Vol. 91, No. 2, 1979 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 28, 1979 Pages 461-467

AMINO AND CARBOXY TERMINAL SEQUENCES OF THE DNA-BINDING PROTEIN HU FROM THE CYANOBACTERIUM SYNECHOCYSTIS PCC [x] 6701 (ATCC 27170)

Alastair AITKEN[*] and Josette ROUVIERE-YANIV°

[*]Service de Physiologie Microbienne and ° Département de Biologie Moléculaire, Institut Pasteur, 25 rue du Dr. roux, 75015 Paris, France

[*]Present address : Department of Biochemistry, University of Dundee, Dundee DD1 4HN, Scotland

Received August 10, 1979

SUMMARY : Amino and carboxy terminal sequences of the DNA-binding protein HU from a cyanobacterium have been determined. The partial amino acid sequence of the cyanobacterial protein is compared to that of the corresponding protein from E. coli. A high degree of similarity in primary structure is detected. The results are interpreted in terms of the large evolutionary distance between E. coli and cyanobacteria to suggest that the protein HU is, like eukaryotic histones, highly conserved in primary structure.

INTRODUCTION :

The discovery of the low molecular weight DNA-binding protein HU, which bears some resemblance to eukaryotic histones, in both E. coli [1] and in cyanobacteria [2], raises the question of whether this protein could represent the prokaryotic equivalent of a histone. The protein HU is associated with the bacterial DNA at low ionic strength and stimulates transcription in vitro . The corresponding proteins from E. coli and cyanobacteria have similarities in amino acid composition and cross-react immunologically [2]. The protein from E. coli has a molecular weight of 9, 500 and contains 90 amino acids [3, 4] while that from cyanobacteria has a molecular weight of around 11, 000 and contains approximately 100 amino acids. The proteins are rich in lysine and alanine and lack tyrosine, cysteine and tryptophan.

The evolutionary divergence of the gram-negative bacterium E. coli and cyanobacteria (which also have a gram-negative cell wall structure) from their common ancestor took place in the Precambrian era, probably

[x] PCC = Culture collection of Cyanobacteria at Pasteur Institute

0006-291X/79/220461-07$01.00/0

around 3×10^9 years ago, according to paleontological evidence[5]. If considerable similarity is found between the amino acid sequences of HU from Synechosystis and E. coli, this would provide strong evidence for a slow rate of evolution of HU, as occurs in most eukaryotic histones [6]. The purpose of this study was to investigate the terminal sequences of HU from a cyanobacterium with a view to revealing possible structural similarities to histones and to determine the degree of amino acid sequence homology with the protein from E. coli.

MATERIAL AND METHODS

Synechocystis PCC 6701 is a unicellular cyanobacterium which is now deposited in the American type culture collection (N° ATCC 27170). The organism was grown essentially as described previously [2] while the protein was purified for use in this study by a new method inspired from the procedure of Alberts and Herrick[7].

Purification of Protein HU - 200 g fresh weight of Synechocystis 6701 (stored at -20°C) was resuspended at 4°C in the minimum volume of 20mM Tris buffer pH 7.9 containing 1.7 M NaCl, 1 mM EDTA and 1 mM mercaptoethanol. The suspension was passed through a French pressure cell once at 1 200 Bar (20,000 pounds $inch^{-2}$, 140 M Pa). The broken cell suspension was diluted to 800 ml in the same buffer, stirred for 30 min to insure thorough homogeneity and centrifuged for 30 min at 8 000 g in a Sorvall centrifuge. Polyethylene glycol (PEG, 30% aqueous solution) was added to the supernatant to give a final concentration of 10 % in PEG. The mixture was stirred for

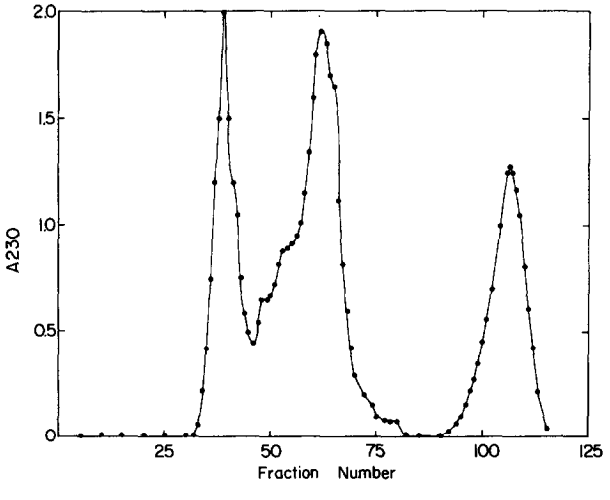

Fig. 1 : Sephadex G-100 gel filtration of HU. The O.4 M salt eluate from the DNA cellulose column was concentrated and applied to a Sephadex G-100 column. Fractions (2.0 ml) were collected and their protein content analyzed by polyacrylamide gel electrophoresis (Fig. 2b). HU was eluted between fractions 48 and 68.

30 min on a magnetic stirrer at 4°C then centrifuged for 14 min at 12,000 g. The supernatant was dialyzed against several changes of 20 mM Tris buffer (pH 7.9) containing 1 mM EDTA and 50 mM NaCl. The solution was centrifuged for 15 min at 12,000 g. The supernatant was made 10 % in glycerol and applied to a column of double stranded DNA cellulose (30 cm x 2.4 cm) prepared as previously described[1]. The column was washed first with the above dialysis buffer. The DNA binding proteins were eluted stepwise with this buffer containing 0.15 M NaCl, 0.4 M NaCl and 2 M NaCl. The protein HU was eluted with the buffer containing 0.4 M NaCl. The fractions containing HU were concentrated by ultrafiltration and further purified by gel filtration on Sephadex G-100 (100 cm x 2.4 cm) (Fig. 1) and by passing through a DEAE cellulose column. The purity of the protein was monitored at each stage of purification by electrophoresis in SDS polyacrylamide gels containing a linear gradient of 10 % to 25 % acrylamide (Fig. 2). The final yield of HU protein was around 10 mg from 200 g fresh weight of cyanobacteria.

Automatic N-terminal sequencing - The protein was sequenced in a Beckman automatic protein sequencer using a dimethylbenzylamine programme [8]. Residues 1 to 27 were identified. The automatic sequencing was done twice on approximately 200 nm aliquots of protein. Phenylisothiocyanate derivatives were identified by thin layer chromatography (tlc) on silicon gel sheets containing internal fluorescent indicator with two solvent system [9] and gas liquid chromatography (g.l.c.). Other methods for amino and carboxy terminal sequencing were as described previously [10, 11].

RESULTS AND DISCUSSION

The results of automatic Edman degradation of the protein are shown in Fig. 3. At steps 15 and 24, evidence for micro-heterogeneity was obtained. The repetitive yield was 94-95 % based on the valine residues. In the investigation of the carboxy terminus of the protein the following amino acids were liberated after digestion with carboxypeptidase A : -leucine, 0.15 ; aspartic acid, 0.18 ; isoleucine, 0.32 ; valine, 0.46 ; alanine, 0.71 (expressed moles/mole protein). These results were consistent with the dansyl amino acids identified on thin layer chromatography.

The amino and carboxy terminal sequences of the cynobacterial HU are shown in Fig. 3(a) aligned with the sequence of E. coli HU variants HU-1 (NS 1) and HU-2 (NS 2) [3, 4] in Fig. 3(b). Identical residues and conservative changes are shown in upper case letters.

Further evidence for the heterogeneity in amino acid sequence seen during automatic sequencing was obtained from the amino acid analysis data for the peptides isolated from the N-terminal region of the protein by chymotryptic and cyanogen bromide digestion (data not shown). The presence of heterogeneity has been established in the amino acid sequence of E. coli

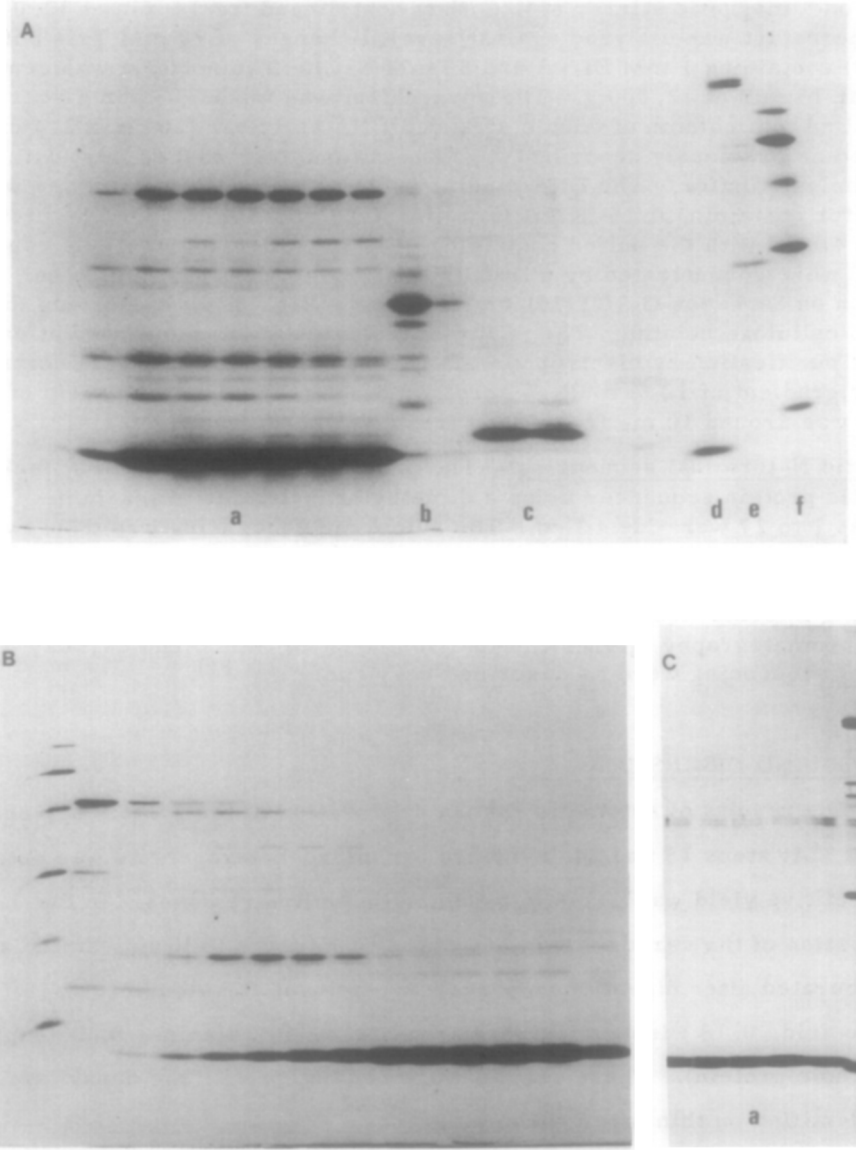

Fig 2 : Polyacrylamide gel electrophoresis monitoring the purification of HU. Aliquots were analyzed on 10-25 % gradient gels.

A. Analysis of proteins eluted from the DNA cellulose column at 0.4 M NaCl (a), 0.15 M NaCl (b) and 2.0 M NaCl (c) ; E. coli HU (d) ; E. coli RNA polymerase (e) ; series of markers (β-galactosidase : 135 K ; phosphorylase B : 94 K ; bovine serum albumine : 68 K ; creatine phosphokinase : 40 K and cytochrome C : 13 K)in (f).

B. Analysis of fractions eluted from Sephadex G-100 column (Fig. 1) ; markers as in A (f)

C. Pure HU obtained after passage through DEAE cellulose column (a) ; HU and RNA polymerase from E. coli (b)

```
                1                   5                  10                  15                  20
(a)     MET-ASN-LYS-Gly-Glu-LEU-VAL-ASP-Ala-VAL-MET-ALA-LYS-ALA-Thr/ASP-VAL-Thr-LYS-Lys-GLN-

(b) HU-1/HU-2  MET-ASN-LYS-Thr/Ser-Gln-LEU-ILE-ASP-Val/Lys-ILE-ALA-Glu/ALA_LYS/Gly-ALA-Glu/ASP_LEU/ILE-SER-LYS-Thr/Ala_GLN/Ala_

                                25
(a)     -ALA-Asp-ALA-PHE/VAL-ILE-Leu-ALA.....................................Leu-ASP-ILE-VAL-Ala

(b) HU-1/HU-2  -ALA-Lys/Gly_ALA/Arg-ALA-LEU-Glu/Asp_Ser/ALA.....................................Lys-ASP-ALA-VAL-Lys/Asn
```

Fig. 3 : Alignment of amino- and carboxy-terminal residues of HU proteins from (a) Synechocystis, (b) E. coli (3, 4). Identical residues and conservative changes are shown in upper case letters. The sequence variants in E. coli HU-1 (NS2) and HU-2 (NS1) are shown above and below the line, respectively.

HU[3, 4]. It has recently been shown that there are differences at 28 positions (out of 90 residues) in the E. coli HU variants NS1 and NS2 [4].

Due to the small amount of cyanobacterial protein available, it is possible that not all variants of the peptide for a particular region were isolated. The yields of these peptides which were isolated were low and some were not completely purified. The amino acid analyses do nevertheless support the proposed N-terminal sequence obtained by automatic Edman degradation (shown in Fig. 3).

It is seen from Fig. 3 that there is 40 % identity in the N-terminal regions of the two proteins (10 3/4 identities out of the 29 residues compared). The presence of microheterogeneity is taken into account in the numerical comparisons. The similarities at positions 12, 13, 20, 23 and 27 are scored 1/2 while that at position 15 is scored 1/4 (threonine/aspartic acid in cyanobacterial HU and glutamic acid/aspartic acid in E. coli HU).

The degree of homology of the proposed C-terminal sequence of Synechocystis HU with that from E. coli HU is similar to that obtained from the N-terminal sequence comparisons. There are 2 residues identical from 5 compared, i.e. 40 % similarity.

The amino- and carboxy-terminal sequence comparisons suggest therefore that there may be about 40 % similarity in the complete amino acid sequence. Since Synechocystis HU is longer by about 10 residues

than E. coli HU, these extra amino acids will be inserted somewhere in the central portion of the sequence. Comparison of the amino acid compositions of the proteins[2] suggests that the sequences could have been even closer. When complete sequences of cyanobacterial HU proteins are known, it is possible that other regions of the molecules will be seen to be more highly conserved.

In spite of the great period of time elapsed since the divergence of gram-negative bacteria and cyanobacteria (approximately 2 to 3 x 10^9 years[5]) there is a large degree of amino acid sequence similarity remaining between the HU proteins of these two organisms, confirming that they have arisen from a common ancestral sequence and have been highly conserved in structure. The results from the preliminary data obtained in this study suggest that the rate of evolution of HU is in the order of 1 % difference in amino acid sequence per 5 x 10^7 years. This slow rate of evolution is comparable to the rates of evolution of histones H2A and H2B for example (1 % difference per 6 x 10^7 years[5]). So far neither the partial amino acid sequence of Synechocystis HU nor the complete sequence of E. coli HU[4] has indicated the presence of regions containing clusters of basic residues which are characteristic of histone proteins.

ACKNOWLEDGEMENTS : We are grateful to Roger Stanier and François Gros for their interest and support. We are endebted to Robert Haselkorn for fruitful discussions and to André de Wolf for running with competence the automatic sequencer. This work was supported by C. N. R. S. (A. T. P. Chromatine).

REFERENCES

1. Rouvière-Yaniv, J. and Gros, F. (1975) Proc. nat. Acad. Sci. USA, 72, 3428-3432
2. Haselkorn, R. and Rouvière-Yaniv, J. (1976) Proc. nat. Acad. Sci. USA 73, 1917-1920
3. Laine, B., Sautière, P., Biserte, G., Cohen-Solal, M., Gros, F. and Rouvière-Yaniv, J. (1978) FEBS Lett. 89, 116-120
4. Mende, L., Timm, B. and Subramanian, A. R. (1978) FEBS Lett 96, 395-398
5. Schopf, J. W. (1970) Biol. Rev. 45, 319-352
6. Wilson, A. C., Carlson, S. S. and White, T. J. (1977), Ann. Rev. Biochem. 46, 573-639
7. Alberts, B. and Herrick, G. (1971) in Methods in Enzymology, Ed. Colowick, S. P. and Kaplan, N. O. (Academic Press, New York) Vol. XXI pp. 198-217

8. Hermodsen, M. A., Ericson, L. H., Titani, K., Neurath, H. and Walsh K. A. (1972) Biochemistry 11, 4493-4502.
9. Bridgen, J., Graffer, A. P., Karger, B. L. and Waterfield, M. D., The Identification of PTH Amino Acids, In "Instrumentation in Amino Acid Sequence Analysis", Ed. R. Perham (Academic Press, New York) pp. 111-145 (1975)
10. Aitken, A. (1975) Biochem. J., 149, 675-683
11. Aitken, A. (1977) Europ. J. Biochem., 78, 273-279.