FEBS 26013 FEBS Letters 518 (2002) 43–47

Identification of a Baeyer–Villiger monooxygenase sequence motif

Marco W. Fraaije[a,*], Nanne M. Kamerbeek[a], Willem J.H. van Berkel[b], Dick B. Janssen[a]

[a] *Laboratory of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands*

[b] *Laboratory of Biochemistry, Department of Agrotechnology and Food Sciences, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands*

Received 11 February 2002; revised 18 March 2002; accepted 21 March 2002

First published online 4 April 2002

Edited by Robert B. Russell

Abstract **Baeyer–Villiger monooxygenases (BVMOs) form a distinct class of flavoproteins that catalyze the insertion of an oxygen atom in a C–C bond using dioxygen and NAD(P)H. Using newly characterized BVMO sequences, we have uncovered a BVMO-identifying sequence motif: FXGXXXHXXXW(P/D). Studies with site-directed mutants of 4-hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens* ACB suggest that this fingerprint sequence is critically involved in catalysis. Further sequence analysis showed that the BVMOs belong to a novel superfamily that comprises three known classes of FAD-dependent monooxygenases: the so-called flavin-containing monooxygenases (FMOs), the *N*-hydroxylating monooxygenases (NMOs), and the BVMOs. Interestingly, FMOs contain an almost identical sequence motif when compared to the BVMO sequences: FXGXXXHXXX(Y/F). Using these novel amino acid sequence fingerprints, BVMOs and FMOs can be readily identified in the protein sequence databank.**

Key words: Baeyer–Villiger monooxygenase; Flavin; Homology; Sequence motif

1. Introduction

Baeyer–Villiger monooxygenases (BVMOs) are flavoenzymes that are able to catalyze a Baeyer–Villiger reaction; using dioxygen and NAD(P)H a ketone function is converted into the corresponding ester (Scheme 1) [1].

Cyclohexanone monooxygenase from *Acinetobacter* NCIMB 9871 is the best characterized BVMO [2,3]. Because of its wide substrate spectrum and stereoselectivity this enzyme has been extensively used for synthetic applications [1,4]. Like other flavoprotein monooxygenases, cyclohexanone monooxygenase forms a hydroperoxyflavin intermediate which is involved in substrate oxygenation [5,6].

Until a few years ago, cyclohexanone monooxygenase was the only BVMO of which the sequence was known. Recently, several other bacterial BVMOs were cloned and sequenced: 4-hydroxyacetophenone monooxygenase [7], steroid monooxygenase [8], cyclododecanone monooxygenase [9] and two other cyclohexanone monooxygenases [10]. Using these data we performed a sequence alignment study which uncovered a conserved sequence motif. The function of this BVMO-specific sequence motif was explored by site-directed mutagenesis of 4-hydroxyacetophenone monooxygenase. Moreover, our search for the presence of conserved sequence motifs also resulted in the disclosure of a superfamily of sequence-related FAD-dependent monooxygenases.

2. Materials and methods

2.1. Materials

Chemicals were purchased from Acros Chimica, Merck, Aldrich, or Sigma. Molecular biology enzymes were purchased from Roche. Oligonucleotides were supplied by Eurosequence BV, Groningen, The Netherlands.

2.2. Site-directed mutagenesis

The Quickchange site-directed mutagenesis kit from Stratagene was used to introduce mutations into the 4-hydroxyacetophenone monooxygenase gene. Successful mutagenesis was confirmed by plasmid sequencing. Mutant proteins were expressed and purified as described for wild type enzyme [7].

2.3. Analytical methods

4-Hydroxyacetophenone monooxygenase activity was determined as described previously [7]. The relative molecular mass of native 4-hydroxyacetophenone monooxygenase was determined by FPLC gel filtration using a Superdex 200 HR 10/30 column (Pharmacia Biotech) running with 50 mM potassium phosphate buffer, pH 7.5. Absorption spectra were recorded at 25°C on a Perkin Elmer Lambda Bio 40 spectrophotometer. Circular dichroism (CD) spectra were recorded at 25°C from 190 to 250 nm using a 0.1-cm cuvette containing 0.1 mg/ml of enzyme (5 mM potassium phosphate, pH 7.5), and an AVIV 62A DS spectrometer.

2.4. Sequence homology analysis

The BLAST programs at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/BLAST/) were used to search for proteins showing sequence similarity. Multiple sequence alignments were made with the ClustalW program at the Centre for Molecular and Biomolecular Informatics (www.cmbi.kun.nl/bioinf/tools). Sequence alignments were visualized using the BOXSHADE 3.21 program (http://www.ch.embnet.org/software/BOX_form.html). TreeView 1.5.2 was used for generating a tree representation using the ClustalW output.

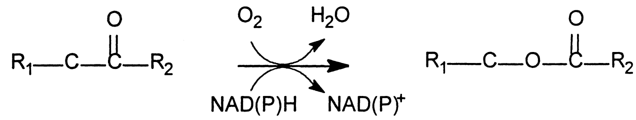
3. Results and discussion

3.1. The BVMO family

A non-redundant database search at the NCBI using the protein sequence of 4-hydroxyacetophenone monooxygenase and the BLASTP program [11] yielded 35 protein sequences with significant similarity ($S > 100$, $E < 1.10^{-20}$). Besides the above-mentioned bacterial BVMOs (sequence identities of 27–33%), this set of homologs contains two fungal gene products that have been shown to be involved in biosynthetic pathways

*Corresponding author. Fax: (31)-50-3634165.
E-mail address: m.w.fraaije@chem.rug.nl (M.W. Fraaije).

0014-5793/02/$22.00
PII: S0014-5793(02)02623-6



Scheme 1. Baeyer–Villiger reactions catalyzed by flavin-containing BVMOs.

of mycotoxins which include Baeyer–Villiger oxidation steps [12,13]. All BVMO homologs found are of bacterial or fungal origin, indicating that BVMOs play a role in a variety of microbial oxidative metabolic pathways. Interestingly, a relatively large number of BVMO homologs was found in pathogenic bacteria (e.g. seven in *Mycobacterium tuberculosis* H37Rv). In addition, by performing a BLAST search in the NCBI database of patented nucleotide sequences we found that a DNA probe (accession nr. L04542) used for specific detection of pathogenic *Mycobacterium avium* isolates [14] encodes a major part of a putative BVMO (35% sequence identity). This indicates that these oxidative enzymes represent attractive targets for drug development. In fact, one of the seven BVMO-related genes from *M. tuberculosis*, was recently shown to be responsible for the activation and therefore the efficacy of the widely used anti-tuberculosis pro-drug ethionamide [15,16]. It has been shown that the *etaA* gene product mediates ethionamide activation by sulfoxidation; a reaction typically catalyzed by BVMOs.

Alignment of the BVMO homologs revealed several conserved regions containing sequence motifs that are known to be involved in dinucleotide cofactor binding [17]. The conserved sequence motifs can also be clearly recognized from the alignment of all above-mentioned biochemically identified BVMOs, as shown in Fig. 1. The presence of two Rossmann folds, as evidenced by two GXGXX(G/A) motifs, clearly discriminates these enzymes from the mechanistically related flavoprotein hydroxylases [18]. Furthermore, we noticed a stretch of conserved residues just before the second Rossmann fold motif. With two exceptions, this FXGXXXHXXXW-(P/D) motif is strictly conserved in all BVMO homologs and represents, apart from the Rossmann fold motif regions, the sequence region with the highest number of conserved residues. In the two aberrant protein sequences from *Caulobacter crescentus* the central histidine is not conserved.

To probe the specificity of the BVMO sequence motif we also used the motif as a seed for a Pattern Hit Initiated BLAST search [19]. By this approach, the above-mentioned sequences could be specifically retrieved. Furthermore, a trawl of the PEDANT genomic sequence database (http://pedant.gsf.de/) using only the BVMO sequence motif as seed for a pattern search resulted in the specific identification of 32 microbial BVMO homologs. These pattern-based searches exemplify the potential of the described sequence motif for efficient annotation of putative BVMO genes.

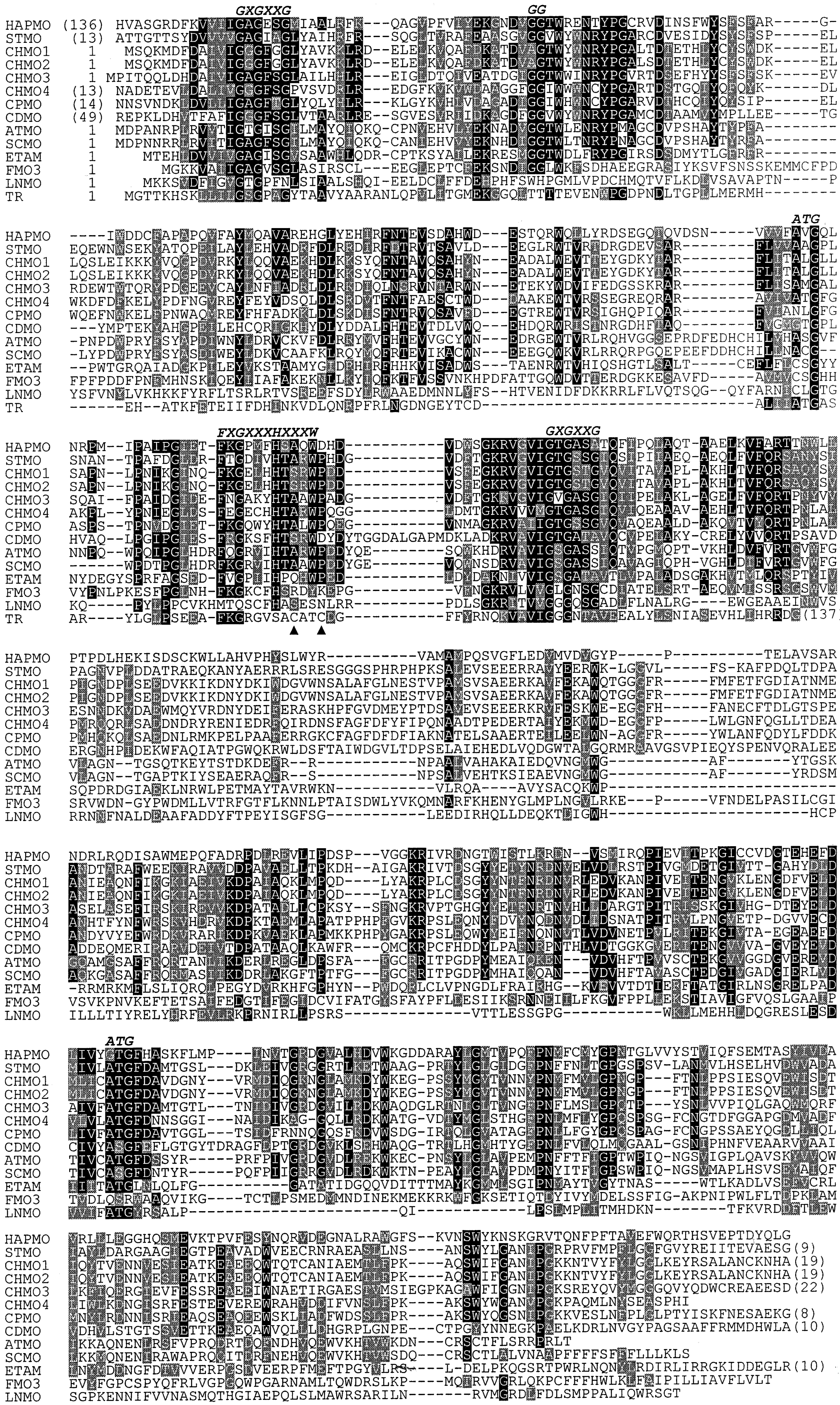
3.2. *BVMOs are sequence-related to flavin-containing monooxygenases (FMOs) and N-hydroxylating monooxygenases (NMOs)*

The sequential organization of the dinucleotide binding sequence motifs observed in known BVMO sequences is reminiscent of two other flavoprotein monooxygenase families: multi-functional FMOs (eukaryotic enzymes oxygenating a broad range of amines and other heteroatom containing compounds) [20] and NMOs (microbial enzymes catalyzing *N*-hydroxylations of long-chain primary amines) [21]. A search in the non-redundant database of NCBI using the PSI-BLAST program [22] confirmed sequence homology between BVMOs, FMOs and NMOs, revealing the presence of more than 150 sequences belonging to one of these three flavoprotein families. Two representative members of the FMO and NMO families have been included in the multiple sequence alignment depicted in Fig. 1. The observed sequence similarity between members of the three different flavoprotein families is reflected in their structural and catalytic properties: all characterized members are single-component FAD-containing monooxygenases of about 50 kDa which specifically use NADPH as electron donor. A dendrogram of all homologous sequences gives insight in the sequence distance relationship within the discovered flavoprotein superfamily nicely illustrating the clustering of sequences in the three distinct monooxygenase families (Fig. 2).

From the three monooxygenase families, members of the FMO family have been most extensively studied as they play an essential role in the detoxification of xenobiotics (e.g. nicotine, drugs) in vertebrates [20]. Vertebrates typically contain five FMO isoforms. Several naturally occurring mutations in human FMO3 have been shown to underlie inheritable trimethylaminuria (also called 'fish-odor syndrome') illustrating the important role of FMOs in metabolic detoxification [23]. Except for the known mammalian FMOs, our sequence similarity search revealed a surprising number of FMO homologs in plants (26 in the *Arabidopsis thaliana* genome). Recently, the function of three of these homologs (YUCCA 1–3) has been identified, confirming a FMO-type reactivity as these enzymes catalyze an *N*-oxygenation of tryptamine yielding a precursor of the plant hormone auxin [24]. Interestingly, in all FMO sequences a similar sequence motif as found for BVMOs can be identified, FXGXXXHXXX(Y/F), differing by only one residue (Fig. 1). Only in the FMO homologs from yeast this sequence signature is not fully conserved. In this respect it is interesting to note that yeast FMO has lost its ability to catalyze typical FMO reactions [25].

→

Fig. 1. Multiple sequence alignment of several members of the flavoprotein monooxygenase superfamily which includes several BVMOs (HAPMO: 4-hydroxyacetophenone monooxygenase from *P. fluorescens* ACB, AF355751; STMO: steroid monooxygenase from *Rhodococcus rhodochrous*, BAA24454; CHMO1–4: several bacterial cyclohexanone monooxygenases, P12015, AAG10021, AAG01289, AAG01290; CDMO: cyclododecanone monooxygenase from *R. ruber*, AAL14233; CPMO: cyclopentanone monooxygenase from *Comamonas testosteroni*, CAD10798; ATMO: aflatoxin biosynthesis monooxygenase from *Aspergillus parasiticus*, AAF26281; SCMO: sterigmatocystin biosynthesis monooxygenase from *Aspergillus nidulans*, Q00730; ETAM: ethionamide-activating monooxygenase from *M. tuberculosis*, CAB06212), a FMO (FMO3 from human, P31513), and an NMO (LNMO: L-lysine N^6-hydroxylase from *E. coli*, P11295). The sequence of thioredoxin reductase (TR) from *E. coli* (66129) is included in the alignment (active-site cysteines indicated by triangles). Common flavoprotein motifs [17] and the newly identified BVMO sequence motif are indicated in the top line. For proper alignment some residues are not shown (in parentheses).


GXGXXG
GG
ATG
FXGXXXHXXXW
GXGXXG
ATG

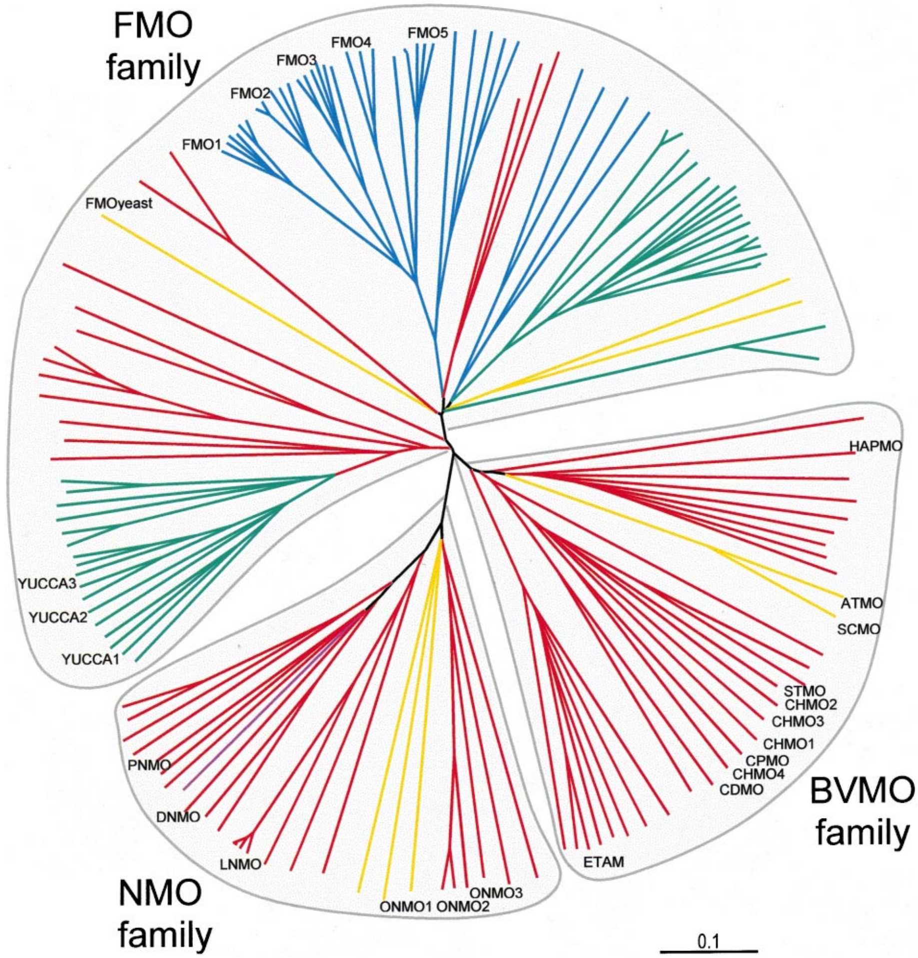


Fig. 2. An unrooted phylogenetic tree of the flavoprotein monooxygenase superfamily. Sequences retrieved from the NCBI protein sequence database using PSI-BLAST were used (January 21, 2002). A bar showing 10% divergence is included. In addition to the monooxygenases that are discussed in the text some other annotated sequences are indicated (PNMO: putrescine hydroxylase from *Bordetella bronchiseptica* Q44740 [29]; DNMO: N^4-diaminopropane monooxygenase from *Sinorhizobium meliloti*, AAK65920 [30]; ONMO1: L-ornithine N^5-monooxygenase from *Ustilago maydis*, P56584 [31]; ONMO2: L-ornithine N^5-monooxygenase from *P. aeruginosa* (Q51548) [32]; ONMO3: L-ornithine N^5-monooxygenase from *Pseudomonas* strain B10 (AAG27518) [26]). The clustering of the three monooxygenase families is indicated by gray shading. The origin of each sequence is indicated by color: animal (blue), plant (green), fungal (yellow), eubacterial (red), and archaebacterial (pink). The accession numbers of all sequences can be found at http:/www.chem.rug.nl/biotechnology.

Members of the NMO family have been less well studied which is partly due to their low affinity for the FAD cofactor [26]. In NMO sequences, only the histidine appears to be strictly conserved in the sequence region corresponding to the location of the above-mentioned BVMO and FMO sequence motifs (Fig. 1) (see [27]). The conservation of this histidine residue throughout the monooxygenase superfamily indicates that it fulfills a role in a shared function within the monooxygenase superfamily.

3.3. Probing the functional role of the newly identified sequence motif

To examine the functional role of the BVMO-specific sequence motif we performed a site-directed mutagenesis study on 4-hydroxyacetophenone monooxygenase from *Pseudomonas fluorescens*. This BVMO has recently been cloned and characterized in our laboratory providing a good opportunity to investigate the observed sequence conservation [7]. As targets for mutagenesis, the conserved His296 and Trp300 residues were chosen. Expression of both W300A and W300Y mutant enzymes resulted in formation of insoluble and inactive protein. This indicates that the conserved tryptophan serves a critical role in attaining and/or maintaining a correctly folded state of the enzyme. The H296A mutant enzyme was expressed in soluble form and could be purified using the procedure used for wild type 4-hydroxyacetophenone monooxygenase. However, during purification the mutant enzyme lost its FAD cofactor to some extent. Holo H296A enzyme could be obtained by saturating the apoprotein fraction with FAD or by purifying the protein in the presence of 50 μM FAD. Both procedures resulted in fully reconstituted H296A as indicated by an A_{280}/A_{440} ratio of 16.8 which is similar to that observed for wild type enzyme. Gel filtration experiments revealed that, similar to wild type enzyme, the holo form of the H296A mutant is present as a dimer. Furthermore, the far-UV CD spectra of wild type 4-hydroxyacetophenone monooxygenase and the H296A mutant were identical, indicating that the mutation did not affect overall protein folding. Kinetic analysis revealed that the H296A mutant showed nearly no activity ($k_{cat} < 0.01$ s^{-1}). Moreover, the flavin spectral properties of the H296A mutant were slightly different from those of wild type 4-hydroxyacetophenone monooxygenase. These data suggest that His296 is located in the vicinity of the flavin cofactor site and is important for catalysis.

A sequence homology search in the protein structure database (PDB) using the 4-hydroxyacetophenone monooxygenase sequence gave a further hint to the functional role of the newly identified fingerprint sequence. Except for homology with short stretches of sequences containing the typical GXGXX(G/A) Rossmann fold fingerprint only one protein sequence in the PDB, that of thioredoxin reductase from *Escherichia coli*, displayed sequence similarity covering a significant length of amino acids (23% sequence identity in 196 residues; see Fig. 1). Thioredoxin reductase is a member of the NAD(P)H-dependent disulfide oxidoreductase family of homodimeric flavoenzymes [28]. In thioredoxin reductase, the flavin cofactor is first reduced by NADPH, after which the flavin transfers the electrons to the active-site disulfide formed by Cys135 and Cys138. From the alignment it can be seen that the identified BVMO sequence motif aligns well with the sequence region of thioredoxin reductase containing the active-site cysteines. This supports the above proposal that the newly identified sequence motif reflects an essential structural element in the active site of BVMOs. Analogous to the active-site cysteines in the disulfide oxidoreductases, the conserved histidine may fulfill a crucial role in catalysis in all members of the identified monooxygenase superfamily through a specific interaction with the flavin cofactor.

Elucidation of the exact functional role of the observed sequence conservation awaits a crystallographic analysis of a member of this novel flavoprotein monooxygenase superfamily as no three-dimensional structure is available for any BVMO, FMO or NMO. Nonetheless, the uncovered fingerprint sequences unambiguously discriminate these enzymes from the mechanistically related flavoprotein hydroxylases and represent valuable tools for functional annotation of genes in the steadily growing genomic sequence database.

References

[1] Willetts, A. (1997) Trends Biotechnol. 15, 55–62.
[2] Chen, Y.C., Peoples, O.P. and Walsh, C.T. (1998) J. Bacteriol. 170, 781–789.
[3] Sheng, D., Ballou, D.P. and Massey, V. (2001) Biochemistry 40, 11156–11167.
[4] Stewart, J.D. (1998) Curr. Org. Chem. 2, 195–216.
[5] Ryerson, C.C., Ballou, D.P. and Walsh, C. (1982) Biochemistry 21, 2644–2655.
[6] Entsch, B. and van Berkel, W.J.H. (1995) FASEB J. 9, 476–483.
[7] Kamerbeek, N.M., Moonen, M.J., van der Ven, J.G., van Berkel, W.J.H., Fraaije, M.W. and Janssen, D.B. (2001) Eur. J. Biochem. 268, 2547–2557.
[8] Morii, S., Sawamoto, S., Yamauchi, Y., Miyamoto, M., Iwami, M. and Itagaki, E. (1999) J. Biochem. (Tokyo) 126, 624–631.
[9] Kostichka, K., Thomas, S.M., Gibson, K.J., Nagarajan, V. and Cheng, Q. (2001) J. Bacteriol. 183, 6478–6486.
[10] Brzostowicz, P.C., Gibson, K.L., Thomas, S.M., Blasko, M.S. and Rouviere, P.E. (2000) J. Bacteriol. 182, 4241–4248.
[11] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[12] Brown, D.W., Yu, J.H., Kelkar, H.S., Fernandes, M., Nesbitt, T.C., Keller, N.P., Adams, T.H. and Leonard, T.J. (1996) Proc. Natl. Acad. Sci. USA 93, 1418–1422.
[13] Yu, J., Chang, P.K., Bhatnagar, D. and Cleveland, T.E. (2001) Appl. Microbiol. Biotechnol. 53, 583–590.
[14] Thierry, D., Vincent, V., Clement, F. and Guesdon, J.L. (1993) J. Clin. Microbiol. 31, 1048–1054.
[15] Baulard, A.R., Betts, J.C., Engohang-Ndong, J., Quan, S., McAdam, R.A., Brennan, P.J., Locht, C. and Besra, G.S. (2000) J. Biol. Chem. 275, 28326–28331.
[16] DeBarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G. and Barry III, C.E. (2000) Proc. Natl. Acad. Sci. USA 97, 9677–9682.
[17] Vallon, O. (2000) Proteins 38, 95–114.
[18] Eppink, M.H., Schreuder, H.A. and van Berkel, W.J.H. (1997) Protein Sci. 6, 2454–2458.
[19] Zhang, Z., Schaffer, A.A., Miller, W., Madden, T.L., Lipman, D.J., Koonin, E.V. and Altschul, S.F. (1998) Nucleic Acids Res. 26, 3986–3990.
[20] Ziegler, D.M. (1990) Trends Pharmacol. Sci. 11, 321–324.
[21] Stehr, M., Diekmann, H., Smau, L., Seth, O., Ghisla, S., Singh, M. and Macheroux, P. (1998) Trends Biochem. Sci. 23, 56–57.
[22] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[23] Dolphin, C.T., Janmohamed, A., Smith, R.L., Shephard, E.A. and Phillips, I.R. (1997) Nat. Genet. 17, 491–494.
[24] Zhao, Y., Christensen, S.K., Fankhauser, C., Cashman, J.R., Cohen, J.D., Weigel, D. and Chory, J. (2001) Science 291, 306–309.
[25] Suh, J.K., Poulsen, L.L., Ziegler, D.M. and Robertus, J.D. (1999) Proc. Natl. Acad. Sci. USA 96, 2687–2691.
[26] Stehr, M., Smau, L., Singh, M., Seth, O., Macheroux, P., Ghisla, S. and Diekmann, H. (1999) Biol. Chem. 380, 47–54.
[27] Ambrosi, C., Leoni, L., Putignani, L., Orsi, N. and Visca, P. (2000) J. Bacteriol. 182, 6233–6238.
[28] Williams Jr., C.H. (1995) FASEB J. 9, 1267–1276.
[29] Giardina, P.C., Foster, L.A., Toth, S.I., Roe, B.A. and Dyer, D.W. (1995) Gene 167, 133–136.
[30] Lynch, D., O'Brien, J., Welch, T., Clarke, P., Cuiv, P.O., Crosa, J.H. and O'Connell, M. (2001) J. Bacteriol. 183, 2576–2585.
[31] Mei, B., Budde, A.D. and Leong, S.A. (1993) Proc. Natl. Acad. Sci. USA 90, 903–907.
[32] Visca, P., Ciervo, A. and Orsi, N. (1994) J. Bacteriol. 176, 1128–1140.