STUDIES ON THE STRUCTURES OF SOME HMG-LIKE NON-HISTONE CHROMOSOMAL PROTEINS FROM TROUT AND CHICKEN TISSUES. COMPARISON WITH CALF THYMUS PROTEINS HMG14 AND 17

J. M. WALKER, E. BROWN, G. H. GOODWIN, C. STEARN and E. W. JOHNS

Chester Beatty Research Institute, Institute of Cancer Research: Royal Cancer Hospital, Fulham Road, London, SW3 6JB, UK

Received 18 February 1980

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 calf thymus, HMG 1, 2, 14 and 17, and some minor HMG protein components of calf thymus chromatin have also been characterised [2]. Recently much attention has been focused on HMG14 and 17 because of their possible role in gene transcription [3,4]. The complete amino acid sequences of HMG17 (from calf thymus and chicken erythrocytes) and HMG14 (from calf thymus) have been determined [19]. The amino acid sequence of a trout testis protein, H6 [7], showed this protein to have considerable sequence homology with calf thymus HMG 17 and this protein is now generally regarded as belonging to the HMG group of proteins. Studies on the HMG proteins have revealed that they are widely distributed. As well as being present in a variety of mammalian tissues [8,9], HMG-like proteins have been shown to be present in avian erythrocytes [10,11], trout testis and trout liver [12,13], wheat and yeast [14] and insects [15,16]. However little is known about the structures of the HMG proteins from many of these sources. It cannot be stated for certain that the above-mentioned proteins are true HMG proteins since their characterisation is based only on gel electrophoretic mobilities and total amino acid analyses. Only when sequence data are available for these proteins can they be rigorously compared to calf thymus HMG proteins.

We have been studying the structure of HMG-like proteins from different species, and we report here partial sequences of two proteins analogous to HMG14 and 17 from trout liver, and the partial sequence of protein HMG14 from chicken erythrocytes. The sequence data on these proteins, when compared with previously determined sequences for HMG14 and 17, strongly suggest that they belong to the HMG14 and 17 group of proteins, and provides information on the evolutionary variability of these proteins.

2. Experimental

2.1. *Isolation of proteins*

Trout liver proteins C and D, and chicken erythrocyte HMG14, were prepared as in [13] and [17] respectively.

2.2. *Pepsin cleavage of chicken erythrocyte HMG14*

Pepsin cleavage of HMG14 was carried out on 15 mg of protein (5 mg/ml in 5% acetic acid) at an enzyme to protein ratio of 1:50 w/w 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 3 days. Peptide P1 remained at the origin while peptides P2 and P3 moved 5 cm and 10 cm respectively.

2.3. *Peptide sequence determinations*

Automated Edman degradations of native proteins were carried out on a Beckman 890C protein sequencer using a 0.1 M quadrol buffer programme with a double cleavage step on each cycle, essentially as described in [18]. The same programme was used for the smaller peptides P2 and P3 but with the addition that polybrene (5 mg) was used as a carrier and taken through

three cycles of the Edman degradation together with 100 nmol of glycyl glycine prior to each sequenator run [18]. PTH derivatives of released amino acids were determined both directly by high pressure liquid chromatography (HPLC) and indirectly by back-hydrolysis to the free amino acid. HPLC was carried out on a DuPont 830 Liquid Chromatogram using a Partisil PX5 ODS column (Whatman). PTH amino acids were eluted with a linear gradient of acetonitrile from 15 to 48% in 0.01 M sodium acetate buffer pH 4.5, over a period of 7 min and then holding at 48% acetonitrile for a further 4 min. Eluted PTH amino acids were identified by their absorbance at 269 nm. Back-hydrolysis of PTH amino acids was carried out in 65% hydriodic acid at 110°C for 24 h. Liberated amino acids were identified on a Rank-Hilger Chromaspek amino acid analyser.

3. Results and discussion

3.1. *Chicken erythrocyte HMG14*

Our initial observations on chicken erythrocyte HMG14 showed the amino-terminal 28 residues to differ from the calf thymus sequence at only four positions [17]. This result, together with the considerable similarity between the amino acid analyses of the two proteins led us to suggest that there would be ~10% sequence variation between the two proteins [17]. At the same time a comparable study on HMG17 from calf and chicken tissue suggested a 5% sequence variation in HMG17 from the two sources and the recently completed sequence of chicken erythrocyte HMG17 [19] (which differs from the calf sequence at only 5 positions) has confirmed this observation. However, the results presented here on the further characterisation of chicken erythrocyte HMG14 show that our initial extrapolations on the structure of chicken erythrocyte HMG14 were incorrect and that the sequence of chicken erythrocyte HMG14 differs considerably from that of the calf protein.

Pepsin cleavage of chicken erythrocyte HMG14 produced 3 peptides (P1 to P3). The amino acid analysis of peptide P1 is shown in table 1 and suggests that this peptide is 89 residues long. The amino acid sequence of the first 30 residues of this peptide has been determined and shows the peptide to overlap

Table 1

The amino acid compositions (% mol) of trout liver proteins C + D [13], trout testis protein H6 [12], proteins HMG14 and 17 from calf thymus and chicken erythrocytes [17] and peptide P1 from chicken erythrocyte HMG14

Amino acids	Trout liver			Calf thymus		Chicken erythrocyte		
	C	D	H6	HMG14	HMG17	HMG14	HMG17	P1
Asp	6.0	8.0	6.7	8.3	12.0	9.3	9.1	11.3 (10)
Thr	2.6	4.1	1.6	4.1	1.2	4.6	3.0	5.4 (5)
Ser	4.4	4.3	5.6	8.0	2.3	5.2	4.3	4.0 (4)
Glu	23.8	21.9	6.1	17.5	10.5	15.6	11.7	14.9 (13)
Pro	10.9	8.4	12.3	8.1	12.9	10.5	12.1	11.5 (10)
Gly	2.8	3.3	7.4	6.4	11.2	5.6	10.0	5.7 (5)
Ala	16.0	16.6	25.4	14.8	18.4	18.0	17.2	16.9 (15)
Cys	–	–	–	–	–	–	–	–
Val	4.2	2.5	3.4	4.0	2.0	tr	2.2	–
Met	tr	tr		–	–	–	–	–
Ile	0.6	–	–	–	–	–	–	–
Leu	0.8	1.0	1.2	2.0	1.0	1.1	1.2	–
Tyr	tr	–	–	–	–	tr	tr	–
Phe	tr	tr	–	–	–	tr	tr	–
His	tr	0.9	–	–	–	1.1	tr	1.1 (1)
Lys	19.6	23.7	23.1	21.1	24.3	24.0	23.6	28.1 (25)
Arg	4.6	4.2	7.2	5.4	4.2	4.1	4.6	1.2 (1)
								89

The values in parentheses for peptide P1 are residues/mol

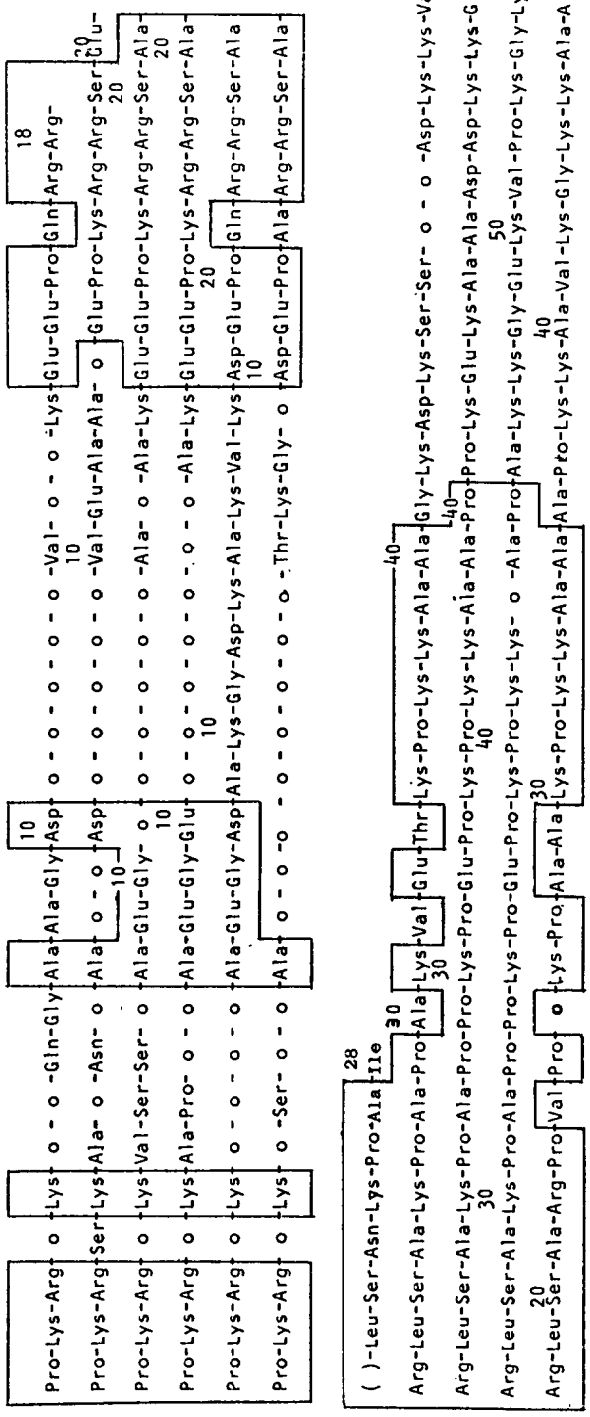

Fig.1. Comparison of the amino terminal sequences of trout proteins C, D and H6, and calf thymus and chicken erythrocyte protein HMG14 and HMG17. The amino terminal sequence of chicken erythrocyte HMG17 shows only 2 differences from calf thymus HMG17, having Thr at residue 9 and Ser at residue 48.

CTHMG14 Glu-Ala-Glu-Glu-Lys-Glu-Ala-Lys-Ser-Asp

ChEHMG14 Ala-Ser-Asp-Asp-Lys-Glu-Ala-Lys-Ser-Glu Peptide P3

Fig.2. Comparison of the C-terminal peptides from calf thymus and chicken erythrocyte HMG14. CT, calf thymus; ChE, chicken erythrocyte.

the previously determined amino terminal sequence, having been produced by cleavage at the leucine residue at position 22 (fig.1). This overlap therefore extends the amino-terminal sequence to residue 52. The amino acid analysis and partial sequence of peptide P2 shows this peptide to be the first 22 residues of the molecule. Peptide P3 was 10 residues long and the sequence of this peptide is shown in fig.2. Since peptide P1 follows peptide P2, peptide P3 must be placed at the C-terminus of the molecule, and indeed this peptide shows considerable sequence homology with the C-terminal 10 residues of calf thymus HMG14 (fig.2). Summation of the sizes of peptide P1 to P3 gives an overall size to the molecule of 121 residues. This is significantly larger than calf thymus HMG14 (100 residues) but it is consistent with the relative SDS gel electrophoretic mobilities of the 2 proteins, where chicken erythrocyte HMG14 runs slower than (i.e. is larger than) calf thymus HMG14.

Although calf and chicken HMG14 show considerable sequence homology up to residue 39 in the chicken sequence (7 changes in 39 residues, fig.1) part of this chicken HMG14 sequence shows greater sequence homology with calf thymus HMG17 than with calf thymus HMG14. The chicken erythrocyte HMG14 sequence residues 12–37 is identical to the sequence of HMG17 residues 17 to 42 but shows three significant differences to the comparable calf thymus HMG14 sequence. This homologous region of HMG17 exhibits one of the main structural features of HMG17, namely a region of high density of proline residues (6 proline residues in 10 residues). In the corresponding calf thymus HMG14 sequence 3 of the proline residues (at residues 30, 32 and 34) are missing. Chicken erythrocyte HMG14 is therefore structurally more similar to calf thymus HMG17 than HMG14 in this region. Further sequence comparison shows residues 40 to 52 in chicken erythrocyte HMG14 to have little sequence homology with calf thymus HMG14. These observations on the sequence differences between calf and chicken HMG14, together with the greater size of chicken erythrocyte HMG14, have revealed considerable differences in protein HMG14 from these two sources. This is in contrast to the strong sequence conservation shown by HMG17 from calf and chicken tissue [19], and suggests that certain aspects of the function of HMG14 in these two tissues might differ.

We have recently isolated HMG14 from chicken thymus and have found the amino acid analysis and amino terminal 28 residues to be identical to that of the chicken erythrocyte proteins (unpublished results). It seems reasonable to assume, therefore, that the chicken erythrocyte HMG14 protein discussed here is representative of HMG14 from chicken tissue in general and is not peculiar to erythrocyte chromatin.

3.2. *Trout liver proteins C and D*

The first 27 residues of protein C were determined and are shown in fig.1. Residue 21 was probably arginine, but was not unambiguously identified because the arginine peak on back hydrolysis at this step was obscured by a large ammonia peak. The sequence of the first 18 residues of protein D is shown in fig.1. Sequence analysis of protein D gave equimolar amounts of both glutamine and glycine at residue 5. Thereafter two residues were obtained in equal amounts at most steps, these being 50% glycine and alanine at cycle 6, 100% Ala at cycle 7, 50% each Gly and Ala at cycle 8 etc. This result was consistent with there being two populations of protein D, one with the N-terminal sequence Pro-Lys-Arg-Lys-Gln-Gly-Ala-Ala-...., and one where the glutamine residue had been deleted giving the sequence Pro-Lys-Arg-Lys-Gly-Ala-Ala-.... Identical results were obtained on a second sequenator run on a separate preparation of protein D, which discounted any error in the sequencing procedure. Additionally, we have recently isolated protein D from trout testis tissue by a somewhat different procedure (unpublished results) and again

the same "staggered" sequence was obtained. Because of this heterogeneity in the sequence analysis of protein D, only 18 residues could be identified unambiguously.

Comparison of the amino acid analyses of proteins C and D with the amino acid analyses of HMG 14 and 17 from other sources (table 1) shows many similarities between these proteins. Like HMG 14 and 17, proteins C and D lack the amino acid cysteine, methionine, tryptophan, phenylalanine and tyrosine, have a very low content of hydrophobic amino acids (<2%), a high content of charged amino acids (>45%) and have a high alanine content. This similarity between proteins C and D, and HMG14 and 17 is further revealed in a comparison of their N-terminal sequences (fig.1). Considerable homology exists between all the proteins shown in fig.1, particularly around the arginine residues. DNA binding studies with HMG17 have previously shown the N-terminal region of HMG17, in particular the arginine residues, to be involved in binding to DNA [20]. It seems likely therefore that the N-terminal regions of all the proteins shown in fig.1 have the common functions of binding to DNA. However, with the limited sequence data available for proteins C and D it is not possible to say which of these proteins correspond to HMG14 and 17. Proteins C and D run in the same position as HMG17 and HMG14 respectively on SDS gel electrophoresis. However, protein H6 (which is present in trout liver) has previously been assumed to be the trout 'equivalent' of calf thymus HMG17 because of the strong sequence homology between the two proteins. It is possible that proteins C and D represent variants of HMG14. Alternatively, proteins C and D may correspond to mammalian HMG14/HMG17, and H6 could be trout specific or be present in mammalian tissues but as yet undetected. Further sequence studies on proteins C and D will allow comparison of the sequences of the C-terminal regions of the molecules, when a more detailed comparison of these proteins can be made.

Acknowledgements

The authors wish to thank Dr J. R. B. Hastings for the amino acid analyses. The provision of trout tissue by Mr C. W. Lonsdale of Latimer Park Farm, Chesham, Bucks., and the assistance of his staff is gratefully acknowledged. This work was supported by a grant to the Chester Beatty Research Institute (Institute of Cancer Research: Royal Cancer Hospital) from the Medical Research Council.

References

[1] Walker, J. M., Goodwin, G. H., Smith, B. J. and Johns, E. W. (1980) In: Comprehensive Biochemistry, Vol. 19B/II (Neuberger, A., ed.), Elsevier/North-Holland, Biomedical Press, (in press).
[2] Goodwin, G. H., Brown, E., Walker, J. M. and Johns, E. W. (1980) Biochim. Biophys. Acta (In press).
[3] Levy, W. B. and Dixon, G. H. (1978) Nucl. Acids Res. 5, 4155–4163.
[4] Weisbrod, S. and Weintraub, H. (1979) Proc. Natl. Acad. Sci. USA 76, 630–634.
[5] Walker, J. M., Goodwin, G. H. and Johns, E. W. (1979) FEBS Lett. 100, 394–398.
[6] Walker, J. M., Hastings, J. R. B. and Johns, E. W. (1977) Eur. J. Biochem. 76, 461–468.
[7] Watson, D. C., Wong, N. C. W. and Dixon, G. H. (1979) Eur. J. Biochem. 95, 193–198.
[8] Goodwin, G. H., Walker, J. M. and Johns, E. W. (1979) in: The Cell Nucleus (Busch, H., ed.) Vol. 4, Academic Press, New York.
[9] Rabbani, A., Goodwin, G. H. and Johns, E. W. (1978) Biochem. J. 173, 497–505.
[10] Rabbani, A., Goodwin, G. H. and Johns, E. W. (1978) Biochem. Biophys. Res. Commun. 81, 351–358.
[11] Sterner, R., Boffa, L. C. and Vidali, G. (1978) J. Biol. Chem. 253, 3830–3836.
[12] Watson, D. C., Peters, E. H. and Dixon, G. H. (1977) Eur. J. Biochem. 74, 53–60.
[13] Rabbani, A., Goodwin, G. H., Walker, J. M., Brown, E. and Johns, E. W. (1980) 109, 294–298.
[14] Spiker, S., Mardian, J. K. W. and Isenberg, I. (1978) Biochem. Biophys. Res. Commun. 82, 129–135.
[15] Alfageme, C. R., Rudkin, G. T. and Cohen, L. H. (1976) Proc. Natl. Acad. Sci. USA 73, 2038–2042.
[16] Franco, L., Montero, F. and Rodriguez-Molina, J. J. (1977) FEBS Lett. 78, 317–320.
[17] Walker, J. M. and Johns, E. W. (1980) Biochem. J. 185, 383–386.
[18] Hunkapiller, M. W. and Hood, L. E. (1978) Biochemistry 17, 2124–2129.
[19] Walker, J. M., Stearn, C. and Johns, E. W. (1980) FEBS Lett. 112, 207–210.
[20] 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.