Overexpression of a designed 2.2 kb gene of eukaryotic phenylalanine ammonia-lyase in *Escherichia coli*

Mathias Baedeker, Georg E. Schulz*

Institut für Organische Chemie und Biochemie, Albert-Ludwigs-Universität, Albertstraße 21, D-79104 Freiburg im Breisgau, Germany

Received 28 June 1999; received in revised form 19 July 1999

Abstract Phenylalanine ammonia-lyase (EC 4.3.1.5) is a key enzyme in the secondary metabolism of higher plants catalyzing the non-oxidative conversion of L-phenylalanine into *trans*-cinnamate. The nucleotide sequence of its 2.2 kb gene was designed for expression in *Escherichia coli* and synthesized in a single reaction from 108 oligonucleotides using assembly PCR. After amplification, the gene was cloned into the expression vector pT7-7 and coexpressed with the chaperone HSP-60 system. The expression system yielded 70 mg of fully active enzyme per liter culture.

© 1999 Federation of European Biochemical Societies.

Key words: Gene synthesis; Phenylpropanoid metabolism; HSP-60 chaperone coexpression; Codon usage; Inclusion body

1. Introduction

Phenylalanine ammonia-lyase (PAL) catalyzes the non-oxidative conversion of L-phenylalanine into *trans*-cinnamate, which is the precursor of a great variety of phenylpropanoids such as lignins, flavonoids, and coumarins [1,2]. Previously, the reaction had been considered a Friedel–Crafts-like reaction [3] using a dehydroalanine formed autocatalytically from a serine. The structure analysis of the homologous enzyme histidine ammonia-lyase [4] (EC 4.3.1.3) following a similar mechanism [3] demonstrated, however, that the reactive electrophile is a 4-methylidene-imidazole-5-one (MIO). MIO is formed from the chain segment Ala-Ser-Gly in an autocatalytic reaction resembling the chromophore formation of the green fluorescent protein [5]. Since it plays a central role in the multibranched phenylpropanoid metabolism, PAL is an extensively studied plant enzyme and a potential target for herbicides [2]. Unfortunately, heterologous expression of the 2.2 kb PAL gene from parsley (*Petroselinum crispum*) in *Escherichia coli* [6] failed to yield enough protein for an intended structure analysis.

Several reasons for low production rates in *E. coli* are presently discussed, one of them is improper codon usage [7,8]. The PAL gene from parsley contains numerous low-usage codons distributed all over the gene. This problem cannot be solved by coexpression of a rare tRNA [9], and it is inefficient to change these codons by site-directed mutagenesis. We therefore decided to redesign the whole gene for the expression in *E. coli* [10]. Previously, gene synthesis had been performed by ligation [11,12], by the *Fok*I method [13], by self-priming PCR [14], and by assembly PCR [15]. We decided to use assembly PCR with a batch synthesis from 108 oligonucleotides as illustrated in Fig. 1. Since *E. coli* shows a tendency to dump eukaryotic proteins in inclusion bodies [16], we also applied a chaperone coexpression system to increase the cytosolic expression [17,18].

2. Materials and methods

2.1. Gene design and synthesis

Parsley contains four distinct PAL genes [19]. The gene coding for PAL1 (716 amino acids, 77828 Da) was de novo designed using the program package HUSAR (EMBL Heidelberg; http://genius.embnet.dkfz-heidelberg.de). We only applied codons preferred by *E. coli* and introduced silent point mutations in order to avoid secondary structures and to also create eight restriction sites. Finally, 108 oligonucleotides (106×40 nt, 1×41 nt, 1×42 nt) were purchased from GIBCO Lifetechnologies (Karlsruhe, Germany) at a scale of 50 nmol.

Every dried down oligonucleotide was resuspended in distilled water to a concentration of 250 μM and 1 μl amounts of each of these solutions were combined to give 108 μl mixture. For the batch synthesis, 0.2 μl of this mixture were added to 20 μl PCR-mix-1 (2 μl 10×*Pfu* buffer, 0.25 mM of each dNTP, 1.25 U of *Pfu* DNA polymerase from Stratagene). The first PCR program was performed with 55 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 30 s (Fig. 1). Subsequently, 6 μl of the batch were added to 14 μl fresh PCR-mix-1 and the second PCR program was started. It consisted of 52 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for (60+cycle number) s. Finally, 2.5 μl of the batch were added to 97.5 μl PCR-mix-2 (10 μl 10×*Pfu* buffer, 0.25 mM of each dNTP, 3.75 U of *Pfu* DNA polymerase from Stratagene) containing both outside primers at concentrations of 1 μM. The primers were oligonucleotides no. 01 and no. 55 representing the 5′-ends of both strands. Finally, the last PCR program (25 cycles at 94°C for 30 s, 55°C for 30 s and 72°C for 300 s) followed.

After checking the product on an analytical 1% agarose gel, the mixture was purified (PCR-Purification Kit, QIAGEN) and digested with *Nde*I and *Hin*dIII. The 2.2 kb band was then extracted from a preparative 1% agarose gel (QIAquick Gel Extraction Kit, QIAGEN) and cloned into vector pT7-7 using standard procedures.

The DNA sequence was determined using a blotter (GATC-1500) with Thermo Sequenase (Amersham). For mutagenesis we applied the mega-primer method [20] and QUICKchange (Stratagene). For coexpression with chaperones we transformed [21] *E. coli* BL21(DE3) harboring pT7-7(PAL) (ampicillin resistance) with the HSP-60 system-containing expression vector pREP4-groESL (kanamycin resistance; P. Caspers, Hoffmann-La Roche).

2.2. Protein expression and purification

E. coli BL21(DE3) transformed with plasmids pT7-7(PAL) and pREP4-groESL was grown at 37°C in 6×250 ml LB medium with 100 μg/ml ampicillin and 25 μg/ml kanamycin. At an OD_{578} of 0.9 the temperature was lowered to 20°C and the cells were induced with 1 mM IPTG, further cultivated for about 20 h and harvested by centrifugation. All subsequent handling was carried out at 4°C. The pellet was resuspended in 30 ml 50 mM KH_2PO_4 pH 7.6, 1 mM PMSF, 250 U of Benzonase (Merck) and sonicated. After centrifugation (60 min at 50 000×*g*) the crude extract was dialyzed overnight

*Corresponding author. Fax: +49-761-203-6161.
E-mail: schulz@bio.chemie.uni-freiburg.de

Abbreviations: MIO, 4-methylidene-imidazole-5-one; PAL, phenylalanine ammonia-lyase; PCR, polymerase chain reaction

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)01000-5

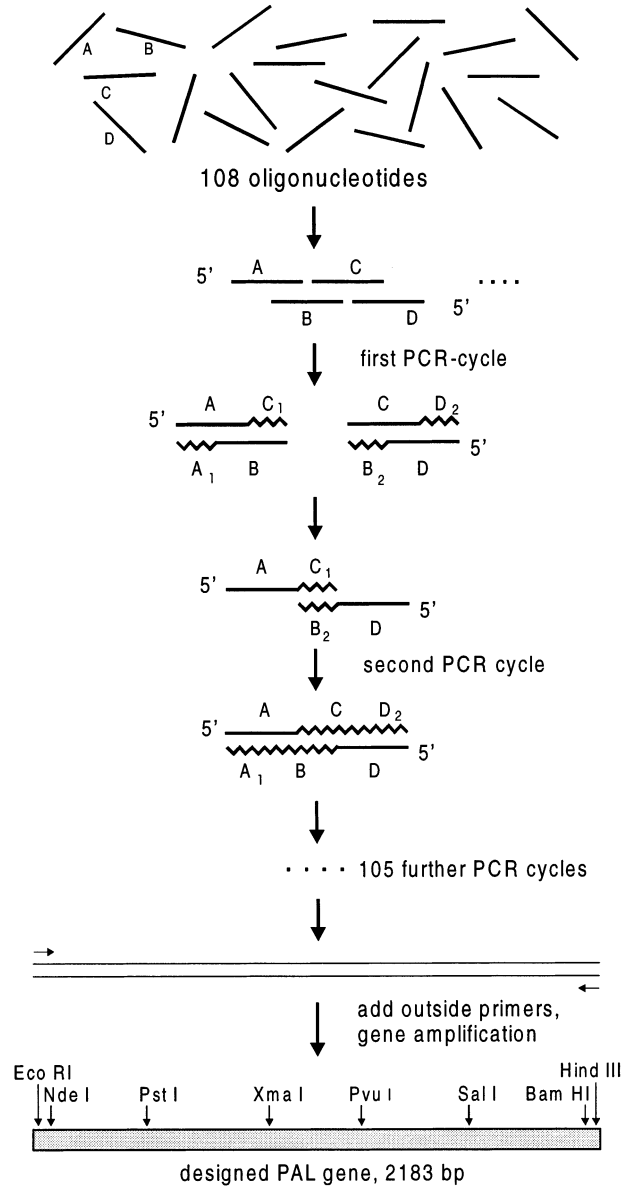

Fig. 1. Synthesis of a designed gene coding for PAL from parsley by the PCR-based gene assembly method [15]. The gene was assembled from a total of 108 oligonucleotides, in general 40-mers. During the first PCR cycle a duplex is formed between adjacent oligonucleotides (e.g. A+B). The following 106 cycles result in the integration of other oligonucleotides giving rise to the full length gene. The gene is then amplified with the two outside primers, i.e. with the first and the last oligonucleotide (no. 01 and no. 55).

against 20 mM KH_2PO_4 pH 7.6 (buffer A) with 1 mM PMSF. Modifying a previously described protocol [6], the protein solution was applied to an ion exchange column (Source Q, Pharmacia) and eluted with buffer A containing 1 M NaCl. The PAL-containing fractions were concentrated and applied to a gel filtration column (Superdex 200, Pharmacia) using 150 mM KH_2PO_4 pH 7.6, 600 mM NaCl and 3% (w/v) maltose.

2.3. Enzyme activity and crystallization

PAL activity was determined by monitoring the reaction product *trans*-cinnamate [22] at 290 nm ($\varepsilon_{290} = 10^4$ M^{-1} cm^{-1}). The assay contained 0.1 M Tris-HCl pH 8.8 and 20 mM L-phenylalanine and was performed at 30°C. For crystallization, the protein was concentrated to 10 mg/ml (Centriprep, Millipore), dialyzed against deionized water, and used in a hanging drop screening procedure [23]. Crystals grew in a 1:1 mixture of the protein with 0.1 M sodium acetate pH 4.6, 2–3% PEG 4000, 10 mM LiCl. They were generally accompanied by protein precipitation.

3. Results and discussion

In order to express milligram amounts of PAL in *E. coli*, we have synthesized a gene coding for the 716 amino acids of PAL by assembly PCR [15]. To our knowledge this is the largest synthetic gene published so far, longer synthetic DNA sequences have been described for entire plasmids [13,15]. For this purpose, we designed 108 oligonucleotides (106×40 nt, 1×41 nt, 1×42 nt) coding for both strands of the gene. We decided to use short oligonucleotides in order to limit synthesis errors. The overlapping stretches of complementary oligonucleotides were always 20 nt in length (Fig. 1). Eight restriction sites (*Eco*RI, *Nde*I, *Pst*I, *Xma*I, *Pvu*I, *Sal*I, *Bam*HI, *Hin*dIII) were introduced to facilitate future cloning and mutagenesis studies. The nucleotide sequence identity between the PAL gene from parsley and the designed synthetic gene was as low as 75%.

Following Stemmer et al. [15], the assembly PCR was carried out in one batch: (i) single stranded ends of complementary oligonucleotides were filled in and increasingly larger fragments were created until the full length gene was obtained after at most 107 cycles (Fig. 1); (ii) the resulting gene was amplified and run on an agarose gel (Fig. 2); (iii) the gene was extracted from the gel, cut with *Nde*I and *Hin*dIII, and cloned into the expression vector pT7-7.

Surprisingly, the agarose gel showed two major bands, one with the expected length of 2.2 kb, the other with 1.1 kb (Fig. 2). To remove the shorter band, we repeated the amplification step, but obtained the same two bands, indicating that the 1.1 kb product contained the correct sequence at both ends. A re-evaluation of all 106 complementary 20 nt seg-

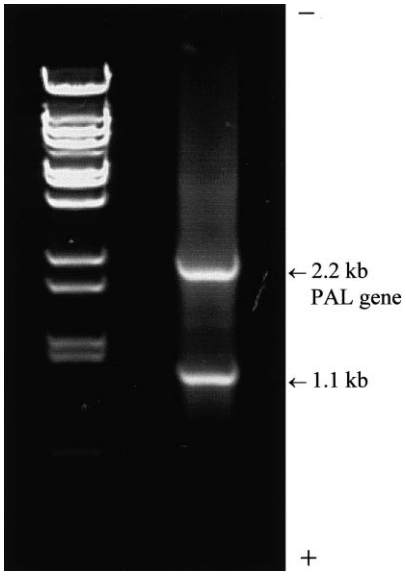

Fig. 2. Analytical agarose gel electrophoresis after gene amplification. Lane 1, λ DNA-*Bst*EII standard (New England Biolabs); lane 2, result of the amplification step showing the 2.2 kb PCR product, the second band is a an artefact probably caused by an incorrect match. It was separated by gel extraction.

```
gggaattcac  atATGGAAAA  CGGTAACGGC  GCTACCACTA  ACGGTCACGT  GAACGGCAAC
GGTATGGACT  TCTGCATGAA  AACCGAAGAT  CCTCTGTACT  GGGGTATCGC  TGCGGAGGCT
ATGACTGGTT  CCCACCTGGA  CGAAGTTAAA  AAGATGGTTG  CTGAATATCG  TAAACCGGTT
GTTAAACTGG  GTGGCGAAAC  TCTGACCATC  TCCCAAGTTG  CTGCAATCTC  TGCTCGTGAC
GGTTCCGGTG  TTACTGTTGA  ACTGTCCGAA  GCTGCGCGTG  CTGGTGTTAA  AGCGTCCTCT
GACTGGGTTA  TGGACTCCAT  GAACAAAGGT  ACCGACTCCT  ATGGCGTTAC  CACTGGTTTC
GGCGCTACCT  CCCATCGTCG  TACCAAACAG  GGGGGTGCAC  TGCAGAAAGA  ACTGATCCGC
TTCCTGAACG  CTGGTATCTT  CGGTAACGGT  TCTGACAATA  CGCTGCCGCA  TTCCGCTACC
CGTGCTGCTA  TGCTGGTTCG  TATCAACACC  CTGCTGCAAG  GTTACTCTGG  TATCCGTTTC
GAAATCCTGG  AGGCTATCAC  GAAATTCCTG  AACCAGAACA  TCACCCCGTG  CCTGCCGCTG
CGTGGTACCA  TCACTGCTTC  CGGCGACCTG  GTTCCACTGT  CCTACATCGC  TGGTCTGCTG
ACTGGTCGTC  CGAACTCTAA  AGCTGTTGGT  CCGACTGGTG  TTATCCTGTC  CCCGGAAGAA
GCGTTCAAAC  TGGCTGGTGT  GGAAGGTGGT  TTCTTTGAAC  TGCAACCGAA  AGAGGGCCTG
GCACTGGTTA  ACGGTACCGC  TGTTGGTTCT  GGTATGGCGT  CCATGGTTCT  GTTCGAAGCT
AACATCCTGG  CTGTTCTGGC  GGAAGTGATG  TCTGCTATCT  TCGCTGAAGT  TATGCAGGGT
AAACCAGAGT  TCACCGACCA  CCTGACTCAC  AAACTGAAAC  ACCACCCGGG  TCAGATCGAA
GCTGCTGCTA  TCATGGAACA  CATCCTGGAC  GGTTCTGCCT  ACGTTAAAGC  TGCTCAGAAA
CTGCACGAAA  TGGACCCGCT  GCAAAAACCG  AAACAGGACC  GTTATGCTCT  GCGTACCTCT
CCACAGTGGC  TGGGCCCGCA  AATCGAAGTT  ATCCGCTCCT  CTACCAAGAT  GATCGAACGT
GAAATCAACT  CTGTTAACGA  CAACCCGCTG  ATCGACGTTT  CCCGCAACAA  AGCTATCCAC
GGTGGTAACT  TCCAGGGGAC  CCCGATCGGC  GTTTCCATGG  ACAACACCCG  TCTGGCTATC
GCAGCTATCG  GTAAACTGAT  GTTCGCTCAA  TTCTCTGAAC  TGGTTAACGA  CTTCTACAAC
AACGGTCTGC  CATCTAACCT  GTCTGGTGGT  CGTAACCCGT  CCCTGGACTA  TGGTTTCAAA
GGTGCAGAAA  TCGCTATGGC  TTCCTACTGT  TCTGAACTGC  AATTCCTGGC  TAACCCGGTT
ACCAACCACG  TTCAGTCCGC  AGAACAGCAC  AACCAAGACG  TTAACTCTCT  GGGTCTGATC
TCTTCTCGTA  AAACCTCTGA  AGCTGTTGAA  ATCCTGAAAC  TGATGTCCAC  TACCTTCCTG
GTTGGTCTGT  GTCAAGCTAT  CGACCTGCGT  CACCTGGAAG  AAAACCTGAA  ATCCACCGTT
AAAAACACCG  TGTCTTCCGT  GGCTAAACGT  GTTCTGACGA  TGGGTGTTAA  TGGAGAACTG
CACCCGTCCC  GTTTCTGCGA  AAAAGACCTG  CTGCGTGTTG  TCGACCGTGA  ATACATCTTT
GCTTACATCG  ACGACCCGTG  CTCCGCTACC  TACCCACTGA  TGCAGAAACT  GCGTCAGACC
CTGGTTGAAC  ATGCTCTGAA  AAACGGTGAC  AACGAACGTA  ACCTGTCTAC  CTCCATCTTC
CAGAAAATTG  CAACCTTCGA  AGATGAACTG  AAAGCTCTGC  TGCCGAAAGA  AGTTGAATCC
GCTCGTGCAG  CACTGGAATC  TGGTAACCCT  GCTATCCCAA  ACCGTATCGA  AGAATGCCGT
TCCTACCCGC  TGTACAAATT  CGTTCGTAAA  GAACTGGGCA  CTGAATACCT  GACCGGTGAA
AAAGTTACCT  CCCCAGGTGA  AGAGTTCGAA  AAAGTTTTCA  TCGCTATGTC  CAAAGGTGAA
ATCATCGACC  CGCTGCTGGA  ATGCCTGGAA  TCCTGGAATG  GTGCTCCGCT  GCCGATCTGC
TAAtgggatc  caacaagctt  gcg
```

Fig. 3. Artificial gene of phenylalanine ammonia-lyase from parsley as designed for expression in *E. coli*. The length is 2183 bp. The translated part is in capital letters. The restriction sites are underlined, the respective restriction enzymes can be deduced from Fig. 1.

ments of the oligonucleotides (Fig. 1) revealed a match between segments no. 30 and no. 82 (10 G-C and 3 T-A pairs) to be the best incorrect fit. This match results in a 1.1 kb artefact and therefore explains the additional band.

After cloning the 2.2 kb band into plasmid pT7-7, digestions with the respective restriction enzymes demonstrated that all six internal restriction sites were present. A more detailed analysis of the DNA sequence showed, however, that the reaction had caused seven point mutations. One of them was a silent mutation, two mutations caused amino acid exchanges, and four single bases were missing. Possibly, the unexpected high error rate was caused by incorrect oligonucleotides. The raw synthetic gene was then corrected to the designed sequence using a standard point mutation procedure. The designed sequence is given in Fig. 3.

An expression of this artificial gene in *E. coli* at 37°C resulted in inclusion body formation; only a small fraction of the enzyme remained in the cytosol and showed activity. This phenomenon is known for the expression of numerous eukaryotic proteins in *E. coli* [16,18]. In order to increase the cytosolic expression, we lowered the incubation temperature to 20°C and added a plasmid carrying the HSP-60 system to the *E. coli* strain containing plasmid pT7-7(PAL). This procedure resulted in a dramatic increase of active PAL. After purification we obtained 70 mg pure PAL per l culture with a specific activity of 2.4 U/mg, which is in the reported range [24].

The difference between the reported and an earlier [6] expression system is illustrated in Fig. 4. The previously described expression with the plant gene was low and no inclusion bodies were produced at 37°C. The expression with the designed gene at 37°C resulted in inclusion bodies containing most of the protein, and it retained about as much active PAL

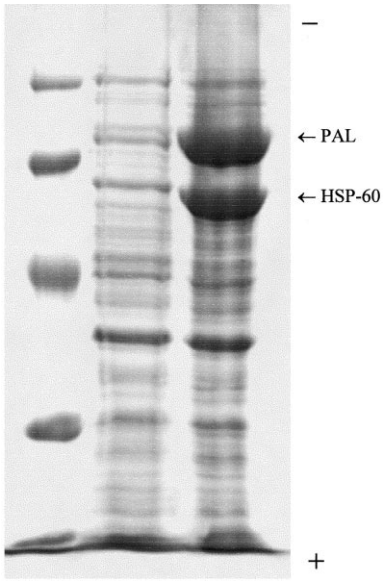

Fig. 4. SDS-PAGE of PAL expression in *E. coli*. Lane 1, molecular mass markers (94, 67, 43, 30 kDa); lane 2, supernatant after cell lysis of a previously described expression system in *E. coli* BL21(DE3) [6] that produced no inclusion bodies at 37°C; lane 3, supernatant after cell lysis of the reported expression system in *E. coli* BL21(DE3) (coexpression of the HSP-60 system) that produced no inclusion bodies.

in the cytosol as the previous expression system. By cultivating at 20°C (factor of 60) and by adding the chaperone system (factor of 2) the amount of cytosolic active PAL increased by a total factor of 120. The resulting enzyme was purified and used for crystallization. The initially obtained crystals diffracted merely to 6 Å resolution and showed some disorder preventing an X-ray structure analysis. Crystal improvement trials are undertaken, among them are mutations for changing the protein surface [25,26]. The successful design and synthesis of the PAL gene confirms the usefulness of this approach for expressing an eukaryotic gene of interest in a simple expression system like *E. coli*.

Acknowledgements: We thank J. Rétey for providing us with his PAL expression system in *E. coli* and P. Caspers for the expression vector pREP4-groESL carrying the HSP-60 system. The work was supported by the Deutsche Forschungsgemeinschaft under SFB-388.

References

[1] Hanson, K.R. and Havir, E.A. (1978) Recent Adv. Phytochem. 12, 91–137.
[2] Hahlbrock, K. and Scheel, D. (1989) Annu. Rev. Plant Physiol. Plant Mol. Biol. 40, 347–369.
[3] Rétey, J. (1996) Naturwissenschaften 83, 439–447.
[4] Schwede, T.F., Rétey, J. and Schulz, G.E. (1999) Biochemistry 38, 5355–5361.
[5] Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J. (1996) Science 273, 1392–1395.
[6] Schuster, B. and Rétey, J. (1995) Proc. Natl. Acad. Sci. USA 92, 8433–8437.
[7] Makrides, S.C. (1996) Microbiol. Rev. 60, 512–538.
[8] Chen, G.-F.T. and Inouye, M. (1990) Nucleic Acids Res. 18, 1465–1473.
[9] Brinkmann, U., Mattes, R.E. and Buckel, P. (1989) Gene 85, 109–114.
[10] Casimiro, D.R., Wright, P.E. and Dyson, H.J. (1997) Structure 5, 1407–1412.
[11] Heyneker, H.L., Shine, J., Goodman, H.M., Boyer, H.W., Rosenberg, J., Dickerson, R.E., Narang, S.A., Itakura, K., Lin, S.y. and Riggs, A.D. (1976) Nature 263, 748–752.
[12] Au, L.-C., Yang, F.-Y., Yang, W.-J., Lo, S.-H. and Kao, C.-F. (1998) Biochem. Biophys. Res. Commun. 248, 200–203.
[13] Mandecki, W. and Bolling, T.J. (1988) Gene 68, 101–107.
[14] Dillon, P.J. and Rosen, C.A. (1990) BioTechniques 9, 298–300.
[15] Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Gene 164, 49–53.
[16] Hartl, F.U. (1996) Nature 381, 571–580.
[17] Dionisi, H.M., Checa, S.K., Krapp, A.R., Arakaki, A.K., Ceccarelli, E.A., Carrillo, N. and Viale, A.M. (1998) Eur. J. Biochem. 251, 724–728.
[18] Vonrhein, C., Schmidt, U., Ziegler, G.A., Schweiger, S., Hanukoglu, I. and Schulz, G.E. (1999) FEBS Lett. 443, 167–169.
[19] Lois, R. and Hahlbrock, K. (1992) Z. Naturforsch. 47c, 90–94.
[20] Sakar, G. and Sommer, S.S. (1990) BioTechniques 8, 404–407.
[21] Hanahan, D. (1983) J. Mol. Biol. 166, 557–580.
[22] Zimmermann, A. and Hahlbrock, K. (1975) Arch. Biochem. Biophys. 166, 54–62.
[23] Jancarik, J. and Kim, S.-H. (1991) J. Appl. Crystallogr. 24, 409–411.
[24] Langer, B., Röther, D. and Rétey, J. (1997) Biochemistry 36, 10867–10871.
[25] Schwede, T.F., Baedeker, M., Langer, M., Rétey, J. and Schulz, G.E. (1999) Protein Eng. 12, 151–153.
[26] Pautsch, A., Vogt, J., Model, K., Siebold, C. and Schulz, G.E. (1999) Proteins Struct. Funct. Genet. 34, 167–172.