The primary structure of protein S14 from the small ribosomal subunit of *Escherichia coli*

M. Yaguchi, C. Roy, R.A.F. Reithmeier*, B. Wittmann-Liebold+ and H.G. Wittmann+

Division of Biological Sciences, National Research Council, Ottawa, Canada and +*Max-Planck-Institut für Molekulare Genetik, D-1000 Berlin 33 (Dahlem), Germany*

Received 31 January 1983

Protein S14 was isolated in pure form from *Escherichia coli* ribosomal 30 S subunits. Its complete amino acid sequence was determined by a combination of various approaches, such as enzymatic and chemical cleavage of the protein chain, isolation of the resulting peptides as well as manual and automatic sequence determination by the Edman degradation technique. The protein has an M_r of 11191 and consists of 98 amino acid residues, 26 of which are basic and 9 acidic. One residue each of cysteine, histidine, tyrosine and tryptophan is present in the protein. The secondary structure of protein S14 as predicted according to 4 different programs shows a long α-helix in the N-terminal region and a short α-helix near the C-terminus of the protein chain. When the amino acid sequence of protein S14 was compared with that of all other *E. coli* ribosomal proteins with computer search programs, only relatively short homologous regions were found. A comparison between protein S14 of *E. coli* and the homologous protein from *Bacillus stearothermophilus* revealed ~35% identity within the protein regions available for comparison.

Amino acid sequence *Molecular mass* *Secondary structure* *α-Helix* *Protein homology* Bacillus stearothermophilus

1. INTRODUCTION

As revealed by immuno-electron microscopy, protein S14 is located in the head of the 30 S particle together with proteins S3, S9, S10, S13 and S19 [1,2]. The finding that protein S14 interacts with proteins S9, S10 and S19 during the assembly of the 30 S subunit [3] is in good agreement with the immuno-electron microscopic results. Furthermore, protein S14 can be crosslinked in situ to proteins S19 [4] and S21 [5] as well as to initiation factors [6]. Affinity labeling studies showed that protein S14 is a part of the binding site for the 3′-end of tRNA on the 30 S subunit [7,8] and that it reacts with puromycin [9] which is an analogue of the 3′-end of tRNAs.

Here, we report the complete determination of the primary structure of protein S14 which consists of 98 amino acids and has a relative molecular mass (M_r) of 11191. Based on the amino acid sequence the secondary structure of the protein is predicted using appropriate computer programs. Furthermore, the primary structure of protein S14 is compared with that of all other ribosomal proteins whose sequences are known.

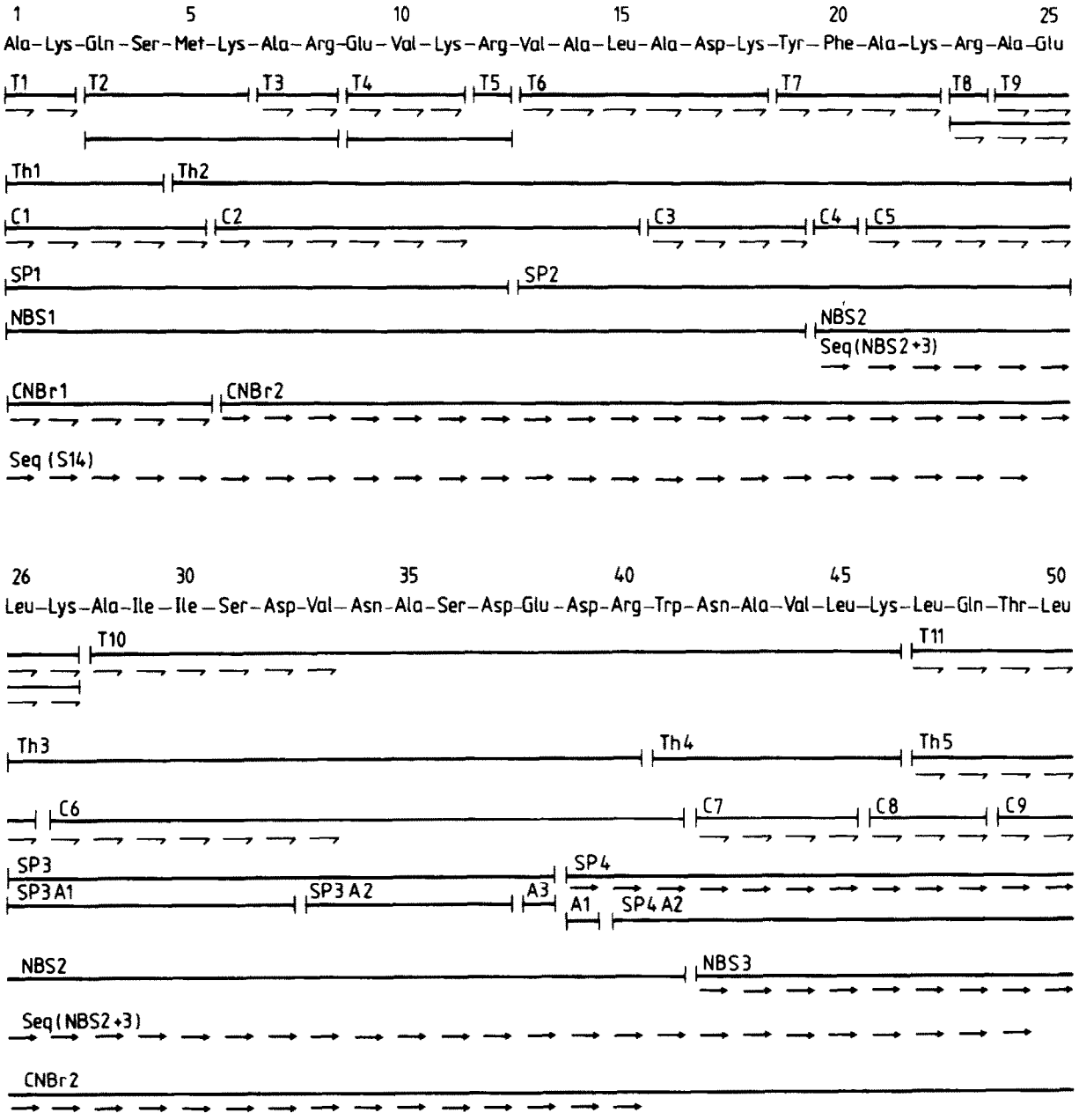
2. MATERIALS AND METHODS

Protein S14 was isolated from *E. coli* K as in [10]. The identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [11].

Performic acid oxidation [12] was done at 0°C for 16 h. Tryptic and chymotryptic digestions were at pH 8.0 and 37°C for 4 h. The thermolytic digestion was at pH 8.0 and 55°C for 4 h. Digestion

* Present address: Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Published by Elsevier Science Publishers
00145793/83/0000–0000/$3.00



with *Staphylococcus aureus* protease was in 50 mM acetic acid (pH 4.0) for 16 h [13]. Large fragments were produced by cyanogen bromide [14], *N*-bromosuccinimide [15], and acetic acid [16] cleavage.

The isolation of large peptides was achieved by gel filtration of various digests (~5 mg each) on Sephadex G-50 or G-75 (superfine) columns (250 × 1.5 cm). HCl (0.01 N) was used for the elution. Smaller peptides in the various fractions eluted from the Sephadex columns were further separated by fingerprinting on cellulose thin-layer plates [17]. The presence of tryptophan-containing peptides on the thin-layer plates was detected with Ehrlich's reagent. Amino acid analyses were performed on a Durrum D-500 amino acid analyzer.

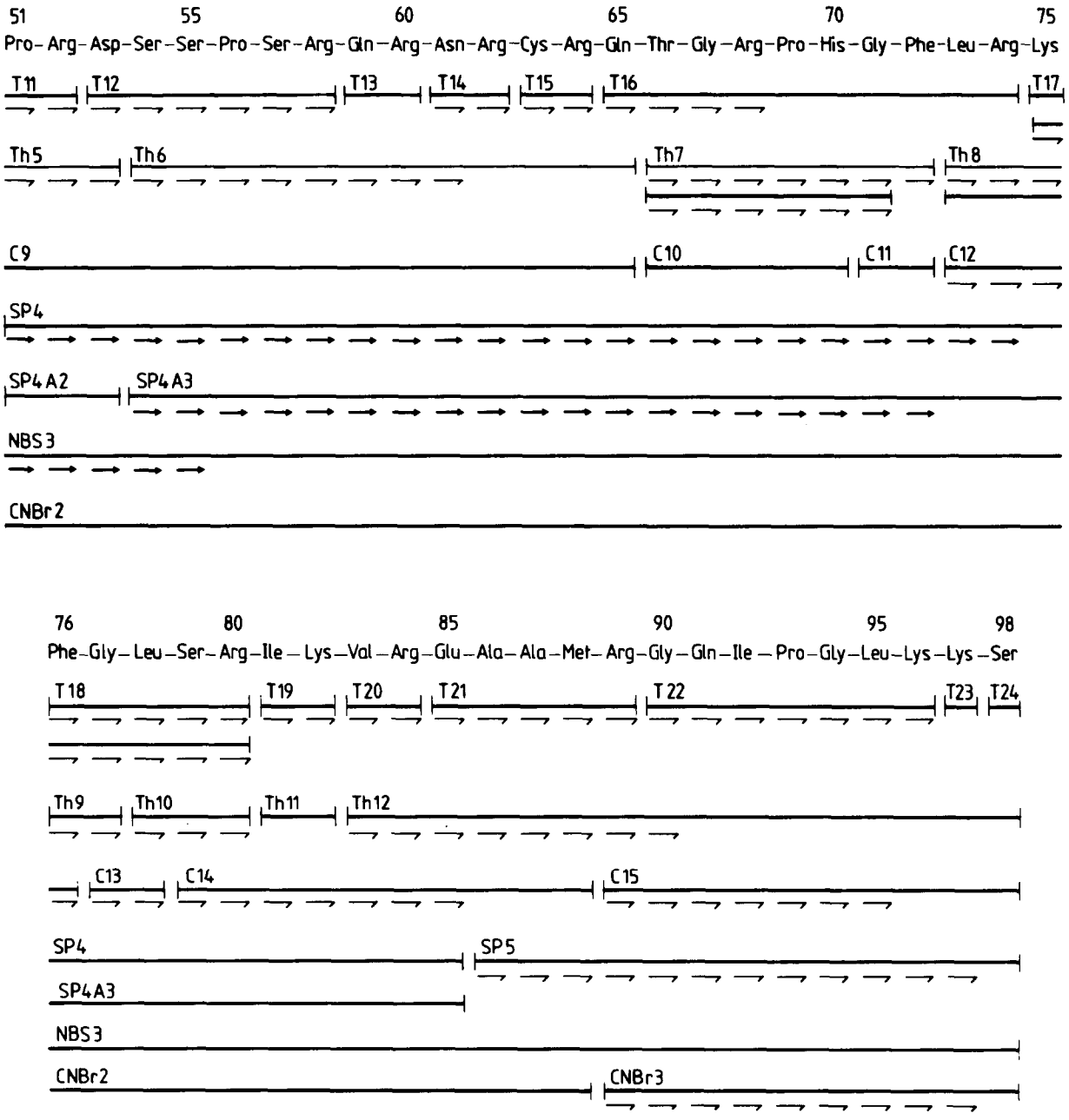


Fig.1. Amino acid sequence of protein S14 from *Escherichia coli* ribosomes: T, tryptic peptide; Th, thermolytic peptide; C, chymotryptic peptide; SP, peptide from digestion with *Staphylococcus aureus* protease; NBS, peptide cleaved with *N*-bromosuccinimide; CNBr, peptide cleaved with cyanogen bromide; ⟶, residue identified with the Beckman sequenator; ⟹, residue identified by manual Edman degradation.

Tryptophan was estimated by the method in [18].

Automatic Edman degradation [19] of protein S14 and its large peptides was made in a Beckman model 890C sequenator utilizing 0.5 M Quadrol protein program 122974 or DMAA peptide program 102974. Polybrene [20] was used as carrier in case of peptides. The amino acid sequence of small peptides was determined by a manual micro-

Edman technique [21] without dansylation. The thiazolinone or PTH derivatives were hydrolysed with 6 N HCl in the presence or absence of 0.1% $SnCl_2$ [22] at 130°C for 20 h, and the amino acid formed was analyzed on a Durrum analyzer. The identification of some PTH derivatives (Trp, Asp, Asn, Glu, Gln) was made by thin-layer chromatography on silica gel plates [23].

3. RESULTS AND DISCUSSION

Large fragments from protein S14 were produced by cyanogen bromide (CNBr2), *N*-bromosuccinimide (NBS2 + 3, NBS3) and *Staphylococcus aureus* protease (SP4). The fragment SP4 (res. 39–85) of the central region was further cleaved by acetic acid into 3 sub-fragments (SP4A1, SP4A2, SP4A3). The automatic sequence determination of the N-terminal regions of protein S14 and the large fragments provided sufficient overlaps in the amino acid sequence of the first 74 residues as shown in fig.1.

All peptides produced by trypsin, chymotrypsin and thermolysin were isolated, and most of them were manually sequenced as shown in fig.1. The sequence of the C-terminal region (res. 73–98) was deduced from the results obtained by the manual sequence determination of the C-terminal cyanogen bromide fragment (CNBr3) and of the *S. aureus* protease fragment (SP5) in addition to tryptic, chymotryptic and thermolytic peptides in this region. The combination of these results allowed the alignment of all the peptides and the determination of the complete amino acid sequence of protein S14 as illustrated in fig.1.

Table 1

Amount of predicted secondary structure of protein S14 (in %)

Prediction according to	α-Helix	β-Sheet	β-Turn	Random coil
[25]	40	0	26	34
[26,27]	28–40	19–32	25	16
[28]	30–48	5–19	29	22
[29]	53	2	34	11

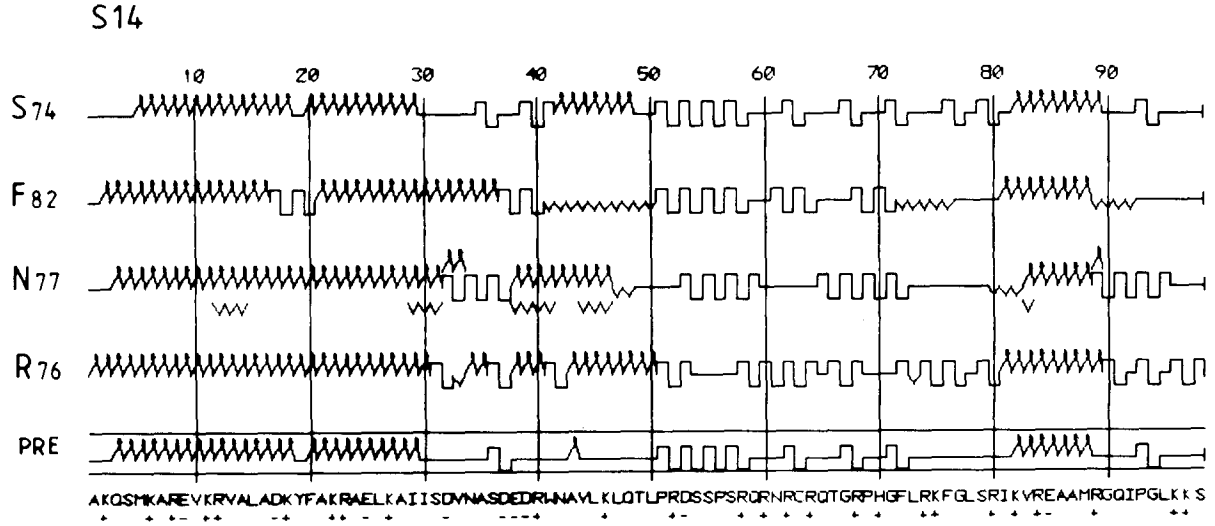


Fig.2. Secondary structure of protein S14 as predicted according to 4 different methods: S, [25]; F, [26] computed as in [27]; N, [28]; R, [29]. PRE summarizes the secondary structure obtained when at least 3 out of the 4 predictions are in agreement. The amino acid sequence of protein S14 is shown in the bottom line in the one-letter code. For details see [24,27] and table 1.

Table 2

Homologous sequences between protein S14 and other ribosomal proteins of *Escherichia coli*

Protein	Positions	Identical or Homologous Sequences
S14	6- 10	Lys-Ala-Arg-Glu-Val
S5	51- 55	Lys-Ala-Arg-Glu-Val
S14	78- 92	Leu-Ser-Arg-Ile-Lys-Val-Arg-Glu-Ala-Ala-Met-Arg-Gly-Gln-Ile
S8	10- 24	Leu-Thr-Arg-Ile-Arg-Asn-Gly-Gln-Ala-Ala-Asn-Lys-Ala-Ala-Val
S14	80- 88	Arg-Ile-Lys-Val-Arg-Glu-Ala-Ala-Met
S10	5- 13	Arg-Ile-Arg-Ile-Arg-Leu-Lys-Ala-Phe
S14	1- 17	Ala-Lys---------Gln-Ser-Met-Lys-Ala-Arg-Glu-Val-Lys-Arg-Val-Ala-Leu-Ala-Asp
S12	1- 19	Ala-Thr-Val-Asn-Gln-Leu-Val-Arg-Lys-Pro-Arg-Ala-Arg-Lys-Val-Ala-Lys-Ser-Asn
S14	81- 85	Ile-Lys-Val-Arg-Glu
S21	3- 7	Ile-Lys-Val-Arg-Glu
S14	9- 13	Glu-Val-Lys-Arg-Val
L3	103-107	Asp-Val-Lys-Lys-Val
S14	24- 29	Ala-Glu-Leu-Lys-Ala-Ile
L6	145-150	Ala-Asp-Leu-Arg-Ala-Tyr
S14	21- 26	Ala-Lys-Arg-Ala-Glu-Leu
L9	49- 54	Ala-Arg-Arg-Ala-Glu-Leu
S14	26- 30	Leu-Lys-Ala-Ile-Ile
L14	58- 62	Leu-Lys-Ala-Val-Val
S14	9- 12	Glu-Val-Lys-Arg
L25	7- 10	Glu-Val-Arg-Lys
S14	24- 27	Ala-Glu-Leu-Lys
L28	68- 71	Ala-Glu-Leu-Arg
S14	5- 27	Met-Lys-Ala-Arg-Glu-Val-Lys-Arg-Val-Ala-Leu-Ala-Asp-Lys-Tyr-Phe-Ala-Lys-Arg-Ala-Glu-Leu-Lys
L29	1- 23	Met-Lys-Ala-Lys-Glu-Leu-Arg-Glu-Lys-Ser-Val-Glu-Glu-Leu-Asn-Thr-Glu-Leu-Leu-Asn-Leu-Leu-Arg
S14	47- 66	Leu-----Gln-Thr-Leu-Pro-Arg-Asp-Ser-Ser-Pro-Ser-Arg-Gln-Arg-Asn-Arg-Cys-Arg-Gln-Thr
L29	43- CT	Leu-Lys-Gln-Val-Arg-Arg-Asp-Val-Ala-Arg-Val-Lys-Thr-Leu-Leu-Asn-Glu-Lys-Ala-Gly-Ala

The amino acid sequence given in fig.1 is in agreement with the DNA sequence of the gene for protein S14 (M. Nomura, personal communication) with the exception of the last residue at position 98. As shown in fig.1, the C-terminal amino acid is serine whereas glycine has been deduced from the DNA sequence. The amino acid analysis of the C-terminal peptides (Th12, C15, SP5,

Table 3

Comparison of protein S14 with all other *E. coli* ribosomal (and related) proteins

The search for homologous sequence stretches was performed with the FORTRAN program SEEK (M. Dzionara, K. Ashman and B. Wittmann-Liebold, unpublished) by comparing all possible segments of protein S14 with all other protein sequences. The segment lengths employed were 15 (table 3a) and 10 residues (table 3b).

Symbols beneath the sequence stretches denote: identical amino acids in identical positions within the segment compared (labelled by a star); non-identical amino acids: related by one nucleotide replacement within their codons (marked by a plus); related by two nucleotide replacements (marked by a minus); unrelated amino acids (not labelled).

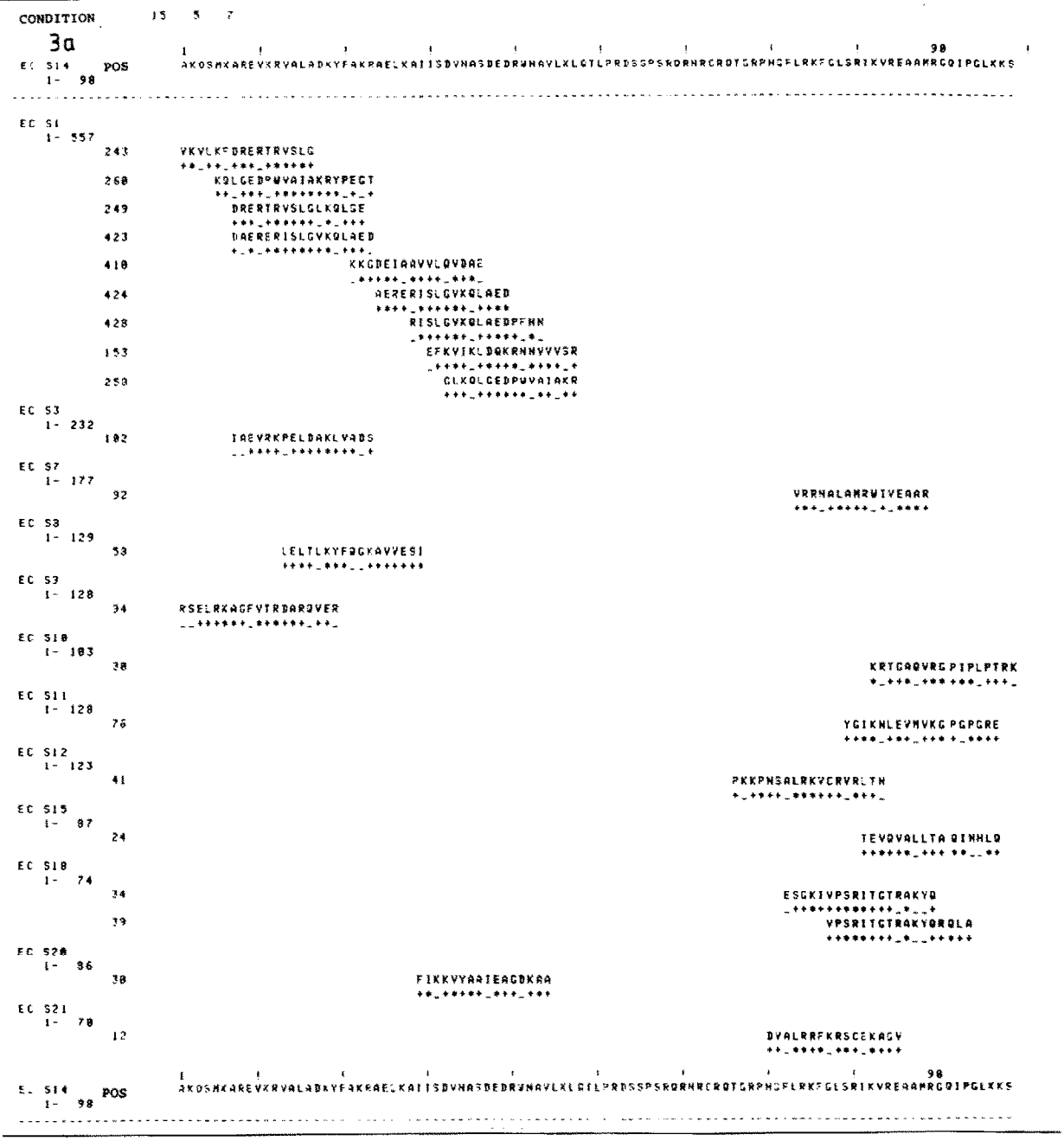
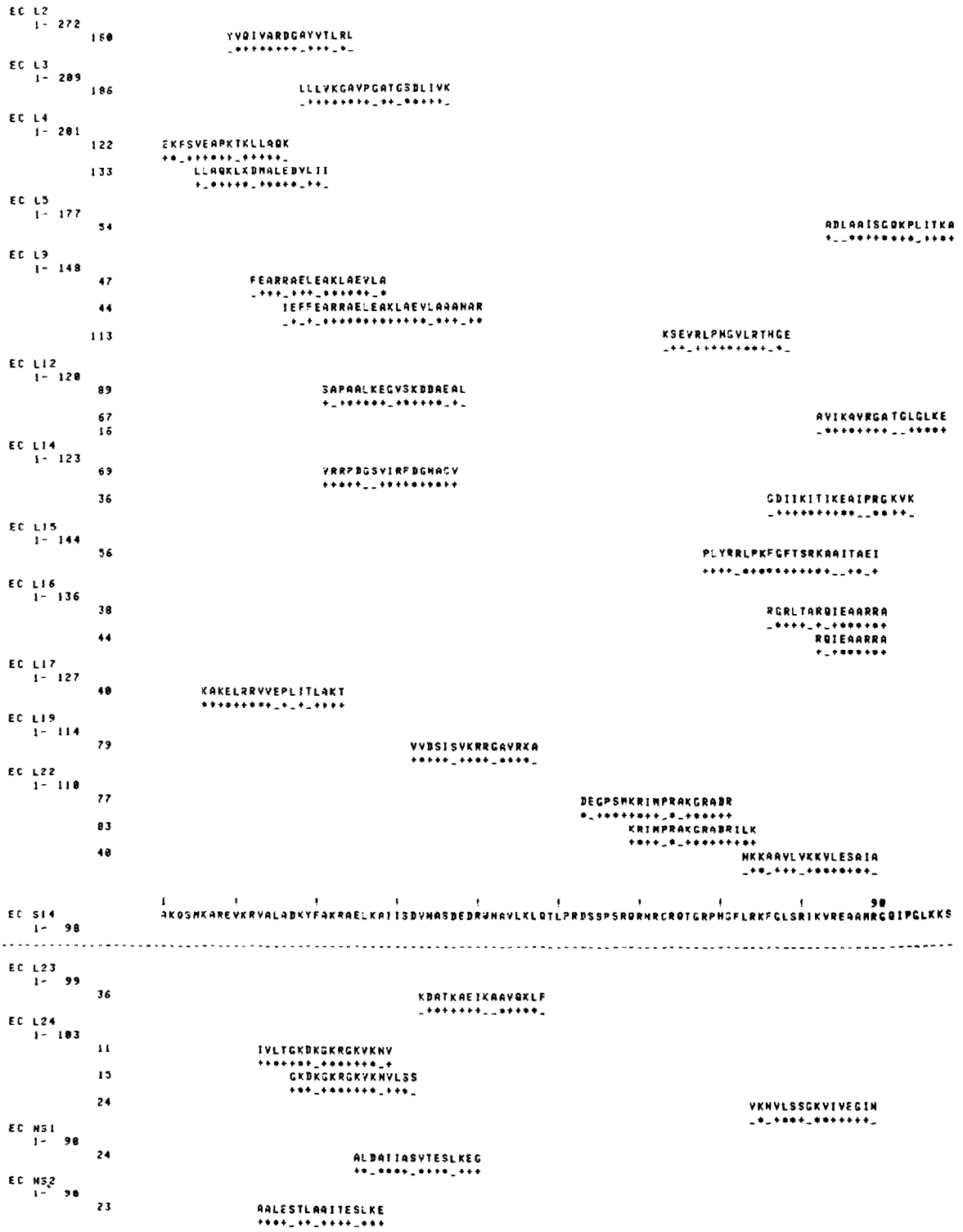


Table 3a (continued)

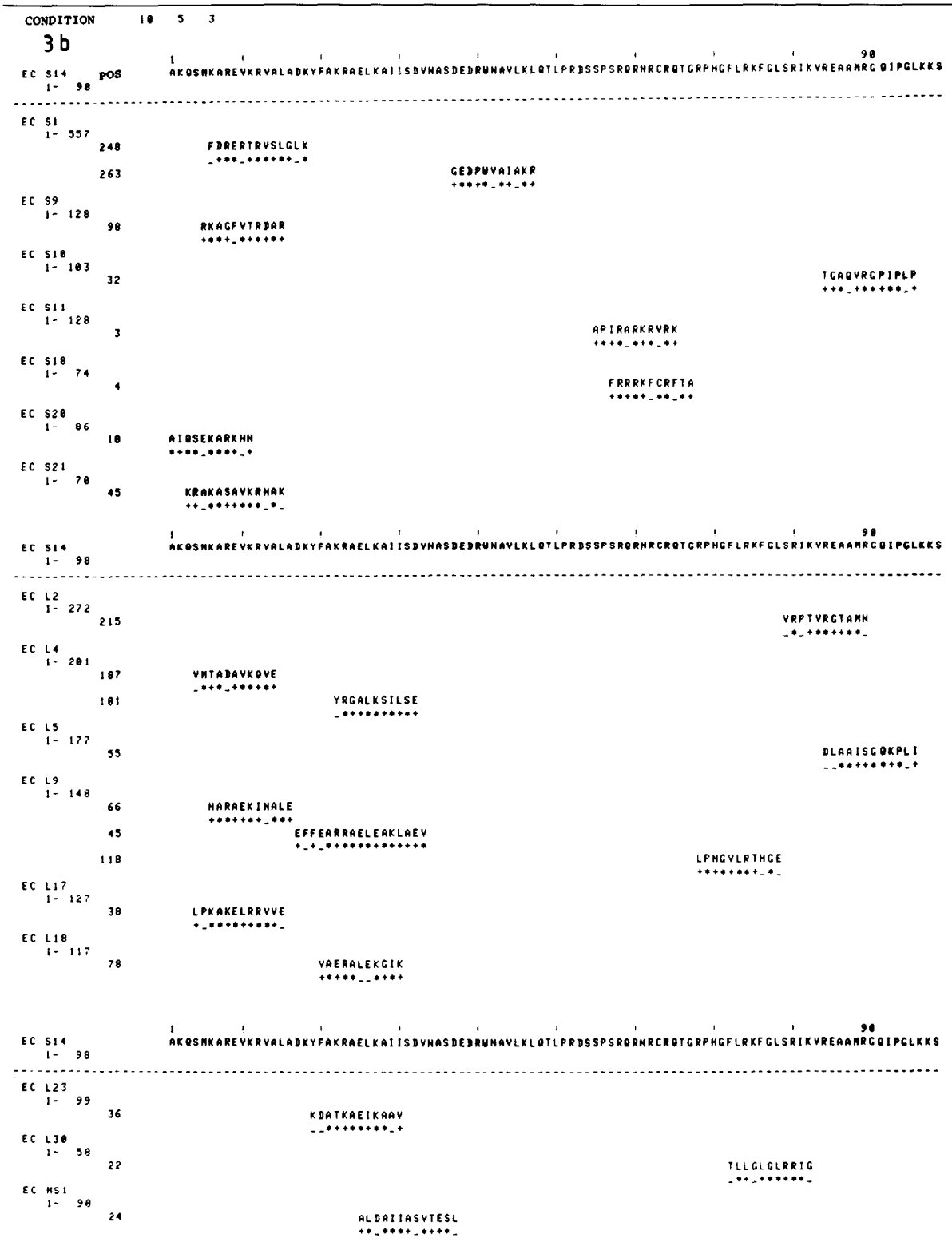


The homologous stretches listed in table 3a were found under the following conditions:

(i) At least 5 identical amino acids at identical positions within a 15-residues segment;

(ii) At least 7 out of the non-identical amino acids of the segment must be related by one nucleotide replacement within their codons.

Table 3b



The stretches given in table 3b were found under more strict conditions:

(i) At least 5 identical out of 10 residues in identical positions of the segments;

(ii) At least 3 of the non-identicals within 10 residues must be related (as above).

CNBr3) demonstrates that these peptides contain 2 glycine and 1 serine residues, and their amino acid sequence analysis indicates that serine is the C-terminal residue. If glycine were at the C-terminus, as deduced from the DNA sequence, these C-terminal peptides should contain 3 glycines and no serine. It is possible that this discrepancy (serine or glycine at the C-terminus) is due to the differences in the *E. coli* strains from which protein S14 or the gene for this protein were isolated. The codons for serine (GG^{U}_{C}) and glycine (AG^{U}_{C}) differ by one nucleotide only.

The amino acid composition derived from the sequence is:

Asp_5, Asn_3, Thr_2, Ser_8, Glu_4, Gln_5, Pro_4, Gly_5, Ala_{11}, Cys_1, Val_5, Met_2, Ile_4, Leu_8, Tyr_1, Phe_3, His_1, Lys_{11}, Arg_{14}, Trp_1

This is in excellent agreement with the data determined from the hydrolysis of the whole protein. The protein consists of 98 amino acid residues, and the M_r calculated from the above composition is 11191. Protein S14 contains 26 basic residues (His + Lys + Arg) and only 9 acidic residues (Asp + Glu), and therefore it is a very basic protein.

Based on 4 programs for the prediction of the secondary structure of proteins (details in [24]) a diagram has been drawn for the secondary structure of protein S14 (fig.2). A long α-helical region of 27–31 residues at the N-terminus (positions 3–29) and a short α-helix consisting of about 7 residues (positions 82–88) near the C-terminus are predicted. Table 1 shows the calculated secondary structure of protein S14 according to the different prediction methods.

Computer searches for homologous regions between protein S14 on the one hand and other ribosomal proteins and related sequences on the other hand gave the results listed in tables 2 and 3. No extended regions of high significance were found. Only some short homologous stretches exist, mainly in the N-terminal (up to position 40) and C-terminal regions (positions 70–98) of protein S14.

A comparison of the primary structure of ribosomal proteins from *Escherichia coli* and other bacteria revealed the following homology between the N-terminal regions of protein S14 from (1) *E. coli* and of the corresponding protein from (2) *Bacillus stearothermophilus* [30]:

(1) AKQSMKAREVKRVALADKYFAKRAELK[27]
(2) AKKSMIAKQ-KRTPKFKVRAYTRTERR

ACKNOWLEDGEMENTS

We thank Joan Logan for excellent technical assistance. This is NRCC publication no. 20892.

REFERENCES

[1] Stöffler-Meilicke, M. and Stöffler, G. (1982) in: Electron Microscopy, vol.3, pp.99–100, Electron Microsc., Hamburg.
[2] Winkelmann, D.A., Kahan, L. and Lake, J.A. (1982) Proc. Natl. Acad. Sci. USA 79, 5184–5188.
[3] Nomura, M. and Held, W.A. (1974) in: Ribosomes (Nomura, M. et al. eds) pp.193–223, Cold Spring Harbor Lab. Press, Long Island NY.
[4] Bode, U., Lutter, L.C. and Stöffler, G. (1974) FEBS Lett. 45, 232–236.
[5] Traut, R.R., Lambert, J.M., Boileau, G. and Kenny, J.W. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp.89–110, University Park Press, Baltimore MD.
[6] Bollen, A., Heimark, R.L., Cazzone, A., Traut, R.R., Hershey, J.W. and Kahan, L. (1975) J. Biol. Chem. 250, 4310–4314.
[7] Girshovich, A.S., Bochkareva, E.S., Kramarov, V.A. and Ovchinnikov, Y.A. (1974) FEBS Lett. 42, 213–217.
[8] Pellegrini, M., Oen, H., Eilat, D. and Cantor, C.R. (1974) J. Mol. Biol. 88, 809–829.
[9] Cooperman, B.S. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp.531–554, University Park Press, Baltimore MD.
[10] Hindennach, I., Stöffler, G. and Wittmann, H.G. (1971) Eur. J. Biochem. 23, 7–11.
[11] Kaltschmidt, E. and Wittmann, H.G. (1970) Analyt. Biochem. 36, 401–412.
[12] Hirs, C.H.W. (1956) J. Biol. Chem. 219, 611–621.
[13] Houmard, J. and Drapeau, G.R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[14] Gross, E. (1967) Methods Enzymol. 11, 238–255.
[15] Ramachandran, L.K. and Witkop, B. (1967) Methods Enzymol. 11, 283–299.
[16] Tsung, C.M. and Fraenkel-Conrat, H. (1965) Biochemistry 4, 793–801.
[17] Yaguchi, M., Wittmann, H.G., Cabezon, T., DeWilde, M., Villarroel, R., Herzog, A. and Bollen, A. (1975) Mol. Gen. Genet. 142, 35–43.

[18] Liu, T.Y. and Chang, Y.H. (1971) J. Biol. Chem. 248, 2842–2848.
[19] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[20] Klapper, D.G., Wilde, G.E. and Capra, J.D. (1978) Analyt. Biochem. 85, 126–131.
[21] Bruton, C.J. and Hartley, B.S. (1970) J. Mol. Biol. 52, 165–178.
[22] Mendez, E. and Lai, S.Y. (1975) Analyt. Biochem. 68, 47–53.
[23] Wittmann-Liebold, B., Geissler, A.-W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[24] Dzionara, M., Robinson, S.M.L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1019–1103.
[25] Burgess, A.W., Ponuswamy, P.K. and Scheraga, H.A. (1974) Israel J. Chem. 12, 239–286.
[26] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[27] Rawlings, N., Ashman, K. and Wittmann-Liebold, B. (1983) Prot. Pept. Res., in press.
[28] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[29] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[30] Matheson, A.T., Möller, W., Amons, R. and Yaguchi, M. (1980) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp.297–332, University Park Press, Baltimore MD.