Structurally and functionally conserved regions of cytochrome *P*-450 reductase as targets for DNA amplification by the polymerase chain reaction

Cloning and nucleotide sequence of the *Schizosaccharomyces pombe* cDNA

John S. MILES

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, U.K.

1. Alignments of the available cytochrome *P*-450 reductase amino acid sequences, and comparison with the crystal structure of ferredoxin–NADP reductase, indicate that two highly conserved regions are of functional importance. 2. Degenerate oligonucleotide primers, based on these sequences, were used in the polymerase chain reaction to amplify a 309 bp fragment of the cytochrome *P*-450 reductase gene from *Schizosaccharomyces pombe* for use as an homologous probe. 3. A 2.6 kb cDNA was cloned from a λ library, and sequencing revealed an open-reading frame of 2034 bp encoding a protein of M_r 76774. This protein shares 38–41 % identity with other eukaryotic cytochrome *P*-450 reductases, and 30 % identity with that of *Bacillus megaterium*. 4. Comparison of the *N*-terminal FMN-binding domain with flavodoxin, and the *C*-terminal FAD- and NADP-binding domain with ferredoxin–NADP reductase, indicates the presence of several functionally conserved regions. 5. The *Sc. pombe* cytochrome *P*-450 reductase gene was shown to contain no introns.

INTRODUCTION

The cytochrome *P*-450-dependent mono-oxygenase (*P*-450) system catalyses the oxidation of lipophilic chemicals through the insertion of one atom of molecular oxygen into the substrate, with the concomitant reduction of the other atom to water. *P*-450s are found throughout Nature and are responsible for a wide range of biotransformations (Schuster, 1989). In yeasts, several *P*-450 systems have been characterized (Karenlampi *et al.*, 1980; Kappeli, 1986). *P*-450_{14DM} (lanosterol 14-demethylase) plays a key role in the biosynthesis of ergosterol, a membrane sterol, and is the target for the azole and triazole antifungal agents (Yoshida & Aoyama, 1991). Biochemical and physiological analysis of the *P*-450 in *Schizosaccharomyces pombe* suggests that it is of this type (Poole & Lloyd, 1974; Bligh & Kelly, 1988). Alkane-assimilating yeasts, such as *Candida tropicalis* and *C. maltosa*, contain *P*-450 systems capable of hydroxylating alkanes and fatty acids (Muller *et al.*, 1991).

P-450 systems can be divided into two groups which differ in the type of electron transfer proteins that supply reducing equivalents during catalysis. The first group, exemplified by camphor hydroxylase (*P*-450_{cam}) from *Pseudomonas putida*, uses both an iron-sulphur protein (ferredoxin) and a flavoprotein (ferredoxin reductase). The second group uses NADP-cytochrome *P*-450 reductase (CPR; EC 1.6.2.4), a flavoprotein of M_r about 78000 associated with the endoplasmic reticulum of many eukaryotic cells, including yeasts (Phillips & Langdon, 1962). Intriguingly, an homologous soluble CPR has been found fused to cytochrome *P*-450 in the catalytically self-sufficient enzyme of *Bacillus megaterium* (Narhi & Fulco, 1986). CPR is unusual in containing 1 mol each of FAD and FMN per mol of protein (Iyanagi & Mason, 1973). It belongs to an expanding family of flavoproteins, with binding sites for both FAD and FMN, that includes the α -subunit of sulphite reductase (α SR)

from *Escherichia coli* and *Salmonella typhimurium* (Ostrowski *et al.*, 1989), and nitric oxide synthase (NOS) from the rat (Bredt *et al.*, 1991).

The amino acid sequences of CPR from *B. megaterium* (Ruettinger *et al.*, 1989), *Saccharomyces cerevisiae* (Yabusaki *et al.*, 1988), *C. tropicalis* (Sutter *et al.*, 1990), rat (Porter & Kasper, 1985; Porter *et al.*, 1990), rabbit (Katagiri *et al.*, 1986) and man (Yamano *et al.*, 1989) have been deduced from the corresponding genes or cDNAs. The primary structure of trout and pig CPRs have been determined directly from the protein (Urenjak *et al.*, 1987; Haniu *et al.*, 1986). Comparing these sequences with flavoproteins of known structure indicates that the *N*-terminal FMN-binding domain of CPR is homologous to bacterial flavodoxins, and the *C*-terminal FAD- and NADP-binding domain is homologous to ferredoxin–NADP reductase (FNR) (Porter & Kasper, 1986; Porter, 1991; Karplus *et al.*, 1991). This suggests an evolutionary origin for CPR based on the fusion of the ancestral genes for flavodoxin and FNR (Porter & Kasper, 1986).

The crystal structure of FNR shows that two sequence motifs, Arg-Leu-Tyr-Ser-Ile and Gly-Thr-Gly-Thr-Gly, are intimately involved in binding FAD and the pyrophosphate moiety of NADP respectively (Karplus *et al.*, 1991). The arginine residue in the first motif hydrogen-bonds with the pyrophosphate of the FAD, the tyrosine makes extensive van der Waal's contact with the flavin and also a hydrogen bond with the ribityl 3' hydroxy, and the serine is in a position to hydrogen-bond with reduced flavin (Karplus *et al.*, 1991). The second motif forms a pyrophosphate binding loop for NADP, and although there is some sequence similarity with other dinucleotide binding folds, the loop in FNR makes a much tighter turn (Karplus *et al.*, 1991). Closer examination of the CPR sequences reveals the presence of two very highly conserved regions: Arg-Tyr-Tyr-Ser-Ile-Ala/Ser-Ser (at positions 454–460 in the rat sequence) and Ile-Met/Leu-

Abbreviations used: CPR, NADP-cytochrome *P*-450 reductase; FDX, flavodoxin; FNR, ferredoxin–NADP reductase; NOS, nitric oxide synthase; *P*-450, cytochrome *P*-450; α SR, α -subunit of sulphite reductase.

The nucleotide sequence data in this paper have been submitted to EMBL, under accession no. X64702

Val/Ile-Gly-Pro-Gly-Thr-Gly-Val/Ile-Ala-Pro-Phe/Leu (positions 529–540), which are believed to be analogous to the functional motifs in FNR. In this paper I have used these structurally and functionally conserved regions of CPR to design degenerate oligonucleotide primers. These primers have been used in the PCR to amplify the appropriate segment of the *Sc. pombe* CPR gene for use as an homologous probe to clone the cDNA. This method should prove useful in cloning CPR cDNAs or genes from other diverse sources, such as plants and insects, without the need to purify the membrane protein; and in cases, such as this one, where DNA–DNA hybridization with heterologous probes would prove unsuccessful.

EXPERIMENTAL

Polymerase chain reaction

Oligonucleotide primers used in the PCRs (Saiki *et al.*, 1988) were P22 (CCGGAATTCGNTAYTAYTCNAT) and P23 [CCGTCGACGGNCCNAYCAT(AGT)AT]. PCRs contained target DNA (*B. megaterium* ATCC14581 genomic DNA or rat cDNA, 100 ng; *Sc. pombe* ED628 genomic DNA, 1 µg) and 1 µg each of P22 and P23 in a 100 µl volume. *Taq* DNA polymerase (Promega) was used according to the manufacturer's instructions, and 35 cycles of 1 min/94 °C, 2 min/50 °C and 3 min/72 °C were performed in a Perkin–Elmer Cetus Thermocycler. Primers used to determine the presence or absence of introns within the CPR gene were P41 (TTCGCATCTGCCTATCA, positions 340–356) and P42 (AAATGGATTAGCTTGAG, positions 1280–1264); and P45 (TTGAGAAGCCTGAAATT, positions 1165–1181) and P47 (ATGATCTTGCTCACGGG, positions 2294–2278). Reaction conditions were the same as above, except that 25 cycles were performed.

Isolation of *Sc. pombe* CPR cDNA clones

The PCR product from the *Sc. pombe* DNA was labelled with [α -³²P]dATP using random priming (Feinberg & Vogelstein, 1983). It was used to screen 50000 plaques from a *Sc. pombe* λZAPII cDNA library, kindly provided by Dr Dallen Young, University of Calgary, Calgary, Alberta, Canada, using the method of Benton & Davies (1977). The final wash conditions were 65 °C in 2 × SSC (1 × SSC = 0.15 M-NaCl/0.015 M-sodium citrate). Individual phage were isolated by serial dilution and rescreeing, and plasmids were recovered following M13 super-infection using the protocol of Blondel & Thillet (1991).

DNA sequence analysis

DNA was sequenced by the chain-termination method using [α -³⁵S]dATP (Sanger *et al.*, 1980; Biggin *et al.*, 1983) and Sequenase version 2.0 (United States Biochemicals). Single-stranded plasmid DNA was sequenced using a series of synthetic oligonucleotides. Specific DNA fragments were subcloned into M13 and sequenced using universal primer. The sequence of both DNA strands was determined unambiguously.

Computer analyses

DNA sequences, and predicted amino acid sequences, were analysed using the University of Wisconsin Genetics Computer Group programmes (Devereux *et al.*, 1984).

Other molecular biology methods

DNA manipulations and bacterial transformations were by standard methods (Sambrook *et al.*, 1989).

Materials

Restriction enzymes, T4 DNA ligase and M13 vectors were from Gibco–BRL or Boehringer. Radiochemicals were from

Amersham, and other reagents were from BDH and Sigma. Oligonucleotides were synthesized on an Applied Biosystems 380B machine.

RESULTS AND DISCUSSION

Highly conserved regions of CPR and design of degenerate PCR primers

Alignments of all the available CPR amino acid sequences (not shown) reveal the presence of two very highly conserved regions: Arg-Tyr-Tyr-Ser-Ile-Ala/Ser-Ser (at positions 454–460 in the rat sequence) and Ile-Met/Leu-Val/Ile-Gly-Pro-Gly-Thr-Gly-Val/Ile-Ala-Pro-Phe/Leu (positions 529–540), the cores of which are invariant (Fig. 1). These sequences are also found in NOS (Bredt *et al.*, 1991). The functional importance of these two regions has been highlighted recently by comparing the CPR sequences with the crystal structure of ferredoxin–NADP reductase (Karplus *et al.*, 1991). This comparison suggests that the Arg-Tyr-Tyr-Ser-Ile and Gly-Thr-Gly-Xaa-Xaa-Pro motifs are involved in binding FAD and the pyrophosphate moiety of NADP respectively. Similar, but not identical, motifs are found in sulphite reductase (Ostrowski *et al.*, 1989), NADH–cytochrome *b₅* reductase (Tomatsu *et al.*, 1989) and NADH–nitrate reductase (Crawford *et al.*, 1988), and are believed to have an analogous functional role.

Degenerate oligonucleotide primers encoding the two conserved regions of CPR were designed as shown in Fig. 1. The degeneracy of primer P22 was minimized by using only TCN serine and CGN arginine codons (AGY and AGR are not used for these serine and arginine codons in any of the known sequences). Restriction sites were added to the 5' end of each sequence to improve the primer selectivity (Mack & Sninsky, 1988) and to facilitate subcloning.

Amplification of CPR DNA sequences

As a control, DNA fragments of the expected sizes were amplified from *B. megaterium* genomic DNA and rat liver cDNA (246 bp and 255 bp respectively; Fig. 2). The correct identity of the *B. megaterium* fragment was confirmed by sequencing. A single fragment of 309 bp was amplified from *Sc. pombe* genomic DNA (Fig. 2). Its sequence encodes a polypeptide fragment with 44% identity to the corresponding region of *S. cerevisiae* CPR. This DNA fragment was then used as an homologous probe to isolate the *Sc. pombe* CPR cDNA.

Isolation and sequencing of *Sc. pombe* CPR cDNA

Approx. 50000 plaques from a *Sc. pombe* cDNA library cloned in λZAPII were screened using the PCR fragment as an homologous probe. Ten positive clones were purified, and plasmid DNA was recovered as described in the Experimental section. Restriction mapping revealed that two clones (pJM37 and pJM43) contained complementary inserts, each of 2.6 kb and of sufficient size to encode CPR. The strategy devised for sequencing these clones is described in Fig. 3, and the sequence is shown in Fig. 4.

The organization of the rat CPR gene has been determined (Porter *et al.*, 1990). The position of the intron–exon boundaries appears to support the idea of a modular structure for CPR generated by fusion of the ancestral genes for flavodoxin and FNR. Thus it was of interest to determine whether the *Sc. pombe* CPR gene contains any introns. As only about one-third of *Sc. pombe* genes contain introns, rather than cloning the gene, it was decided to investigate the gene organization using the PCR. Two pairs of primers (P41/P42, and P45/P47; see the Experimental section) were used to amplify DNA fragments, which span almost all of the CPR coding sequence (see below), both from the

	ArgTyrTyrSerIle	IleMetValGlyPro	GTGVAPF
		LeuIle	I L
Man	CGCTACTACTCCATC	ATCATGGTGGGCCCC	
Rabbit	CGCTACTACTCCATC	ATCATGGTGGGCCCC	
Rat	CGATACTACTCCATT	ATCATGGTGGGCCCC	
<i>S. cerevisiae</i>	CGTTACTACTCTATC	ATCATGATCGGTCCA	
<i>B. megaterium</i>	CGCTATTACTCGATT	ATCATGGTCGGACCG	
<i>(C. tropicalis)</i>	CGTTATTACTCCATT	ATTTTGATTGGTCCA	
	P 22	P 23 (reverse complement)	
	CGTTATTATTCTAT	ATTATGATTGGTCC	
	C C C C	C G C C	
	A A	A A A	
	G G	G G	

Fig. 1. Design of degenerate PCR primers

The two regions of highly conserved amino acid sequence within CPRs are shown with corresponding nucleotide sequence for six species. Only TCN serine and CGN arginine codons are used, thus limiting the degeneracy of primer P22 to 64-fold (CCGGAATTCGNTAYTAYTCNAT). Primer P23 is 96-fold degenerate [CCGTCGACGGNCCNAYCAT(AGT)AT] and is the reverse complement to that shown. Restriction sites (*Eco*RI, P22; *Sal*I, P23) are added to facilitate subcloning. Although the *C. tropicalis* sequence is shown (in parentheses), it was not available when the primers were designed, and was not included in the consensus.

cDNA clone pJM37 and from *Sc. pombe* genomic DNA. The sizes of the products with genomic DNA as the target were the same as with the cDNA, indicating that the *Sc. pombe* CPR gene contains no introns.

Fig. 2. Amplification of CPR DNA sequences

Products (10 μ l) of PCRs using primers P22 and P23 with *B. megaterium* genomic DNA (246 bp; lane 1), *Sc. pombe* genomic DNA (309 bp; lane 2) and rat liver cDNA (255 bp; lane 3) were separated electrophoretically on a 2% agarose gel. Molecular size markers are: M1, kb ladder (Gibco-BRL); M2, Φ X174 digested with *Hae*III; DNA fragment sizes are shown in bp.

Amino acid sequence of *Sc. pombe* CPR and comparison with other flavoproteins

An open-reading frame was detected from positions 474–2507 in the cDNA sequence. It encodes a protein of 678 amino acids, with a predicted M_r of 76774 and all the features expected of a CPR (Fig. 4; and see below). The only other ATG triplet that occurs upstream of position 474 starts a very small open-reading frame (positions 105–128) which encodes a peptide of seven amino acids.

Codon usage in *Sc. pombe* shows a bias in those genes that are highly expressed, such as triose phosphate isomerase and alcohol dehydrogenase, compared with those expressed at a low level, such as cell cycle control genes (Russell, 1989). Analysis of the codon usage in the CPR gene (results not shown) indicates that it is not biased, suggesting that CPR is not expressed at a high level.

The amino acid sequence motifs used to design the PCR primers are at positions 441–445 (Arg-Tyr-Tyr-Ser-Ile) and 534–538 (Ile-Met-Val-Gly-Pro). Comparison of the *Sc. pombe* CPR sequence with those of CPRs from other species shows it to share significant similarity with CPRs from *S. cerevisiae*, *C. tropicalis*, trout, rat, rabbit, man (all between 38 and 41% identity) and *B. megaterium* (30% identity). However, the degree of DNA sequence identity between the CPR genes of *S. cerevisiae* and *Sc. pombe* (49%), and between *C. tropicalis* and *Sc. pombe* (53%), is too low for them to have been successful as heterologous DNA–DNA hybridization probes.

In common with the other eukaryotic CPRs, the *N*-terminal

Fig. 3. Sequencing strategy

The sequence of *Sc. pombe* CPR was assigned unambiguously on both DNA strands using pJM37 and pJM43 single-stranded DNA and a series of oligonucleotide primers (P36–38; P40–47), the start of each arrow indicating the priming site. Additional sequence was from M13 subclones obtained using the internal restriction sites *Kpn*I (K), *Eco*RI (E) and *Xho*I (X), and the PCR-generated DNA fragment. *Not*I (N) sites were used in construction of the λ library. The CPR open-reading frame is indicated.

NotI	
GCGGCCGCAATTGCCAAACAAATAGCTTTTCTCCGAAAGGATTGCTTTTGAAGTCAATTACTTTAGAAGCTTCTAGCATCTATTTG	90
ATTCTTCAATTTGGCATGGTTATTGTGCTCATCATATAAAGCGCAGTTGTTTCATTAATTAAGTATCTTAGCCTACATCTTCTTAT	180
TTTATAGATTCTTTAGAGTATTGGATAAATACTTACTTTTGTTTAAGAGGCTATAGAACCCCTTGTGAGTTTCTTCAATTTTAT	270
TGCTCGCGCTTTAGTGTATCTGCTTCAAAGTGGTTCAAGCTAAAAAAGCATAGTACTGTTGCTCTTATTCGCATCTGCTTATCATCAG	360
ACAATTCGTGTGACTCACTCGTTTTCACGCACATTTACCCCTGGAAATAAAGGCGGATTAAATCCCGAAGTTGAACACTCGTAAAA	450
M K T Y E Y V L L V I I L I L S L C Y F I Y N	
TTGTTATTATATTTTTTGCATTATGAAGACTTATGAATATGTATGCTGGTCATTATTTTGATCCTTAGCTTCTGTTACTTTATTATATA	540
N F L N K P K A P E R R V V A T D S I V E L M E A E K L T A	
ACAATTTTCTCAACAAGCCAAAGGCTCCTGAGAGGCGAGTCGTGGCTACCGACAGTATTGTCGAGTTCATCGAGGCCGAAAGCTCACTG	630
A V F F G S Q T G T A E D F A Y R F S T E A K A N F N L T N	
CCGGGGTTTCTTTGGATCTCAAACTGGTACCGCTGAGGACTTTGCCTACCGTTTTTCAACGGAAGCAAAAGGCAAACTTCAATCTTACCA	720
KpnI	
M V F D L E N Y D L T D L D N F D R S K L L V F F L A T Y G	113
ATATGGTTTTCGATCTTGAGAACTATGATTAACCGACTTGACAACTTTGATAGATCAAACTATTGGTATTCTTCTCGCAACTTACG	810
E G E P T D N A E A F L Q L L E G D D T V F S S G K G I E D	
CTCAAGTGAACTACCGACAATCGGAGGCGCTTTTTCGAACCTCTTGAGGCGCATATACCGTTTTTCAAGTGAAAAAGGCATTGAAG	900
T P F E G I R Y A I F G L G N H T Y E Y Y N A M A K K V D A	
ATATCCCTTCGAAGGAATTCGTTATGCAATTTTGGCCTTGGAAATCATACCTATGAGTATTATAACGCAATGGCTAAAAAGGTAGATG	990
EcoRI	
A M T R L G A T R V G N L G L G D D A A G M L E E D Y L Q W	203
CTGCAATGACTCGTTTACCGCTACTCGTGTGGGCAATCTAGGTTTGGGCGACGATGCCGCTGGTATGCTTGAAGAGGACTATTTACAAT	1080
K D D T L P E I G K L F H L Q E V H K E Y N P M F E V I E K	
GGAGGATGATACTTTACCTGAGATAGGTAAGCTGTTCCATTGCAAGAGTCCACAAGGAATACAATCCTATGTTTGAAGTATTGAGA	1170
P E I S N T S S T V F I G E P S R Q Q L K G N V A S K A P R	
AGCCTGAAATTTCCAACACTAGTTCTACTGCTTTTTGGGAGAACCAAGTCGTGACGAATTTGAAAGGAAATGTTGCATCGAAGCTCCAC	1260
S Q A N P F F S S P V R S L E L F K S G S R N C L H L E L D	
GTTCCTAAGCTAATCCATTTTCTCTCTGCTGATCTCTTGAGCTATTCAAATCTGGAAGTCGTAATTGCTTACATTAGAGTTGG	1350
I A D S G M R Y Q T G D Y A S I C P M N P S Q A V D D L L E	
ATATCGAGACTCTGCTATGCGATATCAACAGGTGATTACGCTTCTATTTGTCTATGAATCTTAGTCAAGCTGTTGATGATCTTCTCG	1440
V L G L K E K R D T V I I V K P I D T L D K A P V L S P T T	
AAGTCTCGGCTCAAAAGAGAAGGGGACTGTAATAATTGTTAAACCTATCGATACTCTCGATAAGGCTCCTGTTTAAAGCTCTACTA	1530
Y D T V F R Y Y Y E I C G I V S R Q L L S F I A P F A P T P	
CCTATGACATGTTTTTTCGCTACTACTATGAAATATGTGTATCGTCTCCAGGCAACTGCTTCTTTTATTGCTCCTTTTCGCTCTACAC	1620
E S K Q E L E K L G N D Y D Y F K K N V V D L H L N L A Q V	
CAGAATCTAAGCAAGAACTCGAGAAATAGGCAACGATTATGACTACTTTAAGAAGAACGTTGTAGACTTGACCTAAACCTTGCTCAAG	1710
XhoI	
L R R V S P D A P F T K L P F S M L L E N M A H M K P R Y Y	443
TCCTTCGTCGTGTTCTCTGACGCTCCATTTACAAAACCTCCATTCTCAATGCTATTGGAAAAATGGCACACATGAAACCGCTTACT	1800
XhoI	
S I S S S S V V H P D K V H V T A V V D K K E W T D K N E I	473
ATTGATTTCTCGAGCAGTGTAGTGCATCCTGACAAAGTTTACGTTACAGCGGTAGTTGACAGAAAGAAATGGACTGATAAGAATCATA	1890
XhoI	
F Y G L T T T N Y L L A H C R H M H G E K I P H P N G L E Y T	503
TCTTTATGGACTAACTACTAAATTTTGTGGCACATTGTCGTATATGACCGGTGAAAAAATCCACATCCAAATGGTTTGGAAATATA	1980
L E G P R K N W T G K I P M F V K K S T F R L A P P D V P I	
CCCTTGAGGGTCTCGTAAAAAATGGACCGAAAAATCCGATGTTTGTAAAGAAATCTACTTTTCGATTGGCTCCTCCCGATGTCCCA	2070
I M V G E G T G V A P F R G F V M E R A N L A S K G V K V A	
TAATAATGCTAGGCGCGGTACTGGTGTAGCCCCATTCCGCGGTTTTGTGATGGAACGGGCTAATTTAGCTTCTAAGGACTTAAAGTAG	2160
K T L L F Y G C Q Y S D K D F L Y K E E W Q Q Y K D V L K D	
CAAAACACTTTTATTTTATGGTTGCCAATATTCTGACAAGGATTTTTATATAAGGAAGAAATGGCAACAAATACAAAGATGTCTTGAAAG	2250
S F E L I T A F S R E Q D H K I Y V Q H R I L E H S D T I A	
ACTCTTTTGAAGTATTACCGCATTTTCCCGTGAGCAAGATCATAAATCTATGTTTACGACCGCTCTTTAGAGCACAGTGATACCATTTG	2340
K L V E E G A A F Y I C G D A D H M A K D V V N A L A S I L	
CTAAACTTGTAAGAAGGGGCTGCTTTTTATATCTGCGGTGATGCTGATCATATGGCTAAAGATGTTGTGATGATTGGCTTCTATTTC	2430
T T V D V D G M K A V K A L R D D N R F F E D T W *	
TGACCACTGTTGATGTTGACGGTATGAAAGCGTTAAAGCTCTTCGAGATGACAATAGATTTTTTGAAGATACCTGGTAAGTTTGGAAAT	2520
TATCCAGACTTTGCGATTACTTCAAGACTTAATGATGCTTTTTCATCTTAGAAATACTTTTGGGCGCGG	2589
NotI	

Fig. 4. cDNA and deduced amino acid sequence of *Sc. pombe* CPR

The nucleotide sequence from pJM37/pJM43 is shown and the positions of restriction sites used in cloning are indicated. An open-reading frame of 678 codons (positions 474–2507) was used to predict the amino acid sequence. The two conserved sequence motifs used in designing PCR primers are underlined. The 309 bp PCR product from *Sc. pombe* genomic DNA consists of nucleotides 1809–2072 (excluding primers). The DNA sequence is numbered from the 5' *NotI* site used in cloning, and the amino acid sequence is numbered from the putative initiating methionine.

region of the protein contains a high proportion of hydrophobic amino acid residues (18 of the first 30 residues are hydrophobic, including the run of nine residues Val-Leu-Leu-Val-Ile-Ile-Leu-Ile-Leu). In the rat the *N*-terminal hydrophobic region is known to attach CPR to the endoplasmic reticulum, and it seems likely that it plays the same role in the *Sc. pombe* enzyme.

There are several regions of similarity which are conserved between all the CPR sequences, including that of *Sc. pombe*, as well as in NOS and α SR (Fig. 5). The functional significance of these regions can be judged by considering homologous proteins of known three-dimensional structure. The *N*-terminal FMN-binding domain of CPR is homologous to flavodoxin, and the C-

(a)				
FDX	YGSTTGNT...STWGD...SSYEY			
	8	58	96	
α SR	SASQTGNA...STQGE...TSYEF			
NOS	YATETGKS...STFGN...RAYPH			
BMEG	YGSNMGTA...ASY-N...KNWAT			
Trout	YGSQTGTG...ATYGE...KTYEH			
Rat	YGSQTGTA...ATYGE...KTYEH			
SCER	YASQTGTA...STYGE...STYEF			
CTRO	FGSQTGTA...ATYGE...STYEF			
SPOM	FGSQTGTA...ATYGE...HTYEF			
	57	110	159	
(b)				
FNR	AVSR...EKMVIQT			
	232	243		
α SR	AWSR...EKIYVQD			
NOS	AYSR...PKKYVQD			
BMEG	AFSR...PKTYVQH			
Trout	AFSR...QKVYVQH			
Rat	AFSR...HKVYVQH			
SCER	AHSR...KKVYVQD			
CTRO	AFSR...KKVYVQH			
SPOM	AFSR...HKIYVQH			
	600	607		
(c)				
FNR	EVY~			
	312			
CYSJ	DVY~			
NOS	DIFGV...			
BMEG	DVWAG~			
Trout	DVWS~			
Rat	DVWS~			
SCER	DVW~			
CTRO	DVW~			
SPOM	DTW~			
	676			

Fig. 5. Functionally important residues conserved between CPR and other flavoproteins

Regions containing functionally-important residues of CPR involved in (a) FMN binding, (b) discriminating between NADP and NAD, and (c) protection of the FAD flavin in the absence of NADP are shown. Residues discussed in the text are marked *. FDX is flavodoxin from *Desulfovibrio vulgaris*, α SR is the α subunit of sulphite reductase from *Salmonella typhimurium*, NOS is nitric oxide synthase from rat. The remaining sequences are CPR from *Bacillus megaterium* (BMEG), trout, rat, *Saccharomyces cerevisiae* (SCER), *Candida tropicalis* (CTRO) and *Schizosaccharomyces pombe* (SPOM). Numbering is for the FDX, FNR and *Sc. pombe* CPR sequences.

cytochrome *c* reductase activity, whereas conversion of Tyr-178 to Asp in rat CPR (equivalent to Tyr-98 in FDX) abolishes both FMN binding and cytochrome *c* reductase activity (Shen *et al.*, 1989).

The nucleotide cofactor specificity of CPR, as for FNR, NOS and α SR, is for NADP rather than NAD. In the crystal structure of FNR with 2'-AMP bound, the residues Ser-234, Arg-235, Lys-244 and Tyr-246 all form hydrogen bonds with the 2' phosphate (Karplus *et al.*, 1991). All of these residues, believed to be important in discriminating between NADP and NAD, are conserved in the alignment with NOS and α SR and correspond to residues 602, 603, 608 and 610 in *Sc. pombe* CPR (Fig. 5b). In FNR it has been suggested that the C-terminal residue (Tyr-314) may fill the nicotinamide binding pocket in the absence of NADP, in order to maintain enzyme structure and protect the flavin (Karplus *et al.*, 1991). An aromatic residue is present in this position in the alignment (Fig. 5c), although the residue is at the extreme C-terminus only in α SR and the yeast CPRs, the other CPRs and NOS having additional sequence extending beyond this residue.

Because of the importance of yeast *P*-450 systems as targets for antifungal agents (Yoshida & Aoyama, 1991), it was of interest to identify any regions of CPR which were significantly different from the mammalian enzymes. Differences in *S. cerevisiae* and *C. tropicalis* CPRs compared with the mammalian enzymes have been noted previously in the NADP-binding domain, where significant gaps are required for optimal alignment (Sutter *et al.*, 1990). Similar gaps are required for optimal alignment of the *Sc. pombe* sequence with the mammalian CPRs in areas between residues 463 and 557, but there are no regions of strong sequence similarity present only in the yeast sequences (results not shown). Further sequences and biochemical analysis of yeast CPRs will be needed to determine whether these differences have any functional significance.

In conclusion, the cDNA for *Sc. pombe* CPR has been cloned using an homologous probe. This was generated by the PCR using degenerate oligonucleotide primers with sequences based on structurally and functionally conserved regions of CPR. This approach should allow cloning of CPR genes or cDNAs from other diverse species where conventional cloning techniques, involving the purification of the enzyme, may be difficult. Comparison of the deduced amino acid sequence with those of other related flavoproteins highlights residues whose importance in cofactor binding may be tested by site-directed mutagenesis experiments.

I thank the Department of Biochemistry, University of Glasgow, and the Nuffield Foundation for financial support of this work. Thanks are also due to Arlene McConnell for technical assistance, and to Dr. Dallan Young, University of Calgary, for supplying the cDNA library.

REFERENCES

- Benton, W. D. & Davies, R. W. (1977) *Science* **196**, 180–182
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Bligh, H. F. J. & Kelly, S. L. (1988) *Yeast* **4**, 85–92
- Blondel, A. & Thillet, J. (1991) *Nucleic Acids Res.* **19**, 181
- Bredt, D. S., Hwang, P. M., Glatt, C. E., Lowenstein, C., Reed, R. R. & Snyder, S. H. (1991) *Nature (London)* **351**, 714–718
- Crawford, N. M., Smith, M., Bellissimo, D. & Davis, R. W. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5006–5010
- Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Feinberg, D. P. & Vogelstein, B. (1983) *Anal. Biochem.* **136**, 6–13
- Haniu, M., Iyanagi, T., Miller, P., Lee, T. D. & Shively, J. E. (1986) *Biochemistry* **25**, 7908–7911
- Iyanagi, T. & Mason, H. S. (1973) *Biochemistry* **12**, 2297–2308
- Kappeli, O. (1986) *Microbiol. Rev.* **50**, 244–258

terminal FAD- and NADP-binding domain of CPR is homologous to FNR. Fig. 5(a) shows sequences within the FMN-binding domains of CPRs, α SR and NOS aligned with that of *Desulfovibrio vulgaris* flavodoxin (FDX) (Watenpaugh *et al.*, 1973). The conserved hydroxy groups of residues Ser-10 (or Thr) and Thr-12 (FDX numbering) form part of a cage of hydrogen bonds around the 5' phosphate of FMN. In the FDX structure the isoalloxazine ring of the FMN is sandwiched by Trp-60 and Tyr-98. These or other aromatic residues are present in equivalent positions in CPR and NOS, although a glutamine residue is present in place of tryptophan in α SR. It is worth noting that an aromatic residue is not found at this position in all flavodoxins. Also, conversion of Tyr-140 into Asp in rat CPR, the residue equivalent to Trp-60 in FDX, does not abolish FMN binding or

- Karenlampi, S. O., Marin, E. & Hanninen, O. O. P. (1980) *J. Gen. Microbiol.* **120**, 529–533
- Karplus, P. A., Daniels, M. J. & Herriott, J. R. (1991) *Science* **251**, 60–66
- Katagiri, M., Murakami, H., Yabusaki, Y., Sugiyama, T., Okamoto, M., Yamano, T. & Ohkawa, H. (1986) *J. Biochem. (Tokyo)* **100**, 945–954
- Mack, D. H. & Sninsky, J. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6977–6981
- Müller, H.-G., Schunck, W.-H. & Kargel, E. (1991) in *Frontiers in Biotransformation* (Ruckpaul, K. & Rein, H., eds.), vol. 4, pp. 87–126, Taylor & Francis, London
- Narhi, L. O. & Fulco, A. J. (1986) *J. Biol. Chem.* **261**, 7160–7169
- Ostrowski, J., Barber, M. J., Rueger, D. C., Miller, B. E., Siegel, L. M. & Kredich, N. M. (1989) *J. Biol. Chem.* **264**, 15796–15808
- Phillips, A. H. & Langdon, R. G. (1962) *J. Biol. Chem.* **237**, 2652–2660
- Poole, R. K. & Lloyd, D. (1974) *Biochem. J.* **138**, 201–210
- Porter, T. D. (1991) *Trends Biochem. Sci.* **16**, 154–158
- Porter, T. D. & Kasper, C. B. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 973–977
- Porter, T. D. & Kasper, C. B. (1986) *Biochemistry* **25**, 1682–1687
- Porter, T. D., Beck, T. W. & Kasper, C. B. (1990) *Biochemistry* **29**, 9814–9818
- Ruettinger, R. T., Wen, L.-P. & Fulco, A. J. (1989) *J. Biol. Chem.* **264**, 10987–10995
- Russell, P. (1989) in *Molecular Biology of the Fission Yeast* (Nasim, A., Young, P. & Johnson, B. F., eds.), pp. 243–274, Academic Press, San Diego
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
- Schuster, I. (ed.) (1989) *Cytochrome P450: Biochemistry and Biophysics*, Taylor & Francis, London
- Shen, A. L., Porter, T. D., Wison, T. E. & Kasper, C. B. (1989) *J. Biol. Chem.* **264**, 7584–7589
- Sutter, T. R., Sangard, D. & Loper, J. C. (1990) *J. Biol. Chem.* **265**, 16428–16436
- Tomatsu, S., Kobayashi, Y., Fukumaki, Y., Yubisui, T., Orii, T. & Sakaki, Y. (1989) *Gene* **80**, 353–361
- Urenjak, J., Linder, D. & Lumper, L. (1987) *J. Chromatogr.* **397**, 123–136
- Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3857–3860
- Yabusaki, Y., Murakami, H. & Ohkawa, H. (1988) *J. Biochem. (Tokyo)* **103**, 1004–1010
- Yamano, S., Aoyama, T., McBride, O. W., Hardwick, J. P., Gelboin, H. V. & Gonzalez, F. J. (1989) *Mol. Pharmacol.* **35**, 83–88
- Yoshida, Y. & Aoyama, Y. (1991) in *Frontiers in Biotransformation* (Ruckpaul, K. & Rein, H., eds.), vol. 4, pp. 127–148, Taylor & Francis, London

Received 4 March 1992/21 April 1992; accepted 22 April 1992