Hypothesis

Towards a new T-fold protein?: The coproporphyrinogen III oxidase sequence matches many structural features from urate oxidase

Nathalie Colloc'h[a,b,*], Jean-Paul Mornon[b], Jean-Michel Camadro[c]

[a] *Université de Caen, CNRS UMR6551, bd. Henri Becquerel, BP5229, 14074 Caen Cedex, France*
[b] *Systèmes Moléculaires et Biologie Structurale, LMCP, CNRS UMR7590, Universités Paris VI et Paris VII, case 115, 4 place Jussieu, 75252 Paris Cedex 05, France*
[c] *Laboratoire d'Ingénierie des Protéines et Contrôle Métabolique, Département de Biologie des Génomes, Institut Jacques Monod, CNRS UMR7592, Universités Paris VI et Paris VII, 2 place Jussieu, 75251 Paris Cedex 05, France*

Received 3 June 2002; revised 22 July 2002; accepted 23 July 2002

First published online 7 August 2002

Edited by Robert B. Russell

Abstract Urate oxidase (UOX) and coproporphyrinogen III oxidase (CPO) are two unusual oxidases as they accomplish their catalytic act with no co-factor nor metal ion. They both require molecular oxygen, and lead to hydrogen peroxide in addition to the product. UOX is composed of two contiguous Tunneling-fold domains and CPO appears to be also divided into two structurally equivalent domains. Moreover, each of these putative domains can be coherently aligned on UOX domains. Although their sequences are very distant, we therefore suggest that functional CPO dimer is built around a tunnel, with the substrate sitting above it, on the N- and C-terminal side. This overall model is supported by mutation data and is coherent with the chemical events expected for substrate processing by CPO.

Key words: Duplication; Oligomer; Oxygen-dependent; Coproporphyria; Hydrophobic cluster analysis

1. Introduction

The Tunneling-fold (T-fold) is a small architecture composed of an antiparallel β-sheet of four sequential strands with a pair of antiparallel α-helices between the second and third strand. This motif has so far been found in five different enzymes built through multimeric association of T-folds crossed by a tunnel [1]. The structural similarity between T-folds is high, although the sequence divergence between them is remarkably large (often below 10% sequence identity). The T-fold possesses the characteristics of a globular domain, with a hydrophobic core and a topohydrophobic network [1–3]. Since the T-fold constitutes a robust structural family and an efficient scaffold to build various multimeric enzymes (from a three-fold to a five-fold symmetry), many as yet undetected T-folds may exist in sequence databases, although so far, no new T-fold protein has been predicted based on sequence comparison alone. It is thus an exciting challenge to identify a putative new member of this structural family. We suggest here that coproporphyrinogen III oxidase (CPO) may resemble the T-fold protein urate oxidase (UOX).

CPO (EC 1.3.3.3) is the sixth enzyme of the heme biosynthesis pathway. This soluble protein catalyzes the sequential oxidative decarboxylation of two propionate groups of coproporphyrinogen III to two vinyl groups of protoporphyrinogen IX [4,5] (Fig. 1). Oxygen-dependent CPOs appear to be remarkably conserved proteins despite variable subcellular locations ranging from the intermembrane space of mitochondria in mammals, the cytosol in yeast and bacteria, to the stroma of plant chloroplasts. Accordingly, the protein sequences share large areas of similarity except for their targeting sequences, which are removed during the import process [6].

UOX (EC 1.7.3.3, uricase), an enzyme belonging to the purine degradation pathway, catalyzes the oxidation of uric acid to allantoin and hydrogen peroxide. UOX from *Aspergillus flavus* is extracted and purified by Sanofi and used in the prevention and treatment of severe hyperuricemia that may occur during chemotherapy [7,8]. The structure has been solved by X-ray crystallography with isomorphous replacement, at a resolution of 2.05 Å [9].

Oxygen-dependent CPO and UOX require molecular oxygen to accomplish their catalytic act on nitrogenous rings which, very remarkably for oxidases, needs no co-factor nor metal ion [9–11]. Moreover, they have similar sizes (around 40000 g/mol), CPO is a dimer and UOX a dimer of dimer, they both lead to hydrogen peroxide in addition of the product and CPO may act through an intermediate hydroxylation state [12] as UOX. The comparison between their sequences was not suggested by a large screening of databanks, no hit being detected between them by, e.g. the sensitive position-specific iterated BLAST (PSI-Blast) program [13]. And no other valid support was proposed by PSI-Blast when using yeast CPO sequence as a query. It is important to note that none of the presently known T-fold proteins used as a query in PSI-blast allow to detect another T-fold protein [1]. The present study was initiated by a direct comparison between their transposed sequences (hydrophobic cluster analysis (HCA) plot [14]) prompted by the very unusual similarity in their mechanism summarized here above. The structural and

*Corresponding author. Fax: (33)-2-31 56 61 99.
E-mail addresses: n.colloch@neuro.unicaen.fr (N. Colloc'h), mornon@lmcp.jussieu.fr (J.-P. Mornon), camadro@ijm.jussieu.fr (J.-M. Camadro).

Abbreviations: UOX, urate oxidase; CPO, coproporphyrinogen III oxidase; T-fold, Tunneling-fold; HCA, hydrophobic cluster analysis

0014-5793/02/$22.00
PII: S0014-5793(02)03123-X

functional results reported in this study further support this hypothesis.

2. Materials and methods

The sequences of UOX and CPO are listed in the Swiss-Prot database [15]. We selected five UOX sequences (*A. flavus*, *Pichia jardinii* (yeast), *Drosophila melanogaster*, pig and soybean) and four CPO sequences (*Saccharomyces cerevisiae* (yeast), *Salmonella typhimurium*, human and tobacco) showing the greatest dissimilarity within each family. The five UOX proteins share between 31 and 51% sequence identity, and the four CPO proteins share between 45 and 66% sequence identity, homogeneity which is a disadvantage for structure prediction. The coordinates of *A. flavus* UOX are accessible in the Protein Data Bank [16,17] under the entry 1uox.

We used 1D screening methods such as the PSI-Blast on the NCBI non-redundant database (scoring matrix Blosum 62) using an *E*-value cutoff of 0.001 [13]. We also used the sensitive 2D method HCA (for a review, see [18]) to determine whether similarities between the proteins do exist. This method highlights the correspondence between sequence and secondary structure and efficiently detects similarities between highly divergent sequences. It has now been established that, due to its intrinsic power of compactness, the helical net representation offers the best correspondence between the positions of the hydrophobic clusters and regular secondary structures, whether α or β [19]. Similarities between sequences were assessed with the sequence identity level, a similarity score computed using the Blossom 62 matrix, and a hydrophobic matching score (HCA score).

The CPO to UOX alignment was further validated by the construction of a 3D model of yeast CPO based on the atomic coordinates of *A. flavus* UOX. We used through the WWW interface 'Bioinformatics Tools Server' [20] the automated program TITO [21] specialized to work at low levels of sequence identity to optimize the alignment, Modeller [22] to build the model, and ProsaII [23] to validate it. The model has then be energetically minimized using CHARMm [24] through Accelrys software (San Diego, CA, USA).

3. Results and discussion

The monomer of UOX is built with two consecutive T-fold domains, giving rise to an antiparallel β-sheet of eight sequential strands, with the four helices layered on the external concave face of the sheet. The root mean square between the two structurally aligned T-fold domains is 1.5 Å for 99 aligned Cα. The dimer is formed by the association of two monomers related by a two-fold axis [9].

The 2D HCA representation of the yeast CPO sequence showed that it may also be divided into two successive and similar domains, with lengths comparable to those of *A. flavus* UOX (137 and 163 amino acids in UOX, 166 and 162 amino acids in CPO). The resulting sequence identity between the two domains is similar to that observed between the UOX domains (7.4% in UOX, 6.9% in CPO). The 2D representation shows a good conservation of the shape of the hydrophobic clusters (Fig. 2A) and consequently of most of the internal hydrophobic faces of regular secondary structure elements. The 1D alignment deduced from the HCA representation and extended to the family reveals a similarity between the two domains, with many conserved hydrophobic and charged positions and a few identical residues (Fig. 2B). Sequence identity levels range from 6.9% to 9.2%, similarity scores from 33.8% to 36.8%, and HCA scores from 38.4% to 51.5%. These scores are just slightly lower than those in the UOX family, with sequence identity levels between 7.4% and 11.4%, similarity scores between 36.1% and 49.3% and HCA scores between 37.7% and 48.4%.

The CPO sequences were then aligned on the UOX sequences in order to check if there is a possible relationship between these two unusual oxidases. The HCA alignment of *A. flavus* UOX and yeast CPO sequences shows good conservation of hydrophobic clusters and again a few identical residues (Fig. 3A). The deduced sequence alignment of the CPO family sequences with the UOX family shows that the hydrophobic and charged properties of residues are fairly well conserved and distributed along the sequences (Fig. 3B). Half of the topohydrophobic positions characterized in the T-fold family [1] are occupied by hydrophobic residues (14 conserved topohydrophobic positions out of 30). CPO and UOX share between 4% and 10% sequence identity, between 32% and 39% similarity, and between 41% and 53% HCA score. These scores are very similar to those observed between all known T-folds [1]. The active loops in one enzyme are either absent or very different in the other enzyme. The small β-sheet (E1 and E2), involved in the catalytic activity of UOX, is absent in CPO. Similarly, the loops between the putative second strand and the first helix in each domain (loops S2–H1 called A1 and S6–H3 called A2 in Fig. 3), probably involved in the catalytic activity of CPO, are longer and quite different from those in UOX. In CPO, an extra stretch of 25 residues is found at the end of each domain (called C1 and C2). These two extra stretches appear highly similar at the sequence level (see

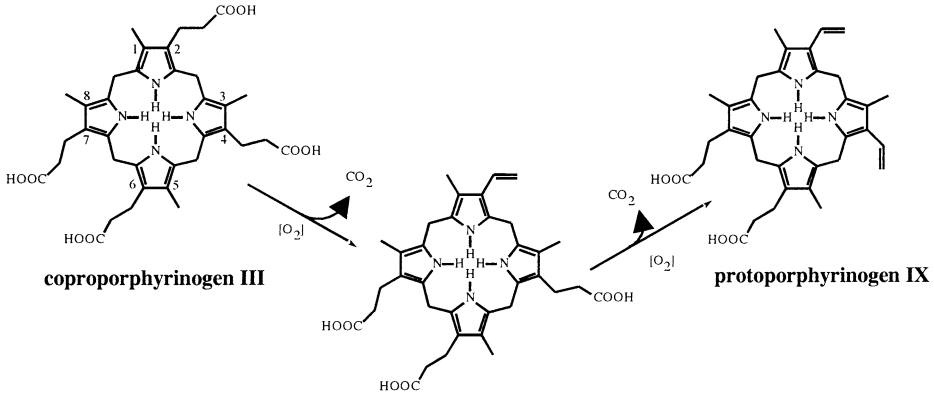

Fig. 1. Reaction mechanism of the stepwise decarboxylation of coproporphyrinogen III to protoporphyrinogen IX. A first propionate group of coproporphyrinogen III is converted to a vinyl group to form the harderoporphyrinogen intermediate, then a second propionate group is converted to a vinyl group to yield the final product, protoporphyrinogen IX. CPO catalyzes both steps.

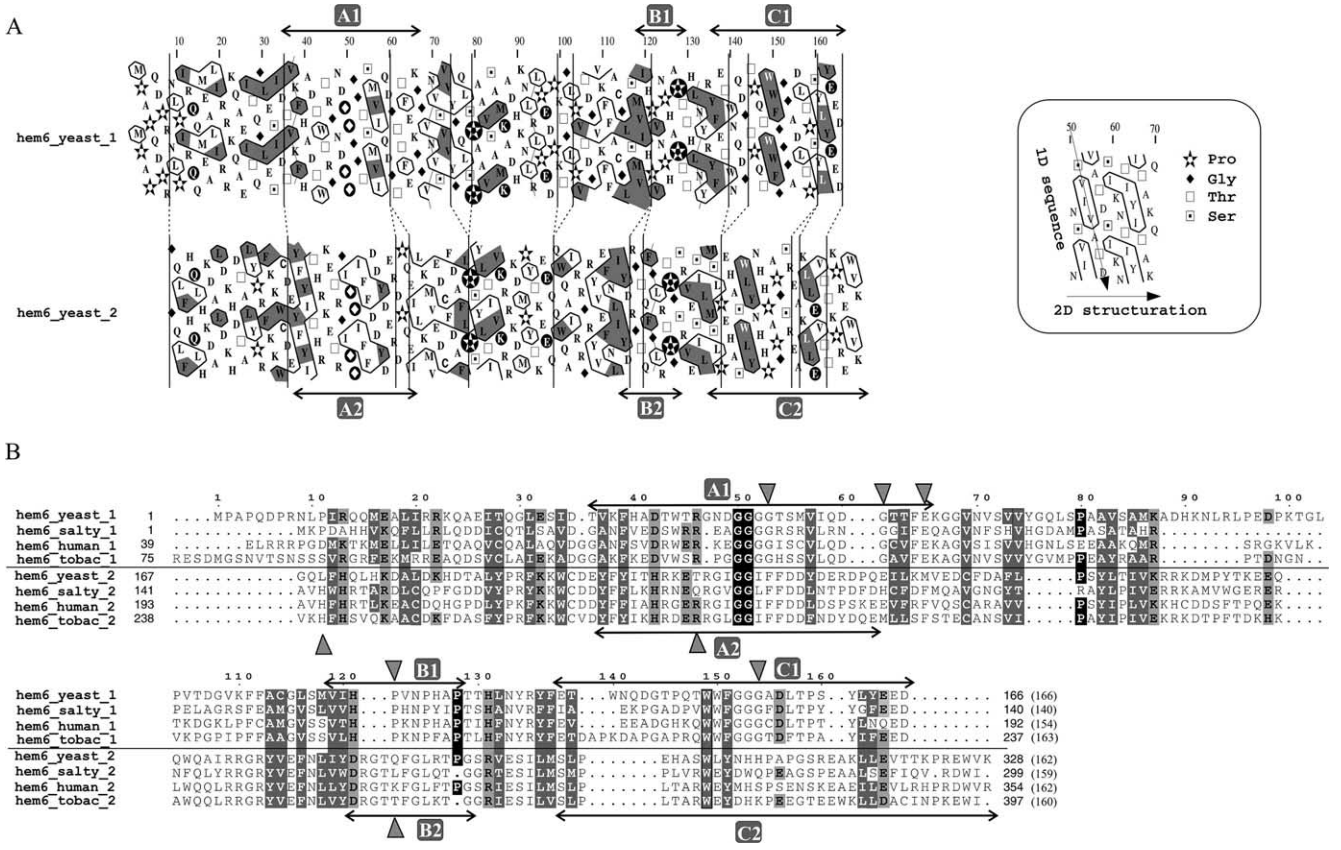

Fig. 2. A: HCA alignment between the two putative similar domains of yeast CPO (Swiss-Prot entry: hem6_yeast). The conserved hydrophobic amino acids are shown in black (similar) or white (identical) on a gray background, and the other identical residues in white on a dark background. Briefly, the sequence is shown on a duplicated α-helical net, on which hydrophobic amino acids are contoured. HCA plots were made using the Drawhca program (Luc Canard, unpublished data). The arrows correspond to important areas described in the text. The insert indicates how to read the sequence and the secondary structure organization in an HCA representation [14]. B: Alignment between putative domains 1 and 2 in four different species of CPO, based on the HCA alignments. The residues or properties are defined as conserved if they appear in six out of eight sequences. The conserved hydrophobicity (VILMFYW to which we added ACT, which can often substitute hydrophobic residues) is shown in white on a gray background, the conserved charged polarity (DEKRH) in black on a gray background and other conservation in white on a black background. The length of each domain is given between parentheses after its sequence. Naturally occurring human mutations are shown with gray triangles. Alignment representations were made using ESPript software [30].

Fig. 2), with many conserved residues. They noticeably contain conserved aromatic residues, and it has been shown that a tyrosine residue is probably involved in the catalytic activity [25]. UOX acts as a tetramer while CPO acts as a dimer [10]. The tetrameric interface of UOX is built with the extremities of four long β-strands (S3′, S4′, S7′ and S8′ in Fig. 3). These regions would be expected to differ between the two enzymes. In particular, the two small strands S4′ and S8′ (called B1 and B2) share no or few sequence similarities with CPO.

A putative 3D model of yeast CPO was built from the coordinates of *A. flavus* UOX. The CPO dimer would assemble in the same way as the UOX dimer, creating a β-barrel of 16 antiparallel β-strands, with eight external helices (Fig. 4A,B). Four regions in CPO were not fully modeled, since their lengths are considerably greater than in UOX: A1 with 12 extra residues, A2 with 13 extra residues (colored in orange in Fig. 4), C1 with 25 extra residues, and C2 with 27 extra residues (colored in red in Fig. 4). These four loops are located on the same side as the N- and C-terminal extremities (Fig. 4B,C). They are highly conserved between CPOs and exhibit patterns associating glycine repeats and small hydrophobic clusters likely associated with short strands. It seems that each of these regions is composed of two small strands which is reminiscent of β-helices or propellers [26]. There is one substrate per dimer, so these eight regions may together constitute on one side of the CPO dimer scaffold a sub-domain devoted to bind and process coproporphyrinogen, particularly through the numerous conserved aromatic residues present in these segments. B1 and B2 loops (colored in green in Fig. 4) which form the tetrameric interface in UOX, absent in dimeric CPO, are certainly different and are likely to curve back toward the top of the dimer (Fig. 4B). A1 and B2 form the interface between two monomers, A2 and B1 form the interface between two T-fold domains within a monomer and interfaces are likely to be flexible regions.

The catalytic reaction of CPO is sequential, with coproporphyrinogen III first undergoing decarboxylation at propionate in position 2 to yield harderoporphyrinogen, followed by decarboxylation of propionate at position 4 to yield the final product of the reaction, protoporphyrinogen IX (for a review, see [27]) (Fig. 1). This unique reaction mechanism raises several questions related to the topology of the active site in the dimer. Substrate specificity has been investigated by using alternative substrates that differ from coproporphyrinogen III by the position of the side chains on the tetrapyrrole nucleus [28,29]. The results suggest that at least two domains on

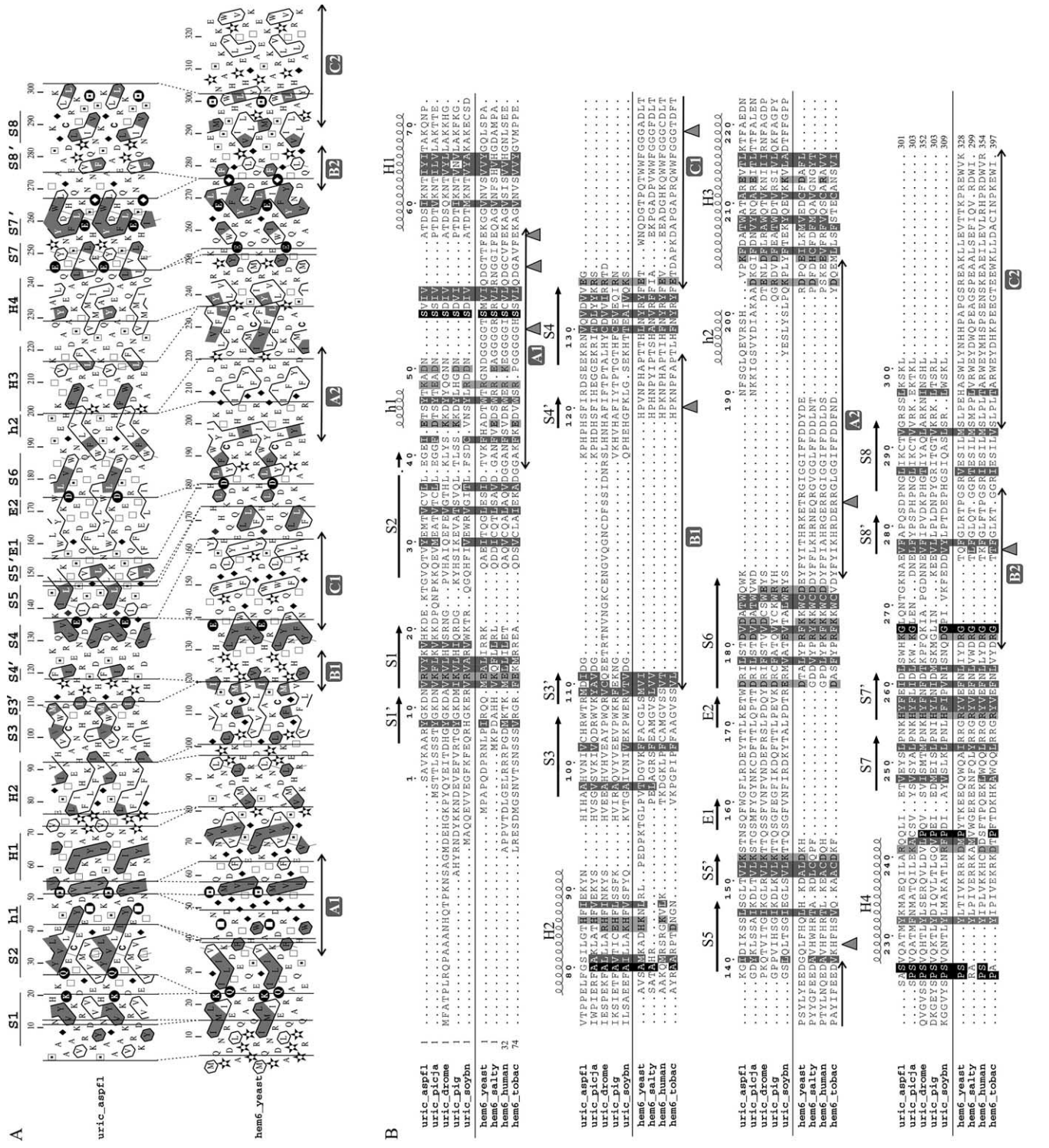

Fig. 3. A: HCA alignment between *A. flavus* UOX (Swiss-Prot entry: uric_aspfl) and yeast CPO. The secondary structure is shown for UOX (from [9]). The segment between 80 and 90 in CPO is rich in alanine and consequently is consistent with the occurrence of a helix as expected for the comparison with UOX (H2). The drawing code is the same as in Fig. 2A. B: Alignment between five UOX sequences and four CPO sequences. The residues or properties are defined as conserved if they appear in seven of nine sequences. The drawing code is the same as in Fig. 2B.

the protein may be involved in substrate recognition, while a third domain may be involved in catalysis per se. A model was proposed in which the reaction intermediate must rotate within the active site to be further processed [28]. When this rotation is not possible due to an inability of the appropriate side chains to position in the binding sites, the reaction can occur only after diffusion of the intermediate from the active site, inversion of its planar axis of symmetry, and repositioning at the active site. This model fits well with the data obtained when studying substrate specificity, but does not take

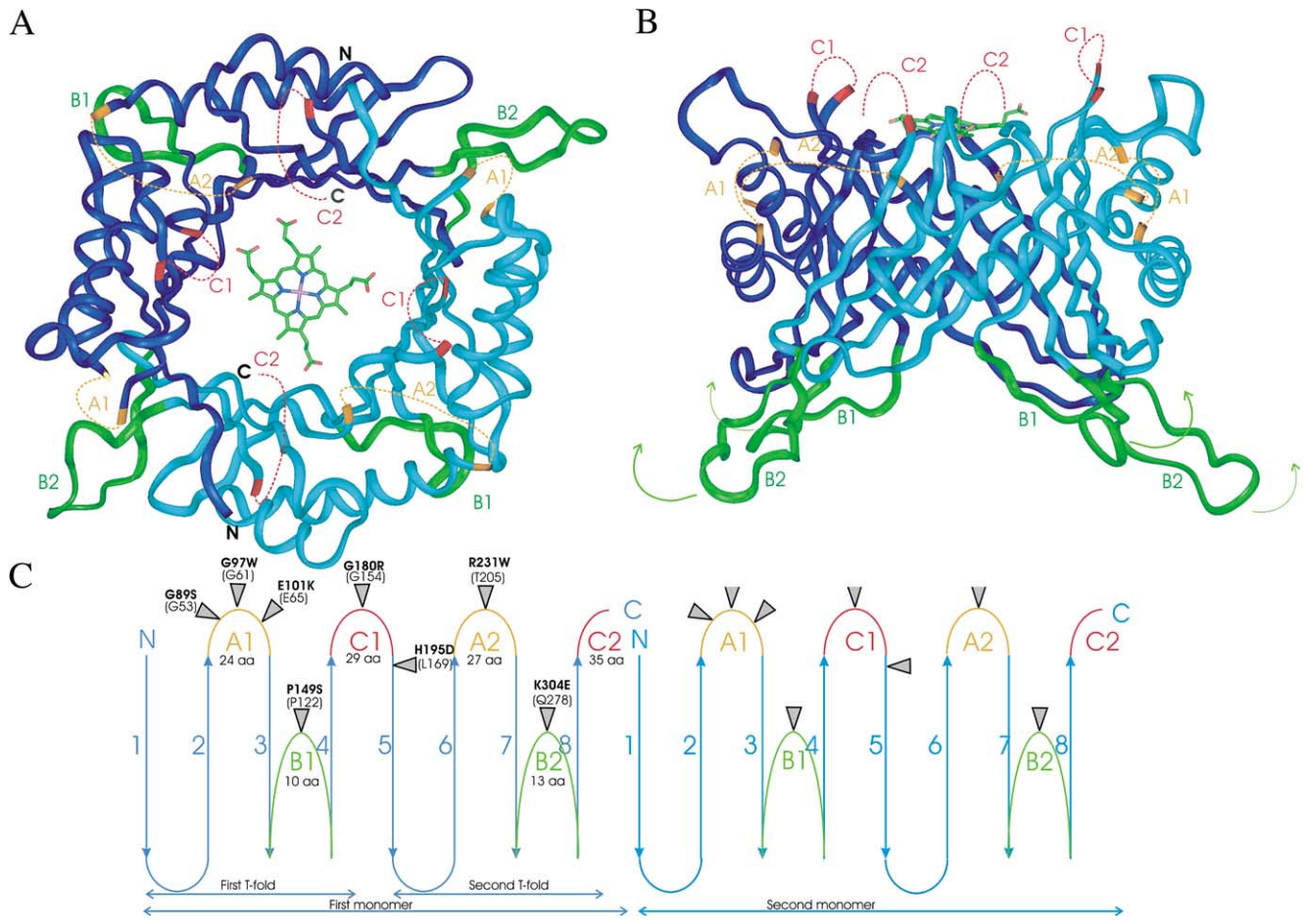

Fig. 4. A: Smooth Cα chain of the putative dimeric model of yeast CPO built on the coordinates of *A. flavus* UOX. The first chain is colored in dark blue, the second chain in pale blue. The long loops A1 and A2 (in orange), C1 and C2 (in red) are almost not modeled and are pictured by dotted lines. Coproporphyrinogen III has been positioned above the dimer, between the 12 putative involved loops. B: View perpendicular to the previous one. The loops B1 and B2 (in green) probably curve back upward in the vicinity of the substrate (as suggested by the arrows). These two figures were made using InsightII (Accelrys, San Diego, CA, USA). C: Schematic view of the 16-stranded β-barrel of the dimer, with the 12 loops colored as above. The eight helices outside the barrel are not shown. The length of A1 and A2, B1 and B2, C1 and C2 are similar, supporting further the gene duplication of CPO. The positions of the naturally occurring human mutations are shown with gray triangles; the numbering corresponds to human CPO and yeast CPO between parentheses.

into account the fact that CPO is active as a homodimer. The two subunits may act synergistically by sharing the binding and recognition sites, thereby forming a highly efficient catalytic system. The duplication of the T-fold domain in each subunit leads to an overall structure of coproporphyrinogen oxidase with a pseudo four-fold symmetry rather than a two-fold symmetry, well designed for decarboxylation of the pseudo-symmetric substrate (Fig. 5). This may explain the choice of a multimeric T-fold structure by CPO to successively perform two similar catalytic processes on a single template.

An analysis of the sites of the naturally occurring mutations in patients with coproporphyria, an inherited defect in CPO activity (NCBI Online Mendelian Inheritance in Man entry #121300 and references therein) leads further support to our hypothesis. These eight sites are involved in the catalytic mechanism and, with the exception of H195D (at the junction of C1 and S5), are located either in un-modeled regions (G89S, G97W, E101K in A1, R231W in A2 and G180R in C1) or in difficult-to-model regions (P149S in B1 and K304E in B2) (see Fig. 4C). These regions as well as the H195D mutation, are mainly close to the active site in the model. Interestingly, the two mutants P149S and K304E (P122 and Q278 in yeast) are at the exact same relative position in each domain (see Fig. 2), which is a further argument for the proposed duplication. The P149S mutation which abolishes activity is in the B1 region at the interface between two T-fold

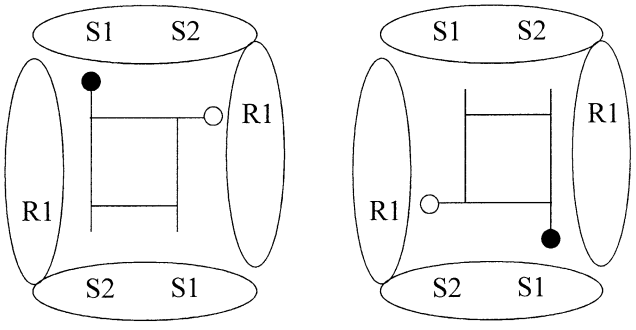

Fig. 5. Postulated topology of CPO subunits and interactions with its substrate, coproporphyrinogen III. Catalytic sites (S1) and binding/recognition sites (S2, R1) may be shared by the two subunits in the active dimer of the enzyme. The substrate may present the propionic residue decarboxylated in the first (black dot) or second step of the reaction (open dot), whatever its orientation relative to the plane of the molecule (left and right schemes).

domains in each subunit. The K304E mutant is impaired in the second step of coproporphyrinogen III decarboxylation, yielding harderoporphyrinogen instead of protoporphyrinogen IX. The B2 region which bears this mutation defines the interface between the two subunits which supports the hypothesis that both subunits of CPO are necessary for correct functioning of the enzyme (Fig. 5).

In conclusion, these common overall functional and structural data for both enzymes constitutes a convergent set of arguments suggesting that CPO may be a new T-fold protein. Others T-folds are probably present in databases but remain hidden by their very high sequence divergence. This property between closely related compact 3D folds may constitute a promising new family that may be useful in improving and evaluating the sensitivity of comparative protein sequence analysis methods.

Acknowledgements: The authors acknowledge fruitful discussions with R. and P. Labbe who initiated this study. N.C. also acknowledges the CRIHAN, the 'Centre Européen de Bioprospectives', and the FEDER to allow her to use the visualization software Insight II (Accelrys, San Diego, CA, USA).

References

[1] Colloc'h, N., Poupon, A. and Mornon, J.P. (2000) Proteins 39, 142–154.
[2] Poupon, A. and Mornon, J.P. (1998) Proteins 33, 329–342.
[3] Poupon, A. and Mornon, J.P. (1999) Theor. Chem. Acc. 101, 2–8.
[4] Sano, S. and Granick, S. (1961) J. Biol. Chem. 236, 1173–1180.
[5] Sano, S. (1966) J. Biol. Chem. 241, 5276–5283.
[6] Kruse, E., Mock, H. and Grimm, B. (1995) Planta 196, 796–803.
[7] Legoux, R., Delpech, B., Dumont, X., Guillemot, J.C., Ramond, P., Shire, D., Caput, D., Ferrara, P. and Loison, G. (1992) J. Biol. Chem. 267, 8565–8570.
[8] Leplatois, P., Le Douarin, B. and Loison, G. (1992) Gene 122, 139–145.
[9] Colloc'h, N., El Hajji, M., Bachet, B., L'Hermite, G., Schiltz, M., Prangé, T., Castro, B. and Mornon, J.P. (1997) Nat. Struct. Biol. 4, 947–952.
[10] Camadro, J.M., Chambon, H., Jolles, J. and Labbe, P. (1986) Eur. J. Biochem. 156, 579–587.
[11] Kohno, H., Furukawa, T., Tokunaga, R., Taketani, S. and Yoshinaga, T. (1996) Biochim. Biophys. Acta 1292, 156–162.
[12] Zaman, Z. and Akhtar, M. (1976) Eur. J. Biochem. 61, 215–223.
[13] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[14] Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) FEBS Lett. 224, 149–155.
[15] Bairoch, A. and Apweiler, R. (1999) Nucleic Acids Res. 27, 49–54.
[16] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shonamouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[17] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) Nucleic Acids Res. 28, 235–242.
[18] Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J.P. (1997) Cell. Mol. Life Sci. 53, 621–645.
[19] Woodcock, S., Mornon, J.P. and Henrissat, B. (1992) Protein Eng. 5, 629–635.
[20] Douguet, D. and Labesse, G. (2001) Bioinformatics 17, 752–753.
[21] Labesse, G. and Mornon, J. (1998) Bioinformatics 14, 206–211.
[22] Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F. and Sali, A. (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 291–325.
[23] Sippl, M.J. (1995) Curr. Opin. Struct. Biol. 5, 229–235.
[24] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) J. Comp. Chem. 4, 187–217.
[25] Yoshinaga, T. and Sano, S. (1980) J. Biol. Chem. 255, 4727–4731.
[26] Callebaut, I. and Mornon, J.P. (1998) Cell Mol. Life Sci. 54, 880–891.
[27] Aktar, M. (1991) in: Biosynthesis of tetrapyrroles, (Jordan, P., Ed.) Vol. 19, pp. 67–100, Elsevier, Amsterdam.
[28] Lash, T., Mani, U., Drinan, M., Chun, Z., Hall, T. and Jones, M. (1999) J. Org. Chem. 64, 464–477.
[29] Elder, G.H., Evans, J.O., Jackson, R.R. and Jackson, A.H. (1978) Biochem. J. 169, 215–223.
[30] Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) Bioinformatics 15, 305–308.