STUDIES ON THE PRIMARY STRUCTURE OF 20 PROTEINS FROM *ESCHERICHIA COLI* RIBOSOMES BY MEANS OF AN IMPROVED PROTEIN SEQUENATOR*

Brigitte WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Berlin-Dahlem, Germany

Received 6 August 1973
Originals received 20 September 1973

1. Introduction

Ribosomes are very complex organelles. *E. coli* ribosomes, for instance, consist of 55 proteins and three RNA molecules. Knowledge of the amino acid sequences of ribosomal proteins is useful for studies on: a) mutants with altered ribosomal proteins; b) protein–RNA interaction; c) homologous structures among ribosomal proteins; d) the topography of ribosomes and e) the tertiary structure of ribosomal proteins.

The protein sequenator developed by Edman and Begg [1] facilitates studies on the primary structure of proteins and supplements conventional methods for determination of amino acid sequences. The Beckman sequencer was considerably improved and equipped with a device for automatic conversion of thiazolinones into PTH amino acids [2]. This paper summarizes the results which were obtained with the improved apparatus in studies on the amino acid sequence of 20 ribosomal proteins from *E. coli*.

2. Materials and methods

Ribosomal proteins were isolated from *E. coli* strain K as previously described [3] and were obtained from Dr. H.G. Wittmann. Lyophilized proteins (4–6 mg each) were dissolved in water and subjected to Edman degradation in a Beckman sequencer which was considerably modified and equipped with a device for automatic conversion of thiazolinones into PTH amino acids [2, 4]. Identification of PTH amino acids was made by thin layer chromatography in several solvent systems [2], by gas chromatography and recently by mass spectrometry with a Varian CH7.

3. Results and discussion

The results obtained by Edman degradation of 20 ribosomal proteins are given in fig. 1. The 18 proteins designated with S belong to the small and the two designated with L to the large subunits [5]. Two proteins from the small subunit, namely S5 and S18, do not yield N-terminal PTH amino acids when analyzed by the sequenator. They probably have blocked N-termini, as has already been found [6] for protein L7. From the 18 analyzed proteins of the small subunit, seven start with alanine, three with methionine, three with proline, two with serine and one each with either threonine, glycine or phenylalanine. The occurrence of methionine as the N-terminal amino acid in 30 S proteins is far less than expected from earlier end group determinations [7, 8]. Methionine, then, occurs more frequently as N-terminus in 50 S proteins than in 30 S proteins. This conclusion agrees with the analysis of end groups in total proteins from both subunits (M. Yaguchi, personal communication).

Comparison of the amino acid sequences (fig. 1) shows that the analyzed regions of the 20 proteins differ markedly in their primary structures. This is in full agreement with immunological studies [9]. It follows

* Paper No. 78 on 'Ribosomal Proteins'. Preceding paper is by E. Deusser et al., J. Mol. Biol. submitted.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
S2	ALA	THR	VAL	SER	MET	ARG	ASP	MET	LEU	LYS	ALA	GLY	VAL	HIS	PHE	GLY	HIS	GLN	THR	ARG	TYR	TRP	ASN	PRO	LYS	MET	LYS	PRO	PHE	L·I	PHE	GLY
S3	GLY	GLN	LYS	VAL	HIS	PRO	ASN	GLY	ILE	ARG	LEU	GLY	ILE	VAL	LYS	PRO	GLY	ASN	SER	THR	GLY	PHE	ALA	ASN	THR	LYS	GLU	PHE	ALA	ASP	ASN	LEU
S4	ALA	ARG	TYR	LEU	GLY	PRO	LYS	LEU	LYS	LEU	SER	ARG	ARG	GLU	GLY	THR	ASP	LEU	PHE	LEU	LYS	SER	GLY	VAL	ARG	ALA	ILE	ASP	THR	LYS	CYS	LYS
S6	MET	ARG	HIS	TYR	GLU	ILE	VAL	PHE	MET	VAL	HIS	PRO	ASP	GLN	SER	GLU	GLN	VAL	PRO	GLY	MET	ILE	GLU	ARG	TYR	THR	ALA	ALA	ILE	THR	GLY	ALA
S7	PRO	LYS	PHE	(SER)	VAL	LEU	GLY	GLN	ALA	LYS	PHE	(ASN)	PRO	ASP	PRO	MET	PHE	GLY	SER	GLU	LEU	LEU	ALA	LYS	PHE	VAL	ASN	VAL	LEU	MET	VAL	ALA
S8	SER	MET	GLN	ASP	PRO	ILE	ALA	ASP	MET	LEU	THR	ARG	ILE	ARG	ASN	GLY	GLN	ALA	ALA	ASN	LYS	ALA	ALA	VAL	THR	MET	PRO	SER	SER	LYS	LEU	LYS
S9	ALA	GLU	ASN	GLN	TYR	TYR	GLY	THR	GLY	ARG	ARG	LYS	SER	SER	ALA	ALA	ARG	VAL	PHE	ILE	LYS	PRO	GLY	ASN	GLY	LYS	ILE	VAL	ILE	ASN	GLN	ARG
S10	MET	GLN	ASN	GLN	ARG	ILE	VAL	PHE	MET	VAL	LEU	ALA	PHE	PHE	ARG	HIS	LEU	(ASP	ILE)	GLN	ALA	THR	ALA	GLU	LEU	VAL	GLU	THR	ALA	(ALA)	GLU	(THR)
S11	PHE	LYS	ALA	PRO	L·I	ARG	ALA	ARG	LYS	ARG	VAL	ARG	LYS	GLN	VAL	SER	(ARG)	GLY	VAL	ALA	HIS	L·I	(HIS)	ALA	SER	PHE	ASN	ASN	THR	L·I	VAL	(THR)
S12	ALA	THR	VAL	ASN	GLN	LEU	VAL	ARG	LYS	PRO	ARG	ALA	ARG	LYS	VAL	ALA	LYS	(LYS)	ASN	VAL	PRO	ALA	LEU	GLU	ALA	CYS	PRO	GLX	SER	ARG	GLY	VAL
S13	ALA	ARG	ILE	ALA	GLY	ILE	ASN	ILE	PRO	ASP	HIS	LYS	HIS	ALA	VAL	ILE	ALA	LEU	THR	SER	ILE	TYR	GLY	VAL	GLY	LYS	THR	ARG	SER	LYS	ALA	L·I
S14	ALA	LYS	GLN	SER	MET	LYS	ALA	ARG	GLU	VAL	LYS	ARG	VAL	ALA	LEU	ALA	ASP	LYS	TYR	PHE	ALA	ALA	LYS	ALA	GLU	L·I	(L·I)	LYS	ALA	L·I	(SER)	ARG
S15	SER	LEU	SER	THR	GLU	ALA	THR	ALA	LYS	ILE	VAL	SER	GLU	PHE	GLY	ARG	ASP	ALA	ASN	ASP	THR	GLY	SER	THR	GLU	VAL	GLN	VAL	ALA	LEU	LEU	THR
S16	MET	VAL	THR	ILE	ARG	LEU	ALA	ARG	HIS	GLY	ALA	LYS	LYS	ARG	PRO	PHE	TYR	GLN	VAL	VAL	VAL	ALA	ASP	SER	ARG	ASN	ALA	ARG	ASN	GLY	ARG	PHE
S17	THR	ASP	LYS	ILE	ARG	THR	LEU	GLN	GLY	ARG	VAL	VAL	SER	ASP	LYS	MET	GLU	LYS	(SER)	L·I	VAL	VAL	ALA	L·I	GLU	ARG	PHE	VAL	LYS	HIS	PRO	L·I
S19	PRO	ARG	SER	L·I	LYS	LYS	GLY	PRO	PHE	L·I	ASP	L·I	HIS	L·I	L·I	LYS	LYS	VAL	GLU	LYS	ALA	VAL	GLU	SER	GLY	ASP	LYS	LYS	PRO	L·I	(PRO)	ARG
S20	ALA	ASN	ILE	LYS	SER	ALA	LYS	LYS	ARG	ALA	ILE	GLN	SER	GLU	LYS	ALA	ARG	LYS	HIS	ASN	ALA	SER	ALA	ARG	SER	MET	MET	ARG	THR	PHE	ILE	LYS
S21	PRO	VAL	ILE	LYS	VAL	ARG	GLU	ASN	GLU	PRO	PHE	ASP	VAL	ALA	LEU	ARG	ARG	PHE	LYS	ARG	SER	CYS	GLU	LYS	ALA	GLY	VAL	LEU	ALA	GLU	VAL	ARG
L12	SER	ILE	THR	LYS	ASP	GLN	ILE	ILE	GLU	ALA	VAL	ALA	ALA	MET	SER	VAL	MET	ASP	VAL	VAL	GLU	LEU	ILE	SER	ALA	MET	GLU	GLU	LYS	PHE	GLY	VAL
L24	ALA	ALA	LYS	ILE	ARG	ARG	ASP	ASP	GLU	VAL	LEU	VAL	ILE	THR	GLY	LYS	ASP	LYS	GLY	LYS	ARG	GLY	LYS	VAL	LYS	ASN	VAL	LEU	SER	SER	GLY	LYS

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
S2	ALA	ALA	ASN																													
S3	ASP	SER	ASP	PHE	LYS	VAL	ARG	GLN	TYR	LEU	THR	LYS	GLU	LEU	ALA	LYS	ALA															
S4	ILE	GLU	GLN	ALA	PRO	GLY	GLN	HIS	GLY	ALA	ARG	LYS	PRO	ARG	LEU	SER	ASP	TYR	GLY													
S6	GLU	GLY	LYS	L·I	(L·I)	HIS	L·I	GLU	ASP	GLY																						
S8	VAL	ALA																														
S9	SER	LEU	GLU	GLN	TYR	PHE	GLY	ARG	GLU	ALA	ALA	ALA	MET	ARG	VAL	VAL	GLN	PRO	SER	GLU	LEU	VAL	MET	ARG	ARG							
S10	GLY	ALA	GLN	LEU	VAL	GLY	PRO	PHE	(LEU)	LEU																						
S12	CYS	THR	ARG	VAL	TYR	THR	THR	THR	PRO	LYS	LYS	PRO	ASN	SER	ALA	LEU	ARG															
S13	LEU	ALA	ALA	ALA	GLY	L·I	ALA	ARG	GLU	VAL	LYS	LEU	THR	THR	LEU	GLU	SER	GLY	GLN	L·I	ARG	THR	L·I	LEU	ARG	GLU	VAL	ALA				
S14	ASN	VAL	(ASN)	ALA																												
S15	ALA	GLN	ILE	ASN	HIS	LEU	GLN	GLY	GLY	PHE	ALA	GLU	ALA	GLU	(ARG	LYS	LYS	ARG	SER	ARG	SER)	PHE	(GLY	L·I)	L·I	MET	VAL						
S16	ILE	GLU	ARG	VAL	GLY	PHE	PHE	ASN	PRO	ILE	ALA	SER	GLU	LYS	GLU	GLU	GLY	THR	ARG	LEU	ASP	LEU	ASP	ARG								
S17	TYR	GLY																														
S19	VAL	ASN	ALA	SER																												
S20	LYS	VAL	TYR	ALA	ALA	ASP	ILE	GLU	ALA	GLU	LYS	ARG	ALA	ALA	GLN	LYS	ALA															
S21	ARG	ARG	GLU	PHE	TYR	GLU	LYS	PRO	ARG	THR	GLU	THR	ARG	LYS	ALA	LYS	ALA	SER	ALA	VAL	LYS											
L12	SER	ALA	ALA	ALA	ALA	VAL	ALA	VAL	ALA	ALA	GLY	PRO	VAL	GLU	ALA	ALA	GLU	GLU	LYS													
L24	VAL	ILE	VAL	GLU	GLY	ILE	ASN	LEU	VAL	LYS	LYS																					

Fig. 1. Amino acid sequence of the N-terminal regions of *E. coli* ribosomal proteins. L·I means leucine or isoleucine.

Table 1
Comparison of identical regions in ribosomal proteins.

Protein	Sequence	Position
S6	Ile–Val–Phe–Met–Val	6–10
S10	Ile–Val–Phe–Met–Val	6–10
S6	Thr–Ala–Ala–Ile –Thr–Gly–Ala	26–32
S10	Thr–Ala–Ala–Glu–Thr–Gly–Ala	28–34
S4	Arg–Lys–Pro–Arg	43–46
S12	Arg–Lys–Pro–Arg	8–11
S21	Lys–Pro–Arg	
S16	Ala–Lys–Lys–Arg	11–14
S20	Ala–Lys–Lys–Arg	6–9
S3	Lys–Pro–Gly–Asn–Ser*	15–19
S9	Lys–Pro–Gly–Asn–	21–24
S14	Val–Ala–Leu	13–15
S21	Val–Ala–Leu	13–15
S15	Val–Ala–Leu	28–30
S13	Ala–Ala–Ala	34–36
L12	Ala–Ala–Ala	34–36
S9	Ala–Ala–Ala	42–44
S2	Ala–Thr–Val	1–3
S12	Ala–Thr–Val	1–3
S17	Lys–Ile –Arg	3–5
L24	Lys–Ile –Arg	3–5

* A similar sequence is present in positions 43–46 of protein S12.

therefore that all ribosomal proteins (except L7 and L12) have unique structures different from each other.

Nevertheless, there are a number of small identical regions among the analyzed proteins as listed in table 1. For instance, proteins S6 and S10 not only start with the same amino acid, methionine, but they also have two (almost) identical regions: Ile–Val–Phe–Met–Val (positions 6–10 in S6 and S10) as well as Thr–Ala–Ala–Ile–Thr–Gly–Ala (positions 26–32 in S6) and Thr–Ala–Ala–Glu–Thr–Gly–Ala (positions 28–34 in S10). These similarities suggest a structural relationship between proteins S6 and S10 and it remains to be seen whether similar functions can be found for the two proteins. Further examples for identical regions among some ribosomal proteins are given in table 1.

Most of the ribosomal proteins are very rich in basic amino acids (10–12). Therefore regions with a relatively high number of arginines and lysines can be expected. Nevertheless some clusters of basic amino acids are remarkable, e.g. the sequences Arg–Lys–Lys–Arg–Ser–Arg (positions 47–52 in S15), Arg–Lys–Pro–Arg–Ala–Arg–Lys–Val–Ala–Lys–Lys (positions 8–18 in S12), Arg–Arg–Phe–Lys–Arg (positions 16–20 in S21), Arg–Ala–Arg–Lys–Arg–Val–Arg–Lys (positions 6–13 in S11), Lys–Ser–Ala–Lys–Lys–Arg (positions 4–9 in S20), Arg–Arg–Arg (positions 32–34 in S21), Arg–Arg–Lys (positions 10–12 in S9) and Lys–Lys–Arg (positions 12–14 in S16).

Apolar amino acids are clustered in several regions, namely positions 31–39 in protein S13, positions 34–43 in L12, positions 26–32 in S7 and positions 6–14 in S10. It can be assumed that these regions are located in the interior of the molecules but this assumption of course needs direct proof by studies on the tertiary structure of the ribosomal proteins, e.g. by X-ray analysis. Knowledge of the amino acid sequence of these proteins is one of the prerequisites for such studies.

Acknowledgements

I thank Mrs. E. Marzinzig and Mr. H. Graffumder for their excellent help in this work which was supported by a grant from the Deutsche Forschungsgemeinschaft.

References

[1] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80.
[2] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem., in press.
[3] Hindennach, I., Stoffler, G. and Wittmann, H.G. (1971) Eur. J. Biochem. 23, 7.
[4] Graffunder, H. and Wittmann-Liebold, B., manuscript in preparation.
[5] Kaltschmidt, E. and Wittmann, H.G. (1970) Proc. Natl. Acad. Sci. U.S. 67, 1276.
[6] Terhorst, C.P., Möller, W., Laursen, R. and Wittmann-Liebold, B. (1972) FEBS Letters 28, 325.
[7] Waller, P.J. (1964) J. Mol. Biol. 10, 319.
[8] Wittmann, H.G., Söffler, G., Kaltschmidt, E., Rudloff, V., Janda, H.G., Dzionara, M., Donner, D., Nierhaus, K., Cech, M., Hindennach, I. and Wittmann-Liebold, B. (1970) FEBS Symp. 21, 33.
[9] Stöffler, G. and Wittmann, H.G. (1971) Proc. Natl. Acad. Sci. U.S. 68, 2283.
[10] Craven, G.R., Voynow, P., Hardy, S.J.S. and Kurland, C.G. (1969) Biochemistry 8, 2906.
[11] Kaltschmidt, E., Dzionara, M. and Wittmann, H.G. (1970) Mol. Gen. Genet. 109, 292.
[12] Mora, G., Donner, D., Thammana, P., Lutter, L. and Kurland, C.G. (1971) Mol. Gen. Genet. 112, 229.