

## **From Lipid Transport to Oxygenation of Aromatic Compounds: Evolution within the Bet v1-like Superfamily**

<http://www.jbsdonline.com>

**Joydeep Chakraborty  
Tapan K. Dutta\***

Department of Microbiology, Bose  
Institute, P-1/12 C.I.T. Scheme VII M,  
Kolkata 700 054, India

### **Abstract**

In absence of significant sequence similarity, remote homology between proteins can be confused with analogy and in such a case, shared ancestry can be inferred in light of certain unique and common features. In the present study, to understand the evolutionary origin of catalytic domain of large subunit of ring-hydroxylating oxygenases (RHOs), belonging to the Bet v1-like superfamily, structure-based phylogenies have been derived from structural alignment of representative proteins of the superfamily. A careful inspection of the structural relatedness of RHOs with the rest of the families showed closest similarity between RHO catalytic domain and PA1206-like protein. In addition, phylogenetic relationship of the Rieske domain of the large subunit of RHOs with functionally and structurally similar proteins has also been elucidated so as to postulate the most possible events leading to the genesis of the large subunit of RHOs.

**Key words:** Structural evolution; Bet v1-like superfamily; Iron-sulfur protein; Rieske protein; Aromatic ring-hydroxylating oxygenase.

### **Introduction**

Bacterial ring-hydroxylating oxygenases (RHOs) belong to a large family of enzymes that have potential utility in bioremediation of diverse aromatic compounds including linked and fused aromatics, aliphatic olefins, and highly substituted aromatics as well as many toxic and/or carcinogenic compounds such as polychlorinated biphenyls and polycyclic aromatic hydrocarbons (PAHs) (1, 2). To cope with the enormous diversity of aromatic compounds in the environment, this enzyme family has evolved remarkably to accept wide range of substrates leading to expansion of catabolic potential. The terminal oxygenase of multi-component RHOs exist either as homo-multimers ( $\alpha_n$ ) or hetero-multimers ( $\alpha_n\beta_n$ ). The large catalytic subunit ( $\alpha$ ) bears an N-terminal Rieske domain with a conserved [2Fe-2S] center and a C-terminal catalytic domain having a conserved mononuclear iron-binding site (2-His-1-carboxylate motif).

Biological sequence analysis with important computational methods such as phylogenetic analysis and classification have been used to infer shared ancestry among proteins (3-17). In the absence of significant sequence similarity, other unique or common features such as protein structure evolution can be employed to unearth shared ancestry. The Structural Classification of Proteins (SCOP) database (18) classifies protein domains with known structures into a hierarchy of structural information where close homologues are grouped into the same family and families with convincing structural and/or functional evidence for common ancestry are grouped

\*Phone: +91-33-2569-3241  
Fax: +91-33-2355-3886  
E-mail: [tapan@bic.boseinst.ernet.in](mailto:tapan@bic.boseinst.ernet.in)

into the same superfamily. Bet v1-like superfamily [SCOP: 55961], named after the white birch (*Betula verrucosa*) major pollen allergen (a member of the PR-10 family of plant pathogenesis-related proteins) (19, 20), is one such superfamily in SCOP (release 1.75, June 2009) having a TATA binding protein (TBP)-like fold [SCOP: 55944] and includes 53 non-identical structures (including mutants, isoforms and those from structural genomic projects). These structures are classified into ten different families, viz., pathogenesis-related protein 10 (PR10)-like, steroidogenic acute regulatory protein-related lipid transfer (START) domain, phosphatidyl inositol transfer protein (PITP), RHO  $\alpha$ -subunit catalytic domain, activator of Hsp90 ATPase homolog 1 (AHSA1) domain, oligoketide cyclase/dehydrase (OK CYC/DH)-like, PA1206-like, Atu1531-like, Smu440-like, and CoxG-like families. Proteins having Bet v1-like fold have been found to be ubiquitously present in all the three domains of life and are mostly involved in lipid transport and/or metabolism and pathogenesis as well as other unique functions such as biosynthesis of secondary metabolites, degradation of aromatic compounds, and antibiotic resistance (21). Functionally well-characterized proteins within the Bet v1-like superfamily belong to PR10-like, START domain, PITP, RHO  $\alpha$ -subunit catalytic domain, OK CYC/DH-like, AHSA1 domain and CoxG-like families while rest of them have been classified only on the basis of structural genomic studies and lack information regarding their functional aspects. Structure of all the members within this superfamily is characterized by a distinct architecture (a common  $\beta$ - $\alpha_2$ - $\beta_6$ - $\alpha$  fold), here referred to as the “Bet v1-like fold” (Figure 1), whereby an anti-parallel  $\beta$ -sheet is wrapped around a long  $\alpha$ -helix forming a large hydrophobic cavity (22). Certain bacterial families are distributed only among a small number of species which suggests that these proteins evolved from a primordial Bet v1-like gene by adopting a specialized function as exemplified by the antibiotic resistance protein (CalC), PA1206 protein and the RHO  $\alpha$ -subunit catalytic domain.

The protein family database, Pfam (release 24, October 2009), categorizes  $\alpha_n\beta_n$  [Pfam: PF00848] and  $\alpha_n$  [Pfam: PF11723] RHOs into two different families, having 2022 and 34 sequences respectively within the Bet v1-like clan [Pfam: CL0209] thus, making RHOs the largest family [SCOP: 55969] within the Bet v1-like superfamily. The Rieske domain of RHO  $\alpha$ -subunit (23), on the other hand, belongs

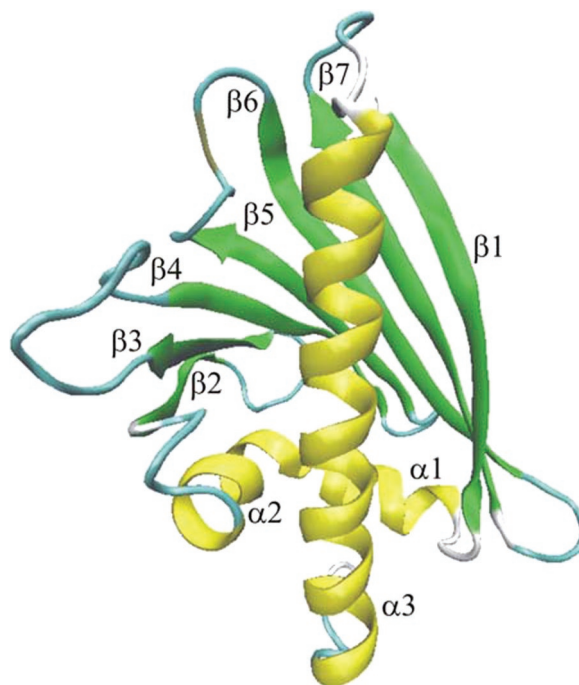


**Figure 1:** Structure of the pollen allergen, Bet v1, from *Betula pendula* (PDB ID: 1BV1) showing the conserved  $\beta$ - $\alpha_2$ - $\beta_6$ - $\alpha$  fold of Bet v1-like superfamily proteins.

to the iron-sulfur protein (ISP) domain superfamily [SCOP: 50022], sharing the same ISP domain fold [SCOP: 50021] with two other families, viz., Rieske ISP and NirD-like. The superfamily includes structures of 22 non-identical proteins from bacteria, archaea and eukaryotes. Proteins bearing Rieske centers are widespread in nature and exhibit different catalytic functions, their occurrence being more common in bacteria and archaea when compared to eukaryotes (24).

Till date, several genes encoding RHOs have been characterized in different species of bacteria, which give useful information regarding diversification of these catabolic genes. But the evolutionary events regarding origin of the terminal oxygenase component of the RHO systems are yet to be revealed. The present study attempts in depicting the possible evolutionary scenarios leading to the origin of the RHO  $\alpha$ -subunit, in light of structural phylogenies and domain relatedness among the homologues of both Rieske and the catalytic domain.

## ***Methods***

### *Retrieval of Structural Homologues and Construction of Dataset*

Structures of Bet v1 homologues as well as those within ISP domain superfamily, classified in SCOP (release 1.75), were obtained from Brookhaven Protein Data Bank (PDB) (25). Bet v1-like families exclusively belonging to eukaryotes (viz. PR10-like and P1TP families) were discarded from further analysis. PDB coordinates for the Rieske and catalytic domains of RHOs were either obtained from ASTRAL database (version 1.75) (26) or extracted separately from the parent PDB file. Moreover, the Vector Alignment Search Tool (VAST) utility (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html>) (27) was used to determine structural homologues of RHOs which were not included in SCOP release 1.75. Searches were made against the minimum-redundancy subset of PDB with structures of both naphthalene 1,2-dioxygenase (PDB ID:1NDO) and 2-oxo-1,2-dihydroquinoline 8-monooxygenase (PDB ID:1Z01) from *Pseudomonas putida* as queries.

### *Phylogenetic Reconstruction and Structural Analyses*

Multiple structural alignments of the Bet v1 homologues and the ISP domain homologues were performed separately using STAMP (28). The initial STAMP alignments were executed using default parameters (-slide 5 and -scanscore 6) and the final alignments were performed with parameters: -slide 7 and -scanscore 4. Pairwise distance matrices based on both the structural distance measures,  $1-Q_H$  and root mean square deviation (RMSD) were then computed using the MultiSeq module in Visual Molecular Dynamics (VMD) program (29, 30), distributed by the NIH Resource for Macromolecular Modeling and Bioinformatics at the University of Illinois (<http://www.ks.uiuc.edu/Research/vmd/>). In structural alignment of proteins, the structurally conserved regions are separated by non-conserved ones, which mostly occur via insertion or deletion. Such regions have been shown to be crucial for inferring the phylogenetic history of a protein family (31). The factor  $Q_H$  accounts for contributions from both structurally aligned regions as well as the structural deviations induced by insertions in each protein in an aligned pair (32). Another set of distance matrices were constructed on the basis of normalized RMSD (superposition RMSD divided by the number of aligned residues) obtained separately from pairwise structural alignments of the 36 Bet v1 homologues as well as the 27 ISP homologues, used in this study, employing the DALI algorithm (33). The distance matrices, thus obtained, were finally used to generate structure-based phylogenetic trees using the program FITCH (34) within the PHYLIP package (35). All the trees were visualized and manipulated using the program Tree Explorer v2.12 (36) and the layouts were modified using a standard graphics program. In addition, the pairwise structural alignments performed with DALI were

used to compute structural distance between certain representative family members of the Bet v1-like superfamily by a measure, called structural distance metric (SDM), which accounts for insertions and deletions (indels) as well as RMSD at the topologically equivalent positions of a pair of aligned structures (37, 38).

#### *Identification of Conserved Domains*

The Conserved Domain Database (CDD), the current version of which is available at <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> (39), was used for identification and comparison of conserved domains present in the structural homologues of both Bet v1-like and ISP domain superfamilies.

#### **Results**

##### *Structural Neighbors of RHO $\alpha$ -subunit Catalytic Domain and Their Taxonomic Distribution*

Since no homologues of bacterial RHOs could be detected in terms of sequence similarity, the next attempt was to study the structural homologues of individual domains of such proteins. As mentioned above, the catalytic domain of bacterial aromatic ring-hydroxylating oxygenase (RHO)  $\alpha$ -subunit belongs to the Bet v1-like superfamily [SCOP: 55969]. Although proteins having the same fold do not always share a common ancestry (40-43), but it seems very unlikely that structures with a complex architecture, as observed within the Bet v1-like superfamily (Figure 1), have evolved independently. Thus to understand their evolutionary relationship and to identify the closest relative of RHO  $\alpha$ -subunit catalytic domain, structural phylogeny of 36 representatives (excluding families belonging exclusively to eukaryotes) from the Bet v1-like superfamily (enlisted in Table I) was computed on the basis of both  $Q_H$  factor (32) as well as normalized RMSD. It may be mentioned here that in addition to contributions from structurally aligned regions (similar to RMSD measures), the factor  $Q_H$  also considers structural deviations induced by insertions in each protein in an aligned pair and thus, ideally reflects evolutionary relationships among structurally similar entities. The phylogenetic tree obtained using the distance measure,  $1-Q_H$ , revealed that the catalytic domain of the steroid-binding  $\alpha_n$  RHO, 3-ketosteroid 9 $\alpha$ -hydroxylase (PDB: 2ZYL) from *Mycobacterium tuberculosis*, and a hypothetical protein, PA1206 (PDB: 2FFS) from *Pseudomonas aeruginosa* strain PAO1 having a conserved domain of unknown function (CD: DUF1857), cluster together and share a common node with other RHOs, further rooted by the self-sacrificing resistance protein, CalC, from *Micromonospora echinospora*; (PDB: 1ZXF) (Figure 2A). The phylogenetic tree computed on the basis of normalized RMSD (Figure 2B) also showed similar results.

The structure-based alignment of both the families (CalC and PA1206) with RHO catalytic domain showed conservation in their secondary structural elements that make up the core Bet v1-like fold except that the CalC protein differs marginally from the core fold by disruption of the C-terminal helix ( $\alpha_2$ ) by a non-helical region, as shown in Figure 3. Moreover, among the two families, PA1206 protein was found to be structurally more close to the  $\alpha$ -subunit catalytic domain of RHOs (especially those of  $\alpha_n$  RHOs), as observed from their respective structural superposition (Figures 4A and 4B) as well as from comparison of their pairwise SDM score (Figure 4C). Further, a BLAST search of the CalC protein against the non-redundant database at the National Center for Biotechnology Information (NCBI, NIH, Bethesda, MD) (<http://www.ncbi.nlm.nih.gov/>) yielded maximum identity with a predicted protein from the closely-related actinomycete, *Salinispora arenicola*, followed by certain uncharacterized bacterial members of the AHSA1 family. Another search was performed to understand the taxonomical distribution of PA1206-like family, especially among prokaryotes, which identified several putative proteins in bacteria (but not in archaea) bearing the PA1206-like domain signature. Most of

these proteins are distributed among the phyla Proteobacteria (especially in strains belonging to genus *Burkholderia*, *Pseudomonas*, *Dechloromonas*, *Aromatoleum*, *Thauera*, *Azoarcus*, *Cupriavidus*, *Ralstonia*, *Janthinobacterium*, *Oxalobacter*, *Limnibacter*, *Methylobium*, *Agrobacterium*, *Geobacter* and *Caulobacter*) with maximum abundance among Burkholderiales and Pseudomonads. Surprisingly,

**Table I**  
Representative structures from Bet v1-like superfamily.

| Family/Protein (length)                                                      | Organism                                         | Function/Substrate                           | Conserved domain family (CDD accession no.) |
|------------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------------|---------------------------------------------|
| Steroidogenic acute regulatory protein-related lipid transfer (START) domain |                                                  |                                              |                                             |
| 1EM2 (229)                                                                   | <i>Homo sapiens</i>                              | Lipid transport/Cholesterol                  | START_STARTD3-like (cd08906)                |
| 1JSS (224)                                                                   | <i>Mus musculus</i>                              | Lipid transport/Cholesterol                  | START_STARTD4-like (cd08902)                |
| 1LN1 (214)                                                                   | <i>Homo sapiens</i>                              | Lipid binding/Phosphatidylcholine            | START_STARTD2-like (cd08910)                |
| 2PSO (237)                                                                   | <i>Homo sapiens</i>                              | -uf-                                         | START_STARTD13-like (cd08909)               |
| 2R55 (231)                                                                   | <i>Homo sapiens</i> <sup>a</sup>                 | Lipid transport                              | START_STARTD5-like (cd08903)                |
| Ring-hydroxylating oxygenase (RHO) $\alpha$ -subunit catalytic domain        |                                                  |                                              |                                             |
| 1NDO (293)                                                                   | <i>Pseudomonas putida</i>                        | Dioxygenase/Naphthalene                      | RHO_alpha_C_NDO-like (cd08881)              |
| 2B1X (279)                                                                   | <i>Rhodococcus</i> sp. NCIMB12038                | Dioxygenase/Naphthalene                      | RHO_alpha_C_NDO-like (cd08881)              |
| 1ULI (270)                                                                   | <i>Rhodococcus</i> sp. RHA-1                     | Dioxygenase/Biphenyl                         | RHO_alpha_C_NDO-like (cd08881)              |
| 1WQL (279)                                                                   | <i>Pseudomonas fluorescens</i> IP01              | Dioxygenase/Cumene                           | RHO_alpha_C_NDO-like (cd08881)              |
| 2BMO (287)                                                                   | <i>Comamonas</i> sp.                             | Dioxygenase/Nitrobenzene                     | RHO_alpha_C_NDO-like (cd08881)              |
| 2CKF (294)                                                                   | <i>Sphingomonas</i> sp. CHY-1 <sup>a</sup>       | Dioxygenase/PAH                              | RHO_alpha_C_NDO-like (cd08881)              |
| 3EN1 (261)                                                                   | <i>Pseudomonas putida</i> <sup>a</sup>           | Dioxygenase/Toluene                          | RHO_alpha_C_NDO-like (cd08881)              |
| 1WW9 (242)                                                                   | <i>Janthinobacterium</i> sp. J3                  | Dioxygenase/Carbazole                        | RHO_alpha_C_DMO-like (cd08878)              |
| 3GOB (218)                                                                   | <i>Stenotrophomonas maltophilia</i> <sup>a</sup> | Demethylase/Dichloromethoxybenzoic acid      | RHO_alpha_C_DMO-like (cd08878)              |
| 2ZYL (264)                                                                   | <i>Mycobacterium tuberculosis</i> <sup>a</sup>   | Hydroxylase/3-Ketosteroid                    | Belongs to HcaE domain family               |
| 1Z01 (279)                                                                   | <i>Pseudomonas putida</i>                        | Monooxygenase/2-Oxoquinoline                 | RHO_alpha_C_DMO-like (cd08878)              |
| Activator of Hsp90 ATPase homolog 1 (AHS1) domain                            |                                                  |                                              |                                             |
| 1X53 (145)                                                                   | <i>Homo sapiens</i>                              | -uf-                                         | SRPBCC_Aha1 (cd08892)                       |
| 1XFS (178)                                                                   | <i>Nitrosomonas europaea</i> ATCC 19718          | -uf-                                         | SRPBCC_CalC_Aha1-like (cd08896)             |
| 1XN5 (146)                                                                   | <i>Bacillus halodurans</i> C-125                 | -uf-                                         | SRPBCC_CalC_Aha1-like (cd07814)             |
| 1XN6 (143)                                                                   | <i>Bacillus cereus</i>                           | -uf-                                         | SRPBCC_CalC_Aha1-like (cd07814)             |
| 1XUV (178)                                                                   | <i>Methanosarcina mazei</i> Go1                  | -uf-                                         | SRPBCC_CalC_Aha1-like_9 (cd07826)           |
| 1Z94 (147)                                                                   | <i>Chromobacterium violaceum</i> ATCC 12472      | -uf-                                         | SRPBCC_CalC_Aha1-like_2 (cd08895)           |
| 2IL5 (171)                                                                   | <i>Staphylococcus aureus</i>                     | -uf-                                         | SRPBCC_CalC_Aha1-like (cd07814)             |
| 2K5G (191)                                                                   | <i>Bordetella parapertussis</i>                  | -uf-                                         | SRPBCC_CalC_Aha1-like_6 (cd08899)           |
| 2NN5 (184)                                                                   | <i>Enterococcus faecalis</i>                     | -uf-                                         | SRPBCC_CalC_Aha1-like_6 (cd08899)           |
| 3ELI (152)                                                                   | <i>Ruegeria pomeroyi</i>                         | -uf-                                         | SRPBCC_CalC_Aha1-like (cd07814)             |
| 1ZXF (155)                                                                   | <i>Micromonospora echinospora</i>                | Self-sacrificing resistance protein/Enediyne | SRPBCC_CalC (cd08891)                       |
| Oligoketide cyclase/dehydrase (OK CYC/DH)-like                               |                                                  |                                              |                                             |
| 1T17 (148)                                                                   | <i>Caulobacter crescentus</i> CB15               | -uf-                                         | COQ10P_like (cd07813)                       |
| 2D4R (147)                                                                   | <i>Thermus thermophilus</i>                      | -uf-                                         | Belongs to SRPBCC domain family             |
| 2RER (173)                                                                   | <i>Streptomyces glaucescens</i>                  | Aromatase-cyclase/Tetracenomyacin            | TcmN_ARO_CYC_like (cd08860)                 |
| PA1206-like                                                                  |                                                  |                                              |                                             |
| 2FFS (157)                                                                   | <i>Pseudomonas aeruginosa</i> PAO1               | -uf-                                         | SRPBCC_DUF1857 (cd08863)                    |
| Atu1531-like                                                                 |                                                  |                                              |                                             |
| 2QPV (156)                                                                   | <i>Agrobacterium tumefaciens</i> C58             | -uf-                                         | Belongs to SRPBCC domain family             |
| 3CNW (148)                                                                   | <i>Bacillus cereus</i> ATCC 14579                | -uf-                                         | PYR_PYL_RCAR_like (cd07821)                 |
| Smu440-like                                                                  |                                                  |                                              |                                             |
| 3IJT (155)                                                                   | <i>Streptococcus mutans</i> <sup>a</sup>         | -uf-                                         | SRPBCC_Smu440_like (cd08862)                |
| CoxG-like                                                                    |                                                  |                                              |                                             |
| 2NS9 (157)                                                                   | <i>Equus caballus</i>                            | Methemoglobin                                | CoxG (cd05018)                              |
| 2PCS (162)                                                                   | <i>Geobacillus kaustophilus</i>                  | -uf-                                         | CoxG (cd05018)                              |

uf: Proteins with unknown function, derived from structural genomics.

<sup>a</sup>Structures not included in SCOP release 1.75.

homologues were identified even in organisms known to possess RHOs, such as *Pseudomonas aeruginosa* strain PAO1 and *Burkholderia xenovorans* strain LB400 which bear toluate and biphenyl dioxygenases respectively.

#### Structural Neighbors of RHO $\alpha$ -subunit ISP Domain

The Rieske domain of RHO  $\alpha$ -subunit (23), as mentioned earlier, belongs to the iron-sulfur protein (ISP) domain superfamily [SCOP: 50022], sharing the same ISP domain fold [SCOP: 50021] with two other families, viz., Rieske ISP and NirD-like. The superfamily includes structures of 22 non-identical proteins from bacteria, archaea and eukaryotes as per SCOP (release 1.75). Figure 2C shows the phylogenetic relationship of 27 structures (enlisted in Table II) representing each family within the ISP domain superfamily, aligned using STAMP (28) and the phylogeny computed on the basis of  $Q_H$  factor (32), while Figure 2D shows their phylogenetic relationship computed on the basis of normalized RMSD. As shown clearly in both the dendrograms, RHO Rieske (ISP) domain, especially those of  $\alpha_n$  type RHOs, cluster with Rieske ferredoxins, the electron transport proteins associated with various RHOs. Such a relatedness is also seen in sequence-based phylogeny obtained using RHO Rieske (ISP) domain and Rieske ferredoxins (data not shown).

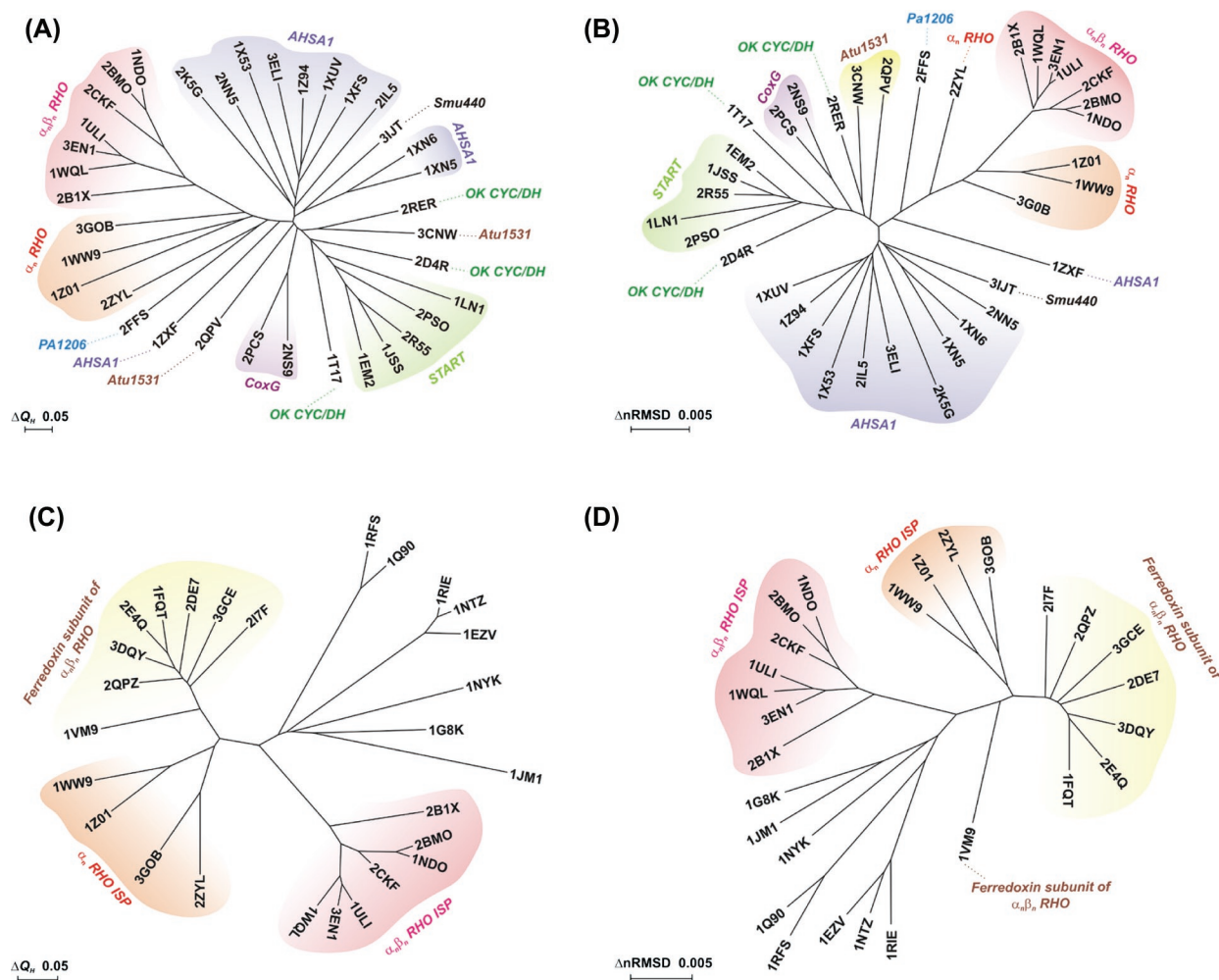


**Figure 2:** Structure-based phylogenetic tree of Bet v1-like superfamily proteins computed on the basis of distance measures,  $1-Q_H$  (A) and normalized RMSD (B), obtained from the alignment of 36 representative structures using the rigid-body alignment methods, STAMP and DALI respectively and that of Iron-Sulfur Protein (ISP) domain superfamily proteins drawn according to the distance measures,  $1-Q_H$  (C) and normalized RMSD (D), obtained from the alignment of 27 representative structures using STAMP and DALI respectively. All the dendrograms were obtained using Fitch method.

Moreover, the phylogenetic trees represented in Figure 2 also point toward the fact that among RHOs, the  $\alpha_n$  RHOs might have evolved earlier and have acted as a link between their ancient precursor and the present day ‘extant’ dioxygenases (both  $\alpha_n$  and  $\alpha_n\beta_n$  types).

## Evolution within the Bet v1-like Superfamily

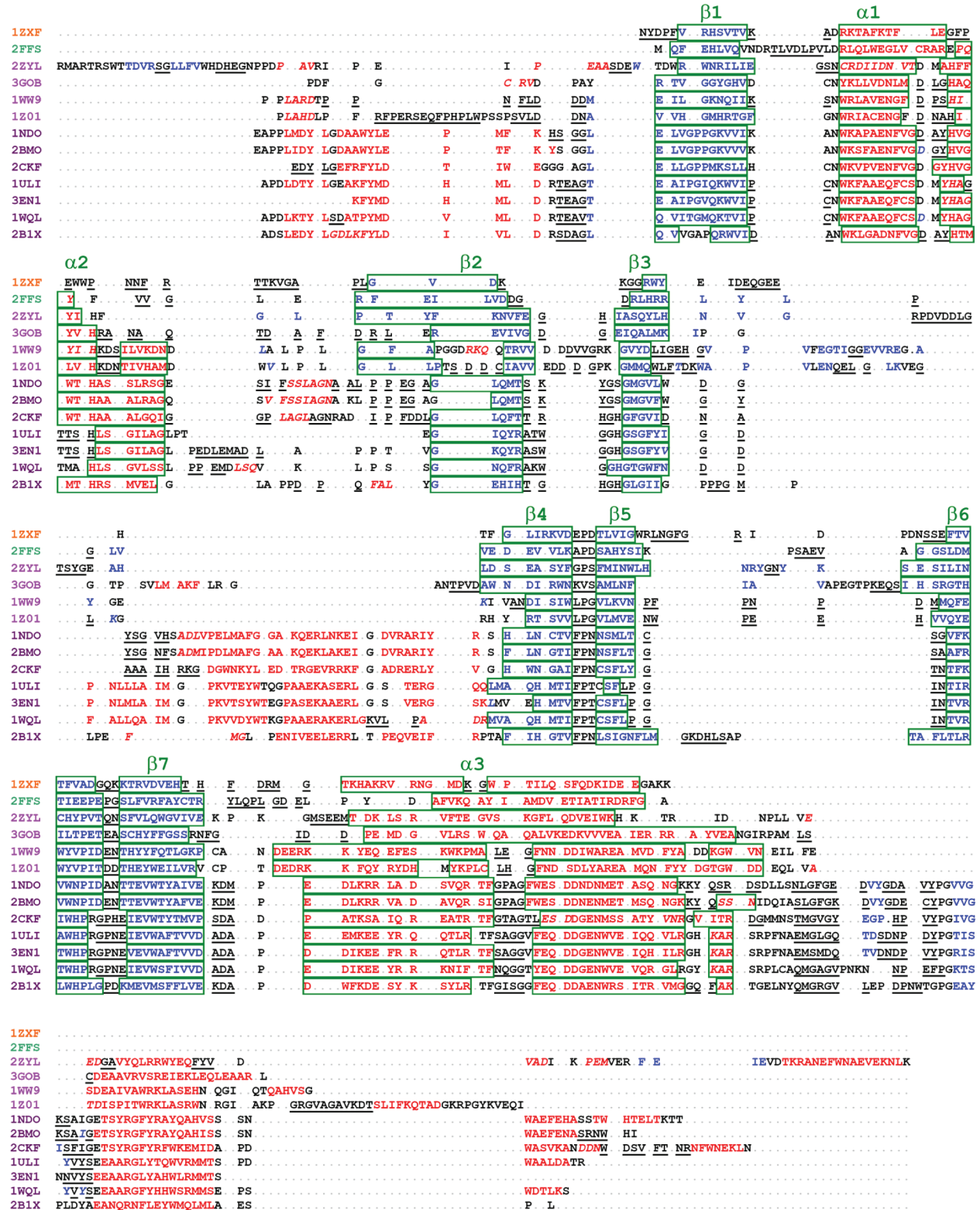


**Figure 3:** Structure-based alignment of PA1206 and CalC proteins with those of RHO  $\alpha$ -subunit catalytic domain family members. Regions that make up the core Bet v1-like fold are represented in green boxes. Each residue is color coded and/or italicized as per its secondary structure information as follows - bold regular: loop; bold red:  $\alpha$ -helix; bold red italics:  $3_{10}$ -helix; bold blue: extended  $\beta$ -strand; bold blue italics:  $\beta$ -bridge; bold underlined: turn. Gaps are denoted by dots.

The conserved domain present in each of the structures included in this study (36 representatives from the Bet v1-like superfamily and 27 from the ISP domain superfamily) were identified from the CDD (39), which are shown in Tables I and II. In CDD, the Bet v1-like superfamily is represented by the SRPBCC (START/RHO\_alpha\_C/PITP/Bet\_v1/CoxG/CalC) domain superfamily [CDD: cd07812] with the most abundant domain being that of the SRPBCC\_CalC\_Aha1-like family [CDD: cd07814] which is found among the AHSA1 family of proteins (Table I). In the CD sub-family hierarchy also, the PA1206-like domain, represented by SRPBCC\_DUF1857 domain family [CDD: cd08863] is placed in between the RHO  $\alpha$ -subunit catalytic domain, represented by the conserved RHO\_alpha\_C domain [CDD: cd00680] and the domains belonging to rest of the Bet v1-like superfamily members considered in this study.

The ISP domain superfamily, on the other hand, is represented by Rieske domain superfamily [CDD: cd03467] in CDD. The RHO  $\alpha$ -subunit ISP domain makes up the largest family (Rieske\_RO\_alpha\_N; CDD: cd03469) within this CD superfamily (Table II). However, the relatedness between RHO ISP domain of  $\alpha_n$  type RHOs and Rieske ferredoxins as observed in our structural phylogenetic studies (Figure 2C and 2D) is not distinct in the CD sub-family hierarchy of Rieske domain superfamily.

### Discussion

It is now a well known fact that three-dimensional structures of proteins are evolutionarily much more conserved, while the sequence identity/similarity between two

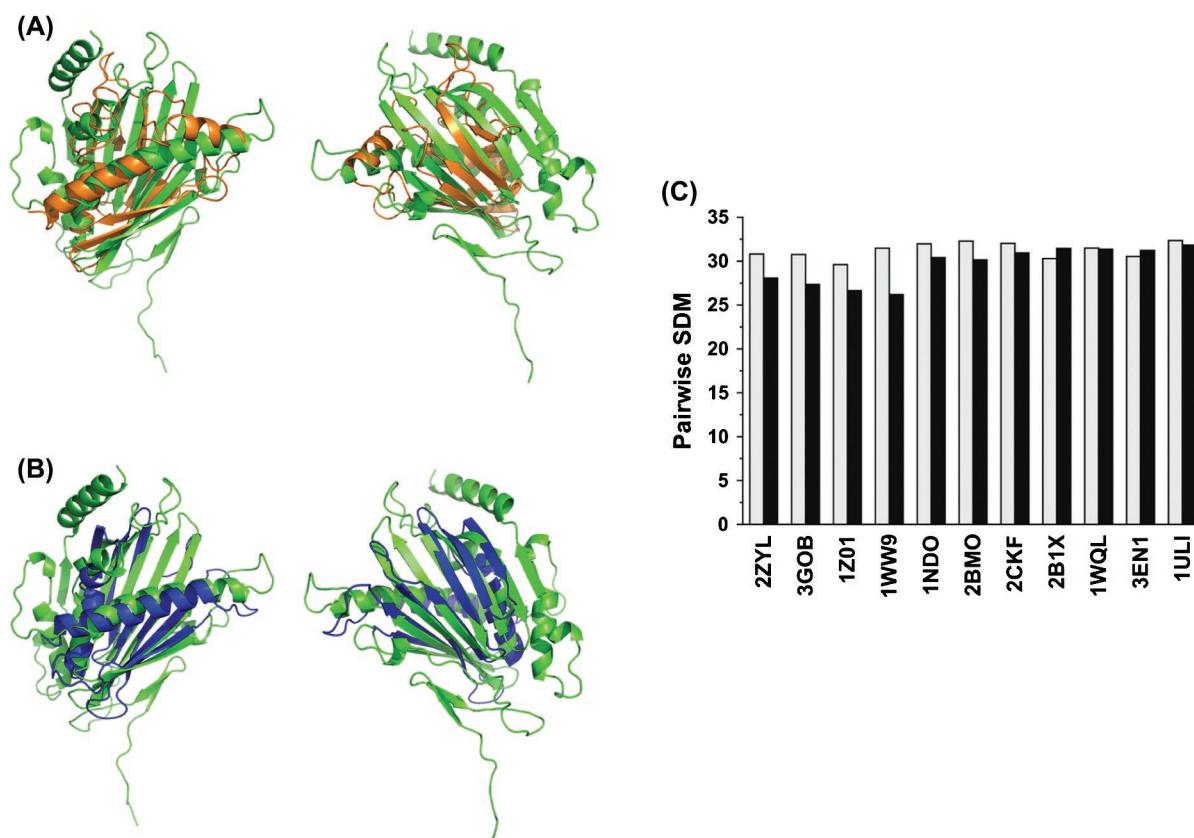


**Figure 4:** Structural superposition of CalC [PDB: 1ZXF] (A) and PA1206 [PDB: 2FFS] (B) proteins with the catalytic domain of 3-ketosteroid 9 $\alpha$ -hydroxylase [PDB: 2ZYL]. Comparison of structural relatedness of both CalC (gray bars) and PA1206 (black bars) proteins with the  $\alpha$ -subunit catalytic domain of RHOs (represented by respective PDB IDs) (C) in terms of pairwise SDM score.

distant homologues may go beyond recognition (44-47). Distant homologues are often accompanied by substantial changes in functional properties, such as altered substrate preferences or even loss of enzymatic activity (48, 49). In absence of significant sequence similarity, remote homology can be confused with analogy and in such a case, shared ancestry can only be inferred in light of several observations, such as extent of overall structural similarity, arrangement of secondary structural elements, unusual structural features, arrangement of functional residues, similarity in solvent accessibility and similar domain architectures (41, 50-52). Stability of the overall spatial conformation of proteins may be attributed to conservation of physico-chemical properties of residues crucial for maintaining the backbone fold (53). One such property is the presence of closed loops which are believed to be the basic structural element of globular proteins (54, 55). Protein structures have been shown to maintain a 'proteomic code' forming prototypes of functional closed loops, called "Elementary Functional Loops" (EFLs) (55-57). It has been suggested that loop fold structure of globular proteins have strong evolutionary implications and that detection of EFLs can be useful in elucidating distant evolutionary relations among protein folds (58, 59). Identification of EFLs in Bet v1-like fold shall be a useful strategy to understand the evolution of such a complex fold, characterized by a large hydrophobic cavity. However, the members within Bet v1-like superfamily have diverged to such an extent, probably due to varying selection pressure, that there is no significant sequence similarity among the families. At the same time, it is also difficult to predict the ancient member of this superfamily although the OK CYC/DH-like family is known to possess the minimal Bet v1-like fold and is thus considered to be the closest relative of the primordial Bet v1-like protein (21).

**Table II**  
Representative structures from ISP domain superfamily.

| Protein (length)                                                                 | Organism                                         | Function/Associated with                   | Conserved domain family (CDD accession no.) |
|----------------------------------------------------------------------------------|--------------------------------------------------|--------------------------------------------|---------------------------------------------|
| <b>Rieske iron-sulfur protein (ISP)</b>                                          |                                                  |                                            |                                             |
| 1G8K (133)                                                                       | <i>Alcaligenes faecalis</i>                      | Arsenite oxidase                           | Rieske_ArOX_sml (cd03476)                   |
| 1RIE (129)                                                                       | <i>Bos taurus</i>                                | Mitochondrial Cyt bc1-complex              | Rieske_cytochrome_bc1 (cd03470)             |
| 1NTZ (124)                                                                       | <i>Bos taurus</i>                                | Mitochondrial Cyt bc1-complex              | Rieske_cytochrome_bc1 (cd03470)             |
| 1EZV (122)                                                                       | <i>Saccharomyces cerevisiae</i>                  | Cyt bc1-complex                            | Rieske_cytochrome_bc1 (cd03470)             |
| 1RFS (139)                                                                       | <i>Spinacia oleracea</i>                         | Chloroplast Cyt bf-complex                 | Rieske_cytochrome_b6f (cd03471)             |
| 1Q90 (127)                                                                       | <i>Chlamydomonas reinhardtii</i>                 | Cyt b6f-complex                            | Rieske_cytochrome_b6f (cd03471)             |
| 1JM1 (204)                                                                       | <i>Sulfolobus acidocaldarius</i>                 | Sulfur oxidase                             | Rieske_SoxF_SoxL (cd03475)                  |
| 1NYK (165)                                                                       | <i>Thermus thermophilus</i>                      | Electron transport                         | Rieske (cd03467)                            |
| 1VM9 (111)                                                                       | <i>Pseudomonas mendocina</i> Kr1                 | Electron transport/Toluene monooxygenase   | Rieske_T4moC (cd03474)                      |
| 1FQT (112)                                                                       | <i>Burkholderia cepacia</i>                      | Electron transport/Biphenyl dioxygenase    | Rieske_RO_ferredoxin (cd03528)              |
| 2I7F (108)                                                                       | <i>Sphingobium yanoikuyae</i> <sup>a</sup>       | Electron transport/Biphenyl dioxygenase    | Rieske_RO_ferredoxin (cd03528)              |
| 2E4Q (109)                                                                       | <i>Pseudomonas sp.</i> <sup>a</sup>              | Electron transport/Biphenyl dioxygenase    | Rieske_RO_ferredoxin (cd03528)              |
| 2QPZ (103)                                                                       | <i>Pseudomonas putida</i> <sup>a</sup>           | Electron transport/Naphthalene dioxygenase | Rieske_RO_ferredoxin (cd03528)              |
| 2DE7 (115)                                                                       | <i>Pseudomonas resinovorans</i> <sup>a</sup>     | Electron transport/Carbazole dioxygenase   | Rieske_RO_Alpha_OMO_CARDO (cd03548)         |
| 3GCE (121)                                                                       | <i>Nocardioide aromaticivorans</i> <sup>a</sup>  | Electron transport/Carbazole dioxygenase   | Rieske_RO_ferredoxin (cd03528)              |
| 3DQY (106)                                                                       | <i>Pseudomonas putida</i> <sup>a</sup>           | Electron transport/Toluene dioxygenase     | Rieske_RO_ferredoxin (cd03528)              |
| <b>Ring Hydroxylating Oxygenase (RHO) <math>\alpha</math>-subunit ISP domain</b> |                                                  |                                            |                                             |
| 1NDO (154)                                                                       | <i>Pseudomonas putida</i>                        | Naphthalene dioxygenase                    | Member of Rieske domain family              |
| 2B1X (162)                                                                       | <i>Rhodococcus sp.</i> NCIMB12038                | Naphthalene dioxygenase                    | Rieske_RO_Alpha_NDO (cd03535)               |
| 1ULI (154)                                                                       | <i>Rhodococcus sp.</i> RHA-1                     | Biphenyl dioxygenase                       | Rieske_RO_Alpha_BPDO_like (cd03472)         |
| 1WQL (157)                                                                       | <i>Pseudomonas fluorescens</i> IP01              | Cumene dioxygenase                         | Rieske_RO_Alpha_BPDO_like (cd03472)         |
| 2BMO (150)                                                                       | <i>Comamonas sp.</i>                             | Nitrobenzene dioxygenase                   | Member of Rieske domain family              |
| 2CKF (127)                                                                       | <i>Sphingomonas sp.</i> CHY-1 <sup>a</sup>       | PAH dioxygenase                            | Member of Rieske domain family              |
| 3EN1 (127)                                                                       | <i>Pseudomonas putida</i> <sup>a</sup>           | Toluene dioxygenase                        | Member of Rieske domain family              |
| 1WW9 (142)                                                                       | <i>Janthinobacterium sp.</i> J3                  | Carbazole dioxygenase                      | Rieske_RO_Alpha_OMO_CARDO (cd03548)         |
| 3GOB (122)                                                                       | <i>Stenotrophomonas maltophilia</i> <sup>a</sup> | Dichloromethoxybenzoic acid demethylase    | Rieske_RO_Alpha_VanA_DdmC (cd03532)         |
| 2ZYL (73)                                                                        | <i>Mycobacterium tuberculosis</i> <sup>a</sup>   | 3-Ketosteroid hydroxylase                  | Rieske_RO_Alpha_KSH (cd03531)               |
| 1Z01 (148)                                                                       | <i>Pseudomonas putida</i>                        | 2-Oxoquinoline monooxygenase               | Rieske_RO_Alpha_OMO_CARDO (cd03548)         |

uf: Proteins with unknown function, derived from structural genomics.

<sup>a</sup>Structures not included in SCOP release 1.75.

Several members of the Bet v1-like superfamily are known to be involved in lipid-binding and/or pathogenesis (21). Catalytic domain of the steroid-binding  $\alpha_n$  RHO, 3-ketosteroid 9 $\alpha$ -hydroxylase (PDB ID: 2ZYL), which is believed to represent the archetypical RHO (60), also acts in cholesterol catabolic pathway of *Mycobacterium tuberculosis* and has been implicated in pathogenesis of the bacterium (61, 62). The catalytic domain of 3-ketosteroid 9 $\alpha$ -hydroxylase may thus be considered most close to the hypothetical pre-oxygenase module having a Bet v1-like fold. Our analysis also shows that among the two proteins, CalC and PA1206, the latter is structurally more close to the RHO  $\alpha$ -subunit catalytic domain family (Figures 2A, 2B and 4) and is distributed mostly among proteobacters, which suggests that either the protein itself, or any of its ancestors, might have been the precursor of the pre-oxygenase module. CalC protein, on the other hand, belongs to AHSA1 family, which makes it (or any of its ancestors) a possible link between the AHSA1 family and the RHO  $\alpha$ -subunit catalytic domain family. A similar inference can also be drawn from the CD sub-family hierarchy of these proteins. However, such inferences are truly putative at this moment as they are solely dependent upon current availability of structural data, e.g., the PDB contains just one known structure each of both CalC protein as well as the PA1206-like family.

Standing at this point, the above lines of analysis can be summed up to sketch the most likely events leading to the evolution of the terminal oxygenase component of RHOs, shown in Figure 5. An ancestral proteobacterial PA1206-like family protein, by spontaneous evolution, attained the Fe<sup>+3</sup>-binding 2-His-1-carboxylate motif required for oxygen activation and further gave rise to a pre-oxygenase module bearing a lipid-binding large hydrophobic cavity as well as a site of oxygen activation. Such a module, upon fusion with an ancestral Rieske ferredoxin-like protein, gave birth to the first catalytic RHO, resembling the present day 3-ketosteroid 9 $\alpha$ -hydroxylase, which further evolved to other  $\alpha_n$  RHOs. These RHOs further took the support of an additional protein (the small  $\beta$  subunit), known to be a distant homologue of nuclear transport factor 2 (NTF2) (63), and evolved into heteromultimeric ( $\alpha_n\beta_n$ ) RHOs, probably for structural stabilization and at the same time, for increasing their catabolic plasticity to accommodate a large spectrum of aromatic compounds.

### Conclusion

Deriving evolutionary scenarios on the basis of existing evidences has always been an information-dependent process. The present study also is not an exception and

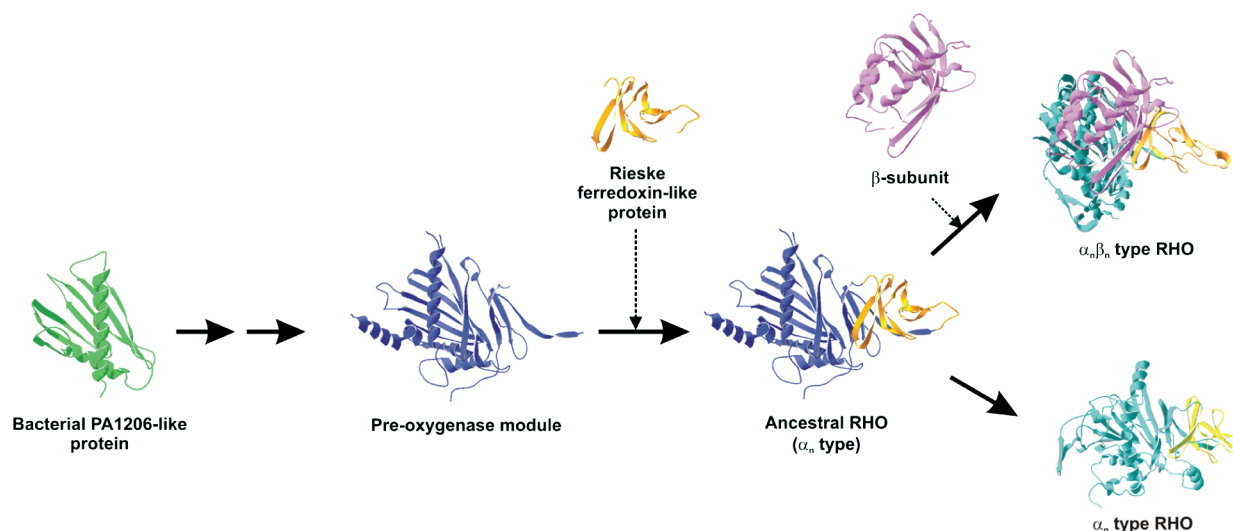


**Figure 5:** Proposed pathway of evolutionary events leading to the genesis of the  $\alpha$ -subunit of aromatic ring-hydroxylating oxygenases (RHOs) and subsequent multimerization with  $\beta$ -subunit. Solid arrows indicate multiple evolutionary steps in between two consecutive structures and dotted arrows indicate fusion or multimerization events.

truly relies on the information content of various biological databases. On the basis of both structural and functional evidences, we have depicted for the first time, the most possible event that has caused the genesis of the catalytic subunit of bacterial aromatic RHOs involved in the degradation as well as in the management of toxicity and/or carcinogenicity of aromatic hydrocarbons in the environment. The Bet v1-like fold, mostly associated with lipid transport and/or pathogenesis, has thus evolved for another crucial job, *i.e.*, oxygenation of aromatic nucleus thus, relieving the environment from various toxic and hazardous chemicals.

## Acknowledgements

The authors acknowledge Bose Institute, Kolkata (India) for financial support.

## References

1. S. M. Resnick, K. Lee, and D. T. Gibson. *J Ind Microbiol* 17, 438-457 (1996).
2. R-H. Peng, A-S. Xiong, Y. Xue, X-Y. Fu, F. Gao, W. Zhao, Y-S. Tian, and Q-H. Yao. *FEMS Microbiol. Rev* 32, 927-955 (2008).
3. F. B. Guo and Y. Lin. *J Biomol Struct Dyn* 26, 413-420 (2009).
4. I. H. Cho, Z. R. Lu, J. R. Yu, Y. D. Park, J. M. Yang, M. J. Hahn, and F. Zou. *J Biomol Struct Dyn* 27, 331-345 (2009).
5. V. Sabbia, H. Romero, H. Musto, and H. Naya. *J Biomol Struct Dyn* 27, 361-369 (2009).
6. J. N. Zhao, Y. Wang, B. S. Zhao, and L. L. Chen. *J Biomol Struct Dyn* 27, 371-380 (2009).
7. P. Mukhopadhyay and T. C. Ghosh. *J Biomol Struct Dyn* 27, 477-488 (2010).
8. P. Anbazhagan, M. Purushottam, H. B. K. Kumar, O. Mukherjee, S. Jain, and R. Sowdhamini. *J Biomol Struct Dyn* 27, 581-598 (2010).
9. P. Putta and C. K. Mitra. *J Biomol Struct Dyn* 27, 599-610 (2010).
10. Z. Liu, Y. Xu, L. Wu, and S. Zhang. *J Biomol Struct Dyn* 28, 97-106 (2010).
11. Z. R. Lu, E. Seo, L. Yan, S. J. Yin, Y. X. Si, G. Y. Qian, Y. D. Park, and J. M. Yang. *J Biomol Struct Dyn* 28, 259-276 (2010).
12. S. Das, S. Mitra, and J. Chakrabarti. *J Biomol Struct Dyn* 28, 239-246 (2010).
13. S. Das, R. Mukherjee, S. Sahoo, R. Thakkar, and J. Chakrabarti. *J Biomol Struct Dyn* 27, 381-390 (2009).
14. S. Zhang and T. Wang. *J Biomol Struct Dyn* 28, 247-258 (2010).
15. W. Huang, Y. Guo, and J. Zhang. *J Biomol Struct Dyn* 26, 599-608 (2009).
16. C. Jia, T. Liu, X. Zhang, H. Fu, and Q. Yang. *J Biomol Struct Dyn* 26, 763-769 (2009).
17. K. Sujatha, A. Mahalakshmi, D. K. Y. Solaiman, and R. Shenbagarathai. *J Biomol Struct Dyn* 26, 771-779 (2009).
18. A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia. *J Mol Biol* 247, 536-540 (1995).
19. H. Breiteneder, K. Pettenburger, A. Bito, R. Valenta, D. Kraft, H. Rumpold, O. Scheiner, and M. Breitenbach. *EMBO J* 8, 1935-1938 (1989).
20. L. C. van Loon, M. Rep, and C. M. Pieterse. *Annu Rev Phytopathol* 44, 135-162 (2006).
21. C. Radauer, P. Lackner, and H. Breiteneder. *BMC Evol Biol* 8, 286 (2008).
22. M. Gajhede, *et al.* *Nat Struct Biol* 3, 1040-1045 (1996).
23. J. S. Rieseke, D. H. MacLennan, and R. Coleman. *Biochem Biophys Res Commun* 15, 338-344 (1964).
24. D. Schneider and C. L. Schmidt. *Biochimica et Biophysica Acta* 1710, 1-12 (2005).
25. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. *Nucleic Acid Res* 28, 235-242 (2000).
26. S. E. Brenner, P. Koehl, and M. Levitt. *Nucleic Acid Res* 28, 254-256 (2000).
27. J. F. Gibrat, T. Madej, and S. H. Bryant. *Curr Opin Struct Biol* 6, 377-385 (1996).
28. R. B. Russell and G. J. Barton. *Proteins* 14, 309-323 (1992).
29. W. Humphrey, A. Dalke, and K. Schulten. *J Mol Graph* 14, 33-38 (1996).
30. J. Eargle, D. Wright, and Z. Luthey-Schulten. *Bioinformatics* 22, 504-506 (2006).
31. A. R. Panchenko and T. Madej. *Proteins* 57, 539-547 (2004).
32. P. O'Donoghue and Z. Luthey-Schulten. *J Mol Biol* 346, 875-894 (2005).
33. L. Holm and C. Sander. *J Mol Biol* 233, 123-138 (1993).
34. W. M. Fitch and E. Margoliash. *Science* 155, 279-284 (1967).
35. J. Felsenstein. *Cladistics* 5, 164-166 (1989).
36. K. Tamura, J. Dudley, M. Nei, and S. Kumar. *Mol Biol Evol* 24, 1596-1599 (2007).
37. M. S. Johnson, M. J. Sutcliffe, and T. L. Blundell. *J Mol Evol* 30, 43-59 (1990).
38. M. S. Johnson, A. Sali, and T. L. Blundell. *Methods Enzymol* 183, 670-690 (1992).
39. A. Marchler-Bauer, A. Panchenko, B. Shoemaker, P. Thiessen, L. Geer, and S. Bryant. *Nucleic Acid Res* 30, 281-283 (2002).
40. O. B. Pitsyn and A. V. Finkelstein. *Q Rev Biophys* 13, 339-386 (1981).
41. L. Holm and C. Sander. *Proc Int Conf Intell Syst Mol Biol* 5, 140-146 (1997).

42. R. B. Russell, M. A. Saqi, R. A. Sayle, P. A. Bates, and M. J. Sternberg. *J Mol Biol* 269, 423-439 (1997).
43. A. N. Lupas, C. P. Ponting, and R. B. Russell. *J Struct Biol* 134, 191-203 (2001).
44. C. Chothia and A. M. Lesk. *EMBO J* 5, 823-826 (1986).
45. L. Holm, and C. Sander. *Science* 273, 595-603 (1996).
46. N. V. Grishin. *J Str Biol* 134, 167-185 (2001).
47. K. Illergard, D. H. Ardell, and A. Elofsson. *Proteins* 77, 499-508 (2009).
48. A. E. Todd, C. A. Orengo, and J. M. Thornton. *J Mol Biol* 307, 1113-1143 (2001).
49. A. E. Todd, C. A. Orengo, and J. M. Thornton. *Structure* 10, 1435-1451 (2002).
50. A. G. Murzin. *Curr Opin Struct Biol* 8, 380-387 (1998).
51. Y. Matsuo and S. H. Bryant. *Proteins* 35, 70-79 (1999).
52. L. N. Kinch and N. V. Grishin. *Curr Opin Struct Biol* 12, 400-408 (2002).
53. B. Matthews. *Adv Protein Chem* 46, 249-278 (1995).
54. I. N. Berezovsky, A. Y. Grosberg, and E. N. Trifonov. *FEBS Lett* 466, 283-286 (2000).
55. I. N. Berezovsky and E. N. Trifonov. *J Biomol Struct Dyn* 19, 397-403 (2001).
56. I. N. Berezovsky and E. N. Trifonov. *J Mol Biol* 307, 1419-1426 (2001).
57. I. N. Berezovsky, V. M. Kirzhner, A. Kirzhner, V. R. Rosenfeld, and E. N. Trifonov. *J Biomol Struct Dyn* 21, 317-325 (2003).
58. I. N. Berezovsky, A. Kirzhner, V. M. Kirzhner, and E. N. Trifonov. *J Biomol Struct Dyn* 21, 327-339 (2003).
59. A. Goncarenko and I. N. Berezovsky. *Bioinformatics* 26, i497-503 (2010).
60. J. K. Capyk, I. D'Angelo, N. C. Strynadka, and L. D. Eltis. *J Biol Chem* 284, 9937-9946 (2009).
61. A. K. Pandey and C. M. Sassetti. *Proc Natl Acad Sci* 105, 4376-4380 (2008).
62. K. C. Yam, et al. *PLoS Pathog.* 5, e1000344 (2009).
63. B. Kauppi, K. Lee, E. Carredano, R. E. Parales, D. T. Gibson, H. Eklund, and S. Ramaswamy. *Structure* 6, 571-586 (1998).

*Date Received: October 11, 2010*

**Communicated by the Editor Ramaswamy H. Sarma**