PRIMARY STRUCTURES OF TWO HOMOLOGOUS RIBOSOME-ASSOCIATED DNA-BINDING PROTEINS OF *ESCHERICHIA COLI*

Liane MENDE, Beate TIMM and Alap R. SUBRAMANIAN
Max-Planck-Institut fur Molekulare Genetik, Abt. Wittmann, D-1000 Berlin-Dahlem, Germany

Received 30 October 1978

1. Introduction

The proteins NS1 and NS2 (NS standing for native subunit) bind specifically to the native 30 S subunit of *Escherichia coli* ribosomes [1]. The two proteins exist in solution as tetramers and have a monomer mol. wt ~9500. The proteins show high similarity in amino acid composition, and they are encoded in separate genes on the *E. coli* chromosome [1].

Electrophoretic and immunological analyses demonstrated that the DNA-binding protein HU [2] as well as the DNA-binding protein HD [3] correspond to mixtures of NS1 and NS2. The proteins belong to one of the four classes of DNA-binding proteins with regard to their molecular weights. It is highly plausible that NS1 and NS2 are two individual forms of a dimorphic DNA-binding protein which specifically binds to native 30 S subunits of the ribosome.

In this paper we present the complete amino acid sequences of the two proteins and show the structural relationship between them.

2. Materials and methods

Proteins NS1 and NS2 were isolated from *E. coli* strains MRE 600 and CP 78 grown in L-broth at 37°C. The conditions of cell-breakage and ribosome isolation were as in [4]. NS1 and NS2 were obtained from the 1 M NH_4Cl wash of ribosomes, using essentially the procedure outlined previously for their purification from native 30 S subunits [1]. The method yielded homogeneous proteins in yield of 0.8–1 mg/g ribosome. Further details and immunological studies with specific antisera will be published elsewhere.

Both proteins were digested with trypsin pretreated with 1-chloro-4-phenyl-3-tosylamidobutane-2-one (from Fa. Merck, Darmstadt or from Worthington Corp., Freehold, NJ) or with *Staphylococcus aureus* protease (Miles Biochemicals, Elkhart). The C-terminal amino acids were determined by digestion of the proteins with carboxypeptidase C (Boehringer, Mannheim). More details of the enzymatic digestions are given in [5,6]. Peptides with C-terminal arginine were obtained by blocking the lysines with ETPA (exo-*cis*-3,6-endoxoΔ4-tetra-hydrophtalic acid anhydride) and subsequent tryptic digestion. In addition, the arginines of NS2 were modified with 1,2-cyclohexandione (Pierce Chemicals, USA) and then digested with trypsin; in this case peptides with C-terminal lysines were obtained. Details of the modification of lysine and arginine residues are described in [7,8].

Small peptides were separated by peptide-mapping on cellulose thin-layer plates and detected by spraying with ninhydrin (Merck, Darmstadt) or fluorescamine (Roche, Basel). The lysine-blocked peptides of both proteins were separated by gel filtration on Sephadex G-50 (superfine) columns with 10% acetic acid. In some cases it was necessary to separate the peptide mixtures from the columns a second time by chromatography [5]. The two long insoluble SP-peptides of NS2 were separated on a G-25 (superfine) column (41 × 1 cm) equilibrated with formic acid : acetic acid : water (3 : 3 : 4).

Amino acid analyses of the proteins and peptides

were performed on a Durrum D-500 analyzer (Palo Alto, CA). Cysteine was determined as cysteic acid after performic acid oxidation of the proteins. For the detection of tryptophan some tryptic peptide maps were developed with Ehrlich's reagent (*p*-dimethylaminobenzaldehyde).

The amino sequences of the peptides were determined by the manual dansyl-Edman method [9,10]. In addition, some very long or hydrophobic peptides, or peptides in very small amounts, were sequenced by a modified Edman-degradation method developed in [11] using 4-*NN*-dimethylaminoazobenzene-4-isothiocyanate (DABITC, Fluka AG, Buchs) in combination with phenylisothiocyanate as coupling reagent.

3. Results

The complete amino acid sequence of NS1 is presented in fig.1 and that of NS2 in fig.2. Detailed results which led to the above sequences will be published elsewhere.

NS1 has a mol. wt 9225 and NS2 of 9535; both proteins consist of 90 amino acid residues. The amino acid compositions derived from the sequences are for protein NS1: Asp_6, Asn_3, Thr_5, Ser_5, Glu_4, Gln_2, Pro_2, Gly_8, Ala_{19}, Val_6, Met_1, Ile_7, Leu_5, Phe_3, Lys_9, Arg_5; and for protein NS2: Asp_3, Asn_4, Thr_7, Ser_4, Glu_7, Gln_4, Pro_2, Gly_6, Ala_{15}, Val_7, Met_1, Ile_5, Leu_7, Phe_3, His_1, Lys_{11}, Arg_3. These data are in good agreement with the amino acid analyses of the hydrolysed proteins. Neither protein contains tryptophan, cysteine or tyrosine. No histidine was found in NS1.

By treatment of the proteins with trypsin 15 peptides in NS1, and 13 peptides in NS2 were obtained. All peptides were sequenced by a manual micro-Edman technique as in section 2. Partial tryptic cleavage occurred after position 61 (Arg–Asn) in both proteins.

After blocking the lysine residues six tryptic peptides of NS1 and only two peptides of NS2 could be sequenced. Therefore, in the case of NS2 it was necessary to block the arginine residues before tryptic cleavage. Ten peptides were obtained. The very large peptide in the middle part of the protein was sequenced to the 18th amino acid residue by the DABITC method. Further alignment was obtained by the total hydrolysis of the peptide and by the sequence of the tryptic and SP peptides.

Both proteins were digested with *Staphylococcus aureus* protease. In NS1 five SP-peptides were found.

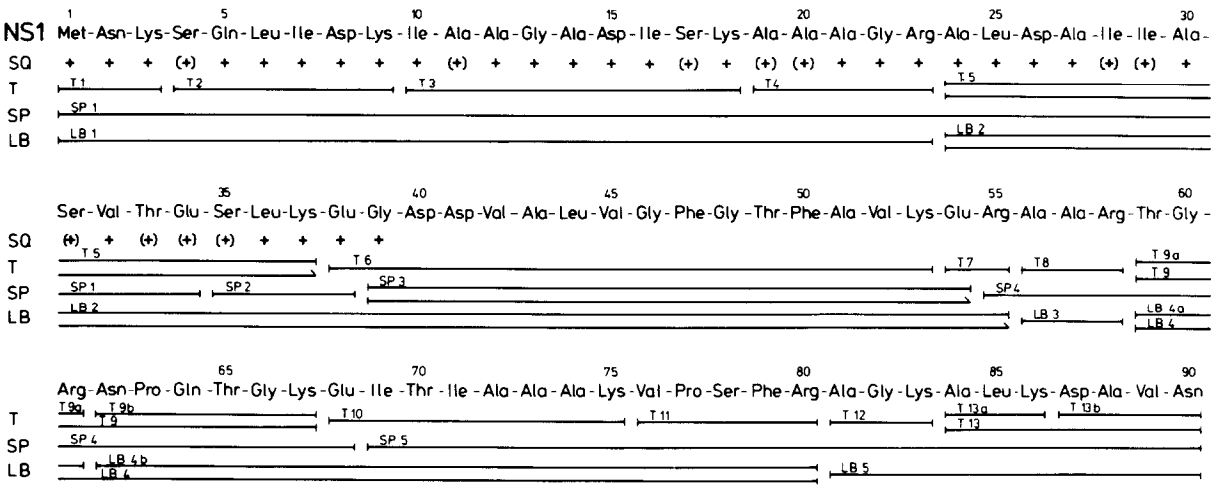

Fig.1. Amino acid sequences of protein NS1. *Abbreviations:* (T) tryptic peptides; (LB) tryptic peptides obtained after blocking the lysine residues; (CHD) tryptic peptides obtained after blocking the arginine residues; (SP) peptides derived from cleavage with *Staphylococcus aureus* protease; +, unambiguously identified by a run in an improved Beckman sequenator; (+), identified by a sequenator run.

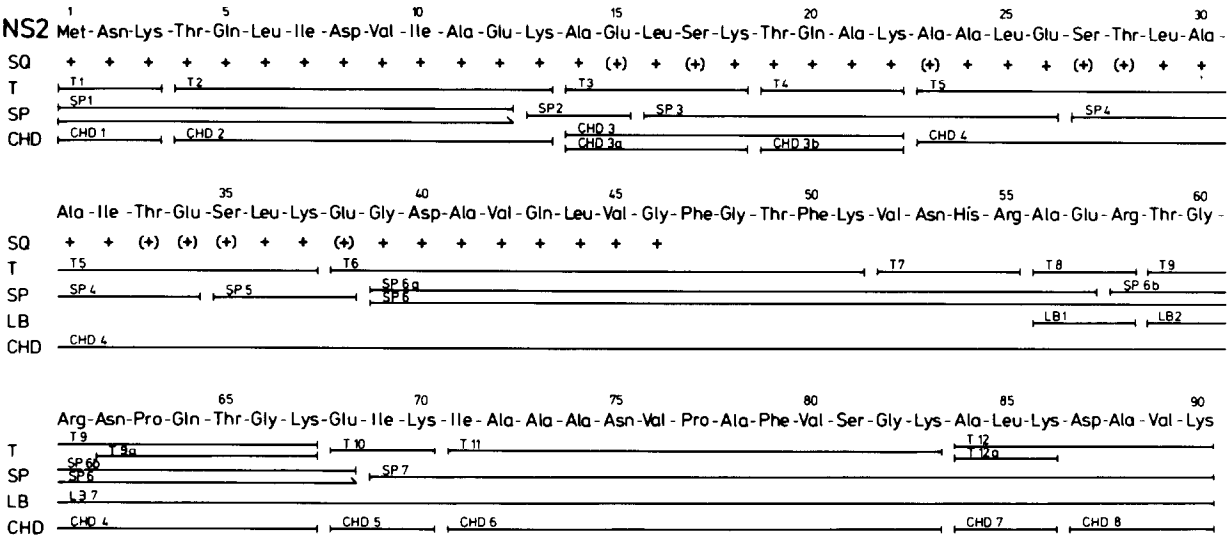

Fig.2. Amino acid sequence of protein NS2. For abbreviations see legend to fig.1.

The peptide SP1 was sequenced to the 26th residue. In NS2 eight SP-peptides were sequenced by the micro-dansyl-Edman method and by the DABITC method. Peptides SP1 and SP6 were separated from each other on a small Sephadex G-25 (superfine) column as described above and then sequenced using the DABITC method.

Combination of the results with the N-terminal sequences of the proteins determined by the automatic liquid-phase Edman degradation gave the alignment of all peptides and the complete amino acid sequences of NS1 and NS2 (fig.1 and fig.2).

4. Discussion

Both proteins, NS1 and NS2, consist of 90 amino acid residues. Each residue of one protein can be aligned with a corresponding residue of the other as shown in fig.3. The amino acids are identical in 62 from 90 residues (i.e., 69%) and different in 28 residues (31%). In the homologous regions the methionine residues as well as the proline and phenylalanine residues are at the same positions to 100%, whilst this value is $\geqslant$ 77% for the Gly, Leu, Val and Arg positions. Alanine is the most abundant amino acid

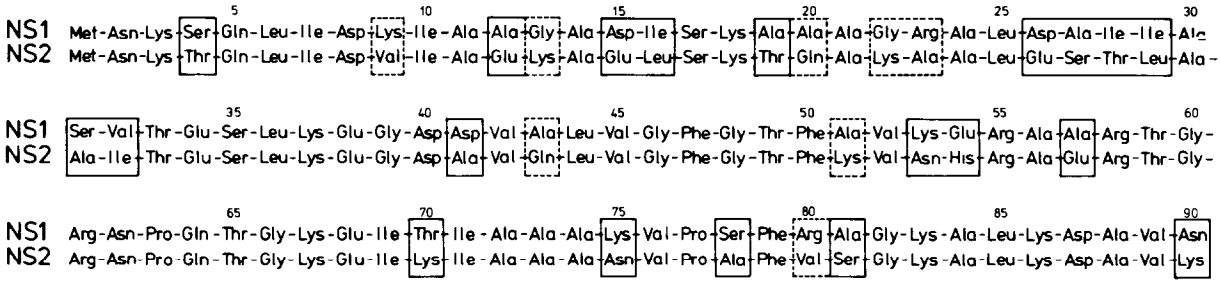

Fig.3. Comparison of the primary structures of proteins NS1 and NS2. Solid boxes contain amino acids whose codons differ by one base change. Codons for amino acids in dotted boxes differ by more than one base.

in both proteins (NS1 21%; NS2 16.6%).

Figure 3 shows the residues which differ between NS1 and NS2 drawn in boxes. Most of the amino acid differences can be explained by a single base change in the corresponding codons. For eight of the amino acid replacements double base changes have to be postulated (shown by dotted boxes in fig.3).

A partial sequence of protein HU has recently been published [12], and it differs from our sequence in several positions. At present the reason for this disagreement remains unclear.

The primary structures of NS1 and NS2 were compared by a computer programme with other nucleic acid binding proteins. The comparison with the calf thymus histones H_1, H_{2a}, H_{2b}, H_3 and H_4 showed no homologies. There was also no similarity to the *lac* repressor protein. Furthermore, the gene 5 protein of phage M13 which binds to single-stranded DNA revealed no homologies to NS1/NS2. The similarities of NS1 and NS2 with 45 *E. coli* ribosomal proteins whose primary structures are known can be explained on a random basis.

Acknowledgements

We thank Dr H. G. Wittmann for encouragement and support, Dr B. Wittmann-Liebold for provision of computer data and Dr P. Woolley for kindly reading this manuscript.

References

[1] Suryanarayana, T. and Subramanian, A. R. (1978) Biochim. Biophys. Acta 520, 342–357.
[2] Rouvière-Yaniv, J. and Gros, F. (1975) Proc. Natl. Acad. Sci. USA 72, 3428–3432.
[3] Berthold, V. and Geider, K. (1976) Eur. J. Biochem. 71, 443–449.
[4] Subramanian, A. R., Haase, C. and Giesen, M. (1976) Eur. J. Biochem. 67, 591–601.
[5] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[6] Mende, L. (1979) in preparation.
[7] Riley, M. and Perham, S. (1970) Biochem. J. 118, 733–739.
[8] Toi, K., Bynum, E., Norris, E. and Itano, H. (1967) J. Biol. Chem. 242, 1036–1043.
[9] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[10] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[11] Chang, J. Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[12] Laine, B., Sautières, P., Biserte, G., Cohen-Solal, M., Gros, F. and Rouvière-Yaniv, J. (1978) FEBS Lett. 89, 116–120.