An insight into the origin and functional evolution of bacterial aromatic ring-hydroxylating oxygenases

Joydeep Chakraborty, Debajyoti Ghosal, Arindam Dutta and Tapan K. Dutta*

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

Communicated by Anna Panchenko

(Received 6 August 2011; final version received 27 February 2012)

Bacterial aromatic ring-hydroxylating oxygenases (RHOs) are multicomponent enzyme systems which have potential utility in bioremediation of aromatic compounds in the environment. To cope with the enormous diversity of aromatic compounds in the environment, this enzyme family has evolved remarkably exhibiting broad substrate specificity. RHOs are multicomponent enzymes comprising of a homo- or hetero-multimeric terminal oxygenase and one or more electron transport (ET) protein(s). The present study attempts in depicting the evolutionary scenarios that might have occurred during the evolution of RHOs, by analyzing a set of available sequences including those obtained from complete genomes. A modified classification scheme identifying four new RHO types has been suggested on the basis of their evolutionary and functional behaviours, in relation to structural configuration of substrates and preferred oxygenation site(s). The present scheme emphasizes on the fact that the phylogenetic affiliation of RHOs is distributed among four distinct ‘Similarity classes’, independent of the constituent ET components. Similar combination of RHO components that was previously considered to be equivalent and classified together [Kweon et al., BMC Biochemistry 9, 11 (2008)] were found here in distinct similarity classes indicating the role of substrate-binding terminal oxygenase in guiding the evolution of RHOs irrespective of the nature of constituent ET components. Finally, a model for evolution of the multicomponent RHO enzyme system has been proposed, beginning from genesis of the terminal oxygenase components followed by recruitment of constituent ET components, finally evolving into various ‘extant’ RHO types.

Keywords: dioxygenase; evolution; biodegradation; iron–sulfur protein; Rieske protein

Introduction

Bacterial aromatic ring-hydroxylating oxygenases (RHOs), bearing Rieske centre, constitute a large family of enzymes, which have potential utility in bioremediation of diverse aromatic compounds in the environment. Degradation by oxygenases (especially dioxygenases) is a major pathway by which aromatic compounds are mineralized in the environment (Gibson & Subramanian, 1984; Gibson & Parales, 2000; Mallick, Chakraborty, & Dutta, 2011). The enzyme family has evolved to expand its catabolic potential in such a way so as to oxidize a wide range of substrates that include 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) (Chen & Que, 1999; Hudlicky, Gonzalez, & Gibson, 1999; Peng et al., 2008; Resnick, Lee, & Gibson, 1996). RHOs are multicomponent enzyme system. All members of the RHO family have one or two soluble electron transport (ET) proteins, viz. ferredoxins and reductases, having flavin and/or Fe–S centers, which transfer electrons from reduced nucleotides, NAD(P)H, to the terminal oxygenase component. Although these are separate polypeptides and are not strongly associated with the oxygenase component, their role in transferring electrons from NAD(P)H to the terminal oxygenase contributes to the overall activity of the RHO (Kauppi et al., 1998; Wolfe, Parales, Gibson, & Lipscomb, 2001). The reduced terminal oxygenase, in turn, catalyzes regio- and stereo-specific addition of both atoms of molecular oxygen into the aromatic nucleus of the substrate to form a cis-dihydrodiol which is then metabolized in subsequent steps to the tricarboxylic acid cycle intermediates. The regio- and stereo-specific additions of molecular oxygen to aromatic compounds to form cis-dihydrodiols is not only the committed and rate-limiting step in the biodegradation process, but also the resulting cis-dihydrodiols are the most interesting...
synths from the standpoint of synthetic chemists (Hud-licky et al., 1999).

The terminal oxygenase mostly comprises of two separate subunits, a large catalytic subunit (α) and a small structural subunit (β) in hetero-multimeric form, αβ. However, certain dioxygenases are devoid of β-subunit, and exist in homo-multimeric form, αα (Mason & Butler, 1997; Nojiri et al., 2005). The α-subunit of RHOs contains an N-terminal iron-sulfur protein (ISP) domain, with a conserved Rieske [2Fe–2S] cluster, however, one of the two Fe ions is bound to the ε-imidazole nitrogens of two histidine residues (Rieske, MacLennan, & Coleman, 1964). On the other hand, rubredoxins represent the simplest form of a sulfur-ligated iron cofactor bearing only one iron atom liganded by four cysteine side chains (Palmer, 1975).

The Conserved Domain Database (CDD) (Marchler-Bauer et al., 2002) classifies the C-terminal catalytic domain of the RHO α-subunit (RHO_alpha_C family; CDD: cd00680) as part of the START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC (SRPBCC) ligand-binding domain superfamily [CDD: cd07812], also known as Bet v1-like superfamily [SCOP: 55961], which spans all the three kingdoms of life. The RHO_alpha_C family comprises of seven well-characterized subgroups, viz. RHO_alpha_C_DMO-like [CDD: cd08878], RHO_alpha_C_AmdDO-like [CDD: cd08879], RHO_alpha_C_AhdA1c-like [CDD: cd08880], RHO_alpha_C_NDO-like [CDD: cd08881], RHO_alpha_C_MupW-like [CDD: cd08882], RHO_alpha_C_CMO-like [CDD: cd08883] and RHO_alpha_C_GbcA-like [CDD: cd08884] as well as three uncharacterized subgroups, RHO_alpha_C_1 [CDD: cd08885], RHO_alpha_C_2 [CDD: cd08886] and RHO_alpha_C_3 [CDD: cd08887]. The N-terminal ISP domain (Rieske_RO_Alpha_N family; CDD: cd03469) family, on the other hand, has been divided into 16 different subgroups. Conserved Rieske ISP domains (Rieske_RO_Alpha_N) have been found to be associated with several members of SRPBCC domain superfamily, showing diverse functions such as pheophorbide a oxygenase, chlorophyll a oxygenase as well as chloroplastic proteins like Tic55 and Ptc52 (Caliebe et al., 1997; Gray et al., 2004). Although distributed mostly among bacteria and in a few archaea, homologues of RHO (with a conserved RHO_alpha_C_CMO-like domain) are often found among plants (Meng, Wang, Zhang, & Nii, 2001).

To date, several genes encoding RHOs have been characterized in different species of bacteria (see supplementary Table S1 for a complete list), which give useful information regarding diversification of these catalytic genes via horizontal gene transfer (van der Meer, de Vos, Harayama, & Zehnder, 1992), gene duplication (Schuler et al., 2009), transposition events (Bosch, Garcia-Valdés, & Moore, 2000; Chablain, Zgoda, Sarde, & Truffaut, 2001; Eaton, Selifonova, & Gedney, 1998; Mattes et al., 2008; Smith, Park, Tiedje, & Mohn, 2007; Sota et al., 2006; Tsuda & Iino, 1990) and DNA rearrangements (Basta, Keck, Klein, & Stolz, 2004). But a complete understanding of the pathways leading to the evolution of RHOs with wide spectrum of substrate specificities and different combinations of ET components is yet to be revealed. Based on structural analyses, we have recently postulated the evolution of RHO α-subunit catalytic domain (RHO_alpha_C) from an ancestral PA1206-like protein, belonging to the Bet v1-like superfamily [SCOP: 55961] having a TATA binding protein (TBP)-like fold [SCOP: 55944] while, the α-subunit ISP domain have been postulated to have evolved from an ancestral Rieske ferredoxin-like protein (Chakraborty & Dutta, 2011). It has also been shown that among RHOs, the αα RHOs might have evolved earlier and have acted as a link between their ancient precursor and the present day ‘extant’ dioxygenases (both αα and αβ types) (Chakraborty & Dutta, 2011). Several attempts have been made to classify RHOs in terms of number of constituent components and the nature of their redox centers (Batie, Ballou, & Correll, 1992; Werlen, Kohler, & van der Meer, 1996) as well as in terms of α-subunit sequence homology (Nam et al., 2001). Kweon et al. (2008) recently claimed a ‘dynamic classification’ scheme for RHOs that takes into account both sequence homologies among the terminal oxygenase components as well as the nature of their constituent ET components encompassing the basis of previous classifications. In the present study, a set of available RHO α-subunit sequences have been analyzed to depict the functional evolution of diverse RHOs with varied substrate specificities and different combinations of constituent ET components, thereby suggesting a modi-
fied classification scheme for RHOs. In addition, an evolutionary model illustrating the events leading to extant multicomponent RHO enzyme system has been proposed.

Methods

**Sequence retrieval and construction of dataset**
RHO α-subunit sequences were identified using both BLASTP as well as an iterative database search with PSI-BLAST program (Altschul et al., 1997; Altschul & Koonin, 1998) against the non-redundant (NR) database at the National Center for Biotechnology Information (NCBI, NIH, Bethesda, MD) using representative sequence from each subgroup of RHO_alpha_C family (CDD: cd00680), as ‘query’. Homologous sequences were also extracted initially from the ‘Genomes’ division of the Entrez retrieval system (http://www.ncbi.nlm.nih.gov/Entrez/Genome/org.html) using the available genome annotation. Additionally, all the protein sequences from complete genomes were searched using genomic BLAST (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), using the program BLASTP, with the above queries, to detect any α-subunit homologue that might have been misannotated. Following retrieval of homologous α-subunit sequence for each RHO, those of respective ET components were also retrieved. The RHO sequences used for phylogenetic studies are listed in supplementary Table S1 while those obtained from complete genome sequences are enlisted in supplementary Table S2.

**Domain architecture of homologues of the ET components obtained from the whole genome analysis was analyzed individually using the Conserved Domain Architecture Retrieval Tool (CDART) at NCBI (Geer, Domrachev, Lipman, & Bryant, 2002).**

**Sequence alignments and phylogenetic studies**
ClustalX v1.81 (Thompson et al., 1997) was used to obtain multiple sequence alignments of either whole protein or individual domains followed by manual adjustment guided by structural information, where necessary. All the parameters were kept as default except for the matrix (BLOSUM series) used for multiple alignments. Phylogenetic trees were constructed by neighbor-joining (NJ) method (Saitou & Nei, 1987) from distance data using the NJ algorithm implemented in ClustalX. The trees were visualized and manipulated using the program TreeExplorer v2.12, a stand-alone version of the same program implemented in MEGA v4.0 (Tamura, Dudley, Nei, & Kumar, 2007) and the layout was modified using a standard graphics program. While constructing the phylogenetic tree, only a representative sequence was taken wherever highly similar sequences were encountered. Phylogenetic trees were also constructed using the maximum likelihood method (Cavalli-Sforza & Edwards, 1967) implemented in the PROML program of PHYLIP package (Felsenstein, 1996). All the trees were calculated using the global rearrangement, randomized species input order options and Jones Taylor Thornton (JTT) matrix as amino acid replacement model. Reliability or robustness

![Figure 1](image-url)
of all the inferred trees was verified by bootstrap analysis with 100 randomly permuted or resampled datasets (bootstrap replicates) generated with SEQBOOT program of the PHYLIP package. Maximum likelihood tree was plotted using the program TreeView (Page, 1996).

**Results**

**Distribution of RHOs among various bacterial lineages**

The initial dataset comprised of 165 RHO α-subunit sequences from 128 bacterial strains, most of which are either from proteobacteria or actinobacteria (supplementary Table S1) while whole genome analysis revealed additional 246 RHO α-subunit homologues from 132 bacterial and seven archaeal strains (supplementary Table S2). Distribution of RHOs in proteobacters was primarily observed among α, β and γ subdivisions whereas, in case of the phylum actinobacteria, RHOs were mostly found to be distributed among the actinomycetales (Figure 1 and supplementary Table S3). Other than proteobacters and actinobacteria, RHOs were also found to be present in few cyanobacteria, firmicutes and even in some archaean strains, mostly euryarchaeota (Figure 1, supplementary Tables S1 and S2). However, the constituent ET genes were either scattered at different loci or were missing in these organisms. Moreover, occurrence of RHOs was found to be limited to just one or two organisms within these phyla (viz. cyanobacteria, firmicutes and crenarchaeota). It is assumed that these genes may have been acquired laterally since the possibility of being remnants of a partially deleted RHO operon is ruled out due to absence of similar gene in any other member of these genera.

Other than prokaryotes, homologues of RHO were also found in strains of *Arabidopsis thaliana*, *Zea mays*, *Oryza sativa*, *Physcomitrella patens* and *Amaranthus tricolor*. This is in concurrence with previous reports regarding homologues of RHOs among plants, as described earlier (Meng et al., 2001).

**‘Modular’ genetic organization**

As shown in Table 1, RHOs are currently classified based on the combination of different types of constituent ET proteins and the presence or absence of homologous β-subunit of the terminal oxygenase (Kweon et al., 2008). Depending upon the nature of the prosthetic

| Table 1. Modified classification scheme for RHOs. |
|-----------------------------------------------|
| Modified classification | Terminal oxygenase components (prosthetic groups) | Ferredoxin (prosthetic groups) | Reductase (prosthetic groups) | Kweon classification |
|--------------------------|--------------------------------------------------|------------------------------|-------------------------------|---------------------|
| Type Iα                  | α[2Fe–2S]Rsk[Fe2+] | –                            | FNRc [FAD/FMN] | Iα                  |
| Type Iαβ                 | α[2Fe–2S]Rsk[Fe2+] | β                            | FNRc [FAD/FMN] | Iαβ                 |
| Type IIα                 | α[2Fe–2S]Rsk[Fe2+] | –                            | FNRn [FAD/FMN] | II                  |
| Type IIαβ                | α[2Fe–2S]Rsk[Fe2+] | β                            | FNRn [FAD/FMN] | IIIα                |
| Type IIIα                | α[2Fe–2S]Rsk[Fe2+] | –                            | FNRn [FAD/FMN] | IIIαβ               |
| Type IVα                 | α[2Fe–2S]Rsk[Fe2+] | –                            | GR [FAD-NAD] | IV                  |
| Type IVαβ                | α[2Fe–2S]Rsk[Fe2+] | β                            | GR [FAD-NAD] | V                   |
| Type Vaβ                 | α[2Fe–2S]Rsk[Fe2+] | β                            | GR [FAD-NAD] |                  |
| Type Vaβ                 | α[2Fe–2S]Rsk[Fe2+] | β                            | GR [FAD-NAD] |                  |
| Type VIαβ                | α[2Fe–2S]Rsk[Fe2+] | β                            | GR [FAD-NAD] |                  |
| Type VIIα                | α[2Fe–2S]Rsk[Fe2+] | –                            | GR [FAD-NAD] |                  |

*Classification scheme suggested by Kweon et al. (2008). Abbreviations: Rsk, Rieske type; Pl, Plant type; Rub, Rubredoxin; [2Fe–2S]Rsk, Cys2His2 Rieske type ISP domain of RHO α-subunit and ferredoxin; [2Fe–2S]Rsk, Cys4 plant type ISP domain of ferredoxin; [Fe]Rub, Cys4 rubredoxin; FAD, Flavin adenine dinucleotide; FMN, Flavin mononucleotide; NAD, Nicotinamide adenine dinucleotide; FNRc, Ferredoxin-NAD reductase fused with a [2Fe–2S]Rsk ISP domain at the C-terminus; FNRn, Ferredoxin-NAD reductase fused with a [2Fe–2S]Rsk ISP domain at the N-terminus; GR, Glutathione reductase; ?, unknown reductase and –, not present.
group, ferredoxin can either be a [2Fe–2S] Rieske type (Rsk), [3Fe–4S] type (3F4S), [2Fe–2S] Plant (or chloroplast)-type (Pl) or simply a Rubredoxin (Rub). On the other hand, the reductase can be of Glutathione reductase (GR)-type or Ferredoxin-NAD+-reductase (FNR)-type with a [2Fe–2S]’Pl’ type ISP domain either at the N-terminus (FNRn) or at the C-terminus (FNRc). RHOs thus seem to possess different Fe–S clusters either as individual proteins (such as ISP domains of ferredoxins) or as fused domains (e.g. ISP domains of α-subunit and FNRs). Interestingly, the Rieske [2Fe–2S] cluster present in ‘Rsk’ type ferredoxin and N-terminal ISP domain of RHO α-subunit share a common rubredoxin-like fold [SCOP: 50021, 57769], whereas the ISP domains of FNR reductases, ‘Pl’ type ferredoxins and ‘3F4S’ type ferredoxins share a common β-grasp fold [SCOP: 54235]. This suggests a possible evolutionary relationship among these proteins. Figure 2(A) shows the relative position of conserved residues in different ISP domains involved in Fe–S cluster binding, while their phylogenetic relationships are shown in Figure 2(B) and (C). The Rieske center-binding domain having conserved residues C–X–H–Xn16,17–C–X–H is believed to have evolved from rubredoxin ISP center having the distinct Fe-binding motif, C–X2–C–X–C–X–C (Link, 1999; Schmidt & Shaw, 2001). Such a trend has also been observed in the phylogenetic tree obtained from these two domains (Figure 2(B)). Moreover, it also appears

Figure 2. (A) Relative position of conserved residues (highlighted in black), involved in Fe–S cluster binding, present in ISP domain of a, RHO α-subunit; b, Rieske type ferredoxin; c, rubredoxin; d, FNR type reductase; e1 and e2, FNRn type reductase; f1 and f2, plant (or chloroplast) type ferredoxin and g, [3Fe–4S] type ferredoxin; (B) neighbor-joining tree of ISP domains having a conserved rubredoxin-like zinc finger at their core [SCOP: 50021, 57769]; (C) neighbor-joining tree of ISP domains having a β-grasp fold [SCOP: 54235]. Both the trees have been rooted with a short-chain dehydrogenase reductase (SDR) from P. putida xyl operon (NCBI: CAC86808). Each entry is represented by the protein name followed by corresponding organism designation within parentheses (see supplementary Table S1 for complete scientific names of the taxa). Values at each node indicate level of bootstrap support based on 100 resampled datasets while bootstrap values below 60% are not shown. Bar represents 0.1 substitutions per amino acid.
from Figure 2(B) that the ISP domain of RHO α-subunit have evolved from a common ancestor of Rieske ferredoxins which probably fused with an ancestral Fe$^{2+}$-binding protein, we call a ‘pre-oxygenase module’, to attain the functionality of RHO α-subunit. A similar shared ancestry is also observed among the two variants of FNR (‘FNRn’ and ‘FNRc’) with conserved residues C–X$_4$–C–X$_2$–C–X$_n$–C and plant-type ferredoxins having conserved residues C–X$_4$–C–X$_2$–C–X$_n$–C (Figure 2(C)). Origin of the two variants of FNR, ‘FNRn’ and ‘FNRc’ appear to have occurred by fusion of a plant-type [2Fe–2S] cluster with either terminus of an ancestral FNR. ‘3FeS’ type ferredoxins are characterized based on the conserved pattern C–X$_5$–C–X$_n$–C (Figure 2(A)), which is believed to have originated by deletion of one of the cysteine residues from the plant-type [2Fe–2S] cluster.

It is worth mentioning that the rubredoxin-involving dioxygenases have been exclusively characterized from the genus Rhodococcus confirming the role of rubredoxin in the ET process, although its reductase partner is still unknown (Kimura et al., 2006; Kulakov, Chen, Allen, & Larkin, 2005; Larkin, Christopher, Kulakov, & Lipscomb, 1999; Maruyama et al., 2005). However, the roles of rubredoxin and rubredoxin reductase as the components of essential ET systems in alkane hydroxylases have already been characterized (van Beilen et al., 2002). Larkin, Christopher, Kulakov, and Lipscomb (1999) observed a variation in the redox protein-binding residues of naphthalene dioxygenase α-subunit (NarA) from Rhodococcus sp. strain NCIMB12038, which is a rubredoxin-involving dioxygenase, with that of the analogous region in dioxygenases involving ferredoxin. Such a variation might be essential for the productive interaction of the terminal oxygenase with rubredoxin. Thus based on their genetic organization, we suggest this class of dioxygenases be referred to as ‘Type Vla’ RHOs and will follow this classification now onwards throughout this study. Moreover, dibenzo- p-dioxynogenase (DxnA1) from Sphingomonas sp. strain RW1, dibenzo thiophene dioxygenase (DbdCa) from Xanthobacter polyaromaticivorans strain 127W, carbazole dioxygenase (CarAa) from Sphingomonas sp. strain KA1, 2-methoxy-3,6-dichlorobenzoic acid (dicamba) monoxygenase (DdmC) from Stenotrophomonas maltophilia strain DI-6, as well as a novel diphenylamine dioxygenase (DpaAa) from Burkholderia sp. strain JS667 have shown to be associated with a ‘Pi’ type ferredoxin component, which either is exclusively involved or may complement the existing ‘Rsk’ type ferredoxin for catalysis (Armengaud, Happe, & Timmis, 1998; Hirano et al., 2006; Shin & Spain, 2009; Shintani et al., 2007). Both ‘Rub’ and ‘Pi’ type ferredoxins associated with RHOs along with other combinations of ET components were ignored in previous classifications (Batie, Ballou, & Correll, 1992; Nam et al., 2001; Werlen, Kohler, & van der Meer, 1996; Kweon et al., 2008), which have been included in this study to reconstruct the classification scheme, as presented in Table 1. Moreover, further modifications were made based on evolutionary vis-à-vis functional aspects of RHOs, which will be discussed in latter sections.

### Substrate specificity and similarity classes

Neighbor-joining trees were constructed individually based on sequences from C-terminal catalytic domain and N-terminal ISP domain of α-subunit of RHOs with the extent of similarity varying from as low as 29.2–99.6% (respective phylogenetic trees are shown in supplementary Figure S1 while the corresponding sequence alignments are shown in supplementary multiple sequence alignment files). The trees were significantly similar to that obtained from alignment of the full length α-subunit protein sequences (Figure 3) but with discrepancies in few branches and that too supported with weak-bootstrap values. Though the artifacts of phylogenetic analysis can never be ruled out completely, but given the strong statistical support for most of the nodes in the phylogenetic tree obtained from full-length protein sequences (Figure 3), this tree was considered for further classification and evolutionary analyses of RHOs. While constructing the phylogenetic tree, representative sequences were used in cases of highly similar entries (identity ≥ 99%), which are listed in supplementary Table S4. Phylogenetic tree constructed using maximum likelihood method gave similar results as obtained from the neighborhood-joining method, but with low-bootstrap support at few deep nodes (data not shown). Substrate preferences of several putative dioxygenases from whole genome sequences and/or multiple oxygenses within the same organism could be assessed based on their clustering (here termed as ‘Similarity class’) with respect to a set of well-characterized RHOs (Figure 3). As we start from deep nodes, the clustering mostly appears to be first largely according to the structural configuration of the substrate in relation to its oxygenating site(s). However, as we gradually proceed towards the subnodes, more compact clustering is observed with increase in structural similarity of the substrates in relation to specific oxygenating site(s), which further resolves according to the species tree, as revealed by 16S rRNA gene phylogeny (data not shown), with occasional events of lateral gene transfer (LGT) between proteobacters and actinobacters and frequent transfers/exchanges between β- and γ-proteobacters (Figure 3).

The α-subunit of bacterial aromatic RHOs can be classified into four similarity classes, A–D, based on their functional affiliations, and into various types based on their constituent components. In the phylogenetic tree presented in Figure 3, Class A comprises of Type IIIαβ,
Figure 3. Neighbor-joining tree of α-subunit of bacterial aromatic RHOs showing their classification based on constituent components and functional affiliations. Each entry is represented by the protein name followed by corresponding organism designation within parentheses (see supplementary Table S1 for complete scientific names of the taxa). Values at each node indicate level of bootstrap support based on 100 resampled datasets while bootstrap values below 60% are not shown. Bar represents 0.1 substitutions per amino acid. An unrelated oxidoreductase from Pseudomonas stutzeri strain A1501 (NCBI: ABP79792) was used as outgroup and position of the root is indicated by an arrow. Highly similar or identical sequences are replaced by a single representative entry (shown in supplementary Table S4), followed by an integer within brackets indicating the number of sequences represented by that entry. Double-headed vertical arrows indicate entries having similar gene organization as shown by a representative entry. Structure of the compounds recognized by each ‘Class’ of RHOs are shown in the right panel along with the site(s) of oxygenation, indicated by a thick arrow(s). Within each panel, the compounds exhibiting site(s) of oxygenation, typical to that class, are shaded in pink while those showing variation in their site(s) of oxygenation are shown without any shade. Nodes pointed in red and numbered as ‘Nx’ (x = 1–22) correspond to those ancestral RHOs whose components have been predicted as shown in Figure 4(B). Abbreviations: α, RHO α-subunit; β, RHO β-subunit; Rsk, Rieske ferredoxin; 3F4S, [3Fe–4S] ferredoxin; Rub, rubredoxin; Pl, Plant type ferredoxin; FNRc, ferredoxin-NADP⁺ reductase with an ISP domain at the C-terminus; FNRn, ferredoxin-NADP⁺ reductase with an ISP domain at the N-terminus and GR, glutathione reductase type.
Type IVαβ, Type V and Type VI RHOs (which we designate here as A-IIIαβ, A-IVαβ, A-V and A-VI, respectively) preferentially dioxygenating at α,β-positions of the aromatic ring with respect to either an adjacent fused aromatic ring or phenyl/alkyl/chloro-substitution or one of the carboxylate functional groups of phthalic acid. Class B comprises of Type Iαβ, Type IIαβ and Type IVαβ RHOs (designated as B-Iαβ, B-IIαβ and B-IVαβ, respectively) preferentially dioxygenating primarily at the vicinal position with respect to amino or carboxylate moiety. While Class C comprises of Type IIIαβ and Type IVαβ RHOs (designated as C-IIIαβ and C-IVαβ, respectively) having entries largely from whole genome sequences and/or organisms, having multiple dioxygenases apart from Type IIIαβ salicylate 5-hydroxylase and two exceptional entries of vicinally dioxygenating Type IVαβ anthranilate dioxygenases. In addition to Type Iα and Type IIIα RHOs (designated as D-Iα and D-IIIα,
respectively, few RHOs with unique combination of ET components were also identified in Class D, which we classify as Type D-IIα (bearing a α-subunit and a ‘FNRn’ type reductase), Type D-IVα (bearing a α-subunit, a ‘Rsk’ type ferredoxin and a ‘GR’ type reductase) and Type D-VIIα (bearing a α-subunit, a ‘Pl’ type ferredoxin and a ‘GR’ type reductase). As shown in Figure 3 (B), the 2-oxo-1,2-dihydroquinoline 8-monoxygenase (OxoO) from Pseudomonas putida strain 86 was found to be a Type D-IIα RHO, the carbazole dioxygenase (CarA) from Nocardioides aromaticivorans strain IC177 was found to belong to Type D-IVα RHO while Type D-VIIα RHOs included the diphenylamine dioxygenase (DpaA) from Burkholderia sp. strain JS667, the carbazole dioxygenases (two copies of CarA) from Sphingomonas sp. strain KA1 as well as dicamba monoxygenase (DdmC) from Stenotrophomonas maltophilia strain DI-6. Although the dioxin dioxygenase (DxnA1) from Sphingomonas sp. strain RW1 and carbazole dioxygenase (CarA) from Sphingomonas sp. strain KA1 are associated with a ‘GR’ type reductase (both the genes located at a distant loci with respect to those of the other components), but only the latter entry was affiliated to the new classification type (D-VIIα) as it significantly clustered with diphenylamine dioxygenase (DpaA) from Burkholderia sp. strain JS667 having similar ET components, arranged in a well-organized gene cluster. ‘Pl’ type ferredoxins were also found within the RHO operon of few other organisms, such as Xanthobacter polymoramicivorans strain 127 W (tdn operon), Pseudomonas putida strain UCC22 (tnm operon) and Delftia acidovorans strain 7N (orf operon), as well as in a genomic entry (Marinobacter algicola strain DG893). But in all the cases, the ‘Pl’ type ferredoxin is present in addition to all the essential components of the respective RHO type. Despite having only a single or very few entries, D-IIα, D-IVα and D-VIIα were assigned as distinct RHO types in the present classification, primarily due to the established involvement of these unique combinations of ET proteins in catalysis as evident from biochemical studies and their common evolutionary origin (Figure 3). In this context, it is very important to mention that RHOs bearing RHO_alpha_C_MupW-like, CMO-like, GbcA-like as well as 1-like, 2-like and 3-like domains clustered separately from a distinct node. However, most of the sequences have been derived from the whole genome annotations and lack complete information regarding their constituent ET components and biochemical function. Thus, these RHOs could not be classified in detail under the present scheme and have been designated together as D‘α since these RHOs are of α_type and also exhibit monoxygenase function (such as MupW and choline monoxygenases, involved in mupirocin and glycine betaine biosynthetic pathways respectively), similar to that observed in various RHOs belonging to Class D (Brouquisse et al., 1989; Cooper et al., 2005; Meng et al., 2001). However, occurrence of other RHO types with distinct combination of ET components in nature cannot be ignored, which will possibly be confirmed from further biochemical and molecular studies of such RHOs in the near future.

Like the other classes (A–C), few RHOs belonging to Class D also perform dioxygenation reactions, but these RHOs preferentially attack the angular positions of the aromatic ring adjacent to the heteroatom similar to carbazole, dibenzofuran, 9-fluorenone and/or dibenzo-p-dioxin dioxygenases belonging to type A-IVαβ RHOs. However, the exceptions primarily lie with Type D-IIIα RHOs, dioxygenating β,γ-positions of the aromatic ring with respect to one of the carboxylate functional groups of phthalic acid or carboxylate group of 3-chlorobenzolate. Apart from these, Class D RHOs also perform diverse monoxygenase reactions, such as vanillate O-demethylase, dicamba monoxygenase, 2-oxo-1,2-dihydroquinoline 8-monoxygenase, toluene sulphonate monoxygenase, choline monoxygenase and the monoxygenase (MupW) involved in mupirocin biosynthesis. It seems that some of the RHOs belonging to similarity Classes A and C favour dioxygenation of heteroaromatics or hetero atom-substituted aromatics at vicinal/angular positions, which is typical of RHOs belonging to Class B and D (Figure 3). This indicates a possible stabilization of the enzyme–substrate interaction via the heteroatom in these dioxygenases. Another exception among type B-IIVαβ RHOs is p-cumate dioxygenase. Rationale of the exceptional site(s) of oxygenation of RHOs belonging to a particular similarity class is discussed in later sections.

**Gene duplication, rearrangements and lateral transfer events**

Several organisms were found to possess multiple copies of the terminal oxygenase gene(s), viz. those coding for large (α) and small (β) subunits. It has been observed that the terminal oxygenases may be located in distinct operons within the same organism, such as ant and xyl operon in Pseudomonas aeruginosa strain PA01, ant and car operon in Pseudomonas resinovorans strain CA10, ben and ohb operon in Burkholderia mallei strain ATCC 23344 and etb and pad operon in Rhodococcus sp. strain RHA1. Such proteins are often placed at distant nodes belonging to the same or different class in the phylogenetic tree (Figure 3), showing traces of LGT. However, in certain cases the terminal oxygenases are found in a single operon as duplicated genes, such as akbA1a and akbA1b in Rhodococcus sp. strain DK17, ethA1a and ethA2a in Rhodococcus sp. strain RHA1, carA1 and carA1a in Sphingomonas sp. strain KAI, Caulobacter sp. strain K31, M. algincola and Chromohalobacter salexigens strain DSM 3043 or even as paralogues, such
as pdoA and pdoX in Mycobacterium sp. strain S65, phnA1a and phnA1b in Cycloclasticus sp. A5, nagAc and nagG in Polaromonas naphthaleneovorans strain C32 and Ralstonia sp. strain U2. In this context, sphingomonads such as Novosphingobium aromaticivorans, Sphingobium yanoikuyae, and several instances of gene rearrangements (Figure 3). For example, Type B-IV operon of Rhodoseudomonas palustris strain No. 7 (psb operon) showed a rearrangement in the relative positions of the α-subunit and the reductase genes. Gene rearrangements were also observed in few organisms having Type C-IVαβ RHO operons (Figure 3). Altered gene order was also found among the Type IIIαβ RHO operons belonging to both Classes A and C. Despite the altered gene arrangement of α-subunit, ferredoxin and reductase genes with respect to each other, as observed in RHO operons, however, the gene encoding the β-subunit, wherever present, was always found to be adjacent to that of the α-subunit. This reveals the possibility of lateral transfer of both the genes together, which is required for their co-evolution and has further been confirmed from the similarities observed in the phylogenetic trees obtained from both α- and β-subunit proteins. However, swapping of certain nodes was observed, though supported by low-bootstrap values (supplementary Figure S2, with corresponding sequence alignment files). It is apparent from Figure 3 that the presence of β-subunit may influence the wide diversification of αβ-type dioxygenases metabolizing large spectrum of aromatics. These observations not only support the integral role of β-subunit in the catalytic activity of αβ type RHOs but also indicate its possible role in substrate recognition and specificity, as suggested earlier (Harayama, Rekik, & Timmis, 1986; Hirose, Suyama, Hayashida, & Furukawa, 1994; Hurtubise, Barriault, & Sylvestre, 1998; Neidle et al., 1991). Moreover, a careful inspection of the published RHO sequences revealed that several RHO operons bear insertion sequence (IS) elements while some of the operons even reside within distinct conjugative transposons, which are indicative of lateral transfer events (few examples shown in supplementary Figure S3).

Discussion

The sudden outburst of industrial revolution has flooded the environment with various toxic and/or carcinogenic aromatic compounds. Many of these are recalcitrant in nature and may pose selective pressure on microbial world. Mere inheritance and evolution of the ancestral genetic material would be insufficient for the microorganisms to cope with such environmental threats unless these organisms share their ‘potencies’ for rapid adaptation against these hazardous compounds. Such adaptation under a short span of time is possible typically with lateral (or horizontal) transfer of genetic elements that encode enzymes capable of catabolizing such compounds, within different microbial communities. The role of LGT in generating novel degradation capabilities and thus increasing metabolic modularity in bacterial communities is now a well-known fact (Dutta & Pan, 2002; van der Meer et al., 1992).

Origin and lateral transfer of RHO operons

Distribution of homologous RHOs (Figure 1) mostly among proteobacteria while a few among actinobacteria, especially within the genus Rhodococcus and Mycobacterium, suggest a possible origin of these operons within the common ancestor of α-, β- and γ-proteobacteria, with frequent lateral transfer among other proteobacteria and between proteobacteria and actinobacteria (Figure 3). Moreover, occurrence of only few copies among cyanobacteria, fircmutes and archaea may be due to occasional transfer of the respective genes probably from proteobacterial ancestor as evident from the reconstructed α-subunit phylogenetic tree (data not shown) of all the entries included in Tables S1 and S2. Nevertheless, the unusual distribution of RHO gene clusters and presence of multiple terminal oxygenases component(s) in sphingomonads attain special attention. Evolutionary relationships among several RHOs in sphingomonads have already been studied (Bosch et al., 2000; Chablain et al., 2001; Eby, Beharry, Coulter, Kurtz, & Neidle, 2001; Williams & Sayers, 1994) and it has been suggested that RHOs in sphingomonads have diverged radically from those of other proteobacteria thereby restricting the genetic exchange, to a certain extent, between sphingomonads and non-sphingomonads (Stolz, 2009).

Nojiri, Shintani, and Omori (2004) reviewed different classes of mobile genetic elements such as conjugative plasmids, integrons and transposons associated with various catabolic gene clusters. They found, as in this pres-
ent study, that several RHO operons are associated with IS elements, while some RHO operons were found to reside within distinct conjugative transposons. Moreover, RHO genes/operon(s) are occasionally encoded in catabolic plasmids (see supplementary Table S1 and Figure 3). Occurrence of these features is indicative of lateral transfer of these catabolic genes within different bacterial communities thus, expanding the fraction of potential aromatic degraders in the environment.

The ‘Network theory’ of biodegradation claims that metabolic networks have a scale-free funnel-like topology whereby several peripheral pathways converge to a set of common intermediates that constitute the network’s ‘hubs’ which, in turn, are involved in a large number of interactions, close to the central metabolism (Jeong, Tombor, Albert, Oltvai & Barabasi, 2000; Trigo, Valencia, & Cases, 2009). According to the ‘Extended Complexity hypothesis’ (Aris-Brosou, 2005) these hubs, having high connectivity, are expected to be more conserved than proteins occupying a peripheral position, which in general, are involved in fewer interactions and are thus more prone to diversifying selection (Aris-Brosou, 2005; Jeong et al., 2000). This accounts for the hierarchical modularity of metabolic pathways. Consistent with these two theories, the lateral transfer of RHO genes, which cover part of the peripheral metabolic pathways, often involves the operon as a whole showing lineage-specific divergences (as in sphi gonomonads and Rhodococcus). However from the available data, it is still not clear at this moment whether the αα and ααββ type RHOs, and their associated catabolic operon as a whole, followed the hierarchical modularity of metabolic pathways discussed above. From the present analysis, it may be postulated that initially the RHO enzyme system possibly evolved as αα type monooxygenases and angular dioxygenases (those belonging to Similarity Class D) which further evolved into ααββ type RHOs. Although it may be possible that RHOs first evolved against simpler mononuclear aromatics and/or their substituted analogues with gradual evolution towards dioxygenation of substrates with complex structures whose initial degradation pathways (peripheral) often converge to the metabolic pathways of simpler substrates (lower pathways), which are either close to or part of the central metabolism.

Modular evolution of the multicomponent system

Phylogeny of the αα-subunit proteins (Figure 3) reveals that majorit y of dioxygenases catalyzing transformation of substrates with similar structural configuration and oxygenation site(s) often cluster together (as indicated by similarity classes) implying that the structural nature of substrate possibly imposes a strong selective pressure over the protein, both evolutionarily and structurally. This is generally brought about by accumulation of selective mutations at different regions of the catalytic domain, especially at the substrate-binding pockets (Dong et al., 2005; Jakoncic, Jouanneau, Meyer, & Stojanoff, 2007; Nojiri et al., 2005), along with insertions and deletions (observed from multiple sequence alignment of homologous dioxygenases, data not shown). Despite having substantial structural differences, hetero-oligomeric (ααββ) or homo-oligomeric (αα) dioxygenases have similar domain architecture and are evolutionarily related. Moreover, a modular architecture of the multicomponent RHO system is apparent (see Table 1) with varying combination of ET proteins. Although it has been reported earlier that the substrate specificity and catalytic activity of the terminal oxygenase has least relevance with the constituent ET machinery (Armengaud, Gaillard, & Timmis, 2000; Zhou, Al-Dulayymi, Baird, & Williams, 2002), but a certain extent of preferential bias of these ET proteins towards the catalytic function of RHOs belonging to different similarity classes (Figure 3) has been observed.

Genesis and/or evolution of individual components of RHOs

Based on structural phylogenetic studies of the Bet v1-like superfamily [SCOP: 55961] proteins, we have recently postulated an evolutionary link between the RHO αα-subunit catalytic domain and the PA1206-like protein (Chakraborty & Dutta, 2011). According to the hypothesis, an ancestral proteobacterial PA1206-like family protein, by spontaneous evolution, attained the Fe+3-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. Moreover, we have shown that among RHOs, the αα RHOs might have evolved earlier and have acted as a link between their ancient precursor and the present day ‘extant’ dioxygenases (both αα and ααββ types), which is also evident from Figure 3. The αα-subunit ISP domain, on the other hand, have been postulated to have evolved from an ancestral Rieske ferredoxin-like protein, both belonging to the ISP domain superfamily [SCOP: 50022], sharing the same ISP domain fold [SCOP: 50021] (Chakraborty & Dutta, 2011). These RHOs further took the support of an additional protein (the small β-subunit) and evolved to hetero-multimeric (ααββ) RHOs, probably for structural stabilization and at the same time, for increasing their plasticity to accommodate a large spectrum of aromatic compounds. The β-subunit is known to be a homologue of nuclear transport factor 2 (NTF2) (Kauppi et al., 1998).

Although studying the evolution of the Fe–S proteins in details is beyond the scope of this work, but to understand the evolution of RHOs, at least we can presume at
this moment that: (i) ‘Rsk’ ferredoxins have evolved from some ancient rubredoxin-like protein and share a common ancestry with the N-terminal ISP domain of RHO α-subunit (all of them having a rubredoxin-like zinc finger at their core), (ii) an ancient gene duplication of ‘PI’ type ferredoxins (having a β-grasp fold) followed by deletion of one conserved cysteine residue might have led to the emergence of ‘3F4S’ type ferredoxins and (iii) ‘FNRn’ and ‘FNRc’ reductases might have evolved by fusions of a ‘PI’ type ferredoxin with an ancient FNR.

A novel scheme for evolution of bacterial multicompartment RHOs

A close inspection of the constituent components of various types of RHOs reveals the possibility of the following parsimonious events of transition (Figure 4(A)): (i) the hypothetical pre-oxygenase module, upon fusion with a primitive Rieske module, evolved to an ancestral α-subunit; (ii) Type Iα RHOs, thus, would have been the most primitive enzymes of this family requiring only a ‘FNRc’ reductase in addition to the α-subunit; (iii) incorporation of a β-subunit facilitated the evolution of Type Iα RHOs to Type Iαβ RHOs while replacement of the ‘FNRc’ with a ‘FNRn’ reductase generated Type Iαn RHOs which, on further recruitment of a ‘Rsk’ ferredoxin, evolved to Type Iαnβ RHOs; (iv) Type Iαβ RHOs, in turn, replaced their reductase component from ‘FNRn’ to ‘GR’ so as to evolve to Type IVα RHOs, while replacement of ‘Rsk’ ferredoxin of the Type IVα RHOs with a ‘PI’ gave rise to Type VIIα RHOs; (v) replacement of ‘FNRc’ reductase to ‘FNRn’, on the other hand, led to the evolution of Type Iαβ RHOs to Type Iαβ RHOs; (vi) recruitment of a ‘Rsk’ ferredoxin further guided the evolution of Type Iαβ towards Type Iαβ RHOs, followed by (vii) replacement of ‘FNRn’ reductase by ‘GR’ to form Type IVαβ RHOs; (viii) replacement of ‘Rsk’ ferredoxin of Type IVαβ by ‘3F4S’ ferredoxin further generated Type Vaβ RHOs and finally, (ix) Type VIαβ RHOs evolved from Type Vaβ RHOs by recruitment of a ‘Rub’ followed by loss of ‘3F4S’ ferredoxin. Although, it is yet unknown whether the reductase component (GR) in Type VIαβ RHOs has been retained or not.

Though it is believed that nature always prefers to evolve in the most parsimonious way involving least changes as depicted in Figure 4(A), but the phylogeny of RHO α-subunit (Figure 3) suggests that the evolution of RHOs have traversed through a more complex path. This is because the evolution of RHOs primarily reflects the evolution of terminal oxygenase components with limited dependence on substitution or gene-gain/loss of the ET components. To depict the possible evolutionary events, different types of RHOs belonging to the four similarity class were first arranged so as to represent each type as a set of constituent genes and originating from a common node (ancestor) shared by a different RHO type. The sets and nodes thus formed were arranged in terms of ‘generation’ of RHOs according to the hierarchy of the branching pattern (Figure 4(B)). To predict the ‘gene set’ of a node, a heuristic algorithm was developed as described below.

If D1n and D2n denote the gene set of two daughter RHO types, D1 and D2 of nth generation with known gene sets (see inset in Figure 4(B)), having a common parent, P1 (of n − 1th generation), then the set of gene(s) vertically inherited and common to both the daughters,  ‫pD1n,D2n‬, is defined by the following equation

\[ p_{D1n,D2n} = D1n \cap D2n \] (1)

and their parental gene set, P1n−1, is expressed as

\[ P1n−1 = p_{D1n,D2n} \cup (P2n−1 \cap (D1n \Delta D2n)) \] (2)

where the symmetric difference, D1nΔD2n, represents the set of gene(s) uncommon to both D1 and D2 and P2n−1 denotes the set of gene(s) belonging to another RHO type, P2 (with known set of genes), from the n − 1th generation sharing a common ancestry with P1 (both having a common ancestor, G).

Gene(s) horizontally acquired by the RHOs, D1 and/or D2, denoted by dDn, are expressed as the complement of P1n−1 relative to D1nΔD2n and the gene(s) lost (dDn) are expressed as the complement of iDn relative to D1nΔD2n, i.e.

\[ i_{Dn} = (D1n \Delta D2n) \setminus P1n−1 \] (3)

and

\[ d_{Dn} = (D1n \Delta D2n) \setminus i_{Dn} \] (4)

In addition, a few approximations were also made such that

If D1n=D2n, then

\[ P1n−1 = D1n = D2n \] (5)

and if D2n ⊆ D1n, then

\[ P1n−1 = D1n \] (6)

and once P1n−1=D1n, then

\[ D2n = D1n \cup (D1n \setminus D2n) \] (7)
Figure 4. Evolution of RHOs: (A) a model of the most parsimonious route, involving the least changes. Arrows indicate transition between RHO types along with gene-gain/loss events. Shades corresponding to each RHO component are the same as used in Figure 3. (B) Proposed scheme for origin and functional evolution of RHOs, suggested by the phylogeny of the RHO α-subunit, and predicted by a heuristic algorithm as described in the text. Numbers at the top denote ‘generation’ of RHOs. The set of genes within each square box represent an ‘extant’ RHO type with conserved RHO_alpha_C domain while each node (designated by ‘Nx’ where x = 1, 2, …, 22) represents the common ancestor of a pair of ‘extant’ RHO types. The predicted set of genes is shown within braces below each node and the type of RHO affiliated to the ancestor of each ‘Similarity Class’ is mentioned within parentheses. The yellow shaded box (inset) shows a generalization of the above model where ‘D1’ and ‘D2’ denote two different RHO types from nth generation, sharing a common ancestry with parent ‘P1’ (from n – 1th generation). Both ‘P1’ and ‘P2’ (both belonging to n – 1th generation), in turn, have a common parent, G, belonging to the n – 2th generation. Known set of genes from D1, D2 and P2 can be used to predict the set of genes for P1. Abbreviations: α, RHO α-subunit; β, RHO β-subunit; Rsk, Rieske ferredoxin; 3F4S, [3Fe–4S] ferredoxin; Rub, rubredoxin; Pl, Plant type ferredoxin; FNRc, ferredoxin-NADP⁺ reductase with a ISP domain at the C-terminus; FNRn, ferredoxin-NADP⁺ reductase with a ISP domain at the N-terminus and GR, glutathione reductase type.
where $D_1 \backslash D_2$ denotes the complement of $D_2$ relative to $D_1$.

After genesis of the first functional enzyme system, the evolutionary scheme could be divided into eight different generations (Figure 4(B)) and 19 nodes. A ‘trace-back’ approach was taken to first predict the components of each peripheral node and then proceeding towards the node of preceding generation using the above algorithm.

In terms of RHO activity, expression of $\alpha$-subunit protein alone would have been non-functional unless its expression was coupled with any ferredoxin and/or reductase protein(s). Thus, following the above approach, it can be hypothesized that the ancestral protein, representing the RHO $\alpha$-subunit with RHO$\alpha$$_C$$DMO$-like domain, combined with a ‘FN$R$’ reductase to give the functional Type Ia RHO (node $N20$, Figure 4(B)), further evolving into the extant Type D-Ia RHOs (all of them having the DMO-like domain). The same protein, in association with ‘Rs’ ferredoxin and ‘GR’ reductase, further evolved into a Type IVa RHO (node $N7$, Figure 4(B)), which was the progenitor of Type D-IIa, D-IIIa, D-IVa and D-VIIa RHOs containing the DMO-like domain and another intermediate Type IVa RHO (node $N14$, Figure 4(B)). This Type IVa RHO further evolved into two variants of RHO where one of the variants remained homo-oligomeric ($\alpha_n$) and is supposed to be the ancestor of RHOs belonging to the Class D$^*$ (those having MupW-, CMO-, GbcA-, 1-, 2- and 3-like RHO$\alpha$$_C$ domains) (node $N15$, Figure 4(B)). While, the other incorporated a $\beta$-subunit and evolved to hetero-oligomeric ($\alpha_n\beta_n$) oxygenases (represented by the ancestral state at node $N11$ (Figure 4(B)) such as those within Class A (with NDO-like domain), Class B (with AntDO-like domain) and Class C (with AhdA1c-like domain) RHOs. Further, according to our predictions the ancestors of RHOs belonging to Class A, B and C were Type V$\alpha$V$\beta$ (node $N7$, Figure 4(B)), Type I$\alpha$$\beta$ (node $N9$, Figure 4(B)) and Type IV$\alpha$V$\beta$ (node $N12$, Figure 4(B)) RHO, respectively. Such an evolution was accompanied by several instances of gene-loss and gene-recruitment events guided by substrate-mediated selection pressure as and when encountered. Xenologous gene displacement, gene rearrangements and LGT (Armengaud et al., 2000) might have played a crucial role in the evolution of RHOs in general.

**Evolutionary classification and functional implications**

The phylogenetic tree shown in Figure 3 is in good accord with those presented by Nam et al. (2001) and Kweon et al. (2008). The present study emphasizes that it is the terminal oxygenase that plays the major role in functional diversification of RHOs, as depicted in our scheme (Figure 4(B)). From our observation, it is quite obvious that both $\alpha$- and $\beta$-subunits of the RHO system have co-evolved and it was observed that the ET components are interchangeable, as also previously investigated (Armengaud et al., 2000; Zhou et al., 2002). Variation in the distribution of ET components among RHOs may have occurred either due to xenologous gene displacement or due to certain biasness towards substrate selection. But whatever might be the reason, similar combination of ET components is usually observed in several similarity classes (Figure 3) and in different evolving ‘generations’ of RHOs (Figure 4(B)). The major discriminating factor that differentiates various RHOs from each other is their substrate preferences in relation to their site(s) of oxygenation, which is largely reflected in their similarity classes (Figure 3). Thus, it can be said that initially the divergence occurred with respect to certain classes of compounds with distinct oxygenating site(s) (interclass divergence) followed by further evolution of each variant (common ancestor of each cluster) towards structurally diverse compounds (intraclass divergence) while maintaining similar oxygenation site(s). Occasionally, encountering a somewhat different structure led to strong selection pressure on the protein thus diversifying it to a greater extent, as can be seen in case of Type A-V$\alpha$V$\beta$ diterpenoid dioxygenase (DitA) and Type A-IV$\alpha$V$\beta$ dibenzofuran (DbfA), carbazole (CarA) and dibenzo-p-dioxin (DxnA) dioxygenases. This is quite obvious because physicochemical properties of the substrate-binding pocket(s) of the $\alpha$-subunit bring about differences in the binding conformations of the substrates. It has been observed earlier in several studies that RHOs perform diverse reactions such as dealkylation, sulphotisation, benzylc monooxygenation as well as desaturation, which reflects variation in the topology of enzyme–substrate interaction in the active site as well as the reaction mechanism involved (Gibson et al., 1995; Guengerich, 1990, 1991; Lehning, Fock, Wittich, Timmis, & Pieper, 1997; Robertson, Spain, Haddock, & Gib- son, 1992). Moreover, altered substrate selectivity as well as mode (monooxygenase or dioxygenase) and specificity (site) of oxygenation were reported from various mutants of the catalytic $\alpha$-subunit with substituted amino acid(s) at the substrate-binding pocket (Uchimura et al., 2008). In this context, dioxygenation of compounds such as carbazole, dibenzofuran, 9-fluorenone and dibenzo-p-dioxin, favouring angular dioxygenation (vicinal to heteroatom-substituted carbon) is attributed to stabilization of the enzyme–substrate interaction via the heteroatom, as shown in case of carbazole dioxygenase from *Janthinobacterium* sp. strain J3 where the interaction is stabilized by hydrogen bonding of carbonyl oxygen of Gly178 with the imino nitrogen of the substrate, favoring dioxygenation at the C9$\alpha$ and C1 positions, as revealed by structural analysis (Nojiri et al., 2005; Uchimura et al., 2008). Similarly, it may be argued that dioxygenation at the vicinal position of benzoate-like
compounds such as benzoate, toluate and anthranilate (or aniline and nitroaromatics) is attributed to enzyme–substrate interaction stabilized by the carboxylate (or amino and nitro) moiety. Thus, this has been observed that the RHOs oxygenating heteroaromatics or hetero atom-substituted aromatics containing highly electronegative heteroatoms such as oxygen and nitrogen prefer the vicinal (or angular) dioxygenation irrespective of their affinity. However, interaction of the heteroatom with the amino acid residue(s) within the substrate-binding pocket of the enzyme, favoring vicinal dioxygenation, can only be understood by detailed structural analysis of the respective enzyme–substrate complex.

Although the grouping of ‘incomplete’ RHOs with ‘standard’ samples according to the similarity index (represented by pairwise distance, PD) suggested by Kweon et al. (2008) is apparently quite significant, but their classification scheme barely indicates that the phylogenetic affiliation of RHOs is mainly determined by their relationship with ET components, as claimed. This is because similar combinations of RHO components are found in distinct similarity classes in the phylogeny of α-subunits (Figures 3 and 4(B)) indicating that such combinations have occurred independently during evolution. So it would be inappropriate to consider, e.g., Type A-IVαβ, Type B-IVαβ and Type C-IVαβ RHOs to be equivalent and be classified together as Type IV RHOs as suggested by Kweon et al. (2008). What distinguishes them from each other are their substrate preferences, which in fact, is the consequence of divergent evolution of the substrate-binding terminal oxygenase, irrespective of the constituent ET components. Although, there exists a certain extent of preferential bias of the ET proteins towards the catalytic function of the RHOs belonging to different similarity classes. Moreover, Kweon et al. (2008) ignored the involvement of rubredoxin (as in Type A-VIαβ RHOs) as one of the ET components, which is exclusively found within the genus Rhodococcus and have been discussed in detail (Kulakov, Chen, Allen, & Larkin, 2005). According to their classification scheme, both indene dioxygenase (NiD) from Rhodococcus sp. strain I24 and diterpenoid dioxygenase (Dita) from Pseudomonas abietanaphila strain BKME-9 (lacking information regarding their ET components and reductase respectively) were affiliated to Type V RHOs. Though from our analysis, Dita does belong to Type A-VIαβ RHOs (due to the presence of a [3Fe-4S] type ferredoxin) but NiDα significantly clustered within Type A-VIαβ RHOs (Figure 3). Previous classification schemes also ignored the involvement of ‘PI’ type ferredoxins as one of the ET components of RHOs, which had already been characterized earlier (Inoue et al., 2004; Shintani et al., 2007) and further substantiated by Shin and Spain (Shin & Spain, 2009) in a recent report on diphenylamine dioxygenase (DpaA) from Burkholderia sp. strain JS667. Thus, in addition to Type VIαβ RHOs, we have identified three more RHO types (designated as Type D-IIα, Type D-IVα and Type D-VIα), which we have classified on the basis of phylogenetic affiliation of the terminal oxygenases as well as their unique combinations of ET components (Table 1).

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 truly relies on the information content of various biological databases. We have already 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 (Chakraborty & Dutta, 2011). The present study attempts to sketch a probable route of evolution of bacterial RHOs beginning from genesis of terminal oxygenase components followed by recruitment of constituent ET components. The proposed modified classification scheme reflects both evolutionary and functional aspects of RHOs, depending on the growing pool of available sequence information. However, recruitment and evolution of downstream genes associated with RHOs, involved in a variety of aromatic metabolic pathways, were ignored as this study focuses exclusively on the origin and functional evolution of the RHO components only and evolution of the complete catabolic operon calls for a subject of separate analysis. Our classification scheme takes into account both the functional evolution of catalytic α-subunit as well as all possible combinations of ET components associated with terminal oxygenase known till date, which ideally reflect the evolutionary and functional aspects of RHOs. However, it is strongly believed that the classification is subject to further modifications as more and more sequence information and/or information regarding involvement of new ET components becomes available. Possibility of some other combination of the existing RHO components, involvement of completely different set of ET proteins known to be associated with other systems or those still uncharacterized/unknown, cannot be denied. In this context, occurrence and characterization of RHOs in unculturable bacteria from metagenomic studies may shed light in updating the present classification and other crucial aspects of RHO evolution.
Acknowledgements
This work was supported in part by a Grant-in-aid from Department of Biotechnology, Government of India (#BT/PR11831/BID/07/290/2009 to T.K.D), and Bose Institute, Kolkata, India.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2012.682208.

References
Altschul, S.F., & Koonin, E.V. (1998). PSI-BLAST—a tool for making discoveries in sequence databases. Trends in Biochemical Sciences, 23, 444–447.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25, 3389–3402.
Aris-Brosou, S. (2005). Determinants of adaptive evolution at the molecular level: The extended complexity hypothesis. Molecular Biology and Evolution, 22, 200–209.
Armengaud, J., Gaillard, J., & Timmis, K.N. (2000). A second [2Fe–2S] ferredoxin from Sphingomonas sp. strain RW1 can function as an electron donor for the dioxin dioxygenase. Journal of Bacteriology, 182, 2238–2244.
Armengaud, J., Happe, B., & Timmis, K.N. (1998). Genetic analysis of dioxin dioxygenase of Sphingomonas sp. Strain RW1: Catabolic genes dispersed on the genome. Journal of Bacteriology, 180, 3954–3966.
Basta, T., Keck, A., Klein, J., & Stolz, A. (2004). Detection and characterization of conjugative degradative plasmids in xenobiotics degrading Sphingomonas strains. Journal of Bacteriology, 186, 3862–3872.
Batie, C.J., Ballou, D.P., & Correll, C.C. (1992). Pthalalate dioxygenase reductase and related flavin-iron-sulfur containing electron transferases. In F. Müller (Ed.), Chemistry and biochemistry of flavoenzymes (pp. 543–556). Boca Raton, FL: CRC Press.
Beinert, H., Holm, R.H., & Münck, E. (1997). Iron-sulfur clusters: Nature’s modular multipurpose structures. Science, 277, 653–659.
Bosch, R., García-Valdés, E., & Moore, E.R.B. (2000). Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from Pseudomonas stutzeri AN10. Gene, 245, 65–74.
Brouquisse, R., Weigel, P., Rhodes, D., Yocum, C.F., & Hanson, A.D. (1989). Evidence for a ferredoxin-dependent choline monooxygenase from spinach chloroplast stroma. Plant Physiology, 90, 322–329.
Caliebe, A., Grimm, R., Kaiser, G., Beck, J.L., Soll, J., & Heins, L. (1997). The chloroplastic protein import machinery contains a Rieske-type iron–sulfur cluster and a mononuclear iron-binding protein. EMBO Journal, 16, 7342–7350.
Cammack, R.C. (1992). Iron-Sulfur Proteins. In R. Cammack & A.G. Sykes (Eds.), Advances in inorganic chemistry (pp. 281–322). San Diego, CA: Academic Press.
Cammack, R., Gay, E., & Shergill, J.K. (1999). Studies of hyperfine interactions in [2Fe–2S] proteins by EPR and double resonance spectroscopy. Coordination Chemistry Reviews, 192, 1003–1022.
Cavalli-Sforza, L.L., & Edwards, A.W.F. (1967). Phylogenetic analysis: Models and estimation procedures. The American Journal of Human Genetics, 19, 233–257.
Chablain, P.A., Zgoda, A.L., Sarde, C.-O., & Truffaut, N. (2001). Genetic and molecular organization of the alkylbenzene catabolism operon in the psychrotrophic strain Pseudomonas putida 01G3. Applied and Environmental Microbiology, 67, 453–458.
Chakraborty, J., & Dutta, T.K. (2011). From lipid transport to oxygenation of aromatic compounds: Evolution within the Bet v1-like superfamily. Journal of Biomolecular Structure and Dynamics, 29, 67–78.
Chen, K., & Que, L. (1999). Cis-dihydroxylation of olefins by a non-heme iron catalyst: A functional model for Rieske dioxygenases. Angewandte Chemie International Edition, 38, 2227–2228.
Cooper, S.M., Cox, R.J., Crosby, J., Crump, M.P., Hotherrs, J., Laosripaiboon, W., ... Thomas, C.M. (2005). Mupirocin W, a novel pseudomonic acid produced by targeted mutation of the mupirocin biosynthetic gene cluster. Chemical Communications, 9, 1179–1181.
Dong, X., Fushinobu, S., Fukuda, E., Terada, T., Nakamura, S., Shimizu, K., ... Wakagi, T. (2005). Crystal structure of the terminal oxygenase component of cumene dioxygenase from Pseudomonas fluorescens IP01. Journal of Bacteriology, 187, 2483–2490.
Dutta, C., & Pan, A. (2002). Horizontal gene transfer and bacterial diversity. Journal of Biosciences, 27, 27–33.
Eaton, R.W., Selifonova, O.V., & Gedney, R.M. (1998). Isopropylbenzene catalytic pathway in Pseudomonas putida RE204: Nucleotide sequence analysis of the ipb operon and neighboring DNA from pRE4. Biodegradation, 9, 119–132.
Eby, D.M., Beharry, Z.M., Coulter, E.D., Kurtz, D.M., & Neidle, E.L. (2001). Characterization and evolution of anthranilate 1,2-dioxygenase from Actinetobacter sp. strain ADP1. Journal of Bacteriology, 183, 109–118.
Fang, G., Rocha, E.P.C., & Danchin, A. (2008). Persistence drives gene clustering in bacterial genomes. BMC Genomics, 9, 4.
Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods in Enzymology, 266, 418–427.
Geer, L.Y., Domrachev, M., Lipman, D.J., & Bryant, S.H. (2002). CDART: Protein homology by domain architecture. Genome Research, 12, 1619–1623.
Gibson, D.T., & Parales, R.E. (2000). Aromatic hydrocarbon dioxygenases in environmental biotechnology. Current Opinion in Biotechnology, 11, 236–243.
Gibson, D.T., Resnick, S.M., Lee, K., Brand, J.M., Torok, D.S., Wackett, L.P., ... Haigler, B.E. (1995). Desaturation, dioxygenation, and monoxygenation reactions catalyzed by naphthalene dioxygenase from Pseudomonas sp. strain 9816-4. Journal of Bacteriology, 177, 2615–2621.
Gibson, D.T., & Subramanian, V. (1984). Microbial degradation of aromatic hydrocarbons. In D.T. Gibson (Ed.), Microbial degradation of organic compounds (pp. 181–252). New York, NY: Marcel Dekker.
Gray, J., Wardzala, E., Yang, M., Reinbothe, S., Hailer, S., & Pauli, F. (2004). A small family of LLS1-related non-heme oxygenases in plants with an origin amongst oxygenic photosynthesizers. Plant Molecular Biology, 54, 39–54.
Guengerich, F.P. (1990). Enzymatic oxidation of xenobiotic chemicals. Critical Reviews in Biochemistry and Molecular Biology, 25, 97–153.
Guengerich, F.P. (1991). Reactions and significance of cytochrome P-450 enzymes. *The Journal of Biological Chemistry*, 266, 10019–10022.

Harayama, S., Rekik, M., & Timmis, K.N. (1986). Genetic analysis of a relaxed substrate specificity aromatic ring dioxygenase, toluate 1,2-dioxygenase, encoded by TOL plasmid pWWO of *Pseudomonas putida*. *Molecular and General Genetics*, 202, 226–234.

Hirano, S., Haruki, M., Takano, Y., Imanaka, T., Morikawa, M., & Kanaya, S. (2006). Gene cloning and *in vivo* characterization of a dibenzoziophiene dioxygenase from *Xanthobacter polyaromaticivorans*. *Applied Microbiology and Biotechnology*, 69, 672–681.

Hirose, J., Suyama, A., Hayashida, S., & Furukawa, K. (1994). Determination of the terminal oxygenase subunit in the biphenyl dioxygenase reactivity pattern toward chlorobiphenyls. *Journal of Bacteriology*, 176, 3283–3284.

Hirano, S., Haruki, M., Takano, K., Imanaka, T., Morikawa, M., & Kanaya, S. (2006). Gene cloning and *in vivo* characterization of a dibenzoziophiene dioxygenase from *Xanthobacter polyaromaticivorans*. *Applied Microbiology and Biotechnology*, 69, 672–681.

Hirose, J., Suyama, A., Hayashida, S., & Furukawa, K. (1994). Construction of hybrid biphenyl (*bph*) and toluene (*tol*) genes for functional analysis of aromatic ring dioxygenases. *Gene*, 138, 27–33.

Hudlicky, T., Gonzalez, D., & Gibson, D.T. (1999). Enzymatic dihydroxylation of aromatics in enantioselective synthesis: Expanding asymmetric methodology. *Aldrichimica Acta*, 32, 35–62.

Hurtubise, Y., Barriault, D., & Sylvestre, M. (1998). Involvement of the terminal oxygenase β subunit in the biphenyl dioxygenase reactivity pattern toward chlorobiphenyls. *Journal of Bacteriology*, 180, 5828–5835.

Huyen, M.A., & Bork, P. (1998). Measuring genome evolution. *Proceedings of the National Academy of Sciences USA*, 95, 5849–5856.

Inoue, K., Wida da, J., Nakai, S., Endoh, T., Urata, M., Ashikawa, Y., … Nojiri, H. (2004). Divergent structures of carboxzole degradative car operons isolated from gram-negative bacteria. *Bioscience, Biotechnology, and Biochemistry*, 68, 1467–1480.

Itoh, T., Takemoto, K., Mori, H., & Gojobori, T. (1999). Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Molecular Biology and Evolution*, 16, 332–346.

Jakoncic, J., Jouanneau, Y., Meyer, C., & Stojanoff, V. (2007). The crystal structure of the ring-hydroxylating dioxygenase from *Sphingomonas CHY-1*. *FEBS Journal*, 274, 2470–2481.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., & Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, 407, 651–654.

Jiang, H., Parales, R.E., Lynch, N.A., & Gibson, D.T. (1996). Site-directed mutagenesis of conserved amino acids in the alpha subunit of tolulene dioxygenase: Potential mononuclear non-heme iron coordination sites. *Journal of Bacteriology*, 178, 3133–3139.

Kauppi, B., Lee, K., Carredano, E., Parales, R.E., Gibson, D. T., Eklund, H., & Ramaswamy, S. (1998). Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase. *Structure*, 6, 571–586.

Kimura, N., Kitagawa, W., Mori, T., Nakashima, N., Tamura, T., & Kamagata, Y. (2006). Genetic and biochemical characterization of the dioxygenase involved in lateral dioxygenation of dibenzofuran from *Rhodococcus opacus* strain SAO101. *Applied Microbiology and Biotechnology*, 73, 474–484.

Kulakov, L.A., Chen, S., Allen, C.C.R., & Larkin, M.J. (2005). Web-type evolution of *Rhodococcus* gene clusters associated with utilization of naphthalene. *Applied and Environmental Microbiology*, 71, 1754–1764.

Kweon, O., Kim, S.-J., Baek, S., Chae, J.-C., Adjei, M.D., Baek, D.-H., … Cerniglia, C.E. (2008). A new classification system for bacterial Rieske non-heme iron aromatic-ring-hydroxylating dioxygenases. *BMC Biochemistry*, 9, 11.

Larkin, M.J., Christopher, C.R.A., Kulakov, L.A., & Lipscomb, D.A. (1999). Purification and characterization of a novel naphthalene dioxygenase from *Rhodococcus* sp. strain NCIMB12038. *Journal of Bacteriology*, 181, 6200–6204.

Lawrence, J.G., & Roth, J.R. (1996). Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics*, 143, 1843–1860.

Lehning, A., Fock, U., Wittich, R., Timmis, K.N., & Pieper, D. H. (1997). Metabolism of chlorotoluenes by *Burkholderia* sp. strain PS12 and toluene dioxygenase of *Pseudomonas putida* F1: Evidence for monooxygenation by toluene and chlorobenzene dioxygenases. *Applied and Environmental Microbiology*, 63, 1974–1979.

Link, T.A. (1999). The structures of Rieske and Rieske-type proteins. *Advances in Inorganic Chemistry*, 47, 83–157.

Mallick, S., Chakraborty, J., & Dutta, T.K. (2011). Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular weight polycyclic aromatic hydrocarbons: A review. *Critical Reviews in Microbiology*, 37, 64–90.

Marchler-Bauer, A., Panchenko, A., Shoemaker, B., Thiessen, P., Geer, L., & Bryant, S. (2002). CDD: A database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Research*, 30, 281–283.

Maruyama, T., Ishikura, M., Taki, H., Shindo, K., Kasai, H., Haga, M., … Misawa, N. (2005). Isolation and characterization of o-xylene oxygenase genes from *Rhodococcus opacus* TKN14. *Applied and Environmental Microbiology*, 71, 7705–7715.

Mason, J.R., & Butler, C.S. (1997). Structure-function analysis of the bacterial ring-hydroxylating dioxygenases. *Advances in Microbial Physiology*, 38, 47–84.

Mates, T.E., Alexander, A.K., Richardson, P.M., Munk, A.C., Han, C.S., Stothard, P., & Coleman, N.V. (2008). The Genome of *Polaromonas* sp. strain JS666: Insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. *Applied and Environmental Microbiology*, 74, 6405–6416.

Meng, Y.L., Wang, Y.M., Zhang, B., & Nii, N. (2001). Isolation of a choline monooxygenase cDNA clone from *Arabidopsis tricloro* and its expressions under stress conditions. *Cell Research*, 11, 187–193.

Nam, J.W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H., & Omori, T. (2001). New classification system for oxygenase components involved in ring-hydroxylating oxygogenations. *Bioscience, Biotechnology, and Biochemistry*, 65, 254–263.

Neidle, E.L., Hartnett, C., Ornston, L.N., Bairoch, A., Rekik, M., & Harayama, S. (1991). Nucleotide sequences of the *Acinetobacter calcoaceticus* benABC genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. *Journal of Bacteriology*, 173, 5385–5395.

Nojiri, H., Ashikawa, Y., Noguchi, H., Nam, J.W., Urata, M., Funamoto, Z., … Omori, T. (2005). Structure of the terminal oxygenase component of angular dioxygenase, carbazole 1,9a-dioxygenase. *Journal of Molecular Biology*, 351, 355–370.

Nojiri, H., Shintani, M., & Omori, T. (2004). Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Applied Microbiology and Biotechnology*, 64, 154–174.
