Isolation and sequence determination of a cDNA clone related to human cytochrome P-450 nifedipine oxidase

(gene cloning/polymorphism/dihydropyridines/calcium blockers)

PHILIPPE H. BEAUNE, DIANE R. UMBENHAUER, RICHARD W. BORK, R. STEPHEN LLOYD, AND F. PETER GUENGERICH*

Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, TN 37232

Communicated by Elizabeth C. Miller, July 18, 1986

Human liver cytochrome P-450_{NF} is the form ABSTRACT of cytochrome P-450 responsible for the oxidation of the calcium-channel blocker nifedipine, which has been reported to show polymorphism in clinical studies. By screening a bacteriophage $\lambda gt11$ expression cDNA library, we isolated two clones: NF95 with an insert length of 0.8 kilobases which gave a stable fusion protein and NF25 with an insert length of 2.2 kilobases. The two clones were both sequenced and shown to be identical in their overlapping section. The sequence of NF25 is 77% similar to that reported for a rat cytochrome "P-450_{PCN}" cDNA (PCN = pregnenolone- 16α -carbonitrile). The similarity decreases to 45-53% when the sequence is compared to human cytochromes P-450 belonging to other families [i.e., "pH P-450(1)," "P1-450," "P3-450," and "P-450_{MP}." The deduced amino acid sequence is 73% similar to that of rat cytochrome P-450_{PCN}, and the first 21 amino acids are identical to those reported for human liver cytochrome "P-450p." Sections of these clones were nick-translated and used as probes for analyses of human mRNA and genomic DNA. The number and size of bands indicate that P-450_{NF} belongs to a multigene family, the so-called pregnenolone-16 α -carbonitrile-inducible family.

Cytochrome P-450 (P-450) plays an important role in the oxidation of drugs and carcinogens as well as endogenous substrates. Interindividual variation in oxidative metabolism can be attributed, at least in part, to the composition of individual P-450 forms, and these differences are partly due to genetic factors. Since 1977, several genetic polymorphisms of drug oxidation have been demonstrated (1-3), and in some cases the involved form of P-450 has been identified in humans (4, 5). Recently Kleinbloesem et al. (6) reported that oxidation of nifedipine, a vasodilator and calcium-channel blocker, was distributed in a polymorphic manner-17% of the Dutch population studied were phenotypically poor metabolizers. Recently we identified and purified the human liver P-450 form that is responsible for this oxidation of nifedipine (7). This protein, P-450_{NF}, was shown to be related or identical to P-450s previously isolated from human liver in this laboratory (8). To better understand the mechanism underlying this polymorphism, we used polyclonal and monoclonal antibodies to screen a human liver bacteriophage λ gt11 cDNA expression library. The selected clones were analyzed, sequenced, and used to prepare nick-translated probes for the analysis of mRNA and genomic DNA.

MATERIALS AND METHODS

Enzymes and Antibodies. Human livers were obtained through the Nashville Regional Organ Procurement Agency,

and protocols were approved by the Vanderbilt Committee for the Protection of Human Subjects. Livers were perfused immediately after circulatory arrest, chilled on ice, and brought to the laboratory. The livers were cut in small pieces, frozen in liquid nitrogen, and stored at -70° C (8).

P-450_{NF} was purified as described, and polyclonal and monoclonal antibodies were produced (7, 9). These antibodies recognized a single band migrating with purified P-450_{NF} when human liver microsomes were electrophoresed and immunoblotted (7, 9). Before use in screening the λ gt11 library, antibodies were adsorbed twice overnight at 4°C with *Escherichia coli* (BNN97) (10) lysate bound to CNBr-activated Sepharose 4B (Pharmacia, Piscataway, NJ) in order to eliminate reaction with *E. coli* proteins; after immunoadsorption, the antibody reactivity toward P-450_{NF} and toward *E. coli* was assayed.

Library Screening. The human liver cDNA phage $\lambda gt11$ expression library was a gift of G. A. Ricca and W. Drohan, Meloy Laboratories (Springfield, VA). It was screened with both polyclonal and monoclonal antibodies basically as described by Young and Davis (10); the nitrocellulose filters were developed with an immunochemical technique (11) using 4-chloro-1-naphthol in place of 3,3'-diaminobenzidine (12). Antisera or ascites fluids were used at dilutions of about 1:50. Positive plaques were checked and plaque-purified by at least two additional rounds of dilution and screening. High-titer phage stocks were purified through CsCl step and equilibrium gradients (13). λ gt11 DNAs were prepared for restriction mapping and subcloning in phage M13 (14). Lysogens were prepared by infection of Y1089 E. coli with λ gt11 phage clones, and colonies were selected for growth at 30°C and not 42°C. Fusion proteins were produced by initially growing cells in exponential phase at 30°C, heat-shocking them at 42°C for 15 min, and then growing them at 37°C in the presence of 10 mM isopropyl- β -D-thiogalactose until the A_{600} stabilized (between 0.5 and 2 hr). Cells were then collected by centrifugation at $10^4 \times g$ for 10 min, and the pellet was solubilized for electrophoresis, with subsequent visualization of bands by silver staining or immunoblotting.

Subcloning and Sequencing. $\lambda gt11$ DNA containing inserts was digested with *Eco*RI, *Hin*dIII, *Sac* I, or *Hin*dIII/*Sac* I and was subcloned in M13 phage as described (15), except that the UT481 strain was used instead of JM101 or JM103. *E. coli* UT481 was constructed by and obtained from C. Lark (Salt Lake City, UT). M13 plaques containing inserts were plaque-purified, and single- and double-stranded DNAs were prepared as described (15); the orientation was checked by asymmetric restriction digestion (i.e., *Pst* I or *Bam*HI).

Single-stranded M13 DNA was used as a template for DNA sequencing by the dideoxy termination method (16). DNA sequencing kits were obtained from New England Nuclear

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviations: P-450, liver microsomal cytochrome P-450; PCN, pregnenolone- 16α -carbonitrile; kb, kilobase(s). *To whom reprint requests should be addressed.



and used according to the supplier's instructions except that the incubation temperature was 37°C and no NaCl was included in the buffer. Additional sequencing primers were 18-mers or 20-mers synthesized with a BioSearch Sam-One Series II DNA Synthesizer; oligonucleotides were purified as described by Lloyd *et al.* (17). Sequences were compared with the aid of computer access to the protein sequence data base of the National Biomedical Research Foundation.[†]

Blotting Analysis of RNA and DNA. Genomic DNA and total RNA were isolated from the same preparation by using the CsCl cushion method described by Chirgwin et al. (18). The RNA was collected as a pellet on the bottom of the centrifuge tube, dissolved in water, and precipitated with 0.1 vol of sodium acetate and 2 vol of ethanol. The DNA was collected as a viscous solution at the interface between the CsCl and the homogenate. This solution was extracted once with phenol/CHCl₃, 1:1 (vol/vol) and once with CHCl₃ and then was precipitated with ethanol. The DNA was dissolved in 10 mM Tris-HCl buffer, pH 8.0/1 mM EDTA, treated with proteinase K (50 μ g/ml) in 0.1% sodium dodecyl sulfate, and reextracted with phenol/CHCl₃ and CHCl₃ as above. The purified DNA was then precipitated with ethanol. For Southern blots, 20 μ g of genomic DNA was cut with various restriction enzymes and electrophoresed through a 0.8% agarose gel. The gel was processed by the Wahl et al. (19) modification of the Southern (20) procedure. The DNA was transferred to GeneScreenPlus (New England Nuclear) and processed as suggested by the supplier. For RNA blot analyses, 15 μ g of RNA was submitted to agarose/ formaldehyde electrophoresis as described (21). RNA was then transferred to GeneScreenPlus (New England Nuclear) and processed as suggested by the supplier. DNA and RNA blots were probed with nick-translated DNA inserts ($\approx 10^8$ $dpm/\mu g$). DNA fragments were purified from agarose gels basically as described (22). Hybridization and washings were performed as described by the supplier. Filters were then autoradiographed with Kodak XAR film (Kodak) with two screens from 3 to 7 days at -70° C.

FIG. 1. Restriction map of clones NF95 and NF25 and sequencing strategy. The hatched region is the coding sequence. Arrows indicate the direction and extent of sequence determined for cDNA fragments subcloned in M13mp19 or M13mp18. Recognition sites for restriction nucleases are indicated. Except for Taq I, these were all determined experimentally and confirmed by computer analysis of the sequence.

RESULTS AND DISCUSSION

Isolation and Characterization of cDNA Clones. A human liver cDNA expression library in bacteriophage $\lambda gt11$ was screened with polyclonal and monoclonal antibodies raised against P-450_{NF}. After the screening of a total of 3×10^5 plaques, 15 clones were positively identified by both antibodies. These clones were characterized by insert length and by the presence of fusion proteins in lysogen preparations. Two clones were selected: NF25, the insert of which was long



FIG. 2. (A and B) Electrophoresis and immunoblotting of $\lambda gt11$ lysogen proteins: immunoblotting with anti-P-450_{NF} (A) and silverstaining of the gel (B). Lanes: λ , E. coli lysogen infected with λ gt11 bacteriophage without insert; 95, E. coli lysogen infected with λ gt11 clone 95; -, cells grown at 30°C pelleted and electrophoresed; +, cells grown at 30°C, heat-shocked 15 min at 42°C, and treated with isopropyl- β -D-thiogalactose. Molecular mass is shown in KDa. P-450 and Gal indicate the migration positions of P-450_{NF} and β -galactosidase, respectively. The arrowhead at the right indicates the migration of the fusion protein. Clone NF25 gave results similar to those in lanes λ . Development of immunoblots with anti- β -galactosidase visualized β -galactosidase in lane λ + and the fusion protein with a higher molecular mass in lane 95+. (C) Various amounts of purified P-450_{NF} (from sample HL 93) (7) were spotted on nitrocellulose with a "dot" apparatus, incubated with a 1:50 dilution of rabbit antibody raised against β -galactosidase/clone NF 95 fusion protein (see text), and developed as described (11, 12).

[†]National Biomedical Research Foundation (1986) Protein Sequence Data Base of the Protein Identification Resource (Washington, DC), Release No. 8.0.

and the second s	Ala GCTC Asp Tyr TT TT Pho GGA C C C C C C C C C C C C C C C C C C	Lew CTC Giy GGA Aan The TGG Val GTG ATC GVal Lys AAA CTT TA ACA GCA GCA GCA GCA CTT TA TYP Net ATC CTT GTG GIN ACA GCA GCA GCA GCA GCA GCA GCA GCA GCA	He ATC CCT Leu Thr ACC Val Giy GGC Val Giy GGC Lys AAA G Ser TCT Leu CTC Glu GAG Ser TCA Leu AAA AA AA GAA GAA GAA GAA GAA GAA GAA	Pro CCA T Ser His CGC Ang C TTG C Blu GAA GAA GAA Val AGA Val AGA Val AGA C Val C C C C C C C C C C C C C C C C C C C	Asp GACC CT Ala Ser TCA A A Ser TCA A A Ser TCA A A Ser Cys TGT C Cys TGT C Cys GGC Gly GGC Gly GGC C C TTT Leu TTA A Ser TTT C A Ser TCA A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA C A A Ser TCA A A Ser TCA A A Ser TCA A A Ser TCA C A A Ser TCA C A A Ser TCA C A Ser TCA C Ser TCA C Ser TCA C Ser TCA C Ser TCA C Ser Ser TCA C Ser TCA C Ser TCA C Ser TCA C Ser TCA C Ser Ser TCA C Ser TCA C Ser C Ser C Ser S Ser S Ser S Ser S Ser S S S S S S	Lew TTG CTC His CAT TY TAC Asp GAT TAT T TPhe Glu GAG AAT GAG GAT Lys AAG GAT C Asp GAT AAT C	Alla G GC A C A The Gly GG A His T T T T T T T T T T T C T Asp G C T Val G T C C T Pro C C T Asn A A C C T T T T T T T T T G T C	Met CTG Leu CTT CTT CTT CTT CTT Gh Gh GTC CCT CCT CCT CCT CCT CT CT CT TGT TG	Giu Gia A Phe T T T T Giy GG C T T C A G G A A G A T Me A C C A Sp G A A T Me A C C A Sp C A Sp C C A Sp C A Sp C C A Sp C C A Sp C C A Sp C C C A Sp C C C C A Sp C C C C C C C C C C C C C C C C C C C	The ACC Lys AAG Phon TTA Lev Pro CCT The ACA The ACA The ACA The See TCT CCA G	Trp TG G G G TG G TG G TG G TG G TG G C Leu Aan A Aa C C C C C C C C C C C C C C C C C	Leu G C T T G C C Val Leu C T T A G A A A A C G G C G G C A G G A C A G A Arg A Arg A Arg Arg Arg Arg Arg Arg Arg Arg Arg Arg	Lew CTC Gily A GGA G GCT A GCT C Alle GCT C A Man A An T Yr T A T Yr T A T Yr T A T C C C C C C C C C C C C C C C C C	Lew CTG Aap GACT T He ATC CCT GA Aap Aap Aap Aap CCT A Aap CCA Ser TCA	Ala GCT A Pro CCT Met ATG Val Thr ACG CA GI GC GC G GGG G GGG G GGG CA A A CA A A A	Val GTC Gly GGG GAA GAT GAA GAT TTG GAT GAT Ala GCC Aap GAC TT The	Ser AGT Val Pro CCT Cys TGT CCA A A TH Pro CCA Leu TTG CCA TT TAC Pro CCC T TY TAC C T Yal	Leu CTG ACA A Lys CAT GTG GAC A CAT CAT CAT CAT CAT CAT CAT CAT CAT	Val G T G Pro C C T Lys A A A Meet A T G G G A G G G G C C A C C C A G C C A C C A C C T C A G G T G Val G T G Pro	Lew CTC Law CTG Lys AAG He ATC AAG AAG AM AAC CTT The ACCG CTT The CTT ACCG CTT CTT ACCG CTC CTG CTG CTG CTG CTG CTG CTG CTG	Lew C T C Pro C C T Tyr T A T Lys A A A A Met A A T C C T G C T G	Tyr T A 1 T A 1 T A 1 T T 1 G G J G G J Thr A C G A G G G G
ana ang ang ang ang ang ang ang ang ang	Tyr Tr Tr Pho GGA GGA C C Val GGA C C C T GGA C C T GGA C C T GGA C C C T GGA C C C C C C C C C C C C C C C C C C	Gly GGA Asso Asso Asso Asso Asso Asso Asso Ass	The ACC Ble ATT GVal Gly GCC Lys AAA G Ser TCT Leu CTC Glu GAG Ser TCA Leu AAA A GAA G AAA C	His C A TR G C T T G C Phe T T G Leu G Leu G Leu G Leu G A A C U G A A C Val Lys A A G Thr A C A Phe T T T A C A C Val C C C C C C C C C C C C C C C C C C C	Ser TCAA A TTCA Ser TCCT CA A Ser TGT CC Cys TGT CC Cys GGC Gly GGC CGly GGC TTTA Ser TTA	His CAT Tyr TAC Asp GAT, Tyr TAT T Phe Glu GAG A ASP GAT C Val GAG GAT C Asp AAT C Asp	Gy GGA Hais T T Tyr Gly GGT TCT Val GTC Val GTC Pro CCT Ason AAC Phe TTT Hile ATC	Levu CTT Lynn Met Ghn GCAA Val GTC CCT CCT CCT CT Val GTC Val TTG TTPhe Cynn CCynn Cynn Cynn Cynn Cynn CCynn CCT CCT CCT CCT CCT CCT CCT CCT CCT C	Phe TTT Gly GGC GGC T TC Glu GAA GA GA C C Asp GAT GAT C C Asp GAT	Lys A A G Phe T T T A Lev Pro C C T Thy A C A The A T T Lev Lev T T G A Met Ser T C T C C C A Pro C C A G	Lys A A G Ys TGT G GT G GT G GT G C Leu A A G A A G C Lys A A G L C C C C C C C C C C C C C C C C C C	Lew C T T A Gin Met d A T A A T A Lys Lew C T G G C G C G G C A G G A C A A C A G A Arg A A C A G A A A C A G A A A C A G A C A G A A A C A C A C A C A C A C A C A C A C	City GGA A Phe GCT C C Ang GCT C C Ang Ang C C Leu T Tyr Tyr TAT Tyr Tyr TAT C C C C C C C C C C C C C C C C C C	He A T T Amp G A C T He A T C Pro C C T G A Amp Amp G A Amp Amp G A Amp C C A Phe T T T C A Ser T C A	Pro T Met ATG G Val The ACG Phe TTT Ser TCA G G G G G G G G G G G G G	Giy GGG Giu GAA G GGT C C GGT C C C C C C C C C C C C C	Pro C Cys T G T T C C Pro C C A T T Pro C C A Leu C T G T T C C T T T A C T Y T A C T Y T A C T Y T A C Y S S S S S S S S S S S S S S S S S S S	The A A A A A A A A A A A A A A A A A A A	Pro C C T Lys A A A Men A T G G G A G G A G G A G G A C C C A G C C C A G A G A G A G A G A G G T G G T G Fro	Leu CTG Lyn AAG He ATC TTTA A IIe TTTTA A CG GAT GAU GAA G GA Phe	Pro CCT Tyr TAT Lys AAA Met ATG TTC Lovu CTG T GTG Amm AACA Lys Lovu CTG	Phe TT1 Gly J GG J Ther A C J A A A GG G Gly J A C G A A C G
ena al al ar to the second and the second al al ar to the second al ar tott al ar tott al ar to the second al ar tott al ar to	Gity GGA GT G GA GT G A GT G A GC C T Gity GGA GA A GA A GA A GA A GA A GA A GA A	Am A C C C C C C C C C C C C C C C C C C	Be AT G G G G G G G C C C C C C C C C C C C C	Leu TTG C TTG Leu Glu GAA Val Lys AAG Val Lys AAG Thy ACA Phe TTT AGA Val C AGA C GLU GLU GAA	Ser TCCA AAT AAT AAT AAT AAT AAT AAT AAT AA	Tyr TAC Amp GAT, Tyr TAT TAT T Phe GA GA GA GA C C C Amp GAT C C Amp GA TAT C C Amp	His CAT CAT Tyr GGT GGT CT Asp GAT CT Val CT CT Pro CCT Asn AAC Phe TTT Ilie ATC	Lya A A G Gin G C A Val G C C C C C C C C C C C C C C C C C C C	Giy G G C G T G M C G G M T Met T T C G M G A T C G M C C M G A T C C M G A T C C M G A T C C M G C T C T C T C T C T C T C T C T C T C	Phe TTTA A Leu PCCT Thr ACA TTP TGG Hie ATT Leu TTGA Met C C C C T C C C T Pro C C C T Pr C C T T P T G G C T T P C C T T P C C C T T P C C C T T P C C C T T P C C C T P C C C T P C C C T P C C C T P C C C T P C C C T P C C C C	Cys TGT TGT Trp GTGC Leu Asn AAG GCC AAA GLys AAAG Leu CTC TT TCC G GCC C AAA GU GCC C C C C C C C C C C C C C C C C	Met A T GG A A Lys C T G T T Phe Arg C G G C A G A Arg A A A C Arg A A A A C Arg A A A A C A A C A C T G C G G C A C A A A A C T G C G G C A C A A A A C A A A C T G C A C A A A A C T G C G G C A C A A A A C T G C G G C A C A A A A C A A A C T G C G G C A C A C A A A C A C A A A C A C A A C A C A A C A C A C A	Phe TTT Ala GCT C Leu TTA AGG C Leu TTA TTY TY TAT Val GTC G GCT C Leu C TC C C C C C C C C C C C C C C C C C	Amp GAC T Ile ATC Proo CGA Amp GGA GGA GIV GGA Phe TTT T CCA	Met A T G G Val Thr A C G Phe T T T C A C G G C G G C G G G C G G G C C A A A C A C	Glu GA A G GA T G GA T C GG T C Leu T T G G T G T He Ala G C C Asp GA C T T TA A Asp	Cys TGT C CCT A A The Pro CCA T CCA Leu CTG CCA TTG TTG CCA TTG CCA T TY TAC C T TY TAC C T T T T T C T T C C T C T C C T C C T C C T C C T C C T C C T C C T C C C T C C C C C T C	His CAT Asp GACA Giu Val GTG Ser TCT C Val GTG GTG A Ser A GC Phe	Lys A A A Met A T G Gly G G A G G G A G C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C C C A C	Lys A A G Ble A T C Phe T T T A Ble T T T A Ble A T C T T T A C G Asm T C T Yr G A T G Bu A G C Bu C C C C C C C C C C C C C C C C C C C	Tyr TAT Lys A A A G Niet A TG TTC Loru C T Ual G TG Asm A A A A Lys Loru C T	Giy GG / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C / A C
ys al TG GT GT GG GG GG GG GG GG GG SG SG SG SG SG SG	Val GTGA A Leu CTA Ala GCCT Gly GAA GAA GAA A GAA A A GAA A A GCA A CTA A CTA A A GCA A CTA A A GCA A A GCA A A A CTA A A A CTA A A A CTA A A A CTA A A A	Trp T G G Val G T G Val Lyns A A A G A A A G A A A G A A C T T C T T C T T C T T C T A C T T C T A C T T C T A C T T C T A C T A C T A C T A C T G	Giy G G C G Lys A A A G Ser T C T C T C T C T C T C T C T C T C T C	Phe T T T T G Leu Gla A G A Val Lys A A G The A C A Phe T T T A C A Val C A G A A G A G A A	Tyr TAT T Phe Cys TGT C C Ala GCT Gly GGA GGA GGY GGA C TTT Lev TTA Ser	Asp GAT Tyr TAT The Glu GAGA AA Lys AAG GAG AA GAG C C Asp C Asp C Asp C Asp	Giy GGT Ser TCT Asp GAT Val GTC T Pro CCT Asn AAC Phe TTT Ile ATC	Gin CAA Val GTC , GNu GAA GTC C C C C C C C C C C C C C C C C C C	Gin GA A T A T Met Phe T T C Giu G A A G Giu He A T C C A ap G A T C A ap G A T C Val G A T C Val G A T C Val G A T C C C C C C C C C A T T C C A T C A T C C C I U C C A T C C C I U C C I U C C C I U C C C C C C	Pro CCT Thr ACA Trp TGG ATT Levu TACA Met Ser CCA Pro CCG Phe	Val GT G C Leu Asn A A C Lys A A A G C C C A A A C G C C C L S C C L S S C C C C C C C C C C	Leu CTG TTPhe Ang CGG Ang AGA Gin CAG Gin CAG Asp GAC A A Lys Asn AAC	Ala GCT C Ang AGG Leu TTA Tyr Tyr Tyr TAT Val GTC G C C C Leu C T C	lie ATC Pro CGA Asp CGA CGA Gly GGA Gly GGA Phe TTT Pro CCA Ser TCA	The A C A G Phe T T T See T C A G C Ala G C G C Ala G G G G C Ala C A G I U S G G C A L S S C A C A C A C A C A C A C A C A C A	Asp GAT C Gly GGT C Leu TTG GTG AT GCC Val GCC Asp GACC T Thr ACA	Pro C C T A A Thr Pro C C A Leu C T G Leu T T G T J T A C C C C C C C C T	Asp GACA Glu Val GTG Ser TCT C Val GTG A Ser A GC Phe TTT	Met A T G G G A G G A C C A C C Ang A G A A G A A T G Uss Met A T G Val G T G Pro	He ATC Phe TTTA Ile Thr ACC G Asn AATT Tyr Asp GAT Glu GAA G	Lys AAA G Niet ATG TTC Lou CTG T GTG Aam AACA Lys Lou C	The A C / A A A A A A A A A A A A A A A A A
al TG AAA GT AAA GG GG CA SG AA CC CC SG AA CC T CA T T T	Lev CTA Ala GCC GGA GGA GGA GGA GGA GGA GGA GGA GGA	Val GTG Be ATC GVal Lys AAA GCA GCA GCA GCA CTT The ACA Law CTT TA TY TY Met ATG	Lys A A A Ger T C T Leu C T C Ghu G A G Ser T C A Leu T T A Ghu G A A G A A G Lys A A A G	Giu GAA ATA C Val Lys AAG Thr ACA Phe TTT AGA Val GAA	Cys TGTGC Alla GCT Glu GGA Gly GGA Cly GGA Phe TTT Leu TTA	Tyr T A T Phe G A G Lys A T G Lys A A G C T G C Asp G T G C Asp A A T C Asp	Ser TCT Amp GAT GTC T Pro CCT Am AAC Phe TTT Ille ATC	Val GTC , Glu GA G CCTC Val GTC Val GTC Val TTGT Phe Cys TGTC C	Phe TTC Glu GAAA Hie ATC Thr ACC Asp GAC T Asp GAT Val GTG	Thr ACA Trp TGG Ille ATT Leu TTGA Met Ser TCT C C C C C C C C C C C C C C C C C	Asn A A C Lys A A G G C C A A G U U ys A A A C T C T T Phe T T C G Leu	Ang C G G Ang A G A Gin C A G G A C A G A Sap G A C A Sap A A C A Sap C A C A Sap C A C A Sap C A C A Sap C A C A C A C A C A C A C A C A C A C A	Ang AGG C Leu T T A T Tyr T A T Val G T C G Asn A A T C Leu C T C G	Pro C C T G A Asp Arg C G A A Gly G G A Phe T T T Pro C C A Ser T C A	Phe TTT Ser TCA GA GA CA Asp GG GI CAA A A G I Use ATA	Gly G G T C Leu T T G G T G A T Ile G C C Asp G A C T Thr A C A	Pro CCA Leu CTG Leu TTG Tyr TAC Pro CCC T	Val GTG Ser TCT C Val GTG A GTG A Ser TTT	Giy G G A G C C A C C A G A A G A A G A A G A A G A G T G G T G	Phe TTTT A lie Thr ACCG Asn AATT TCC Tyr GAT Glu GAA G Phe	Net ATG Phe TTC Leu CTG T Val GTG Aam AAC A Lys Leu C	Lys A A i G G (Gly Thr A C (A ng A ng A ng A ng A ng A ng A ng A ng
er GT GT GT GG GG AA A C C S A A C C T C T C T C T C T C T C T	Ala GCCT Gly GGA GAA GAA GAA GAA GAA GAA GAA A GAA A GAA A GAA A GAA A G CCT T C T C T C I Y C CA A C C C T	Be ATC Val Lys A G Arg Ala G C G Thr ACA Leu C TT TA A S T Vr Met A TG G In	Ser T C T Leu C T C Glu G A G Ser T C A Leu T T A G N G A A G Lys A A A G	Be A T A Val Lys A A G Thr A C A Phe T T T A T A G A Val G A A	Ala GCT Glu GAG Gly GGC Cly GGA Phe TTT Leu TTA	Glu G A G A Lys Met A T G Lys A A G Vel G T G G A T C Asp G A T C C Asn A A T	Asp GAT Val GTC T Pro CCT Assn AAC Phe TTT Ille ATC	Giu G A A G C C T C Val G T C Val `Leu T T G T T Phe Cys T G T C	Giu GAA G Ile ATC Thr ACC GAC T GAC T GAT GAT G	Trp TGG ILe ATT TTG A Met Ser TCT C CCA G Phe	Lys A A G G C C A A G U Lys A A A Leu C T C T Phe T T C G Leu	Ang A G A Gin C A G G A C A A Lys Asn A A C Phe T T T C	Leu TTA Tyr Tyr TAT Val GTC G Asn AAT C Leu CTC T G	Arg CGA A Gly GGA Phe TTT Pro CCA Ser TCA	Ser T C A G C Ala G A G A C Gly G G G G T Gln C A A A A G Lys Ibe A T A	Leu TTG GTG AT Ile Ala GCC Asp GAC T Thr ACA	Leu CTG Leu TTG Tyr TAC Pro CCC T	Ser TCT C Val GTG A Ser AGC Phe TTT	Pro C C A C C A G A A G A A G Lys Met A T G Val G T G Pro	Thr ACC G Aan AAT T C Tyr Asp GAT Ghu GAA G Phe	Phe TTC Levu CTG T Val GTG Asm Asm Lys Levu CTC	Thr AC Ang AG Lys Be AT C
er GGT GGG GGG AAA C C C C C C C C C C C C C C	Gly G G A G ha G A Ser A G C Lys A A G A A G A A G A A G C A A G C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C A C C C A C C C C A C	Lys A A A G A A G C A He A C A Leu C T T T G Leu C T T T A T T Yr Net A T G	Leu CTC Glu GAG Ser TCA Leu TTA Glu GAA GAA G AAA C	Lys A A G The A C A Phe T T T Arg A G A G T A A G G T A A G G Met G A A	Giu G A G Giy G G C Giy G G A Phe T T T Leu T T A Ser	Met A T G Lys A A G G T G G A G G A G C Asn C Asn C	Val GTC T Pro CCT Asn AAC TTT Ile ATC	Pro CCT C Val GTC G Val `Leu T T Phe Cys T GT C	lie ATC Thr ACC Asp GAC T Asp GAT Val GTG	lie ATT Leu TTG A Met Ser TCT C C A G C C A G Pho	Ala GCC AA Glu Lys AAA CTC T Phe TTC G Leu	Gin CAG GAC A SA Lys Asn AAC Phe TTT C	Tyr TAT GTC G Asn AAT C Leu CTC T G	Giy G G A Phe T T T Pro C C A Ser T C A	Anp GAT C Gly GGG T Gin CAA A G Lys Lie ATA	Val GTG AT Ile Ala GCC Asp GAC T Thr ACA	Leu TTG Tyr TAC Pro CCC T	Val GTG A Ser AGC Phe TTT	Ang A G A A G Lys Met A T G Val G T G	Aan A A T T C Tyr Aap G A T G A T G Iu G A A G	Low CTG T Val GTG Asm AAC A Lys Lew CTC	Arg A G Lys Ne A T C
ng GGG AAA hn CCT CC Sa AG CCT CC Sa AG CCT T C T C T C T C T	Glu GAA G Ser AGC Lys AAG ATT AGA ATT AGG ASer TCT	Ala GCA G Thr ACA Leu CTT TA A Tyr Net ATG	Giu G A G Ser T C A Leu T T A Giu G A A G Lys A A A G	The ACA Phe TTT Arg AGA Val GA Met GAA	Giy G G C Giy G G A Phe T T T Leu T T A Ser	Lys A A G Val G T G G A T C Asn A A T C	Pro CCT Asn AAC Phe TTT Ille ATC	Val GTC Be ATC G Val `Leu TG T Phe Cys TGT C	Thr ACC Asp GAC T Asp GAT Val GTG	Lev TTG A Met Ser TCT C Pro CCA G Phe	Lys AAA CTC T Phe TTC G Leu	Asp GAC A A Lys Asn AAC Phe TTT C	Val GTC G Asn A A T C Lew CTC T G	Phe TTT Pro CCA Ser TCA	Gly GGG T CAA A Lys Ile ATA	Ala GCC Asp GAC T Thr ACA	Tyr TAC Pro CCC T	Ser AGC Phe TTT	Met ATG Vaj GTG Pro	Anap GAT Glu GAA G	Vai GTG Asm AAC A Lys Leu CTC	Be AT The AC
hr CC Vs Vs CA CA CA CA T CA T CA T	Ser AGC Lys AAG A A T A G G G G G G	The ACA Leu CTT TG Leu CTT TA Tyr Met ATG	Ser T C A Leu T T A Glu G A G Lys A A A G	Phe TTT Arg AGA Val GTA A G Met Glu GAA	Gly G G A Phe T T T Leu T T A	Val GTG Asp GAT C Asn AAT C	Asn AAC Phe TTT Ile ATC	lie ATC G Val Leu TTG The Cys TGT C	Asp GAC T Asp GAT Val GTG	Ser TCT C Pro CCA G Phe	Levu CTC T Phe TTC G Levu	Asn AAC Phe TTT C	Asn AAT C Leu CTC T G	Pro CCA Ser TCA	Gin CAA A G Lys ile ATA	Asp GAC T	Pro C C C T Val	Phe T T T	Val GTG Pro	Glu G A A G	Asm AAC A Lys Leu CTC	
	Lys A A G A T T A G G A G G A G G A G G Ser T C T G	Leu CTT TG Leu CTT TA Tyr Met ATG	Leu TTA Glu GAA G Lys AAA G	Ang AGA GTA AG Met Glu GAA	Phe TTT Leu TTA	Asp GAT C Asn AAT C	Phe TTT Ile ATC	Leu TTG T Phe Cys TGT C	Asp GAT Val GTG	Pro CCA G	Phe TTC G Leu	Phe TTT C	Leu CTC T G	Ser TCA	lle A T A	Thr ACA	Val	Pha	Pro	Phe	Levu CTC	
	Be ATTA A Ang AGG A G Ser TCT G	Leu CTT TA Tyr Met ATG	Glu G A A G Lys A A A C	Val GTA A G Met Glu GAA	Leu T T A	Asa A A T C	, lle ∧ T C	Cys TGT C	Val GTG	Phe					G Val	G T Vel	GTC C Lev	ŤŤŤ	CCA	TTC		AT C Thr
	Ang AGG A Ser TCT G	Met ATG	Lys A A A G	Giu G A A	Ser				Met	TTT C	CCA	Arg AGA A Lys	Glu GAA T Asp	Val GTT TCA Ser	Thr ACA T Ile	Asın AAT GA Galu	Phe TTT	Leu TTA C Phe	Ang A G A A Lys	Lys A A A	Ser TCT T Phe	Vel G T d
	Ser TCT G	Gla			AGT CC Thr	Ang CGC	Lew CTC G	Glu GAA T Asp	Assp GAT TC Ser	Thr ACA GTG Val	Gln CAA G	Lys A A G	His CAC T	Ang CGA	Val GTG	Assp GAT	Phe TTC T	Leu CTT	Gin CAG	Leu CTG	Met ATG	lle A T 1
	Ale	CAG T His	Asn A A T	GAT Assp	Ser TCA T	Lys A A A	Glu GAA C Assp	Thr ACT AA Lys	Glu G A G A	Ser TCC T	His CAC T	Lys A A A C Thr	Ala GCT C	Leu CTG A	Ser тсс	Assp GAT	Leu CTG A	Gilu GAG	Leu CTC A	Vel GTG ACA	Ala G C C	Gin CA
Ĩ.	ATT C	NE ATC T	Phe TTT	ile ATT	Phe TTT	Ala GCT	Ghy G G C A	Тут ТАТ	Glu GAA	Thr ACC C Pro	Thr ACG C	Ser A G C	Ser AGT C	Vel GTT ACA Thr	Leu CTC T	Ser TCC	Phe TTC T	ile ATT GC Val	Met ATG C	Tyr TAT C	GNU GAA TCC	Leu CTO
c /	Ther ACT	His CAC	Рто ССТ А	Assp GAT	Vel GTC ACA Th r	Gin CAG	Gln CAG A Lys	Lys A A A	Leu CTG	Gin CAG	Glu G A G	Glu G A A G	A T T C	Asp GAT C	Ala GCA AGG Araz	Val GTT C Ala	Leu TTA CG	Pro CCC	Asen A A T	Lys A A G	Ala GCA	Рто СС/ 1
C /	The ACC	Тут ТАТ	Asp GAT	The ACT	Vei GTG	Leu CTA A G Met	Gin CAG GA Glu	Met ATG	Glu GAG A	Tyr TAT C	Leu CTT G	Assp GAC T	Met A T G	Val GTG	Val GTG T	Asn A A T	Glu G A A	Thr ACG C	Leu CTC	Ang AGA	Leu TTA G	Phe TTC AT
°	ile ATT	Ala G C T G Gly	Met ATG AT Asn	Ang A G A	Leu CTT	Glu G A G	Ang A G G A	Val GTC	Cys TGC T	Lys A A A	Lys A A A	Asp GAT	Vel GTT	Glu G A G A	ile A T C	Asn A A T	Gly G C C T	Met A T G G Val	Phe TTC T	ile ATT G Mat	₽ то ССС	Lys A A A
у с 1	Trp TGG CA Ser	Vel GTG	V∎I GTG C	Met ATG	Ne ATT	Pro C C A	Ser AGC TCT	Туғ ТАТ	Ala G C T	Leu (**T T	His CAC	Ang CGT	Asp GAC T	Pro CCA	Lys AAG C Gin	Tyr TAC C His	Trp TGG	Thr ACA C Pro	Giu G A G	Pro CCT	GNU GAG	Lys A A G G A Chu
ç c	Lew CTC G Arg	Pro C C T A	Giu G A A	Ang AGA G	Phe TTC	Ser AGC	Lys A A G	Lys AAG G Glu	Asn AAC	Lys A A G	Asp GAC Gly	Asn A A C G Ser	lle ATA T	Asp GAT	Pro CCT	Tyr TAC T	ile ATA G Val	Tyr TAC T	Thr A_CA CTG Lanu	Pro C C C	Phe TTT	Ghy G G A
T G	Giy G G A	Pro C C C	Ang AGA G	Asn A A C	Cys TGC	A T T	Giy G G C	Met ATG	Ang AGG	Phe TTT	Ala GCT	Leu CTC	Met A T G	Asn AAC T	Met ATG	Lys A A A	Lew CTT C	Ala GCT	Leu CTA C	He ATC CT	Ang AGA A	Vel GTC T
т с 6	Gin CAG A		Phe TTC	Ser TCC	Phe TTC	Lys AAA CG Gin	Pro CCT	Cys TGT	Lys A A A G	Giu G A A	Thr ACA	Gin CAG	lle ATC A	Pro CCC T	Leu CTG	Lys A A A	Leu TTA	Ser AGC	Leu TTA G Arr	Gly G G A C A G	Giy G G A	Leu CTT
	Gin CAA	Рто С С А	Glu G A A	Lys A A A	Рто ССС	Val GTT A	Val GTT A	Leu CTA	Lys A A G	Val GTT	Glu GAG T Val	Ser T C A C Pro	Ang AGG C	Anap GAT	Giy G G C A A	Thr ACC T	Val GTA A	Ser AGT C	Giy G G A	Ab <u><u><u></u></u><u><u></u><u></u><u></u><u></u><u></u><u></u><u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u><u></u></u></u></u>	•••• T G A	
		A ATT Trip CA Ser Leu C CA Gin C Gin C AA	A ATT GET Gly Trop Vel CA Set CA Set CA Set Gly Pro C CTC CCT Ang A Gly Pro C Gln Asn C A A C C A C A	A ATT GET ATG GT ATG GIY AM TOP Val Val CA GT GT CA CC Ser C CT C CCT GAA ANT A GUY Pro Glu GUA AAC TTC A AAC TTC GUA CCA GAA	A ATT GAT ATG AGA GIY AM TOP Val Val Met G TGG GTG GTG ATG CA CT GAA AGA Ser C CTC CCT GAA AGA AATG GIY Pro Arg Am GGA CCC AGA AAA G AAC TTC TCC Gin Pro Giu Lys CAA CCA ACA AAA	A ATT GAT ATG AGA CTT GT ATG AGA CTT Gly Am G TGG GTG GTG ATG ATT CA GTG GTG GTG ATG ATT CAN Pro Giu Ang Phe C CTC CCT GAA AGA TTC GIN Pro Ang Am Cys GGA CCC AGA AAA CTC Gin Pro Giu Lys Pro GAA CCA ACA AAA CCC	A ATT GRAT AND AND AND AND ATT GRAT AND	A ATT GET ATG ANE LEW UND ANG GA AT GA AGA CTT GAG ANG GY AM GIY AM GTOG GTG GTG ATG ATG ATT CCA AGG CCA GTG GTG ATG ATT CCA AGG TGG GTG GTG GTG ATG ATT CCA AGG CCA GAA AGA TC AGC AAG ANG TGGA CCC AGA AGA CTGC ATT GGC GIN CAG AAC TTC TCC TTC AAA CCT GIN AM Pho Ser Pho Lys Pro CAG AAC TTC TCC CT C AAA CCT GIN CAG AAA CTCC GTT GTT CAG AAC CCA GAA AAA CCC GTT GTT AAC AAA CCC GTT GTT	A ATT GET ATG AND LEW UND ATT VAI GIN ANN Met LIE Pro Ser Tyr GIY ANN Met LIE Pro Ser Tyr CA GG GTG GTG ATG ATT CCA AGC TAT CA Ser Pro GIN AND Phe Ser Lys Lys C GTC CCT GAA AGA TC AGC AAG AAG G AND GAA AGA TC AGC ATT GGC ATG GIN ANN PHE Ser Phe Lys Pro Cys C GIN ANN PHE Ser Phe Lys Pro Cys C AG AAC TTC TC TC AAA CCT TGT GIN ANN PHE Ser Phe Lys Pro Cys GIN CAG AAC TTC TC C GTT GTT CTA AC AAA CCC GTT GTT CTA AC AAA ACC GTT GTT CTA AC AAA AC C GTT GTT CTA AC AAA AC AC AC AAA AC AC AAA AC AAA AC AC	A ATT GET ATG AND LEW LEW CAN ARG Val Cya GAT GG AGA CTT GAG AGG GTC TG GUY ANN AN TO GUY ANN Met He Pro Ser Tyr Ala TGG GTG GTG GTG ATG ATT CCA AGC TAT GCT CA GU AND TC AGG AAG ATT CAGC AAG AAG AAG AAG AAG TTC AGC AAG AAG AAG AAG AAG TGC ATT GGC ATG AGG GU AND Pro AND AAG CYN HE GUY Met AND GA CCC AGA AAA CCG ATT GGT AAA A CCA GAA AAA CCG GT GT CTA AAG A CA CA GAA AAA CCG GT GT CTA AAG A ATT GGA CAG AAG AAG CYN LLW LYS A CA CCA GAA AAA CCG GT GT CTA AAG A ATT GGA CAG AAA CCG GT GT CTA AAG A ATT CTA AAG AC THO HE HE HE	A ATT GET ATG AND LEW LOW ATG Val Cya Lya GT ATG AGA CTT GAG AGG GTC TCA AAA GUY AAN TTP GUY AAN Met Be Pro Ser Tyr Ala Leu CA GG GTG GTG ATG ATT CCA AGC TAT GCT CTT CA Ser Pro Glu Ang Phe Ser Lya Lya Aan Lya GU AAG AAC AAG TTC AGC AAG AAG AAC AAG GGU ANG GA AGA TTC AGC AATG AGG TTT GAA ACC AGA AAA CCC GTT GT AAA GAA GGN Pro GLU Lya Pro Val Val Leu Lya Val CAA CCA GAA AAC CTG GTT CTA AAG GTT TMP Be be	A ATT GET ATG AND LYS LEW CHU ANG YAL CYS LYS LYS GU AA AAA GIY AM MEL IIE PRO SET TYT AIA LEW HIS TGG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CA CCA CC AGA AGA TTC AGC AAG AAG AAC AAG GAC GIV GIV AM CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR GIV AM CONTRACTOR C	A ATT GRAT AND AND LYS PRO LYS LYS LYS LYS ASD GT ATG AGA CTT GAG AGG GTC TGC AAA AAA GAT GUY AM GUY AM GT GG GTG GTG ATG ATT CCA AGG TAT GCT CTT CAC CGT CA GT GG GTG GTG GTG ATG ATT CCA AGG TAT GCT CTT CAC CGT CA GT GG GTG GTG GTG ATG ATT CCA AGG AAG AAG AAG AAG AAG GT GG AAA AGA TTC AGC AAG AAG AAG AAG AAG AAG AAG GU GUY Ser GUA AND PRO AND CYS LYS AND AND CTT GCT CTC GUA AND PRO AND CYS LYS AND AND CTT GCT CTC GUA AND PRO AND CYS LYS AND AND AND CTT GCT CTC GUA AND PRO AND CYS LYS AND AND AND CTT GCT CTC GUA AND PRO AND CYS LYS AND CYS LYS GUU THT GUA GUA AND PRO AND CYS LYS AND CYS LYS GUU THT GUA GUA AND PRO SER PRO CYS LYS GUA AGA AAC CGG GUA CCC AGA AAA CCC TGC ATT GGC ATG AGG TTT GCT CTC GUA AND PRO GUA LYS PRO VAI CTT CTA AAG GTT GAG TCA ACA CCA GAA ACC CC GTT GTT CTA AAG GTT GAG TCA ACA AC AAC CC GTT GTT CTA AAG GTT GAG TCA ACA AC AAC CA GTT GTT CTA AAG GTT GAG TCA ACA AC	A ATT GRAT AND AND LYS PRO VAL CAN AND CAL LYS LYS LYS AND VAL GAT GAT ATT GAG AGG GTC TGC AAA AAA GAT GTT GUY AND CAL AND CAL CAG GTG GTG AGA CAC ATT CCA AGG GTC TAT GCT CTT CAC CGT GAC CA CAG ATT CCA AGG TAT GCT CTT CAC CGT GAC TCT CAC CGT GAC TCT CAC CGT GAC AGG AAG AAG AAG AAG AAG AAG AAG AAG	A ATT GRAT ARE LEW GAU ARE Val Crys Lys Lys Asp Val Glu Grat Are AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG Gly Asn Trop Val Val Met He Pro Ser Tyr Ala Lew His Are Asp Pro CGA GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA Ser Pro Glu Are Pro Ser Lys Lys Asn Lys Asp Asn He Asp C GC CCT GAA AGA TTC AGC AAG AAG AAG AAC AAG GAC ATA GAT G A AG A AGA TTC AGC AAG AAG AAG AAC AAG GAC ATA GAT G A AG A AGA TTC AGC AAG AAG AAG AAC AAG GAT CTC ATG AAC G Glu Gly Ser Gla Are	A ATT GRAT ARG ANG CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC GI AAT GU AAN GU AAN GTO GAG ATG ATT CCA AGG GTC TGT CAC CGT GAC CCA AAG TTO GAA AGA AGA TT CCA AGC AAG GI CCC AGA AGA TT CCA AGC AGG AAG AAG AAG AAG AAG GU GU GU GU GU GU GU GU GU GU GU GU GU GU GU	A ATT GRAT ARG ANG CTU GAG AGG GTC Yal Cya Lya Lya AAA GAT GTT GAG ATC AAT GIN AM GIN AM GI	A ATT GRAT AND AND LYS FOR CTT GAG AGG GTC TYS LYS LYS ASP VAI GRU HE ANN GRU GY ANN ATA GAT GTT GAG ATC AAT GGG GTG GTG ATG AGT CTT GAG AGA GAT GAT GAT GAT GTT GAG ATC AAT GGG TG GTG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG CAG GTG GTG GTG GAT ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG GTG GTG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG GTG GTG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG GTG GTG GTG GTG ATG ATG AAG AAC AAG GAC AAC AAA AAA GAT CTT CCA CGT GAA AGA TTC AGC AAG AAG AAC AAG GAC AAC AAA GAT CTT TGG GRU GRU GRU GRU GRU GRU GRU GRU GRU G	A ATT GRET AND AND LEV UND ANG VAL CYS LyS AS AND VAL GIU HE AS CAT GIV MEL G AT AT AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGG ATG GIV ANN TTO VAL VAL HE AND TTO TO	A ATT GRUT AND ANG LEU UND ANG GTC TGC YS Lys Lys AAA GAT GTT GAG ATC AAT GGG ATG TT GAG ATG TT GAG ATG GTT GAG AGA CTA GAG ATG TT GAG ATG ATT GAG ATG TT GAG ATG AT	A ATT GET ATG AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGG ATG TTC ATT GT AT GT ATG AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGG ATG TTC ATT GT AT GT ATG GT GTG ATG ATT CCA AGC TAT GCT TT CAC CGT GAC CCA AAG TAC TGG ACA GAG CCT C GIN His Pro C TG GTG GTG GTG ATG ATT CCA AGC TAT GCT TT CAC CGT GAC CCA AAG TAC TGG ACA GAG CCT C TC C GIN His Pro C TG GAA AGG TTC AGC AAG AAC AAG GAC AAC ATA GAC CCT TT CAC GT GAC ACA AGC TAC ATA TAC ACA CCC GIN His Pro C TG GAA AGG TTC AGC AAG AAC AAG GAC AAC ATA TAC ACA CCT C GIN GIV Ser T GGA AGG AAC TGC ATT GGC ATG GGC ATG AGG TTT GCT CTC ATG AAA CTT GC CT C GIN AM T GGA AGG AAC TGC ATT GGC ATG GGC ATG GGC AAC ATG AAC ATG AAC ATG AAA CTT GC CT TAC GIN AM T GGA AGG AAC TGC TT GT AAG GGC TT GCT CTC ATG AAC ATG AAA CTT GCC CT GIN AM T GGA AAG AAC TGC TT GT AAG GGC AAC ACA ATG AAC ATG AAC ATG AAA CTT GCC CT GIN AM T GGA AAC TTC TCC TTC AAG GGC ATG GGG ATC CCC GTG AAA TTA AGC TT GCA GIN CIV SUB COL CTC CTG AAA TTA AGC CTT GCT TGC ATG AAAC ATG AAA CTT GCC CT GIN AM T GGA AAC TTC TCC TGC ATG GGC ATG GGA ACA ACA ACA AGG GAT GGC AAA TTA AGC TTA GGA ATT GGA AAC TTC TCC TGC AAG GGA ACA ACA CAG ATC ATG AAA CTT GCC CT GIN AM T GGA AAC TTC TCC TCC TGT AAA GG TTT GTT GAG AAC ACA CAG ATC AGG GAT GGC AAA TTA AGC TTA GGA ATT GGA AAA CCC GGN GIN TT GTT GAA AGG GAT GGC ATG GGA TGGC GTT AAC TTA GGA AGC GGA TGGC GAT GGC ATG GGA AGC GTT GAA CCA AGG GAT GGC ATC GGA AGC GTT GGA ACC CT GGA AGC GTT GAA CCA AGG GAT GGC ATC GGA AGC GTT GAA CCA AGG GAT GGC ATC GGA AGC CT GAA AGT GGA ACC CT GAA AGT GGA ACC CT GGA AGC CT GAA CCA AGG GAT GGC GAT GGC GTT GGA AGC CT GAA CCA AGG GAT GGC ATC GGA AGC CT GAA AGA CCT GGT GTT CTA AAG GTT GAG CCA AGG GAT GGC GAT GGC ACC GTA AGT GGA ACC CT GAA AGA CCT GGA AAA CCA AAG CCC GTT AAC CCA AGG GGAT GGC GAT GGC GTA AGG GAT GGC GTA AGG GAT GGC GTA AGG GAT GGC GTA AGG GAT GGC ACC GTA AGG GGT AAC CCA AGG GAT GGC GAT GGC GTA AGG GAT GGC GTA AGG GGA GCC GTA ACC CA AGG GTT GAA CCA AGG GAT GGC GAT GGC GAT GGC ACC GTA AGG GAT GGC ACC AGG G	A ATT GET ANTO ANT LEV LAV ANT CAG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGY MEL PHE BE PRO GY ANT AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGY MEL PHE BE PRO GY ANT AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGY MEL PHE BE PRO GY ANT AGA CTT GAG AGG GTC TGC AAA AAA GAT GTT GAG ATC AAT GGY ATG TTC ATT CCC GY ANT ANT CAA TC CA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG ACA GAG CCT GAG GG GTG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG ACA GAG CCT GAG GG GTG GTG GTG ATG ATT CCA AGC TAT GCT CTT CAC CGT GAC CCA AAG TAC TGG ACA GAG CCT GAG GG GTC CCT GAA AGA TTC AGC AAG AAG AAG AAC AAG GGC AAC ATA GAT CCT TAC ACC TTC TC GAN GA AGA TTC AGC AAG AAG AAC AAG GGC AAC ATA GAT CTT TC TTC ATT AC ACC CT TT GAN GA AGA AGA AGA ATT GGC ATG AGG TTT GCT CTC ATG AAA CTT GCT CTA AG CTT GIV GIV Ser VII LEV LYS LYS GIN THE GIT CTC ATG AAA CTT GCT CTA AG CTT AC GGN AGA AGC TTC TGC ATG AGG TTT GCT CTC ATG AAA CTT GCT CTA AG CT GGN GG GA CCC AGG AAC CTT CTC TGT AAA GAA ACA CAG ATC ATG AAA CTT GAAA CTT GCT CTA AGG GAT GGG GGN GGN CCC AGG AAC CTC CTC TGT AAA GAA ACA ACA CAG ATC ATG AAA CTT GCT CTA AGG GAT GGG ACC GTA AGC GGG GGN GGN CCC AGA AAA CCT TGC TGT AAA GAA ACA CAG ATC ATG AAA CTT GCT TTC LYS LEV LYS CIN THE LYS LEV ANT AGC TTG GGA GGN CCC AGA AAA CCT TCC TTGT AAA GAA ACA CAG GAT CAG ATC ATG AAA CTT GGC ACC GTA AGC GGG GGN CCC AGA AAA CCC C CT TGT AAA GAA ACA CAG GAT CAG GAT CCC GTA AGC GAA ACC CAG ATC AGG GAT GGC GCC GTA AGC GGA AAA CCC GGA AAA ACC CAG GAT GAG GGC GCC GTA AGC GGA AAA CCC GTA AAA CCC TTG TGT AAAG GTT GAG GGC ACC GGA AAA CCC GTA AAA CCC GTA AAA CCC TTGT AAAG GTT GAG TGA AGG GAT GGC GCC GTA AAA CCC GTA AAA CCC GTA AAA CCC TTGT AAAG GTT GAG GTT GAG GGC ACC GTA AAA CCC GTA AAAA CCC GTA AAA CCC GTA AAA CCC GTA AAA CCC GTA AAAA CCC GTA AAAA CCC GTA AAA CCC GTA AAA CCC GTA AAAA CCC GTA AAA CCC GTA AAAA CCC GTA A

FIG. 3. Nucleotide and deduced amino acid sequences of NF25 (upper line) and comparison with rat $P-450_{PCN}$ sequence (24) (lower line). Only the differences from NF25 are indicated in the rat $P-450_{PCN}$ sequence. Initiation and termination codons are boxed. The cysteine-containing peptide thought to bind heme (25) is underlined. Dashes indicate the three-nucleotide gap introduced to increase the base and amino acid match.

enough to code for the entire protein (Fig. 1), and NF95, which gave a very strong antibody response and produced a fusion protein recognized by both anti- β -galactosidase and monoclonal and polyclonal anti-P-450_{NF} antibodies (Figs. 1 and 2). The increase in molecular weight of the fusion protein (compare β -galactosidase) correlated well when compared to

the length of the insert (Fig. 2 A and B). Antibodies were raised against the fusion protein from NF95 by transferring the fusion protein to nitrocellulose paper, solubilizing the nitrocellulose in dimethylsulfoxide, and injecting this material into rabbits (23). These antibodies recognized P-450_{NF} (Fig. 2C). NF25 contained two internal *Eco*RI sites, which were used for subcloning into M13, while NF95 did not contain any internal *Eco*RI sites. Three probes were prepared by nick-translation of restriction fragments of insert DNA (probes 1, 2, and 3; see Fig. 1) obtained after subcloning in M13. These were used to confirm the alignment of the subinserts of NF25 by Southern blotting of λ gt11 DNA that had been cut with *Eco*RI and electrophoresed through a 1% agarose gel. Probe 1 recognized the same subinsert as probe 2 (data not shown). All other λ gt11 clones selected with antibodies were recognized by probe 1.

Sequence Analysis of NF95 and NF25. NF25 and NF95, in phage M13, were sequenced by the dideoxy termination method (16) using the strategy indicated in Fig. 1. Each region represented by an arrow was sequenced between three and eight times. The sequence is shown in Fig. 3 along with the deduced amino acid sequence. The NF25 insert is 2.2 kilobases (kb) long and the sequenced portion represents 1606 nucleotides, including 68 nucleotides in the 5' untranslated region, 29 nucleotides in the 3' noncoding region, and an open reading frame of 1509 nucleotides, which corresponds to a protein of 503 residues. The termination codon is TGA, the most common termination codon published in known human and mouse sequences (26). The position -3upstream from the initiation codon is a purine, as for all known eukaryotic sequences (27).

The deduced amino acid sequence was compared to those of two P-450s, rat P-450_{PCN} (24) (PCN = pregnenolone-16 α carbonitrile) and human P-450p (28); the latter is known to be highly similar or identical to P-450_{NF} by comparison of the procedures of purification (8, 28) and by immunochemical cross-reactivity (7, 28). The first 21 amino acids for NF25 (deduced sequence) and P-450p (28) are identical (Fig. 4). P-450_{PCN} (24) and NF25 share 77% nucleotide similarity, and only a one-codon gap has to be introduced (P-450_{PCN} has 504 amino acids instead of 503 for NF25). The NF25 nucleotide sequence shares 53%, 45%, 45%, and 47% similarity, respectively, with human pH P-450(1) (31), P_1 -450 (26), P_3 -450 (32), and P-450_{MP} cDNA coding sequences (D.R.U., R.S.L., and F.P.G., unpublished results) (5), respectively, as analyzed with these programs, indicating that $P-450_{NF}$ belongs to a family different from the phenobarbital- and 3-methylcholanthrene-inducible families-the rat and rabbit PCN-inducible P-450 gene family (25). A more extensive comparison of the sequences of the N-terminal region and C-terminal cysteine-containing peptide, a highly conserved region, is shown in Fig. 4. All P-450s of the PCN-inducible family

(human P-450_{NF}, rat P-450p, rat P-450_{PCN}, and rabbit P-450 LM3c) have similar N-terminal sequences, but they are not similar to human P_1 -450 and P-450_{MP}. The heme ligand cysteine-containing peptide is more conserved within the gene families than within species: the nucleotide sequence similarity in this region can reach 96% within a family $(P-450_{NF} \text{ versus } P-450_{PCN}, \text{ Fig. 4})$. The cDNA-deduced amino acid composition was compared with the one obtained with purified P-450_{NF} (7); the difference index of 7.5 shows that they fit well (33). The amino acids that show the greatest deviation (glycine and alanine) are those that are known to generally yield discrepancies between amino acid analysis and cDNA-predicted sequences (34). Comparison of rat P-450_{PCN} and P-450_{NF} compositions gave a difference index of 3.7, confirming the strong relationship between the two proteins. The deduced molecular weight of the P-450_{NF} protein is slightly higher than the one estimated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis, as is the case for other hydrophobic P-450s, which may bind more sodium dodecyl sulfate than soluble proteins (34). A complete restriction map was obtained by computer analysis of the sequence and fitted perfectly with the experimental map (BamHI, Cla I, EcoRI, HindIII, Kpn I, Pst I, Pvu I, Sac I, and Sal I). NF95 was also sequenced and, in the overlapping region with NF25, no differences were seen, indicating that both clones were probably derived from the same mRNA.

mRNA and Genomic DNA Analysis. Genomic DNA was cut with three different restriction enzymes, and the fragments were analyzed by Southern blotting (Fig. 5). Several bands were obtained with the three different cuts and the two nonoverlapping probes used (probes 2 and 3), even when relatively high-stringency hybridization conditions were used, indicating that $P-450_{NF}$ very likely belongs to a multigene family. The size (50–60 kb with the two probes) and the numbers of bands practically exclude the existence of a single gene. Liver samples HL 32 and HL 34 showed some small differences in Southern blotting patterns (Fig. 5A); comparison with three other liver DNAs has not shown any specific pattern for poor or extensive metabolizers to date.

The multiplicity of P-450_{NF} genes was confirmed by Northern blot analysis (Fig. 5B), which shows that probe 1 recognizes three or four bands, indicating that there are several transcripts, the sizes of which (\approx 2–4 kb) are all compatible with the size of the protein and the lengths of other P-450 mRNAs (25). Clearly the major size class is about 2.0 kb, although the number of distinct transcripts within this

Human P-450_{NF} Human P-450p Rat P-450 PCN Rat P-450p Rat P-450 LM3c Human P₁-450 Human P-450_{MP} B Human P-450_{NF} Rat P-450 PCN Human P₁-450 Human P₃-450 Human pH P-450(1)

A



FIG. 4. Comparison of deduced amino acid sequences in two regions: N-terminal region (A) and the heme-binding cysteine-containing peptide (B) (25). Important similarities between the sequences are boxed. Amino acid sequences were obtained from the following sources in parentheses: rat P-450_{PCN} (24), rat P-450p (29), rabbit P-450 LM3c (30), human P-450p (28), pH P-450(1) (31), human P₁-450 (26), human P₃-450 (32), human P-450_{MP} (D.R.U., R.S.L., and F.P.G., unpublished results).



FIG. 5. DNA and RNA analysis by blotting. Microsomes prepared from human liver samples HL 32 (lanes a) and HL 34 (lanes b) oxidized nifedipine at rates of 0.88 and 0.65 nmol of product formed per min/nmol of P-450, respectively. (A and B) Southern DNA blots. DNA was loaded (20 μ g) and hybridized with probe 3 (A) and probe 2 (B). (C) RNA blots. RNA was loaded (15 μ g) and hybridized with probe 1. Probes 1 and 2 gave similar results. Restriction enzymes used are indicated at the tops of the lanes. Ribosomal RNA markers are indicated at 18S and 28S. Sizes are shown in kb.

zone is unknown, and we do not know conclusively which mRNA size-class codes for the protein. Our previous work suggested that catalytic activity is related to the amount of immunochemically detectable $P-450_{NF}$ (7), and these results have been confirmed. Analysis of mRNA levels in nine liver samples (with probe 2) showed a 20-fold variation [after normalization of mRNA hybridization with an oligomer (50-mer) complementary to the 5' end of human serum albumin mRNA (35)]. However, the mRNA levels could not be correlated to either levels of nifedipine oxidase activity or immunochemically determined $P-450_{NF}$ in microsomes prepared from this set of samples.

CONCLUSIONS

While cDNA clones have been previously obtained for human P-450s (26, 31, 32), in no case has a sequence been obtained that is related to a P-450 protein that has been isolated and characterized in terms of physical properties and catalytic activity. We obtained a sequence for a cDNA related to a major human liver P-450 that appears to be involved in an oxidation polymorphism where a significant population is affected (6). The protein has been suggested to be inducible in humans (28), and other work in our laboratory indicates that P-450_{NF} is involved in a number of reactions, including 6B-hydroxylation of testosterone and androstenedione, estradiol 2- and 4-hydroxylation, aldrin epoxidation (7), benzphetamine N-demethylation (7, 8), quinidine 3- and N-oxidation (36), and the oxidation of nifedipine and 17 other dihydropyridine analogs to pyridine compounds (37). Several related sequences are found in this gene family, and we cannot state with certainty that the coding sequence deduced

from the cDNA is identical to that of P-450_{NF}. However, the cloning and sequence determination should provide a basis for further studies on the basis of regulation of activities related to this P-450 in humans.

This research was supported in part by U.S. Public Health Service Grants CA 30907 and ES 00267. D.R.U. is the recipient of U.S. Public Health Service Individual Research Award ES 05340. F.P.G. is a Burroughs Wellcome Scholar in Toxicology (1983-1988).

- Maghoub, A., Dring, L. G., Idle, J. R., Lancaster, R. & Smith, R. L. 1. (1977) Lancet ii, 584-586
- Küpfer, A. & Preisig, R. (1983) Semin. Liver Dis. 3, 341-354. 2.
- Wedlund, P. J., Aslanian, M. D., McAllister, C. B., Wilkinson, G. R. & Branch, R. A. (1984) Clin. Pharmacol. Ther. 36, 773-780. 3.
- Distlerath, L. M., Reilly, P. E. B., Martin, M. V., Davis, G. G., Wilkinson, G. R. & Guengerich, F. P. (1985) J. Biol. Chem. 260, 4. 9057-9067
- 5. Shimada, T., Misono, K. S. & Guengerich, F. P. (1986) J. Biol. Chem. 261, 909-921.
- 6. Kleinbloesem, C. H., van Brummelen, P., Faber, H., Danhof, M., Vermeulen, N. P. E. & Breimer, D. D. (1984) Biochem. Pharmacol. 33, 3721-3724.
- 7. Guengerich, F. P., Martin, M. V., Beaune, P. H., Kremers, P., Wolff, T. & Waxman, D. J. (1986) J. Biol. Chem. 261, 5051-5060.
- Wang, P. P., Beaune, P., Kaminsky, L. S., Dannan, G. A., Kadlubar, 8. F. F., Lar 5375-5383. Larrey, D. & Guengerich, F. P. (1983) Biochemistry 22,
- 9. Beaune, P., Kremers, P., Letawe-Goujon, F. & Gielen, J. E. (1985) Biochem. Pharmacol. 34, 3547-3552
- Young, R. A. & Davis, R. W. (1983) Proc. Natl. Acad. Sci. USA 80, 10. 1194-1198.
- Guengerich, F. P., Wang, P. & Davidson, N. K. (1982) Biochemistry 21, 11. 1698-1706.
- Guengerich, F. P., Dannan, G. A., Wright, S. T., Martin, M. V. & 12. Kaminsky, L. S. (1982) Biochemistry 21, 6019-6030.
- Yamamoto, K. R., Alberts, B. M., Benzinger, L., Lawhorne, L. & Treiber, G. (1970) Virology 40, 734–744. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) in Molecular Clon-13.
- 14. ing: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Messing, J. (1983) Methods Enzymol. 101, 20-79. 15.
- Sanger, F. S., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. 16. Sci. USA 74, 5463-5467
- 17. Lloyd, R. S., Recinos, A., III & Wright, S. T. (1986) Biotechniques 4, 8-10.
- 18. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979) Biochemistry 18, 5296-5300.
- Wahl, G. M., Stern, M. & Stark, G. R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687. 19. 20.
- Southern, E. M. (1975) J. Mol. Biol. 98, 503-517. Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977) Proc. Natl. Acad. Sci. 21.
- USA 74, 5350-5354
- Benson, S. A. (1984) Biotechniques 2, 66-69. 22.
- Knudsen, K. A. (1985) Anal. Biochem. 147, 285-288. 23.
- Gonzalez, J., Nebert, D. W., Hardwick, J. P. & Kasper, C. B. (1985) J. 24. Biol. Chem. 260, 7435-7441.
- 25. Adesnik, M. & Atchison, M. (1986) CRC Crit. Rev. Biochem. 19, 247-305.
- 26. Jaiswal, A. K., Gonzalez, F. J. & Nebert, D. W. (1985) Science 228, 80-83.
- 27. Kozak, M. (1984) Nature (London) 308, 241-246.
- 28. Watkins, P. B., Wrighton, S. A., Maurel, P., Schuetz, E. G., Mendez-Picon, G., Parker, G. A. & Guzelian, P. S. (1985) Proc. Natl. Acad. Sci. USA 82, 6310-6314.
- 29. Wrighton, S. A., Maurel, P., Schuetz, E. G., Watkins, P. B., Young, B. & Guzelian, P. S. (1985) Biochemistry 24, 2171-2178.
- 30. Koop, D. R., Morgan, E. T., Tarr, G. E. & Coon, M. J. (1982) J. Biol.
- Chem. 257, 8472-8480. Phillips, I. R., Shepard, E. A., Ashworth, A. & Rabin, B. R. (1985) Proc. Natl. Acad. Sci. USA 82, 983-987. 31.
- Quattrochi, L. C., Okino, S. T., Pendurthi, U. R. & Tukey, R. H. 32. (1985) DNA 4, 395-400.
- Metzger, H., Shapiro, M. B., Mosimann, J. E. & Vintar, J. E. (1968) Nature (London) 219, 1166-1169. 33.
- Black, S. D. & Coon, M. J. (1986) in Cytochrome P-450: Structure, 34. Mechanism, and Biochemistry, ed. Ortiz de Montellano, P. R. (Plenum, New York), pp. 161–216. Dugaiczyk, A., Law, S. W. & Dennison, O. E. (1982) Proc. Natl. Acad.
- 35. Sci. USA 79, 71-75.
- Guengerich, F. P., Müller-Enoch, D. & Blair, I. A. (1986) Mol. 36. Pharmacol., in pres
- 37. Böcker, R. H. & Guengerich, F. P. (1986) J. Med. Chem., in press.