THE PRIMARY STRUCTURE OF THE NUCLEOSOME-ASSOCIATED CHROMOSOMAL PROTEIN HMG 14

John M. WALKER, Graham H. GOODWIN and Ernest W. JOHNS

Chester Beatty Research Institute, Institute of Cancer Research: Royal Cancer Hospital, London, SW3 6JB, England

Received 13 February 1979

1. Introduction

Chromatin contains a group of non-histone proteins called the high mobility group (HMG) proteins [1]. There are four main HMG proteins in thymus, HMG 1, 2, 14 and 17, which have all been shown to be present in isolated nucleosomes [2]. All four of these proteins have been isolated in a pure form from both pig and calf thymus [3–5], and an HMG protein from trout testis, HMG-T, has also been isolated [6]. Because of the quantities of the HMG proteins present in the nucleus (10^5–10^6 molecules of each protein) we feel that the HMG proteins are structural proteins, possibly involved in the higher ordered structure of the chromatin, and not involved in specific gene control. As part of our characterisation of the HMG proteins we are determining the amino acid sequences of the four calf thymus proteins. We have published partial sequences for HMG 1 and 2 [6,7] and the complete amino acid sequence of calf thymus HMG 17 [8]. We now report the complete amino acid sequence of calf thymus HMG 14.

2. Experimental

2.1. *Isolation of HMG 14*

Calf thymus HMG 14 was prepared as in [4].

2.2. *Proteolytic cleavage of HMG 14*

HMG 14 was subjected to enzymatic digestion with the following four enzymes:

(a) Pepsin digestion of HMG 14 was carried out on 15 mg protein (10 mg/ml in 1 N acetic acid) at an enzyme to protein ratio of 1 : 50 at 37°C for 4 h. The cleavage products (designated P) were loaded directly onto Whatman 3MM paper and separated by descending paper chromatography in butanol : pyridine : acetic acid : water (15 : 10 : 1 : 12, v/v/v/v) for 6 days. Peptides were numbered in order from the origin (peptide P1 remained at origin) and, if necessary, further purified by high-voltage paper electrophoresis at pH 3.5.

(b) V8 protease (Staphylococcal protease) digestion was carried out on 15 mg protein (10 mg/ml in 0.2 M ammonium bicarbonate) at an enzyme to protein ratio of 1 : 50 at 37°C for 24 h. Cleavage products (designated V) were separated and further purified as described for the pepsin digest. Peptide V1 moved ~1 cm from the origin.

(c) Thermolytic digestion was carried out on 15 mg protein (5 mg/ml in 0.2 M ammonium bicarbonate) at an enzyme : protein ratio of 1 : 50 at 37°C for 2 h. Cleavage products (designated TL) were separated and, where necessary, further purified as for the pepsin digest. Peptide TL1 moved ~1 cm from the origin.

(d) Tryptic digestion was carried out on 10 mg protein (5 mg/ml in 0.2 M ammonium bicarbonate) at an enzyme : protein ratio of 1 : 100 at 37°C for 24 h. Cleavage products (designated T) were initially separated by high-voltage paper electrophoresis at pH 3.5, then, where necessary, further purified by descending paper chromatography.

2.3. *Peptide sequence determinations*

Automated Edman degradations were carried out

Elsevier/North-Holland Biomedical Press

on a Beckman 890C protein sequencer using a 0.1 M quadrol buffer programme [9], with a double cleavage step on each cycle. Polybrene (5 mg) was used as a carrier and taken through three cycles of the Edman degradation together with 100 nmol of glycyl glycine [9] prior to each sequenator run. PTH amino acid derivatives were identified by high-pressure liquid chromatography on a DuPont 830 liquid chromatograph using a Zorbax ODS column as in [10] and by back hydrolysis with 50% hydriodic acid to the free amino acid.

3. Results and discussion

Residues 1–35 were determined by sequenator analysis of 3 mg total protein. The remainder of the sequence was derived from sequenator analysis of peptides P1, V1, TL1 and T10. The first 30 residues of peptide P1 were determined giving residues 24–53 in the total molecule. The total sequence of peptide V1 gave residues 34–64 and the total sequence of peptide TL1 gave residues 65–100. The only overlap not produced by these peptides at residues 64–65 was provided by tryptic peptide T10. The amino acid analysis of these major peptides together with that of total HMG 14 is shown in table 1. In addition to the major peptides a variety of smaller peptides were isolated from the four digests. These peptides, sequenced by the manual dansyl-Edman technique, confirm a large proportion of the residues found by sequenator analysis. In particular the production of the tryptic peptide Ser–Asp confirms TL1 to be the C-terminal peptide as does the production of free aspartic acid from the V8 protease digest, produced by cleavage at Ser 99. Some of these smaller peptides are shown in the composite sequence for HMG 14 in fig.1. It is interesting to note that the Staphylococcal protease, in addition to causing cleavage at glutamic acid residues (its normally reported specific-

Table 1
Amino acid composition and N-terminal analyses of HMG 14 and the major pepsin, thermolysin, V8 protease and tryptic peptides obtained from HMG 14

Amino Acid	HMG 14	P1	TL1	V1	T10
Asp	8.3 (8)	3.2 (3)	5.5 (6)	2.5 (2)	1.1 (1)
Thr	4.1 (4)	1.9 (2)	1.9 (2)	1.8 (2)	1.0 (1)
Ser	8.0 (9)	2.7 (3)	2.6 (3)	1.9 (2)	0.2 (0)
Glu	17.5 (18)	5.8 (6)	10.3 (10)	3.0 (3)	4.1 (4)
Pro	8.1 (7)	3.6 (3)	2.5 (2)	1.3 (1)	– –
Gly	6.4 (6)	4.1 (4)	1.4 (1)	4.3 (4)	0.2 (0)
Ala	14.8 (16)	7.9 (8)	5.4 (5)	4.4 (4)	2.1 (2)
Cys	– –	– –	– –	– –	– –
Val	4.0 (4)	2.4 (3)	1.1 (1)	1.0 (1)	1.0 (1)
Met	0.1 (0)	– –	– –	– –	– –
Ile	0.3 (0)	– –	– –	– –	– –
Leu	2.0 (2)	0.1 (0)	1.0 (1)	0.1 (0)	– –
Tyr	0.2 (0)	– –	– –	– –	– –
Phe	0.3 (0)	– –	– –	– –	– –
His	0.2 (0)	– –	– –	– –	– –
Lys	21.1 (21)	11.7 (13)	5.3 (5)	11.2 (11)	1.1 (1)
Arg	5.4 (5)	1.1 (1)	0.1 (0)	0.7 (1)	– –
N-terminus	Pro	Ser	Val	Thr	Glx
Total residues	100	46	26	31	10
Position in sequence	1-100	24-69	65-100	34-64	62-71

The composition of each peptide is given as the molar ratios of the amino acids. No corrections are made for hydrolytic losses. Figures in parentheses are the numbers of residues found

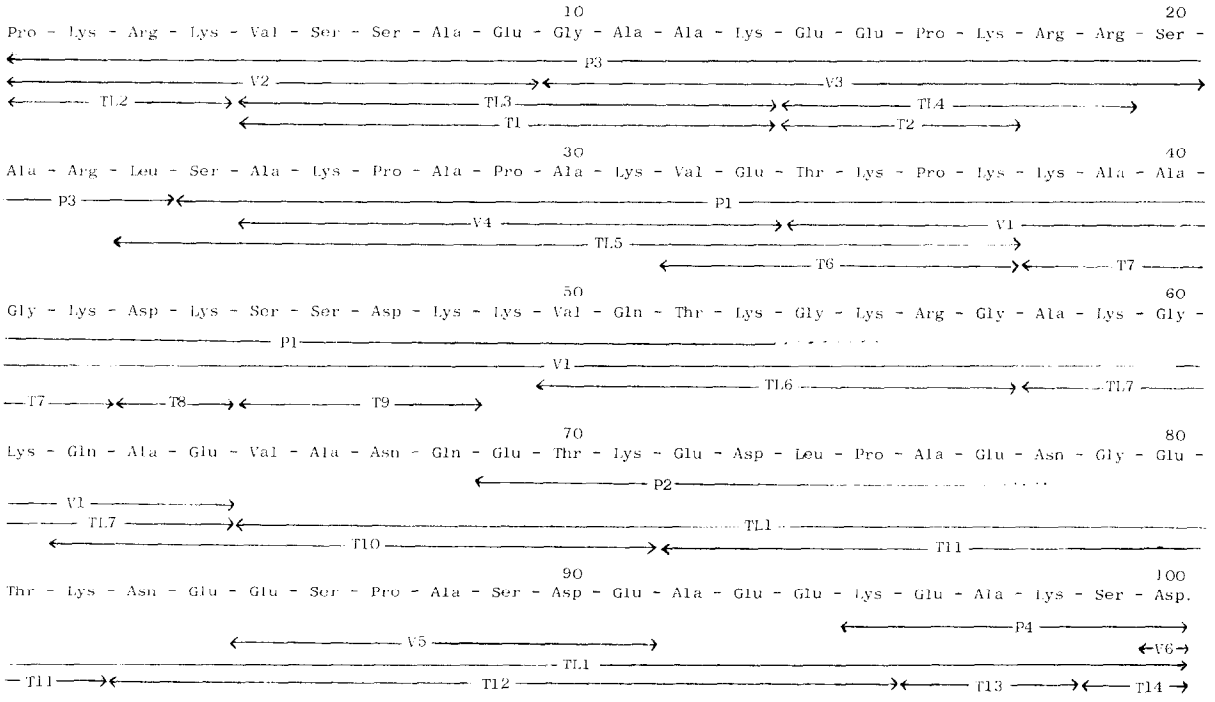

Fig.1. The amino acid sequence of HMG 14.

ity) at positions 9, 33 and 64, cleavage at three serine residues at positions 20, 24 and 99 also occurred. This is in agreement with our observation [8] that the Staphylococcal protease will cleave at certain serine residues.

The total amino acid sequence of HMG 14 is shown in fig.1. The sequence consists of 100 amino acid residues, and shows the charged residues to be irregularly distributed throughout the molecule. SDS–gel electrophoresis of HMG 14 shows this protein to run just behind HMG 17 (89 residues), which is consistent with the molecule being 100 residues long. 85% of the basic amino acids (5 arginine + 17 lysine residues) are concentrated in the first 61 residues. This situation is similar to that found with the histones where the N-terminal regions are rich in basic amino acids. However, in contrast to the histones, the C-terminal region of the molecule (residues 62–100) contains a high proportion of acidic amino acids (14 glutamic and aspartic acid residues, but only 4 lysine residues). This overall negative charge in the C-terminal region of the molecule is a feature not found with any of the histones. It is, however, also a feature of protein HMG 17, where the last 30 residues of the molecule contain 7 acidic residues and only 4 basic residues [8]. Comparison of the sequence of HMG 14 with that of the histones shows some short regions of homology with rabbit thymus histone H1 and the erythrocyte specific histone H5, all of which are also shared with non-histone protein HMG 17 (see fig.2). Comparison with the known sequences of non-histone chromosomal proteins reveals considerable sequence homology between the N-terminal regions of HMG 14 and HMG 17, and to a lesser extent with the trout specific protein H6 [11] (see fig.2). We have shown that the principal DNA-binding segment of HMG 17 is between about residues 15 and 40 [12]. Since it is this region that shows greatest sequence homology between the three above-mentioned proteins (fig.2), it seems likely that this region of common sequence represents a common DNA binding site in all three proteins. In contrast, the C-terminal regions show only one reasonable length of seven residues homology between the three

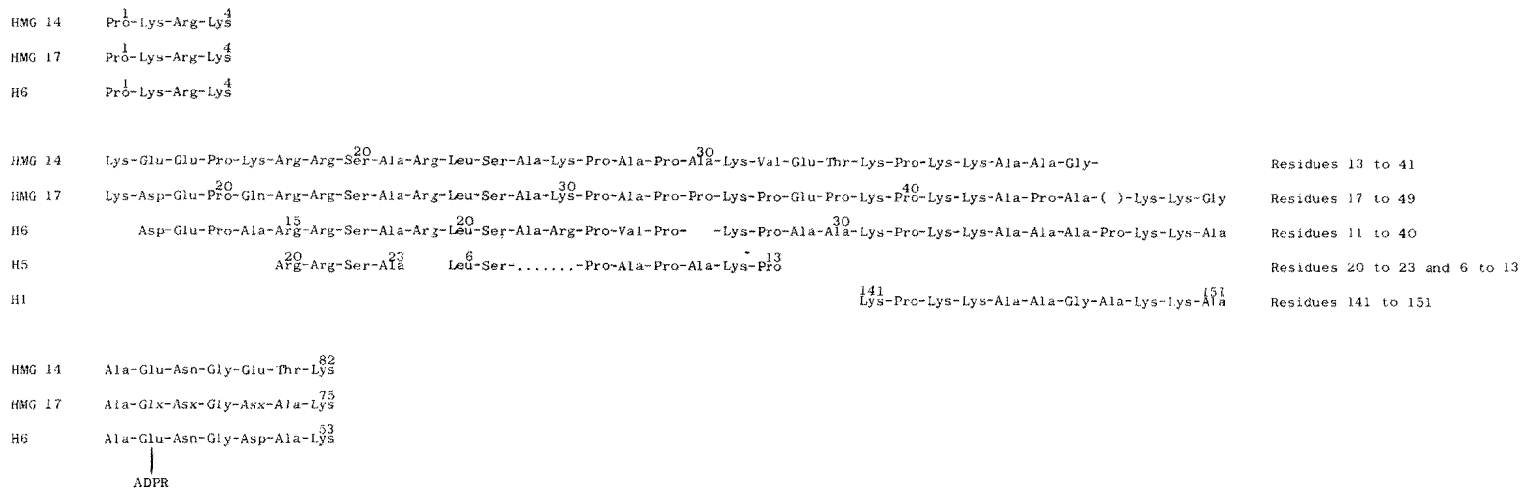

Fig.2. Regions of sequence homology between HMG 14 and other chromosomal proteins.

proteins, which is also shown in fig.2. It is interesting to note that in this homologous region the glutamic acid at residue 48 in H6 has been shown to be linked to poly-ADP-ribose [11]. Poly-ADP-ribosylation of HMGs 14 and 17 at the homologous glutamic acid residues is therefore a possibility that we are actively investigating.

Although DNA binding is an obvious function for the N-terminal regions of HMGs 14 and 17, we have little idea what functions the acidic C-terminal regions of these molecules are performing. We have recently shown non-histone chromosomal proteins HMG 1 and 2 also to have overall negatively charged C-terminal sequences [13]. At present our sequence results for the high mobility group proteins represent a large part of the available sequence data for mammalian non-histone chromosomal proteins. It will therefore be interesting to see if the feature of a negatively charged C-terminal region, which we have now shown to be present in all four HMG proteins, proves to be a common feature of other nucleosome-associated non-histone chromosomal proteins.

Acknowledgements

The authors wish to thank Miss Carol Wright for excellent technical assistance. This work was supported by grants to the Chester Beatty Research Institute (Institute of Cancer Research: Royal Cancer Hospital) by the Medical Research Council.

References

[1] Johns, E. W., Goodwin, G. H., Walker, J. M. and Sanders, C. (1975) in: The Structure and Function of Chromatin, Ciba Found. Symp. 28, pp. 95–112, Elsevier Excerpta Medica, Amsterdam, New York.
[2] Goodwin, G. H., Woodhead, L. and Johns, E. W. (1977) FEBS Lett. 73, 85–88.
[3] Goodwin, G. H., Nicolas, R. H. and Johns, E. W. (1975) Biochim. Biophys. Acta 405, 280–291.
[4] Goodwin, G. H., Rabbani, A., Nicolas, R. H. and Johns, E. W. (1977) FEBS Lett. 80, 413–416.
[5] Sanders, C. (1975) PhD Thesis, London University.
[6] Walker, J. M., Goodwin, G. H., Johns, E. W., Wietzes, P. and Gaastra, W. (1977) Int. J. Pept. Prot. Res. 9, 220–223.
[7] Walker, J. M., Parker, B. P. and Johns, E. W. (1978) Int. J. Pept. Prot. Res. 12, 269–276.
[8] Walker, J. M., Hastings, J. R. B. and Johns, E. W. (1977) Eur. J. Biochem. 76, 461–468.
[9] Hunkapiller, M. and Hood, L. (1978) Biochemistry 17, 2124–2133.
[10] Zimmerman, C., Appella, E. and Pisano, J. (1976) Anal. Biochem. 77, 569–574.
[11] Levy, W. B., Wong, N. C. W., Watson, D. C., Peters, E. H. and Dixon, G. H. (1977) Cold Spring Harbor Symp. 42, 793.
[12] Abercrombie, B. D., Kneale, G. G., Crane-Robinson, C., Bradbury, E. M., Goodwin, G. H., Walker, J. M. and Johns, E. W. (1978) Eur. J. Biochem. 84, 173–177.
[13] Walker, J. M., Hastings, J. R. B. and Johns, E. W. (1978) Nature 271, 281.