THE AMINO- AND CARBOXY-TERMINAL AMINO ACID SEQUENCES OF PROTEIN HU FROM *ESCHERICHIA COLI*

Bernard LAINE, Pierre SAUTIÈRE and Gérard BISERTE
Unité 124 INSERM et Institut de Recherches sur le Cancer, Cité Hospitalière, 59000 Lille

and

Michel COHEN-SOLAL
Unité 91 INSERM, Hôpital Henri Mondor, 94010 Creteil

and

François GROS and Josette ROUVIÈRE-YANIV
Département de Biologie Moléculaire, Institut Pasteur, 25, rue du Docteur Roux, 75015 Paris, France

Received 7 March 1978

1. Introduction

The *Escherichia coli* DNA binding protein HU appears as a single polypeptide chain with mol. wt 9000–10 000 when determined by SDS–polyacrylamide gel electrophoresis. Protein HU is an alanine- and lysine-rich protein which lacks cysteine, tyrosine and tryptophan [1]. Closely related proteins have been isolated from blue–green algae [2] and *Bacillus subtilis* (F. Le Hegarat and J. R.-Y., unpublished results). Recent studies have shown that HU is associated with the bacterial nucleoid isolated at low ionic strength [3–5]. We therefore, postulated that this protein could play a role in prokaryotes similar to that of histones in the condensation of eukaryotic DNA.

This paper presents the amino-terminal sequence of the protein HU, determined simultaneously by automated Edman degradation of the protein and structural studies of tryptic peptides.

Abbreviations: SDS, sodium dodecyl sulfate; TPCK, tosylphenylalanyl chloromethyl ketone; TLCK, tosyllysyl chloromethyl ketone; DFP, diisopropyl phosphofluoridate; PMSF, phenylmethane sulfonylfluoride; PTH, phenylthiohydantoin; DMBA, dimethylbenzylamine

The carboxy-terminal sequence established from structural studies of chymotryptic and tryptic peptides from native protein is also reported.

2. Materials and methods

Trypsin (TPCK-treated) was from Worthington. Chymotrypsin (TLCK-treated), carboxypeptidase A (DFP-treated) and carboxypeptidase B (PMSF-treated) were purchased from Merck.

All the reagents used for automated Edman degradation were purchased from Beckman Instruments with the exception of *N,N'*-dimethylbenzylamine and *n*-propanol from Pierce Chemical Co.

For manual Edman degradation, all reagents were sequanal grade from Pierce Chemical Co, except benzene and ethyl acetate which were purchased from S.D.S, Peypin, France. *N-O*-bis (trimethylsilyl)-acetamide and standard PTH-derivatives were from Pierce Chemical Co. All other chemicals were the highest grade commercially available. Silica thin-layer plates (Kieselgel 60 F 254) were purchased from Merck.

2.1. *Protein purification*

The protein HU was isolated from *E. coli* (strain

 Elsevier/North-Holland Biomedical Press

W 3350, Institut Pasteur collection) as in [1], except that the final purification step consisted of a gel filtration chromatography on a Biogel P 60 column (200 × 2.5 cm) eluted with 0.01 N HCl, 0.05 M NaCl, 0.02% NaN_3, instead of the usual DEAE-cellulose chromatography. The purity of the protein was assessed by SDS–gel electrophoresis as in [1].

The amino acid composition of the protein was established on 24 h, 48 h and 72 h hydrolysates.

2.2. *COOH-terminal amino acid sequence determination*

Protein HU (200 nmol) dissolved in 2 ml 0.1 M ammonium bicarbonate, pH 8.0, was digested with carboxypeptidase B (2 units carboxypeptidase B/100 nmol protein) at 40°C. After 2 h digestion, a 0.4 ml aliquot was taken and mixed with 1 ml 30% acetic acid and freeze-dried. The remainder of the digest was then subjected to carboxypeptidase A hydrolysis (24 units carboxypeptidase A/100 nmol protein). After 2 h and 4 h digestion, 0.4 ml aliquots were added to 1 ml 30% acetic acid and freeze-dried. The released amino acids were analysed on the amino acid analyser.

2.3. *Enzymatic digestions*

The protein was dissolved in 0.1 M *N*-methylmorpholine acetate buffer at pH 7.5, and digested at 37°C for 4 h with trypsin. The final enzyme-to-substrate ratio as 1/50 (w/w).

Chymotryptic digestion was performed in 0.1 M ammonium bicarbonate, at pH 8.0, and 37°C for 2 h using an enzyme-to-substrate ratio of 1/50 (w/w).

The peptides were fractionated on Chromobeads P column (Technicon) with pyridine-formate and pyridine-acetate buffers and purified as in [6].

Peptides were sequenced by the manual Edman degradation with direct identification of the PTH-amino acids by thin layer micro-chromatography [7]. A ninhydrin spray reagent for color differentiation of PTH-amino acids was also employed for selected samples [8].

Automated Edman degradation [9] was performed in a Beckman 890 B Sequencer equipped with a nitrogen flush and an undercut cup. All the degradations were performed using 0.8 M DMBA as coupling buffer and a program in [10]. Dithiothreitol (Sigma) was added at a concentration of 10 mg/liter to the 1-chlorobutane. After conversion from the phenylthiazolinones, phenylthiohydantoins were identified by thin layer micro-chromatography [7,8] and by gas-chromatography [11] on a Beckman GC-45 unit. Alternatively the PTH-amino acids were converted back to the amino acids by hydrolysis in 6 N HCl for 24 h at 150°C in vacuo. In the aqueous phase, PTH-histidine and PTH-arginine were identified by a spot reaction on Whatman 3 MM paper with Pauly reagent and phenanthrene quinone reagent, respectively [12].

2.4. *Nomenclature*

The tryptic and chymotryptic peptides are lettered T- and C-, respectively. Each peptide from the amino-terminal sequence is numbered according to its position in the polypeptide chain. Peptides from the carboxy-terminal sequence are followed by a letter indicating their relative position in this sequence.

3. Results

The amino acid composition of the protein HU is presented in table 1. On the basis of three residues of phenylalanine present in the protein, as evidenced by structural studies (B. L. and P. S., unpublished results) the total number of amino acid residues is 90–91 which corresponds to calculated mol. wt 9500. This value is in good agreement with that estimated by SDS gel electrophoresis. The protein HU is rich in alanine and lysine and is devoid of cysteine, tyrosine and tryptophan.

3.1. *Amino-terminal sequence*

Automated Edman degradation of the protein HU was performed in 2 different runs, with samples of 480 nmol and 530 nmol, respectively. The repetitive yield calculated from the leucine residues at positions 6, 16 and 25 was 95% in both experiments. Positive identifications were achieved through step 37 by thin layer chromatography and gas chromatography. Isoleucine residues at positions 28 and 29 were confirmed by the presence of isoleucine and alloisoleucine on the amino acid analyser after hydrolysis in 6 N HCl at 150°C for 24 h.

There is evidence of polymorphism at positions 4, 9, 13, 15 and 26: threonine and serine were identified at position 4, lysine and valine at position 9, lysine and glycine at position 13, aspartic acid and

Table 1
Amino acid composition of protein HU from *Escherichia coli* (strain W 3350)

	mol %	Residues[a]	
Aspartic acid	8.91	8.05	(8)[b]
Threonine[c]	6.81	6.16	(6)
Serine[c]	4.53	4.10	(4)
Glutamic acid	9.54	8.62	(9)
Proline	2.55	2.30	(2)
Glycine	7.18	6.49	(6–7)
Alanine[d]	18.50	16.73	(17)
Valine	7.35	6.65	(7)
1/2 Cystine	0	0	
Methionine	1.18	1.06	(1)
Isoleucine[e]	6.59	5.96	(6)
Leucine	6.86	6.20	(6)
Tyrosine	0	0	
Phenylalanine	3.32	3.00	(3)
Lysine[d]	11.12	10.05	(10)
Histidine	0.86	0.78	(1)
Arginine	4.71	4.25	(4)
Tryptophan[f]	0	0	
Total	100.01		90–91
Lys/Arg	2.36		

[a] The results are expressed in no. residues/mol protein
[b] Number in parentheses is the nearest integer
[c] Values for threonine and serine were obtained by linear extrapolation to zero hydrolysis time
[d] The differences in the content of alanine and lysine between this composition and [1] are probably due to the use of different strains of *E. coli*.
[e] 72 h hydrolysis values
[f] Determined spectrophotometrically

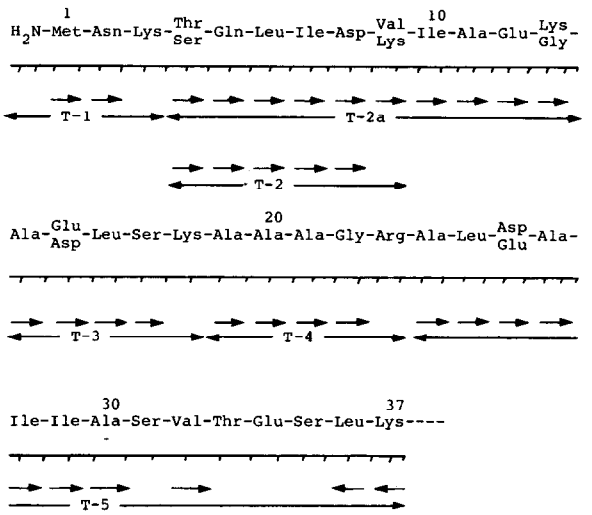

Fig.1. The amino-terminal amino acid sequence of protein HU from *E. coli*. Sequence determined by automated Edman degradation of the protein is indicated as (⌐¬¬¬). Methods used for determination of the sequence of peptides are indicated as follows: manual Edman degradation (→) and carboxypeptidases A and B hydrolysis (←). For each tryptic peptide the lysine or the arginine residue was assigned to the COOH-terminus from the specificity of the trypsin.

glutamic acid at positions 15 and 26.

The sequence of the first 37 residues of the protein HU thus established by automated Edman degradation was confirmed by structural studies of the peptides isolated from a tryptic digest of the native protein (fig.1).

The isolation of the peptides T-2, T-2a and T-3 confirms the polymorphisms at positions 4, 9, 13 and 15 observed previously in the automated Edman degradation of the protein and shows that threonine and serine at position 4 are respectively related to valine and lysine at position 9, to lysine and glycine at position 13 and to glutamic acid and aspartic acid at position 15. The tryptic peptide with glycine at position 13 and aspartic acid at position 15 was not isolated. The polymorphism at position 26 was observed uniquely in the automated Edman degradation. The tryptic peptide with glutamic acid at position 26 was not isolated.

3.2. *Carboxy-terminal sequence*

The hydrolysis of the protein HU with the carboxypeptidase B for 2 h did not release any basic residue. The hydrolysis of the protein HU with the carboxypeptidase A for 4 h released the following residues (mol/mol protein): asparagine 0.48; valine 0.13.

The carboxy-terminal sequence (11 residues) of the protein HU was deduced from the study of tryptic and chymotryptic peptides from native protein (fig.2).

From the specificity of trypsin, the tryptic peptide T-z lacking a lysine or arginine residue must be assigned to the carboxy terminal position in the protein HU. The sequence of the peptide T-z established as follows: Asp–Ala–Val–Asn, correlates with the data provided by the carboxypeptidase digestion of the protein. The tryptic peptides T-x, T-y and T-z were ordered with the overlap of the chymotryptic peptide Ch-z.

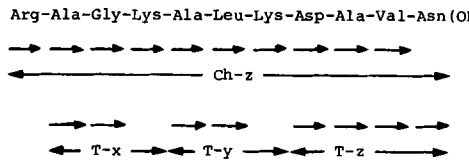

Fig.2. The carboxy-terminal sequence of protein HU from *E. coli*. (→) Denotes sequences established by manual Edman degradation.

4. Discussion

The sequence of the first 37 amino acids of the protein HU contains 19 hydrophobic residues. Neither aromatic residues, nor proline residues are present in this sequence. A sequence of 13 residues without any basic amino acid is found between arginine 23 and lysine 37. Furthermore no cluster of basic amino acids is found in this sequence. This mode of distribution is different from that observed in the histones where basic residues are generally clustered. However, in the COOH-terminal part of the protein, the distribution of the basic amino acids is according to the scheme B–X–Y–B–X–Y–B where B is a basic residue and X and Y are neutral residues. This is analogous to that found in the amino terminal sequences of the histones H2A, H2B, H3 and H4.

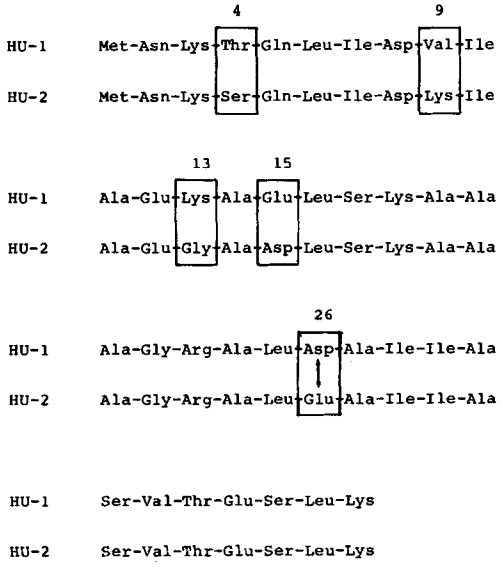

Fig.3. Amino-terminal amino acid sequence of the two molecular species HU-1 and HU-2 of protein HU from *E. coli* (strain W 3350). Incertainty, indicated by arrows, at position 26 cannot be eliminated.

The most interesting feature of the sequence is the presence of polymorphism at positions 4, 9, 13, 15 and 26. The substitutions at positions 4, 15 and 26 are conservative, while the substitutions at positions 9 and 13 are not. However, these two substitutions do not modify the net charge of the protein, since the valine to lysine change at position 9 is balanced with the lysine to glycine change at position 13. The structural studies show that polymorphisms 4, 9, 11 and 13 follow a strict alignment within the same molecule (fig.3). The alignment is such that there are apparently two variants of the HU primary sequence suggesting that there are at least two genes coding for the HU protein. The alignment of polymorphism 26 has not yet been determined (fig.3).

A small DNA-binding protein HD isolated [13] in *E. coli* (strain H 570) has a strong homology with the protein HU. Despite some differences in amino acid composition, namely in their alanine, tyrosine and histidine contents, the two proteins have identical NH_2 terminal amino acid sequences: Met–Asn–Lys–Thr [13]. Furthermore, immunological cross reactivity was observed between HU and HD (V. B. and J. R.-Y., unpublished observations).

Acknowledgements

The skilful technical assistance of Mrs D. Belaïche, M. J. Dupire, S. Quief, C. Valentin and of Mr E. Laval is gratefully acknowledged. We thank Drs A. Aitken, N. Kjeldgaard and M. Yaniv for valuable discussion. This research was supported by grant no. 745.059.02 of the Institut National de la Santé et de la Recherche Médicale, by grant no. 75.7.0187 of the Délégation Générale de la Recherche Scientifique et Technique and by Grants L.A. no. 88, L.A. no. 268 and A.T.P. no. 2873 of the Centre National de la Recherche Scientifique.

References

[1] Rouvière-Yaniv, J. and Gros, F. (1975) Proc. Natl. Acad. Sci. USA 72, 3428–3432.

[2] Haselkorn, R. and Rouvière-Yaniv, J. (1976) Proc. Natl. Acad. Sci. USA 73, 1917–1920.
[3] Rouvière-Yaniv, J., Gros, F., Haselkorn, R. and Reiss, C. (1977) in: The Organization and Expression of the Eukaryotic Genome (Bradbury, E. M. and Javaherian, K. eds) p. 211, Academic Press, New York.
[4] Varshavsky, A. J., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. and Georgiev, G. P. (1977) Nucleic Acids Res. 4, 2725–2745.
[5] Rouvière-Yaniv, J. (1978) Cold Spring Harbor Symp., Vol. XLII, in press.
[6] Sautière, P., Moschetto, Y., Dautrevaux, M. and Biserte, G. (1970) Eur. J. Biochem. 12, 222–226.
[7] Cohen-Solal, M. and Bernard, J. L. (1973) J. Chromatog. 80, 140–143.
[8] Roseau, G. and Pantel, P. (1969) J. Chromatog. 44, 392–395.
[9] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[10] Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H. and Walsh, K. A. (1972) Biochemistry 11, 4493–4501.
[11] Pisano, J. J. and Bronzert, T. J. (1969) J. Biol. Chem. 244, 5597–5607.
[12] Yamada, S. and Itano, H. A. (1966) Biochim. Biophys. Acta 130, 538–540.
[13] Berthold, V. and Geider, K. (1976) Eur. J. Biochem. 71, 443–449.