Volume 267, number 2, 193-198 FEBS 08588 July 1990

Nucleotide sequence of four genes encoding ribosomal proteins from the 'S10 and spectinomycin' operon equivalent region in the archaebacterium *Halobacterium marismortui*

Evelyn Arndt

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, 1000 Berlin 33 (Dahlem), FRG

Received 8 May 1990

Four genes encoding ribosomal proteins HmaS17, HmaL14, HmaL24 and HS3, have been identified in the λEMBL3 clone PP*7 from a genomic library of the archaebacterium *Halobacterium marismortui*. The clone contains genes from the 'S10 and spectinomycin' operon equivalent region. Three of the deduced proteins are homologous to the corresponding *Escherichia coli* and *Methancoccus vannielii* S17, L14 and L24 proteins, as well as to eukaryotic proteins from rat or yeast. HS3 was identified as an extra protein corresponding to the gene product for orfc in *M. vannielii* and the eukaryotic ribosomal protein RS4 from rat. The equivalence of HmaL24 (HL16) and *E. coli* L24, which share only 28% identical amino acid residues, could now be shown by localizing the HmaL24 gene at the same position in the cluster.

Ribosomal protein; Nucleotide sequence; Gene organization; Archaebacteria; Evolution

1. INTRODUCTION

Ribosomes are essential cell components, present in all organisms. Therefore, the constituents of the ribosome, the diverse rRNAs and proteins, form an excellent tool for studying molecular evolution with respect to the conservation of functionally important structures.

Comparisons of amino acid sequences of ribosomal proteins from all 3 kingdoms, the eubacteria, eukaryotes and archaebacteria, have been carried out [1-4]. These studies become especially valuable if full sets of ribosomal proteins are available also for representative organisms derived from archaebacteria and eukaryotes. For this purpose, the determination of ribosomal protein sequences of the halophilic archaebacterium *Halobacterium marismortui* have been extended, either by using protein-chemical methods to determine the amino acid sequence, or by cloning and sequencing of the appropriate genes [5,6]. In general, the halobacterial ribosomal proteins can be subdivided into 4 different groups: proteins which are sequence-related to eubacterial, as well as to eukaryotic counterparts, proteins which have homologues in either eubacteria or eukaryotes, and proteins for which no counterparts were found at all [3]. For most of the proteins which can be compared to eubacterial and eukaryotic homologues, the degree of homology (alignment score) is higher to the eukaryotic ones [2,5,6].

Correspondence address: E. Arndt, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, 1000 Berlin 33 (Dahlem), FRG

Recently, studies on the gene organization in archaebacteria [7-11] showed a similar ribosomal protein gene order to that in eubacteria, except for some deletions or insertions of additional protein genes. Some of these extra proteins show significant homologies to eukaryotic ribosomal proteins. The length and conservation of amino acid sequences typical of eukaryotic counterparts on the one hand, and a gene organization such as in eubacteria on the other, substantiate the hypothesis that archaebacteria represent an intermediate evolutionary stage between eubacteria and eukaryotes.

In this paper, the gene organization, the nucleotide sequences and the deduced amino acid sequences of four ribosomal proteins (HmaS17, HmaL14, HmaL24 and HS3) of the linked S10/spc operon equivalent region in *H. marismortui* will be presented and compared to homologous structures in eubacteria, eukaryotes and archaebacteria.

2. MATERIALS AND METHODS

2.1. *Materials*

The λEMBL3 clone containing a ~15 kb insert DNA was isolated from a genomic library of *H. marismortui* as described [6]. Sequencing vectors, enzymes and other materials were used as in [6].

2.2. *DNA sequence analysis*

The nucleotide sequence of the *H. marismortui* DNA fragment described here was determined by subcloning appropriate restriction fragments which were sequenced in both directions using the dideoxy chain termination method [12].

2.3. *Computer analysis*

The deduced amino acid sequences were compared with proteins in

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

the National Biomedical Research Foundation (NBRF, release 23, 1990) protein data base and the RIBO-data base available in our institute using the computer programs from the University of Wisconsin Genetic Computer Group (Version 6.1, August 1989) [13] on a VAX/VMS computer. The program ALIGN was used to obtain maximal homology between related proteins [14]. The mutation data matrix was applied as scoring matrix, and a break penalty of 20 was employed.

2.4. *Nomenclature*

In the case where ribosomal proteins from *H. marismortui* are significantly homologous to those of *Escherichia coli* they are denoted by the prefix 'Hma', e.g. HmaS17. If no equivalent sequence was detected to the eubacterial kingdom they are designated by the prefix 'H', e.g. HS3, according to their migration in the two-dimensional gel electrophoresis systems [15,16]. The deduced amino acid sequences of the open reading frames of *H. marismortui* and *M. vannielii* which have no equivalent proteins in eubacteria and were not detected in the two-dimensional gel system were named HMprot3, MVprotB and MVprotC.

3. RESULTS AND DISCUSSION

3.1. *DNA sequence*

The screening of the λEMBL3 clone PP*7 from the genomic library and the nucleotide sequence to the SalI site in position 5638 (encoding the ribosomal proteins HmaL3, HL6, HmaL23, HmaL2, HmaS19, HmaL22, HmaS3, HmaL29; the putative gene products from orf1, orf2 and a part of orf3), have been described in [6].

Fig. 1 shows the nucleotide sequence (2118 bp) containing the end of the HmaL29 gene (up to position 5636), the complete orf3 (positions 5639–5920), as well as the 3′-joining region up to position 7718. The genes for the ribosomal proteins HmaS17 (positions 5914–6249), HmaL14 (positions 6252–6647), HmaL24 (positions 6655–7014) and HS3 (positions 7014–7715), which are described in this work, are located in the latter.

As has been observed for some other ribosomal protein genes of *H. marismortui*, the genes for orf3/HmaS17, HmaS17/HmaL14 and HmaL24/HS3 are overlapping. Only for the HmaL14 and the HmaL24 genes an intergenic region between their stop and start codon is present, 4 base pairs in length. There is, however, no space for regulatory sequences such as promoter or terminator structures between the genes, so that this gene cluster is probably regulated by the promoter located before orf1 [6] (P in Fig. 6).

3.2. *Comparisons of the deduced proteins with those from other organisms*

The amino acid composition and molecular masses of the translated proteins HMprot3, HmaS17, HmaL14, HmaL24 and HS3 are summarized in Table I. In the following, comparisons of these proteins with counterparts of archaebacteria, eubacteria and eukaryotes will be presented.

3.2.1. HMprot3

No homologous protein from eubacteria or eukaryotes has so far been found for the putative gene product encoded by orf3 but it shows significant homology to the protein MVprotB (encoded by orfb) in the corresponding gene cluster of *M. vannielii* [17]. An alignment of the *H. marismortui* HMprot3 and the *M. vannielii* MVprotB is presented in Fig. 2. The putative proteins share 36% identical amino acids (alignment score 14.8 S.D. units). Their occurrence in both archaebacterial species is an indication that they might be translated proteins whose function is still not known.

3.2.2. HmaS17

The amino acid sequence of the purified protein HmaS17 (HS14) has been determined previously by protein-chemical methods [18]. In that work, relationships between the proteins EcoS17 from *E. coli* and RS11 from rat have been presented. Subsequent alignment of HmaS17 and MvaS17 which is now available from the archaebacterium *M. vannielii* [17] results in 44% identical amino acid residues and an alignment score of 17.0 SD units, respectively.

The deduced amino acid sequence from the nucleotide sequence determined in this work differs from the protein sequence as follows: in position 88, cysteine was found instead of serine and in position 92, serine instead of proline. In addition, in the C-terminal region a glycine and an aspartic acid have to be introduced in the sequence after the aspartic acid in position 108, so that the total number of amino acid residues in HmaS17 is 111.

3.2.3. HmaL14

The N-terminus of this protein deduced from the translation of the DNA sequence corresponds to the N-terminal sequence of the purified protein HL27 (T. Hatakeyama, personal communication). The complete sequence of the protein could be obtained from the DNA translation. The proteins MvaL24 from *M. vannielii* [17], YL17 from yeast [19] and EcoL14 from *E. coli* [20] are the counterparts of HmaL14 (HL27) sharing 52%, 44% and 37% identical amino acid residues and alignment scores of 52.2, 28.9 and 18.3 SD units, respectively. The alignment of the L14 ribosomal protein family is shown in Fig. 3. The high conservation might be an indication for the functionally important role of this protein in the ribosome.

3.2.3. HmaL24

The HmaL14 gene is followed by an open reading frame whose gene product was identified to code for HL16, a protein which has been sequenced by protein-chemical methods [21]. In that work, any homology between HL16 and EcoL24 was considered negligible because of the relatively low number of identical amino acids and the low alignment score (28%, 2.8 SD units).

Fig. 1. Nucleotide sequence of a DNA fragment encoding ribosomal protein genes from the S10/spectinomycin operon region of *H. marismortui*. The deduced amino acid sequences are written below the DNA strand. Orf3 corresponds to positions 5639–5920, HmaS17 to positions 5914–6249, HmaL14 to positions 6252–6647, HmaL24 to positions 6655–7014 and HS3 to positions 7014–7715. The initiator methionine is not present in the mature protein in the case of HmaS17 [18], HmaL24 [21], and HS3 (our observation).

After the location of its gene in the cluster, it became probable that HL16 is the homologue of EcoL24. Since protein MvaL24 of *M. vannielii*, whose gene is placed in the same position in the equivalent gene cluster [17], can be easy aligned to HL16, it was clear that HL16 belongs to the L24 family and was therefore named HmaL24. MvaL24 and HmaL24 share 47% identical amino acid residues and an alignment score of 28.5 SD units, in-

Table I

Amino acid composition and molecular masses of the *H. marismortui* proteins HMprot3, HmaS17, HmaL14, HmaL24 and HS3

AS	HMprot3	HmaS17	HmaL14	HmaL24	HS3
Ala	7	9	13	8	19
Cys	1	4	1	–	–
Asp	8	10	7	14	19
Glu	7	10	10	16	28
Phe	2	2	2	1	7
Gly	6	7	13	6	26
His	2	3	2	3	6
Ile	2	1	7	2	11
Lys	1	6	9	7	8
Leu	10	7	7	10	15
Met	3	2	3	–	1
Asn	2	3	4	4	10
Pro	6	6	6	4	8
Gln	1	4	4	5	7
Arg	7	7	12	13	9
Ser	5	5	6	6	10
Thr	11	7	10	6	13
Val	10	16	15	13	28
Trp	2	–	–	–	1
Tyr	–	2	1	–	7
Total	93	111	132	119	233
M_r	10179	12142	14195	13519	25119

dicating the highly significant homology between these archaebacterial proteins. Interestingly, MvaL24 and EcoL24 share a considerably high degree (37%) of identical amino acid residues and a significant alignment score (6.6 SD units), leaving no doubt about their common descent.

The ribosomal protein RL26 from rat [22] and the N-terminal fragment of yeast ribosomal protein YeaL33 [23] were found to be eukaryotic counterparts of HmaL24. Fig. 4 shows the alignment of HmaL24 with the homologous proteins, whose complete sequences are known.

The amino acid sequence for HmaL24 deduced from the DNA translation agrees well with the sequence determined by Edman-degradation [21], except for one difference at position 20, namely a histidine instead of an arginine.

3.2.4. HS3

The protein sequence translated from the gene following the HmaL24 gene was found to correlate well with the N-terminal sequence of HS3 from the 30S subunit of the *H. marismortui* ribosome (data not shown). The initiator methionine in position 1 of the

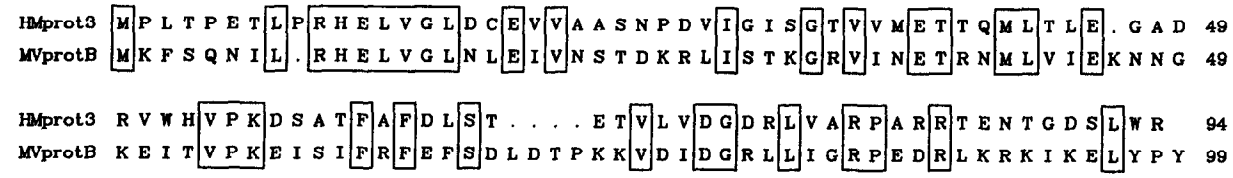

Fig. 2. Alignment of the putative amino acid sequences deduced from orf3 (HMprot3) from *H. marismortui* and orfb (MVprotB) from *M. vannielii* [17].

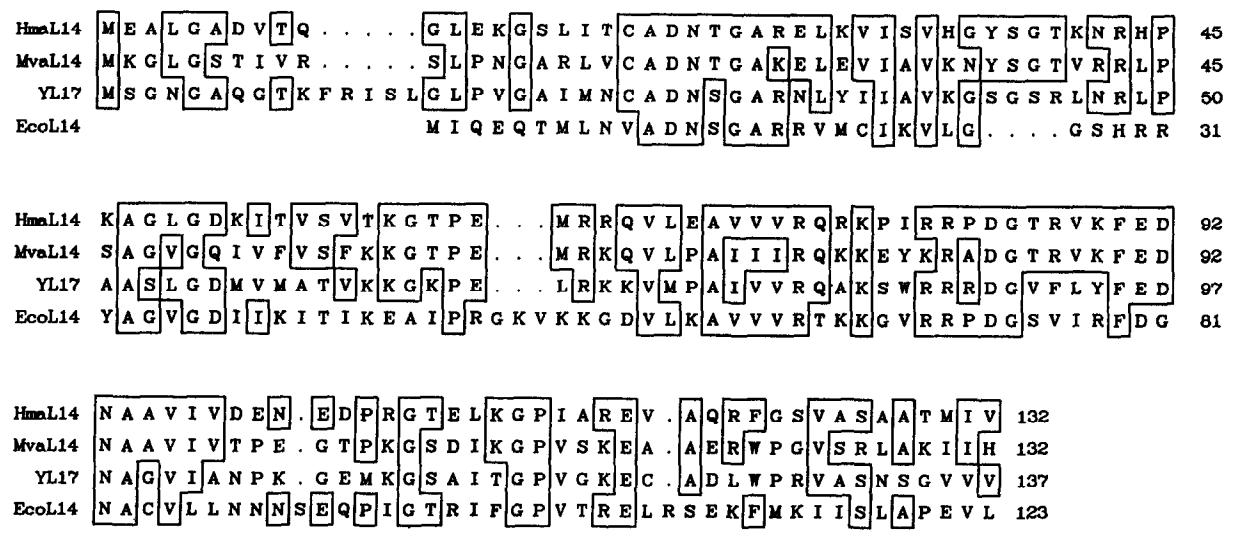

Fig. 3. Alignment of the ribosomal protein HmaL14 with the archaebacterial MvaL14 [17], the eukaryotic YL17 [19] and the eubacterial EcoL14 [20].

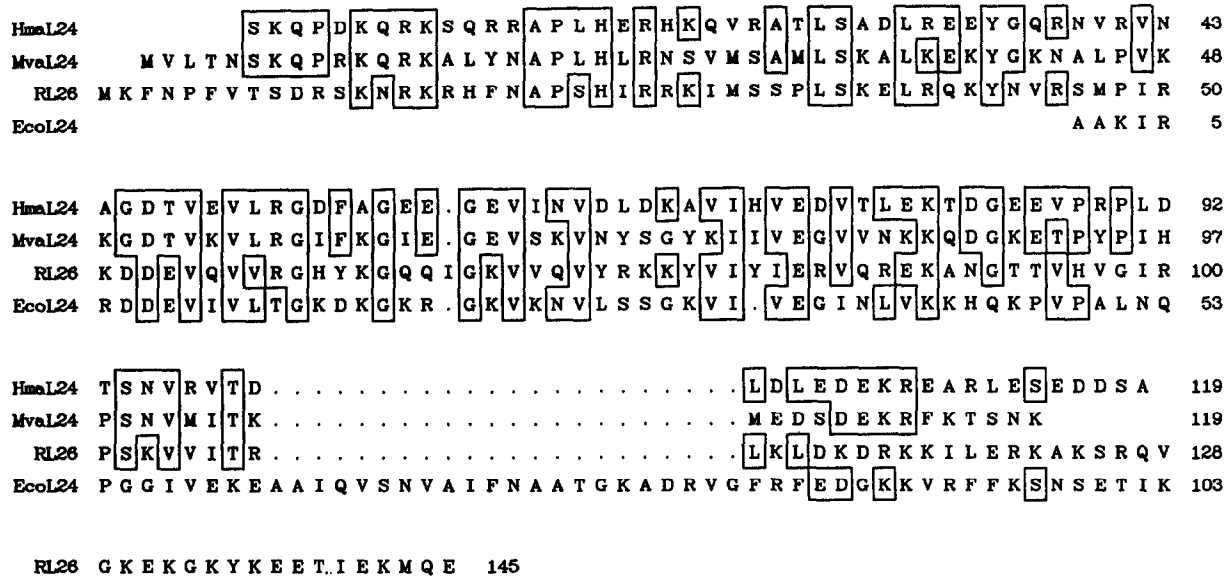

Fig. 4. Alignment of the ribosomal protein HmaL24 with its counterparts MvaL24 [17], RL26 [22] and EcoL24 [24].

deduced protein is cleaved off during or after translation because it is not present in the mature protein. Fig. 5 shows the alignment of HS3 with its corresponding archaebacterial protein deduced from orfc of *M. vannielii* and the eukaryotic protein RS4 from rat [25]. HS3 and its counterparts MVprotC and RS4 share 37% and 28% identical amino acid residues and alignment scores of 47.7 and 20.6 SD units.

Interestingly, RS4 has been reported to be located at the interface between the ribosomal subunits [26–28]

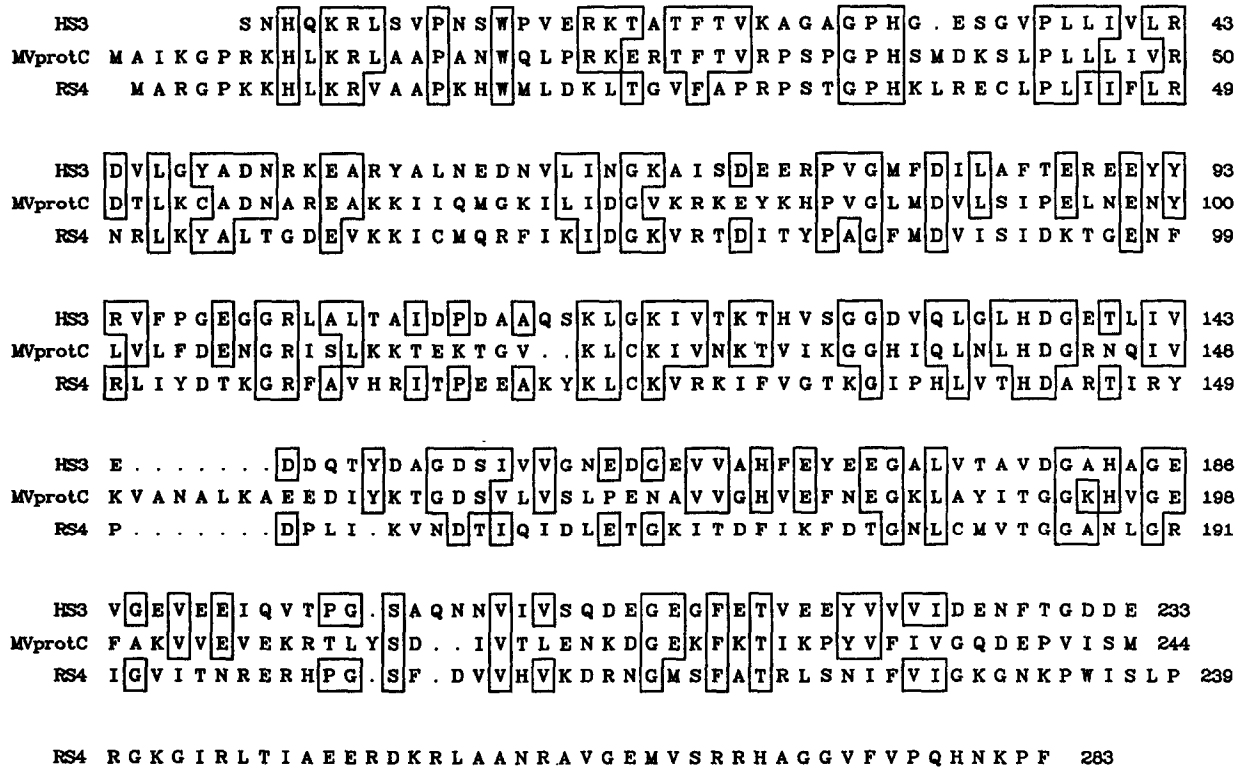

Fig. 5. Alignment of the ribosomal protein HS3 from *H. marismortui* with the homologous proteins MVprotC (deduced from orfc) from *M. vannielii* [17] and RS4 from rat [25].

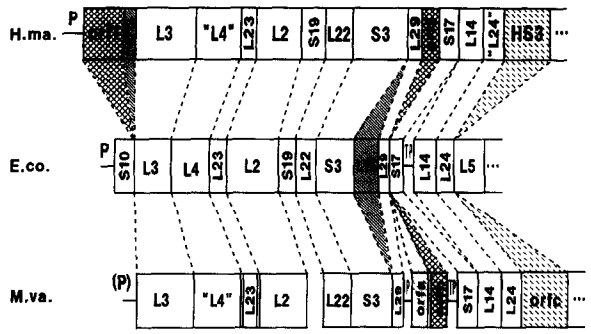

Fig. 6. Gene organization of the S10/spectinomycin operon region of *H. marismortui* in comparison to that of the eubacterium *E. coli* and the related archaebacterium *M. vannielii*.

and is thought to form part of the domain for the binding of mRNA in the initiation stage of the peptide synthesis. It is likely that HS3 has an identical function in the archaebacterial ribosome.

3.3. *The organization of the genes in the cluster*

The investigated genes of the ribosomal proteins HmaS17, HmaL14 and HmaL24 of *H. marismortui* appear in the same gene order as their equivalent genes EcoS17, EcoL14 and EcoL24 in the S10 and spc operon of *E. coli*. However, in *H. marismortui*, the S10/spc operon region seems to be fused in contrast to the two separate transcription units found in the same region in the related archaebacterium *M. vannielii* [17]. A similar situation has been described for the transcription units from the eubacteria *E. coli*, *Micrococcus luteus* and *Mycoplasma capricolum*. The separation of the S10 and spc operon in *M. luteus* and *E. coli* is different from the apparent fusion of the S10/spc operon in *M. capricolum* [29]. These results show that operon structures are not necessarily conserved and may undergo variations during evolutional processes.

Two additional genes in the S10/spc region of *H. marismortui* have no apparent equivalents in *E. coli*, namely orf3 and HS3. It could be shown that HS3 codes for a ribosomal protein from the small subunit of the ribosome. The obvious equivalence of HS3 and orfc from *M. vannielii* [17] substantiate the occurrence of additional ribosomal proteins in archaebacteria. Since orf3 from *H. marismortui* is related to orfb from *M. vannielii*, a corresponding gene for orfa from *M. vannielii* [17] is missing in *H. marismortui*. Fig. 6 shows the S10 and spc operon region from *H. marismortui* in comparison with that of *E. coli* and *M. vannielii*.

Acknowledgements: I thank Dr. H.G. Wittmann for his continuous interest and encouragement.

REFERENCES

[1] Matheson, A.T. and Yaguchi, M. (1982) Zentralbl. Bacteriol. Abt. I Orig. B Hyg. C3, 192–199.
[2] Kimura, M., Arndt, E., Hatakeyama, T., Hatakeyama, T. and Kimura, J. (1989) Can. J. Microbiol. 35, 195–199.
[3] Wittmann-Liebold, B., Köpke, A.K.E., Arndt, E., Krömer, W., Hatakeyama, T. and Wittmann, H.G. (1990) in: The Structure, Function and Evolution of Ribosomes (W.E. Hill and Warner, J.R. eds), NSF, Washington, DC, in press.
[4] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G., eds) pp. 326–361.
[5] Hatakeyama, T., Kaufmann, F., Schroeter, B. and Hatakeyama, T. (1989) Eur. J. Biochem. 185, 685–693.
[6] Arndt, E., Krömer, W. and Hatakeyama, T. (1990) J. Biol. Chem. 265, 3034–3039.
[7] Auer, J., Lechner, K. and Böck, A. (1989) Can. J. Microbiol. 35, 200–204.
[8] Shimmin, L.C., Newton, C.H., Ramirez, C., Yee, J., Downing, W.L., Louie, A., Matheson, A.T. and Dennis, P.P. (1989) Can. J. Microbiol. 35, 164–170.
[9] Spiridonova, V.A., Akhmanova, A.S., Kagramanova, V.K., Köpke, A.K.E. and Mankin, A.S. (1989) Can. J. Microbiol. 35, 153–159.
[10] Köpke, A.K.E. and Wittmann-Liebold, B. (1989) Can. J. Microbiol. 35, 11–20.
[11] Arndt, E. and Kimura, M. (1988) J. Biol. Chem. 263, 16063–16068.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[14] Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) NBR Report 820501-08710 National Biochemical Research Foundation, Washington, DC.
[15] Hatakeyama, T. and Kimura, M. (1988) Eur. J. Biochem. 172, 703–711.
[16] Arndt, E., Breithaupt, G. and Kimura, M. (1986) FEBS Lett. 194, 227–234.
[17] Auer, J., Spicker, G. and Böck, A. (1989) J. Mol. Biol. 209, 21–36.
[18] Kimura, J. and Kimura, M. (1987) J. Biol. Chem. 262, 12150–12157.
[19] Leer, R.J., van Raamsdonk-Duin, M.M.C., Hagendoorn, M.J.M., Mager, W.H. and Planta, R.J. (1984) Nucleic Acids Res., 6685–6700.
[20] Kimura, M. and Wittmann-Liebold, B. (1982) Biochem. Int. 4: 567–574.
[21] Hatakeyama, T., Hatakeyama, T. and Kimura, M. (1988) FEBS Lett. 240, 21–28.
[22] Paz, V., Olvera, J., Chan, Y.L. and Wool, I.G., FEBS Lett. 251, 89–93.
[23] Otaka, E., Higo, K. and Itoh, T. (1984) Mol. Gen. Genet. 195, 544–546.
[24] Wittmann-Liebold, B. (1979) FEBS Lett. 108, 75–80.
[25] Devi, K.R.G., Chan, Y.L. and Wool, I.G. (1989) Biochim. Biophys. Acta 1008, 258–262.
[26] Nygard, O. and Nika, H. (1982) EMBO J. 1, 357–362.
[27] Uchiumi, T., Kikuchi, M. and Ogata, K. (1986) J. Biol. Chem. 261, 9663–9667.
[28] Westermann, P. and Nygard, O. (1983) Biochim. Biophys. Acta 741, 103–108.
[29] Ohama, T., Muto, A. and Osawa, S. (1989) J. Mol. Evol. 29, 381–395.