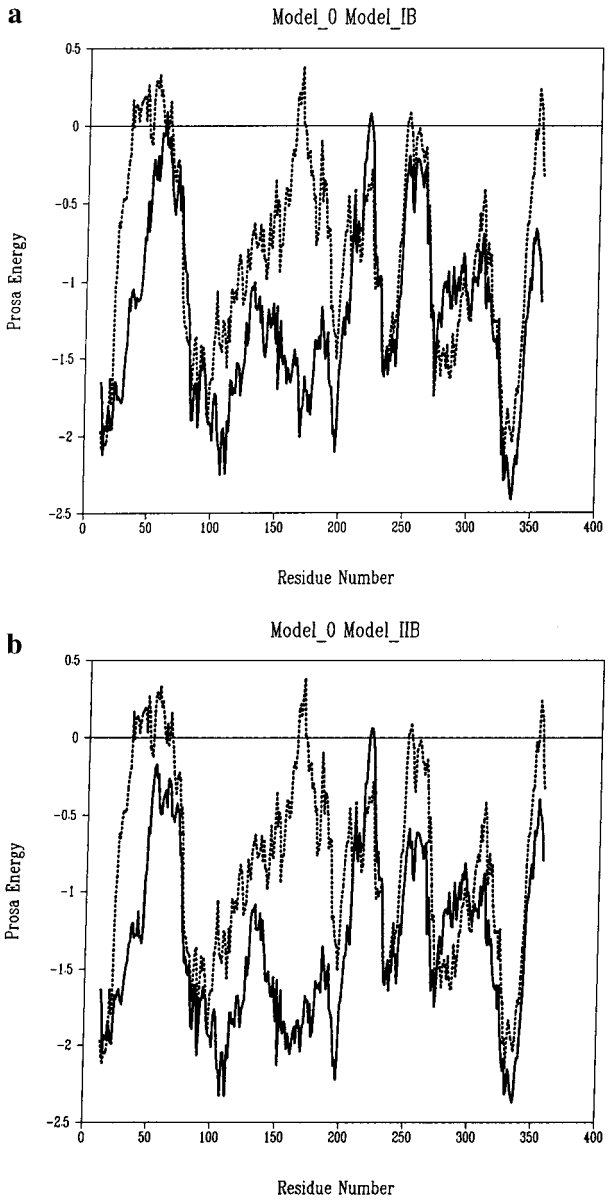

FIGURE 2: Comparison of the Prosa energy plots of different models of CYP119. (a) Initial model, model 0 (dotted line), vs final model, model IB (solid line). (b) Initial model, model 0 (dotted line), vs final model, model IIB (solid line).

by the decrease (i.e., more attractive) in residue−residue interaction energies. In addition, for all regions in model IB, the Prosa interaction energies are attractive, indicating no bad backbone contacts in this model. Similar improvement was also obtained for the second model as demonstrated in Figure 2b in which the Prosa energy plots for model 0 (dotted line) and model IIB (solid line) were compared.

The improvement of the overall quality of the models was also reflected in the values of the Prosa normalized *z* scores listed in Table 2. A value of this *z* score of >0.70 is suggested by the authors of this analysis program to be indicative of a good structure. The Prosa normalized *z* scores for the crystal structures of the P450 enzymes used as templates are all ∼1.0 (*9*). The 0.923 and 0.956 values of model IB and model IIB, respectively, listed in Table 2 approach the values for the crystal structures, indicating the two models to be of very good quality in backbone conformations.

Table 2: Overall Quality of the Two Final Models of CYP119 Assessed by Various Methods

	model IB	model IIB	recommended value for good structures
Prosa[a]	0.923	0.956	>0.7
Whatif[b]	−1.288	−1.189	>−1.0
Procheck[c]	−0.21	−0.22	>−0.50

[a] Normalized *z* score. [b] Quality control value. [c] Overall score as listed in the "Procheck Summary" output.

Further evaluation of the overall quality of the two final models was performed using the criteria embedded in the Whatif and Procheck programs. The results of the evaluation, together with the recommended value for a good protein structure using these two programs, are also listed in Table 2. Compared to the overall quality of the template proteins evaluated in our previous work, the qualities of these two models appear to be very satisfactory based on the recommendations. In addition, the results show that these two final models are of very similar quality, even though model IB and model IIB differ dramatically in the B−C variable region.

Molecular Dynamics Simulations

Using these two fully energy-minimized models of CYP119, molecular dynamics simulations were then performed for two purposes: (1) to further examine the quality of the model structures by checking their stability over long simulations at room temperature (300 K) and (2) to examine the effect of increased temperature on the unfolding of the models by performing the simulations at 390 K, 26 K higher than the melting temperature of CYP119.

Figure 3a shows the plots of the radius of gyration at every 20 ps interval along the 1000 ps trajectory at 300 K and 390 K for model IB. As seen in this figure, for both temperatures, the fluctuations of the radius of gyration are not significant, but they are slightly larger at the earlier stage for the 390 K simulation than for the 300 K simulation. However, the radii of gyration for both simulations converge to similar values.

Figure 3b shows the plots for the Cα rmsd between the minimized structure and each structure obtained at every 20 ps interval along the 1000 ps trajectory at both temperatures. The Cα rmsd values obtained for the 390 K simulation are significantly larger than those obtained for the 300 K simulation. The average deviations of snapshots from the last 500 ps are 1.62 and 2.50 Å for the 300 and 390 K simulations, respectively.

Panels a and b of Figure 4 show the same plots obtained for model IIB. Similar behaviors of the radius of gyration and the Cα rmsd are seen for this second model. The average Cα rmsds of snapshots from the last 500 ps are 1.96 and 2.20 Å for the 300 and 390 K simulations, respectively. The difference in the average Cα rmsd values between 300 and 390 K is smaller for the second model than for the first model.

These results show that the two model structures are quite stable at 300 K for the duration of the 1000 ps simulation at room temperature. It thus provides further support for the good overall quality of these two models.

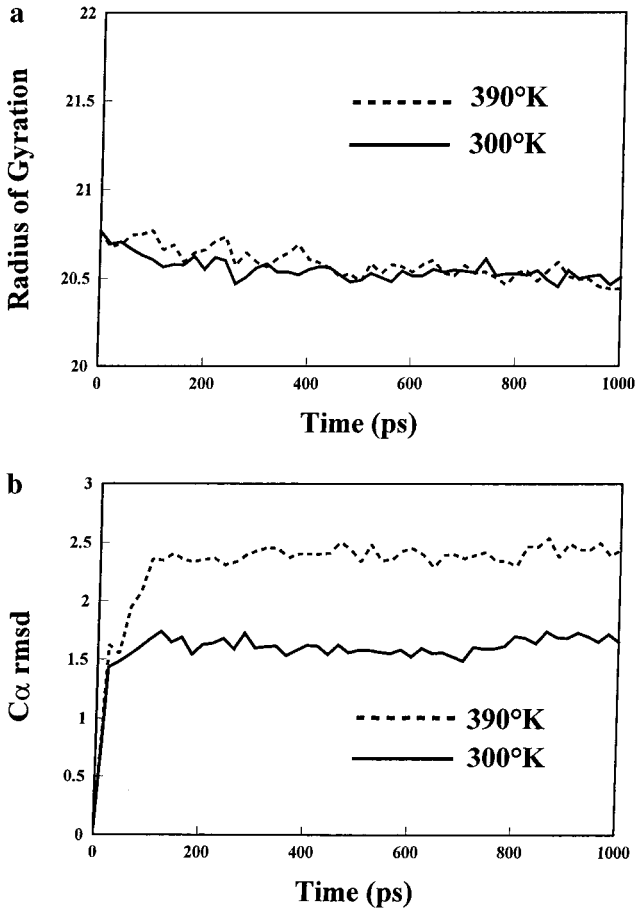

FIGURE 3: Comparison of the MD trajectories at 300 and 390 K for model IB: (a) radii of gyration and (b) Cα root-mean-square deviations from the energy-minimized structure.

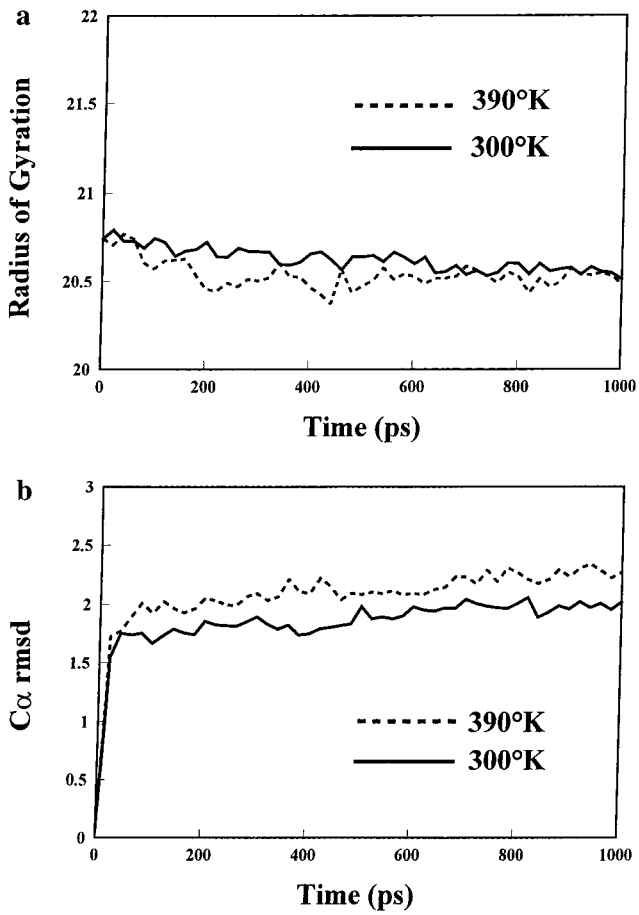

FIGURE 4: Comparison of the MD trajectories at 300 and 390 K for model IIB: (a) radii of gyration and (b) Cα root-mean-square deviations from the energy-minimized structure.

The only indication of decreased stability of the two models of CYP119 at 390 K, 26 K higher than the melting temperature, is the increased Cα rmsd at the higher temperature. There is no other indication of significant unfolding at 390 K during the 1000 ps of simulations. The most likely factor contributing to this behavior is the fact that the time scale of unfolding is much longer than the 1000 ps used in these simulations.

Although it is not possible to observe significant differences in protein conformation from the 1 ns MD simulations of the thermophilic CYP119 at 300 and 390 K, the possibility remains that differential behavior of this thermophile and the mesophilic CYP101 at these two temperatures would provide additional insight into their relative thermal stabilities.

To further assess this possibility, 1 ns MD simulations of the crystal structure of CYP101 were performed at the same two temperatures (300 and 390 K). As for CYP119, one of these temperatures is below and one above the reported melting temperature (54 °C = 327 K) of CYP101 (*5*). The simulation protocols used for CYP101 at these two temperatures were identical to those used for the CYP119 simulations.

The results of these simulations are shown in Figure 5a,b. Figure 5a shows the radius of gyration of CYP101 along the simulations for these two temperatures. As in the simulations of CYP119, the magnitude and fluctuations of the radius of gyration obtained from the 300 and 390 K simulations are quite similar.

Figure 5b shows the plots of the Cα rmsds for simulations at these two temperatures. As in the calculation of the Cα rmsds for the two CYP119 models, the Cα rmsd values are calculated between the fully energy-minimized crystallographic structure of CYP101 (PDB 1phc) and each structure obtained at 20 ps intervals along the 1000 ps trajectory. As seen in this figure, the deviations from the energy-minimized crystallographic structure are much more similar for the 300 and 390 K simulations than they are for the two models of CYP119 shown in Figures 3b and 4b. The average deviations of snapshots from the last 500 ps are 1.65 and 1.71 Å for the 300 and 390 K simulations, respectively. These results taken together indicate that neither the thermophile nor the mesophile exhibits any significant denaturation at temperatures above their observed melting point and provide additional support for the the possibility that the origin of this deficiency is that the time scale of unfolding is much longer than the 1000 ps used in these simulations.

There is, however, the possibility that the minute constraint k (=0.001 kcal/Å^2) imposed on the oxygen atoms of all water molecules could also be contributing to the observed artificial stability. To further investigate this possibility, a 1 ns MD simulation of the meosphilic CYP101 with a known X-ray structure was performed at a much higher temperature of 600 K. The following question was addressed. Would evidence for significant changes in protein conformation be obtained at this temperature, despite the constraint on the

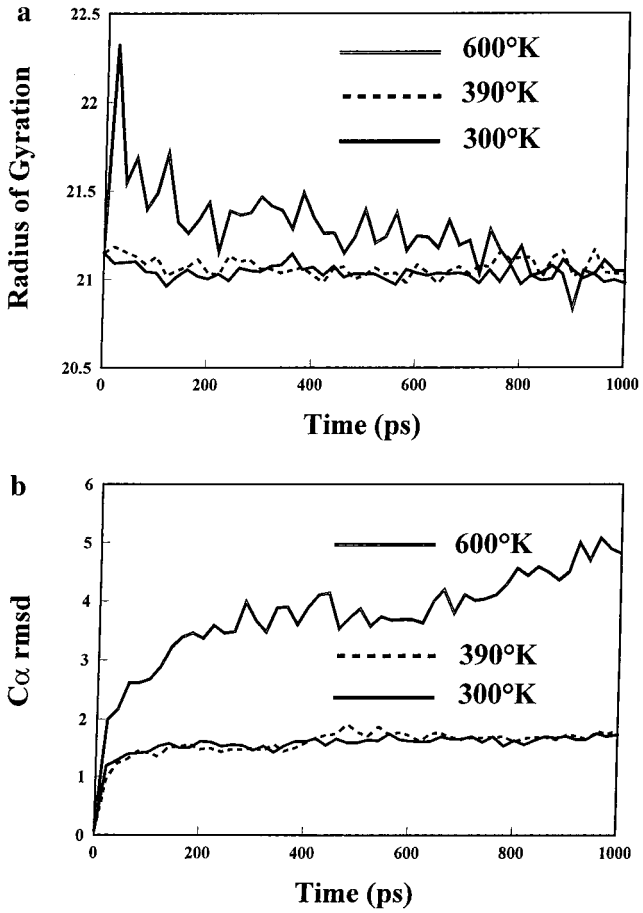

FIGURE 5: Comparison of the MD trajectories at 300, 390, and 600 K for the energy-minimized crystal structure of CYP101: (a) radii of gyration and (b) Cα root-mean-square deviations from the energy-minimized structure.

oxygen atoms of the water molecules? The protocols used for this simulation were identical to those used at the lower-temperature, 390 K, simulations.

The results are shown in panels a and b of Figure 5, in comparison with those obtained at 300 and 390 K. As seen in this figure, both the radius of gyration and the Cα rmsd in the 600 K simulation indicate instability and possibly the beginning of unfolding of CYP101. The Cα rmsd value obtained for the 600 K simulations increases throughout the simulation and reached a value of 5.0 Å near the end of the 1000 ps simulation. The average deviation of snapshots from the last 500 ps for this 600 K simulation is 4.2 Å. These large deviations are indicative of a significant change in protein structure despite the small constraint on the oxygen atoms of the water molecules. In addition, along this same trajectory, many water molecules are leaving the surface of CYP101. For example, at the end of the 1 ns simulations, ~40 out of the 191 crystallographic water molecules evaporated with the farthermost one more than 80 Å away from the surface of the system. The significant change in protein structure and the evaporation of water molecules at 600 K indicate that the very small constraint that is used does not impose significant artificial stability of CYP101. It is thus concluded that the lack of robust signs of unfolding for CYP119 and for CYP101 above melting temperatures (390 K) is mainly due to the relatively short time scale (1000 ps) used in these simulations.

Possible Factors Contributing to Enhanced Stability of CYP119

The enhanced thermo- and barostabilities of CYP119 compared to those of the mesophilic CYP101 have been detected experimentally in two laboratories (*5*, *6*). To the best of our knowledge, the stabilities of the other P450s, including CYP102, CYP107A, CYP108, and CYP55A1, with known structures at elevated temperature and pressure have not been reported. However, since none of these four P450s were known to come from an extremophilic organism, it is likely that these enzymes do not exhibit enhanced stability.

Using a model structure of CYP119 built with the automated program Modeler, McLean et al. (*5*) reported the clustering of aromatic residues around W281 as the only feature found in their model which was not seen in the crystal structures of CYP101 and CYP107A. In this study, possible origins of such thermo- and barostabilities of CYP119 were investigated by comparing the amino acid compositions and the 3D structures of CYP119 with those of the five template CYP450s.

For consistency in presenting the results from sequence analysis and structure analysis, only those residues resolved in the crystal structures were included. Thus, residues 10−414 from PDB 1phc of CYP101, residues 1−457 from monomer A of PDB 2hpd of CYP102, residues 2−404 from PDB 1oxa of CYP107A, residues 1−190 and 207−428 from PDB 1cpt of CYP108, residues 5−403 from PDB 1rom of CYP55A1, and all 368 residues of CYP119 were included for analysis.

Possible Role of Salt Bridges. Table 3 shows the analysis of the amino acid composition of the five template P450s and of CYP119. The residues are divided into three groups: charged residues (DEKRH), uncharged polar residues (CGSTNQY), and hydrophobic residues (AIVLMPWF) (*2*). Although His residues are counted as charged residues in this table, their actual protonation states would depend strongly on the conditions. The number and percentage of each residue are given as well as the subtotal and percentage of each group. Columns 2−11 list the results for CYP101, CYP102, CYP107A, CYP108, and CYP55A1, respectively. Column 12 shows the average percentage for each residue type over the five template P450s. Columns 13 and 14, show the results for CYP119.

As seen in this table, there is a larger overall population (32.4%) of charged residues in CYP119 compared to those of the five template P450s with an average of 27.0%. The population of negatively charged residues, Asp and Glu, is higher than the population of the positively charged residues, Arg and Lys, in all the CYP450s. However, these populations of oppositely charged residues are most evenly matched in CYP119. This enhanced balance implies that among the six P450s that were examined, CYP119 should have the largest relative population of salt bridges.

Examination of the number of salt bridges was thus carried out by using the crystal structures of the five template P450s (1phc for CYP101, both monomers of 2hpd and 2bmh for CYP102, 1oxa for CYP107A, 1cpt for CYP108, and 1rom for CYP55A1) and the two model structures (model IB and model IIB) of CYP119. In this work, a salt bridge is considered to be formed if any side chain oxygen atom of Glu (two O atoms) and Asp (two O atoms) is within 3.5 Å

Table 3: Population of the 20 Amino Acids in the Five P450s with Known Structures and in CYP119[a]

	CYP101		CYP102		CYP107A		CYP108		CYP55A1			CYP119	
	no.	%	no.	%	no.	%	no.	%	no.	%	average %	no.	%
						Charged Residues							
D	22	5.4	33	7.2	34	8.4	31	7.5	23	5.8	6.9	23	6.3
E	32	7.9	36	7.9	30	7.4	24	5.8	25	6.3	**7.1**	35	**9.5**
K	12	3.0	36	7.9	7	1.7	17	4.1	26	6.5	**4.6**	26	**7.1**
R	26	6.4	20	4.4	34	8.4	25	6.1	16	4.0	**5.9**	28	**7.6**
H	13	3.2	13	2.8	8	2.0	11	2.7	7	1.8	2.5	7	1.9
subtotal	105	25.9	138	30.2	113	27.9	108	26.2	97	24.4	**27.0**	119	**32.4**
						Polar Residues							
C	8	2.0	3	0.7	2	0.5	5	1.2	5	1.3	1.1	1	0.3
G	25	6.2	25	5.5	28	6.9	26	6.3	18	4.5	5.9	18	4.9
S	23	5.7	17	3.7	23	5.7	23	5.6	21	5.3	5.2	25	6.8
T	17	4.2	22	4.8	21	5.2	22	5.3	21	5.3	5.0	18	4.9
N	12	3.0	19	4.2	6	1.5	16	3.9	19	4.8	3.5	18	4.9
Q	22	5.4	22	4.8	10	2.5	12	2.9	21	5.3	**4.2**	5	**1.4**
Y	9	2.2	12	2.6	9	2.2	14	3.4	8	2.0	**2.5**	14	**3.8**
subtotal	116	28.7	120	26.3	99	24.5	118	28.6	113	28.5	27.4	99	27.0
						Hydrophobic Residues							
A	30	7.4	33	7.2	39	9.7	38	9.2	41	10.3	**8.8**	12	**3.3**
I	25	6.2	22	4.8	17	4.2	25	6.1	18	4.5	**5.2**	30	**8.2**
V	24	5.9	24	5.3	36	8.9	21	5.1	31	7.8	6.6	23	6.3
L	42	10.4	50	10.9	48	11.9	36	8.7	40	10.0	10.4	41	11.1
M	10	2.5	14	3.1	4	1.0	14	3.4	8	2.0	2.4	5	1.4
P	30	7.4	26	5.7	26	6.5	26	6.3	30	7.5	**6.7**	17	**4.6**
W	5	1.2	5	1.1	4	1.0	5	1.2	2	0.5	1.0	5	1.4
F	18	4.4	25	5.5	17	4.2	21	5.1	19	4.8	4.8	17	4.6
subtotal	184	45.4	199	43.6	191	47.4	186	45.1	189	47.4	**45.9**	150	**40.9**
total	405	100.0	457	100.0	403	100.0	412	100.0	399	100.0	100.0	368	100.0

[a] The sources of the sequence for each P450 are PDB 1phc for CYP101 (405 residues), monomer A of PDB 2hpd for CYP102 (457 residues), PDB 1oxa for CYP107A (403 residues), PDB 1cpt for CYP108 (412 residues), PDB 1rom for CYP55A1 (399 residues), and model structures of CYP119 (368 residues).

Table 4: (a) Number of Salt Bridges Found in the Five P450s with Known Structure and the Two Models of CYP119 and (b) Salt Bridges in CYP119 That Are Formed between Residues Far Apart in Sequence and Their Stability over 1000 ps Molecular Dynamics Simulations[a]

(a)	R−E	R−D	K−E	K−D	H−E	H−D	total	total/sequence length
CYP101	23	15	2	4	4	4	52	12.8
CYP102[b]	15−21	13−17	2−8	5−9	0	0	39−50	8.6−10.9
CYP107A	21	14	1	6	0	0	42	10.4
CYP108	13	10	3	3	3	1	33	8.0
CYP55A1	12	4	10	3	0	0	29	7.3
CYP119 model IB	25	12	4	9	0	0	50	13.6
CYP119 model IIB	27	7	4	11	0	0	49	13.3

(b)	residue	location	residue	location	minimized[c]	IB (300 K)	IB (390 K)	IIB (300 K)	IIB (390 K)
unique	R256	β1-4	E352	β4-2	yes	no	no	no	no
	K358	β4-2	D149	F helix	yes	yes	yes	yes	yes
	K348	C terminus	D3	N terminus	no	no	no	yes	yes
nonunique	R9	A helix	E289	K′ helix	yes	yes	no	yes	yes
	R9	A helix	E296	meander	yes	yes	yes	yes	yes
	R27	B helix	E290	K′ helix	yes	yes	yes	yes	yes
	R227	I helix	D345	β4-1	yes	yes	yes	yes	yes
	R233	J helix	D301	meander	yes	yes	yes	yes	yes
	R324	L helix	E245	K helix	yes	yes	yes	yes	yes
	R337	β3-3	D108	D helix	yes	yes	yes	yes	yes
	R363	β3-2	E114	β3-1	yes	yes	yes	yes	yes

[a] Yes indicates stable and no not stable over MD simulations. [b] The variations represent results from two monomers in p4502hpd.ent and two monomers in p4502bmh.ent. [c] All except one salt bridge are found in the two energy-minimized models of CYP119.

of any side chain nitrogen atom of Arg (three N atoms), Lys (one N atom), and His (two N atoms). With this definition, more than one salt bridge could be formed between the positively charged residues and the negatively charged residues.

Table 4a lists the total number of salt bridges and the normalized population of salt bridges (number of salt bridges per sequence length) found in each structure. As seen in this table, the number of salt bridges found in each of the two models of CYP119 is greater than in all but CYP101 of the five template P450s. Furthermore, the normalized populations of salt bridges are higher in these two models of CYP119 than in any of the five template P450s.

To further probe the significance of the differences in salt bridges, the number of the salt bridges in the mesophile

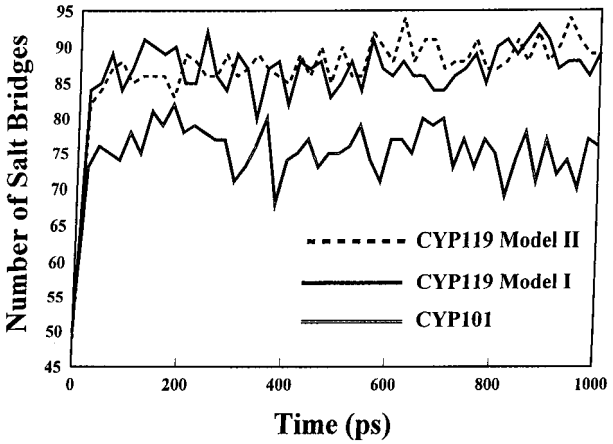

FIGURE 6: Comparison of the number of salt bridges for the two models of CYP119 and for CYP101 during the 1000 ps molecular dynamics simulations at 300 K.

CYP101 and the thermophile CYP119 found during the 1000 ps of MD simulations was determined and compared. Figure 6 summarizes the results obtained at 300 K. As seen in this figure, although the numbers of salt bridges in the two models of CYP119 and of CYP101 are initially nearly identical, the numbers of salt bridges during the MD are much larger in the two CYP119 models than that in CYP101. Similar behavior was seen for the 390 K simulations. These results thus indicate that the contribution of salt bridges to the stability of the enzyme could be larger in CYP119 than in CYP101 when the dynamical behaviors are taken into account.

Of all the stable salt bridges found in the two models of CYP119, those connecting residues that are far apart in sequence could be particularly important for stability. This subset of 11 salt bridges is listed in Table 4b for the two models of CYP119. The location of all these residues, as deduced from Table 1, is also given in this table. Almost all the partners participating in these salt bridges are surface residues. The finding of more salt bridges on the surface of CYP119 is not new in known thermophilic proteins (*31*). A recent MD simulation of the hyperthermophilic protein Sac7d also suggests that surface salt bridges may contribute to protein stability at elevated temperatures (*32*).

Table 4b also indicates whether the salt bridge is or is not unique in CYP119. As seen in this table, there are three salt bridges that are present in CYP119 and absent in all the templates. The remaining eight are nonunique salt bridges that are found in the energy-minimized models of CYP119. Two of the unique salt bridges, one between Arg256 and Glu352 and the other between Asp149 and Lys358, are found in both energy-minimized models of CYP119 (models IB and IIB). Arg256 is located in β1-4, and Glu352 is located in β4-2. Asp149 is located in the F helix, and Lys358 is located in β4-2. None of the five template P450s contain both charged residues in these positions needed to form a salt bridge.

As seen in Table 4b, all except one of the salt bridges are present in the two energy-minimized models of CYP119. It was also important, however, to determine not only the presence but also the stability of these salt bridges. Thus, the stabilities of the salt bridges listed in Table 4b were monitored over the 1000 ps molecular dynamics simulations at 300 and 390 K. The results are summarized in Table 4b. In this table, "yes" means that the salt bridge is stable and "no" means that it is not. As seen in this table, except for the salt bridges between Arg256 and Glu352, between Arg9 and Glu289, and between Asp3 and Lys348, all salt bridges listed in this table are stable in both models and at both temperatures.

A major finding in the MD simulations was the presence of a salt bridge in model II of CYP119 at both 300 and 390 K between Asp3, a residue at the N terminus, and Lys348 near the C terminus. This salt bridge was not present in either of the energy-minimized models (model IB and model IIB). The formation of this salt bridge between Asp3 and Lys348 in CYP119 is unique. Only CYP101 contains charged residues E40 and H391 in both aligned positions, and they do not form a salt bridge in the crystal structure or during the MD simulations at 300 and 390 K. Given the fact that Asp3 and Lys348 are located far apart and at the two termini of the sequence of this enzyme, their contributions to the stability could be significant.

The combined results from the comparative analysis of salt bridges of CYP119 and of the other five P450s and from MD simulations of CYP119 and of CYP101 thus suggest the importance of salt bridges in the stability of the CYP119 enzyme. The residues listed in Table 4b that participate in salt bridge formation between residues far apart in sequence are thus suggested for future mutagenesis studies for probing their role of the overall stability of CYP119.

Possible Role of the Different Distribution of Hydrophobic Residues. As shown in Table 3, there is a smaller overall population (40.9%) of hydrophobic residues in CYP119 compared to those of the five template P450s with an average of 45.9%. This change is due mainly to the net result of the following differences: (a) the significant decrease in the population of Ala residues from 8.8% of the average of the five P450s to 3.3% in CYP119 and (b) an increase in the population of Ile residues from 5.2% of the average of the five template P450s to 8.2% of CYP119.

The decrease in the relative population of Ala residues and the increase in the population of Ile residues in CYP119 as compared to those of the five template P450s are intriguing. Examination of the structures of the five P450s shows a significant population of Ala residues on the surface of each of the five template proteins. Such a population is not seen in the two models of CYP119. It is thus possible that by reducing the number of Ala residues on the surface area and increasing the number of Ile residues in the interior, an improved hydrophobic packing could be achieved in the interior of the protein, thus contributing to the enhanced stability of CYP119.

Possible Role of the Decrease in the Population of Pro Residues. As shown in Table 3, there is a significant decrease in the relative population of Pro residues from 6.7% for the average of the five template P450s to 4.6% for CYP119. The presence or absence of a proline residue has a particularly large effect on protein structure. Specifically, proline-induced kinks in helical segments are very common in known protein structures.

As shown in Table 1, there are five conserved prolines in the templates that are either deleted or mutated in CYP119. These positions are marked with the # sign in Table 1. One of these, in particular, a deletion of a conserved Pro located between the I and J helices, could be important for stability. In all five templates, this conserved Pro residue causes a

~90° kink between the two helices. The I and J helices are very structurally conserved among the five template P450s with known structures. The deletion of the Pro connecting these two helices could be significant for structure stability.

To determine if this missing Pro is a characteristic of only CYP119, an examination of the multiple-sequence alignment of 65 P450s reported by Nelson et al. (*33*) was performed. It was found that the Pro between the I helix and the J helix is highly conserved, missing in only seven of the 65 P450s that were included. Thus, while not completely clear, it is possible that the absence of this proline is a factor in the stability of CYP119. Mutagenesis studies at this position will be very useful for clarifying the effect of the Pro on stability.

Possible Role of the Decrease in the Population of Gln Residues. A decrease in the population of Gln residues from 4.2% for the average of the five P450s to 1.4% for CYP119 was seen in Table 3. From the multiple-sequence alignment in Table 1, it is seen that Gln residues in the five template P450s are often replaced with charged residues in CYP119. For example, Q46 and Q69 of CYP101 are replaced with R9 and R27, respectively, in CYP119; Q27 of CYP102 is replaced with D3 in CYP119; Q62 and Q82 of CYP108 are replaced with E31 and D61, respectively, in CYP119, etc. This observation for CYP119 is consistent with a general observation made for thermophiles by Jaenicke and Bohm (*2*) that the increase in the population of charged residues seen in thermophiles is primarily at the expense of uncharged polar residues and glutamine seems to be significantly discriminated against in thermophiles. A tentative explanation of this phenomenon was also given by Jaenicke and Bohm (*2*), who concluded that the rate of deamindation of glutamine increased at higher temperatures. However, the same argument could also apply to asparaines which do not appear to be discriminated against. Thus, the reasons for the discrimination against glutamine remain unclear.

Possible Role of Clusters of Aromatic Residues. Since the clustering of aromatic residues was suggested as one feature in the model by McLean et al. (*5*) as the possible origin of enhanced stability, the relative populations of the aromatic residues were determined by summing up the relative populations of Phe, Tyr, and Trp residues listed in Table 3. These combined residues comprise 7.8, 9.2, 7.4, 9.7, and 7.3% of all the residues in CYP101, CYP102, CYP107A, CYP108, and CYP55A1, respectively. For CYP119, these three aromatic residues comprise 9.8% of all the residues. Thus, while the relative population of aromatic residues in CYP119 is significantly higher than in CYP101, CYP107A, and CYP55A1, it is similar to that in CYP102 and CYP108. Thus, the population of aromatic residues alone does not seem to be a determinant of the enhanced stability of CYP119.

An analysis was then performed to examine how these aromatic residues cluster in the structures of the five template P450s and the two models of CYP119. For an aromatic residue to be considered in a cluster, this residue must find at least one other aromatic residue in this cluster within a Cα−Cα distance cutoff.

Table 5a lists the results using a Cα−Cα cutoff distance of 7.0 Å. It was found that with this cutoff no cluster in any of the six P450s contains more than five aromatic residues. The results obtained using a 10.0 Å cutoff are also listed in Table 5b. As seen in this table, the largest clusters found among all six P450s are in the two models of CYP119. They contain 20 and 19 aromatic residues, respectively, much larger than any cluster in the templates. These results thus indicate a distinctive structure feature of CYP119 as compared to the structures of the five template P450s, and it could be an important factor for the enhanced stability of CYP119.

Table 5: Number of Clusters Found and Number of Aromatic Residues in Each Cluster for the Five P450s with Known Structures and for the Two Models of CYP119[a]

(a)	no. of clusters	no. of residues in each cluster
CYP101	8	4, 4, 3, 2, 2, 3, 2, 2
CYP102	9	3, 2, 3, 2, 2, 2, 3, 2, 2
CYP107A	8	2, 2, 4, 3, 2, 2, 2, 2
CYP108	9	2, 5, 2, 2, 3, 3, 2, 2, 2
CYP55A1	5	2, 2, 3, 2, 2
CYP119 model IB	8	2, 3, 2, 4, 4, 3, 2, 3
CYP119 model IIB	8	2, 3, 2, 4, 4, 3, 2, 3

(b)	no. of clusters	no. of residues in each cluster
CYP101	7	6, 4, 4, 4, 5, 2, 2
CYP102	9	3, 8, 4, 6, 7, 3, 3, 2, 2
CYP107A	7	8, 3, 2, 2, 6, 2, 2
CYP108	8	2, 3, 9, 11, 2, 4, 2, 2
CYP55A1	5	4, 3, 7, 3, 2
CYP119 model IB	4	19, 2, 2, 3
CYP119 model IIB	4	20, 2, 2, 4

(c)	no. of residues	no. of residues in each cluster
cluster 1	19	Y2, W4, F5, Y15, Y16, W21, F24, Y26, Y28, F36, F39, F225, F228, W231, Y250, Y277, W281, F292, F298
cluster 2	2	F334, F338
cluster 3	2	Y66, Y202
cluster 4	3	F144, W147, Y168

(d)	no. of residues	no. of residues in each cluster
cluster 1	20	Y2, W4, F5, Y15, Y16, W21, F24, Y26, Y28, F36, F39, Y46, F225, F228, W231, Y250, Y277, W281, F292, F298
cluster 2	2	F334, F338
cluster 3	2	Y66, Y202
cluster 4	4	F144, W147, F162, Y168

[a] (a) Cα−Cα distance cutoff of 7.0 Å, (b) Cα−Cα distance cutoff of 10.0 Å, (c) the residues in each cluster of model IB of CYP119 using the 10.0 Å Cα−Cα distance cutoff, and (d) the residues in each cluster of model IIB of CYP119 using the 10.0 Å Cα−Cα distance cutoff.

Parts c and d of Table 5 list the residues in each cluster of model IB and model IIB, respectively. The aromatic residues in the largest cluster in each of the two models are located near either the N terminus (Y2−Y46), the I−J helical region (W231−Y250), or the K′ helix region (Y277−F298). Three residues (Y2, W4, and F5) at the N terminus are located in positions that connect the aromatic residues near the N terminus and those in the I−J and K helical region.

Cluster 4, formed among aromatic residues in the F and G helical regions, including F144, W147, and Y168, could also be important. The three pairwise Cα−Cα distances among these three residues are within 6.5 Å. It is thus worthwhile to probe the contribution of these three residues to the stability of CYP119. The combined results suggest that these three residues in the F−G helix region together with those 19 or 20 aromatic residues in the largest cluster

Table 6: Substrate Binding Site Residues within 10.0 Å of the Heme Iron

model IB	I130, Y202, L205, L206, I208, A209, G210, N211, E212, T213, T214, L217, P253, V254, T257, I282, L354
model IIB	R65, M68, L69, I130, Y202, L205, L206, I208, A209, G210, N211, E212, T213, T214, L217, L248, P253, V254, T257, I282, L354

could be the target for mutagensis studies for probing their contribution to the thermo- or barostability of CYP119.

Identification of Binding Site Residues in CYP119

The active site for substrate metabolism of P450 enzymes is located on the so-called distal side of the heme iron opposite the attachment of the proximal cysteine ligand. The natural substrate(s) of CYP119 is not known. Nevertheless, a binding site pocket was identified consisting of those amino acids within 10.0 Å of the heme iron atom on the distal side of the heme plane. Table 6 lists those active site residues identified using this criterion for model IB and model IIB.

As seen in Table 6, most of the active site residues predicted from these two models are the same. However, a few residues such as R65, M68, L69, and L248 in model IIB identified as being within 10.0 Å of the heme iron atom are not found in model IB. Three of them are located in the loop between the B helix and the C helix. As described earlier, model IB and model IIB differ significantly in the region between the B helix and the C helix. This B−C region is highly variable among known P450s. This region of model IB was obtained by using CYP107A for the sequence alignment and structure building. The resulting model thus mimics CYP107A with the B−C loop extruding out to the solvent. Thus, none of the residues in this region come close to the heme iron. On the other hand, the B−C region of model IIB was obtained by using CYP108 for sequence alignment and structure building. The resulting model thus mimics the structure of CYP108 in this region.

Figure 7a,b shows the binding site of these two models. The main difference between the two binding sites is the presence of R65, M68, and L69 in the vicinity of the C ring of the protoporphyrine in model IIB which are absent in model IB. However, in terms of the space available directly above the four rings of the heme unit, similar topologies are seen between the two models with the same qualitative trend: D > C > A > B.

McLean et al. (*5*) also reported the active site of their model structure. The residues near their active site include Y66, T67, M68, L69, S148, V151, Y168, L205, T213, R256, and T257. However, the criterion used to choose the residues to be shown in their figure was not reported, nor were any details of the model construction and assessment. Thus, it is not possible to deduce the origin of the differences in the active site reported and the two active sites found here.

Identification of Surface Residues in CYP119 That Could Be Involved in Redox Partner Interaction

Since the natural electron donor of CYP119 is not known, the surface region of CYP119 that may interact with redox partners was probed using two alternative partners, putidaredoxin and the FMN domain of P450BM3 (CYP102). Both models of CYP119 were examined for such interaction. As described in detail in Methods, a combined strategy of the protein−protein docking program GRAMM (*27*, *28*) and a statistical analysis of the prediction (*29*) was used for this purpose.

Using putidaredoxin as the probe enzyme for interaction with model IB of CYP119, it was found that all 100 highest-scored predictions locate the proximal side, i.e., the side of the heme plane opposite the substrate binding site, of the CYP119 model structure for binding with putidaredoxin. Similarly, for the prediction of complex formation between CYP119 model IB and the FMN domain of CYP102, all 100 highest-scored predictions locate the proximal side of CYP119 for binding.

The left panel of Table 7a lists the surface residues in CYP119 that are likely to be in close contact with putidaredoxin. The residues listed in the first column in this panel were obtained using a Cα−Cα distance cutoff of 8.0 Å and a frequency of appearance cutoff of 40%. Additional charged residues obtained using a Cα−Cα distance cutoff of 11.0 Å and a frequency of appearance cutoff of 40% are listed in the second column. The right panel of Table 7a lists the residues in model IB of CYP119 found to interact with the FMN domain of CYP102 obtained using the same cutoff criteria as those for the prediction with CYP119 model IB and putidaredoxin. Similarly, the left and right panels of Table 7b list those residues in model IIB of CYP119 that may be in close contact with putidaredoxin and with the FMN domain of CYP102.

Examination of the residues listed in Table 7 produces two results. (1) Most of the residues of CYP119 predicted to be in close contact with putidaredoxin and with the FMN domain of CYP102 are identical. These results imply the same or highly overlapping binding sites for these two redox partners. (2) The results obtained by using model IB are very similar to those obtained by using model IIB. Given the fact that the major difference in backbone conformation between these two models is in the loop between the B helix and the C helix, which is not on the proximal side where binding with redox partners is predicted, the similar results that are obtained are not surprising.

Upon examination of the positions in sequence for these redox-binding residues, it was found that they are scattered among the B helix, the C and C* helices, and the β-bulge/L helix regions as indicated with the + sign in the multiple-sequence alignment (Table 1). Examination of the only known P450−redox partner complex structure of all P450s, i.e., the complex between the heme domain and the FMN domain of CYP102, shows that those residues in the heme domain of CYP102 that have close contact with the FMN domain are also located in these sequence regions (*26*). In particular, the β-bulge/L helix residues that are identified are in proximity to the heme unit.

Among those residues listed in Table 7, Arg80 of CYP119 is aligned (see Table 1) with His100 of CYP102 and with Arg112 of CYP101. There is particularly robust evidence that both of these residues in these two templates play an important role in the recognition of redox partners. Specifically, in the only known structure of a complex (*26*), His100 of CYP102 forms the only salt bridge with a residue in its

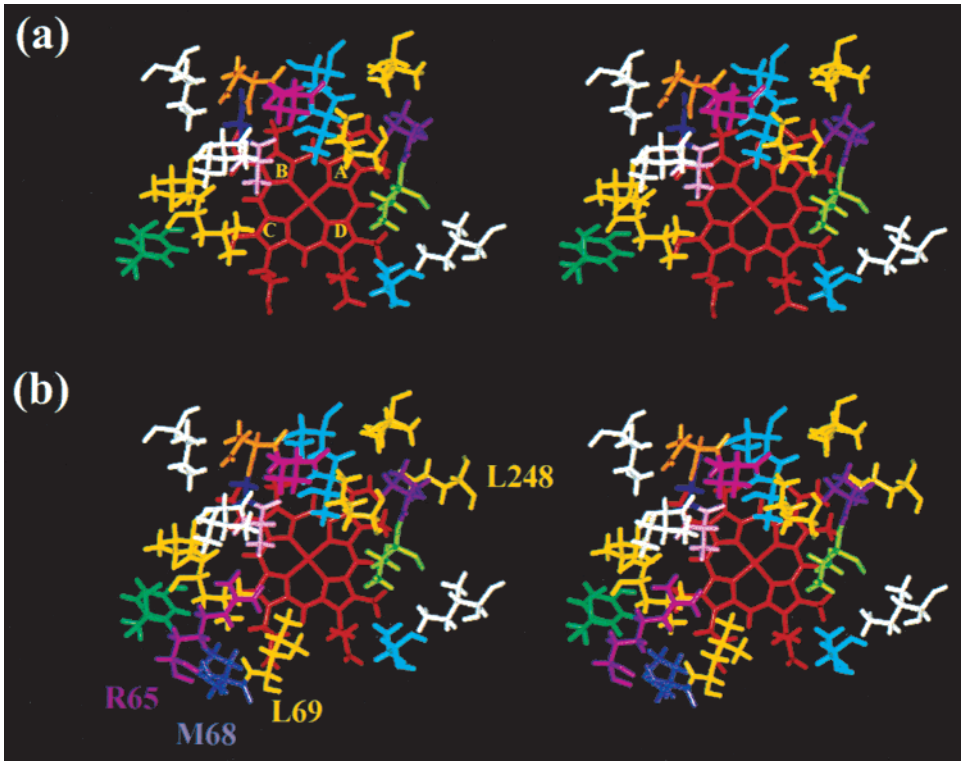

FIGURE 7: Stereoview of the substrate binding site of CYP119 looking from above the heme unit: (a) model IB and (b) model IIB. Rings A−D of the heme group are denoted in panel a. The color scheme of residues is as follows starting from T257 (cyan) near the D ring and going counterclockwise: (a) T257 (cyan), I282 (white), V254 (green), L354 (yellow), P253 (purple), L217 (yellow), T213 (cyan), T214 (cyan), E212 (magenta), N211 (brown), A209 (pink), G210 (blue), I208 (white), I130 (white), L206 (yellow), L205 (yellow), and Y202 (green). The color scheme in panel b is the same as in panel a, and the four additional residues are denoted.

redox partner, E494 of the FMN domain. Mutagenesis studies of Arg112 of CYP101 have established its important role for binding and electron transfer with putidaredoxin (*34*−*37*). Thus, the corresponding residue, R80, in CYP119 is likely to play an important role, too.

These comparisons, taken together, lend support to the predictions for binding site residues of CYP119 that interact with redox partners. Mutation of these residues and the determination of the effect on the binding of putidaredoxin and the FMN domain of CYP102 to CYP119 are thus suggested by these studies.

CONCLUSION

In this work, homology modeling of the only known thermophilic and barophilic P450 enzyme, CYP119 from *S. solfataricus*, was performed using the strategies previously developed in this laboratory. Two models that differ mainly in the loop structure between the B helix and the C helix were constructed. The two models were examined by three protein analysis programs, including Prosa, Whatif, and Procheck, and were found to be of good quality.

Molecular dynamics simulations of these two models at room temperature (300 K) and at an elevated temperature (390 K), which is above the reported melting temperature of CYP119, for 1000 ps were performed to examine the stability. On the basis of the size of the radius of gyration and the Cα rmsd, both models were found to be quite stable at room temperature for the length of the 1000 ps simulation. The radii of gyration for the two models at 390 K are very similar to that obtained at 300 K. The Cα rms deviations obtained for the two models at 390 K are larger than those obtained at 300 K. However, the average deviations of 2.5 and 2.2 Å from snapshots of the last 500 ps of simulations for model I and model II, respectively, at 390 K are still too small to indicate significant unfolding.

Comparative studies of MD simulations of CYP101 at 300 and 390 K show behavior similar to that of CYP119. However, MD simulation of CYP101 at 600 K for 1000 ps shows a significant Cα rms deviation with an average of 4.2 Å for the last 500 ps, thus indicating the beginning of unfolding. It is thus concluded that the lack of an indication of unfolding for CYP119 and CYP101 at 390 K is mainly due to the relatively small length of 1000 ps, not the protocols used for the simulations. Nevertheless, the stabilities of CYP119 at both 300 and 390 K being similar to those of CYP101 provide further support of the high quality of the two models.

Possible origins of the thermo- and barostability of CYP119 were investigated by examining the amino acid compositions and structural features of CYP119 and comparing them with those of the five P450 enzymes with known structures. A higher population of charged residues and of Ile residues and a lower population of Ala, Gln, and Pro residues were found for CYP119 than for the five template P450s. A few unique salt bridges formed by the surface-charged residues were found that could not be found in any of the five template P450s. The decrease in the population of Ala residues and the increase in the population of Ile residues are likely to improve packing in the interior of CYP119. The population of aromatic residues was also

Table 7: (a) List of Residues in CYP119 Model IB That May Have Close Contact with Residues in Putidaredoxin and the FMN Domain of P450BM3 and (b) List of Residues in CYP119 Model IIB That May Have Close Contact with Residues in Putidaredoxin and the FMN Domain of P450BM3[a]

(a)

putidaredoxin		FMN domain of P450BM3	
8.0 Å cutoff	11.0 Å cutoff	8.0 Å cutoff	11.0 Å cutoff
D77	K30	N34	R27
R80	E31	D77	K30
S81	H76	R80	E31
A84	E78	S81	E78
D85	H307	A84	H307
S88	H315	D85	H315
P89	R324	S88	
P306		P89	
G313		N305	
I314		P306	
L316		S312	
L318		G313	
P321		I314	
		L316	
		P321	

(b)

putidaredoxin		FMN domain of P450BM3	
8.0 Å cutoff	11.0 Å cutoff	8.0 Å cutoff	11.0 Å cutoff
N34	K30	N34	R27
D77	E31	D77	K30
R80	H76	R80	E31
S81	E78	S81	H76
A84	H307	A84	E78
D85	H315	D85	K242
S88	R324	S88	H307
P89		P89	H315
P306		N305	R324
S312		P306	
G313		S312	
I314		G313	
L316		I314	
L318		L318	
P321		P321	

[a] All residues obtained using the 8.0 Å Cα−Cα cutoff and additional charged residues obtained using the 11.0 Å Cα−Cα cutoff are listed.

analyzed, and it was found that CYP119 has the densest aromatic cluster of any of the five template P450s. Thus, this characteristic is likely to contribute to the enhanced stability of CYP119.

As pointed out by Jaenicke et al. (*1*, *2*), enzymes have to maintain a well-balanced compromise between stability and flexibility for optimum function. Although extremophilic enzymes are more stable than the mesophilic enzymes, such differences do not exceed ∼100 kJ/mol, which is the equivalent of a few noncovalent interactions. The few characteristics identified here, including the increase in the extent of surface salt bridge formation, improved packing in the interior, and formation of denser aromatic cluster could be part of the origins of the enhanced stability of CYP119. The deletion of a conserved Pro between the I and J helices in CYP119 was not found to be unique, since quite a few other P450s also miss this conserved Pro. Mutagenesis studies of some of the residues listed for salt bridge formation and aromatic clustering and the insertion of the missing Pro would help in assessing the contribution of these factors in the stability of CYP119.

The enhanced stability of CYP119 makes it a useful target for biotechnology application. The optimum function of CYP119 will rely on both the binding of substrates and the interaction with the redox partner to activate this enzyme. Thus, the identification of binding site residues and surface residues for redox partner interactions was also reported and suggested here for future mutagenesis studies.

ACKNOWLEDGMENT

This work was inspired by communication with Dr. Paul Ortiz de Montellano. We also appreciate use of the Cray T3E supercomputer at Pittsburgh and at the San Diego Supercomputing Centers, where the molecular dynamics simulations were performed.

REFERENCES

1. Jaenicke, R. (1991) *Eur. J. Biochem. 202*, 715−728.
2. Jaenicke, R., and Bohm, G. (1998) *Curr. Opin. Struct. Biol. 8*, 738−748.
3. Wright, R. L., Harris, K., Solow, B., White, R. H., and Kennelly, P. J. (1996) *FEBS Lett. 384*, 235−239.
4. Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., and Scholz, I. (1980) *Arch. Microbiol. 125*, 259−269.
5. McLean, M. A., Maves, S. A., Weiss, K. E., Krepich, S., and Sligar, S. G. (1998) *Biochem. Biophys. Res. Commun. 252*, 166−172.
6. Koo, L., and Ortiz de Montellano, P. R. (1999) private communication.
7. Nelson, D. R., Koymans, L., Kamataki, T., Stegeman, J. J., Feyereisen, R., Waxman, D. J., Waterman, M. R., Gotoh, O., Coon, M. J., Estabrook, R. W., Gunsalus, I. C., and Nebert, D. W. (1996) *Pharmacogenetics 6*, 1−42.
8. Sali, A., and Blundell, T. L. (1993) *J. Mol. Biol. 234*, 779−815.
9. Chang, Y. T., and Loew, G. H. (1996) *Protein Eng. 9*, 755−766.
10. Chang, Y. T., Stiffelman, O. B., Vakser, I. A., Loew, G. H., Bridges, A., and Waskell, L. (1997) *Protein Eng. 10*, 119−129.
11. Chang, Y. T., and Loew, G. H. (1999) *Proteins: Struct., Funct., Genet. 34*, 403−415.
12. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1986) *Biochemistry 25*, 5314−5322.
13. Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A., and Deisenhofer, J. (1993) *Science 261*, 731−736.
14. Li, H., and Poulos, T. L. (1995) *Acta Crystallogr. D51*, 21−32.
15. Cupp-Vickery, J. R., and Poulos, T. L. (1995) *Nat. Struct. Biol. 2*, 144−153.
16. Hasemann, C. A., Ravichandran, K. G., Peterson, J. A., and Deisenhofer, J. (1994) *J. Mol. Biol. 236*, 1169−1185.
17. Park, S. Y., Shimizu, H., Adachi, S., Shiro, Y., Iizuka, T., Nakagawa, A., Tanaka, I., Shoun, H., and Hori, H. (1997) *FEBS Lett. 28*, 346−350.
18. Sippl, M. J. (1993) *J. Comput.-Aided Mol. Des. 7*, 473−501.
19. Holm, L., and Sander, C. (1991) *J. Mol. Biol. 218*, 183−194.
20. Dunbrack, R. L., Jr., and Karplus, M. (1994) *Nat. Struct. Biol. 1*, 334−340.
21. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1995) *AMBER*, version 4.1, University of California, San Francisco.
22. Zhao, D., Gilfoyle, D. J., Smith, A. T., and Loew, G. H. (1996) *Proteins: Struct., Funct., Genet. 26*, 204−216.
23. Vriend, G., and Sander, C. (1993) *J. Appl. Crystallogr. 26*, 47−60.
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
25. Pochapsky, T. C., Ye, X. M., Ratnaswamy, G., and Lyons, T. A. (1994) *Biochemistry 33*, 6424−6432.

26. Sevrioukova, I. F., Li, H.-Y., Zhang, H., Peterson, J. A., and Poulos, T. L. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 1863−1868.
27. Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., and Vakser, I. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2195−2199.
28. Vakser, I. A. (1996) *Biopolymers 39*, 455−464.
29. Chang, Y. T., and Loew, G. H. (1999) *Protein Eng.* (submitted for publication).
30. Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A., and Deisenhofer, J. (1995) *Structure 3*, 41−62.
31. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., Freeman, New York.
32. de Bakker, P. I. W., Hunenberger, P. H., and McCammon, J. A. (1999) *J. Mol. Biol. 285*, 1811−1830.
33. Nelson, D. R. (1995) in *Cytochrome P450. Structure, mechanism, and biochemistry* (Ortiz de Montellano, P. R., Ed.) 2nd ed., Appendix A, pp 575−606, Plenum Press, New York.
34. Stayton, P. S., and Sligar, S. G. (1990) *Biochemistry 29*, 7381−7386.
35. Koga, H., Sagara, Y., Yaoi, T., Tsujimura, M., Nakamura, K., Sekimizu, K., Makino, R., Shimada, H., Ishimura, Y., Yura, K., Go, M., Ikeguchi, M., and Horiuchi, T. (1993) *FEBS Lett. 331*, 109−113.
36. Nakamura, K., Horiuchi, T., Yasukochi, T., Sekimizu, K., Hara, T., and Sagara, Y. (1994) *Biochim. Biophys. Acta 1207*, 40−48.
37. Unno, M., Shimada, H., Toba, Y., Makino, R., and Ishimura, Y. (1996) *J. Biol. Chem. 271*, 17869−17874.

BI991966U