Cloning of a potential cytochrome P450 from the Archaeon *Sulfolobus solfataricus*

Rhonda L. Wright, Kathryn Harris, Barbara Solow, Robert H. White, Peter J. Kennelly*

Department of Biochemistry and Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0308, USA

Received 1 March 1996

Abstract A gene, *CYP119*, for a potential cytochrome P450 has been isolated and sequenced from the extreme acidothermophilic archaeon *Sulfolobus solfataricus*. The gene predicts a polypeptide of 368 amino acids containing the consensus heme-binding sequence Phe-Gly-Xaa-Gly-Xaa-His-Xaa-Cys-Xaa-Gly-Xaa$_3$-Ala-Arg-Xaa-Glu. It most closely resembles the cytochrome P450s found in the bacterium *Bacillus subtilis*, with which it shares 129 identical amino acid residues (35%). This first sequence of a potential archaeal cytochrome P450 represents an important step in tracing the complex evolutionary history of this biologically important enzyme family.

Key words: Cytochrome P450; Evolution; Archaea; Archaebacteria; CYP119

1. Introduction

The cytochrome P450s comprise a large family of heme-containing monooxygenases ubiquitously distributed among aerobic organisms (reviewed in [1–3]). These enzymes participate in the oxidation of a wide range of endogenous and exogenous organic compounds. While the oxidation of foreign substances by cytochrome P450 can serve as an important step in their neutralization, in man and other organisms the oxidation of foreign compounds often results in the creation of highly toxic and/or carcinogenic metabolites — hence these enzymes have been the subject of intense scrutiny. The genes for several hundred cytochrome P450s have now been sequenced. They suggest a complex evolutionary history that crosses phylogenetic divisions, an evolutionary history that is only partially understood [2,3]. One of the barriers to unravelling the chain of gene duplication, gene transfer, and other events is the lack of any data concerning the cytochrome P450s from the third phylogenetic domain, the Archaea. Thus, when we accidentally encountered the partial clone of a gene predicting a potential cytochrome P450 from the extreme acidothermophilic archaeon *Sulfolobus solfataricus*, we seized upon the opportunity to begin filling a major gap in the evolutionary history of this important enzyme superfamily.

2. Materials and methods

2.1. Materials

Purchased materials included custom oligonucleotides (DNAgency, Malvern, PA); radiochemicals (DuPont/New England Nuclear, Boston, MA); nitrocellulose and nylon membranes (Micron Separations, Inc., Westboro, MA); restriction endonucleases, T4 DNA ligase, pGEM-3Z, Prime-a-Gene DNA labelling kit, and Erase-A-Base kit (Promega, Madison, WI); Sequenase version 2.0 DNA sequencing kit, pUC vectors, and shrimp alkaline phosphatase (US Biochemicals, Cleveland, OH); GeneAmp PCR core reagent kit (Perkin Elmer, Branchburg, NJ); *E. coli* DH5α cells (Life Technologies, Inc., Gaithersburg, MD); Long Ranger gel solution (J.T. Baker, Phillipsburg, NJ); and Deep Vent DNA polymerase (New England Biolabs, Beverly, MA). *E. coli* 7118 was the generous gift of Dr. Dennis Dean, Virginia Tech. All other reagents were purchased from Sigma Chemical Co. (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA).

2.2. Cloning procedures

S. solfataricus (ATCC 35091) was grown, genomic DNA isolated therefrom, and genomic DNA libraries constructed as described by Leng et al. [4]. Restriction enzyme digests, Southern blots, and the identification and isolation of DNA clones were performed using standard procedures [5]. Oligonucleotide probes were radiolabelled with ^{32}P using the Prime-A-Gene kit following the manufacturer's protocols. DNA sequencing was performed by the dideoxy method of Sanger et al. [6] using a Sequenase version 2.0 kit following the manufacturer's protocol.

2.3. PCR techniques

The following oligonucleotide primers were used for PCR: primer A, 5′-d(GGGAATTCCRTDTAYGGWVHHCARTGG)-3′; primer B, 5′-d(GGGATCCCSWHGCDATRTTRAANGG)-3′; primer C, 5′-d(CCCGGATCCATTTATCTGAATTCCCAAGG)-3′; and primer D (SP6 promoter primer), 5′-d(GATTTAGGTACACTATAG)-3′. 'Touchdown' PCR was performed by the procedure of Roux [7] as modified by Leng et al. [4]. The synthesis of the 920 bp primer was performed according to the standard PCR protocol of Leng et al. [4], with the exception that the quantity of DNA used was reduced from 1000 to 500 ng and the quantity of each primer was increased from 100 to 500 pmol.

3. Results and discussion

The gene for the potential archaeal cytochrome P450 was encountered during an attempt to isolate clones for a thymidylate synthase gene from the extreme acidophilic archaeon *S. solfataricus*. Briefly, touchdown PCR amplification of genomic DNA from *S. solfataricus* was performed using degenerate oligonucleotide primers A and B, which were modelled after the sequence of thymidylate synthase from *E. coli*. The 300 bp product that resulted was used to probe a genomic DNA library contained in plasmid pGEM-3Z that had been constructed using the restriction enzyme *Hin*dIII. Several potential clones were isolated and their ends sequenced. A computer search of DNA databases using the BLASTX program revealed that the 5′-end of one clone, ≈1.5 kb in length, predicted a gene product possessing significant sequence homology to the members of the cytochrome P450 family. It also was apparent that it encoded only a portion of the presumed structural gene, specifically the first 294 amino acid residues of the protein. Using PCR primers C and D, a 920 bp probe encompassing this partial clone was synthesized and utilized to probe a second library of *S. solfataricus* geno-

*Corresponding author. Fax: (1) (540) 231-9070.
E-mail: pjkennel@vt.edu

SSDI S0014-5793(96)00322-5

```
TAA AAC CAT TTA AAT ATC GTA TTT TTT CTA TAT CTA TAC TTT AAG TTA AGA TAT ATC TAT TTA TTT

ATC TGA ATT CCC AAG GTA TTA CCT ATG TAT GAC TGG TTT AGT GAG ATG AGA AAG AAA GAC CCA GTG
                ** *** *** **  M   Y   D   W   F   S   E   M   R   K10 K   D   P   V

TAT TAT GAC GGA AAC ATA TGG CAG GTG TTT TCC TAT AGG TAC ACA AAG GAG GTT TTA AAC AAC TTT
 Y   Y   D   G   N   I20 W   Q   V   F   S   Y   R   Y   T   K30 E   V   L   N   N   F

TCG AAA TTC TCC TCA GAT CTT ACA GGT TAC CAC GAA AGG CTT GAG GAC CTA AGG AAC GGT AAA ATA
 S   K   F   S   S40 D   L   T   G   Y   H   E   R   L50 E   D   L   R   N   G   K   I

AGG TTC GAC ATC CCC ACT AGG TAT ACC ATG CTG ACC TCA GAT CCC CCT CTC CAT GAT GAG TTA AGA
 R   F60 D   I   P   T   R   Y   T   M   L   T70 S   D   P   P   L   H   D   E   L   R80

TCA ATG TCA GCA GAT ATA TTC TCG CCT CAA AAG CTA CAG ACA CTT GAG ACA TTT ATT AGG GAG ACC
 S   M   S   A   D   I   F   S   P   Q90 K   L   Q   T   L   E   T   F   I   R100 E   T

ACC AGA AGC CTA TTA GAC TCA ATT GAC CCT AGG GAA GAC GAC ATA GTG AAG AAG TTA GCT GTT CCA
 T   R   S   L   L   D   S   I110 D   P   R   E   D   D   I   V   K   K120 L   A   V   P

CTA CCA ATA ATA GTT ATC TCA AAA ATA TTG GGT CTC CCA ATT GAA GAT AAG GAG AAG TTC AAA GAG
 L   P   I   I   V   I130 S   K   I   L   G   L   P   I   E   D140 K   E   K   F   K   E

TGG TCA GAC TTA GTC GCA TTC AGG TTG GGT AAG CCT GGA GAA ATA TTT GAG CTA GGT AAG AAG TAC
 W   S   D   L150 V   A   F   R   L   G   K   P   G   E160 I   F   E   L   G   K   K   Y

CTT GAG TTA ATA GGT TAT GTG AAG GAT CAT CTA AAT TCA GGG ACC GAA GTG GTC AGC AGA GTT GTC
 L   E170 L   I   G   Y   V   K   D   H   L   N180 S   G   T   E   V   V   S   R   V   V190

AAC TCA AAC CTC TCA GAC ATA GAG AAA CTC GGA TAC ATT ATT TTA CTT CTC ATA GCG GGT AAT GAG
 N   S   N   L   S   D   I   E   K   L200 G   Y   I   I   L   L   L   I   A   G210 N   E

ACT ACA ACT AAC TTA ATA TCA AAC TCT GTT ATT GAC TTC ACT AGG TTT AAC CTG TGG CAG AGG ATA
 T   T   T   N   L   I   S   N220 S   V   I   D   F   T   R   F   N   L230 W   Q   R   I

AGG GAA GAG AAC CTC TAC CTT AAG GCT ATC GAA GAG GCT TTA AGG TAT TCT CCT CCT GTG ATG AGG
 R   E   E   N   L   Y240 L   K   A   I   E   E   A   L   R   Y250 S   P   P   V   M   R

ACT GTA AGA AAG ACT AAG GAA AGA GTG AAA TTG GGT GAT CAG ACT ATT GAA GAG GGA GAG TAC GTT
 T   V   R   K260 T   K   E   R   V   K   L   G   D   Q270 T   I   E   E   G   E   Y   V

AGA GTA TGG ATA GCC TCA GCA AAC AGG GAC GAG GAG GTG TTT CAT GAC GGA GAG AAG TTC ATC CCT
 R   V280 W   I   A   S   A   N   R   D   E   E290 V   F   H   D   G   E   K   F   I   P300

GAC AGG AAT CCG AAC CCA CAC TTA AGC TTT GGG TCT GGA ATA CAT CTG TGT TTA GGT GCT CCT TTG
 D   R   N   P   N   P   H   L   S   F310 G   S   G   I   H   L   C   L   G   A320 P   L

GCT AGA TTA GAG GCA AGA ATA GCA ATT GAG GAA TTT TCA AAA AGG TTT AGG CAC ATT GAG ATA TTG
 A   R   L   E   A   R   I   A330 I   E   E   F   S   K   R   F   R   H340 I   E   I   L

GAT ACT GAA AAA GTT CCA AAT GAA GTG CTG AAT GGT TAT AAG AGA CTA GTG GTC AGG TTG AAG AGT
 D   T   E   K   V   P350 N   E   V   L   N   G   Y   K   R   L360 V   V   R   L   K   S

AAT GAA TAA TGC AGA GAG GTT TCT CCA CTT AAT GTT AAA GGA GGG AAG AAA GGT GTT ATT TCT GTT
 N   E  end
```

Fig. 1. Complete DNA and inferred amino acid sequence of *S. solfataricus* cytochrome P450. This sequence, along with portions of the 3′ and 5′ untranslated regions not shown above for reasons of space, are listed in the GenBank Data Base under accession number U51337.

mic DNA. This library was contained in plasmid pUC119 and had been constructed using the restriction enzyme *Xba*I. This resulted in the isolation of a ≈2.5 Kb fragment encoding the entire open reading frame for a potential protein possessing a strong resemblance to the members of the cytochrome P450 family.

The open reading frame for the potential archaeal cytochrome P450, *CYP119*, predicts a polypeptide of 368 amino

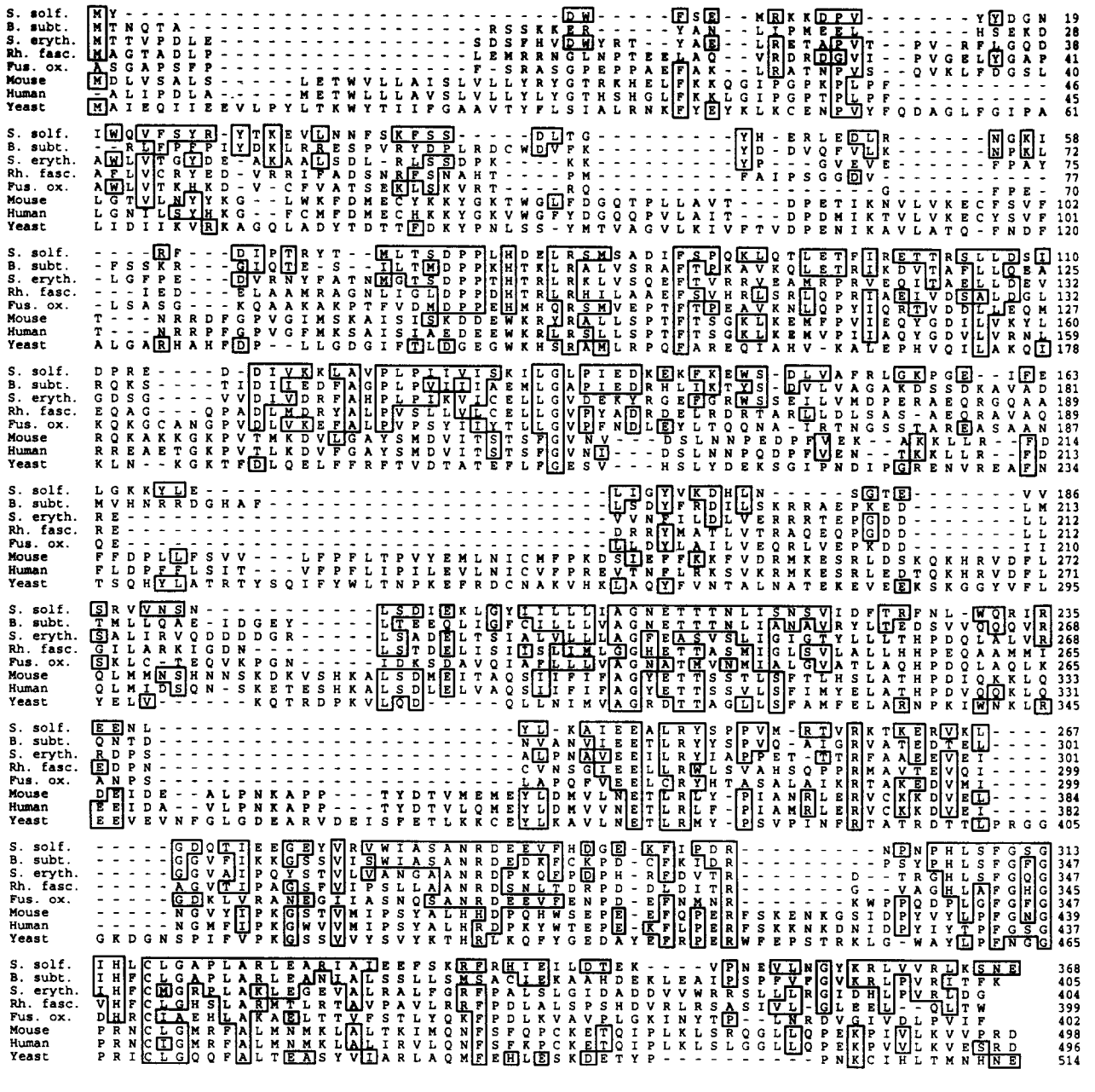

Fig. 2. Comparison of the inferred amino acid of the archaeal cytochrome P450 with other members of the cytochrome P450 family. Shown is the derived amino acid sequence of *S. solfataricus* (S. solf.) cytochrome P450 aligned with the corresponding regions of cytochrome P450s from three bacteria: CYP109 from *Bacillus subtilis* (B. subt. [10]), CYP107A1 from *Saccharopolyspora erythrea* (S. eryth. [11]), and ORF1 from the *fas* locus in *Rhodococcus fascians* (Rh. fasc. [12]), and four eukaryotes: CYP55 from the fungus *Fusarium oxysporum* (Fus. ox. [13]), CYP3A11 from mouse (mouse [14]), CYP3A4 from human (human [15]), and CYP56 from the yeast *Candida maltosa* (yeast [16]). Sequences were aligned using the BLASTX program from the National Center for Biotechnology Information. Identities with the archaeal cytochrome P450 are enclosed by boxes.

acids (Fig. 1). The first of the two potential N-terminal methionines, Met^1 and Met^8, was assigned as the initiator methionine based on two factors. First, the codon for Met^1 is located adjacent to a strong potential ribosome binding sequence, AAGGTAATTCC, as is typically the case for many archaeal genes [8]. Second, the sequence between Met^1 and Met^8 contains several amino acid residues conserved in other cytochrome P450s, most notably Phe^5, but also including Asp^3, Trp^4, and Glu^7 (Fig. 2). A sequence, AAACCATTTAAA, bearing a strong resemblance to the consensus promoter sequence for sulfur-dependent, thermophilic archaeons, AAANNTTTAAA [8], is located 77 nucleotides upstream of the predicted initiator methionine. A potential consensus transcription termination sequence, TTATTTCT [9], lies in a pyrimidine-rich region 48 bp downstream from the translation termination codon.

The predicted protein product lacks any evidence of an N-terminal membrane targeting sequence. This, along with sequence identity comparisons, indicates that it should be classified as one of the soluble, 'prokaryotic' (sometimes called B type [3]) cytochrome P450s found in many bacteria and some 'simple' eukaryotes, e.g. CYP55 of the fungus *Fusarium oxysporum* (Fig. 2). Its closest known homolog is CYP109 from *Bacillus subtilis*, with which it shares 35% amino acid sequence

Table 1
Sequence identity between *S. solfataricus* CYP119 and selected cytochrome P450s

Cytochrome P450	Organism	Identities	% identity
CYP109	*B. subtilis*	129	35
CYP107A1	*S. erythrea*	107	29
ORF1 of *fas* operon	*R. fascians*	85	23
CYP55	*F. fascians*	82	22
CYP3A11	Mouse	70	19
CYP56	*C. maltosa*	70	19
CYP3A4	Human	68	18

The number of amino acid identities observed between CYP119 from *S. solfataricus* and the cytochrome P450s described in Fig. 2 are tabulated. Percent identity levels were calculated by dividing the number of identical amino acids by the total number of amino acids in the *S. solfataricus* sequence, 368.

identity — 129 identical residues out of 368 (Table 1). CYP119 possesses a complete consensus heme-binding sequence, Phe-Gly-Xaa-Gly-Xaa-His/Arg-Xaa-Cys-Xaa-Gly-Xaa$_3$-Ala-Arg-Glu [1], centered around Cys^{317} near the C-terminal end. This fact, and the observation that the area of resemblance between the potential archaeal protein and a set of representative cytochrome P450s (Fig. 2) is spread quite generally throughout its sequence, suggests that the gene product very likely constitutes a functional enzyme. What might this function be? At this point in time it is difficult to tell, however, it is noteworthy that in its natural habitat *S. solfataricus* — which is an obligate aerobe — obtains energy primarily via the oxidation of elemental sulfur to sulfuric acid. Cytochrome P450s are already known to participate in sulfoxide formation [1] and have been postulated to comprise part of nature's most ancient respiratory systems [3]. Hence, it is tempting to speculate that the potential product of this open reading frame might participate in this metabolically vital sulfur oxidation process.

What can this archaeal gene sequence tell us about the evolution of the cytochrome P450 family? Even though we have ventured for the first time into the archaeal phylogenetic kingdom, the putative cytochrome P450 from *S. solfataricus* looks very much like any other prokaryotic cytochrome P450. In fact, construction of an abbreviated dendrogram containing a few representative bacterial and eukaryotic cytochrome P450s fails to betray any evidence of its archaeal status (Fig. 3). Like CYP55 from the fungus *F. oxysporum*, it groups among several bacterial cytochrome P450s as if it were derived from another bacterium rather than from an archaeon. Since the apparent lineage of this archaeal cytochrome P450 does not follow phylogeny, this suggests the occurrence of a horizontal gene transfer event. Two things seems apparent concerning the nature of this event. Since the Archaea do not contain subcellular compartments or organelles, acquisition cannot have proceeded via the incorporation of material from the genome of a bacterial endosymbiont into its genome, as could have been possible for a eukaryote such as *F. oxysporum*. Second, the first bifurcation in phylogeny is the separation of the Bacteria from the lineage that eventually gave rise to the Archaea and Eukaryotes. Although this means that *S. solfataricus* and *F. oxysporum* are much more closely related to each other than either is to any bacterium, *CYP119* and *CYP55* do not cluster together — once again in defiance of phylogeny. This suggests the occurrence of one or more horizontal gene transfer events after the separation of the archaeal and eukaryotic lineages one from another. However, the specific partners involved and the direction of the transfer, bacterium to eukaryote, archaeon to bacterium, etc. remains cryptic. The addition of future cytochrome P450 genes from the Archaea should do much to resolve these issues.

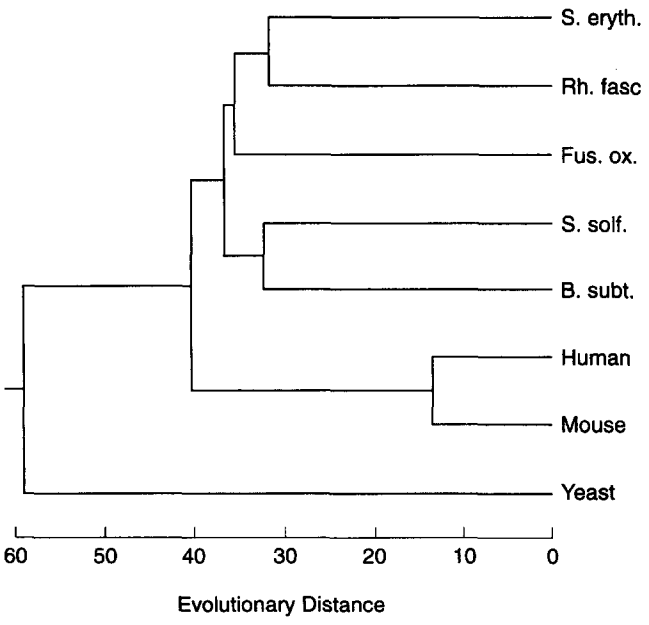

Fig. 3. Phylogenetic relationship of *S. solfataricus* cytochrome P450 to representative bacterial and eukaryotic cytochrome P450s. This tree was constructed using the Lasergene computer program from DNASTAR, Inc. (Madison, WI) and includes all of the cytochrome P450s shown in Fig. 2. The abbreviations used are explained in the legend to Fig. 2.

Acknowledgements: This work was supported by Grant 3502 from the Council for Tobacco Research, USA. We thank Daniel W. Nebert, University of Cincinnati, and David R. Nelson, University of Tennessee at Memphis, for their assistance with CYP gene nomenclature.

References

[1] Nebert, D.W. and Gonzalez, F.J. (1987) Annu. Rev. Biochem. 56, 945–993.
[2] Nebert, D.W. and Nelson, D.R. (1991) Methods Enzymol. 206, 3–11.
[3] Degtyarenko, K.N. and Archakov, A.I. (1993) FEBS Lett. 332, 1–8.
[4] Leng, J., Cameron, A.J.M., Buckel, S. and Kennelly, P.J. (1995) J. Bacteriol. 177, 6510–6517.
[5] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[6] Sanger, F., Nicklen, F. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[7] Roux, K.H. (1994) Biotechniques 16, 812–814.

[8] Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) CRC Crit. Rev. Microbiol. 16, 287–338.
[9] Reiter, W.-D., Palm, P. and Zillig, W. (1988) Nucleic Acids Res. 16, 2445–2458.
[10] Ahn, K.S. and Wake, R.G. (1991) Gene 98, 107–112.
[11] Haydock, S.F., Dowson, J.A., Dhillon, N., Cortes, J. and Leadlay, P.F. (1991) Mol. Gen. Genet. 230, 120–128.
[12] Crespi, M., Vereecke, D., Temmerman, W., von Montagu, M. and Desomer, J. (1994) J. Bacteriol. 176, 2492–2501.
[13] Kizawa, H., Tomura, D., Oda, M., Fukamizu, A., Hoshino, T., Gotoh, O., Yasui, T. and Shoun, H. (1991) J. Biol. Chem. 266, 10632–10637.
[14] Yanagimoto, T., Itoh, S., Muller-Enoch, D. and Kamataki, T. (1992) Biochim. Biophys. Acta 1130, 329–332.
[15] Gonzalez, F.J., Schmid, B.J., Umemo, M., McBride, O.W., Hardwick, J.P., Meyer, U.A., Gelboin, H.V. and Idle, J.R. (1988) DNA 7, 79–86.
[16] Schunck, W.H., Kaergel, E., Gross, B., Wiedmann, B., Gaestel, M. and Mueller, H.G. (1989) Biochem. Biophys. Res. Commun. 161, 843–850.