Volume 224, number 1, 65–70 FEB 05271 November 1987

Primary structures of three highly acidic ribosomal proteins S6, S12 and S15 from the archaebacterium *Halobacterium marismortui*

Junko Kimura, Evelyn Arndt and Makoto Kimura

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 19 August 1987

The amino acid sequences of three extremely acidic ribosomal proteins, S6, S12, and S15, from *Halobacterium marismortui* have been determined. The sequences were obtained by the sequence analysis of peptides derived by enzymatic digestion with trypsin, *Stapylococcus aureus* protease and chymotrypsin, as well as by cleavage with dilute HCl. The proteins, S6, S12 and S15, consist of 116, 147 and 102 amino acid residues, and have molecular masses of 12251, 16440 and 11747 Da, respectively. Comparison of the amino acid sequences of these proteins with ribosomal protein sequences of other organisms revealed that halobacterial protein S12 has homology with the eukaryotic protein S16A from *Saccharomyces cerevisiae*, while S15 is significantly related to the *Xenopus laevis* S19 protein. No homology was found between these halobacterial proteins and any eubacterial ribosomal proteins.

Amino acid sequence; Ribosomal protein; Sequence comparison; (*Halobacterium marismortui*)

1. INTRODUCTION

Halobacteria, as the name indicates, are organisms living in a very high salt environment. They are phylogenetically grouped into the proposed third kingdom, archaebacteria, because of some of their biochemical traits [1]. The halobacterial ribosomes possess several characteristics which distinguish them from those of other organisms. Firstly, they require high salt concentrations to maintain stable 70 S, 50 S and 30 S particles [2] and to produce full functions [3]. When ribosomes are suspended in low-ionic strength buffers, most of the proteins are released from the ribosomal RNA (rRNA) core. Secondly, it is well known that ribosomal proteins from nonhalophiles are mostly basic and that this character of ribosmal proteins probably plays an important role in the interaction with rRNAs. In contrast, most, if not all, of the halophilic ribosomal proteins are acidic, having average isoelectric points of 3.9 [2]. These findings suggest that the structures of halophilic ribosomal proteins and the mechanisms by which the proteins interact with rRNAs have undergone drastic changes in adapting to extreme halophilic environments. Thirdly, halobacterial ribosomes share both the eukaryotic and eubacterial features: the amino acid sequences of proteins S14 and S16 from *Halobacterium marismortui* 30 S ribosome subunits [4], of the A-protein from *H. cutirubrum* ribosomes [5], and of a number of 5 S rRNAs from halophilic ribosomes [6] show more relatedness to the eukaryotic counterparts than to the eubacterial ones. On the other hand, the primary and secondary structures of the 16 S rRNA from halobacterial ribosomes and their gene organization appear to be more similar to the eubacterial

Correspondence address: J. Kimura, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), Germany.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

counterparts than to eukaryotic ones [7–9]. The halophile ribosomes, therefore, seem particularly interesting for studies, not only on the molecular basis of the adaptation to an extreme halophilic environment, but also should give insight into the molecular evolution of the ribosomal proteins.

In recent analyses, we have determined the complete amino acid sequences of four ribosomal proteins: S11, S14, S16 and S17, from 30 S subunits of *H. marismortui* [4,10,11]. These results allowed us to identify unambiguously the corresponding proteins from either eubacterial and/or eukaryotic ribosomes, and also to evaluate the evolutionary changes in amino acid sequences between halophilic and nonhalophilic ribosomal proteins. Here, we have extended this information through the sequence determination of three extremely acidic ribosomal proteins, S6, S12 and S15 from *H. marismortui*, and compared their sequences with those of ribosomal proteins of other organisms.

2. MATERIALS AND METHODS

2.1. *Purification of ribosomal proteins*

Ribosomal proteins S6, S12 and S15 of *H. marismortui* were extracted from 30 S subunits and then separated on DEAE-cellulose column chromatography as described in [10].

2.2. *Sequence determination*

The proteins were cleaved enzymatically with trypsin, chymotrypsin, and *Staphylococcus aureus* protease, as in [12], and some peptides were cleaved at aspartyl residues with dilute acid (220 μl concentrated HCl was diluted to 100 ml with deionized and Millipore filtered water, pH 2 + 0.04) using the procedure of Inglis [13]. The resulting peptides were separated by a fingerprinting method [12] or on reverse-phase HPLC with a Vydac C18 column using an acetonitrile gradient in 0.1% trifluoroacetic acid, and effluents were monitored by absorption at 220 nm. Amino acid analyses were performed on an HPLC system using *o*-phthaldialdehyde as a derivatized reagent [14]. Sequence determination was carried out by the DABITC/PITC double-coupling method [15]. The amino-terminal sequences and some long peptides were sequenced with a solid-phase sequencer (LKB 4020) as described in [12].

2.3. *Computer analysis*

The amino acid sequences of proteins S6, S12 and S15 were compared to all ribosomal proteins included in the NBRF Protein Sequences Database (release 11, December, 1986) and also to the amino acid sequences of ribosomal proteins in our own file, with program RELATE and ALLIGN [16] on a VAX/VMX computer. Hydropathic analyses were performed with the program PEPPLOT [17] included in UWGCG (University of Wisconsin Genetics Computer Group, Version 4.1, August 1986).

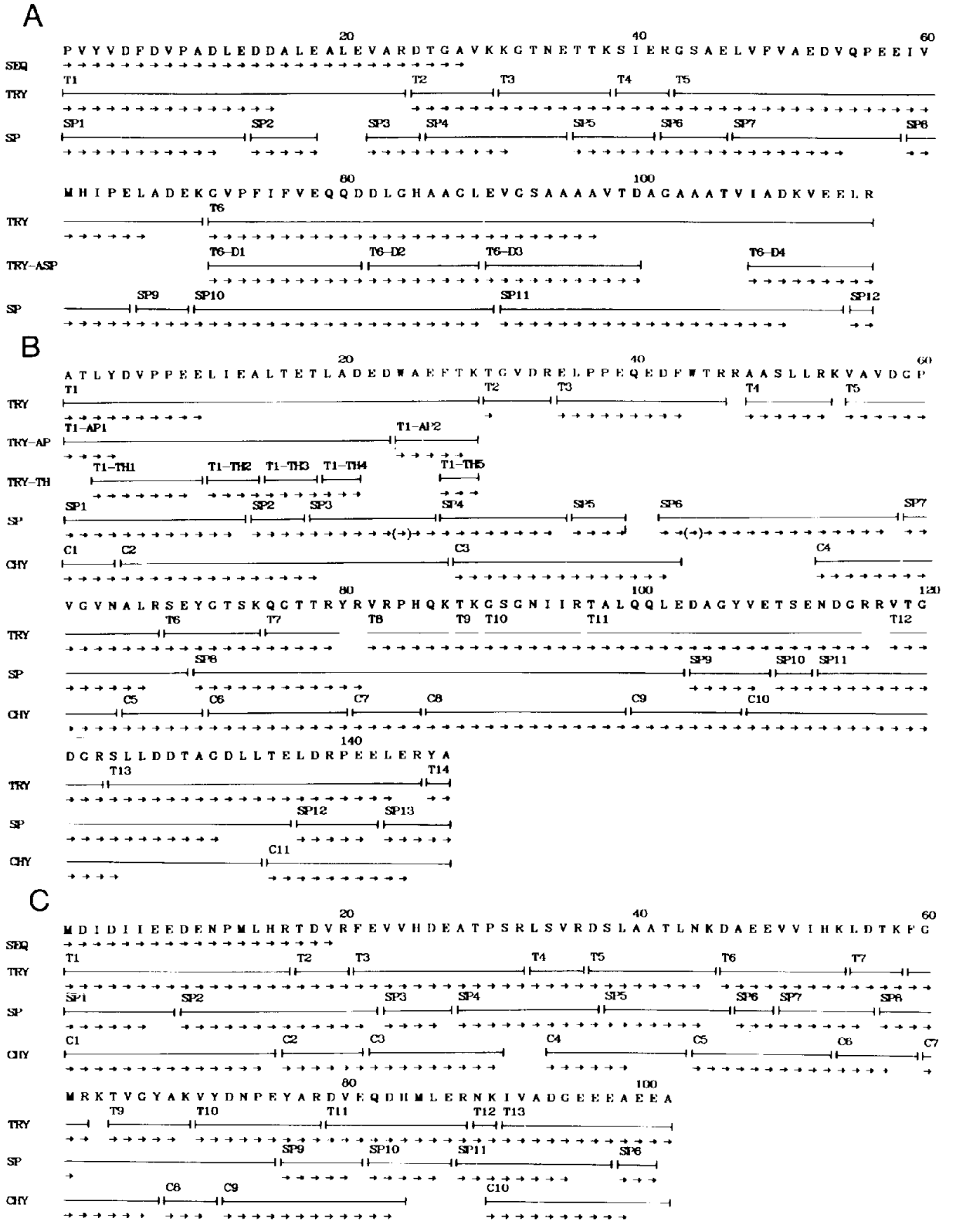
3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The complete amino acid sequences of ribosomal proteins S6, S12 and S15 from *H. marismortui* are presented in fig.1. The sequences have been derived mainly by sequence analysis of tryptic peptides, which were aligned by sequencing overlapping peptides from *S. aureus* protease digestion of the proteins. Other sequence information was obtained from chymotryptic peptides of the proteins, and from thermolytic and acid cleavage of the peptides. The resultant sequence information established the complete primary structures of the proteins, as shown in fig.1.

Problems in these analyses were encountered at the amino-terminal region of the S12 protein due to the occurrence of consecutive proline residues at positions 7 and 8, which tend to cause incomplete Edman degradations. Further, extended sequence determination of the carboxyl-terminal region of S6, through the tryptic peptide T6 and *S. aureus* protease peptide SP11, proved to be difficult due

Fig.1. Amino acid sequences of proteins S6 (A), S12 (B) and S15 (C) from *H. marismortui*. SEQ indicates degradations of intact protein. TRY, SP and CHY designate peptides derived from digestions of proteins with trypsin, *S. aureus* protease and chymotrypsin, respectively. TRY-AP and TRY-TH indicate peptides obtained by acid cleavage and thermolytic digestion of tryptic peptides, respectively. Arrows (→) indicate the cycles of the DABITC/PITC double-coupling method or solid-phase Edman degradations.


A
B
C
SEQ
TRY
SP
TRY-ASP
TRY-AP
TRY-TH
CHY

to the large size of the peptides. The sequences were finally determined with smaller subdigested peptides produced by acid cleavage and thermolytic digestion of peptide T1 in S12, and by acid cleavage of peptide T6 in S6, as shown in fig.1. In addition, the Trp-44 residue in protein S12 was identified by the optical absorbance of tryptic peptide T3, although the Edman degradation of the *S. aureus* protease peptide SP4 could yield no amino acid at residue 44 in the S12 protein.

The amino acid compositions of the three proteins and their molecular masses calculated from the primary structures are given in table 1. The amino acid compositions derived from the sequence data are in good agreement with those from the amino acid analyses of the hydrolysed proteins (table 1).

The three proteins contain a large number of acidic amino acids randomly distributed throughout the molecules. From the number of acidic and basic residues in the three proteins, net charges of -18, -10.5 and -12 are calculated for S6, S12 and S15, respectively, assuming that the His residue carries a charge of $+0.5$ [18].

Examination of the amino acid sequences for hydropathy revealed a general hydrophilic character of the proteins. In particular, the amino-terminal region (positions 20–45) in S6, the central 20 residues (positions 70–90) in S12 and the carboxyl-terminal 30 amino acid residues in S15 are predominantly hydrophilic. Presumably, these hydrophilic regions might be involved in structural stabilization of the protein in the presence of the high molarity of the cytoplasmic salt.

3.2. *Sequence comparison*

The complete amino acid sequences of the three proteins studied in this paper were compared with other ribosomal protein sequences using the Dayhoff programs, as described in section 2. First, the comparison of the three proteins with eubacterial ribosomal proteins, including all 52 *E.*

Table 1

Amino acid compositions and molecular masses of proteins S6, S12 and S15 from *H. marismortui*

Amino acid	Protein					
	S6		S12		S15	
	Sequence	Hydrolysate	Sequence	Hydrolysate	Sequence	Hydrolysate
Asp	12	18.0	13	17.0	12	16.0
Asn	1		3		4	
Thr	7	5.3	15	12.9	5	5.1
Ser	3	2.9	6	5.9	3	2.9
Glu	15	22.0	17	20.8	15	17.5
Gln	3		5		1	
Pro	5	n.d.	7	n.d.	3	n.d.
Gly	8	6.4	12	12.7	3	3.0
Ala	18	20.0	12	11.5	9	10.6
Val	15	14.9	9	7.3	10	10.4
Met	1	0.9	0	0	4	2.9
Ile	5	4.7	3	2.3	5	5.4
Leu	8	8.0	16	14.0	6	6.0
Tyr	1	1.0	5	4.6	3	3.7
Phe	4	5.1	2	2.7	2	2.6
His	2	1.9	1	1.0	4	2.9
Lys	5	3.4	5	5.0	6	4.8
Arg	3	3.1	14	14.2	7	5.3
Trp	0	n.d.		n.d.	0	n.d.
Total	116		147		102	
M_r	12 251		16 440		11 747	

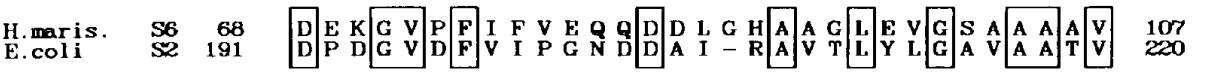


Fig.2. Sequence homology of regions in the *H. marismortui* protein S6 (positions 68–107) and the *E. coli* protein S2 (positions 191–220). The sequence stretches were aligned with the program ALIGN using the mutation data matrix with a break penalty of 10.

coli proteins, was performed with the program RELATE (segment length 20). The segment comparison score thus obtained suggested a possible relatedness for only one protein pair, namely S6 from *H. marismortui* and S2 from *E. coli* ribosomes (at the level of 3.579 SD units). However, in this case, further analysis using the program ALIGN revealed that the homology was restricted to only one part, giving the alignment score of 2.11, as presented in fig.2, and that long insertions and deletions are required to align two protein chains for a maximum homology. Therefore, the relatedness of the two proteins obtained by the program RELATE seems to be somewhat uncertain.

Further sequence homology search with the segment length of 10 residues did not reveal a significant relation to any eubacterial ribosomal protein sequences. Hence, the three proteins cannot be correlated to any eubacterial ribosomal proteins. On the other hand, in previous studies, four ribosomal proteins, S11, S14, S16 and S17, from *H. marismortui* were found to be homologous to eubacterial ribosomal proteins [4,10,11]. This finding suggests that the amino acid residues of the highly acidic proteins described here may have been drastically substituted during evolution.

The sequences of the three proteins S6, S12 and S15 were also compared with those of eukaryotic ribosomal proteins. This comparison demonstrated that S12 from *H. marismortui* exhibits a significant homology to the yeast *S. cerevisiae* ribosomal protein S16A [19] at the level of 12.086 SD units, and that *H. marismortui* S15 appears to

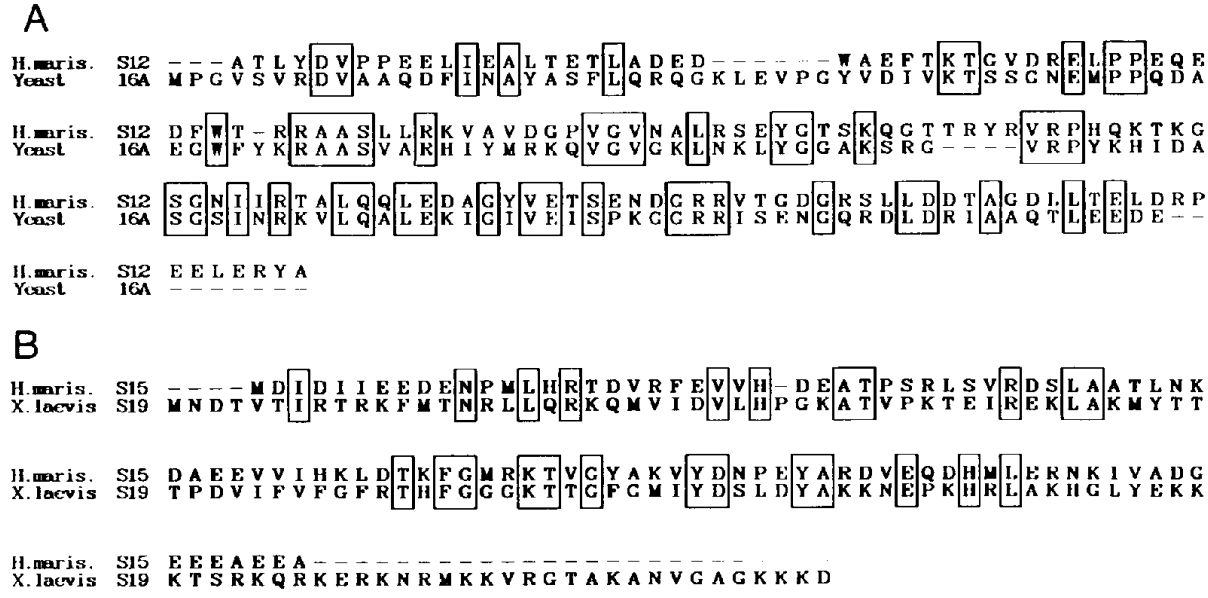


Fig.3. Comparison of the primary structures of S12 and S15 from *H. marismortui* ribosomes with ribosomal proteins from eukaryotes. (A) Alignment of ribosomal protein S12 from *H. marismortui* (HS12) with protein S16A from *S. cerevisiae* (YS16A) [19]. (B) Alignment of ribosomal protein S15 from *H. marismortui* (HS15) with protein S19 from *X. laevis* (XS19) [20]. The sequences were aligned with the program ALIGN using the mutation data matrix with a break penalty of 20.

be related to the *Xenopus laevis* ribosomal protein S19 [20], giving 9.124 SD units. Accordingly, *H. marismortui* proteins S12 (HS12) and S15 (HS15) were aligned with yeast protein S16A (YS16A) and *X. laevis* S19 (XS19) for a maximum homology by the program ALIGN, as given in fig.3. This comparison shows that HS12 and HS15 can easily be aligned to proteins YS16A and XS19, respectively, with a few insertions and deletions. Protein HS12 has 35% identical residues (ALIGN score: 23.26) with YS16A, and HS15 shares 24% identical residues with XS19 (ALIGN score: 12.88). This is strong statistical evidence that HS12 and HS15 are related to YS16A and XS19, respectively, since two proteins which give an ALIGN score of more than 3.0 SD units are considered to be strongly related to each other [16].

There is no indication from our computer search programs that protein S6 is homologous to any of the known eukaryotic or eubacterial ribosomal proteins. However, it is possible that this negative finding is due to insufficient sequence data for eukaryotic ribosomal proteins, and that protein S6 might be related to a eukaryotic protein whose sequence has not yet been sequenced.

Although it is still premature at present for any definite conclusions concerning the evolutional relationship of halophilic ribosomal proteins, the results from both the present and previous analyses [4,5] indicate that the primary structures of halobacterial ribosomal proteins appear to be more related to eukaryotic than to eubacterial proteins. On the other hand, it is well known that the size of the chromosomal DNA and the gene organization in halobacteria are of the eubacterial type [7–9]. This observation suggests that a divergence of the ribosomal proteins into the eubacterial, halobacterial (archaebacteria) and eukaryotic type had taken place before divergence of the 'urkaryotes' occurred, possibly by drastic genetic events, e.g. by gene duplications. Further comparative sequence studies on ribosomal proteins, especially from archaebacteria, will provide more valuable information on the molecular evolution of ribosomes.

ACKNOWLEDGEMENTS

We thank Dr H.G. Wittmann for his interest and encouragement, and Dr J. McDougall for critical reading of the manuscript.

REFERENCES

[1] Woese, C.R. and Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088–5090.
[2] Strom, A.R., Hasnain, S., Smith, N., Matheson, A.T. and Visentin, L.P. (1975) Biochim. Biophys. Acta 383, 325–337.
[3] Shevack, A., Gewitz, H.S., Hennemann, B., Yonath, A. and Wittmann, H.G. (1985) FEBS Lett. 184, 68–71.
[4] Kimura, J. and Kimura, M. (1987) J. Biol. Chem., in press.
[5] Matheson, A.T. and Yaguchi, M. (1982) Zbl. Bakt. Hyg., I. Abt. Orig. C3, 192–199.
[6] Hori, H. and Osawa, S. (1980) in: Genetics and Evolution of RNA Polymerase, t-RNA and Ribosomes (Osawa, S. et al. eds) pp. 539–551, University of Tokyo Press, Tokyo.
[7] Gupta, R., Lanter, J.M. and Woese, C.R. (1983) Science 221, 656–659.
[8] Hui, I. and Dennis, P.P. (1985) J. Biol. Chem. 260, 899–906.
[9] Mankin, A.S., Kagramanova, V.K., Teterina, N., Rubtsov, P.M., Belova, E.N., Kopylov, A.M., Baratova, L.A. and Bogdanov, A.A. (1985) Gene 37, 181–189.
[10] Arndt, E., Breithaupt, G. and Kimura, M. (1986) FEBS Lett. 194, 227–234.
[11] Kimura, M. and Langner, G. (1984) FEBS Lett. 175, 213–218.
[12] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide Sequence Analysis (Birr, C. ed.) pp. 49–72, Elsevier, Amsterdam, New York.
[13] Inglis, A.S. (1983) Methods Enzymol. 91, 324–332.
[14] Ashman, K. and Bosserhof, A. (1985) in: Modern Methods in Analytical Protein Chemistry (Tschesche, H. ed.) pp. 15–171, De Gruyter, Berlin.
[15] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[16] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol. 5, suppl. 3, National Biomedical Research Foundation, Washington, DC.
[17] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[18] Offord, R.E. (1966) Nature 211, 591–593.
[19] Molenaar, C.M.T., Woudt, L.P., Jansen, A.E.M., Mager, W.H. and Planta, R.J. (1984) Nucleic Acids Res. 14, 7345–7358.
[20] Amaldi, F., Beccari, E., Bozzoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) Gene 17, 311–316.