© 1991 Federation of European Biochemical Societies 00145793/91/\$3.50 *ADONIS* 001457939100653K

Isolation and nucleotide sequence of the extracellular acid protease gene (ACP) from the yeast *Candida tropicalis*

Giuseppe Togni¹, Dominique Sanglard², Rocco Falchetto³ and Michel Monod¹

¹Laboratoire de Mycologie, Service de Dermatologie, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland, ²Institut für Biotechnologie ETH-Hönggerberg, 8093 Zürich, Switzerland and ³Laboratorium für Biochemie, ETH-Zentrum, 8092 Zürich, Switzerland

Received 18 March 1991; revised version received 30 April 1991

The extracellular acid protease of Candida tropicalis was purified from the supernatant fraction of culture medium containing bovine serum albumin as nitrogen source and the NH₂-terminal amino acid (aa) sequence of the protein was determined. The gene for the acid protease (ACP) was isolated using a pool of synthetic oligonucleotides as a probe and a segment of the deduced as sequence was found to be in agreement with the NH₂-terminal as sequence of the protein. The deduced as sequence of ACP is similar to the as sequence of proteases of the pepsin family. The nucleotide sequence of the 5' portion of this gene revealed a coding sequence for a 60 residue propeptide containing two Lys-Arg amino acid pairs that have been identified as sites for peptidase processing of several exported peptides and proteins. The final Lys-Arg site occurs at the junction with the mature extracellular form of the acid protease.

Candida tropicalis; Protease; DNA sequence; Secretion C. tropicalis acid protease (EC 3. 4. 23. 6.)

1. INTRODUCTION

Three opportunistic species of the genus Candida, C. albicans, C. tropicalis and C. parapsilosis secrete an acid protease (ACP) when protein is the sole nitrogen source in the growth medium [1]. The proteases of the three species possess common and species-specific antigenic sites [2]. For C. albicans and C. tropicalis, secretory proteolytic activity is considered to be a major factor of virulence [2-6]. Although C. parapsilosis synthesizes large amounts of ACP in vitro, as do C. albicans and C. tropicalis, it appears to be weakly virulent in vitro [5] and in vivo [7]. Consistent with this observation it has been shown that the extracellular ACP of C. albicans and C. tropicalis, but not ACP of C. parapsilosis, is expressed during phagocytosis by murine peritoneal macrophages [2,6]. C. albicans and C. tropicalis are medically more important as opportunistic organisms causing infections in immunocompromised patients. In particular, in patients with leukemia, the most commonly found infecting Candida species is C. tropicalis [8,9].

We are interested in the isolation and characterisation of the Candida acid protease genes in order to

Abbreviations: aa, amino acid; ACP, extracellular acid protease; ACP, extracellular acid protease gene; BSA, bovine serum albumin; kb, kilobase; kDa, kilodalton; ORF, open reading frame; PAS, periodic acid-Schiff reagent

Correspondence address: M. Monod, Laboratoire de Mycologie, Service de Dermatologie, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland

elucidate the molecular basis of their regulation and to determine their function in virulence by making directed mutants by gene disruption. Here we report the isolation and the nucleotide sequence of the *C. tropicalis* ACP gene (ACP).

2. MATERIALS AND METHODS

2.1. Strains and growth conditions

C. tropicalis ATCC 750 was used and propagated on Sabouraud agar plates. For enzyme production, the strains were grown in a liquid medium containing 0.2% bovine serum albumin (BSA) and 1.17% yeast carbon base (YCB; Difco). The medium was adjusted to pH 4.0 with HCl. Flasks containing 500 ml medium were inoculated with approximately 10⁷ cells and incubated for 4 days at 30°C on an orbital shaker at 200 rpm.

2.2. Enzyme purification

The supernatant of a 500 ml culture was separated from yeast cells by centrifugation at $5000 \times g$. Sufficient ammonium sulphate was dissolved at room temperature in the supernatant to give a 65% saturated solution. The precipitated proteins were collected by centrifugation and resuspended in distilled water at 1:100 of the original volume. Insoluble material was removed by centrifugation at $5000 \times g$ for 5 min and the supernatant was dialyzed against 15 mM Nacitrate (pH 5.6), for 2 h. Subsequently, the protease in the dialyzate was adsorbed to a 2 ml volume column of cellex E cellulose (Biorad) equilibrated with 15 mM Na-citrate (pH 5.6). After washing the column with the same buffer, the enzyme was eluted with 80 mM Nacitrate, pH 5.6, and fractions with enzymatic activity were retained and pooled.

Enzymatic activity was assayed using as a substrate 0.1% BSA in 50 mM Na-acetate (pH 3.6) in a total volume of 1 ml. After precipitation of undigested substrate by trichloroacetic acid (4% final concentration) and centrifugation, soluble peptides remaining in the supernatant were measured by the Lowry procedure [10]. For practical purposes, one unit of enzyme activity was defined as that producing an increase in absorbance (A 220 nm) of 0.001/min.

Protein concentrations were measured by the method of Lowry [10] with BSA as a protein standard.

Proteins extracts were analyzed by SDS-PAGE according to Laemmli [11] using a 12% polyacrylamide gel. Gels were either stained with Coomassie brilliant blue R-250 or were stained for glycoproteins with the periodic acid-Schiff reagent (PAS) [12].

2.3. Effect of proteinase inhibitors

The inhibitors were all obtained from Sigma (St. Louis, MO, USA). Concentrated stock solutions of the inhibitors were added to 0.5 ml of enzyme solution to give the following final concentrations: EDTA (2 mM), iodoacetamide (2 mM), dithiothreitol (DTT; 2 mM), phenylmethylsulphonyl fluoride (PMSF; 2 mM), soybean trypsin inhibitor (SBTI; 200 μ g/ml) and pepstatin (1 μ g/ml). After incubation of the enzyme-inhibitor mixtures for 20 min at 30°C, followed by the addition of 0.5 ml of 0.2% BSA, assays were performed as described above. Stock solutions of EDTA, iodoacetamide, DTT and SBTI were prepared in water. Pepstatin and PMSF were dissolved in ethanol and isopropanol, respectively.

2.4. Amino acid sequencing

The ACP was first separated on Tricine-SDS-PAGE [13] and then transferred onto PVDF-Immobilon membranes (Millipore). After staining with Coomassie blue, the protein band was cut out and analysed by a gas phase sequenator 470A (Applied Biosystems) with on-line PTH-derivative analyser 120A.

2.5. C. tropicalis ACP gene cloning

E. coli P2392: hsdR514 (rk-, mk+) supE44, supF58, lacY1, galK2, galT22, metB1, trpR55, (P2) was used to grow bacteriophage λDASHII (Stratagene, La Jolla, CA). Plasmid standard transformations were performed in E. coli DH5 (supE44, hsdR17, recA1, endA1, gyrA96, thi1, relA1). Approximately 50000 bacteriophage \(\lambda DASHII \) plaques of a previously constructed C. tropicalis genomic library [14] were immobilized on nylon membranes (Gene Screen Plus, Dupont). These filters were probed with a pool of 33mer oligonucleotides corresponding to the codons of the determined ACP N-terminal sequence in a solution containing 5 × SSC, 1% SDS and 5 × Denhardt's at 37°C for 24 h. The oligonucleotides were labelled with $[\gamma\text{-}^{32}P]ATP$ (Amersham) using polynucleotide kinase (Boehringer). The membranes were exposed to X-ray film after successive washes at 40°C, 50°C and 55°C in 5 × SSC and 1% SDS. Bacteriophages showing positive signals after these washes were purified and their DNA was isolated in small scale preparations by high-speed centrifugation [15]. Agarose gel electrophoresis of bacteriophage \(\lambda DASHII \) DNA subjected to enzyme restriction and Southern blotting were performed according to standard protocols [15].

2.6. DNA subcloning and sequencing

The plasmid pMTL 21 [16] was used for subcloning DNA fragments. For sequencing, plasmid DNA was isolated and purified according to Maniatis [15]. After denaturation (0.2 N NaOH and 2 mM EDTA), neutralisation and precipitation, DNA was annealed with the reverse primer (Biofinex, Praroman, Switzerland), SK primer (USB, Cleveland, USA), or synthetic designed oligonucleotide-primers (Microsynth, Windisch, Switzerland). Following annealing, the sequence was determined with [35S]dATP (Amersham) using a sequenase version 2.0 sequencing kit (USB, Cleveland, USA) according to the recommendations and protocols of the supplier.

3. RESULTS

3.1. Isolation and partial characterization of ACP

In a typical experiment, a total of 23600 U of proteolytic activity was produced in one liter of BSA medium after 4 days of growth at 30°C. The C. tropicalis ACP was purified as described in section 2 with a yield of 43%. The purified enzyme showed a

single protein band in SDS-PAGE gel stained with Coomassie blue (Fig. 1, track 1) with an estimated molecular mass of 40 kDa. This band also stained with PAS reagent indicating that the enzyme was a glycoprotein. The specific activity of the purified protein was 5000 U/mg. The enzyme was the major protein secreted by the yeast as shown by the protein profile of the 65% saturated ammonium sulphate precipitate of the culture supernatant (Fig. 1, track 2).

The protease activity of the purified enzyme was totally inhibited by pepstatin whereas it was not affected by any of the other enzyme inhibitors tested here. The effects of the various protease inhibitors suggested that the *C. tropicalis* ACP, like those of *C. albicans* and *C. parapsilosis* [2], belonged to the pepsin family (EC 3.4.23.6).

3.2. Cloning and DNA sequence of the putative C. tropicalis ACP gene

The initial 28 amino acid residues of the N-terminus of the protein were determined to be Ser-Asp-Val-Pro-Thr-Thr-Leu-Ile-Asn-Glu-Gly-Pro-Ser-Tyr-Ala-Ala-Asp-Ile-Val-Val-Gly-Ser-Asn-Gln-Gln-Lys-Gln-Thr.

A pool of 33mer oligonucleotides was designed to detect ACP in the C. tropicalis genomic library. The design was based on the N-terminal aa sequence of the protein and on the codon usage of three other sequenced genes of the same C. tropicalis strain [14,17,18] using the most probable codons for the N-terminal residues 2-12 with the least degeneracy. The following mixture of oligonucleotides was synthesized:

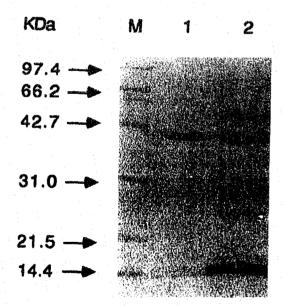


Fig. 1. Electrophoresis of *C. tropicalis* ACP preparations (12% SDS-PAGE). (1) One µg purified enzyme sample. (2) Proteins precipitated from the culture medium by 65% saturated ammonium sulphate. The gels were stained with Coomassie brilliant blue R-250. M, molecular mass markers (phosphorylase B, 97.4 kDa; bovine serum albumin, 66.2 kDa; ovalbumin, 42.7 kDa; bovine carbonic anhydrase, 31.0 kDa; soybean trypsin inhibitor, 21.5 kDa; lysozyme, 14.4 kDa).

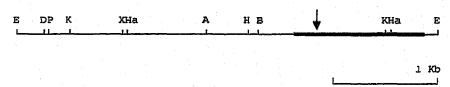
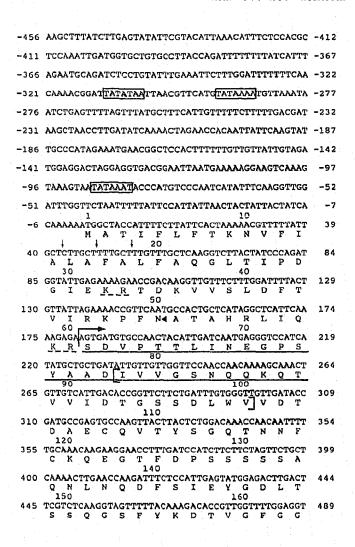


Fig. 2. Restriction map of the 4.0 kb EcoRI fragment obtained from the positive bacteriophages. The ACP ORF is represented by the filled region. An arrow indicates the position of hybridization of the oligonucleotide probe. Both strands have been sequenced between the HindIII and EcoRI sites (A, Accl; B, Bg/II; D, Dral; E, EcoRI; H, HindIII; Ha, HaeIII; K, Kpnl; P, PvuII; X, Xhol).

GAT GTT CCA $AC_C^TAC_C^TTT_A^G$ ATT AA_C^TGAA GGT CCA.

The screening of the C. tropicalis λ DASHII genomic library led to the identification of two clones which hybridized with the designed oligonucleotide probe. Restriction enzyme digestion of purified DNA revealed that the two clones carried similar but not identical

yeast sequences. A 4 kb EcoRI fragment present in both clones was recognized by the designed probe (data not shown). This fragment, of which a restriction map is shown in Fig. 2, was subcloned in pMTL 21, generating the plasmid pMTL 21-E4. The sequence hybridizing with the oligonucleotides could be localised on a stretch of 1400 bp between a HindIII and a HaeIII



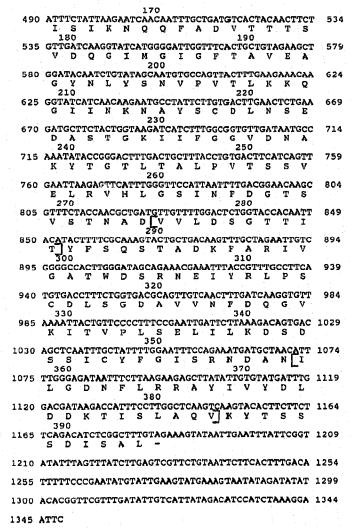


Fig. 3. Nucleotide sequence of the *C. tropicalis ACP*. The deduced as sequence is shown below the nucleotide sequence. The N-terminal as sequence of the mature enzyme begins at residue 61 and is preceded by a vertical bar and a horizontal arrow. The 28 as sequence from N-terminus of the protein deduced by as sequencing is underlined. The two KEX2-like peptidase processing sites found in the propeptide as sequence are dotted and putative cleavage sites within the signal sequence for secretion are indicated by an arrow. A possible N-linked glycosylation site in the propeptide is marked by a filled triangle and the three domains showing strong homology with proteins member of the pepsin family are delimited with brackets. Possible TATA sequences upstream of the large ORF are boxed.

site. The entire fragment together with an additional 400 bp between the *HaeIII* and *EcoRI* sites were sequenced.

3.3. Sequence analysis

One long open reading frame (ORF) of 1182 bp coding for a protein of 394 aa was found, starting 457 bp downstream the HindIII site. The sequence of this long ORF and its 5'- and 3'-flanking regions along with the corresponding aa sequence is shown in Fig. 3. The N-terminal aa sequence of the mature ACP obtained by Edman amino acid sequencing is preceded by a polypeptide of 60 aa and, in addition, by a pair of Lys-Arg residues just upstream of the N-terminal aa of the mature form of ACP. Precursors of hormones in mammalian systems [19], α -factor in S. cerevisiae [20] and secreted alkaline protease in Yarrowia lipolytica [21] are known to be processed by peptide cleavage immediately after pairs of basic residues. The membrane bound proteinase ysc F encoded by the gene KEX2 in S. cerevisiae is specific for cleaving peptides after the Lys-Arg or Arg-Arg tandem sequences [20,22]. A similar endopeptidase is likely therefore to be present in C. tropicalis. Another potential peptidase processing site Lys-Arg pair is found in position 32-33 in the precursor propeptide upstream of the N-terminus of the mature protein.

Inspection of the N-terminal amino acid sequence beginning from the Met-1 residue suggested the existence of a signal peptide in the protease precursor polypeptide. This signal peptide begins with an hydrophobic core of 7 aa and, 6–10 aa downstream, contains 3 putative signal peptidase cleavage sites in accordance with the -3-1 van Heijne's rule [23]. They are indicated by an arrow on Fig. 3. In summary, the structure of the leader propeptide sequence upstream of the mature form of the *C. tropicalis* ACP appeared to be similar to that of the secretory alkaline protease of *Y. lipolytica* [21].

The protein molecule generated by cleavage at the KEX2-like cleavage site has a calculated molecular mass of 35.7 kDa which is 4 kDa less than the size estimated from the electrophoretic mobility on SDS-PAGE. The difference may be explained by the fact that the mature ACP is a glycoprotein. Indeed, it is known that the molecular masses of glycoproteins are overestimated on SDS-PAGE [24]. The C. tropicalis ACP was shown to be a mannose protein [2], but glycosylation sites on the mature protein have not been investigated. The only putative N-glycosylation site present in the ORF was found in the propeptide sequence, so that the mature protein must contain O-glycosylated residues and not N-glycosylated ones.

3.4. The 5'- and 3'-flanking regions of the putative ACP gene

The 5' region of most eukaryotic genes including

those of yeast shows the presence of the model sequence $TATA_A^TA_A^T$ involved in mRNA transcription start-site selection [25]. In the upstream region three such sequences occur at -391, -292 and -88 constituting possible TATA boxes. A well-conserved purine residue at -3 observed in the proximity of ATG start codon of most yeast genes is also present in ACP [26]. No transcription termination consensus sequence $TAAATAA_G^A$ [27] is found at the 3'-flanking region of the ORF as it was found for other genes of C. tropicalis [14,17,18].

3.5. Similarity with proteins of the pepsin family

The deduced amino acid sequence of the putative C. tropicalis ACP translation product was compared with sequences in the FASTP Protein Identification Program [28]. A relevant degree of similarity was observed with proteases of the pepsin family. High scores were obtained with an acid protease from Saccharomycopsis fibuligera [29], Rhizopus aspartic proteases [30,31], the human pepsinogen A precursor [32], the protease A and the barrier protein from Saccharomyces cerevisiae

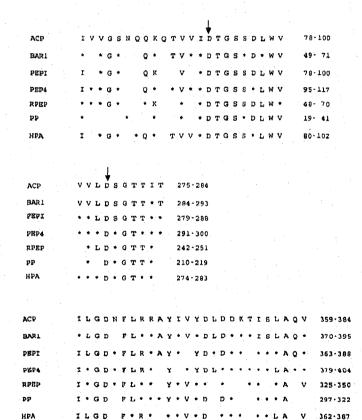


Fig. 4. Domains of homology between the deduced as sequence of C. tropicalis ACP and other protease members of the pepsin family. Only identical amino acids are written by the one letter code. Conservative changes are marked by a star and aspartic acid residues of the reactive site of pepsin-like proteases are indicated by an arrow. (BAR1, S. cerevisiae barrier protein [34]; PEP1, Saccharomycopsis fibuligera ACP [29]; PEP4, S. cerevisiae proteinase A [33]; RPEP, Rhizopus chinensis ACP [30]; PP, peniciliopepsin [36]; HPA, human pepsinogen A [32]).

[33,34]. In particular, upon alignment of the ACP with these pepsin-like proteases three conserved regions were observed (Fig. 4). Two of these regions contain the two reactive aspartic acid residues of the active site of the pepsin-like proteases [34].

4. DISCUSSION

This is the first report of a Candida gene coding for ACP. It has been suggested that extracellular acid proteases of Candida species are involved in the pathogenicity of these organisms. Now that the C. tropicalis ACP is available, such a hypothesis can be tested by a genetic approach. We intend to disrupt this gene in C. tropicalis and to examine the pathogenicity of ACP-deficient C. tropicalis mutants in vivo. Furthermore, experiments are also planned to examine the maturation of the protein and the control of transcription of this gene at the molecular level.

An aspartyl protease gene has also been identified in *C. albicans* through hybridisation with the vacuolar *S. cerevisiae* protease A *PrA* (=*PEP*4) probe [35]. We isolated the ACP from a strain of *C. albicans* (no. 74 in our collection) using the same protocol as that described here for the *C. tropicalis* enzyme. The 10 aa residues of the N-terminus of the *C. albicans* enzyme were determined to be Glu-Ala-Val-Pro-Val-Thr-Leu-Ile-Asn-Glu. This sequence differed by 3 aa from the homologous sequence in *C. tropicalis* ACP, and was not found in the reported aa sequence of the *C. albicans* protease gene. We conclude that the gene previously isolated does not code for the *C. albicans* ACP.

Acknowledgements: We thank Dr J. Tschopp for sequencing the ACP from C. albicans, Dr L. Applegate and Dr H. Pooley for assistance with the English, R. Sahli and D. Pittet for helpful discussions, and J. Chevalley for secretarial assistance.

REFERENCES

- [1] Rüchel, R., Uhlemann, K. and Böning, B. (1983) Zentralblatt für Bakteriologie Mikrobiologie und Hygiene I. Abt. Orig. A 255, 537-548.
- [2] Rüchel, R., Böning, B. and Borg, M. (1986) Infect. Immun. 53, 411–419.
- [3] MacDonald, F. and Odds, F.C. (1983) J. Gen. Microbiol. 129, 431-438.
- [4] Kwon-Chung, K.J., Lehman, D., Good, C. and Magee, P.T. (1985) Infect. Immun. 49, 571-575.
- [5] Borg, M., Kirk, D., Baumgarten, H. and Rüchel, R. (1984) Sabouraudia 22, 357-367.
- [6] Borg, M. and Rüchel, R. (1990) J. Med. Vet. Mycol. 28, 3-14.

- [7] Bistoni, F., Vecchiarelli, A., Cenci, E., Sbaraglia, G., Perito, S. and Cassone, A. (1984) Sabouraudia 22, 409-418.
- [8] Komshian, S.V., Uwaydah, A.K., Sobel, J.D. and Crane, L.R. (1989) Rev. Infect. Dis. 11, 379-390.
- [9] Horn, R., Wong, B., Kiehn, T.E. and Armstrong, D. (1985) Rev. Infect. Dis. 7, 646-655.
- [10] Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) J. Biol. Chem. 193, 265-275.
- [11] Laemmli, U.K. (1970) Nature 227, 680-685.
- [12] Zacharius, R.M. and Zell, T.E. (1985) Anal. Biochem. 30, 148-152.
- [13] Schägger, H. and Von Jagow, G. (1987) Anal. Biochem. 166, 368-379.
- [14] Seghezzi, W., Sanglard, D. and Fiechter, A. (1991) (submitted).
- [15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory Press (2nd edn.), Cold Spring Harbor, NY.
- [16] Chambers, S.P., Prior, S.E., Barstow, D.A. and Milton, N.P. (1988) Gene 68, 139-149.
- [17] Sutter, T., Sanglard, D. and Loper, J.C. (1990) J. Biol. Chem. 265, 16428-16436.
- [18] Sanglard, D. and Loper, J.C. (1989) Gene 76, 121-136.
- [19] Rholam, M., Nicolas, P. and Cohen, P. (1986) FEBS Lett. 207, 1-6.
- [20] Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J. (1984) Cell 37, 1075-1089.
- [21] Davidow, L.S., O'Donnell, M.M., Kaczmarek, F.S., Pereira, D.A., DeZeeuw, J.R. and Franke, A.E. (1987) J. Bacteriol. 169, 4621-4629.
- [22] Rendueles, P.S. and Wolf, D.H. (1988) FEMS Microbiol. Rev. 5, 17-46.
- [23] Von Heijne, G. (1986) Nucleic Acids Res. 14, 4683-4690.
- [24] Strayer Leach, B., Collawn, J.F. and Fish, W.W. (1980) Biochemistry 19, 5734-5741.
- [25] Corden, J., Wasylyk, B., Buckwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
- [26] Hamilton, R., Wanatabe, C.K. and De Boer, H.A. (1987) Nucleic Acids Res. 15, 3581–3593.
- [27] Bennetzen, J.L. and Hall, D.H. (1982) J. Biol. Chem. 257, 3018-3031.
- [28] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435-1441.
- [29] Hirata, D., Fukui, S. and Yamashita, I. (1988) Agric. Biol. Chem. 52, 2647-2649.
- [30] Delaney, R., Wong, R.N.S., Meng, G., Wu, N. and Tang, J.
- (1987) J. Biol. Chem. 262, 1461–1467. [31] Horiuchi, H., Yanai, K., Okazaki, T., Takagi, M. and Yano, K.
- (1988) J. Bacteriol. 170, 272-278.
 [32] Sogawa, K., Fujii-Kuryiyama, Y., Mizukamu, Y., Ichihara, Y. and Takahashi, K. (1983) J. Biol. Chem. 25, 5306-5311.
- [33] Ammerer, G., Hunter, C.P., Rothman, J.H., Saari, G.C., Valls, L.A. and Stevens, T.H. (1986) Mol. Cell. Biol. 6, 2490-2499.
- [34] MacKay, V.L., Welch, S.K., Insley, M.Y., Manney, T.R., Holly, J., Saari, G.C. and Parker, M.L. (1988) Proc. Natl. Acad. Sci. USA 85, 55-59.
- [35] Lott, T.J., Page, L.S., Boiron, P., Benson, J. and Reiss, E. (1989) Nucleic Acids Res. 17, 1779.
- [36] James, M.N.G. and Sielecki, A.R. (1983) J. Mol. Biol. 163, 299-361.