Protein Science (1997), 6:2454–2458. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding

MICHEL H.M. EPPINK,[1] HERMAN A. SCHREUDER,[2] AND WILLEM J.H. VAN BERKEL[1]

[1]Department of Biochemistry, Agricultural University, Wageningen, The Netherlands
[2]Core Research Functions, Building G865A, Hoechst Marion Roussel, D-65926 Frankfurt, Germany

(RECEIVED May 19, 1997; ACCEPTED July 7, 1997)

Abstract: A novel conserved sequence motif has been located among the flavoprotein hydroxylases. Based on the crystal structure and site-directed mutagenesis studies of *p*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens*, this amino acid fingerprint sequence is proposed to play a dual function in both FAD and NAD(P)H binding. In PHBH, the novel sequence motif (residues 153–166) includes strand A4 and the N-terminal part of helix H7. The conserved amino acids Asp 159, Gly 160, and Arg 166 are necessary for maintaining the structure. The backbone oxygen of Cys 158 and backbone nitrogens of Gly 160 and Phe 161 interact indirectly with the pyrophosphate moiety of FAD, whereas it is known from mutagenesis studies that the side chain of the moderately conserved His 162 is involved in NADPH binding.

Keywords: fingerprint; flavoprotein family; NADPH-binding; *p*-hydroxybenzoate hydroxylase; sequence alignment

Flavoprotein hydroxylases are monooxygenases that catalyze the insertion of one atom of molecular oxygen into the substrate, using pyridine nucleotides as external electron donor (van Berkel & Müller, 1991). These enzymes play an important role in the biodegradation of lignin-derived aromatic compounds as well as environmental pollutants, and in the biosynthesis of sterols, antibiotics, and plant hormones. They lack a known fingerprint sequence for NAD(P)H binding, but possess two fingerprint motifs for the FAD binding. The first FAD motif identifies the dinucleotide binding $\beta\alpha\beta$-fold, which binds the ADP moiety of FAD (Wierenga et al., 1986), whereas the second motif represents residues that are in contact with the riboflavin moiety of FAD (Eggink et al., 1990).

PHBH (EC 1.14.13.2) is the prototype of FAD-dependent hydroxylases, and the only enzyme in this class of flavoproteins for which a three-dimensional structure is known in atomic detail (Schreuder et al., 1989). The strictly NADPH-dependent enzyme catalyzes the *ortho*-hydroxylation of 4-hydroxybenzoate into 3,4-dihydroxybenzoate via the transient stabilization of an oxygenated flavin intermediate (Entsch & van Berkel, 1995). The structure of PHBH is unusual because there is no NADPH-binding domain. So far, crystallographic analysis did not reveal a structure of the enzyme complexed with NADPH, and soaking experiments with the coenzyme analogue ADPR resulted in displacement of FAD by ADPR (van der Laan et al., 1989). Site-directed mutagenesis studies have pointed to the involvement of Arg 44 (Eppink et al., 1995) and His 162 (Eppink et al., 1997) in NADPH binding. From this and the properties of other mutants, a model for the mode of coenzyme binding was proposed (van Berkel et al., 1997).

In the past few years, the number of flavoprotein hydroxylase cloned genes has increased tremendously, and about 50 amino acid sequences are known currently. Therefore, and in view of the unknown binding mode of NADPH in this class of flavoenzymes, it was of interest to search for the presence of conserved sequence motifs. This report describes the identification of a novel sequence motif in flavoprotein hydroxylases, which appears to be important for the binding of both FAD and NAD(P)H.

Discussion: Sequence alignments have classified a number of gene products to flavoprotein hydroxylases (Kälin et al., 1992; Kukor & Olsen, 1992; Nakahigashi et al., 1992; Blanco et al., 1993; Filippini et al., 1995; Haigler et al., 1996; Marin et al., 1996; Seibold et al., 1996; Tsuji et al., 1996; Yang et al., 1996). These sequence data, together with that of PHBH from *Pseudomonas fluorescens* (van Berkel, 1992), were the starting points for a thorough screening of different databases. This search was performed with BEAUTY, which is an BLAST-enhanced alignment utility that integrates multiple biological information resources (Worley et al., 1995). From the 50 collected sequences, small groups were generated based on the different types of substrates: *p*-hydroxybenzoate (phb), 2-hydroxybiphenyl (biph), phenol (phe), salicylate (sal), *p*-aminobenzoate (pab), polyketide (poly), 4-methyl-5-nitrocatechol (cat), epoxide (epox), 2-methyl-3-hydroxypyridine-5-carboxylic acid (oxy), and a group of monooxygenases (mono) for which the function is largely unknown. A multiple sequence alignment was per-

---

Reprint requests to: W.J.H. van Berkel, Department of Biochemistry, Wageningen Agricultural University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands; e-mail: willem.vanberkel@fad.bc.wau.nl.

Table 1. *Multiple alignment of the three consensus sequences in the flavoprotein hydroxylases*[a]

Enzyme (strain)	FAD fingerprint (1)	Conserved motif	FAD fingerprint (2)	Reference
1 phb (*Pseudomonas fluorescens*)	5-VAIIGA**G**PS**G**LLLG-18	153-DYIAGC**DG**FHGISR-166	279-GRLFLA**GD**AAHIVPPTGAKGLNLAASDVSTL-309	(Weijer et al., 1982)
2 phb (*Pseudomonas aeruginosa*)	5-VAII**G**A**G**PS**G**LLLG-18	153-DYIAGC**DG**FHGISR-166	279-GRLFLA**GD**AAHIVPPTGAKGLNLAASDVSTL-309	(Entsch et al., 1988)
3 phb (*Pseudomonas fluorescens*)	8-VAII**G**A**G**PS**G**LLLG-21	156-DYIAGC**DG**FHGVAR-169	282-GRLFLL**GD**AAHIVPPTGAKGLNLAASDVSTL-312	(Shuman & Dix, 1993)
4 phb (*Rhizobium legominosarum* B155)	5-VAII**G**S**G**PS**G**LLLG-18	153-DFIAGC**DG**FHGASR-166	279-GRLFLV**GD**AAHIVPPTGAKGLNLAASDVHYL-309	(Wong et al., 1994)
5 phb (*Rhizobium legominosarum* MNF300)	5-VAII**G**S**G**PS**G**LLFG-18	153-DFIAGC**DG**FHGVSR-166	279-GRLFLV**GD**AAHIVPPTGAKGLNLAASDVHYL-309	(Wong et al., 1994)
6 phb (*Acinetobacter calcoaceticus*)	5-VAII**G**S**G**PA**G**LLLG-18	152-DFIAGC**DG**YHGVCR-165	278-GKLFLA**GD**AAHIVPPTGAKGLNLAASDIAYL-308	(DiMarco et al., 1993)
7 phb (*Pseudomonas species* CBS3)	9-VGII**G**A**G**PA**G**LLLS-22	156-DYVVGC**DG**FHGPSR-169	282-GRLLLA**GD**AAHIVPPTGAKGLNLAVNDVRLL-312	(Seibold et al., 1996)
8 biph (*Pseudomonas azelaica* HBP1)	9-VLIV**G**A**G**PA**G**AMSA-22	172-KYIIGA**DG**AHSLVA-185	306-GRVFCM**GD**AVHRHTPMGGLGLNTSVQDAYNL-336	(Schmid & van der Meer, 1997)
9 phe (*Flavobacterium* species)	17-VLIV**G**G**G**PT**G**LIAA-30	167-RWVIGA**DG**VRSRVR-180	291-GNVFLA**GD**AAHCHSPSGGSGMNVGMQDAFNL-321	(Orser et al., 1993)
10 phe (*Sphingomonas chlorophenolica*)	17-VLIV**G**G**G**PT**G**LIAA-30	167-RWVIGA**DG**VRSRVR-180	291-GNVFLA**GD**AAHCHSPSGGSGMNVGMQDAFNL-321	(Ederer et al., 1996)
11 phe (*Alcaligenes eutrophus*)	9-VLVV**G**T**G**PA**G**ASAG-22	172-KYLIGA**DG**ANSRVV-185	304-GRVFCA**GD**AVHRHPPTNGLGSNTSIQDSFNL-334	(Perkins et al., 1990)
12 phe (*Ralstonia eutropha*)	12-VLVV**G**T**G**PA**G**AAST-25	175-KYLIGA**DG**ANSRIV-188	307-KRVFCM**GD**AVHRHPPTNGLGSNTSIQDAFNL-337	(van der Meer, 1997)
13 phe (*Pseudomonas putida* pEST4011)	12-VLVV**G**S**G**PA**G**AAST-25	175-KYLIGA**DG**ANSQVV-188	307-NRVFCM**GD**AVHRHPPTNGLGSNTSIQDAFNL-337	(Kõiv et al., 1996)
14 phe (*Pseudomonas* species EST1001)	37-VLIV**G**S**G**PA**G**SSAA-50	200-KYLIGA**DG**ARSKVA-213	321-GRVCCA**GD**AIHKHPPSHGLGSNTSIQDSYNL-351	(Nurk et al., 1991)
15 phe (*Trichosporon cutaneum*)	10-VLIV**G**A**G**PA**G**LMAA-23	219-KYVIGC**DG**GHSWVR-232	350-ERVFIA**GD**ACHTHSPKAGQGMNTSMMDTYNL-380	(Kälin et al., 1992)
16 poly (*Streptomyces coelicolor*)	23-VLVV**G**G**S**LV**G**LSTS-36	184-DYLVAA**DG**PRSPVR-197		(Blanco et al., 1993)
17 poly (*Streptomyces halstedii*)	17-VLVV**G**G**S**LV**G**LSTS-30	171-DYLVAA**DG**PRSPVR-184	306-GRVFLA**GD**SAHEMSPTGAFGSNTGIQDAHNL-336	(Blanco et al., 1993)
18 poly (*Streptomyces peucetius*)	10-VLVV**G**G**G**LG**G**LSTA-23	173-RYLVAA**DG**PRSAIR-186	301-GPVLLV**GD**AAKVTPPTGGMGGNTAIGDGFDV-331	(Filippini et al., 1995)
19 poly (*Streptomyces peucetius caesius*)	10-VLVV**G**G**G**LG**G**LSTA-23	173-RYLVAA**DG**PRSAIR-186	301-GPVLLV**GD**AAKVTPPTGGMGGNTAIGDGFDV-331	(Hong et al., 1994)
20 poly (*Streptomyces purpurascens*)	8-VLVV**G**A**G**LG**G**LSTA-21	173-GYLVGA**DG**NRSLVR-186	301-GRVFLA**GD**AAKVTPPTGGMSGNAAVADGFDL-331	(Niemi & Mäntsälä, 1995)
21 poly (*Streptomyces glaucescens*)	19-VLIV**G**G**G**LT**G**LSAA-32	191-RYLIAA**DG**VRSPVR-204	328-GRVFLA**GD**AAHVHPPAGAFGANGGIQDAHNL-358	(Decker et al., 1993)
22 poly (*Streptomyces fradiae*)	5-VIVA**G**A**G**PT**G**LMLA-18	147-PYLVGC**DG**GRSTVR-160	270-GRVLLA**GD**AAHIHLPAGGQGMNTGIQDAVNL-301	(Decker & Haag, 1995)
23 poly (*Streptomyces venezuelae*)	28-VVVI**G**A**G**PT**G**LMLA-41	170-RYAVGC**DG**ERTTVR-183	290-GRVLWA**GD**AAHQQMPIGGQALNLGLQDAVNL-320	(Yang et al., 1996)
24 poly (*Streptomyces aureofaciens*)	1-MVVA**G**A**G**PT**G**LMLA-14	137-RYLVGC**DG**G-STVR-149	246-GRVLLA**GD**AAHIHLPAGGQGLNLGFQDAVNL-276	(Dairi et al., 1995)
25 cat (*Burkholderia species* DNT)	14-VLIV**G**G**S**MV**G**LSTA-27	180-RYLVAS**DG**WRSQRR-193	306-GRIFLR**GD**AAHVVPPYGGFGGNTGVQDAHNL-336	(Haigler et al., 1996)
26 oxy (*Pseudomonas* MA-1)	14-AEVA**G**G**G**FA**F**LTAA-27	150-DLIVGA**DG**VGSKVR-163	281-GKVALV**GD**AAHAMCPALAQGAGCAMVNAFSL-311	(Chaiyen et al., 1997)
27 sal (*Pseudomonas putida* NAH7)	10-IGIV**G**G**G**IS**G**VALA-23	157-DLLIGA**DG**IKSALR-170	308-GRVVLI**GD**AAHAMLPHQGAGAGQGLEDAYFL-338	(You et al., 1991)
28 sal (*Pseudomonas putida*)	10-IGII**G**G**G**IS**G**VALA-23	156-DLLIGA**DG**IKSALR-169	307-GRVVLI**GD**AAHAMLPHQGAGAAQGLEDAYFL-337	(Darby et al., 1995)
29 sal (*Pseudomonas putida* KF715)	13-VAIV**G**G**G**IS**G**LALA-26	158-DLLIGR**DG**IKSALR-171	308-GRVALI**GD**AAHAMLPHQGAGAGQGLEDAYFL-338	(Lee et al., 1996)
30 sal (*Pseudomonas putida* S-1)	8-VAVI**G**G**G**IA**G**TALA-21	154-DVAIAA**DG**IKSSMR-167	305-GRVALI**GD**AAHAMLPHQGAGAGQGLEDAYFM-335	(Suzuki et al., 1996b)
31 sal (*Sphingomonas* species AJ1)	3-IAII**G**A**G**IG**G**LALA-16	153-DVVIGA**DG**VRSVIR-166	286-GPAVLI**GD**AAHAMLPHHGQGANTSIEDACVL-316	(Iwabuchi, 1997)
32 pab (*Agaricus bisporus*)	9-VAIV**G**A**G**IV**G**LTLA-22	175-DILVGA**DG**IKSTLR-188	326-GRVFLM**GD**AAHAMLPHLGAGAHVGMEDAYIL-356	(Tsuji et al., 1996)
33 epox (*Lycopersicon esculentum*)	88-VLVA**G**G**G**IG**G**LVFA-101	242-DLLVGA**DG**IRSKVR-255	371-GRVTLL**GD**SVHAMQPNLGQGGCMAIEDSYQL-401	(Burbidge, 1997)
34 epox (*Nicotiana plumbaginifolia*)	82-VLVA**G**G**G**IG**G**LVFA-95	236-DLLVGA**DG**IRSKVR-249	365-GRVTLL**GD**SVHAMQPNLGQGGCMAIEDSYQL-395	(Marin et al., 1996)
35 epox (*Capsicum annuum*)	83-VLVA**G**G**G**IG**G**LVFA-96	236-DLLVGA**DG**IRSKVR-249	366-GRVTLL**GD**SVHAMQPNLGQGGCMAIEDSYQL-396	(Bouvier et al., 1996)
36 epox (*Saccharomyces cerevisiae*)	21-AIVI**G**A**G**VI**G**PCVA-34	203-HLTFIC**DG**IFSRFR-216	328-TGMCVI**GD**ALNMRHPLTGGGMTVGLHDVVLL-358	(Jandrositz et al., 1991)
37 epox (Rat)	125-VIII**G**S**G**VL**G**SALA-138	278-PLTVVA**DG**LFSKFR-290	401-RGVLLL**GD**AYNLRHPLTGGGMTVALKDIKIW-431	(Sakakibara et al., 1995)
38 mono (*Rhodococcus equi*)	4-VIIV**G**A**G**PT**G**LMLA-17	145-RYLVGC**DG**GRSTVR-158	271-DRVFLA**GD**AAHIHPPMGGQGLNLGVQDAFNL-301	(Andersen et al., 1997)
39 mono (*Pseudomonas aeruginosa*)		50-SLVVGA**DG**IRSSVR-63	178-GRVVLI**GD**AAHAMTPNLGQGAAMALEDAFLL-208	(Kinscherf & Willis, 1995)
40 mono (*Escherichia coli* K-12 CS520)	18-VAIA**G**A**G**PV**G**LMMA-31	168-QWLVAC**DG**GASFVR-181	288-DRVLLA**GD**AAHIMPVWQGQGYNSGMRDAFNL-318	(Ferrandez et al., 1996)
41 mono (*Escherichia coli* K-12 W3110)	18-VAIA**G**A**G**PV**G**LMMA-31	168-QWLVAC**DG**GASFVR-181	288-DRVLLA**GD**AAHIMPVWQGQGYNSGMRDAFNL-318	(Kawamukai, 1996)
42 mono (*Escherichia coli* K-12 visC)	6-VAIV**G**G**G**MV**G**LAVA-19	150-RLVIGA**DG**ANSWLR-163	274-HRLALV**GD**AAHTIHPLAGQGVNLGFMDAAEL-304	(Nakahigashi et al., 1992)
43 mono (*Escherichia coli* K-12 ubiH)	3-VIIV**G**G**G**MA**G**ATLA-16	147-RVLVAA**DG**THSALA-160	270-HRTVLV**GN**AAQTLHPIAGQGFNLGMRDVMSL-300	(Nakahigashi et al., 1992)
44 mono (*Saccharomyces cerevisiae*)	5-VAII**G**A**G**LV**G**CLAA-18	162-DFVIGC**DG**AYSATR-175	307-GKAILL**GD**AAHAMVPFYGQGMNCGFEDVRIL-337	(Domdey et al., 1994)
45 mono (*Caenorhabditis elegans* R07B7.4)	4-VVIA**G**G**G**LV**G**SANA-17	155-DLFLAC**DG**AHSSIR-168	291-DKLVLM**GD**AAHAMVPFNGQGVNCGFEDCLVL-321	(Wilson et al., 1994)
46 mono (*Caenorhabditis elegans* R07B7.5)	4-VAIA**G**A**G**LV**G**ALNA-17	171-DLILAC**DG**AYSAVR-184	308-DNLVLM**GD**AAHAMVPFYGQGMNCGFEDCLVF-338	(Wilson et al., 1994)
47 mono (*Bacillus subtilis*)	4-MLIA**G**G**G**IG**G**LSAA-17	149-DILAGF**DG**IHSVVR-162	278-GRVIIG**GD**AAHAGAPTLAQGAAMAIEDAIVL-308	(Sekiguchi, 1996)
48 mono (*Mycobacterium tuberculosis* H37Rv)	4-VVVS**G**A**S**VA**G**TAAA-17	150-DLVIGA**DG**LHSNVR-163	282-GRVALV**GD**AGYCCSPLSGQGTSVALLGAYIL-312	(Barrell et al., 1996)
Consensus sequence	Vhhh**G**s**G**hh**G**hhhs	chhhss**DG**xcSxhR	GxhhLh**GD**AAHxxxPxxGxGxNxsxxDsxxL	

[a] This multiple alignment was obtained from 48 sequences with MACAW and ClustalW using the BLOSUM62 matrix. The consensus profiles shown underneath the alignment include strictly conserved residues as well as those profiles in which there are not more than 10 violations. Uppercase letters in the profile are amino acid residues, lowercase letters and symbols are: h = hydrophobic residues; s = small residues; c = charged residues; x = all residues; − = gap.

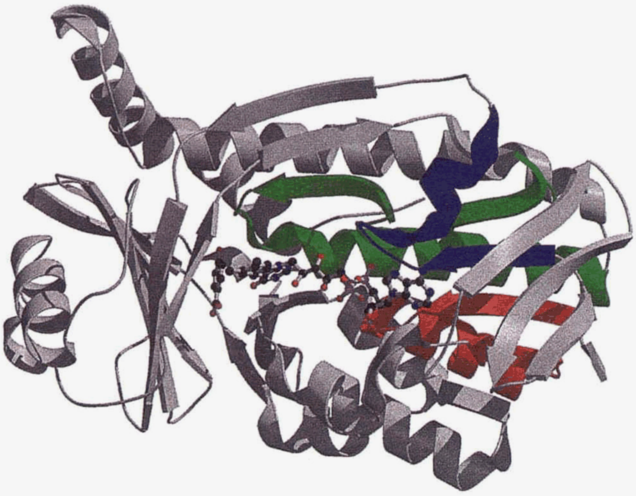

Fig. 1. Ribbon structure of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens.* MOLSCRIPT (Kraulis, 1991) picture highlighting the conserved regions of the protein. GXGXXG sequence in red; DG sequence in blue; GD sequence in green. The FAD and aromatic substrate are depicted in ball and stick representation.

formed with the programs MACAW (Schuler et al., 1991) and ClustalW (Thompson et al., 1994) using the Blosum matrixes (Henikoff & Henikoff, 1992). From the alignment of 50 flavoprotein hydroxylase sequences, three conserved regions could be deduced, which are shown in Table 1.

The first FAD fingerprint sequence, shown in Table 1, is the well-known Rossmann fold or $\beta\alpha\beta$-fold (containing the GXGXXG sequence), a common motif among FAD- and NAD(P)H-dependent oxidoreductases (Wierenga et al., 1986). In PHBH, this fingerprint (residues 5–19) is important for binding the ADP moiety of FAD (Fig. 1). The structural properties of this dinucleotide binding fold were reported more than 10 years ago (Wierenga et al., 1983, 1985, 1986).

The second FAD binding motif contains the GD sequence (Table 1) with the highly conserved Asp residue, which contacts the O3′ of the ribose moiety of FAD (Eggink et al., 1990). This common fingerprint sequence among the family of FAD-dependent oxidoreductases differs somewhat between the disulfide oxidoreductases and flavoprotein hydroxylases because the latter enzymes have more conserved residues downstream from the GD sequence (DiMarco et al., 1993). In PHBH, this fingerprint sequence (residue 278–308) is located partly at the *re*-side of the isoalloxazine ring of FAD, near the binding site of the aromatic substrate (Fig. 1; Schreuder et al., 1989).

Table 1 shows that the newly defined DG amino acid sequence is highly conserved among all flavoprotein hydroxylases studied. In PHBH, this short sequence motif comprises strand A4 and the N-terminal part of helix H7 (residues 153–166) of the FAD binding domain, and is situated near the cleft leading toward the active site (Fig. 1). Strand A4 (residue 154–157) is completely buried and multiple contacts are made with residues of both the FAD binding domain and a long excursion of the substrate binding domain. However, as one of the referees pointed out, one could argue that this excursion, together with the FAD and the interface domain, forms one large globular domain and that the contacts of strand A4 are important for maintaining the integrity of this domain. The large turn (residues 158–163) that connects strand A4 and helix H7 contains the strictly conserved residues Asp 159 and Gly 160. This Gly 160 faces the putative NADPH binding cleft and its Phi/Psi angles (62.1/−174.8) are allowed for glycines, whereas they are disallowed for other residues. Also, a side chain at this position would probably hinder binding of the cofactor. The structurally important and tightly packed residues Asp 159 and Gly 160 form hydrogen bonds with the backbone atoms of residues 163–165 (Fig. 2). Indirect hydrogen bonds exist between the backbone oxygen of Cys 158 and backbone nitrogen of Gly 160, with the pyrophosphate moiety of FAD via protonated water molecules (Schreuder et al., 1989). From site-directed mutagenesis studies, it is known that replacement of Cys 158 by Ser decreases the affinity for FAD, probably by influencing the solvation of the pyrophosphate moiety of FAD (van der Bolt et al., 1994). Mutagenesis studies also revealed that His 162 is very important for the binding of NADPH (Eppink et al., 1997). Table 1 shows that this position in the conserved sequence motif almost always contains a positively charged residue. Chemical modification of salicylate hydroxylase has suggested that Lys 165, the equivalent of His 162 in

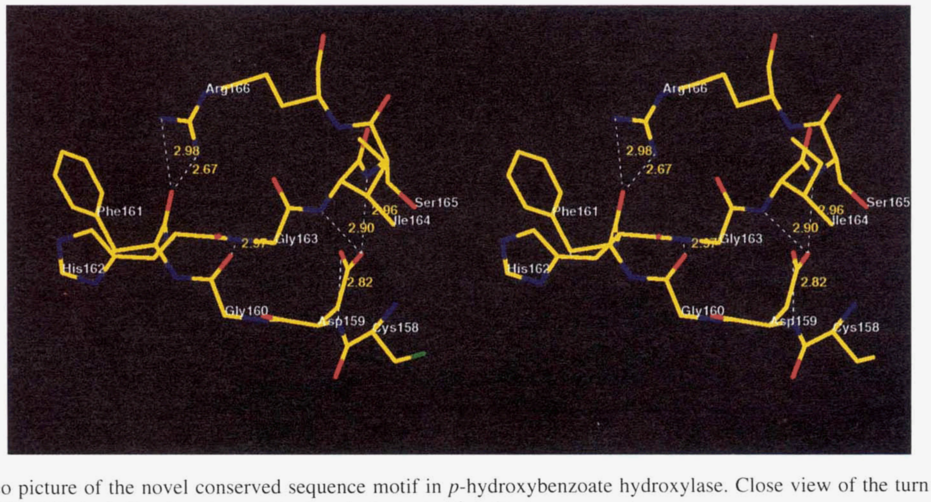

Fig. 2. Stereo picture of the novel conserved sequence motif in *p*-hydroxybenzoate hydroxylase. Close view of the turn region from amino acid 158–166, including the strong intradomain hydrogen bonds.

PHBH, is important for binding the pyrophosphate moiety of NADH (Suzuki et al., 1996a). Phe 161 in PHBH is not conserved in the fingerprint. Mutagenesis studies confirmed that replacement of Phe by Ala weakens NADPH binding, but that Phe 161 is not structurally important (van Berkel et al., 1997). Helix H7 is not regular (Schreuder et al., 1989) but, as in regular α-helices, all peptide dipoles point in approximately the same direction, giving rise to an overall helix dipole moment (Hol et al., 1978). This helix H7 (residues 164–169) is located near the protein surface (Fig. 1). The highly conserved residue Arg 166 in this helix forms strong inter- and intradomain contacts with the backbone oxygens of Phe 161 and Ala 287 (Fig. 2). Substitution of Arg 166 by Ser led to significant structural changes in the Cα-backbone and destabilization of the mutant (van Berkel et al., 1997).

In conclusion, a unique short amino acid sequence motif for flavoprotein hydroxylases is presented that seems to serve a dual function. Crystallographic analysis and site-directed mutagenesis studies of PHBH from *P. fluorescens* suggest that this sequence is involved indirectly in binding the pyrophosphate moiety of FAD and that it is also necessary for the recognition of the NADPH cofactor. Although the mode of NADPH binding in PHBH is still unknown, helix H7 might be involved in binding the pyrophosphate moiety of the pyridine nucleotide cofactor. There are two common characteristics of a dinucleotide binding fold (Wierenga et al., 1985) that probably also occur here. (1) A glycine residue near the N-terminus of a helix, to allow close contact with the pyrophosphate moiety: Gly 160 is located at such a position near the N-terminus of helix H7. (2) Favorable interaction of the helix dipole with the negatively charged pyrophosphate moiety: In the proposed model, the pyrophosphate moiety of NADPH is located near the positive end of the dipole of helix H7 (van Berkel et al., 1997). The newly identified fingerprint is highly specific for flavoprotein hydroxylases. Running BEAUTY with the sequence DFLVGADGIHSXVR (based on the alignment results and where X denotes all possibilities) yielded 48 of 52 flavoprotein hydroxylases present in the databases, and 5 unrelated proteins. Our fingerprint recognizes all different types of flavoprotein hydroxylases that are encoded by a single gene, something the current fingerprints cannot recognize. This shows that our fingerprint is able to detect unambiguously flavoprotein hydroxylases, which will allow the identification of such enzymes among the millions of genes that are produced currently by the large scale whole-genome sequencing efforts.

References

Andersen SJ, Quan S, Gowan B, Dabbs ER. 1997. Monooxygenase-like sequence of a *Rhodococcus equi* gene conferring increased resistance to rifampin by inactivating this antibiotic. *Antimicrob Agents Chemother 41*:218–221.

Barrell BG, Rajandream MA, Walsh SV. 1996. *Mycobacterium tuberculosis* monooxygenase gene. Direct submission to GenBank (MTCY50).

Blanco G, Pereda A, Brian P, Méndez C, Chater KF, Salas JA. 1993. A hydroxylase-like gene product contributes to synthesis of a polyketide pigment in *Streptomyces halstedii*. *J Bacteriol 175*:8043–8048.

Bouvier F, d'Harlingue A, Hugueney P, Marin E, Marion-Polin A, Camara B. 1996. Xanthophyll biosynthesis. Cloning, expression, functional reconstitution, and regulation of β-cyclohexenyl carotenoid epoxidase from pepper (*Capsicum annuum*). *J Biol Chem 271*:28861–28867.

Burbidge A. 1997. *Lycopersicon esculentum* mRNA for zeaxanthin epoxidase. Direct submission to GenBank (LEZEAXAT).

Chaiyen P, Ballou DP, Massey V. 1997. Gene cloning, sequence analysis, and expression of 2-methyl-3-hydroxypyridine-5-carboxylic acid (MHPC) oxygenase. *Proc Natl Acad Sci USA*. pp 7233–7238.

Dairi T, Nakano T, Aisaka K, Katsumata R, Hasegawa M. 1995. Cloning and nucleotide sequence of the gene responsible for chlorination of tetracycline. *Biosci Biotech Biochem 59*:1099–1106.

Darby RM, Bi YM, Doig S, Firek S, Draper J. 1995. *Pseudomonas putida* nahG gene for salicylate hydroxylase. Direct submission to GenBank (PPNAHGG).

Decker H, Haag S. 1995. Cloning and characterization of a polyketide synthase gene from *Streptomyces fradiae* Tü2717, which carries the genes for biosynthesis of the angucycline antibiotic urdamycin A and a gene probably involved in its oxygenation. *J Bacteriol 177*:6126–6136.

Decker H, Motamedi H, Hutchinson CR. 1993. Nucleotide sequences and heterologous expression of *tcmG* and *tcmP*, biosynthesis genes for tetracenomycin C in *Streptomyces glaucescens*. *J Bacteriol 175*:3876–3886.

DiMarco AA, Averhoff BA, Kim EE, Ornston LN. 1993. Evolutionary divergence of *pobA*, the structural gene encoding *p*-hydroxybenzoate hydroxylase in an *Acinetobacter calcoaceticus* strain well-suited for genetic analysis. *Gene 125*:25–33.

Domdey H, Gassenhuber H, Obermaier B, Piravandi E. 1994. *Saccharomyces cerevisiae* chromosome II reading frame ORF YBL098w. Direct submission to GenBank (SCYBL098W).

Ederer MM, Crawford RL, Orser CS. 1996. *Sphingomonas chlorophenolica* strain ATCC 33790 PCP-4-monooxygenase(pcpB) gene. Direct submission to GenBank (SCU60175).

Eggink G, Engel H, Vriend G, Terpstra P, Witholt B. 1990. Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. *J Mol Biol 212*:135–142.

Entsch B, Nan Y, Weaich K, Scott KF. 1988. Sequence and organization of *pobA*, the gene coding for *p*-hydroxybenzoate hydroxylase, an inducible enzyme from *Pseudomonas aeruginosa*. *Gene 71*:279–291.

Entsch B, van Berkel WJH. 1995. Structure and mechanism of *p*-hydroxybenzoate hydroxylase. *FASEB J 9*:476–483.

Eppink MHM, Jacobs D, van Berkel WJH. 1997. Involvement of His162 in NADPH binding of *p*-hydroxybenzoate hydroxylase. In: Stevenson K, Massey V, Williams Ch, eds. *Flavins and flavoproteins XII*. Calgary: University Press. pp 315–318.

Eppink MHM, Schreuder HA, van Berkel WJH. 1995. Structure and function of mutant Arg44Lys of 4-hydroxybenzoate hydroxylase. Implications for NADPH binding. *Eur J Biochem 231*:157–165.

Ferrandez A, Garcia JL, Diaz E. 1996. *Escherichia coli* mhp cluster for 3-hydroxyphenylpropionic acid degradation. Direct submission to GenBank (ECMHP).

Filippini S, Solinas MM, Breme U, Schlüter MB, Gabellini D, Biamonti G, Colombo AL, Garofano L. 1995. *Streptomyces peucetius* daunorubicin biosynthesis gene, *dnrF*: Sequence and heterologous expression. *Microbiology 141*:1007–1016.

Haigler BE, Suen WC, Spain JC. 1996. Purification and sequence analysis of 4-methyl-5-nitrocatechol oxygenase from *Burkholderia* sp. strain DNT. *J Bacteriol 178*:6019–6024.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA 89*:10915–10919.

Hol WGJ, van Duijnen PTh, Berendsen HJC. 1978. The α-helix dipole and properties of proteins. *Nature (Lond) 273*:443–446.

Hong YS, Hwang CK, Hong SK, Kim YH, Lee JJ. 1994. Molecular cloning and characterization of the aklavinone 11-hydroxylase gene of *Streptomyces peuceticus* subsp. *caesius* ATCC 27952. *J Bacteriol 176*:7096–7101.

Iwabuchi T. 1997. *Sphingomonas* sp. DNA for salicylate hydroxylase. Direct submission to GenBank (AB000564).

Jandrositz A, Turnowsky F, Högenauer F. 1991. The gene encoding squalene epoxidase from *Saccharomyces cerevisiae*: Cloning and characterization. *Gene 107*:155–160.

Kälin M, Neujahr HY, Weissmahr RN, Sejlitz T, Jöhl R, Fiechter A, Reiser J. 1992. Phenol hydroxylase from *Trichosporon cutaneum*: Gene cloning, sequence analysis, and functional expression in *Escherichia coli*. *J Bacteriol 174*:7112–7210.

Kawamukai M. 1996. *Escherichia coli* genes for MhpR, MhpA, MhpB, MhpC, MhpD, MhpE and MhpF, complete sequence of the mhp operon. Direct submission to GenBank (D86239).

Kinscherf TG, Willis DK. 1995. *Pseudomonas aeruginosa* FAD binding protein homolog gene. Direct submission GenBank (PAU29897).

Kõiv V, Marits R, Heinaru A. 1996. Sequence analysis of the 2,4-dichlorophenol hydroxylase gene *tfdB* and 3,5-dichlorocatechol 1,2-dioxygenase gene *tfdC* of 2,4-dichlorophenoxyacetic acid degrading plasmid pEST4011. *Gene 174*:293–297.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Lee J, Oh J, Min KR, Kim Y. 1996. Nucleotide sequence of salicylate hydroxylase gene and its 5′-flanking region of *Pseudomonas putida* KF715. *Biochem Biophys Res Commun 218*:544–548.

Marin E, Nussaume L, Quesada A, Gonneau M, Sotta B, Hugueney P, Frey A, Marion-Poll A. 1996. Molecular identification of zeaxanthin epoxidase of *Nicotiana plumbaginifolia*, a gene involved in abscisic acid biosynthesis and corresponding to the ABA locus of *Arabidopsis thaliana. EMBO J 15*:2331–2342.

Nakahigashi K, Miyamoto K, Nishimura K, Inokuchi H. 1992. Isolation and characterization of a light-sensitive mutant of *Escherichia coli* K-12 with a mutation in a gene that is required for the biosynthesis of ubiquinone. *J Bacteriol 174*:7352–7359.

Niemi J, Mäntsälä P. 1995. Nucleotide sequences and expression of genes from *Streptomyces purpurascens* that cause the production of new anthracyclines in *Streptomyces galilaeus. J Bacteriol 177*:2942–2945.

Nurk A, Kasak L, Kivisaar M. 1991. Sequence of the gene (*pheA*) encoding phenol monooxygenase from *Pseudomonas* sp. EST1001: Expression in *Escherichia coli* and *Pseudomonas putida. Gene 102*:13–18.

Orser CS, Lange CC, Xun L, Zahrt T, Schneider BJ. 1993. Cloning, sequence analysis, and expression of the *Flavobacterium* pentachlorophenol-4-monooxygenase gene in *Escherichia coli. J Bacteriol 175*:411–416.

Perkins EJ, Gordon MP, Caceres O, Lurquin PF. 1990. Organization and sequence analysis of the 2,4-dichlorophenol hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4. *J Bacteriol 172*:2351–2359.

Sakakibara J, Watanabe R, Kanai Y, Ono T. 1995. Molecular cloning and expression of rat squalene epoxidase. *J Biol Chem 270*:17–20.

Schmid A, van der Meer JR. 1997. 2-Hydroxybiphenyl 3-monooxygenase from *Pseudomonas azelaica* HBP1. Direct submission to GenBank (PAU73900).

Schreuder HA, Prick PAJ, Wierenga RK, Vriend G, Wilson KS, Hol WGJ, Drenth J. 1989. Crystal structure of the *p*-hydroxybenzoate hydroxylase–substrate complex refined at 1.9 Å resolution. *J Mol Biol 208*:679–696.

Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet 9*:180–190.

Seibold B, Matthes M, Eppink MHM, Lingens F, van Berkel WJH, Müller R. 1996. 4-Hydroxybenzoate hydroxylase from *Pseudomonas* sp. CBS3. Purification, characterization, gene cloning, sequence analysis and assignment of structural features determining the coenzyme specificity. *Eur J Biochem 239*:469–478.

Sekiguchi J. 1996. *Bacillus subtilis* monooxygenase gene. Direct submission to GenBank (YfnL).

Shuman B, Dix TA. 1993. Cloning, nucleotide sequence and expression of a *p*-hydroxybenzoate hydroxylase isozyme gene from *Pseudomonas fluorescens. J Biol Chem 268*:17057–17062.

Suzuki K, Mizuguchi M, Gomi T, Itagaki E. 1996a. Identification of a lysine residue in the NADH-binding site of salicylate hydroxylase from *Pseudomonas putida* S-1. *J Biochem 117*:579–585.

Suzuki K, Mizuguchi M, Ohnishi K, Itagaki E. 1996b. Structure of chromosomal DNA coding for *Pseudomonas putida S-1* salicylate hydroxylase. *Biochim Biophys Acta 1275*:154–156.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

Tsuji H, Oka T, Kimoto M, Hong YM, Natori Y, Ogawa T. 1996. Cloning and sequencing of cDNA encoding 4-aminobenzoate hydroxylase from *Agaricus bisporus. Biochim Biophys Acta 1309*:31–36.

van Berkel WJH, Eppink MHM, van der Bolt FJT, Vervoort J, Rietjens IMCM, Schreuder HA. 1997. *p*-Hydroxybenzoate hydroxylase: Mutants and mechanism. In: Stevenson K, Massey V, Williams Ch, eds. *Flavins and flavoproteins XII.* Calgary: University Press. pp 305–314.

van Berkel WJH, Müller F. 1991. Flavin-dependent monooxygenases with special references to *p*-hydroxybenzoate hydroxylase. In: Müller F, ed. *Chemistry and biochemistry of flavoenzymes* 2. Boca Raton, Florida: CRC Press. pp 1–29.

van Berkel WJH, Westphal AH, Eschrich K, Eppink MHM, de Kok A. 1992. Substitution of Arg214 at the substrate-binding site of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens. Eur J Biochem 210*:411–419.

van der Bolt FJT, Drijfhout MC, Eppink MHM, Hagen WR, van Berkel WJH. 1994. Selective cysteine → serine replacements in *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens* allow the unambiguous assignment of Cys211 as the site of modification by spin-labeled *p*-chloromercuribenzoate. *Protein Eng 7*:801–804.

van der Laan JM, Schreuder HA, Swarte MBA, Wierenga RK, Kalk KH, Hol WGJ, Drenth J. 1989. The coenzyme analogue adenosine 5-diphosphoribose displaces FAD in the active site of *p*-hydroxybenzoate hydroxylase. An X-ray crystallographic investigation. *Biochemistry 28*:7199–7205.

van der Meer JR. 1997. Chlorophenol monooxygenase from *Ralstonia eutropha* JMP134. Direct submission to GenBank (REU16782).

Weijer WJ, Hofsteenge J, Vereijken JM, Jekel PA, Beintema JJ. 1982. Primary structure of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens. Biochim Biophys Acta 704*:385–388.

Wierenga RK, De Maeyer MCH, Hol WGJ. 1985. Interaction of pyrophosphate moieties with α-helixes in dinucleotide binding proteins. *Biochemistry 24*:1346–1357.

Wierenga RK, Drenth J, Schulz G. 1983. Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. *J Mol Biol 167*:725–739.

Wierenga RK, Terpstra P, Hol WGJ. 1986. Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol 187*:101–107.

Wilson R et al. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans. Nature 368*:32–38. Direct submission GenBank (CER07B7).

Wong CM, Dilworth MJ, Glenn AR. 1994. Cloning and sequencing show that 4-hydroxybenzoate hydroxylase (*pobA*) is required for uptake of 4-hydroxybenzoate in *Rhizobium leguminosarum. Microbiology 140*:2775–2786.

Worley KC, Wiese BA, Smith RF. 1995. BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Research 5*:173–184.

Yang K, Han L, Ayer SW, Vining LC. 1996. Accumulation of the angucycline antibiotic rabelomycin after disruption of an oxygenase gene in the jadomycin B biosynthesis gene cluster of *Streptomyces venezuelae. Microbiology 142*:123–132.

You IS, Ghosa D, Gunsalus IC. 1991. Nucleotide sequence analysis of the *Pseudomonas putida* PpG7 salicylate hydroxylase gene (*nahG*) and its 3′-flanking region. *Biochemistry 30*:1635–1641.