BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(23)

Analysis of multi-domain hypothetical proteins containing iron-sulphur clusters and fad ligands reveal rieske dioxygenase activity suggesting their plausible roles in bioremediation

Nitish Sathyanarayanan & Holenarasipur Gundurao Nagendra*

Department of Biotechnology, Sir M. Visvesvaraya Institute of Technology, Krishnadevarayanagar, Hunasamaranahalli, Bangalore 562 157; Holenarasipur Gundurao Nagendra - Email: nagshaila@gmail.com; *Corresponding author

Received September 30, 2012; Accepted November 05, 2012; Published November 23, 2012

Abstract:
'Conserved hypothetical' proteins pose a challenge not just for functional genomics, but also to biology in general. As long as there are hundreds of conserved proteins with unknown function in model organisms such as *Escherichia coli*, *Bacillus subtilis* or *Saccharomyces cerevisiae*, any discussion towards a 'complete' understanding of these biological systems will remain a wishful thinking. Insilico approaches exhibit great promise towards attempts that enable appreciating the plausible roles of these hypothetical proteins. Among the majority of genomic proteins, two-thirds in unicellular organisms and more than 80% in metazoa, are multi-domain proteins, created as a result of gene duplication events. Aromatic ring-hydroxylating dioxygenases, also called Rieske dioxygenases (RDOs), are class of multi-domain proteins that catalyze the initial step in microbial aerobic degradation of many aromatic compounds. Investigations here address the computational characterization of hypothetical proteins containing Ferredoxin and Flavodoxin signatures. Consensus sequence of each class of oxidoreductase was obtained by a phylogenetic analysis, involving clustering methods based on evolutionary relationship. A synthetic sequence was developed by combining the consensus, which was used as the basis to search for their homologs via BLAST. The exercise yielded 129 multi-domain hypothetical proteins containing both 2Fe-2S (Ferredoxin) and FNR (Flavodoxin) domains. In the current study, 40 proteins with N-terminus 2Fe-2S domain and C-terminus FNR domain are characterized, through homology modelling and docking exercises which suggest dioxygenase activity indicating their plausible roles in degradation of aromatic moieties.

Background:
Over the last decade, more than 150 complete genomes of diverse bacteria, archaea and eukaryotes have been sequenced, and many more are currently in the pipeline **[1].** It is well known that, in any newly sequenced bacterial genome, as many as 30–40% of the genes do not have an assigned function **[2].** This figure is even higher for archaeal and eukaryotic genomes and for the relatively large genomes of bacteria with a complex life style, such as *Anabaena*, *Streptomyces,* etc **[3, 4]**.

'Conserved hypothetical' proteins pose a challenge not just to functional genomics, but also to biology in general **[5].** As long as there are hundreds of conserved proteins of unknown function even in model organisms, such as *Escherichia coli*, *Bacillus subtilis* or *Saccharomyces cerevisiae*, any discussion of a 'complete' understanding of these organisms as biological systems will remain in the realm of wishful thinking. Although it appears likely that the central pathways of information processing and metabolism are already known, crucial elements of these systems could still be lurking among the 'conserved hypotheticals', and important mechanisms of signalling and stress response, in all likelihood, would remain undiscovered **[6].**

Aromatic compounds are widely distributed throughout the biosphere predominantly in the form of recycled material **[7].**

Because of the inherent thermodynamic stability of the aromatic ring, natural turnover of these compounds is slow and instead relies on complex microbial degradation pathways. With aromatic compounds comprising >25% of the earth's biomass, these pathways play a crucial role in the biogeochemical carbon cycle. However, despite the abundance of microbial degraders, man-made aromatic pollutants are often recalcitrant to existing bioprocessing pathways. As a result, these xenobiotic compounds, many of which are derived from the processing of crude oil, persist in the environment causing irreversible damage to the biosphere **[7].**

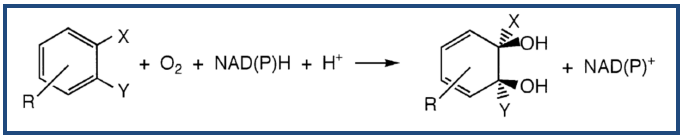

Figure 1: Reaction of ring cleavage mediated by RDO

Aromatic ring-hydroxylating dioxygenases, also called Rieske dioxygenases (RDOs), are class of multi-domain proteins that catalyze the initial step in microbial aerobic degradation of many aromatic compounds. Two hydroxyl groups are introduced into the aromatic ring yielding cyclic cis-dihydrodiols or cis-diol carboxylic acids **(Figure 1)** [Substituents X and Y can be hydrogen atoms or any of several other groups] **[8, 9].**

More than three dozen distinct RDOs have been identified. RDOs consist of a reductase, an oxygenase and in some cases, an additional ferredoxin that mediates electron transfer between the former two components. The oxygenase component catalyzes the insertion of both atoms of molecular oxygen into the aromatic substrate, which is believed to occur at a mononuclear iron site and to be accompanied by electron insertion from a Rieske-type [2Fe–2S] centre. Either the reductase or, where present, the intermediary ferredoxin component, supplies the two electrons from NAD(P)H to the dioxygenase **[10].** RDOs have been empirically classified according to the various combinations of subunits and electron transfer co-factors involved in reducing the oxygenase component **[10, 11]** as mentioned in **Table 1 (see supplementary material)**.

Here we present a protocol to data mine and computationally characterize redox hypothetical proteins possessing multiple domains. Most proteins consist of multiple domains, and domains determine the function and evolutionary relationships of proteins **[12].** Thus, it is important to understand the principles of domain combinations and their associated inter domain interactions especially, in hypothetical proteins.

Primarily, 2Fe-2S (Ferredoxins) and FMN/FAD (Flavodoxins) were considered due to their vital and diverse roles in biological systems, the most important amongst it being their role in Electron Transport Mechanisms. Ferredoxins are small, acidic, electron transfer proteins that are ubiquitous in biological redox systems. Members of the 2Fe-2S ferredoxin family have a general core structure consisting of beta(2)-alpha-beta(2). They are proteins of around one hundred amino acids with four conserved cysteine residues to which the 2Fe-2S cluster is ligated **[13].** Flavoenzymes have the ability to catalyse a wide range of biochemical reactions. They are involved in the dehydrogenation of a variety of metabolites, in electron transfer from and to redox centres, in light emission, in the activation of oxygen for oxidation and hydroxylation reactions. About 1% of all eukaryotic and prokaryotic proteins are predicted to encode a flavin adenine dinucleotide (FAD) or Flavin Mono Nucleotide (FMN)-binding domains which are involved in electron transport **[14].**

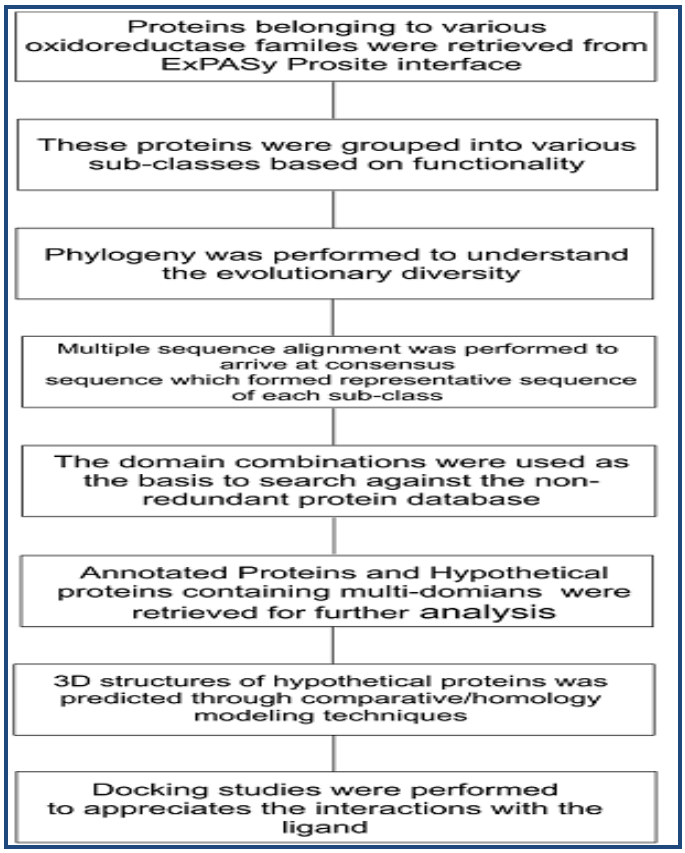

Figure 2: Protocol

Methodology:

The proteins belonging to oxidoreductase (Ferredoxin, Flavodoxin) families were retrieved from the ExPASy Prosite interface **[15].** However, engineered and mutated sequences were not considered to avoid redundancy. Additionally, only reviewed sequence from Uniprot containing a structural entry were considered. Binding sites of all the proteins belonging to the same group were analyzed in order to arrive at a consensus pattern through multiple sequence alignment. Extended regions which had no information with the other sequences were clipped to strengthen the alignment. The protocol adopted is shown in **(Figure 2)**.

The search for Ferredoxin family (PDOC00175) yielded 14 sequences with 2Fe-2S binding signature. As there existed heterogeneity within the group, the sequences were clustered based on phylogenetic analysis. The sequence alignment was performed through ClustalW **[16]** and the tree was obtained using MEGA (NJ method) **[17].** The tree obtained is shown in (**Figure 3A)**. Further to the clustering, multiple sequence alignment was performed using Multalin **[18],** for all the 3 clusters (groups) to obtain a representative sequence containing strong signatures. The multiple sequence alignment of sequences belonging to group 1 yielded better consensus

compared to the other clusters, which is as depicted in (**Figure 3B)**.

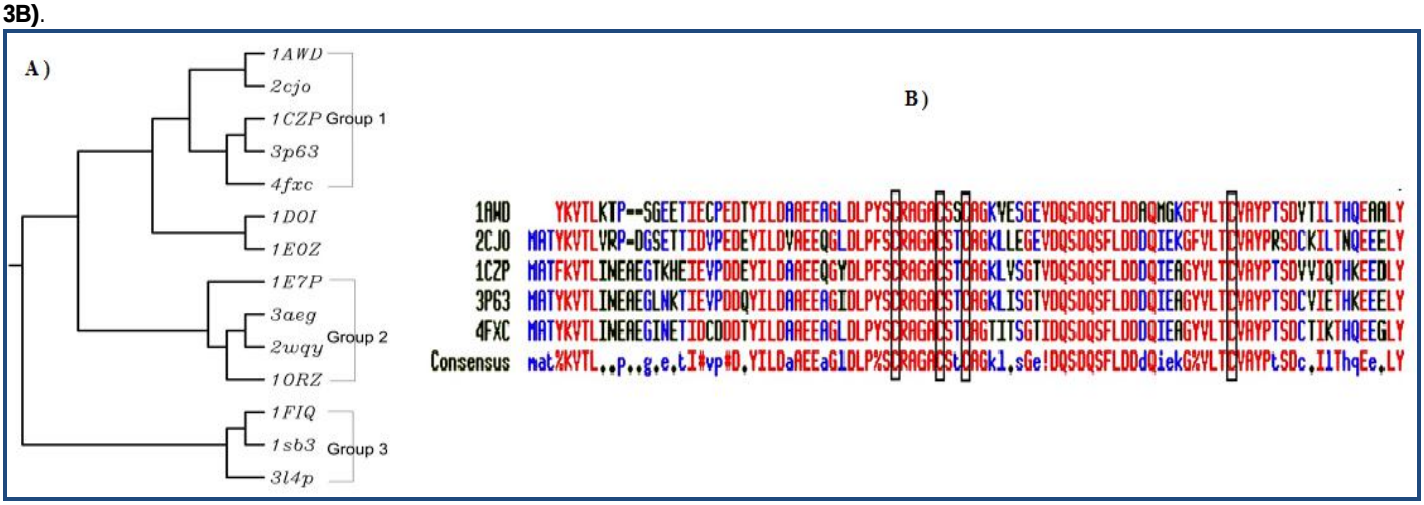

Figure 3: A) Phylogenetic tree of 2Fe-2S family; **B)** MSA of group 1 of 2Fe-2S family

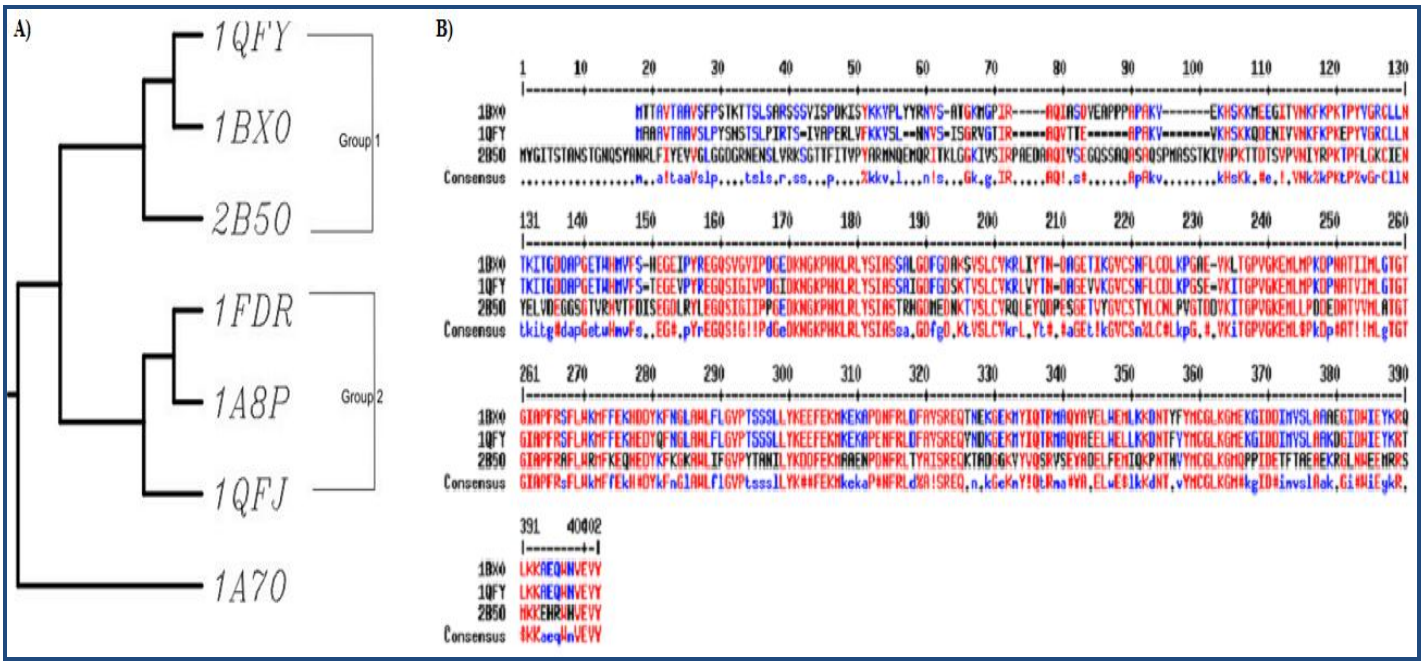

Figure 4: A) Phylogenetic tree for FNR reductase family; **B)** MSA of group 1 of FNR reductase family

Similarly, the search for flavodoxin family (FNR reductase - PDOC51384) yielded 7 sequences, whose Phylogenetic tree is shown in **(Figure 4A)**. When multiple sequence alignments of both the clusters were critically analyzed, the MSA of group 1 exhibited strong signatures of the FNR domain when compared to cluster 2, which is depicted in **(Figure 4B).**

Thus, a consensus of the cluster of sequences from group 1, for both the 2Fe-2S and FNR domains respectively, were considered as possible representative patterns, towards generating the probable synthetic sequence, which was used as the basis for BLAST tool **[19]** search against the non-redundant database. Interestingly, this approach yielded 2078 sequences, and clearly contained both 2Fe-2S and FNR domains when analysed through the conserved domain database (CDD) **[20].** Amongst these 2078 sequences, 129 belonged to that special class of hypothetical proteins, which were taken up for further characterization and analysis.

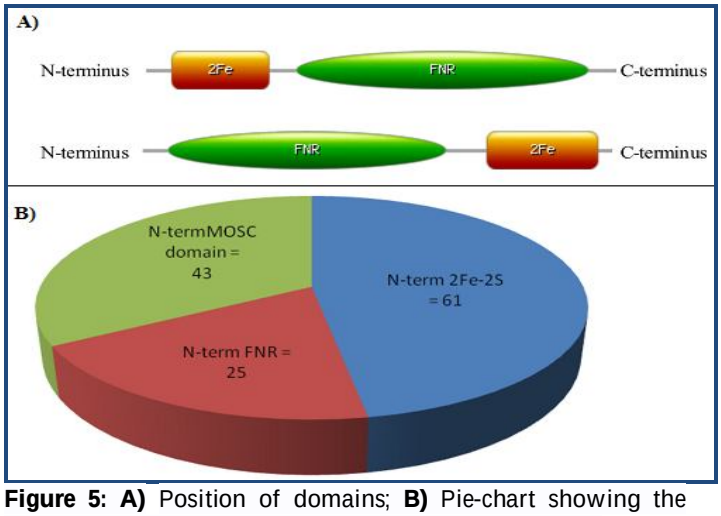

Figure 5: A) Position of domains; **B)** Pie-chart showing the distribution of domains in the 129 hypothetical proteins.

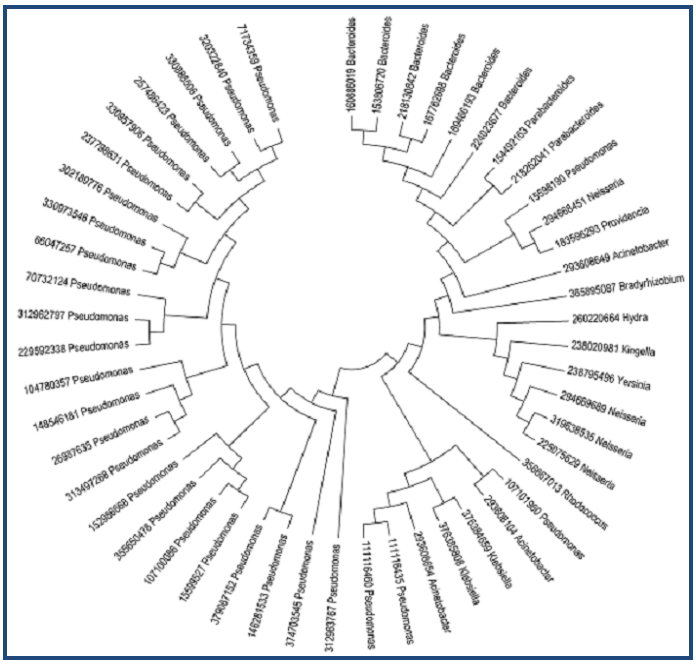

Figure 6: Phylogenetic tree of the hypothetical proteins containing N-terminus 2Fe-2S and C-terminus FNR domain.

Results and Discussion:

Upon critical evaluation of the 129 multi-domain hypothetical sequences through CDD, significant differences in the location of 2Fe-2S domain, relative to other domains, were found. Of these 129 sequences, 61 contained an N-terminus 2Fe-2S and a C-terminus FNR domain while this order was reversed in 25 sequences as shown in the **(Figure 5A).**

The remaining 43 sequences contained an N-terminus MOSC domain **[21]** (pfam03473 and pfam03476) which is a super family of beta-strand-rich domains identified in the molybdenum cofactor sulfurase and several other proteins from both prokaryotes and eukaryotes. The MOSC domain is predicted to be a sulfur-carrier domain that receives sulfur abstracted by the pyridoxal phosphate-dependent NifS-like enzymes, on its conserved cysteine, and delivers it for the formation of diverse sulfur-metal clusters. The pie chart in **(Figure 5B)** illustrates the distribution of the domains amongst these 129 proteins.

In the current study, 61 sequences containing N-terminus 2Fe-2S and C-terminus FNR domains are only considered. The remaining sequences, 25 of which contain an N-terminus FNR and C-terminus 2Fe-2S domains, and 43 of them containing MOSC domain will be considered for detailed analysis in near future. The phylogenetic analysis of the 61 sequences containing an N-terminus 2Fe-2S and C-terminus FNR domains is depicted in **(Figure 6)**, which exhibits the domination of the genus *Pseudomonas* (46% amongst 61 sequences).

The sequences were searched against the PDB database (using the PDB BLAST tool) towards identification of a suitable template. This yielded 1KRH (which has 338 amino acid residues), a benzoate dioxygenase (BenC), from *Acinetobacter* sp. strain ADP1 at 1.5A resolution. BenC contains an iron–sulphur and a FAD domain **[10].** The [2Fe–2S] domain is similar to that of plant ferredoxins, and the FAD and NADH domains are similar to members of the ferredoxin:NADPH reductase superfamily.

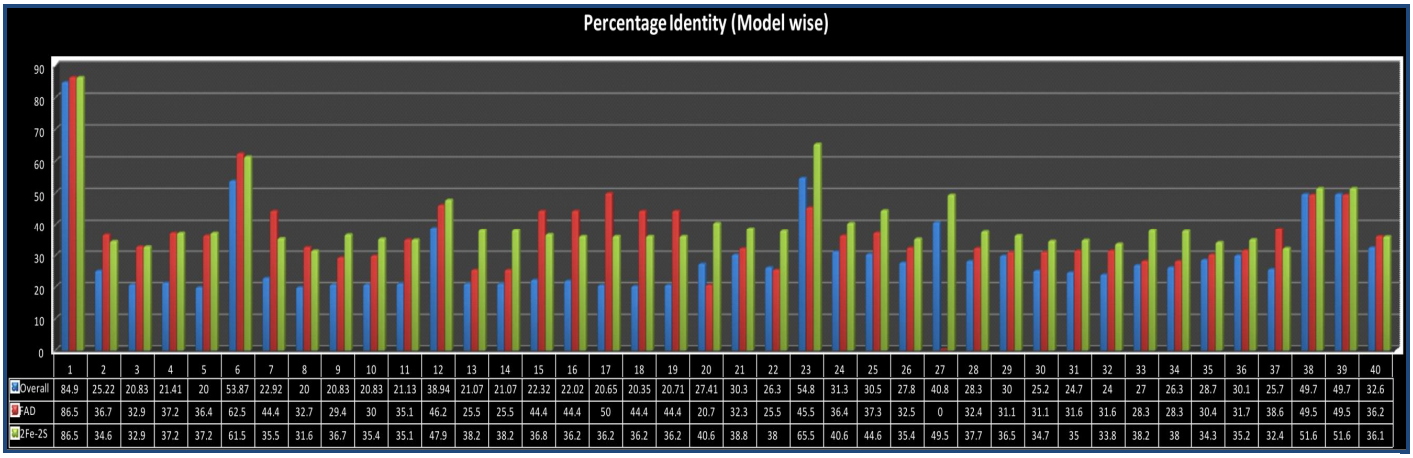

Figure 7: Bar graph showing the overall sequence identity (blue), identity at FAD binding region (red) and 2Fe-2S binding region (green) against the model 1KRH (Please see table 3 for cross-reference).

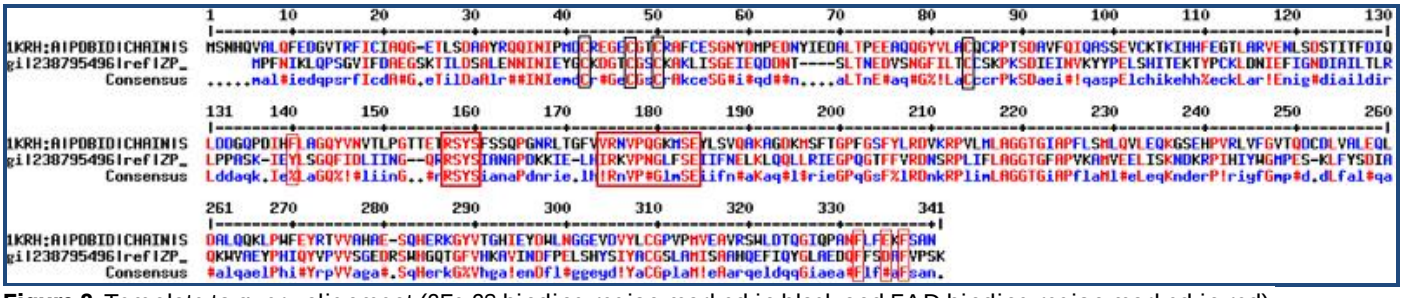

Figure 8: Template to query alignment (2Fe-2S binding region marked in black and FAD binding region marked in red).

Of these 61 sequences, 21 sequences had very low (<20%) sequence identity with the template 1KRH, and hence were discarded from further analysis due to lack of clarity. The remaining 40 sequences were considered with confidence for homology modelling exercises, as they exhibited high similarity with 1KRH. The overall sequence identity between the query and template was between 20-30% for most sequences, except 7 of them which was in the higher range of 40-70%. However, in spite of lower overall identities will the template, the appreciation with the patterns at domain regions was indeed revealing. The **(Figure 7)** shows the distribution of the overall sequence identity, identities at the FAD and 2Fe-2S binding regions for each sequence, which clearly illustrates the conservation at critical regions of functional relevance.

The FNR family contains two conserved motifs, viz., (R-x-Y-[ST]) where positively charged Arg residue forms hydrogen bonds to the pyrophosphate oxygen atom and (G-x(2)-[ST]-x(2)-L-x(5)-G-x(7)-P-x-G) which is the phosphate-binding motif **[14]**. Similarly, 4 conserved Cys residues at positions i, i+5, i+8 and variable i+38 is required for binding of 2Fe-2S ligand **[13]**. Both the FAD and 2Fe-2S binding regions are highly conserved in all the 40 models.

In view of the poise in the signatures between the template and the 40 target sequences, model building exercises were carried out with Swissmodel automated mode **[22]**. The RMSD between the modelled structure and template for the Ca- atoms confirmed the quality of the models in spite of seemingly low sequence identity (refer table 3 and figure 10), in addition to the satisfaction of various criteria calculated using ProCheck **[23]**. Individual models were analysed for the binding of ligands through docking studies which was performed using FlexX algorithm **[24]**. As a case study, modelling of GI ID 238795496 is illustrated below, to define the quality of the structural and functional aspects of these hypothetical protein sequences.

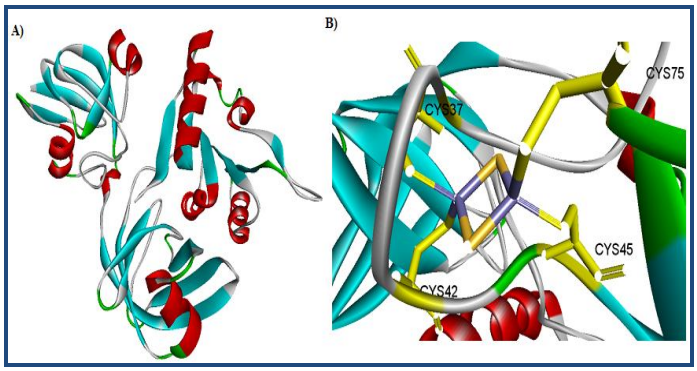

Figure 9: A) Modelled structure of GI : 238795496; **B)** 2Fe-2S interaction with Cysteins (C37, C42, C45 and C75) in the model

1KRH based model for the GI ID 238795496

The query protein 238795496 from *Yersinia mollaretii* ATCC 43969 was successfully modelled using SWISS model interface, where the overall identity between the query and template is 25.22%. The alignment between the template and query is shown in **(Figure 8)**. In spite of the low overall sequence identity, it can be appreciated that the binding regions of 2Fe-2S and FAD exhibit conservation up to 35%.

The RMSD for C-alpha atoms between the modelled structure and template is found to be 0.53 Å (for 93.2% of the atoms superposed). The quality of the model was assessed with PROCHECK (Ramachandran map) which showed that 96.8% of the residues were in allowed regions and only 3.2% non-critical residues in disallowed regions. Interestingly, none of these residues in the outlier regions belong to the functionally important regions of the model. The 2Fe-2S and FAD ligands were docked into the model and all the interactions were found similar to that of the template. The binding of 2Fe-2S Ligand and FAD are shown in **(Figures 9A, B and Figure 10A, B, C).**

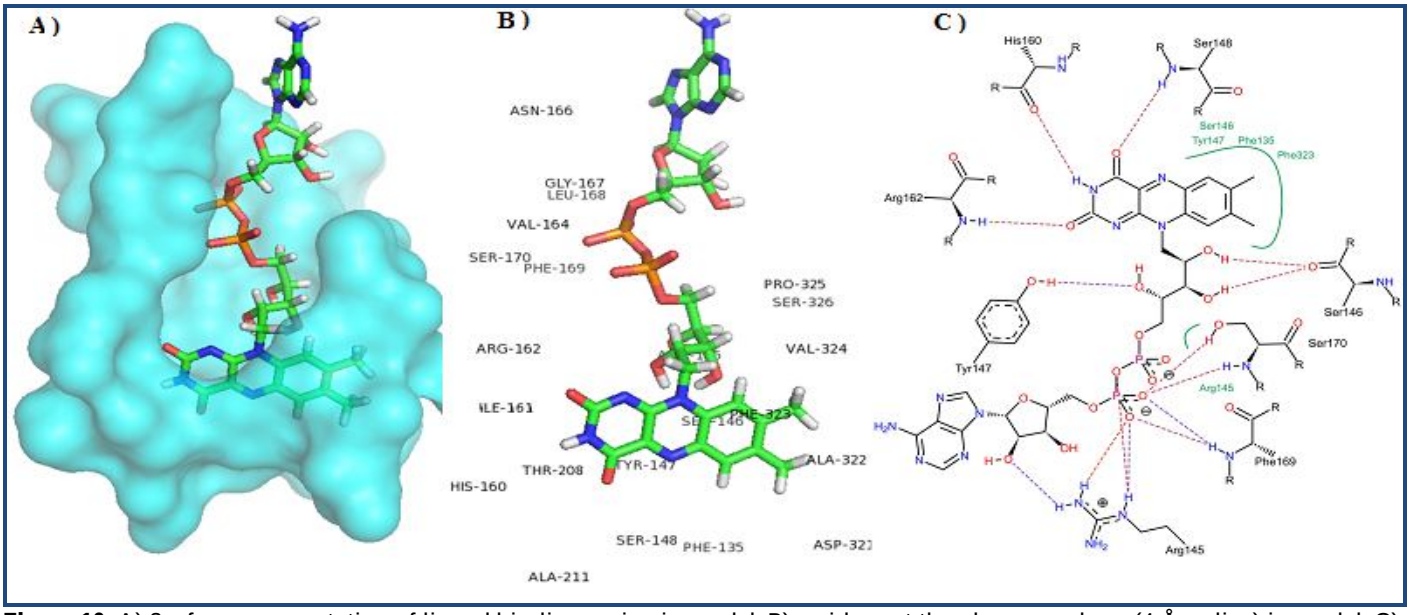

Figure 10: A) Surface representation of ligand binding region in model; B) residues at the pharmacophore (4 Å radius) in model; C) 2D representation of the ligand-residue interaction in model.

Table 2 (see supplementary material) summarizes the residues forming the Pharmacophore (within 4 Å radius) for FAD ligand in template, FAD ligand redocked to template and the model, where good conservation is observed. The docking of the FAD

to the template (using the program FlexX) was done to re-confirm the ligand binding pose, and normalize the artefacts due to the software, if any. The residues highlighted in bold forms H-bonds with the FAD, which further reiterates decent bind of the ligand.

The modelled and docked structures were deposited at the Protein Model Data Bank (PMDB) **[25]** where all the 40 models were judged to possess clashes within acceptable limits. **Table 3 (see supplementary material)** summarises the details of all the models generated with IKRH (which contains 338aa) as the template.

Conclusion:

129 hypothetical proteins from across the prokaryotic genomes have been data mined, and the 3D description of 40 sequences has been derived with confidence. The statistics related to comparative modelling and docking studies (with acceptable energy values) have revealed a strong interaction of redox ligands, viz., 2Fe-2S and FAD, which further strengthens the argument that these proteins may be involved in cleavage of aromatic compounds.

Though degradation of aromatic compounds by *Pseudomonas* is a well established fact **[26, 27]**, characterization of hypothetical sequences from *Pseudomonas* in the present study could aid in better understanding of these microbial systems. Additionally, large number of other bacterial systems containing these dioxygenases have also been mined and characterized in the present investigations, which could provide insights into their degradation properties.

Thus, this study on multi-domain hypothetical proteins could prove critical in two ways viz., in understanding the mechanism of uptake of nutrients which contain aromatic ring structures and hence enabling engineering of these proteins towards effective degradation of harmful xenobiotics.

Acknowledgment:

The authors would like to thank R Sowdhamini and Jagadeesh Kumar D for the technical discussions and Sir MVIT for the infrastructural support. NS would specifically like to thank Srignanakshi, Ojal Sarah Paul and Harshal Soni for the assistance provided during the initial stages of the work.

References:

[1] Galperin MY & Koonin EV, *Nucleic acids res.* 2004 **32**: 5452 [PMID: 15479782]
[2] Bork P, *Genome Res.* 2000 **10:** 398 [PMID: 10779480]
[3] Bentley SD *et al*. *Nature.* 2002 **417:** 141[PMID: 12000953]
[4] Kaneko T *et al*. *DNA Res.* 2001 **8**: 205 [PMID: 11759840]
[5] Galperin MY, *Compa Funct Genomics.* 2001 **2**: 14 [PMID: 18628897]
[6] Galperin MY & Koonin EV, *Nucleic Acids Res.* 2004 **32**: 5452 [PMID: 15479782]
[7] Gibson & Harwood CS, *Annu Rev Microbiol*. 2002 **56**: 345 [PMID: 12142480]
[8] Bertini I *et al*. *Coordination Chemistry Reviews*. 1996 **151**: 145
[9] Butler CS & Mason JR, Adv *Microbial Physiology*. 1996 **38:** 47 [PMID: 8922118]
[10] Karlsson A *et al*. *Journal of molecular biology*. 2002 **318**: 26 [PMID:12051836]
[11] Correll CC *et al Science*. 1992 **258:** 1604 [PMID: 1280857]
[12] Vogel C *et al*. *Curr Opin struct Biol*. 2004 **14:** 208 [PMID: 15093836]
[13] Valentine RC, *Bacteriological reviews.* 1964 **28:** 497
[14] Joosten V & WJ Van Berkel, *Curr Opin Chem Biol*. 2007 **11:** 195 [PMID: 17275397]
[15] Bairoch A, *Nucleic Acids Res*. 1993 **21:** 3097 [PMID: 8332530]
[16] Thompson JD *et al*. *Nucleic Acids Res*. 1994 **22**: 4673 [PMID: 7984417]
[17] Kumar S *et al. Brief Bioinform.* 2008 **9:** 299 [PMID: 18417537]
[18] Corpet F, *Nucleic Acids Res.* 1988 **16**: 10881 [PMID: 2849754]
[19] Altschul SF *et al*. *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
[20] Marchler-Bauer A, *Nucleic Acids Res*. 2011 **39**: D225 [PMID: 21109532]
[21] Anantharaman V & Aravind L, *FEMS Microbiol Lett*. 2002 **207**: 55 [PMID: 11886751]
[22] Schwede T *et al. Nucleic Acids Res.* 2003 **31**: 3381 [PMID: 12824332]
[23] Laskowski RA *et al. Journal of Applied Crystallography.* 1993 **26**: 283
[24] Kramer B *et al. Proteins.* 1999 **37**: 228 [PMID: 10584068]
[25] Castrignano T, *Nucleic Acids Res*. 2006 **34:** D306 [PMID: 16381873]
[26] Gibson DT *et al. Biochemistry*. 1970 **9:** 1631
[27] Cowles CE *et al. J Bacteriol.* 2000 **182:** 6339 [PMID: 11053377]

Edited by P Kangueane

Citation: Nitish & Nagendra, Bioinformation 8(23): 1154-1161 (2012)

BIOINFORMATION

open access

Supplementary material:

Table 1: Components of Rieske dioxygenases.

Class	Reductase	Intermediate electron transfer component	Oxygenase	Examples
IA	FMN Cys4[2Fe–2S]	None	Cys2His2 [2Fe–2S] e^{2+}	Phthalate dioxygenase
IB	FAD Cys4[2Fe–2S]	None	Cys2His2 [2Fe–2S] e^{2+}	Benzoate dioxygenase
IIA	FAD	Cys_4[2Fe–2S]	Cys2His2 [2Fe–2S] e^{2+}	Dibenzofuran 4,4a-dioxygenase
IIB	FAD	Cys_2His_2[2Fe–2S]	Cys2His2 [2Fe–2S] e^{2+}	Biphenyl dioxygenase
III	FAD Cys4[2Fe–2S]	Cys_2His_2[2Fe–2S]	Cys2His2 [2Fe–2S] e^{2+}	Naphthalene dioxygenase

Table 2: Residues forming the Pharmacophore (with in 4 Å radius) of FAD

Sl. No.	Residues in 1KRH	Residues in FAD re-docked structure	Residues in the model of GI : 238795496
1	Y144	Y144	F135
2	**R156**	**R156**	**R145**
3	**S157**	**S157**	**S146**
4	Y158	**Y158**	**Y147**
5	**S159**	S159	**S148**
6	**V172**	**V172**	**H160**
7	V173	V173	I161
8	**R174**	**R174**	**R162**
9	V176	V176	V164
10	Q178	Q178	N166
11	G179	G179	G167
12	**K180**	**K180**	L168
13	**M181**	**M181**	**F169**
14	**S182**	**S182**	**S170**
15	T220	T220	T208
16	A223	A223	A211
17	E333	E333	D321
18	K334	K334	A322
19	**F335**	F335	F323
20	S336	S336	V324
21	A337	A337	P325
22	N338	**N338**	S326

Table 3: Summary of 40 models

Sl.No	Multi domain hypo protein	Species	Number of amino acids	RMSD Å for Ca Atoms [% of atoms superposed]	PMDB id
1	293608654	Acinetobacter sp. SH024	338	0.06 [100%]	PM0077745
2	238795496	Yersinia mollaretii ATCC 43969	330	0.53 [93.2%]	PM0078394
3	229592338	Pseudomonas fluorescens SBW25	310	0.51 [81.6%]	PM0078546
4	148546181	Pseudomonas putida F1	306	0.51 [89.5%]	PM0078547
5	313497268	Pseudomonas putida F1	306	0.62 [86.2%]	PM0078655
6	111116435	Pseudomonas putida	336	0.21 [98.5%]	PM0078548
7	70732124	Pseudomonas fluorescens Pf-5	312	0.50 [83.9%]	PM0078549
8	26987635	Pseudomonas putida KT2440	306	0.51 [86.2%]	PM0078551
9	15599527	Pseudomonas aeruginosa PAO1	308	0.56 [81.4%]	PM0078553
10	152986668	Pseudomonas aeruginosa PA7	309	0.56 [81.2%]	PM0078554
11	312963767	Pseudomonas fluorescens WH6	322	0.58 [90.9%]	PM0078556
12	107101960	Pseudomonas aeruginosa PACS2	340	0.46 [97.6%]	PM0078656
13	146281533	Pseudomonas stutzeri A1501	344	0.39 [75.8%]	PM0078557
14	66047257	Pseudomonas syringae pv. syringae B728a	312	0.43 [82.3%]	PM0078657
15	302189776	Pseudomonas syringae pv. syringae 642	312	0.50 [82.6%]	PM0078658
16	257486423	Pseudomonas syringae pv. tabaci ATCC 11528	312	0.45 [83.9%]	PM0078559
17	237798631	Pseudomonas syringae pv. oryzae str. 1_6	312	0.50 [85.8%]	PM0078560
18	320322840	Pseudomonas syringae pv. Glycinea	312	0.54 [85.2%]	PM0078561
19	71734359	Pseudomonas syringae pv. phaseolicola 1448A	312	0.49 [85.8%]	PM0078661
20	319638535	Neisseria mucosa C102	334	0.41 [93.1%]	PM0078662
21	104780357	Pseudomonas entomophila L48	306	0.45 [81.5%]	PM0078562
22	107100086	Pseudomonas aeruginosa PACS2	308	0.46 [81.5%]	PM0078563
23	111116460	Pseudomonas putida	336	0.21 [98.5%]	PM0078564
24	225075629	Neisseria flavescens NRL30031 / H210	362	0.41 [85.3%]	PM0078565
25	238020981	Kingella oralis ATCC 51147	340	0.29 [88.5%]	PM0078663

226	260220664	Curvibacter putative symbiont of Hydra magnipapillata	341	0.26 [92.9%]	PM0078566
27	293608104	Acinetobacter sp. SH024	344	0.18 [99.1%]	PM0078568
28	293608649	Acinetobacter sp. SH024	353	0.31 [88.6%]	PM0078664
29	294669669	Neisseria elongata subsp. glycolytica ATCC 29315	336	0.32 [91.9%]	PM0078571
30	312962797	Pseudomonas fluorescens WH6	310	0.40 [82.6%]	PM0078572
31	330888506	Pseudomonas syringae pv. mori str. 301020	312	0.44 [82.1%]	PM0078573
32	330957906	Pseudomonas syringae pv. maculicola str. ES4326	312	0.37 [84.2%]	PM0078575
33	330973546	Pseudomonas syringae pv. aceris str. M302273	312	0.37 [83.3%]	PM0078665
34	355650478	Pseudomonas sp. 2_1_26	308	0.44 [81.8%]	PM0078578
35	356667013	Rhodococcus opacus PD630	332	0.36 [92.6%]	PM0078585
36	365895087	Bradyrhizobium sp. STM 3843	346	0.35 [89.0%]	PM0078586
37	374703545	Pseudomonas sp. S9	312	0.45 [83.9%]	PM0078587
38	376384659	Klebsiella oxytoca 10-5243	338	0.21 [96.7%]	PM0078666
39	376385808	Klebsiella oxytoca 10-5245	338	0.21 [96.7%]	PM0078667
40	379067152	Pseudomonas stutzeri ATCC 14405	291	0.45 [87.9%]	PM0078592