J Mol Model (2004) 10:69–75
DOI 10.1007/s00894-003-0169-2

ORIGINAL PAPER

**Saikat Chakrabarti · Jaisurya John · Ramanathan Sowdhamini**

# Improvement of comparative modeling by the application of conserved motifs amongst distantly related proteins as additional restraints

Received: 13 August 2003 / Accepted: 5 November 2003 / Published online: 23 December 2003


**Abstract** Protein comparative modeling has useful applications in large-scale structural initiatives and in rational design of drug targets in medicinal chemistry. The reliability of a homology model is dependent on the sequence identity between the query and the structural homologue used as a template for modeling. Here, we present a method for the utilization and conservation of important structural features of template structures by providing additional spatial restraints in comparative modeling programs like MODELLER. We show that root mean square deviation at $C^{\alpha}$ positions between the model and the corresponding experimental structure and the quality of the models can be significantly improved for distantly related systems by utilizing additional spatial restraints of the template structures. We demonstrate the influence of such approaches to homology modeling during distant relationships in understanding functional properties of protein such as ligand binding using cytochrome P450 as an example.

**Keywords** Comparative modeling · Template structure · Spatial restraints · Structural motifs

## Introduction

The high-throughput sequencing of entire genomes of model organisms is providing an overwhelming flood of sequence information that is exponentially on the increase. In comparison, the rate of structure determination of proteins is low. Thus, the structural gap—the difference between the number of known sequences and the number of experimentally determined three-dimensional structures—is widening rapidly. As a consequence, homology modeling, using methods like MODELLER, [1] serves as a convenient platform for the structural understanding of new proteins. The availability of models can often shed some light on the function and its underlying mechanism. [2] They can also provide insight to design experiments and suggest possible leads for rational drug design.

The CASP modeling experiments [3, 4] reiterate the fact that model building by homology can provide remarkably good results if the sequence identity is high between the protein to be modeled and the template (>75%). But at low sequence identities (<40%), the reliability of homology model is low. There are structural deviations between superfamily members that may be reflected not only in loop regions but also as shifts in secondary structural positions that are hard to model using comparative modeling techniques. This has spurred many researchers into improving the accuracy of model-building-by-homology for addressing distant relationships. Here, we present a method for the utilization and conservation of important structural features of template structures by providing additional spatial restraints in MODELLER. [1] We have tested this strategy using proteins for which structural information is available and deposited in the Protein Data Bank (PDB). The root mean square (RMS) deviation between the model and the corresponding PDB structure, measured at $C^{\alpha}$ positions, can be significantly improved by utilizing additional spatial restraints for the structurally conserved regions of the template structures. Models derived using spatial restraints for a hypothetical sequence fare better in structure validation techniques. Proteins belonging to superfamilies have similar structure and function [5, 6] despite poor sequence identity. We utilize the spatial orientation patterns (distance and torsion angles) of the structurally conserved regions or motifs identified within superfamilies [7] in comparative modeling to improve the accuracy of the models. We show that accurate models, utilizing spatial restraints from other superfamily members, can lead to better binding of the ligand molecule and

S. Chakrabarti · R. Sowdhamini (✉)
National Centre for Biological Sciences (TIFR),
UAS-GKVK campus, Bellary Road, 560065 Bangalore, India
e-mail: mini@ncbs.res.in
Tel.: +91-80-3636421/8 Ext. 4240/1
Fax: +91-80-3636665

J. John
A.K. College of Engineering, Madurai, India

this is shown using members of cytochrome P450 family as an example. The availability of accurate models has application in biologically significant problems such as substrate specificity and areas such as drug design.

## Materials and methods

Structural motifs for the superfamilies were obtained from the SMoS database. [7] The aligned protein superfamily databases [6, 7] were scanned to identify triads of proteins within a superfamily where two members are equidistant in terms of sequence identity and RMS deviation from the third one. Three-dimensional modeling of the proteins was performed using the program MODELLER. [1] RMS deviations between the protein structures were obtained by STAMP. [8] PSI_BLAST [9] was used to identify homologous proteins for superfamily members against the non-redundant sequence database at NCBI. Fold prediction methods, like 3D-PSSM, [10] GenThreader, [11] were used to predict the three-dimensional structure for new protein sequences with no characterized structure or function. VERIFY3D, [12] PROCHECK [13, 14] and MOLPROBITY [15] are used for structure validation of the model for the hypothetical sequence. Compatibility to accommodate ligand binding is assessed, in the cytochrome P450 superfamily, by the rigid-body superposition of ligand from the query crystal structure and examining the root mean square deviation of the $C^{\alpha}$-positions of all residues within 5 Å from the ligand between the crystal structure and model coordinates.

## Results and discussion

### Structural motifs for the superfamilies

A superfamily is a hierarchical classification, which contains proteins of different families and subfamilies with similar fold. [6, 16] These proteins might have very low sequence identities but retain similar structure through well-conserved secondary structural parts. On the basis of criteria, such as amino acid preference, solvent accessibility, secondary-structure content, hydrogen-bonding pattern, non-polar interaction and residue packing, several conserved regions of the proteins belonging to the same superfamily have been identified. The presence of all these structural parameters at equivalent alignment positions in a superfamily is examined to obtain residue segments with high structural conservation. These segments are termed as "structural motifs" and are integrated into a mainstream database of motifs called SMoS. [7] In this study, we have utilized the spatial orientations of the motifs, i.e. the virtual distance and torsion angular patterns of the structural motifs for any given superfamily structure. Only the $C^{\alpha}$ positions were considered at the motifs and inter-motif distance and angular patterns were calculated. The average distances and angles of the motifs were represented in a matrix format for all proteins considered within a superfamily. These matrices contain the spatial information for structurally important regions of all the superfamilies.

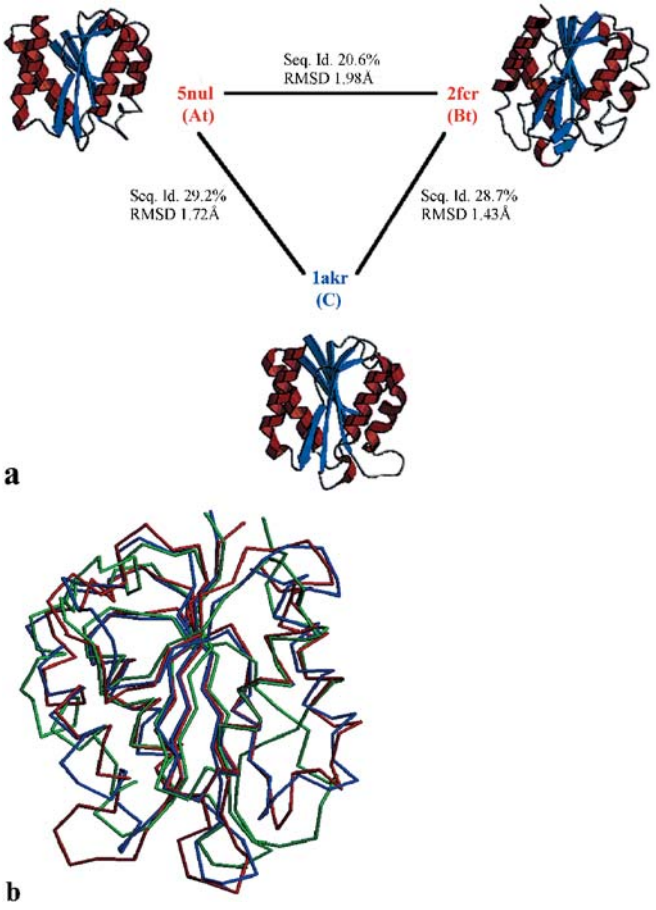


**Fig. 1** Example of a modeling triad. **a** Structures of flavodoxin from *Clostridium beijerinckii* (PDB code, 5nul) and flavodoxin from *Chondrus crispus* (PDB code, 2fcr) that are equidistant in sequence identity (Seq. ID.) and structural similarity (RMSD) to the third flavodoxin member (from *Desulfovibrio vulgaris*, PDB code, 1akr). Structural representations of the proteins are generated using MOLSCRIPT. [19] **b** $C^{\alpha}$ representation of the superimposed coordinates of the original and both model structures of flavodoxin from *Desulfovibrio vulgaris* (PDB code, 1akr) generated with and without the spatial restraints. 1akr PDB structure is shown in *red* where the models with and without the spatial restraints are shown in *blue* and *green*, respectively. This figure has been generated using RASMOL [20]

### Choice of modeling triads: distantly related proteins belonging to the same superfamily

Figure 1a shows one example of a modeling triad that includes flavodoxin from *Clostridium beijerinckii* (PDB code, 5nul) and flavodoxin from *Chondrus crispus* (PDB code, 2fcr) that are almost equidistant in sequence similarity and distantly related to the third flavodoxin member (from *Desulfovibrio vulgaris*, PDB code, 1akr). Fold prediction results also shows 5nul and 2fcr as the top two hits for the 1akr sequence. Although the structure of *D. vulgaris* is known, we do not use this information and instead examine the efficacy of the modeling strategy by comparing the model with the structure of 1akr. Due to

**Table 1** RMS deviation at $C^{\alpha}$ positions between models and the corresponding PDB structure. The protein notations (A, B and C) are as in Fig. 1a

| Models of C | RMS deviation with 1akr structure | RMS deviation with 5nul structure |
|---|---|---|
| Br[a] | 1.25 Å | 1.82 Å |
| B[b] | 1.49 Å | 1.92 Å |
| AB[c] | 1.79 Å | 1.00 Å |
| Models of C | RMS deviation with 1akr structure | RMS deviation with 2fcr structure |
| Ar[a] | 1.55 Å | 1.75 Å |
| A[b] | 1.79 Å | 1.79 Å |
| AB[c] | 1.79 Å | 1.55 Å |

[a] Br and Ar: models of 1akr built based on two template structures [2fcr (Bt) and 5nul (At)] separately, using superfamily distance patterns as distance restraints
[b] B and A: model based on same templates without using the motif distance pattern as distance restraints
[c] AB: model built on both template structure (both 5nul and 2fcr simultaneously) without using the motif distance restraints

**Table 2** RMS deviation at $C^{\alpha}$ positions between models and PDB structures

| Name of the superfamily | Models generated using spatial restraints | Models generated without using spatial restraints (single template) | Models generated without using spatial restraints (two templates) |
|---|---|---|---|
| Flavoproteins | 1.25<br>1.55 | 1.49<br>1.79 | 1.79<br>1.79 |
| Chromo domain-like | 1.25<br>1.82 | 1.32<br>1.9 | 1.84<br>1.84 |
| PRTase-like | 1.84<br>1.9 | 1.9<br>1.94 | 1.85<br>1.85 |
| DHS-like NAD/FAD-binding domain | 1.84<br>2.06 | 1.83<br>2.18 | 1.95<br>1.95 |
| DNA ligase/mRNA capping enzyme, catalytic domain | 2.09<br>2.19 | 2.29<br>2.39 | 2.27<br>2.27 |
| DsRNA-binding domain-like | 1.73<br>1.72 | 1.83<br>1.83 | 1.85<br>1.85 |
| "Helical backbone" metal receptor | 2.53<br>2.71 | 2.7<br>2.84 | 2.82<br>2.82 |
| Glutathione synthetase ATP-binding domain-like | 2.03 | 2.15 | 2.6 |
| Ribosome inactivating proteins (RIP) | 2.03 | 2.14 | 2.25 |

similar values of sequence identities, one can assume that the 1akr sequence has an equal probability of attaining both the 5nul and the 2fcr structure. In this case, either one of the two structural homologues could be selected as the reference template structure for three-dimensional modeling. It is known that employing more than one template, where sequence identity is poor (<25% sequence identity), does not usually lead to a better homology model. [17]

## Application of spatial restraints in the generation of homology models: tests using known examples

One of the structural homologues was used as a template and the spatial information between structural motifs observed in the second structural homologue was provided as restraints. The distance and angle restraints are provided in the top file of MODELLER under the "special_restraints" routine. For each MODELLER run, one template structure together with the average spatial orientation pattern of the structural motifs for the whole superfamily was provided. In the example shown in Fig. 1a, the three-dimensional model of the 1akr sequence was built using the 5nul structure as template and utilizing the spatial orientation patterns of the important structural motifs for both 5nul and 2fcr. Figure 1b shows a superimposition of the 1akr structure and both the models based on 2fcr structure, with and without the spatial restraints of the 5nul structure. The resulting three-dimensional model was indeed closer to the 1akr structure in terms of RMS deviation at $C^{\alpha}$ positions compared to the 1akr model without the spatial orientation restraint (Table 1). Interestingly, the resultant structure obtained by the present strategy is closer to the crystal structure (1akr) than an alternate strategy where both 5nul and 2fcr are simultaneously utilized as templates.

Similarly, nine modeling triads were identified from the superfamily alignment databases and three-dimensional models were generated using superfamily spatial restraints starting from reliable structure-based sequence alignments. The spatially restrained models and the

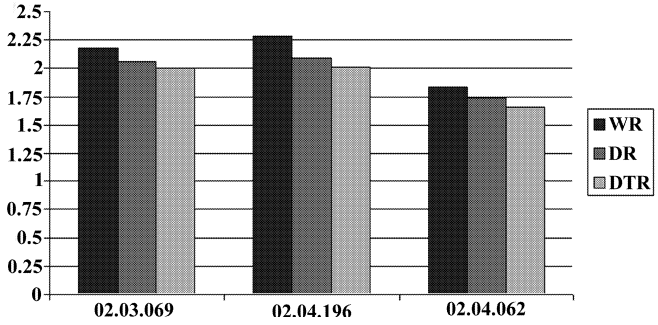


**Fig. 2** The application of spatial restraints in comparative modeling. The overall accuracy of the three-dimensional models, in terms of RMS deviation at $C^{\alpha}$ atoms, improves when distance restraints (DR) alone and both the distance and torsion angle orientation patterns (DTR—distance and torsion angle restraints) are used as additional spatial restraints for modeling as compared to a default homology modeling run without restraints (WR). Three-dimensional models of the proteins from the superfamilies DHS-like NAD/FAD-binding domain (code: 02.03.069), DNA ligase/mRNA capping enzyme, catalytic domain (code: 02.04.196) and dsRNA-binding domain-like (code: 02.04.062) are generated using spatial restraints for the structural motifs and compared with the models generated without any restraints

models without spatial restraints generated using single and both templates were compared for RMS deviation (measured at $C^{\alpha}$ positions) to the corresponding PDB structural entry. Table 2 consolidates this information for the nine triads. In most instances, the RMS deviations for the distance-restrained models are closer to their corresponding PDB structure when compared against the models without restraints using single or both templates. Figure 2 shows that the overall accuracy of the three-dimensional models improves even further when both the distance and torsion angle orientation patterns are used as additional spatial restraints for modeling. This approach inherently assumes that the alignment used in modeling runs would be accurate and reliable. Since accurate alignments for distantly related proteins are hard to obtain, this can in turn influence the quality of the model. In order to obtain reliable alignments of distantly related proteins where structural information may not be available for all them, we have developed a method that guides the alignment by fixing conserved regions or motifs (Saikat Chakrabarti, Prem Anand, Nitin Bharadwaj and R. Sowdhamini, unpublished results).

## Three-dimensional model of hypothetical protein CAC12685.1

In the modeling runs for the nine triads where structural information was known for the query, the primary quality check was to compare the derived model with the structural data. Another RMSD value was recorded that is independent of the query structure, i.e. to compare the derived model with respect to the other member of the triad that has not been directly used as a template in modeling (secondary template). This additional parameter can enable the choice of a better model in cases where there is no structural information available for the query, for example, a hypothetical protein from a genome database.

The phosphohistidine domain superfamily includes pyruvate phosphate dikinase (PDB code 1dik) and enzyme I of PEP: sugar phosphotransferase system (PDB code 1zym). CAC12685.1 is a hypothetical protein from *Thauera aromatica*, identified by BLAST searches (see Materials and methods for details) to be distantly related to both 1dik and 1zym (E values: 41e-01 and 66e-01, respectively). Fold prediction results of CAC12685.1 by 3D-PSSM [10] and Gen_Threader [11] also show its high structural compatibility to both the pyruvate phosphate dikinase (1dik) and PEP enzyme I (1zym) structures (Table 3). Thus pyruvate phosphate dikinase, PEP enzyme and CAC12685.1 form a perfect modeling triad as discussed before, having an average sequence identity between 30–35% among each other. Three-dimensional models for CAC12685.1 have been generated utilizing 1dik and 1zym as templates, separately, with the spatial restraint patterns for the structural motifs identified for phosphohistidine superfamily. Figure 3 shows a cartoon representation of the three-dimensional model structure CAC12685.1 using 1dik as template (primary template) and utilizing spatial orientation restraints for both 1dik and 1zym (secondary template). Table 4 lists the RMS deviation for the models with respect to the secondary template that shows the utilization of spatial restraints results in a model that is closer to the secondary template.

**Table 3** Fold prediction results for the hypothetical protein CAC12685.1

| Fold prediction method | PDB code | Superfamily | Family | Protein |
|---|---|---|---|---|
| | 1dik_2 | Phosphohistidine Domain | Pyruvate phosphate Dikinase | Pyruvate phosphate Dikinase |
| | 1zyma | Phosphohistidine Domain | Enzyme I of PEP | Enzyme I of PEP |
| 3D_PSSM | 2dika | Phosphohistidine Domain | Pyruvate phosphate Dikinase | Pyruvate phosphate Dikinase |
| | 1dik_3 | Phosphohistidine Domain | Pyruvate phosphate Dikinase | Pyruvate phosphate Dikinase |
| | 1ave | Annexin | Annexin | Annexin |
| | 1dik_2 | Phosphohistidine Domain | Pyruvate phosphate Dikinase | Pyruvate phosphate Dikinase |
| | 1zyma | Phoshohistidine Domain | Enzyme I of PEP | Enzyme I of PEP |
| Gen_Threader | 1d8c | Malate synthase G | Malate synthase G | Malate synthase G |
| | 1ecxa | PLP dependent transferase | Cystathone synthase | Selenocysteine lyase |
| | 1cr6b | HAD_like | Epoxide hydrolase | Epoxide hydrolase |

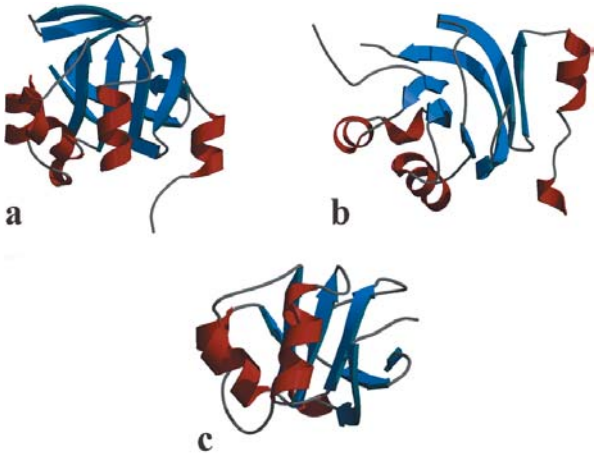


**Fig. 3** Cartoon representation of the three-dimensional model of the hypothetical protein CAC12685.1 (**c**) from *Thauera aromatica* that belongs to the phosphohistidine superfamily using pyruvate phosphate dikinase (**a**) (PDB code: 1dik) as template (primary template) and utilizing spatial orientation restraints for both pyruvate phosphate dikinase and enzyme i (**b**) (PDB code: 1zym) (secondary template). Figures have been generated using the program MOLSCRIPT. [20]

**Table 4** RMS deviation at $C^{\alpha}$ positions between models and PDB structures

| Models of CAC12685.1 | 1dik structure (secondary template) |
|---|---|
| Ar | 1.35 Å |
| A | 1.43 Å |
| AB | 1.41 Å |
| Models of CAC12685.1 | 1zym structure (secondary template) |
| Br | 1.37 Å |
| B | 1.45 Å |
| AB | 1.41 Å |

Three-dimensional modeling of the hypothetical sequence CAC12685.1, utilizing the spatial orientation pattern of the structural motifs, provides a reasonable model that is structurally closer to both the structural templates of the superfamily. The RMS deviation at $C^{\alpha}$ positions with their corresponding PDB structure is lower when compared to models built without spatial restraints. The model of the hypothetical sequences obtained using various schemes were examined by three independent structure validation methods like VERIFY3D, [12] PROCHECK [13, 14] and MOLPROBITY criteria [15] and the results were compared between models obtained from different schemes for the hypothetical sequence. Models that utilize spatial restraints from the other superfamily members acquire a better profile by VERIFY3D (Fig. 4), suggesting a better structural compatibility to the hypothetical sequence. VERIFY3D scores above the threshold of 0.0 indicate good local structural environments to individual residues. Local errors that VERIFY3D identifies in the C-terminal part of the hypothetical protein are removed in the restrained-derived model (shown on the right in Fig. 4). The percentage of

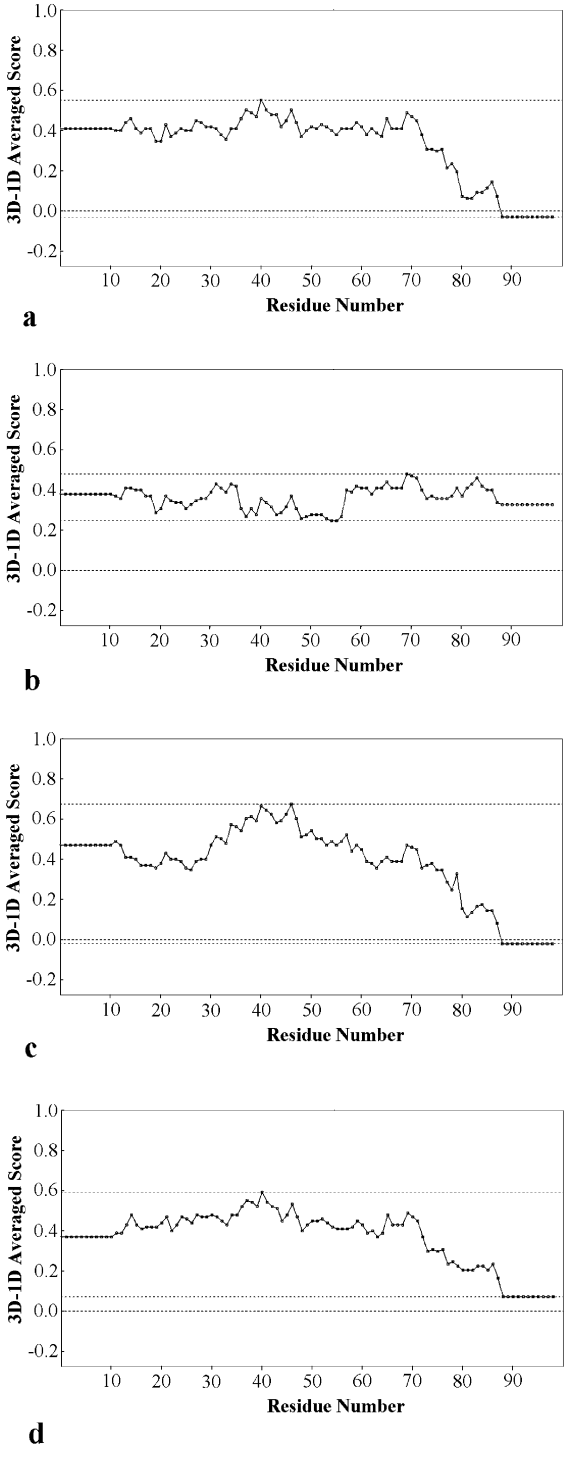


**Fig. 4** VERIFY3D [12] plots of homology models of the hypothetical protein (code: CAC12685.1). See Fig. 3 for the templates chosen for modeling. **a** Model generated using template A (PDB code: 1dik) alone. **b** Model generated using template A and spatial restraints from both A and B. **c** Model generated using template B (PDB code: 1zym) alone. **d** Model generated using template B and spatial restraints from both A and B. Local errors (shown as reduced VERIFY3D scores) observed in the models generated using single templates **a** and **c** are rectified after including spatial restraints from the secondary template (as seen in **b**)

residues in Ramachandran-disallowed regions are also lower for the spatial restraint-derived model as examined using PROCHECK. 71% of residues were within strictly allowed regions in Ramachandran maps for the models derived from default runs, whereas in the spatial-restrained models, 77% of residues were in strictly allowed regions. 4.5% and 1% of residues were in disallowed regions for the unrestrained and restrained models, respectively. In addition, residue–residue contacts were much more preferable for the restrained-derived model by MOLPROBITY criteria. For the restraint-derived models, the number of $C^{\beta}$–$C^{\beta}$ contacts that were within 0–0.125 Å from preferred values is about 79% compared to 71% of residues in unrestrained models.

## Accurate homology models for better ligand binding: cytochrome P450 superfamily as an example

The improvement in the accuracy of a model can influence the understanding of the biological properties of the protein such as ligand binding and substrate specificity. For example, let us consider hemethiolate proteins belonging to the superfamily of cytochrome P450 that are involved in the oxidative degradation of toxins and a variety of endogenous compounds. They recognize hydrophobic ligands such as camphor and imidazole and are characterized by their high substrate specificity. Members of this superfamily retain high overall structural similarity despite poor sequence identity. We chose a triad from this superfamily: cytochrome P450–102 domain from *Bacillus megatarium* (PDB code 1jpz), cytochrome P450–cyp119 from *Fusarium oxysporum* (PDB code 1io7) and cytochrome P450–51 from *Mycobacterium tuberculosis* (PDB code 1e9x) (Fig. 5a) of poor pairwise sequence identity. We have chosen 1jpz as the query (Q) and no structural information about the query is utilized in the modeling exercise. All members of this triad employ protoporphyrin IX containing Fe as the cofactor. Cytochrome P450–102 (Q) recognizes *N*-palmitoylglycine as the substrate and the crystal structure of Q is in the substrate-bound form. One of the templates (A, 1io7) is bound to the cofactor alone but is in the substrate-unbound form, whereas the other template (B, 1e9x) structure is determined bound with the cofactor as well as the substrate (4-phenyl-1H-imidazole). Figure 5b shows $C^{\alpha}$-trace of template A with structural templates or conserved regions marked. Although none of the conserved regions are close to the substrate-binding site, large loop variations could be modeled depending on the choice of template and the inclusion of these conserved regions as spatial restraints (see below for details).

The crystal structures of three cytochrome P450 proteins are shown near the substrate-binding site (Fig. 5c: query (1jpz, shown in *gray*), template A (1io7, shown in *green*) and template B (1e9x, shown in *red*) have distinct loop conformations in this region). In the crystal structure of 1jpz, one of the substrate binding regions adopts a helical conformation at the backbone.

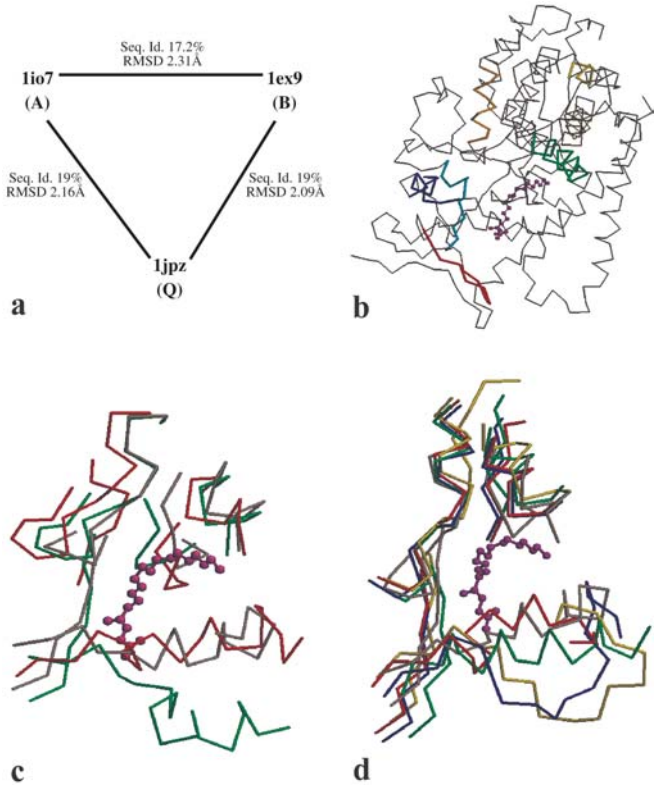


**Fig. 5** Comparative modeling on a triad from cytochrome p450 superfamily and ligand binding properties. **a** Three proteins from cytochrome P450 superfamily have been chosen by consulting CAMPASS database. [16] The sequence identity and structural similarity, as measured by overall root mean square deviation, are provided in a pairwise manner. One of them, 1jpz, is chosen as the query for homology modeling. No structural information of 1jpz is used in the entire modeling exercise. **b** Backbone representation of the crystal structure of cytochrome P450–102 domain from *Bacillus megatarium* (PDB code 1jpz). In certain modeling schemes, the structural templates were employed as spatial constraints and are marked in different colors. The bound ligand molecule is also shown. None of the restraints are proximate to the ligand binding site. **c** Superimposition of the crystal structures of cytochrome P450–102 from *Bacillus megatarium* (PDB code 1jpz, (Q), shown in *gray*), cytochrome P450-cyp119 from *Fusarium oxysporum* (PDB code 1io7, (A), shown in *green*) and cytochrome P450–51 from *Mycobacterium tuberculosis* (PDB code 1e9x, (B), shown in *red*). Ligand binding regions alone are shown for the sake of clarity. The three proteins adopt different loop conformations at this region, especially at the loop to the right. To note that the structure of A has been determined in the absence of substrate but in the co-factor bound form, whereas template B and query Q are in the substrate bound form. The substrate of Q is superposed at the ligand-binding site. **d** Results of homology modeling using 1jpz as (Q) and choice of 1io7 (A) and 1e9x (B) as two possible templates. Four different schemes were used for modeling using MODELLER (1). In each case, the close-up of the model conformations at the ligand-binding region is shown and compared with the crystal structure of Q, shown in *gray*). The model derived from template A alone is shown in *yellow*, model derived from template A including restraint is shown in *red*, model derived from template B alone is shown in *green* and model derived from template B including spatial restraints is shown in *blue*. The substrate of Q is shown at the ligand-binding site. The models and the query crystal conformation are projected using RASMOL [20] after rigid-body superposition and best fit. **b**, **c** and **d** are provided in similar orientations. The model derived from template A including spatial restraints by considering the conformation of template B is closest to the crystal conformation of Q, especially at the ligand-binding region

The crystal conformation of 1e9x (template B) that binds to a different substrate also adopts a helical conformation in this loop region. However, template A, which is in the substrate-free form, adopts a different unstructured conformation in this region (Fig. 5c).

Four different modeling schemes were adopted with and without spatial restraints (see legend to Fig. 5d for details). On an average, the four models generated using different approaches, record an RMSD of 1.44 Å around the substrate-binding region. When the coordinates of the substrate are superposed on the model using the coordinates as in 1jpz, of the various modeling schemes, the model derived from A as template together with spatial restraints obtained from template B gives rise to better binding to the substrate (model shown in *red* in Fig. 5d), whereas the model derived using only template A alone gives rise to very poor substrate binding (model shown in *yellow*; Fig. 5d). Interestingly, simple homology modeling using template B alone (complexed form; template structure shown in *red* in Fig. 5c; model shown in *green* in Fig. 5d) or template B together with spatial restraints from template A (model shown in *blue*) fail to give rise to accurate conformations at one of the substrate-binding loops.

In the case of modeling during distant relationships, rather than choosing multiple distant templates in the distance geometry approach or using only one distantly related template, the accuracy of the model can be improved by using one template and imposing spatial restraints as observed in the other structural members of that superfamily. The availability of a better starting model can lead to improved modes of interaction with small molecules by subsequent docking algorithms. Usually, docking of ligands and the design of inhibitors require rational modeling and rigorous molecular dynamics around important loops, as was applied earlier [18] for the model of P450arom, which catalyzes the conversion of steroids to esterogens. The problem of accurate modeling is especially relevant for cytochrome P450, where the mammalian proteins are membrane-bound, hard to crystallize and the bacterial structural homologues are distantly related for homology modeling.

We have developed a strategy for improving the quality of the models generated by comparative modeling through utilization of additional spatial restraints of structurally important regions of the templates. In this approach, the utilization of one template, together with the spatial orientation (distance and torsion angle) of motifs from the SMoS database [7] or any other conserved region, on known examples improves the quality of the resulting models. In this paper, we show that the utilization of the spatial restraints of more than one template, where there is a distant relationship with any of the structural homologues, results in more accurate three-dimensional model of protein sequences. This strategy can be employed, in general, to overcome the inherent limitation of comparative modeling methods when using multiple distantly related templates.

**Acknowledgements** RS is a recipient of a Senior Research Fellowship awarded by Wellcome Trust. Financial and infrastructural support from NCBS (TIFR) is also acknowledged.

## References

1. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
2. Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Sali A (2000) Nat Struct Biol 7(Suppl):986–990
3. Dunbrack RL, Gerloff DL, Bower M, Chen X, Lichtarge O, Cohen FE (1997) Fold Des 2:R27–R42
4. Mosimann S, Meleshko R, James MNG (1995) Proteins: Struct, Funct, Genet 23:301–317
5. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) J Mol Biol 247:536–540
6. Mallika V, Bhaduri A, Sowdhamini R (2002) Nucleic Acids Res 30:284–288
7. Chakrabarti S, Venkataramanan K, Sowdhamini R (2003) Nucleic Acids Reshttp://www3.oup.co.uk/nar/database/summary/275
8. Russell RB, Barton GJ (1992) Proteins: Struct, Funct, Genet 14:309–323
9. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Nucleic Acids Res 25:3389–4002
10. Kelley LA, MacCallum RM, Sternberg MJ (2000) J Mol Biol 299:499–520
11. Jones DT (1999) J Mol Biol 287:797–815
12. Eisenberg D, Luthy R, Bowie JU (1997) Methods Enzymol 277:396–404
13. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Proteins: Struct, Funct, Genet 12:345–364
14. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Cryst 26:283–291
15. Lovell SC, Davis IW, Arendall 3rd WB, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Proteins: Struct, Funct, Genet 50:437–450
16. Sowdhamini R, Burke DF, Huang JF, Mizuguchi K, Nagarajaram HA, Srinivasan N, Steward RE, Blundell TL (1998) Structure 6:1087–1094
17. Srinivasan N, Blundell TL (1993) Protein Eng 6:501–512
18. Graham-Lorence S, Amarneh B, White RE, Peterson JA, Simpson ER (1995) Protein Sci 4:1065–1080
19. Kraulis P (1991) J Appl Crystallogr 24:945–949
20. Sayle RA, Milner-White EJ (1995) Trends Biochem Sci 20:374–375