The EMBO Journal vol.14 no.9 pp.2034–2042, 1995

Three-dimensional structure of the bifunctional protein PCD/DCoH, a cytoplasmic enzyme interacting with transcription factor HNF1

Ralf Ficner, Uwe H.Sauer, Gunter Stier and Dietrich Suck[1]

EMBL, Structural Biology Programme, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

[1]Corresponding author

R.Ficner and U.H.Sauer contributed equally to this work.

The bifunctional protein pterin-4a-carbinolamine dehydratase (PCD)/dimerization cofactor of HNF1 (DCoH) is a cytoplasmic enzyme involved in the tetrahydrobiopterin regeneration and is found in complex with the transcription factor HNF1 in liver cell nuclei. An atypical hyperphenylalaninemia and the depigmentation disorder vitiligo are related to a deficiency of PCD/DCoH activity. The crystal structure of PCD/DCoH was solved by multiple isomorphous replacement and refined to a crystallographic *R*-factor of 20.5% at 2.7 Å resolution. The single domain monomer comprises three α-helices packed against one side of a four-stranded, antiparallel β-sheet. The functional enzyme is a homo-tetramer of 222 symmetry where each of the monomers contributes one helix to a central four helix bundle. In the tetramer two monomers form an eight-stranded, antiparallel β-sheet with six helices packing against it from one side. The concave, hydrophobic surface of the eight-stranded β-sheet with its two protruding loops at either end is reminiscent of the saddle-like shape seen in the TATA-box binding protein. PCD/DCoH binds as a dimer to the helical dimerization domain of dimeric HNF1 forming a hetero-tetramer possibly through a mixed four helix bundle.

Key words: biopterin/crystal structure/four helix bundle/hyperphenylalaninemia/phenylalanine hydroxylation

Introduction

The distribution of pterin-4a-carbinolamine dehydratase (PCD) in mammalian tissues is broad, with the highest levels in the liver and kidney (Davis *et al.*, 1992a). Initially PCD was purified and characterized as phenylalanine hydroxylase stimulating protein (Huang *et al.*, 1973), but was later shown to be an enzyme involved in the phenylalanine hydroxylation system (Lazarus *et al.*, 1983). PCD catalyses the conversion of 4a-hydroxytetrahydrobiopterin to quinoid-dihydrobiopterin, which is the first of two steps in the regeneration of tetrahydrobiopterin (BH_4) (Bailey *et al.*, 1993) (Figure 1). BH_4 is an essential cofactor for several mono-oxygenases, such as the aromatic amino acid hydroxylases (Nichol *et al.*, 1985), glycerol ether mono-oxygenase (Kaufman *et al.*, 1990) and nitric oxide synthase (Bredt and Snyder, 1994). The three aromatic amino acid hydroxylases convert phenylalanine, tyrosine and tryptophan to tyrosine, dihydroxyphenylalanine and 5-hydroxytryptophan, respectively. Since these hydroxylases synthesize the precursors of the neurotransmitters dopamine, epinephrine, norepinephrine and serotonin, a lack of the cofactor BH_4 causes neurological disorders and mental retardation which are the clinical symptoms of hyperphenylalaninemia or phenylketonuria (Scriver *et al.*, 1989). It has long been recognized that this inherited disease is caused by a malfunction of the phenylalanine catabolism due to a deficiency either of the phenylalanine hydroxylase (PAH) or of one of the enzymes involved in the *de novo* synthesis or regeneration of BH_4 (Scriver *et al.*, 1989; Blau *et al.*, 1993). A rare, mild form of hyperphenylalaninemia is due to the absence of PCD activity, which is characterized by excretion of 7-substituted pterins in the urine of affected patients (Curtius *et al.*, 1990; Davis *et al.*, 1991; Adler *et al.*, 1992; Citron *et al.*, 1993). In a non-enzymatic process the 4a-hydroxytetrahydrobiopterin isomerizes to 7-substituted pterins which act as inhibitors of PAH (Adler *et al.*, 1992; Davis *et al.*, 1992b). In addition, the depigmentation disorder vitiligo is found to be related to a deficiency of PCD in the affected epidermis. The resulting accumulation of 7-substituted pterins in the epidermal cells is proposed to inhibit the PAH catalyzed synthesis of tyrosine and therefore the biosynthesis of melanin pigments (Schallreuter *et al.*, 1994).

The bifunctional character of PCD became obvious after its sequence was shown to be identical to a protein previously described as dimerization cofactor of HNF1 (DCoH) (Citron *et al.*, 1992; Hauer *et al.*, 1993). DCoH was initially co-purified with the transcription factor hepatocyte nuclear factor 1 (HNF1) from rat liver nuclear extracts (Mendel *et al.*, 1991a). From the 1:1 ratio of the DCoH–HNF1 complex it was concluded that DCoH binds as a dimer to dimeric HNF1 thus forming a hetero-tetrameric complex. Immunoprecipitation experiments showed that DCoH binds to the dimerization domain of HNF1 and v-HNF1. Stabilization of dimeric HNF1 and enhancement of its transcriptional activity were proposed functions of DCoH (Mendel *et al.*, 1991a; Hansen and Crabtree, 1993).

HNF1 (also called LFB1 or HNF-1α) and v-HNF1 (or LFB3, HNF-1β) belong to a family of homo- and hetero-dimeric transcription factors regulating the expression of many genes (De Simone and Cortese, 1991; Tronche and Yaniv, 1992; Hansen and Crabtree, 1993). Three consecutive domains are involved in the DNA binding of HNF1 (Chouard *et al.*, 1990; Nicosia *et al.*, 1990; Mendel *et al.*, 1991b; Tomei *et al.*, 1992). The homo- and hetero-dimerization of HNF1 and v-HNF1 is mediated by the N-terminal 32 amino acid residues and leads to an increased DNA binding affinity. The second domain is a region

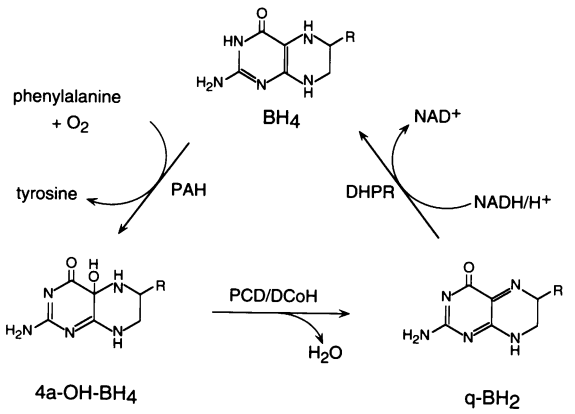

Fig. 1. Role of PCD/DCoH in the regeneration of tetrahydrobiopterin (BH_4, R = 1′,2′-dihydroxypropyl). During hydroxylation of aromatic amino acids, e.g. by phenylalanine hydroxylase (PAH), the cofactor BH_4 is converted to the carbinolamine 4a-hydroxytetrahydrobiopterin (4a-OH-BH_4). BH_4 is regenerated in two steps by pterin-4a-carbinolamine dehydratase (PCD) and dihydropteridine reductase (DHPR).

distantly related to the A-box of the POU-specific domain and followed by an atypical homeodomain of known three dimensional structure (Ceska *et al.*, 1993; Leiting *et al.*, 1993). The latter two are both necessary and sufficient for specific DNA recognition.

Recently, the gene of a prokaryotic homologue (*phhB*) of PCD/DCoH was discovered as part of the phenylalanine hydroxylase operon in *Pseudomonas aeruginosa*. Like the mammalian PCD/DCoH the *phhB* gene product is bifunctional. It shows PCD activity and apparently controls the expression of the prokaryotic phenylalanine hydroxylase (Zhao *et al.*, 1994).

The cDNAs of DCoH/PCD from human, rat and mouse encode a protein of 104 amino acids (Mendel *et al.*, 1991a), but the protein expressed in liver is post-translationally modified. The N-terminal methionine is removed and the new N-terminal residue, alanine, becomes acetylated (Hauer *et al.*, 1993).

The functional enzyme PCD isolated from liver as well as overexpressed in *Escherichia coli* exists exclusively as a homo-tetramer in solution (Huang *et al.*, 1973; Hauer *et al.*, 1993; Ficner *et al.*, 1995). The recombinant PCD has catalytic activity comparable with that of PCD purified from liver (Citron *et al.*, 1992; Thöny *et al.*, 1993).

Here we present the X-ray crystal structure at 2.7 Å resolution of the tetrameric form of recombinant PCD/DCoH from rat liver, whose sequence is identical to the human protein. Implications for the active site (based on the protein structure) and the interaction with HNF1 will be discussed.

Results

The crystallized PCD/DCoH is a homo-tetramer of 222 symmetry. Each monomer folds into a compact, single domain which can be classified as an open-faced α−β sandwich structure (Richardson and Richardson, 1989; Orengo and Thornton, 1993). The sequential connectivity of the three α-helices and four β-strands follows the scheme α1-β1-β2-α2-β3-β4-α3 (Figure 2). The α-helices are packed from one side against the antiparallel, four-

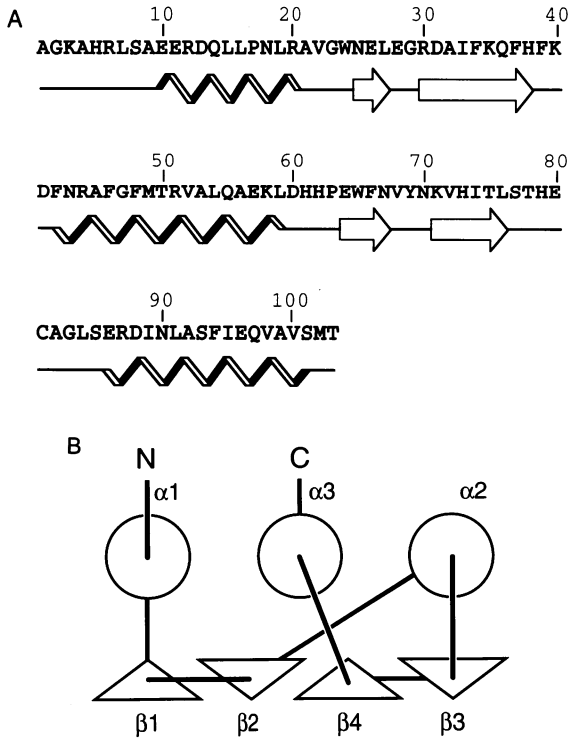

Fig. 2. (**A**) Amino acid sequence and secondary structure elements of PCD/DCoH. The secondary structure was assigned by DSSP (Kabsch and Sander, 1983). (**B**) Topology cartoon of PCD/DCoH.

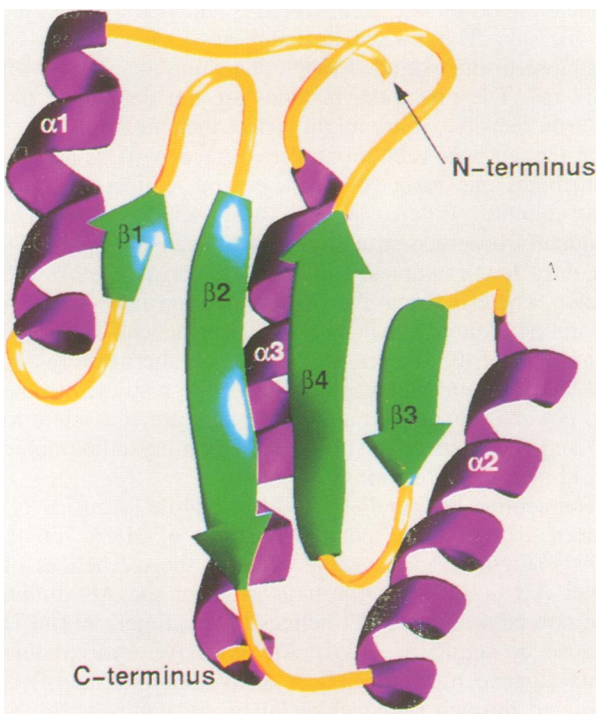

Fig. 3. Ribbon rendering of a PCD/DCoH monomer. Three α-helices are packed from one side against a four stranded, antiparallel β-sheet. The figure was generated using RIBBONS (Carson, 1987).

stranded β-pleated sheet which displays the usual left-handed twist (Figure 3). The helices are aligned almost parallel with respect to the strands of the β-sheet. Helices α2 and α3 together with strands β2, β3 and β4 form a

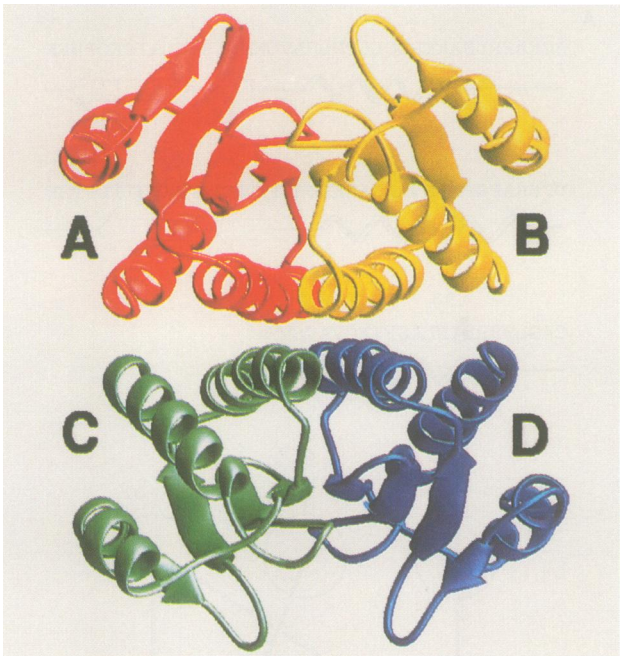

Fig. 4. RIBBONS (Carson, 1987) representation of a PCD/DCoH tetramer. The tetramer has 222 symmetry and each monomer contributes its α2 helix to a central antiparallel four helix bundle. The tetramer can be viewed as a dimer of dimers with extended contacts between monomers A and B (shown in red and yellow) and monomers C and D (green and blue).

compact hydrophobic core, which is shielded by helix α1 on one side. The N-terminal tail formed by residues 4–8 lies close to the extended loop connecting strand β4 with helix α3. The first three residues are not defined in the electron density, which might reflect their flexibility.

In the tetramer each monomer (termed A, B, C and D) contributes one helix (α2) to a central, antiparallel four helix bundle (Figure 4). The four helix bundle is a frequently observed structural motif (Harris *et al.*, 1994), but only few examples, e.g. p53 (Lee *et al.*, 1994), are known where it forms the oligomerization interface of a tetrameric protein. The PCD/DCoH tetramer can be viewed as a dimer of the dimers AB and CD or alternatively AC and BD. The subsequent analysis will be restricted to the two possible dimers AB and AC, which are equivalent to CD and BD, respectively, due to the non-crystallographic 2-fold symmetry relations.

The central, antiparallel four helix bundle forms an 'X' pattern according to the classification of Harris *et al.* (1994) (Figure 5). The angle between the α2 helices of dimer AB is ~172°, while it is 142° for the AC dimer. The non-adjacent, parallel helices of monomers A and D enclose an angle of −39°. An extensive hydrophobic contact between the α2 helices of monomers A and B is mediated through the stacking of phenylalanine residues 42 and 46 of both monomers (residue numbers according to Hauer *et al.*, 1993). Dimer AB is additionally stabilized through the formation of a continuous eight-stranded, antiparallel β-sheet providing four main-chain hydrogen bonds between residues Trp65 and Asn67 of both monomers. The eight-stranded β-sheet spans an arch bordered by two protruding loops, each with a Glu residue on its tip (Figure 6). Remarkably, the hydrophobic residues Leu27, Phe34, Phe66 and Val68 are highly solvent, exposed and form extended hydrophobic patches on the concave surface of the eight-stranded β-sheet. Less extensive contacts are formed between monomers A and C. In addition to hydrophobic contacts several polar interactions involve residues Asn43(A)−Arg51(C), Thr50(A)− Thr50(C), and Arg51(A)−Asn43(C). Furthermore, electrostatic contacts involving Glu57 and Lys58 exist between monomers A and D or B and C. Upon association of monomers A and B a common surface area of ~750 Å^2 is removed from solvent. This is in good agreement with interface values found for dimers of comparable molecular weight (Miller *et al.*, 1987). The buried surface area of the alternative dimer AC is 534 Å^2 which is ~220 Å smaller than expected. Hence, the dimer AB is likely to be more stable and possibly the interaction partner of dimeric HNF1.

The asymmetric unit of the crystal cell contains two PCD/DCoH tetramers. Helix α1 of a monomer belonging to the second tetramer in the asymmetric unit, packs tightly against the concave eight-stranded β-sheet formed by monomers A and B, whereas the concave β-sheet of monomers C and D remains entirely solvent exposed.The eight monomers of the asymmetric unit show only subtle structural variations with root mean square (r.m.s.) differences of their Cα positions ranging from 0.4 to 0.6 Å. The most obvious difference concerns helix α1 where a proline residue (Pro18) introduces flexibility, which allows different degrees of helix bending in each monomer depending on its environment within the crystal.

The structure of the PCD monomer was compared with all protein structures deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) using a distance matrices based search (Holm and Sander, 1993). Several distantly related structures were found which contained either the complete or partial folding motif of PCD/DCoH. The best agreement is seen for part of domain II of the biotin synthetase/biotin repressor (Wilson *et al.*, 1992) which shows a similar topology, but the coordinates r.m.s. difference is 3.6 Å for 92 Cα atoms.

Discussion

Structural implications for enzymatic activity

The structure presented is that of a tetrameric PCD apoenzyme. Only residue Cys81 has so far been reported as a tentative active site residue, but there are recent contradictory experimental results. In one case two mutations were found in the alleles of a PCD gene from a hyperphenylalaninemic child who excretes large amounts of 7-substituted pterins (Citron *et al.*, 1993). The missense mutation in one allele exchanges Cys81 for Arg and by a nonsense mutation in the second allele the codon for Glu86 is replaced by a premature termination codon. In the latter mutation it seems unlikely that the resulting truncated protein, which is missing the entire C-terminal helix α3, can maintain a stable structure with enzymatic activity. The observed hyperphenylalaninemia in this child leads to the conclusion that the mutation of the only Cys residue to Arg causes an inactive enzyme. However, recent experiments show that recombinant PCD with Cys81 mutated to Arg is only slightly less active than the wild type enzyme (S.Köster, B.Thöny, P.Macheroux, H.C.Curtius, C.W.Heizmann, W.Pfleiderer and S.Ghisla, in preparation).

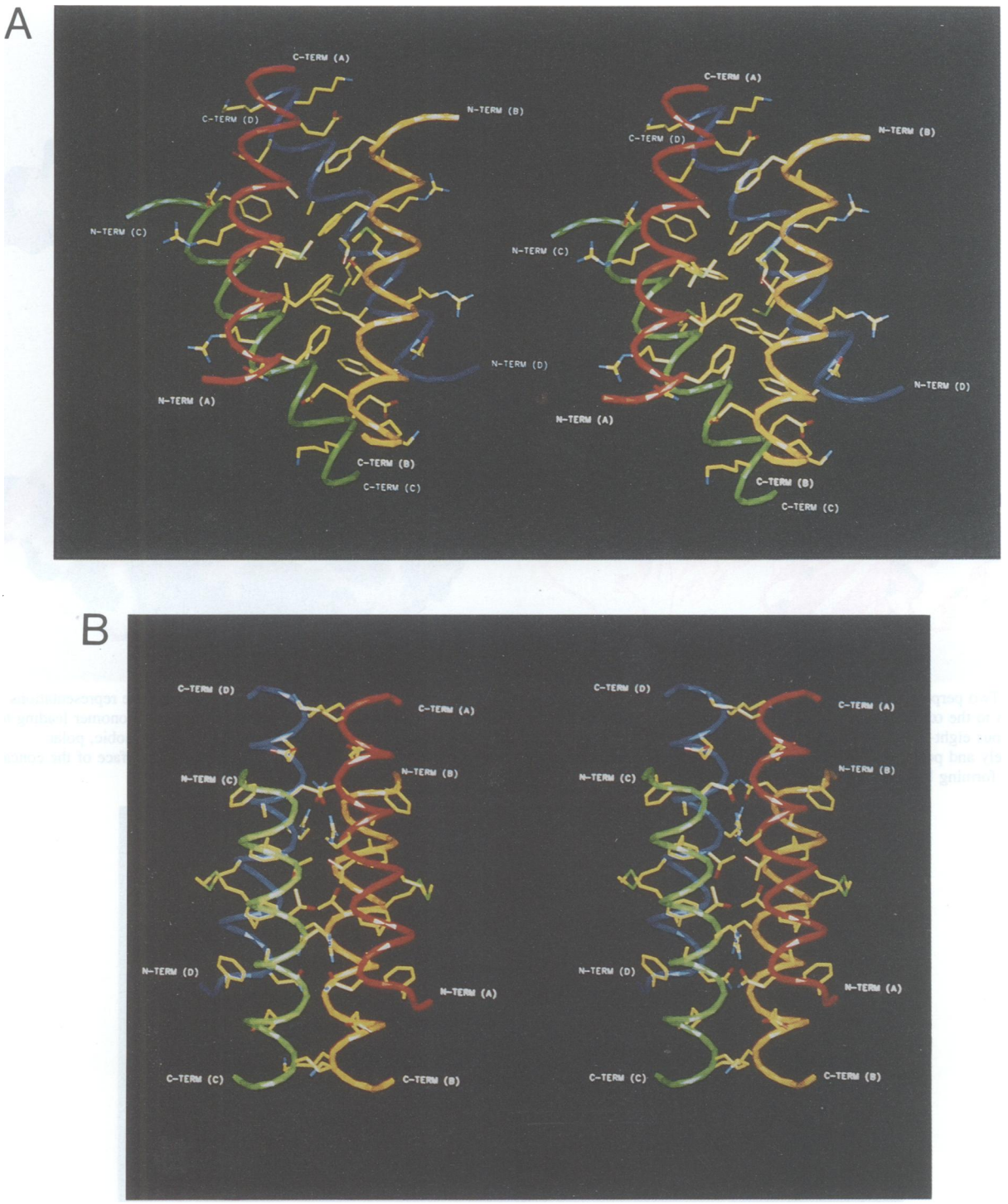

Fig. 5. Two perpendicular stereo views of the four helix bundle in the PCD/DCoH tetramer. Each monomer contributes its α2 helix to a central antiparallel four helix bundle. The α2 helices of monomers A and B (shown in yellow and red, respectively) and the corresponding helices of monomers C and D (green and blue, respectively) interact tightly via stacking of Phe residues 42 and 46 of both subunits. Several polar interactions exist between the helices of monomer A and C or B and D, involving residues Asn43, Thr50 and Arg51.

In addition, modifying the Cys residue chemically does not inactivate the enzyme (S.Köster *et al.*, in preparation), suggesting that Cys81 is not directly involved in catalysis, but presumably located close to the substrate binding site.

Another hint as to the location of the active site might come from fluorescence spectroscopic data. Besides the substrate 4a-hydroxytetrahydrobiopterin several neo- and biopterines bind to PCD and significantly change the fluorescence of one tryptophan residue (S.Köster *et al.*, in preparation). The observed quenching effect indicates a change in the environment of a tryptophan residue by either direct interaction with the bound pterin or indirectly due to a conformational change of the protein induced by the pterin. The two tryptophan residues of PCD, Trp24 and Trp65, are both part of the hydrophobic core of the monomer. Trp24 terminates the loop connecting helix α1

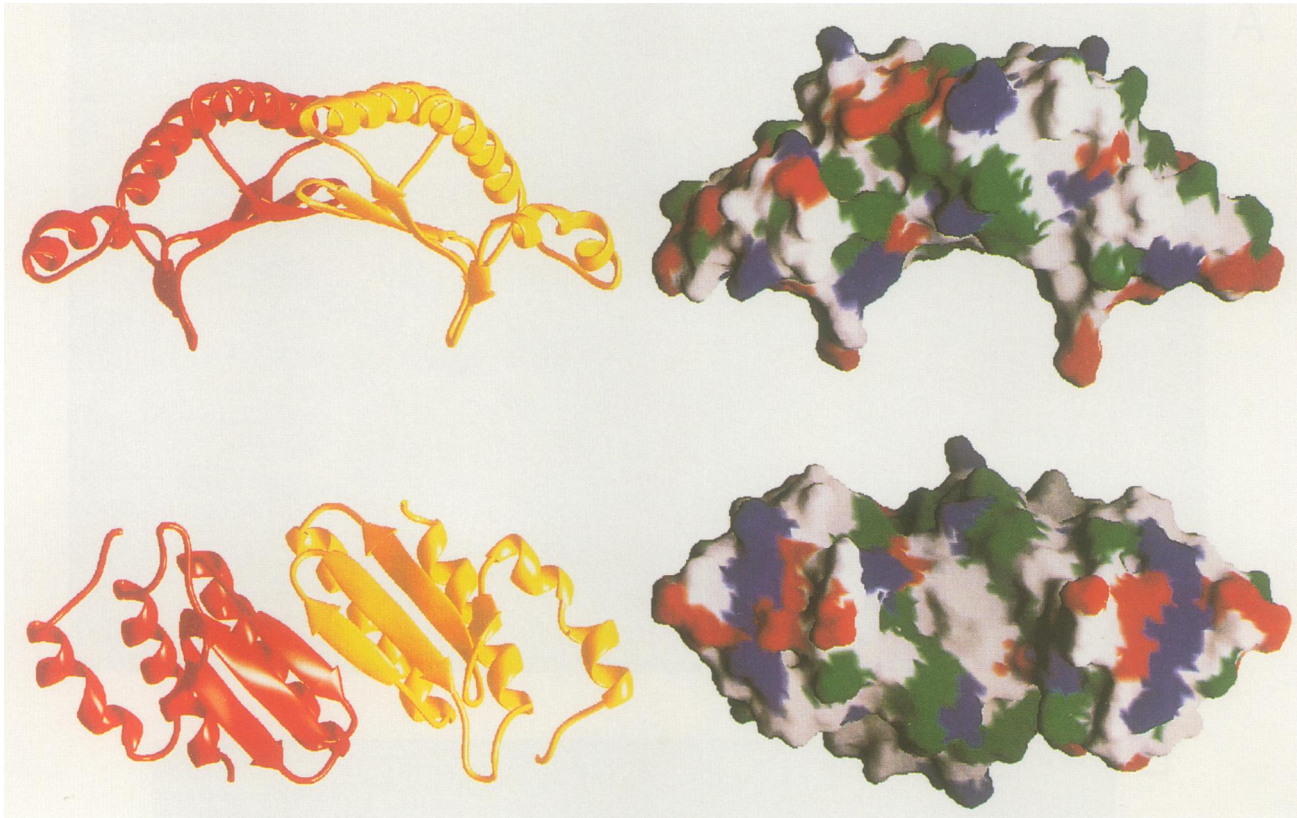

Fig. 6. Two perpendicular views of the AB dimer as RIBBONS (Carson, 1987) plots and GRASP (Nicholls *et al.*, 1991) surface representations. In addition to the α2 helix contacts the dimer is stabilized by the formation of hydrogen bonds between the β-sheets β3 of each monomer leading to a continuous eight-stranded, antiparallel β-sheet. In the surface presentation the colours grey, green, red and blue indicate hydrophobic, polar, negatively and positively charged surface residues, respectively. Four Phe, two Leu and two Val residues are exposed on the surface of the concave β-sheet forming large hydrophobic patches.

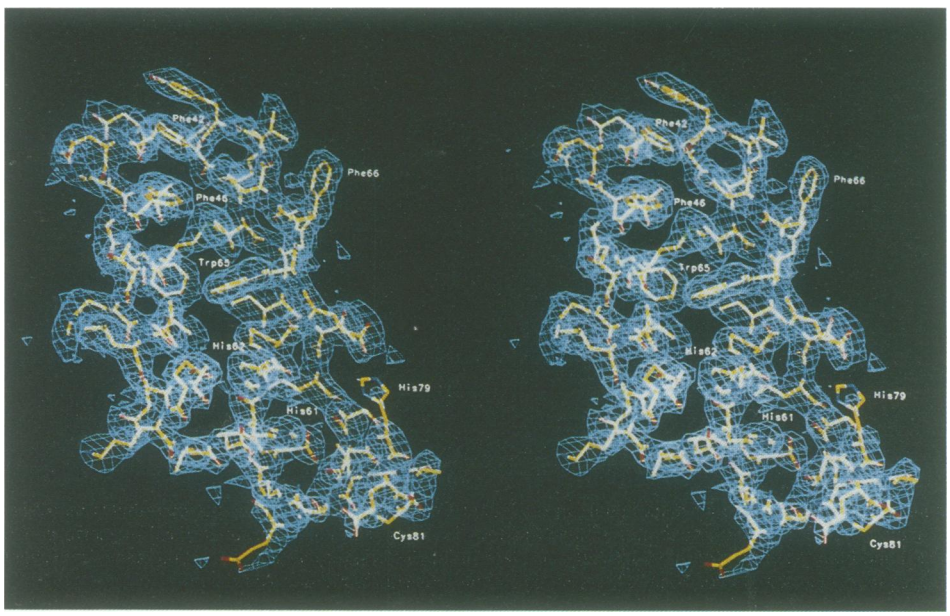

Fig. 7. Stereo view of a $2|F_o|-|F_c|$ electron density map contoured at 1.5σ in the region of α-helix α2 and β-strand β3 which are both part of the AB dimer interface. His61 and His62 are located in the loop connecting helix α2 and strand β3, and His79 and Cys81 are part of the loop connecting strand β4 and helix α3. The side chain of His79 is not defined in the electron density reflecting its flexibility. The putative active site is located in the cleft between these two loops.

and strand β1 and reaches into the hydrophobic core between helix α3 and strand β2. Trp65 is part of strand β3 and is completely excluded from the solvent accessible surface of the tetramer.

The comparison of the crystal structure of PCD/DCoH with three other pterin binding enzymes of known structure, dihydrofolate reductase (DHFR; Bolin *et al.*, 1982; Brown *et al.*, 1993), dihydropteridine reductase (DHPR;

Table I. Data collection and phasing statistics

	Native 1	Native 2	$EtHgPO_4$	CMPS	MMA	$HgCl_2$
Temperature	4°C	−173°C	−173°C	−173°C	−173°C	−173°C
Resolution (Å)	15−2.7	15−3.3	15−3.3	15−3.3	15−5.5	15−5.5
Observed reflections	97383	38688	41996	28492	15254	11549
Unique reflections	34118	16248	16691	12937	4001	3769
Completeness	0.94	0.86	0.87	0.69	0.96	0.88
R-sym	0.089	0.060	0.119	0.103	0.135	0.158
R_{iso}			0.199	0.201	0.248	0.271
R_{Cullis}			0.64	0.64	0.60	0.64
Phasing power			1.78	1.86	2.28	2.45

R-sym, $\Sigma_h \Sigma_i | I_{ih} - \langle I_h \rangle | / \Sigma_h \Sigma_i \langle I_h \rangle$, where $\langle I_h \rangle$ is the mean intensity of the *i* observations of reflection *h*.
R_{iso}, mean isomorphous difference $\Sigma | |F_{PH}| - |F_H| | / |F_P|$
R_{Cullis}, $\Sigma | |F_{PH} - F_P| - |F_H| | / \Sigma |F_{PH} - F_P|$
Phasing power, $\Sigma |F_H| / \Sigma (||F_P| e^{(i\alpha)} + F_H| - |F_{PH}|)$
where $|F_{PH}|$ and $|F_P|$ are the observed derivative and native structure factor amplitudes, F_H is the calculated structure factor amplitude, and F is the calculated phase angle.

Varughese *et al.*, 1992) and 6-pyruvoyl tetrahydropterin synthase (PTPS; Nar *et al.*, 1994) gives almost no hint as to the location of the active site of PCD. The only common pterin binding motif exists between DHFR and PTPS (H.Nar, personal communication) and is the interaction of the carboxylate group of either an Asp or Glu residue with the pterin NH_2 group at ring position 2 and the ring N−H at position 3.

The analysis of the PCD surface reveals a possible pterin binding site. It is located between the two loops connecting helix α2 with strand β3 and strand β4 with helix α3 (Figure 7). The former loop contributes the residues Asp60, His61, His62, Pro63 and Glu64 to the tentative active site, the latter the residues Leu76, His79 and Glu80. The side chain of Cys81 is not part of the tentative pterin binding site, but is located in the vicinity, and its solvent accessible sulfhydryl group is the binding site for the mercury compounds used for the X-ray structure determination. Replacing Cys81 with Arg might cause structural changes in the loop affecting the position of His79 and consequently reducing the enzymatic activity, as observed in the PCD mutant. Remarkably, the possible catalytic centre is formed by the three histidine residues His61, His62 and His79, each of which could catalyse the dehydration reaction. The residues Asp60, Glu64 and Glu80 surround the putative active site, probably facilitating pterin−protein interactions. The distance from the centre of the putative active site to the centre of the indole ring of Trp65 is ~10 Å, which could explain the observed quenching of the tryptophan fluorescence by an energy transfer process. The location of the proposed active site is additionally supported by the alignment of the mammalian PCD sequence to the PCD sequence of *P.aeruginosa* (Zhao *et al.*, 1994), where His61, His62, Pro63 and His79 are conserved.

Tetramer stability and interaction with HNF1

Initially DCoH was isolated from rat liver nuclear extracts in complex with HNF1. The DCoH−HNF1 complex was shown to contain both proteins in a 1:1 ratio, suggesting a hetero-tetrameric composition based on the known dimerization of HNF1 (Mendel *et al.*, 1991a). A hetero-tetramer consisting of a DCoH and a HNF1 dimer is also consistent with gel filtration experiments of the complex obtained by co-expression of PCD/DCoH together with the DNA binding domain of HNF1 in *E.coli* (data not shown). The enzyme PCD, either isolated from liver, or the recombinant form(PCD/DCoH) prepared from *E.coli*, exists exclusively as a stable homo-tetramer. Hence, formation of the DCoH−HNF1 complex either induces or requires a dissociation of PCD into dimers. Two alternative dimers, one formed by the monomers A and B, the other by monomers A and C, are possible (Figure 4). The interface of the putative dimer AB consists of the hydrogen bonded network of the common β-sheet and an almost perfect stacking of four phenylalanine side chains between the two, non-crystallographically related α2 helices from each monomer (Figure 5). Monomers A and B cover a common surface that is ~220 $Å^2$ larger than the one covered by the alternative dimer AC. The larger buried surface area as well as the side- and main-chain interactions suggest that the dimer AB is the more stable one. Furthermore, the four glycine residues at position 47 of each monomer generate a central cavity within the four helix bundle weakening the interaction between the AC and BD monomers. Dissociation of the tetramer into dimers must be accompanied by a disruption of the central four helix bundle and will most likely lead to the formation of the AB and CD dimers.

The surface of the dimer AB offers two possible docking sites for the dimerization domain of HNF1 (Figure 6). One possibility arises following the disruption of the central four helix bundle, which could be re-established as a mixed four helix bundle formed by DCoH and HNF1. Peptides representing the N-terminal 32 residue

Table II. Refinement statistics

Resolution (Å)	10.0–2.7
Number of atoms (non-hydrogen)	6504
Number of reflections (working set)	27065
Working *R*-factor	0.205
Number of reflections (test set)	3061
R_{free}	0.279
Deviations from ideal geometry	
bonds (Å)	0.007
angles (°)	1.6
Mean temperature factor ($Å^2$)	20.3

The R_{free} (Brünger, 1992b) was calculated from a random selection of reflections constituting ~10% of the data, the working *R*-factor was calculated with the remaining intensities.

dimerization domain of HNF1 were shown to fold into two helical segments and apparently form a four helix bundle in solution (Pastore *et al.*, 1991, 1992). Therefore, a mixed four helix bundle might be a favourable way for HNF1 to interact with DCoH.

Alternatively, HNF1 could bind to the concave, eight-stranded β-sheet of the AB dimer (Figure 6). The interaction of the β-sheet with an α-helix is observed in the crystal packing of the PCD/DCoH. One tetramer of the asymmetric unit packs with one of its α1 helices tightly against the β-sheet of the second tetramer. Although we cannot exclude this as the docking site for HNF1, our current experimental data concerning the formation of the HNF1−DCoH complex do not support this. The HNF1−DCoH complex was isolated from liver and characterized by co-expression in an *in vitro* reticulocyte system (Mendel *et al.*, 1991a) and was obtained as well by co-expression in *E.coli* (data not shown), whereas no complex is formed *in vitro* by mixing dimeric HNF1 and tetrameric PCD/DCoH, both independently purified (data not shown). This might indicate that the easily accessible concave β-sheet surface of PCD is not the docking site for HNF1.

The concave shape of the eight-stranded antiparallel β-sheet with two protruding loops on either side and the presence of several hydrophobic residues on its solvent exposed surface is reminiscent of the TATA-box binding protein (TBP) (Nikolov *et al.*, 1992). This resemblance together with the proposed regulatory function of DCoH on HNF1-controlled gene expression (Hansen and Crabtree, 1993) raises the question as to whether PCD/DCoH interacts with DNA. The putative regulative involvement of DCoH in the transcription of specific genes is strongly supported by a recent report showing that the PCD/DCoH homologue found in *P.aeruginosa* is required for the expression of prokaryotic PAH (Zhao *et al.*, 1994). Whether this prokaryotic PCD interacts directly with DNA or with another DNA binding protein is not yet known. Comparing the composition and location of charged, polar and hydrophobic residues located on the β-sheet surface of PCD/DCoH with the corresponding surface in the TBP−DNA complex (Kim and Burley, 1994), deems it unlikely that DCoH interacts with DNA. In addition, there is currently no experimental evidence for DCoH binding to DNA. However, the shape of the eight-stranded β-sheet and the extended hydrophobic patches on its surface clearly indicate a docking site for a biological macromolecule.

Materials and methods

Crystallization and data collection

Recombinant rat liver PCD/DCoH was expressed in *E.coli*, purified and crystallized as described previously (Ficner *et al.*, 1995). Crystals belong to space group $P3_221$ with unit cell dimensions at 4°C of $a = b = 106.2$ Å, $c = 197.1$ Å, $\gamma = 120°$ and at −173°C of $a = b = 103.8$ Å, $c = 195.3$ Å, $\gamma = 120°$.

A complete data set from a single native crystal was collected at 4°C using synchrotron radiation ($\lambda = 1.0$ Å), recorded on the prototype MAR research 22 cm image plate detector mounted on the EMBL X31 beam line at DESY, Hamburg. In an extensive search for heavy atom derivatives, crystals were soaked in a variety of heavy atom compounds. Diffraction data of soaked crystals were collected on a rotating anode X-ray generator at a wavelength of 1.54 Å and cooling to 4°C on a MAR research 30 cm image plate detector. For processable data sets, the isomorphous differences to the native data set were negligible. However, crystals soaked with four different mercury compounds [*p*-chloromercuriphenylsulfonic acid (CMPS), ethylmercuric phosphate ($EtHgPO_4$), mercuric chloride ($HgCl_2$), methylmercuric acetate (MMA)] became extremely sensitive to X-ray exposure, leading to a rapid decay of their diffraction patterns at 4°C and complete data sets could not be collected. In order to extend their diffraction life time, further data collection was performed at −173°C using the Oxford cryostream cooler (Oxford Cryosystems). To facilitate flash freezing, crystals were stepwise equilibrated with a cryo-buffer containing 20% sucrose in addition to 65% saturated ammonium sulfate buffered with 0.1 M MES−NaOH at pH 6.4. The mercury compounds were added at 0.5–1.0 mM concentrations to the cryo-buffer. Crystals were mounted free standing in a film (Teng, 1990) supported by a rayon fibre loop of ~1 mm diameter and flash-cooled in a dry nitrogen stream at −173°C. In this manner complete data sets of one native crystal as well as crystals soaked in the aforementioned four mercury compounds could be recorded with isomorphous differences suitable for use as heavy atom derivatives (Table I). Cryo data to a resolution of 3.3 Å were collected on a rotating anode generator ($\lambda = 1.54$ Å) using focusing mirrors and detected on a Siemens Xentronics X100-A area detector.

Structure determination

All diffraction data were processed and scaled with XDS (Kabsch, 1988). Crystallographic computing was carried out using the CCP4 program suite (Collaborative Computational Project, 1994) unless otherwise stated. The structure was solved by a combination of multiple isomorphous replacement (MIR) and density modification techniques including solvent flattening, histogram matching as well as additional 8-fold non-crystallographic symmetry (NCS) averaging.

The difference Patterson maps of the $EtHgPO_4$ and $HgCl_2$ derivatives were interpreted using the program VERIFY (S.Roderick, unpublished) using data from 15 to 5 Å resolution. Eight unique mercury atom positions were identified suggesting two tetramers per asymmetric unit. Assuming eight monomers (8 × 11.9 kDa) in the asymmetric unit leads to a V_m coefficient of 3.3 Å^3/Da, corresponding to 65% solvent content which agrees well with values typical for protein crystals (Matthews, 1968). The heavy atom positions were confirmed and refined by the programs VECREF and MLPHARE, and single isomorphous replacement (SIR) phases were calculated at 5 Å. No additional sites were found either in the difference Patterson or in the difference Fourier map. Cross-phased Fourier maps showed the same eight positions for all the mercury derivatives. The thermal parameters of the heavy atoms were set to 30 Å^2 and the occupancies and positions were refined with MLPHARE. Differences between the four derivatives mainly concern the occupancies of corresponding positions. The final phasing statistics are summarized in Table I. A 3.3 Å MIR map was calculated and drastically improved by solvent flattening and histogram matching using the β release of the program DM (K.D.Cowtan, unpublished). The first solvent flattened MIR map, calculated in space group $P3_121$, showed electron density for left handed α-helices, hence, the calculations were repeated by changing the space group to the enantiomorphic space group $P3_221$ and inverting the signs of the heavy atom coordinates.

Initial polyalanine fragments of the eight monomers were manually built using O (Jones *et al.*, 1991). Based on these fragments the NCS operators were determined. A mask with 5 Å radius around one monomer was created and 8-fold NCS averaging of the electron density map was performed using MAMA and RAVE (Kleywegt and Jones, 1994) resulting in an improved electron density map to 3.3 Å resolution. Additional residues could be included and a partial model of 83 alanine residues per monomer was subjected to rigid-body refinement and subsequent Powell minimization applying the force field parameters from Engh and Huber (1991) and NCS restraints in XPLOR (Brünger, 1992a), which reduced the initial *R*-factor from 46.0 to 34.5%. Model phases were combined with MIR phases by using the program SIGMAA and the averaging procedure was repeated. In the second round of model building the side chains of 97 residues could be positioned into the improved density and subsequent refinement reduced the *R*-factor to 31.0%. For further refinement and model building the 2.7 Å native data set collected at 4°C was used. Since the cell dimensions for the 4°C data set were slightly larger compared with the cell constants of crystals collected at −173°C, the orientation of the previous model was first optimized by a rigid body refinement against the 2.7 Å data. After subsequent positional refinement an improved electron density map at 2.7 Å resolution was obtained which allowed three additional residues to be included. Further refinement of the new model reduced the crystallographic *R*-factor to 22.3% and the free *R*-value (Brünger, 1992b) to 26.7%, including data from 10.0 to 2.7 Å. After releasing the NCS restraints the corresponding

current R-values are 20.5 and 27.9%, respectively. At present, the model comprises residues 4–103 for each monomer, but no solvent molecules are included. The coordinates have been deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The current refinement parameters are summarized in Table II. The dihedral angles of the polypeptide backbone all lie in allowed regions of the Ramachandran plot. The quality of the current model was analysed with the program PROCHECK (Laskowski *et al.*, 1993) revealing that at 2.7 Å resolution all parameters show smaller than expected deviations from the ideal. Interhelical angles were calculated with the program EDPDB (C.X.-J. Zhang, unpublished). Screening the Brookhaven Protein Data Bank for structurally related proteins was done with the program DALI (Holm and Sander, 1993) keeping the sequential order of the secondary structure elements.

Acknowledgements

We are grateful to B.Thöny, C.W.Heizmann (Zürich) and S.Ghisla (Konstanz) for providing us with unpublished biochemical data of wildtype and mutated PCD and for stimulating discussions. We thank M.Yaniv (Paris) for providing us with PCD/DCoH cDNA. S.Roderick (New York) and C. X.-J.Zhang (Eugene) supplied us with the programs VERIFY and EDPDB, respectively. We particulary acknowledge the permanent support of our colleague T.A.Ceska (Heidelberg), the help of L.Holm (Heidelberg) with the program DALI, and the support of the staff at the EMBL outstation (Hamburg) during data collection.

References

Adler,C., Ghisla,S., Rebrin,I., Haavik,J., Heizmann,C.W., Blau,N., Kuster,T. and Curtius, H.-C. (1992) 7-substituted pterins in humans with suspected pterin-4a-carbinolamine dehydratase deficiency. *Eur. J. Biochem.*, **208**, 139–144.

Bailey,S.W., Boerth,S.R., Dillard,S.B. and Ayling,J.E. (1993) The mechanism of cofactor regeneration during phenylalanine hydroxylation. In Ayling,J.E., Nair,M.G. and Bough,C.M. (eds), *Chemistry and Biology of Pteridines and Folates*. Plenum Press, New York, pp. 47–54.

Bernstein,F.C., Koetzle,T., Williams,G., Meyer,E.,Jr, Brice,M., Rodgers,J., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: A computer-based archival-file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.

Blau,N., Thöny,B., Heizmann,C.W. and Dhondt,J.-L. (1993) Tetrahydrobiopterin deficiency: from phenotype to genotype. *Pteridines*, **4**, 1–10.

Bolin,J.T., Filman,D.J., Matthews,D.A., Hamlin,R.C. and Kraut,J. (1982) Crystal structures of *Escherichia coli* and *Lactobacillus casei* dihydrofolate reductase refined at 1.7 Å resolution. I. General features and binding of methotrexate. *J. Biol. Chem.*, **257**, 13650–13662.

Bredt,D.S. and Snyder,S.H. (1994) Nitric oxide: a physiologic messenger molecule. *Annu. Rev. Biochem.*, **63**, 175–195.

Brown,K.A., Howell,E.E. and Kraut,J. (1993) Long-range structural effects in a second-site revertant of a mutant dihydrofolate reductase. *Proc. Natl Acad. Sci. USA*, **90**, 11753–11756.

Brünger,A.T. (1992a) X-PLOR, Version 3.1 Yale University Press, New Haven, CT.

Brünger,A.T. (1992b) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Carson,M. (1987) Ribbon models of macromolecules. *J. Mol. Graphics*, **5**, 103–106.

Ceska,T.A., Lamers,M., Monaci,P., Nicosia,A., Cortese,R. and Suck,D. (1993) The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.*, **12**, 1805–1810.

Chouard,T., Blumenfeld,M., Bach,I., Vandekerckhove,J., Cereghini,S. and Yaniv,M. (1990) A distal dimerization domain is essential for DNA binding by the atypical HNF1 homeodomain. *Nucleic Acids Res.*, **18**, 5853–5863.

Citron,B.A., Davis,M.D., Milstien,S., Gutierrez,J., Mendel,D.B., Crabtree,G.R. and Kaufman,S. (1992) Identity of 4a-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transregulator of homeodomain proteins. *Proc. Natl Acad. Sci. USA*, **89**, 11891–11894.

Citron,B.A., Kaufman,S., Milstien,S., Naylor,E.W., Greene,C.L. and Davis,M.D. (1993) Mutation in the 4a-carbinolamine dehydratase gene leads to mild hyperphenylalaninemia with defective cofactor metabolism. *Am. J. Hum. Genet.*, **53**, 768–774.

Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Curtius,H.-C., Adler,C., Rebrin,I., Heizmann,C. and Ghisla,S. (1990) 7-substituted pterins: formation during phenylalanine hydroxylation in the absence of dehydratase. *Biochem. Biophys. Res. Commun.*, **172**, 1060–1066.

Davis,M.D., Kaufman,S. and Milstien,S. (1991) Conversion of 6-substituted tetrahydropterins to 7-isomers via phenylalanine hydroxylase-generated intermediates. *Proc. Natl Acad. Sci. USA*, **88**, 385–389.

Davis,M.D., Kaufman,S. and Milstien,S. (1992a) Distribution of 4a-hydroxytetrahydropterin dehydratase in rat tissues. Comparison with the aromatic amino acid hydroxylases. *FEBS Lett.*, **302**, 73–76.

Davis,M.D., Ribeiro,P., Tipper,J. and Kaufman,S. (1992b) '7-Tetrahydrobiopterin,' a naturally occurring analogue of tetrahydrobiopterin, is a cofactor for and a potential inhibitor of the aromatic amino acid hydroxylases. *Proc. Natl Acad. Sci. USA*, **89**, 10109–10113.

De Simone,V. and Cortese,R. (1991) Transcriptional regulation of liver-specific gene expression. *Curr. Opin. Cell Biol.*, **3**, 960–965.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Ficner,R., Sauer,U.H., Ceska,T.A., Stier,G. and Suck,D. (1995) Crystallization and preliminary crystallographic studies of recombinant dimerization cofactor of transcription factor HNF1/pterin-4α-carbinolamine dehydratase from liver. *FEBS Lett.*, **357**, 62–64.

Hansen,L.P. and Crabtree,G.R. (1993) Regulation of the HNF-1 homeodomain proteins by DCoH. *Curr. Opin. Gen. Dev.*, **3**, 246–253.

Harris,N.L., Presnell,S.R. and Cohen,F.E. (1994) Four helix bundle diversity in globular proteins. *J. Mol. Biol.*, **236**, 1356–1368.

Hauer,C.R., Rebrin,I., Thöny,B., Neuheiser,F., Curtius,H.-C., Hunziker,P., Blau,N., Ghisla,S. and Heizmann,C.W. (1993) Phenylalanine hydroxylase-stimulating protein/pterin-4α-carbinolamine dehydratase from rat and human liver. Purification, characterization, and complete amino acid sequence. *J. Biol. Chem.*, **268**, 4828–4831.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Huang,C.Y., Max,E.E. and Kaufman,S. (1973) Purification and characterization of phenylalanine hydroxylase-stimulating protein from rat liver. *J. Biol. Chem.*, **248**, 4235–4241.

Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kabsch,W. (1988) Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallogr.*, **21**, 916–924.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kaufman,S., Pollock,R.J., Summer,G.K., Das,A.K. and Hajra,A.K. (1990) Dependence of an alkyl glycol-ether mono-oxygenase activity upon tetrahydrobiopterins. *Biochim. Biophys. Acta*, **1040**, 19–27.

Kim,J.L. and Burley,S.K. (1994) 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.*, **1**, 638–653

Kleywegt,G.J. and Jones,T.A. (1994) Halloween ... masks and bones. In Bailey.S., Hubbard,R. and Waller,D. (eds), *From First Map to Final Model. Proceedings of the CCP4 study weekend, 6–7 January 1994*. SERC Daresbury Laboratory, Daresbury, UK, pp.59–66.

Laskowski,R.A., McArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lazarus,R.A., Benkovic,S.J. and Kaufman,S. (1983) Phenylalanine hydroxylase stimulator protein is a 4a-carbinolamine dehydratase. *J. Biol. Chem.*, **258**, 10960–10962.

Lee,W., Harvey,T.S., Yin,Y., Yau,P., Litchfield,D. and Arrowsmith,C.H. (1994) Solution structure of the tetrameric minimum transforming domain of p53. *Nature Struct. Biol.*, **1**, 877–890.

Leiting,B., De Francesco,R., Tomei,L., Cortese,R., Otting,G. and Wüthrich,K. (1993) The three-dimensional NMR-solution structure of the polypeptide fragment 195–286 of the LFB1/HNF1 transcription factor from rat liver comprises a non-classical homeodomain. *EMBO J.*, **12**, 1797–1803.

Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.

Mendel,D.B., Khavari,P.A., Conley,P.B., Graves,M.K., Hansen,L.P.,

Admon,A. and Crabtree,G.R. (1991a) Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. *Science*, **254**, 1762–1767.

Mendel,D.B., Hansen,L.P., Graves,M.K., Conley,P.B. and Crabtree,G.R. (1991b) HNF-1α and HNF-1β (vHNF-1) share dimerization and homeo domains, but no activation domains, and form hetero-dimers in vitro. *Genes Dev.*, **5**, 1042–1056.

Miller,S., Lesk,A.M., Janin,J. and Chothia,C. (1987) The accessible surface area and stability of oligomeric proteins. *Nature*, **328**, 834–836.

Nar,H., Huber,R., Heizmann,C.W., Thöny,B. and Bürgisser,D. (1994) Three-dimensional structure of 6-pyruvoyl tetrahydropterin synthase, an enzyme involved in tetrahydrobiopterin biosynthesis. *EMBO J.*, **13**, 1255–1262.

Nichol,C.A., Smith,G.K. and Duch,D.S. (1985) Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. *Annu. Rev. Biochem.*, **54**, 729–764.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein Struct. Funct. Genet.*, **11**, 281–296.

Nicosia,A., Monaci,P., Tomei,L., De Francesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) A myosin-like dimerization helix and an extra-large homeodomain are essential elements of the tripartite DNA binding structure of LFB1. *Cell*, **61**, 1225–1236.

Nikolov,D.B., Hu,S-H., Lin,J., Gasch,A., Hoffmann,A., Horikoshi,M., Chua,N.-H., Roeder,R.G. and Burley,S.K. (1992) Crystal structure of TFIID TATA-box binding protein. *Nature*, **360**, 40–46.

Orengo,C.A. and Thornton,J.M. (1993) Alpha plus beta folds revisited: some favoured motifs. *Structure*, **1**, 105–120.

Pastore,A., De Francesco,R., Barbato,G., Castiglione Morelli,M.A., Motta,A. and Cortese,R. (1991) ^{1}H resonance assignment and secondary structure determination of the dimerization domain of transcription factor LFB1. *Biochemistry*, **30**, 148–153.

Pastore,A., De Francesco,R., Castiglione Morelli,M.A., Nalis,D. and Cortese,R. (1992) The dimerization domain of LFB1/HNF1 related transcription factors: a hidden four helix bundle? *Protein Engng*, **5**, 749–757.

Richardson,J.S. and Richardson,D.C. (1989) Principles and patterns of protein conformation. In Fasman,G.D. (ed.), *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum Press, New York, pp.1–99.

Scriver,C.R., Kaufman,S. and Woo,S.L.C. (1989) The hyperphenylalaninemias. In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The metabolic Basis of Inherited Disease*. McGraw-Hill, New York, pp.495–546.

Schallreuter,K.U., Wood,J.M., Pittelkow,M.R., Gütlich,M., Lemke,K.R., Rödl,W., Swanson,N.N., Hitzemann,K. and Ziegler,I. (1994) Regulation of melanin biosynthesis in the human epidermis by tetrahydrobiopterin. *Science*, **263**, 1444–1446.

Teng,T.Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. *J. Appl. Crystallogr.*, **23**, 387–391.

Thöny,B., Neuheiser,F., Hauer,C. and Heizmann,C.W. (1993) Molecular cloning and recombinant expression of the human liver phenylalanine hydroxylase stimulating factor revealed structural and functional identity to the dimerization cofactor for the nuclear transcription factor HNF1-α. In Ayling,J.E., Nair,M.G. and Bough,C.M. (eds), *Chemistry and Biology of Pteridines and Folates*. Plenum Press, New York, pp.103–106.

Tomei,L., Cortese,R. and De Francesco,R. (1992) A POU-A related region dictates DNA binding specifity of LFB1/HNF1 by orienting the two XL-homeodomains in the dimer. *EMBO J.*, **11**, 4119–4129.

Tronche,F. and Yaniv,M. (1992) HNF1, a homeoprotein member of the hepatic transcription regulatory network. *Bioessays*, **14**, 579–587.

Varughese,K.I., Skinner,M.M., Whiteley,J.M., Matthews,D.A. and Xuong,N.H. (1992) Crystal structure of rat liver dihydropteridine reductase. *Proc. Natl Acad. Sci. USA*, **89**, 6080–6084.

Wilson,K.P., Shewchuck,L.M., Brennan,R.G., Otsuka,A.J. and Matthews,B.W. (1992) *Escherichia coli* biotin holoenzyme synthetase/*bio* repressor crystal structure delineates the biotin- and DNA binding domains. *Proc. Natl Acad. Sci. USA*, **89**, 9257–9261.

Zhao,G., Xia,T., Song,J. and Jensen,R.A. (1994) *Pseudomonas aeruginosa* possesses homologues of mammalian phenylalanine hydroxylase and 4α-carbinolamine dehydratase/DCoH as part of a three-component gene cluster. *Proc. Natl Acad. Sci. USA*, **91**, 1366–1370.

Received on January 25, 1995; revised on February 24, 1995