Isolation and Characterization of the Alkane-inducible NADPH-Cytochrome P-450 Oxidoreductase Gene from *Candida tropicalis*

IDENTIFICATION OF INVARIANT RESIDUES WITHIN SIMILAR AMINO ACID SEQUENCES OF DIVERGENT FLAVOPROTEINS*

(Received for publication, April 9, 1990)

Thomas R. Sutter§, Dominique Sanglard†, and John C. Loper‡**

From the §Department of Environmental Health and ‡Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524

The gene coding for the *Candida tropicalis* NADPH-cytochrome P-450 oxidoreductase (CPR, NADPH:ferrocytochrome oxidoreductase, EC 1.6.2.4) was isolated by immunoscreening of a *C. tropicalis* &gt;111 expression library and colony hybridization of a *C. tropicalis* genomic library. The *C. tropicalis* CPR gene produces a 2.35-kilobase mRNA transcript, levels of which were shown to be increased 16-fold in cells grown on tetradecone relative to cells grown on glucose as the sole carbon source. A 3-kilobase DNA fragment was sequenced, including 554 and 397 base pairs of 5'- and 3'-noncoding sequence, respectively. A single open reading frame of 2040 base pairs was identified and predicts a 76,683-Da polypeptide of 680 amino acid residues. The deduced *C. tropicalis* CPR amino acid sequence was compared with each of the CPR sequences reported from other organisms and invariant residues were identified. Multiple pairwise alignments of divergent members of protein families, previously recognized for their sequence similarities in their respective binding domains for FMN, FAD, and NADPH, have allowed identification of a subset of these invariant residues. From these analyses we infer the importance of 25 of the 680 amino acid residues.

Microsomal NADPH-cytochrome P-450 oxidoreductase (CPR, NADPH:ferrocytochrome oxidoreductase, EC 1.6.2.4)

* Portions of this work were supported by a grant from the United States Environmental Protection Agency and computer services were provided by BIONET National Computer Resource for Molecular Biology. A portion of this work was presented at the 16th International Congress of Genetics, Toronto, Canada, August 20–27, 1988. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M35199.

§ Present address: Chemical Industry Institute of Toxicology (CIIT), 6 Davis Dr., P. O. Box 12137, Research Triangle Park, NC 27709.

† Present address: Dept. of Biotechnology, Swiss Federal Institute of Technology ETH-Honggerberg, 8093 Zurich, Switzerland.

‡ Present address: Dept. of Biotechnology, Swiss Federal Institute of Technology ETH-Honggerberg, 8093 Zurich, Switzerland.

** To whom correspondence and reprint requests should be addressed: Dept. of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, 231 Bethesda Ave., Cincinnati, OH 45267-0524.

The abbreviations used are: CPR, NADPH-cytochrome P-450 oxidoreductase; IPTG isopropyl-β-D-galactopyranoside; FDX, flavodoxin; FNR, ferredoxin: NADP⁺ oxidoreductase (EC 1.18.1.2); P-450, cytochrome P-450. bp, base pair(s).
sequence identify a subset of invariant residues within each of the binding domains of FMN, FAD, and NADPH. These residues are discussed in light of the current biochemical information pertaining to these proteins, and as future targets for site-directed mutagenesis.

EXPERIMENTAL PROCEDURES

Materials—Reagents, enzymes, and standards were purchased from the following sources: affinity purified, [125]I-labeled protein A and [α-32P]dATP, Amersham Corp.; protein molecular weight standards, Bio-Rad; nitrocellulose, Schleicher & Schuell; β-galactosidase antibody, Cooper Biomedical/Cappel Labs; DNA sequencing primer 1216, New England Biolabs; avian myeloblastosis virus reverse transcriptase, Life Sciences, Inc.; RNA molecular weight standards, Gibco/BRL; bovine albumin fraction V, Sigma; tetradecane, Fisher. All other materials were purchased as the best available grade from commercial suppliers.

Strains, Media, and Plasmids—The C. tropicalis strain (ATCC 750) was grown as described previously (37). The vector pAg11 and recipient Echerichia coli strains Y1090(r-) and Y1089(r-) were purchased from the Promega Corporation. E. coli AA102 and plasmid pAA3.7x (38) have been described (25, 26). Conditions for the growth, transformation, and infection of the indicated E. coli strains have been reported (25, 27, 37).

Agt11 Library Screening—The C. tropicalis Agt11 expression library was constructed and screened as described previously (37) with antibody previously produced in this laboratory using purified C. tropicalis CPR protein (39). Briefly, polyclonal antisera against purified C. tropicalis CPR protein were raised in female New Zealand White rabbits (40). Control sera were collected from these rabbits prior to immunization, and all sera were treated at 56 °C for 20 min to inactivate complement and stored in aliquots at −80 °C. Immuno-screening of the expression library (41) used the Blotto procedure (42). Positive signals were detected by autoradiography after [125]I-protein A labeling. For lysogen production of each positive plaque, strains were propagated on agar plates containing appropriate polyvalent antiserum overnight at 4 °C. The filter was immersed in 1 ml of 0.1 M Tris-HCl, pH 7.5, 0.1 M NaCl, 10 mM MgSO4. E. coli Y1089 was infected at a multiplicity of infection of 2.5 and incubated overnight at 30 °C. Lysogens were picked and grown to mid-log phase in 5 ml of LB medium. Phage production was then induced by temperature shift to 44 °C for 30 min. Production of the β-galactosidase fusion proteins was induced by incubating for 1 h at 37 °C in the presence of 10 mM IPTG. Cells were pelleted by centrifugation, resuspended in 0.2 ml of Tris-buffered saline (50 mM Tris-HCl, pH 8.0, 150 mM NaCl), 0.5 mM phenylmethylsulfonyl fluoride, and disrupted by freezing and thawing, three times. Bacterial proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and immunoblot procedures. Affinity purification of CPR-specific antibodies using the fusion proteins produced by recombinant phage clones was as described (37, 43). E. coli Y1090 was infected and plated to yield 106 plaques/100-mm Petri dish. The phage were induced for 4 h at 42 °C, covered with a nitrocellulose filter prewet in 10 mM IPTG, and then incubated for an additional 8 h at 37 °C. The filter was removed, blocked with Blotto reagent, and incubated with the appropriate polyclonal antiserum overnight at 4 °C. The filter was washed several times with Tris-buffered saline, cut into pieces, and the antibody eluted twice with successive 3-ml volumes of 0.2 M glycine HCl, pH 3.0. The combined eluate was neutralized with 2 ml of 1 M Tris-HCl, pH 8.0, and diluted with 20 ml of a 225 mM NaCl solution containing 1.5 g of dry milk and 1 μl of antifreeze A. This neutralized antibody fraction was then used for Western immunoblotting.

Protein Methods—Polyacrylamide gel electrophoresis was performed according to Laemmli (44), using a 6% stacking and 10% running gel. Proteins were visualized by a silver staining method (45). For Western immunoblot (46), proteins were separated by electrophoresis and transferred to nitrocellulose using a Trans-blot cell, according to the manufacturer’s instructions (Bio-Rad). Western immunoblot procedure employed the Blotto technique (42). Signals were detected by autoradiography after [125]I-protein A labeling. Antiserum were diluted 1:100 and 1 × 106 cpm of [32P]protein A was used for each antibody. Signals were determined by the method of Lowery et al. (47) using bovine albumin as the standard.

Recombinant DNA Methods—Unless indicated, recombinant DNA procedures were as described (48), or according to the instructions of the supplier. E. coli plasmid DNA was isolated using the rapid boiling method (49). The C. tropicalis genomic library has been described (37). Hybridization experiments used DNA restriction fragments that were labeled with [α-32P]dATP to high specific activities by a random primer method (50). Genomic library plasmid screening was performed as described previously (29, 51). Plasmid DNA sequencing by the chain-termination method (52) followed the protocol used in this laboratory (25, 26, 27). Plasmid deletion constructs for DNA sequence analysis were generated using a transposon-mediated method (25, 38). Where overlapping deletions were not obtained, the sequence was determined using oligodeoxyribonucleotide primers synthesized using a DNA Synthesizer (Applied Biosystems). RNA was isolated (53), and Northern analysis was performed as described previously (37). For slot blot analysis, the x-ray film was preflashed and exposed for 18 h. Scanning densitometry employed the Hoef er Scientific Instruments GS300.

Amino Acid Sequence Comparisons—These analyses were performed using an optimal alignment algorithm (54) as implemented in the DNANALYZE package (G. Werneck, University of Cincinnati). Parameters were set as: gaps, −4; matches, 2; similar matches, 0.1; mismatches, −2. The deduced amino acid sequence of the C. tropicalis CPR gene was compared to all of the protein sequences listed in the Protein Identification Resource Protein Sequence Data Base, Release 11.0 (National Biomedical Resource Foundation, Washington, D. C.) using XFASTP (54). Multiple alignments were constructed after examination and optimization of a number of pairwise alignments using the DNANALYZE package. This procedure was similar to an approach recently used in this laboratory to identify significant segments of similarity of cytochrome P-450 proteins (55). Sources of the amino acid sequences used were rat (21), rabbit (56), pig (57), fish (58), human (59), S. cerevisiae (28), C. tropicalis (this work), and Bacillus megaterium (60); NADPH-sulfite reductase (EC 1.8.1.2) sequence = Salmonella typhimurium (61); FDIX sequences = Desulfovibrio vulgaris (62), and Acetobacter calcoaceticus (63); FNR sequences = spinach (64) and the algae Spirulina (65); human NADH-cytochrome b, reductase (EC 1.6.2.2) sequence (66).

RESULTS

Gene Isolation—Immunoscreening of a C. tropicalis Agt11 expression library with polyclonal antisera raised against the purified C. tropicalis CPR protein yielded eight positive clones, identified as clones 1, 2, 4, 6, 7, 9, 12, and 13. Immunoblots of cell extracts retrieved three classes of CPR immunoreactive proteins, having apparent molecular weights of 157, 155, or 145 kDa (Fig. 1, lanes 1–5). Parallel samples treated with β-galactosidase antibody detected the same fusion products (Fig. 1, lanes 6–10). Thus, each of these phage clones contains a DNA insert oriented in the proper transcriptional reading frame of the lacZ gene segment of Agt11. The fusion protein of clone 13 showed the largest addition in apparent molecular weight relative to β-galactosidase. The increase, 41 kDa, represents approximately 55% of the C. tropicalis CPR protein coding region.

To confirm the CPR identity from these clones, the fusion proteins produced by clones 13 and 2 were used to generate antisera against purified CPR protein. These antisera were used to identify positive clones from the remaining library (Fig. 1, lanes 3–5). Western immunoblotting of the indicated recombinant phage proteins were separated by electrophoresis and transferred to nitrocellulose using a Trans-blot cell, according to the manufacturer’s instructions (Bio-Rad). Western immunoblot procedure employed the Blotto technique (42). Signals were detected by autoradiography after [125]I-protein A labeling. Antiserum were diluted 1:100 and 1 × 106 cpm of [32P]protein A was used for each antibody. Signals were determined by the method of Lowery et al. (47) using bovine albumin as the standard.

Recombinant DNA Methods—Unless indicated, recombinant DNA procedures were as described (48), or according to the instructions of the supplier. E. coli plasmid DNA was isolated using the rapid boiling method (49). The C. tropicalis genomic library has been described (37). Hybridization experiments used DNA restriction fragments that were labeled with [α-32P]dATP to high specific activities by a random primer method (50). Genomic library plasmid screening was performed as described previously (29, 51). Plasmid DNA sequencing by the chain-termination method (52) followed the protocol used in this laboratory (25, 26, 27). Plasmid deletion constructs for DNA sequence analysis were generated using a transposon-mediated method (25, 38). Where overlapping deletions were not obtained, the sequence was determined using oligodeoxyribonucleotide primers synthesized using a DNA Synthesizer (Applied Biosystems). RNA was isolated (53), and Northern analysis was performed as described previously (37). For slot blot analysis, the x-ray film was preflashed and exposed for 18 h. Scanning densitometry employed the Hoef er Scientific Instruments GS300.

Amino Acid Sequence Comparisons—These analyses were performed using an optimal alignment algorithm (54) as implemented in the DNANALYZE package. This procedure was similar to an approach recently used in this laboratory to identify significant segments of similarity of cytochrome P-450 proteins (55). Sources of the amino acid sequences used were rat (21), rabbit (56), pig (57), fish (58), human (59), S. cerevisiae (28), C. tropicalis (this work), and Bacillus megaterium (60); NADPH-sulfite reductase (EC 1.8.1.2) sequence = Salmonella typhimurium (61); FDIX sequences = Desulfovibrio vulgaris (62), and Acetobacter calcoaceticus (63); FNR sequences = spinach (64) and the algae Spirulina (65); human NADH-cytochrome b, reductase (EC 1.6.2.2) sequence (66).
affinity purify CPR antibody. This antibody was then immunoblotted with purified CPR and with the E. coli lysates of clones 2 and 13. As shown in Fig. 2, each affinity-purified CPR antibody detects only the pure CPR and fusion proteins, thus demonstrating the specificity of the antigen-antibody relationship.

DNA of clone 13 was isolated for analysis. Restriction endonuclease digestion using EcoRI, the linker site used for library construction, revealed a 820-bp recombinant-specific fragment. This fragment was 32P-labeled and used as a probe in Southern hybridization of the EcoRI restricted DNA of the other seven clones. Clone 1 was shown to contain both the 820-bp fragment and two other recombinant-specific EcoRI fragments of 2000 and 545 bp.

Each of these three restriction fragments was gel-purified from clone 1 and used as a hybridization probe both to construct the genomic map (Fig. 3A), and to isolate and map the recombinant clone pTS1 (Fig. 3B) from a C. tropicalis genomic library. Based upon both the pTS1 map and the fusion protein product size of the Agt11 recombinants, the approximate position and orientation of the CPR gene was assigned as indicated in Fig. 3A.

Nucleotide Sequence of the C. tropicalis CPR Gene—From clone pTS1, the 3.7-kilobase pair SpeI fragment (Fig. 3B) containing the entire CPR structural gene was subcloned into pAA 3.7X (38). The sequencing strategy is shown in Fig. 3C.

The complete nucleotide sequence of the CPR gene is presented in Fig. 4. One TATA element is present at position -355 relative to the ATG. This sequence, TATATAT, matches the proposed consensus (67). Long stretches of poly(dT) sequences, often observed in many S. cerevisiae promoters (68), are also found in this 5' -flanking region and resemble promoter elements involved in constitutive transcription of some yeast genes (69). A single long open reading frame of 2040 bp is identified. The 680 deduced amino acid residues predict a polypeptide of 76,683 Da. The TACTAAC sequence, suggesting that this gene has no introns. Multiple copies of the tripartite sequence TAG . . . TATGT . . . TTT, seen for isocytochrome c, a moderately expressed gene of S. cerevisiae (70), are not present.

The codon usage of CPR is similar to that of other genes from S. cerevisiae and from C. tropicalis (25, 26, 75). A quantitative measure of nonrandom codon usage, the Codon Bias Index, compares the codon bias in the gene of interest with the bias seen in highly expressed yeast genes (76). The Codon Bias Index for CPR is 0.48, similar to the value of 0.47 seen for isocytochrome c, a moderately expressed gene of S. cerevisiae (77).

mRNA Analysis—The 820-bp EcoRI fragment containing only CPR coding sequence from pTS1 (Fig. 3B) was used to probe Northern blots of mRNA isolates from glucose- and tetradecane-grown cells. As shown in Fig. 5A, this fragment hybridized to a 2.35-kilobase pair mRNA transcript, an appropriate size for an mRNA encoding the CPR protein. Slot blot analysis of mRNA from glucose-grown and tetradecane-grown C. tropicalis cells shows that the CPR mRNA is elevated 16-fold in cells grown on tetradecane (Fig. 5B). This closely parallels the 20-fold increase of CPR enzyme activity in tetradecane-grown cells seen by Bertrand et al. (78). This large induction of CPR in C. tropicalis is in marked contrast to the mammalian system where CPR induction is usually limited to 2-fold (79). These studies establish that the isolated CPR gene is transcribed in cells grown on glucose or on n-alkane, and along with evidence for a single CPR gene from genomic Southern analysis, support previous work suggesting that only one form of CPR exists in C. tropicalis (39, 78).

Amino Acid Sequence Comparisons—An alignment of the rat and C. tropicalis CPR amino acid sequences is shown in Fig. 6. Also each of the eukaryotic CPR sequences cited under "Experimental Procedures" was examined by pairwise align-
Fig. 4. Nucleotide sequence of the CPR gene. Nucleotide numbering and amino acid numbering of the open reading frame are shown at the right and the left, respectively. Underlined are a putative TATA box in the 5'-flanking region and multiple transcription termination and polyadenylation elements in the 3'-noncoding region.

The B. megaterium CPR sequence was excluded from this alignment since it exists as part of a P-450-CPR fused protein and we can not tell to what extent the differences in amino acid sequence reflect this fact. However, sequence data for B. megaterium CPR was used in the analyses of cofactor binding site domains shown in Figs. 7-9.

In addition, alignments for the binding domains of FMN (Fig. 7), FAD (Figs. 8 and 9) and NADPH (Fig. 9) were constructed from multiple pairwise alignments of divergent members of flavoprotein families. The invariant residues identified in these sets of alignments are discussed below. No significant similarities were identified between any of the sequences of Figs. 7-9 and the reported sequence for adrenodoxin (NADP+-ferridoxin) reductase (82).

DISCUSSION

Earlier comparisons of primary amino acid sequences of CPRs with those of other flavoproteins, as well as chemical modification studies of these enzymes, indicated that for the rat CPR, residues 77-228, 267-678, and Cys-566 are important in the binding of FMN, FAD, and NADPH, respectively (9, 19, 21, 66, 83). In this work we have extended those analyses by incorporating newly reported sequences, including

1 The rat CPR deduced amino acid sequence was the first complete sequence reported (21), and since the mammalian CPR amino acid sequences are approximately 90% similar (66, 67, 99), discussions of all CPR sequences will be referenced to the specific rat amino acid residues.

The small bold symbols in Fig. 6 identify the invariant amino acid residues among all those alignments. The rat and C. tropicalis CPR protein show 34% identity, with an overall similarity of 75% if one considers conservative amino acid substitutions (80). Except for the amino-terminal region, the hydrophathy plots (81) of the yeast and mammalian CPR proteins are very similar. Like S. cerevisiae (28), the C. tropicalis protein contains fewer hydrophilic residues than mammalian CPR. How-

ment to the C. tropicalis CPR sequence. The invariant residues identified in these sets of alignments are discussed below. No significant similarities were identified between any of the sequences of Figs. 7-9 and the reported sequence for adrenodoxin (NADP+-ferridoxin) reductase (82).
16432 C. tropicalis P-450 Reductase

FIG. 3. mRNA analysis. Total RNA, from glucose-grown (g) and tetradecane-grown (t) C. tropicalis was probed with the 820-bp EcoRI fragment of the CPR gene. A, Northern blot of 5 μg of RNA/lane. Numbers on the left show in kilobase pair units the position of RNA standards. B, slot blot analysis. The numbers on the right are the dilution coefficients of the amounts of mRNA/slot, where 1 is equivalent to 2.5 μg. Relative intensity of the signals, determined by scanning densitometry, indicates that the cellular concentration of CPR mRNA (t) is 16-fold higher than CPR mRNA (g).

FIG. 6. Alignment of the rat and C. tropicalis CPR amino acid sequences. Residues are identified by the standard single-letter amino acid code. The specific residues aligned are indicated at the margins of each sequence. Identical residues of the rat (21) and C. tropicalis CPR proteins are underscored by stars. Matches of this alignment that are also conserved when all other known CPR proteins are aligned are shown as small bold letters. The amino-terminal transmembrane segment (81, 115) is overlined (rat residues 25-43). Highly conserved portions of the FMN binding region are boxed (residues 79-202). Proposed FAD binding regions are bracketed (residues 471-678). Residues identified in alignments presented for the binding domains of FMN (Fig. 7), FAD1 (Fig. 8), and FAD2 and NADPH (Fig. 9) are underscored by diamonds.

FIG. 7. The FMN binding domain: amino acid sequence alignment of FMN-containing flavoproteins from taxonomically diverse sources. This alignment was constructed from multiple pairwise alignments of the indicated source sequence as described under "Experimental Procedures." The organism abbreviations are: Ct, C. tropicalis; Bm, B. megaterium; St, S. typhimurium; Du, D. vulgaris; Au, A. vinelandii; SiR, NADPH-sulfite reductase. Residues are identified by the standard single-letter amino acid code. Invariant residues are identified as small bold letters and are underscored by diamonds.

FIG. 8. The FAD1 binding domain. This alignment was constructed as indicated for Fig. 7. The additional organism abbreviations are: Sp, spinach; Spr, spirulina; SiR, NADPH-sulfite reductase. Invariant residues are identified as small bold letters and are over underscored by diamonds.

FIG. 9. The FAD2 and NADPH binding domains. This alignment was constructed as indicated for Fig. 7. The additional organism abbreviations are: Sp, spinach; Spr, spirulina; Hum, human; b5R, NADH-cytochrome b reductase; SiR, NADPH-sulfite reductase. Invariant residues are identified as small bold letters and are over underscored by diamonds.

the C. tropicalis CPR sequence reported here, and by including sequences from other flavoprotein families, with specific members being chosen on the basis of their belonging to taxonomically divergent organisms. We reasoned that, by constructing multiple alignments from reiterative pairwise optimal alignments of these chosen amino acid sequences, we might identify subsets of invariant amino acid residues within the similar sequences. Alignment of the C. tropicalis CPR amino acid sequence with seven eukaryotic CPR sequences revealed that 147 of its 680 amino acid residues are invariant. These invariant residues are shown as small bold letters and are over underscored by diamonds.

Downloaded from http://www.jbc.org/ on October 28, 2017
Since biochemical studies have demonstrated the importance of specific residues not identified by our methods, our discussion of each cofactor binding domain considers both sets of data.

The Transmembrane Domain—The common region of the transmembrane segment is identified by an overline in Fig. 6, residues 25–43. Two transmembrane residues, Leu-33 and Tyr-40 are invariant in each reported CPR sequence. Leu-33 is also present in the corresponding position of the NADPH-sulfite reductase primary sequence (61). The 19-residue transmembrane segment of C. tropicalis CPR resembles a simple eukaryotic transmembrane segment (84) with respect to hydrophobicity and charge surrounding the hydrophobic segment, although it is shorter than the analogous mammalian CPR segments. Thus both C. tropicalis (this work) and S. cerevisiae (28) CPR polypeptides have shorter amino-terminal sequences than their mammalian counterpart(s). Despite these differences, both the rat CPR (85) and the C. tropicalis CPR (29) proteins, when expressed in S. cerevisiae, are localized to the microsomal fraction.

The FMN Binding Domain—Based upon data obtained using a fluorescent cysteine probe and tryptic peptide mapping, it has been shown that the FMN binding domain of CPR is in the amino-terminal portion of this protein (18). The highly conserved segments of this domain (previously identified, see Ref. 21) reside between residues 79 and 202 and are boxed in Fig. 6.

Evidence in this figure are several discrete blocks of similarity within this domain that are common among all CPR sequences. The multiple alignment presented in Fig. 7 identifies six invariant residues: Ser-86, Gly-89, Ala-95, Phe-171, Gly-174, and Gly-196. The FXD serine corresponding to CPR Ser-86 has been shown by x-ray studies to side-chain bond to a phosphate oxygen of FMN (86). Alignments of nine flavodoxin sequences (62, 87) show that the glycine corresponding to CPR Gly-89 is invariant in that series. Moreover, where crystallographic structures were determined, this residue was shown to define a limit of the first helix (87).

It was suggested that CPR residues Tyr-140 and Tyr-178 are functionally similar to residues Trp-60 and Tyr-95 of the D. vulgaris FDX protein (21). In the latter, these residues stabilize the FMN prosthetic group through stacking interactions between the tryosyl and tryptophanyl aromatic side chains and the flavin isoalloxazine ring (86). Nuclear polarization studies also provide evidence that in CPR, a tryosyl residue is involved in FMN binding (20). Recently, the functional importance of rat CPR residues Tyr-140 and Tyr-178 was determined by site-directed mutagenesis (17). Substitution of these residues by phenylalanine resulted in proteins whose cytochrome c reductase activity and FMN binding were comparable to wild type. However, replacement by aspartate, a charged amino acid, resulted in a significant decrease in activity (either Tyr-140 or Tyr-178) or in FMN binding (only Tyr-178).

As is shown in Fig. 7, both Tyr-140 and Tyr-178 and their adjacent residues are invariant among all CPR primary sequences, except in the E. megaterium sequence, where the tyrosine corresponding to CPR Tyr-178 is replaced by another aromatic amino acid, tryptophan. By contrast, in non-CPR polypeptides the position corresponding to CPR Tyr-140 is variable and has been positionally substituted by aromatic (tryptophan), polar (glutamine) or aliphatic (leucine) amino acid residues. These analyses suggest that while residues Phe-171 to Asn-189 may provide the functional core for binding of the FMN isoalloxazine ring, the segment of residues from Thr-139 to Gly-143 is more likely to be important to the reductase activity of this CPR protein. This inference, based upon the alignments presented in Figs. 6 and 7, is consistent with the information obtained by site-directed mutagenesis (17), and provides a rationale for further mutagenesis studies of these CPR segments.

The many hydrophobic residues found in these FMN binding regions are believed to shield and stabilize the isoalloxazine ring (19). Of the 44 phenylalanine residues in the C. tropicalis CPR, only seven are positionally conserved among all CPR proteins (Fig. 6). That 4 of these 7 residues, 135, 152, 171, and 181, occur in this region suggests that phenylalanine may be of particular importance in these positions of the CPR protein. The phenylalanine corresponding to CPR Phe-171 is invariant in the alignment presented in Fig. 7, and by this criterion is a target for site-directed mutagenesis.

The CPR-P-450 Charge-pairing Domains—The mechanism of electron transfer from microsomal CPR to P-450 is not known, but studies of electron transfer complexes have shown the importance of complementary electrostatic interactions (89–93). For CPR, chemical modification of carboxylic amino acid residues demonstrated the importance of electrostatic interaction to electron transfer from CPR to both physiologic (P-450) and nonphysiologic (cytochrome c) acceptors (93–95). CPR and P-450 form a functional electron transfer complex through the interaction of carboxyl residues of CPR and amino residues of P-450 (93, 96), and formation of this complex is facilitated by substrate, NADPH, and lipid (93).

Insight into the nature of these interactions has come from studies of cytochrome c and flavodoxin (89, 97). FDX and cytochrome c interact through the complementarity of charged residues near the partially exposed surfaces of their respective flavin and heme prosthetic groups (98, 99). Using computer modeling, it was shown that salt linkages were favored between the lysines of the heme-containing crevice of cytochrome c and the aspartate and glutamate residue side chains of Clostridium MP FDX. In a stereoscopic view of this model, two groups of acidic residues positioned on either side of the flavin of FDX were localized to the segment at amino acid residues 55–65, on one side, and to residue 13 and residue 120, on the other (99).

In Fig. 6, three discrete groups of acidic residues are positioned conservatively among all CPR sequences. These are: 1) Asp-113, Asp- or Glu-116, Asp-118; 2) Glu-142, Asp- or Glu-144, Asp-147; 3) Asp-207, Asp-208, Asp- or Glu-213, Glu-214, Asp-215. Positionally, groups 2 and 3 are reminiscent of the FDX model, and support similar predictions of secondary structure (19, 83) and chain flexibility (83). Moreover, chemical cross-linking was achieved in a study of the rat CPR-horse cytochrome c protein complex, and amino acid sequence analysis revealed that the cross-link occurred between Lys-13 of cytochrome c and one of 6 side chain carboxyl residues of CPR (100). Five of these 6 residues are the invariant members of group three, above. Therefore, the third group of acidic residues contains likely candidates for electrostatic charge pairing of CPR with P-450. However, none of the other charged groups should be discounted, since it is now accepted that in addition to specific residue interaction, the sum of the charge pair interactions and steric components also contribute to net complex binding and electron transfer (89, 95, 97).

Although it was generally believed that aromatic side chains were not involved in electron transfer intermediates (91), computer modeling predicted (101), and site-directed mutagenesis demonstrated, that the substitution of a non-aromatic amino acid residue at position 82 of cytochrome c resulted in a 10-fold decrease in electron transfer (102, 103). There are too many aromatic residues in this region of the CPR poly-
peptide to speculate on the importance of any one residue. However, by applying positional conservation as a criterion, the number of candidate aromatic residues in this CPR region is reduced by more than one-half (Fig. 6). Moreover, 1 aromatic residue, Phe-171, is identified as an invariant residue (see Fig. 7).

The FAD Binding Domain—X-ray structure analysis of FNR has shown that this protein contains two distinct domains that bind FAD and NADP⁺, and that the former is localized to the amino-terminal half of this protein (104). Alignment of the amino acid sequence of the FNR polypeptide with CPR and other flavoprotein sequences had permitted tentative identification of the segments of the CPR polypeptide involved in FAD binding, but these studies were limited by the low similarity of the sequences used (9, 19, 57).

The alignments presented here in Figs. 8 and 9 strengthen these earlier analyses, and identify 6 invariant amino acid residues in two discrete segments of the CPR polypeptide (overlined in Fig. 6), that are likely to be involved in the functional FAD domain of this protein. These residues are: His-302, Tyr-314, Gly-317 (Fig. 8), and Arg-454, Tyr-456, Ser-460 (Fig. 9).

The NADPH Binding Domain—The glycine-rich segment (CPR residues Gly-531 to Gly-536 in Fig. 6), characteristic of a nucleotide binding domain (105, 106), has been localized to the NADPH binding site of FNR (104). This conclusion is supported by sequence and secondary structure predictions for FNR (64) and CPR (19).

The alignment presented in Fig. 9 identifies 13 invariant residues in this NADPH (NADH) binding domain. These residues are: Gly-488, Gly-534, Thr-535, Gly-536, Pro-539, Leu-562, Leu-574, Arg-597, Tyr-604, Gln-606, Cys-630, Gly-631, and Met-636. Four of these residues (Gly-534, Thr-535, Gly-536, and Pro-539) occur in the glycine-rich region of the conserved nucleotide binding segment, where they define a unique subset of this segment (Fig. 9). The consensus sequence for this subset is Gly-Thr-Gly-Ile/Val-Xaa-Pro. The first glycine of this consensus corresponds to the residue believed to provide space for the close approach of the pyrophosphate moiety of the dinucleotide (105, 107).

Data from several chemical modification studies provide evidence that this nucleotide binding domain is structured from a large number of residues in the primary amino acid sequences aligned in Fig. 9. Residues localized to the NADPH or NADH binding domains of their respective proteins include: Cys-472, pig CPR (83); Lys-110, steer liver NADH-cytochrome b₅ reductase (108); Lys-116, spinach FNR (109); Cys-565, pig and human CPR (57, 110); Lys-244, spinach FNR (111); Cys-283, steer liver NADH-cytochrome b₅ reductase (112).

In Fig. 9, these protected residues are not invariant among the sequences aligned. Moreover, even within a given protein family, substitution of the protected residue can occur. For example, the protected cysteine (57, 110) of mammalian CPR is replaced by serine in the S. cerevisiae CPR sequence (28, 29). Also, some residues such as leucine are not easily modified in a specific manner and their importance cannot be easily addressed by such approaches. Thus, in addition to chemical modification experiments, alignments such as those reported here provide useful information concerning the importance of specific amino acid residues.

Structural Domains—CPR is a complex protein with multiple structural domains to bind FMN, FAD, and NADP⁺. Presumably these domains must interact in a highly specific manner to allow effective electron transfer from NADPH to P-450. The three most positionally conserved amino acid residues among all CPR proteins are tyrosine, proline and glycine. This was calculated by dividing the number of times a residue was positionally conserved among all CPR proteins (Fig. 6) by the number of times that amino acid occurs in the protein. As discussed for the FMN binding domain, tyrosine may have a special role in the stabilization of the CPR flavin prosthetic groups. Glycine and proline are often important to the secondary and tertiary structures of proteins. Eight of the 25 invariant residues identified in the alignments in Figs. 7, 8, and 9 are glycines. Proline, however, is uniquely identified within the CPR alignments (Fig. 6). Two interesting segments of residues occur at residues Pro-275 to Pro-281 and Pro-360 to Pro-364 (Fig. 6). The function of these invariant residues is unknown, but they are potential candidates for amino acid segments of the CPR polypeptide that are involved in folding of this protein.

Unique Yeast Sequences—In the alignment of the rat and C. tropicalis CPR sequences, three gaps are required for optimal alignment of the residues in the NADPH binding domain between residues Arg-495 to Val-513 (Fig. 6). This is also true for the S. cerevisiae sequence (28). The significance of this is unknown. It could reflect the understanding that unlike the mammalian CPR, the yeast CPR must interact with only a small number of P-450 proteins. Alternatively, it could reflect differences in the cellular environment in which these proteins function. Nonetheless, this difference in primary amino acid structure between yeast and mammalian CPR might provide the basis for a yeast-specific CPR inhibitor targeted to the NADPH binding domain. Such structural differences provide the basis for the antibiotic effectiveness of trimethoprim through its selective inhibition of the bacterial dihydrofolate reductase (113).

Alkaline assimilation by C. tropicalis is a trait shared by numerous Candida species, including the human pathogen C. albicans, and by species among several other genera of yeasts widely distributed in the environment (114). Our isolation and sequence determination of the C. tropicalis CPR gene is a further step toward characterization of the cytochrome P-450-catalyzed alkaline hydroxylaction which initiates this assimilation (20), as is our demonstration that steady state levels of this gene's transcript are increased in response to growth to tetradecane. In addition, we have used the deduced amino acid sequence of this lower eukaryotic CPR gene to examine primary protein structure among representatives of this highly conserved oxidoreductase and also with respect to other more functionally diverse flavoproteins. These comparisons, when considered in light of available biochemical information, suggest specific hypotheses for the importance of the identified invariant amino acid residues, hypotheses currently testable by site-directed mutagenesis.

Acknowledgments—Our thanks to Dr. Vernon Kalb for providing the yeast genomic library; to Greg Wernke for aid in performing the computer-assisted analyses; to Melissa Thomas, Jennifer Schroeder, and G. Wernke for their assistance in the preparation of Figs. 6; to M. Thomas for the preparation of this manuscript. We also thank Dr. Elmer Price for his critical review of this manuscript; Dr. V. Kalb, Dr. Rick Thompson, Dr. Chien Chen, Tom Turri, and Chitta Dey for their helpful discussions and assistance during the course of this research, and Eartell Brownlow for her technical assistance.

REFERENCES
C. tropicalis P-450 Reductase

Additions and Corrections

Vol. 265 (1990) 9628–9633

An 18-kDa glycoprotein from bovine nasal cartilage. Isolation and primary structure of small cartilage-derived glycoprotein.

Peter J. Neame, James T. Treep, and Carmen N. Young

Page 9629, Fig. 1: The illustrated structure of the protein did not agree with either the miniprint or the text. The correction involves the addition of a threonine in front of the first N-linked oligosaccharide and the removal of an aspartate (D) at what appeared to be residue 42 in the original figure. This changes the apparent numbering of amino acids and positions of substituted amino acids prior to amino acid 43 (phenylalanine). The figure legend remains unchanged; N-linked oligosaccharides are shown as filled triangles and the O-linked oligosaccharide is shown as a filled oval. The disulfide bonds are shown as lines. The disulfide bonds at amino acids 69 and 70 may be the other way around (see Miniprint). The sequence as submitted to the National Biomedical Research Foundation is correct.

Vol. 265 (1990) 16428–16436

Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from Candida tropicalis. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins.

Thomas R. Sutter, Dominique Sanglard, and John C. Loper

The name of an author was misspelled. The corrected name is shown above.

We suggest that subscribers photocopy these corrections and insert the photocopies at the appropriate places where the article to be corrected originally appeared. Authors are urged to introduce these corrections into any reprints they distribute. Secondary (abstract) services are urged to carry notice of these corrections as prominently as they carried the original abstracts.
Isolation and characterization of the alkane-inducible NADPH-cytochrome P-450 oxidoreductase gene from Candida tropicalis. Identification of invariant residues within similar amino acid sequences of divergent flavoproteins.

T R Sutter, D Sanglard, J C Loper and D Sangard


Access the most updated version of this article at http://www.jbc.org/content/265/27/16428

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/27/16428.full.html#ref-list-1