Complete amino acid sequence of ribosomal protein S14 from *Bacillus stearothermophilus* and homology studies to other ribosomal proteins

Elke Herfurth, Ulrike Briesemeister, Brigitte Wittmann-Liebold*

Max-Delbrück-Centrum für Molekulare Medizin, Abteilung Proteinchemie, Robert-Rössle-Straße 10, 13125 Berlin, Germany

Abstract The complete amino acid sequence of protein S14 from the small subunit of *Bacillus stearothermophilus* was determined by N-terminal sequence analysis and by sequencing of overlapping peptides obtained from enzymatic digestions. Protein S14 consists of 60 amino acid residues with a molecular mass of 7148 Da. It has a high content of basic amino acids and a predicted isoelectric point of 11.46. Protein S14 contains two pairs of cysteines in the carboxyl-terminal region, presumably linked by two sulphur bridges. A comparison between protein S14 of *B. stearothermophilus* and homologous proteins from other organisms revealed highly conserved carboxyl-termini for this protein in eubacteria, archaebacteria and eukaryotes.

Key words: Ribosomal protein; Protein sequencing; Evolution; *Bacillus stearothermophilus*

1. Introduction

A comparision of the primary structures of ribosomal proteins derived from different organisms allows information to be derived about their evolutionary relationships. Such comparison will identify regions of high evolutionary conservation which may correspond to regions of similar function in related components of the different ribosomes.

All ribosomal protein and rRNA sequences are completely known for the Gram-negative eubacterium *Escherichia coli* [1,2], whereas in all other organisms of the eubacterial, archaebacterial and eukaryotic kingdom many protein sequences but no completed sets are available [3]. Some ribosomal proteins of the Gram-positive eubacterium *Bacillus stearothermophilus* show a high degree of sequence similarity (with up to 70% identities) to ribosomal proteins of *E. coli*, whereas others share fewer identities (of 20–40%) with the corresponding proteins [4].

So far, 43 *B. stearothermophilus* proteins have been sequenced [4]. Here we report the identification, characterization and the primary structure as determined by amino acid sequencing of ribosomal protein S14 from the small subunit of *B. stearothermophilus*. We discuss the degree of similarities of protein S14 from *B. stearothermophilus* to all known ribosomal S14 sequences from other organisms.

The location of the *E. coli* equivalent protein S14 has been determined by immune electron microscopy [5]. It is located in the head of the 30S subunit together with proteins S3, S9, S10, S13, and S19. Protein S14 is a part of the binding site for the 3'-end of the tRNA [6,7]. Furthermore, protein S14 could be crosslinked to S19 [8], and S21 [9] as well as to initiation factors [10].

*Corresponding author. Fax: (49) (30) 9406 3869.

2. Materials and methods

2.1. Protein Isolation

Cells of *B. stearothermophilus* strain 799 were grown, their ribosomes isolated, and subunits prepared as described [11]. Ribosomal proteins were extracted from the 30S subunits by 66% acetic acid according to [12]. The protein mixture was dialysed against 2% acetic acid containing 6 mM β-mercaptoethanol and lyophilized. Proteins were separated by reversed HPLC (RP-HPLC) on a labotatory-packed Vydac (The Separation Group, Hesperia, CA, USA) C_4 column (4 × 250 mm; 5 μm, 300 Å) using a gradient of acetonitrile in aqueous 0.1% TFA. The HPLC equipment was purchased from Knauer, Berlin. The effluent was monitored at 220 nm. Protein S14 was identified on a 15% SDS-PAGE [13] and by N-terminal sequence analysis.

2.2. Fragmentation of protein S14 and peptide separation

All fragmentations were carried out on 2–4 nmol purified protein in a buffer volume of 200 μl. Enzymatic digestions with endoproteinases Glu-C and Lys-C (both Boehringer Mannheim) were performed for 20 h at 37°C with an enzyme/subtrate ratio of 1 : 20 in 50 mM ammonium acetate (pH 4.1) and in 50 mM NH_4CO_3 (pH 7.8) buffer, respectively. Fragmentation with α-chymotrypsin (Boehringer Mannheim) was done with an enzyme/substrate ratio 1 : 50 for 4 h at 37°C in 100 mM *N*-methylmorpholine acetate buffer (pH 8.1). Digestions by enzymes were stopped through immediate injection of the peptide mixtures into the HPLC system. All peptide separations were performed by RP-HPLC on a Vydac C18 column (4 × 250 mm; 5 μm, 300 Å) as described above.

2.3. Amino acid analysis

For amino acid analysis protein S14 was hydrolysed in constant boiling HCl (Sigma, St. Louis, MO) for 22 h at 110°C. The amino acids were analysed automatically with an amino acid analyser (Waters, USA) using *o*-phthaldialdehyde precolumn derivatization as described previously [14]. Oxydation of cysteines was performed by performic acid oxydation at −20°C for 4 h [15] and the hydrolysis and analysis was done as above.

2.4. Protein sequence analysis

Automated N-terminal sequence analysis of the entire protein BstS14 and of the purified peptides was performed on the modular protein sequenator, model 810, from Knauer (Berlin, Germany) using the sequencing protocols for the attachment via the C-terminal carboxyl group [16]. Pyridylethylation was performed by reducing 2–5 nmol of the entire protein in 200 μl guanidine hydrochloride in 0.1 M Tris-HCl (pH 8.5) in the presence of 5% β-mercaptoethanol at 50°C for 3 h. Subsequently cysteines were alkylated at RT by the addition of 15 μl vinylpyridine (Fluka, Switzerland). The protein was desalted by RP-HPLC and subjected to N-terminal sequence analysis.

2.5. Mass analysis

The determination of the molecular mass was performed on the mass spectrometer, model TSQ 700, from Finnigan (Bremen, Germany) in a triple stage quadrupol instrument equipped with electrospray ionization. The whole protein was dissolved in 50 μl 1% acetic acid (v/v) in 50% methanol (v/v) and injected into the capillary of the analyser.

2.6. Computer analysis

For comparative sequence analysis the programme ALIGN [17], FASTA [18] and GAP [19] were used to obtain maximal homology between related proteins. Computer searches were performed on a VAX/VMS 8600 in the Protein Identification Resource databank release 25 (National Biomedical Research Foundation, Washington, DC, USA) and the ribosomal databank (B. Wittmann-Liebold, unpublished results) The pI value of the protein was calculated by the programmn ISOELECTRIC from the University of Wisconsin Genetics Computer Group software package (release 5.0 and 6.2, 1990).

3. Results and discussion

3.1. Purification and characterization of protein S14

The total protein mixture of the 30S subunit from B. stearothermophilus was separated by RP-HPLC (Fig. 1). The purity of all peaks was checked by SDS-PAGE (Fig. 2) and by N-terminal sequence analysis. The second peak, indicated with an arrow in Fig. 1, contained the protein S14.

The amino acid composition of protein S14 was determined (data not shown), and a pI of 11.46 predicted from the sequence. Due to the hydrophilic and basic character the protein eluted at low concentration of organic solvent. The molecular mass of protein S14 is low; hence its migration is fast in electrophoresis (Fig. 2).

3.2. Amino acid sequence determination

The complete primary structure was determined by N-termi-

nal sequence analysis of the intact protein and overlapping peptides produced by different specific fragmentations, as shown in Fig. 3. The sequence was established using peptide sequences generated by endoproteinases Glu-C, Lys-C and cleavage by α-chymotrypsin. Cysteines were unequivocally identified as pyridylethyl-cysteine after pyridylethylation of the purified protein and sequencing from the N-terminus to position 47 after covalent attachment via the carboxyl-terminus to arylamino-membranes. Besides, the cysteine content deduced from the protein sequence was consistent with the composition obtained by amino acid analysis after performic acid oxidation.

Protein S14 is composed of 60 amino acid residues, and it has a calculated mass of 7148 Da.

To verify the sequence of protein BstS14 a mass spectrometric analysis with an electrospray mass spectrometer was carried out. A mass of 7144.0 Da was found. The mass determined exactly corresponds to the calculated mass (7148 Da) of protein BstS14, if the four cysteine residues of protein BstS14 are linked by sulphur bridges.

3.3. Comparative sequence analysis

The amino acid sequence of BstS14 was compared to other ribosomal sequences as described in section 2. Homologous counterparts were found in eubacteria, archaebacteria and eukaryotes. As indicated in Fig. 4 and Table 1 protein BstS14 is homologous to BsuS14 from Bacillus subtilis [20], McaS14 from Mycoplasma capricolum [21], TaqS14 from Thermus aquaticus [22], EcoS14 from Escherichia coli [23,24], VfaS14 from mitochondria of Vicia faba [25], OsaS14 from chloroplasts of Oryzia sativa [26], YcaS14 from mitochondria of Saccharomyces cerevisiae [27], MvaS14 from Methanococcus vannielii [28], and to HmaS14 from Haloarcula marismortui [29].
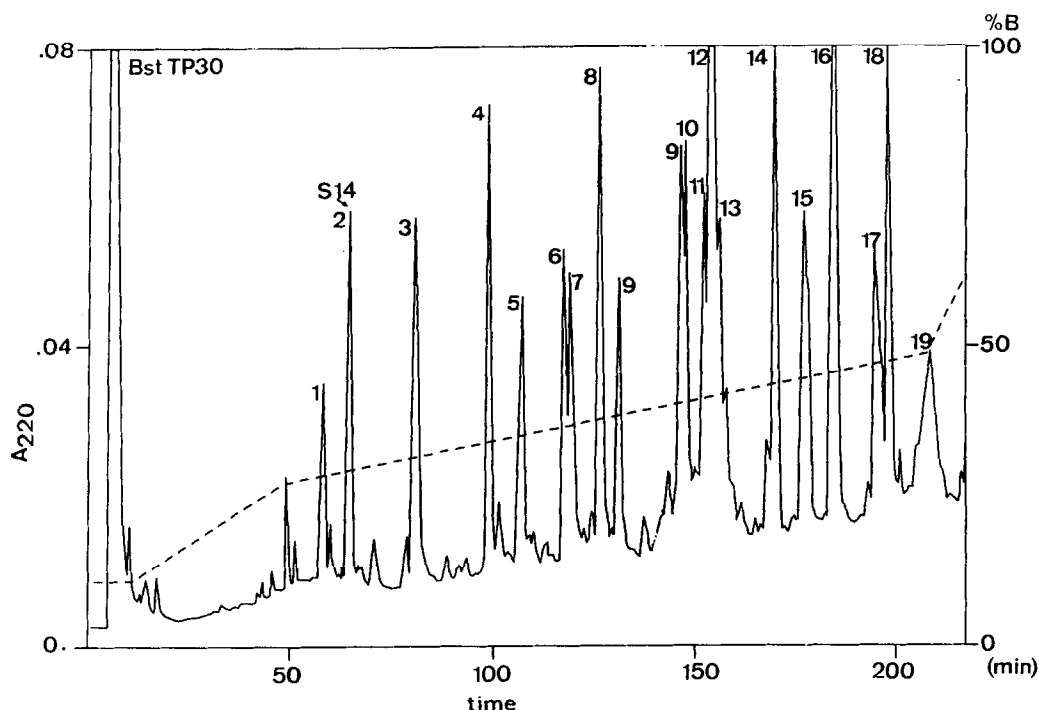
Fig. 1. Purification of BstS14 by RP-HPLC. 1 mg of B. stearothermophilus 30S ribosomal proteins were dissolved in 2% acetic acid and separated on a Vydac C$_4$ column. The proteins were eluted at 40°C at a flow rate of 0.5 ml/min in a gradient made of 0.1% aqueous TFA and 0.1% TFA in acetonitrile. The dotted line shows the gradient conditions. Measurements were made at 220 nm; range 0.08. The position of protein S14 is indicated by an arrow.
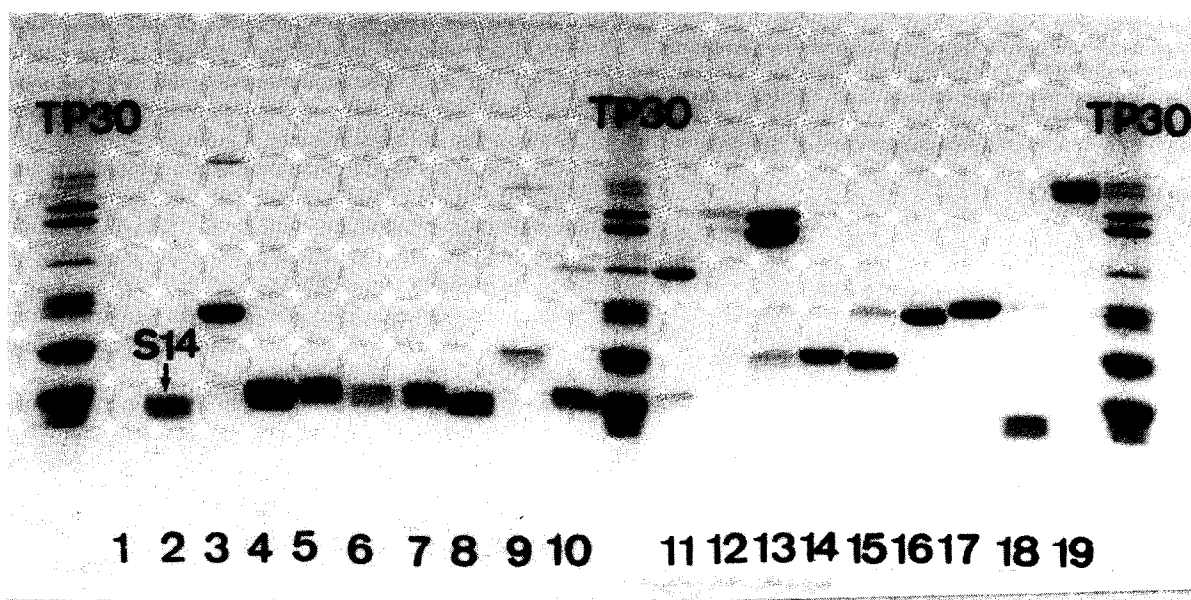
Fig. 2. Analysis of RP-HPLC fractions by one-dimensional SDS-PAGE. Lanes 1–19 correspond to the marked peaks in Fig. 1. 50 μg of total 30S ribosomal proteins (TP 30) was used as reference. Protein BstS14 is indicated.

Alignments of amino acid sequences show large insertions in the central regions for proteins EcoS14, for the protein S14 from chloroplasts (VfaS14), for the mitochondrial protein S14 from *Oryzia sativa* (OsaS14), for the archaebacteria HmaS14 and MvaS14, and for mitochondrial S14 from *Saccharomyces cerevisiae* (YeaS14). They are absent in S14 proteins from *Bacil-*

Fig. 3. Amino acid sequence of BstS14. Sequenced peptides are indicated by lines. Arrows represent positively identified PTH-amino acids. Cysteines were detected as pyridylethyl-cysteines. The enzymes used were Glu C-enzyme (Glu-C), lysyl-endoproteinase (Lys C) and α-chymotrypsin (Chy).

lus stearothermophilus, *Bacillus subtilus*, *Mycoplasma capricolum* and *Thermus aquaticus*. These data and the results presented for the percentage of identical amino acids in Table 1 indicate a evolutionary conservation for the N-terminal sequences of the Gram-positive bacteria, whereas the same region shows little homology to the sequences of Gram-negative bacteria. However, *E. coli* S14 is essential for in vitro reconstitution of functionally active 30S subunits and it is known, that the *E. coli* S14 can be replaced by the *B. stearothermophilus* homolog to form active 30S particles [30].

It is obvious that the N-termini of the S14 proteins are conserved for the Gram-positive bacteria (BstS14, BsuS14, McaS14) and TaqS14, and that the carboxyl-termini of the S14 proteins are strongly conserved through all three kingdoms, namely the amino acid sequence from position 80–118, respectively. Thus it might be expected, that the highly conserved carboxyl-terminal region of protein S14 is important for the interaction with the 16S RNA and is involved in the initiation function of active ribosomes.

Interestingly, all four cysteine residues of BstS14, BsuS14, McaS14 and TaqS14 are arranged in pairs with the motif Cys-X_2-Cys in the carboxyl-terminal region. In BstS14, BsuS14, McaS14, and TaqS14 these two groups of cysteine residues are separated by 12 amino acids. Usually, cysteine residues rarely occur in ribosomal protein sequences. So far conserved spacing of cysteine residues could only be observed in protein L31 from *E. coli*, and in proteins L32, L33 from *B. stearothermophilus* [31]. A possible function may be the formation of disulphide bridges which would stabilize the tertiary structure of the proteins. The difference in molecular masses deduced from sequence and mass spectrometry is underlining this assumption.

In summary, it becomes evident that only complete sequence sets of ribosomal proteins and their comparisons will identify regions of high evolutionary conservation in the different ribosomes. This may lead to an understanding of their functional implication in protein biosynthesis on the ribosome.
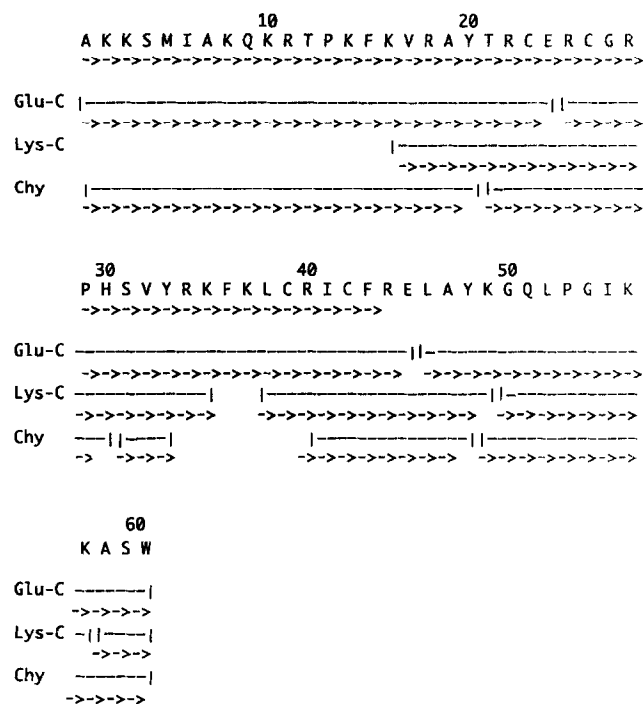
```
          1                                                    40
BstS14    . . . . . . . . . . . . . . . . . . A K K S  M I A K Q K R T P K  F K V R A Y T . . .
BsuS14    . . . . . . . . . . . . . . . . M A K K S  M I A K Q Q R T P K  F K V Q E Y T . . .
McaS14    . . . . . . . . . . . . . . . . M A K K S  L K V K Q A K H P K  F N V R N Y T
TaqS14    . . . . . . . . . . . . . . . . M A R K A  L I E K A K R T P K  F K V R A Y T . . .
EcoS14    . . . . . . . . . . . . . . A K Q S  M K A R E V K R V A  L A D K Y F A K R A
VfaS14    . . . . . . . . . . . . M S E K  R N I R D H K R R L  L A A K Y E L R R K
OsaS14    . . . . . . . . . . . . M A K K S  L I Q R E R K R Q K  L E Q K Y H L I R R
YeaS14    M G N F R F P I K T  K L P P G F I N A R  I L R D N F K R Q Q  P K E N E I L V K S
MvaS14    . . . . . . . . . .
HmaS14    . . . . . . . . . .

          41                                                   80
BstS14    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R
BsuS14    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R
McaS14    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R
TaqS14    . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . R
EcoS14    E L K A I I S D V . . N A S D E D R W N  A V L K L Q T L P R  D S S P S R Q R N R
VfaS14    L Y K A F C K D S . . D L P S D W D K  L R Y K L S K L P R  N S S F A R V R N R
OsaS14    S S K K K I R S K V Y  P L S L S E K T K  M R E K L Q S L P R  N S A P T R L H R R
YeaS14    . . . L K F I A R N M  N L P T K L R L E  A Q L K L N A L P N  Y M R S T Q I K N R
MvaS14    . . . . . . . . . . . . . . . . . . . M T K E P F K T K  Y G Q G S K V C K R
HmaS14    . . . . . . . . . . . . . . M S E S E T T D  E P D S E T A S S E  R T G Q L E S C Q R

          81                                                   116
BstS14    C E R C G R P H S V  Y R K F K L C R I C  F R E L A Y K G Q L  P G I K K A S W
BsuS14    C E R C G R P H S V  I R K F K L C R I C  F R E L A Y K G Q I  P G V K K A S W
McaS14    C N H C G R P H A V  L K K F G I C R L C  F R K F A Y E G Q I  P G I K K A S .
TaqS14    C V R C G R A R S V  Y R Y F G L C R T C  L R E L A H K G Q L  P G V K K A S W
EcoS14    C R Q T G R P H G F  L R K F G L S R I K  V R E A A M R G Q I  P G L K K G . .
VfaS14    C I S T G R P R S V  Y E L F R I S R I V  F R S L A S R G P L  M G I K K S S W
OsaS14    C F L T G R P R A N  Y R D F G L S G H I  L R E M V Y A C L L  P G A T R S S W
YeaS14    C V D S G H A R F V  L S D F R L C R Y Q  F R E N A L K G N L  P G V K K G I W
MvaS14    C G R K G P G I I R  K Y G L D L C R Q C  F R E L A P K . . .  L G F K K Y D .
HmaS14    C G R . E Q G L V G  K Y D I W L C R Q C  F R E I S R G . . .  M G F R K Y S .
```
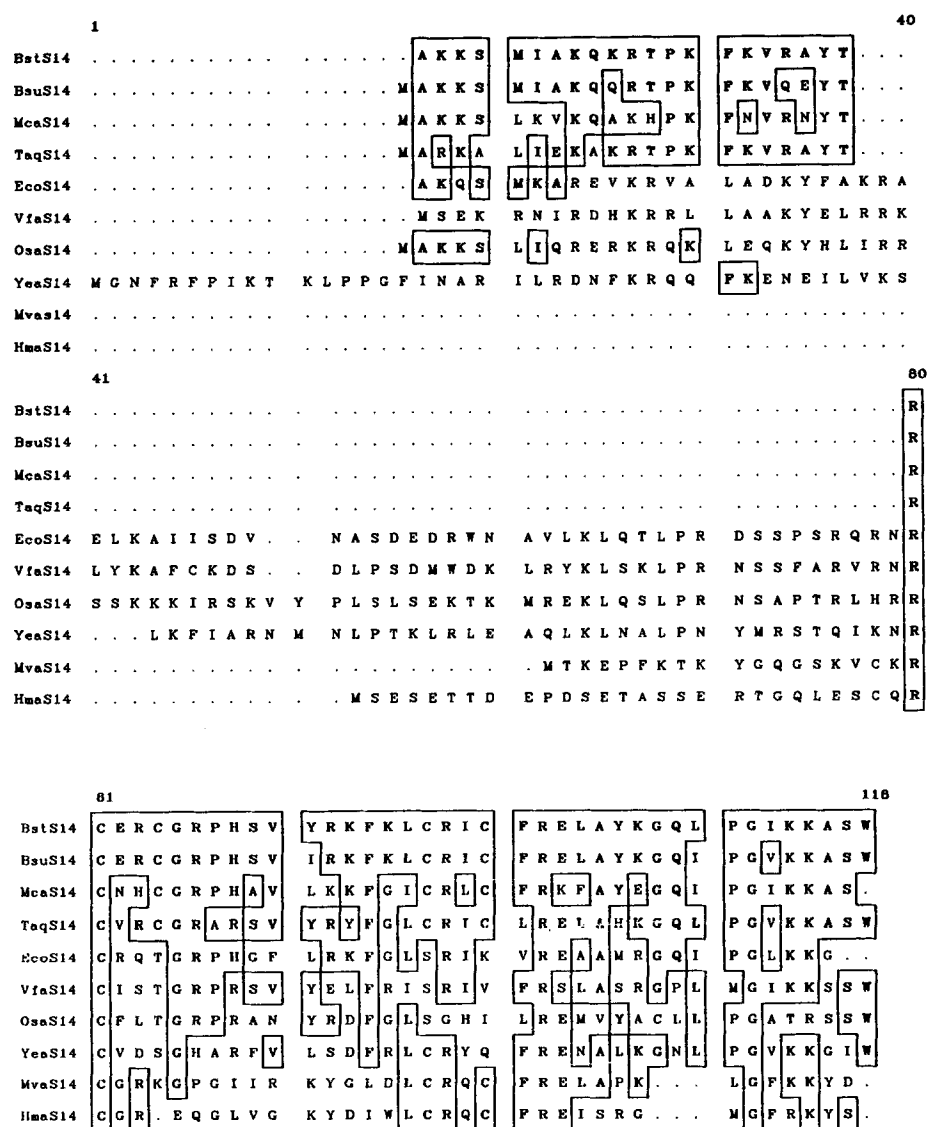
Fig. 4. Sequence similarity of protein BstS14 to related proteins. Identical residues are boxed. Nearly all S14 proteins, except proteins S14 from E. coli [23] and from B. stearothermophilus (this paper) were deduced from the respective DNA sequence. Bst, Bacillus stearothermophilus, Bsu, Bacillus subtilis, Mca, Mycoplasma capricolum, Taq, Thermus aquaticus, Eco, Escherichia coli, Vfa, Vicia faba, Osa, Oryzia sativa, Yea, mitochondrial ribosomal protein of Saccharomyces cerevisiae, Mva, Methanococcus vannielii, Hma, Haloarcula marismortui.

Table 1
Percent identity (%) in the ribosomal protein family S14 of each pair of proteins homologous to BstS14

	YeaS14	OsaS14	VfaS14	EcoS14	HmaS14	MvaS14	McaS14	TaqS14	BsuS14	BstS14
BstS14	38	40	45	47	39	44	67	78	90	–
BsuS14	38	39	38	50	38	50	65	72	–	–
TaqS14	48	41	41	39	28	38	55	–	–	–
McaS14	40	37	43	43	28	40	–	–	–	–
MvaS14	26	23	28	34	40	–	–	–	–	–
HmaS14	20	21	34	23	–	–	–	–	–	–
EcoS14	22	41	39	–	–	–	–	–	–	–
VfaS14	31	39	–	–	–	–	–	–	–	–
OsaS14	27	–	–	–	–	–	–	–	–	–
YeaS14	–	–	–	–	–	–	–	–	–	–

Percentage of identities have been ascertained by the computer program GAP [19]. For abbreviations, see legend to Fig. 4.

References

[1] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G., Eds.) pp. 326–361, Springer, Heidelberg, New York.

[2] Noller, H.F., Asire, M., Barta, A., Douthwaite, S., Goldstein, T., Gutell, R.R., Moazed, D., Normanly, J., Prince, J.B., Stern, S., Triman, K., Turner, S., Van Stolk, B., Wheaton, B., Weiser, B. and Woese, C.R. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G., Eds.) pp. 143–163, Springer, Heidelberg, New York.

[3] Wittmann-Liebold, B., Köpke, A.K.E., Arndt, E., Krömer, W., Hatakeyama, T. and Wittmann, H.G. (1990) in: The Ribosome: Structure, Function & Evolution (Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R., Eds.) pp. 598–616, American Society for Microbiology, Washington, DC.

[4] Arndt, E., Scholzen, T., Krömer, W., Hatakeyama, T. and Kimura, M. (1991) Biochimie 73, 657–668.

[5] Lake, J.A., Pendergast, M., Kahan, L. and Nomura, M. (1974) Proc. Natl. Acad. Sci. USA 71, 4688–4692.

[6] Girshovich, A.S., Bochkareva, E.S., Kramarov, V.A., and Ovchinikov, Y.A. (1974) FEBS. Lett. 42, 213–217.

[7] Pellegrini, M., Oen, H., Eilat, D. and Cantor, C.R. (1974) J. Mol. Biol. 88, 809–829.

[8] Bode, U., Lutter, L.C. and Stöffler, G. (1974) FEBS Lett. 45, 232–236.

[9] Traut, R.R., Lambert, J.M., Boileau, G. and Kenny, J. W. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G., Craven, G.R., Davies, J., Kahan, L. and Nomura, M., Eds.) pp. 89–110, University Park Press, Baltimore.

[10] Bollen, A., Heimark, R.L., Cazzone, A., Traut, R.R., Hershey, J.W. and Kahan, L. (1975) J. Biol. Chem. 250, 4310–4314.

[11] Brockmöller, J. and Kamp, R.M. (1986) Biol. Chem. Hoppe-Seyler 367, 925–935.

[12] Hardy, S.J.S., Kurland, C.G., Voynow, P. and Mora, G. (1969) Biochemistry 7, 2897–2905.

[13] Laemmli, U.K. (1970) Nature 227, 680–685.

[14] Ashman, K. and Bosserhoff, A. (1985) in: Modern Methods in Protein Chemistry (Tschesche, H., Ed.) pp. 155–171, Vol. 2, W. de Gruyter, Berlin, New York.

[15] Henschen, A. (1986) in: Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. and Erdmann, V.A., Eds.) pp. 244–255, Springer, Berlin, New York.

[16] Herfurth, E., Pilling, U. and Wittmann-Liebold, B. (1991) Biol. Chem. Hoppe-Seyler 372, 351–361.

[17] Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) NBRF Report 820501-08710 Program Version 1.0 National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC, USA.

[18] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.

[19] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.

[20] Henkin, T.M., Moon, S.H., Mattheakis, L.C. and Nomura, M. (1989) Nucleic Acids Res. 17, 7469–7486.

[21] Ohkubo, S., Muto, A., Kawauchi, Y., Yamao, F. and Osawa, S. (1987) Mol. Gen. Genet. 210, 314–322.

[22] Jahn, O., Hartmann, R.K. and Erdmann, V.A. (1991) Eur. J. Biochem. 197, 733–740.

[23] Yaguchi, M., Roy, C., Reithmeier, R.A.F., Wittmann-Liebold, B. and Wittmann, H.G. (1983) FEBS Lett. 154, 21–30.

[24] Ceretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M. and Nomura, M. (1983) Nucleic Acids Res. 11, 2599–2616.

[25] Wahleithner, J.A. and Wolstenholme, D.R. (1988) Nucleic Acids Res. 16, 6897–6913.

[26] Coté, J.-C. and Wu, R. (1988) Nucleic Acids Res. 16, 10384–10384.

[27] Myers, A.M., Crivellone, M.D., and Tzagoloff, A. (1987) J. Biol. Chem. 262, 3388–3397.

[28] Auer, J., Spicker, G. and Böck, A. (1989) J. Mol. Biol. 209, 21–36.

[29] Scholzen, T. and Arndt, E. (1991) Mol. Gen. Genet. 228, 70–80.

[30] Higo, K., Held, W., Kohan, L. and Nomura, M. (1973) Proc. Natl. Acad. Sci. USA 70, 944–948.

[31] Kruft, V., Kapp, U. and Wittmann-Liebold, B. (1991) Biochimie 73, 855–860.