Volume 245, number 1,2, 253–260 FEB 06913 March 1989

The complete amino acid sequence of ribosomal protein S18 from the moderate thermophile *Bacillus stearothermophilus*

John McDougall, Theodora Choli, Volker Kruft, Ulrike Kapp and Brigitte Wittmann-Liebold

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

Received 27 December 1988

The amino acid sequence of ribosomal protein S18 from *Bacillus stearothermophilus* has been completely determined by automated sequence analysis of the intact protein as well as of peptides derived from digestion with *Staphylococcus aureus* protease at pH 4.0 and cleavage with cyanogen bromide. The carboxy-terminal region was verified by both amino acid analyses of chymotryptic peptides and by mass spectrometry from the terminal region. The protein contains 77 amino acid residues and has an M_r of 8838. Comparison of this sequence with the sequences of the S18 proteins from tobacco and liverwort chloroplasts and *E. coli* shows a relatively high similarity, ranging from 42 to 55% identical residues with the *B. stearothermophilus* S18 protein. The regions of homology common to all four proteins consist of several positively charged sections spanning the entire length of the protein.

Ribosomal protein S18; Amino acid sequence; Primary structure; (*Bacillus stearothermophilus*)

1. INTRODUCTION

The primary structures of numerous prokaryotic ribosomal proteins have been determined to date and compared to counterparts in other organisms (for example see [1–3]). The results of such comparisons allow us to correlate homologous proteins to reveal regions that have been evolutionarily conserved, probably for reasons of either tertiary and quaternary structure or function. In addition, a comparative approach may allow the identification of sites for RNA or protein interaction, or regions that may have arisen by gene duplication [4]. In using a comparative approach, it has been especially valuable to examine ribosomal proteins that are located in a well-defined region of the 70 S ribosome. One such protein of particular interest is the ribosomal protein S18, which is located on the platform region of the *E. coli* ribosome [5]. This is one of the best mapped areas known in the ribosome and is made up of proteins S6, S11, S15 and S18, along with the decoding region of the 16 S ribosomal RNA (rRNA). In this region, codon-anticodon recognition is thought to occur (see [6]). The *E. coli* S18 protein (EcoS18) has been positioned both by immunological and affinity labeling [7] as well as by neutron scattering [8]. Crosslinking data have placed EcoS18 10–20 Å from protein S1 and 0–10 Å from protein S6 [9], while neutron scattering data have further located it with respect to the distance from the center of masses of proteins S6 (33 Å) and S11 (70 Å) [8]. EcoS18 has also been crosslinked to messenger RNA (mRNA), 16 S rRNA and IF-1 [7,10,11], and is thought to be located in the tRNA A-site of the ribosome [12]. In addition, EcoS18 appears to be involved in conformational changes both during the activation of the ribosome and also when the ribosome interacts with EF-G [13,14]. Due to the central importance of this area of the ribosome in general, and the importance of EcoS18 specifically, comparative studies on the S18 proteins have been initiated.

In this paper, the complete amino acid sequence of the protein S18 from *B. stearothermophilus* is described. This sequence is compared to the primary sequence of the S18 protein from *E. coli*

Correspondence (present) address: J. McDougall, College of Fishery Sciences, University of Tromsø, Tromsø, Norway

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

[15], tobacco chloroplast [16] and *Marchantia polymorpha* (liverwort) chloroplast [17]. As well, regions of the S18 protein are compared to other ribosomal proteins that have been found to have significant similarity. The results of these comparisons are discussed with respect to proposed functions within the ribosome and the nature of the conserved amino acid residues.

2. MATERIALS AND METHODS

2.1. *Materials*

The growth of *B. stearothermophilus* cells (strain 799), the isolation of 70 S ribosomes and the preparation of ribosomal subunits were as described previously [18]. *Staphylococcus aureus* protease was from Miles and chymotrypsin was from Boehringer, Mannheim.

2.2. *Protein nomenclature*

Proteins from *B. stearothermophilus*, tobacco and *Marchantia polymorpha* chloroplast are numbered according to the homologous proteins in *E. coli*, as determined by either their N-terminal or full protein sequence. When necessary, the prefix Bst (*B. stearothermophilus*), Eco (*E. coli*), Nta (tobacco chloroplast) or Mpo (*Marchantia* chloroplast) is added to differentiate between the similar proteins.

2.3. *Protein isolation*

Ribosomal proteins from *B. stearothermophilus* were extracted from the 30 S subunit as described by Hardy et al. [19], dialyzed extensively against 2% acetic acid containing 10 mM 2-mercaptoethanol and then lyophilized. The proteins were separated by reverse phase HPLC (RP-HPLC) on a Vydac C4 (300 Å, 5 μm) column using a 2-propanol gradient in aqueous TFA, similar to that described by Hirano et al. [20]. Peak fractions were pooled and analyzed by micro two-dimensional gel electrophoresis [21]. The resultant BstS18 protein-containing peak, which was contaminated with small amounts of BstS12 and BstS19 proteins, was rechromatographed on RP-HPLC using a LiChrospher C8 (100 Å, 5 μm) column with an acetonitrile gradient in aqueous TFA. The purified BstS18 protein was then subjected to sequence analysis.

2.4. *Sequence determination and amino acid analyses*

Purified BstS18 protein was cleaved enzymatically with *S. aureus* protease as well as chemically with cyanogen bromide in 70% TFA as described in [22]. The resulting peptides were separated by RP-HPLC (TSK ODS 120T, 5 μm column) with an acetonitrile gradient in aqueous TFA. Amino acid analyses of the intact protein and the peptides were performed using an RP-HPLC system after precolumn derivatization with *o*-phthaldialdehyde (OPA) [23]. Protein and peptide sequence determination was performed on an Applied Biosystems pulsed-liquid phase sequencer, model 477A, equipped with the model 120 PTH-aa analyzer. Samples were dissolved in 100% TFA and applied to a TFA-treated polybrene-coated glass filter that had been pre-cycled, as described previously [24]. The determination of the carboxy-terminal sequence was verified by additional methods: the cyanogen bromide peptide CB2 (fig.1) was digested with chymotrypsin and the resultant two small peptides were purified by RP-HPLC (as described above) for amino acid analysis. In addition, the CN2 peptide was analyzed by mass spectrometry as described in [20]. Cysteine content was determined by amino acid analysis after performic acid oxidation, as described in [25], and measured against standard proteins.

3. RESULTS AND DISCUSSION

3.1. *The amino acid sequence of protein BstS18*

The complete amino acid sequence of protein S18 from the *B. stearothermophilus* ribosome was determined by automated N-terminal sequence

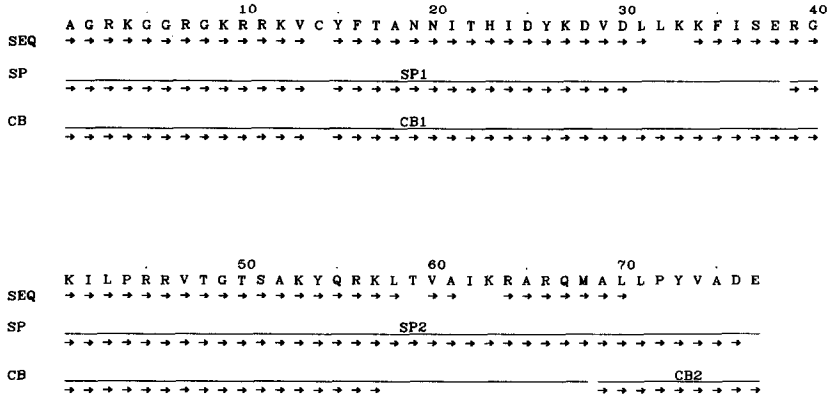


Fig.1. Amino acid sequence of protein S18 from *B. stearothermophilus*. Sequence data of the intact protein and peptides by automated protein sequencer analysis is indicated (→). SEQ indicates the direct degradation of the intact protein. SP and CB indicate peptides derived from cleavage with *Staphylococcus aureus* protease and cyanogen bromide, respectively.

Table 1

Amino acid composition and M_r of the S18 ribosomal protein family as derived from the sequences

Amino acid	BstS18	EcoS18	MpoS18	NtaS18
Ala	7	7	3	4
Phe	2	3	2	7
Lys	10	6	6	10
Pro	2	2	3	4
Thr	5	6	2	6
Cys	1	1	0	0
Gly	6	3	2	3
Leu	6	5	10	12
Gln	2	5	3	6
Val	5	2	2	1
Asp	4	3	1	3
His	1	1	0	0
Met	1	0	3	2
Arg	10	12	13	16
Tyr	4	6	1	1
Glu	2	2	3	4
Ile	5	6	7	8
Asn	2	1	6	5
Ser	2	3	8	9
Total	77	74	75	101
M_r	8838	8855	8880	12053

Table 2

Degree of similarity of individual proteins, as percent identity, calculated for each pair of protein S18

	B. stearothermophilus	*E. coli*	*M. polymorpha*
B. stearothermophilus	–		
E. coli	55	–	
M. polymorpha	47	38	–
N. tabacum	42	33	72

analysis of the intact protein up to position 66 and of several peptides, as summarized in fig.1. Identification of most of the residues was confirmed by at least two independent sequencer runs.

Digestion of BstS18 with *S. aureus* protease at pH 4.0 yielded two peptides which were separated by RP-HPLC, sequenced and aligned with the intact protein sequence. The CB2 peptide was obtained by cleaving either the SP2 fragment or the intact BstS18 protein; the resultant peptides were separated by RP-HPLC and the sequences were

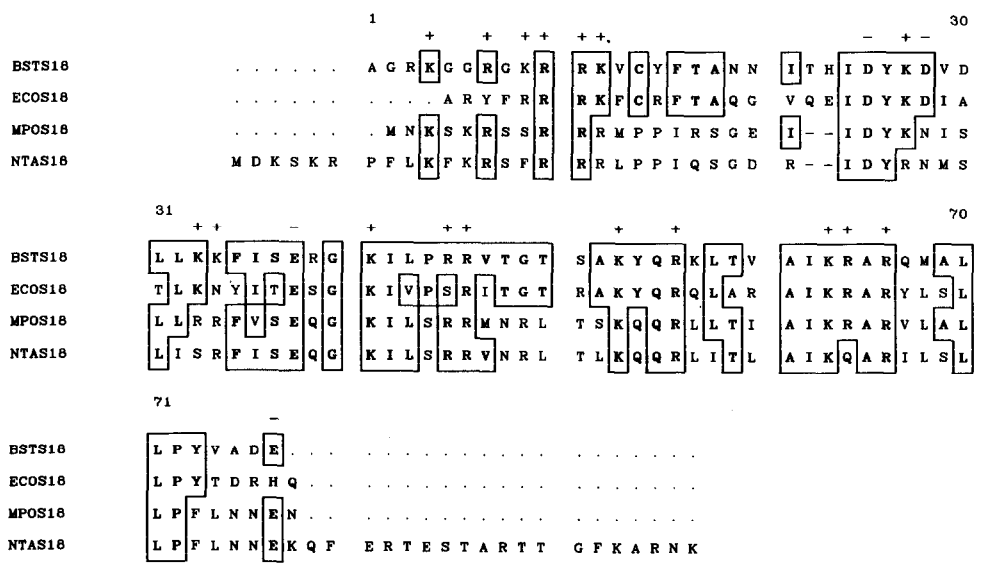


Fig.2. Comparison of the amino acid sequences of S18 from *B. stearothermophilus* (Bst) ribosomes with the homologous sequences from *E. coli* (Eco) [15], and from the chloroplasts of *Marchantia polymorpha* (Mpo) [16] and *Nicotiana tabacum* (Nta) [17]. Identical residues are boxed and sites having conserved charges are marked above the sequences (+ or −). Maximum homology was obtained by the program ALIGN [26] using the standard mutation data matrix, 50 random runs and a break penalty of 20. Numbering is based on the residues of the protein S18 from *B. stearothermophilus*.

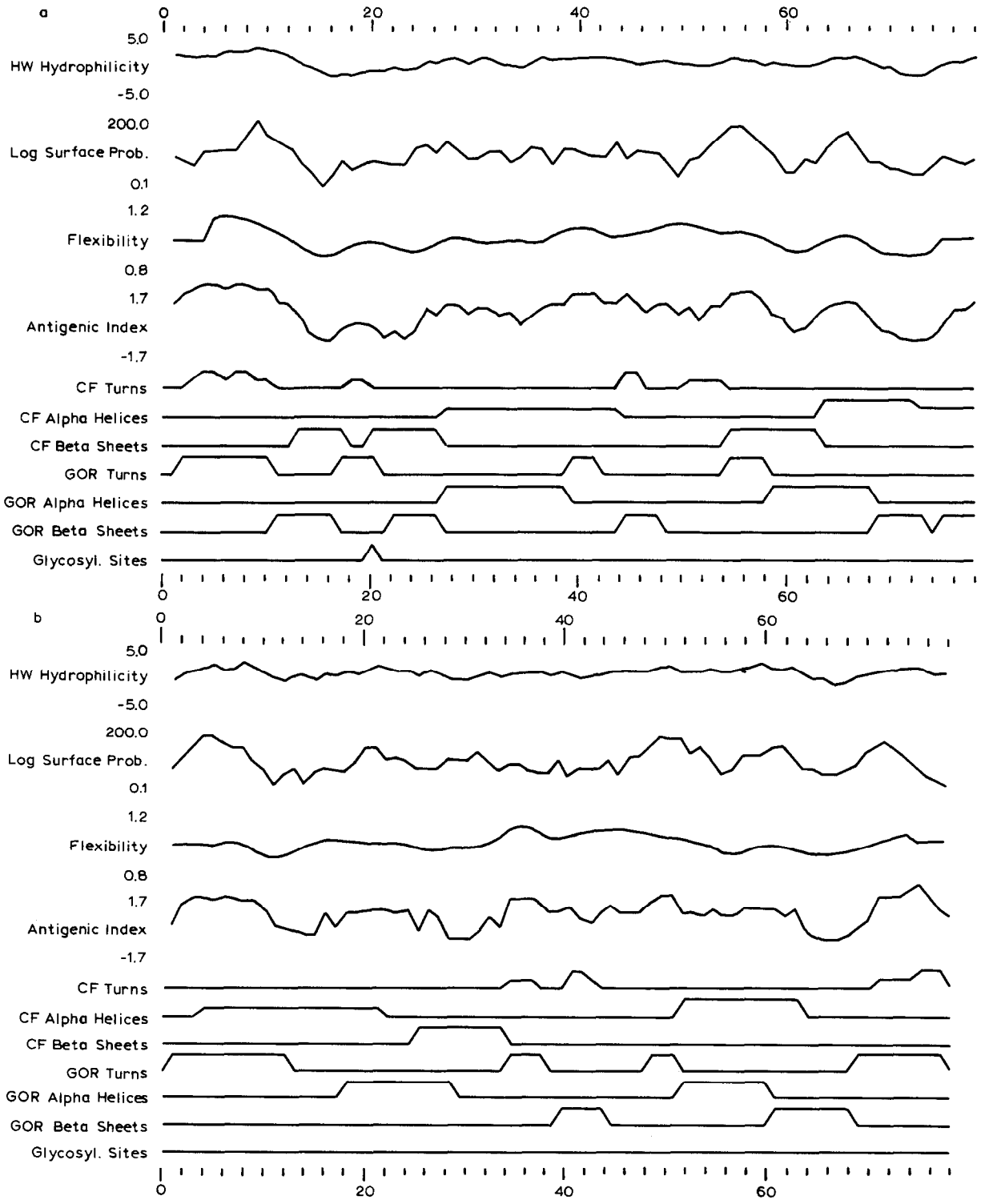
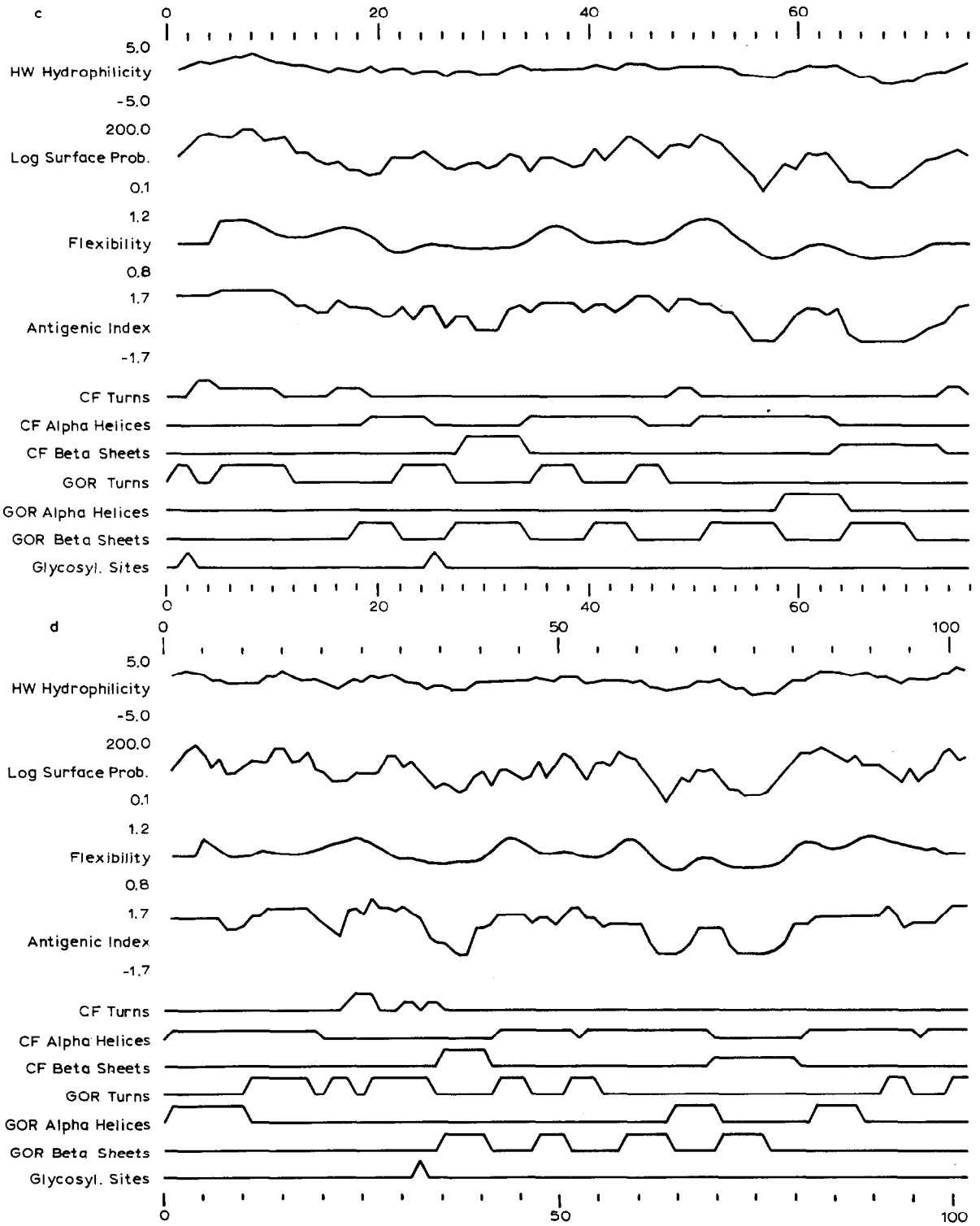


Fig.3. Secondary structure predictions and higher order structural components of BstS18 (a), EcoS18 (b), MpoS18 (c) and NtaS18 (d). The secondary structures were calculated according to the methods of Chou and Fasman [29] and Garnier et al. [30]. The prediction



of higher order structural components was performed on a Vax 8600 computer with programs from the Genetics Computer Group [31]. Numbering along the top and bottom of each figure denotes the residue number within the protein.

determined. Since small peptides such as CB2 do not bind well to the polybrene-coated glass fiber support used in the sequencer, the yields for the last few residues at the carboxy-terminus were low and the amino acid identification was difficult. To verify the sequence and confirm the terminal residues, the CB2 fragment was digested with chymotrypsin; the resulting two peptides were separated by RP-HPLC and then subjected to amino acid analysis. In addition, mass spectrometric analysis of the CB2 peptide confirmed the amino acid composition derived from OPA amino acid analysis of this peptide and the terminal chymotryptic peptides. All amino acid analyses were in agreement with the sequence determination of the CB2 and SP2 fragments and verify the terminal glutamic acid residue. The residue at position 14 of the protein could not be identified by repeated automated sequence analyses as the PTH-amino acid derivative suggested a cysteine residue at this position. The presence of this residue was confirmed by performic acid oxidation of the protein.

From these results, the amino acid sequence of BstS18 was unambiguously determined. As given in table 1, protein BstS18 contains 77 amino acid residues and has a calculated M_r of 8838.

3.2. *Comparison of amino acid sequences*

The amino acid sequence of protein S18 from *B. stearothermophilus* has been aligned with previously sequenced S18 proteins from *E. coli*, tobacco chloroplast and *M. polymorpha* chloroplast (fig.2). As can be seen, BstS18 can be easily aligned with these proteins, and they all show a relatively high similarity to each other (table 2). The highest sequence alignment scores are found between the more evolutionarily related two eubacterial proteins EcoS18 and BstS18 (55%), and the two chloroplast proteins NtaS18 and MpoS18 (72%). The higher degree of similarity found between the two chloroplast-encoded S18 proteins probably indicates the more recent divergence of the two organelles' genomes from each other. In fact, the overall degree of similarity amongst all four proteins indicates a rather low rate of evolutionary change in comparison to other ribosomal proteins (see [1,2]). This may reflect the importance of the protein's structure or function within the ribosome and would be in agreement with the proposed location of S18 in the decoding region of the ribosome. Interestingly, there has also been a significant conservation of charged residues (fig.2), which, because of the large number of basic amino acids in this protein, may indicate an RNA binding function of these regions.

Most of the conserved sequence regions of the proteins occur in several stretches along the entire protein length. In comparing BstS18 to EcoS18, the former contains four extra residues at the N-terminus and has one less residue at the C-terminus but no insertions or deletions are found. The only significant change occurs at the C-terminus, where the overall positively charged region in EcoS18 is acidic in BstS18. Both proteins contain a conserved cysteine residue at position 14, while the single methionine and histidine residues in BstS18 are not conserved. As expected, the two chloroplast S18 proteins are quite similar to each other and have numerous identical residues that are not shared by the two eubacterial proteins. As well, the chloroplast S18 proteins have a two-residue deletion at position 22/23 of BstS18 and they do not contain the cysteine that is conserved in EcoS18 and BstS18, nor any histidine. Interestingly, although MpoS18 is of approximately the same length as the two eubacterial proteins, the homologous NtaS18 protein has six additional N-terminal and twenty additional C-terminal residues when compared to BstS18. Protein S18 from *E. coli* was found to have an acetylated N-terminal alanine [15], while the BstS18 protein has no modified N-terminal residue. Since the protein sequences of the two chloroplast S18 proteins were derived from DNA sequence analysis, it is not known whether either of these proteins contain modified residues.

As would be expected, the secondary structure predictions for the four S18 proteins are quite similar (fig.3). A comparative analysis of the four S18 proteins shows very similar higher order structural features and surface topography along the entire protein length. In general, the proteins have several regions of relatively high hydrophilicity, e.g. an extended region towards the N-terminus. There appear to be three major areas of the four proteins that are predicted to be on the surface; the regions encompass residues 7–12, 53–57 and 63–67 of BstS18. All of these regions are relatively rich in basic amino acids, which again suggests a

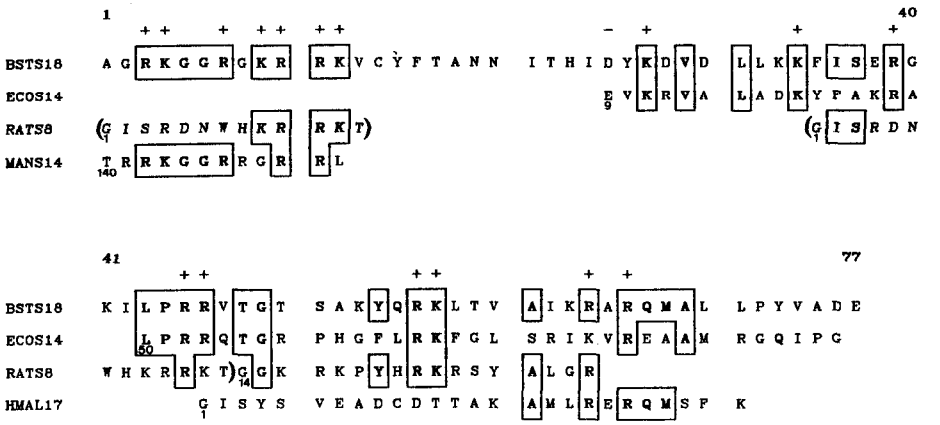


Fig.4. Primary structure similarities between the S18 protein from *B. stearothermophilus* (BstS18) and ribosomal proteins S14 from *E. coli* (EcoS14) [32], the C-terminal sequence of S14 from human (ManS14) [33], and the N-terminal sequences of S8 from rat liver (RatS8) [34] and L27 from *H. marismortui* (HmaL17) [35]. Identical residues are boxed and residues having conserved charges are marked above the sequence (+ or −). The first residue number of each sequence is indicated below the residue. The part of the RatS8 protein found to have similarity to two different regions of BstS18 is indicated in these positions by brackets. The program ALIGN was used, as described in fig.2.

possible RNA-binding function. In addition to the similar structural features that are common to the four S18 proteins, additional features can be seen that are common to only the eubacterial or chloroplast protein pairs, again reflecting the higher sequence homologies found within these pairs. The results of these structural analyses again argue for an evolutionarily conserved three-dimensional structure and function.

It is interesting to note that, using the ALIGN programs [26], protein BstS18 also shows a significant similarity to ribosomal proteins EcoS14 and human S14 protein (ManS14) as well as the N-terminus regions of the partially sequenced RatS8 and *Halobacterium marismortui* L17 proteins (fig.4). In addition, BstS18 shows a lower but still significant similarity to several other eubacterial and archaebacterial ribosomal proteins, as has been shown previously for other eubacterial ribosomal proteins [27]. At present, it is not clear whether these similarities are structurally, functionally or evolutionarily derived. However, a high proportion of the residues found to be homologous to BstS18 are basic residues. Since several of these proteins have been mapped or crosslinked to rRNA, mRNA and tRNA (for a review see [28]), RNA binding may be responsible for the weak sequence homology that has been detected. Similar results have also been found for the EcoS18, NtaS18 and MpoS18 proteins (results not shown). More information on both the primary and higher order structure of the ribosomal proteins and their function within the ribosome is still required before the nature and origin of these sequence similarities can be understood.

Acknowledgements: We would like to thank Dr H.G. Wittmann for his support, Dr K. Eckart for performing the mass spectrometry experiments and D. Kamp for performing the amino acid analyses.

REFERENCES

[1] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.326–361, Springer-Verlag, Heidelberg.
[2] Kimura, M. and Kimura, J. (1987) FEBS Lett. 210, 91–96.
[3] Strobel, O., Köpke, A.K.E., Kamp, R.M., Böck, A. and Wittmann-Liebold, B. (1988) J. Biol. Chem. 263, 6538–6546.
[4] Wittmann-Liebold, B., Ashman, K. and Dzionara, M. (1983) FEBS Lett. 154, 31–41.
[5] Stöffler, G. and Stöffler-Meilicke, M. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.28–46, Springer-Verlag, Heidelberg.

[6] Oakes, M., Henderson, E., Scheinman, A., Clark, M. and Lake, J.A. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.47–67, Springer-Verlag, Heidelberg.
[7] Cooperman, B.S., Hall, C.C., Kerlavage, A.R., Weitzmann, C.J., Smith, J., Hasan, T. and Friedlander, J.D. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.362–378, Springer-Verlag, Heidelberg.
[8] Ramakrishnan, V., Capel, M., Kjeldgaard, M., Engelman, D.M. and Moore, P.B. (1984) J. Mol. Biol. 174, 265–284.
[9] Lambert, J.M., Boileau, G., Cover, J.A. and Traut, R.R. (1983) Biochemistry 22, 3913–3920.
[10] Greuer, B., Osswald, M., Brimacombe, R. and Stöffler, G. (1987) Nucleic Acids Res. 15, 3241–3255.
[11] Boileau, G., Butler, P., Hershey, J.W.B. and Traut, R.R. (1983) Biochemistry 22, 3162–3170.
[12] Lelong, J.C., Maschler, R., Crepin, M., Jeantet, C., Tischendorf, G.W. and Gros, F. (1979) Biochimie 61, 881–889.
[13] Ginzburg, I., Miskin, R. and Zamir, A. (1973) J. Mol. Biol. 79, 481–494.
[14] Gudkov, A.T. and Gongadze, G.M. (1984) FEBS Lett. 176, 32–35.
[15] Yaguchi, M. (1975) FEBS Lett. 59, 217–220.
[16] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[17] Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[18] Brockmöller, J. and Kamp, R.M. (1986) Biol. Chem. Hoppe-Seyler 367, 925–935.
[19] Hardy, S.J.S., Kurland, C.G., Voynow, P. and Mora, G. Biochemistry 7, 2897–2905.
[20] Hirano, H., Eckart, K., Kimura, M. and Wittmann-Liebold, B. (1987) Eur. J. Biochem. 170, 149–157.
[21] Brockmöller, J. and Kamp, R.M. (1986) in: Advanced Methods in Protein Sequence Analysis (Wittmann-Liebold, B. ed.) pp.34–44, Springer-Verlag, Heidelberg.
[22] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed.) pp.49–72, Elsevier, Amsterdam.
[23] Ashman, K. and Bosserhof, A. (1985) in: Modern Methods in Analytical Protein Chemistry (Tschesche, H. ed.) pp.155–171.
[24] Hewick, R.H., Hunkerpiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[25] Hirs, C.H.W. (1967) Methods Enzymol. 11, 197–199.
[26] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, Vol.5, Suppl.3, National Biomedical Research Foundation, Washington, DC.
[27] Wittmann-Liebold, B., Ashman, K. and Dzionara, M. (1984) Mol. Gen. Genet. 196, 439–448.
[28] Hardesty, B. and Kramer, G. (eds) (1986) Structure, Function and Genetics of Ribosomes, Springer-Verlag, Heidelberg.
[29] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–147.
[30] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[31] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[32] Yaguchi, M., Roy, C., Reithmeier, R.A.F., Wittmann-Liebold, B. and Wittmann, H.G. (1983) FEBS Lett. 154, 21–29.
[33] Rhoads, D.D., Dixit, A. and Roufa, D.J. (1986) Mol. Cell. Biol. 6, 2774–2783.
[34] Wittmann-Liebold, B., Geissler, A.W., Lin, A. and Wool, I.G. (1979) J. Supramol. Struct. 12, 425–433.
[35] Walsh, M.J., McDougall, J. and Wittmann-Liebold, B. (1988) Biochemistry 27, 6867–6876.