Amino acid sequence of the N-terminal domain of calf thymus histone H2A.Z

Dorothy J. Ball, Clive A. Slaughter*, Preston Hensley+ and William T. Garrard

*Division of Molecular Biology, The Department of Biochemistry, *The Department of Microbiology, The University of Texas Health Science Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75235* and *+The Department of Biochemistry, Georgetown University Medical Center, Washington, DC 20007, USA*

Received 25 January 1983

The minor histone H2A subtype, H2A.Z, has been purified to homogeneity from calf thymus and subjected to automated Edman degradation. The sequence of the first 30 amino acids possesses only 60% homology with major H2A subtypes of the same tissue. This sequence difference is more extreme than that exhibited between evolutionarily distant major H2A subtypes. However, an analysis of secondary structure reveals that H2A.Z and major H2A subtypes exhibit the same general topographical features within their N-terminal domains.

Histone H2A.Z *Nucleosome* *Evolution* *Histone variant* *Chromatin* *Amino acid sequence*

1. INTRODUCTION

Although the nucleosome contains an evolutionarily conserved protein core of 8 histone molecules [1], a considerable complexity exists in histone octamers due to the presence of non-allelic histone variants [2,3]. One of these variants, termed M1 and more recently H2A.Z [4,5] has been proposed to be a member of the histone H2A family [5,6]. This protein appears to be evolutionarily conserved [7], is enriched in transcriptionally active chromatin [8], and exhibits unique synthetic patterns during early development and the cell cycle [7,9]. To determine how these novel biological properties might be reflected in primary structure, we subjected purified calf thymus H2A.Z to automated Edman degradation and report here the first sequence information on this protein.

2. MATERIALS AND METHODS

2.1. *Purification of histone H2A.Z*

Calf thymus core histones [10,11] were electrophoretically separated on preparative Triton–acid–urea slab gels (15 × 14 × 0.3 cm; 20 mg/gel), and 0.5 cm wide strips cut parallel to the direction of electrophoresis were stained with Coomassie blue to locate histone H2A.Z [12]. Appropriate unstained regions were excised, minced, and electroeluted protein [13] was dialyzed against distilled water at 4°C and lyophilized. Resulting crude H2A.Z was dissolved in 8 M urea, 5% acetic acid, 5% 2-mercaptoethanol and separated electrophoretically on preparative acid–urea slab gels (15 × 14 × 0.3 cm; 0.6 mg/gel). Resulting protein was recovered as described above. The gel systems used have been described in [14–16].

2.2. *Amino acid analysis*

Hydrolysates were fractionated on an Aminco Aminalyzer high-performance liquid chromatography system as in [17] except that the column was run at 49°C without ammonia traps. Individual amino acids were quantitated by comparing integrated peak areas with those of standards. Values for valine and isoleucine were taken from a 72 h hydrolysate, while values for serine and threonine were estimated by extrapolation to zero hydrolysis time.

Published by Elsevier Science Publishers
00145793/83/0000–0000/$3.00

2.3. *Automated sequence analysis*

Automated sequencing was performed using a Beckman 890C amino acid sequencer, modified by addition of a cold trap and a 0.1 M Quadrol program with the non-protein carrier polybrene [18,19]. The phenylthiohydantoin amino acid derivatives were identified by high-pressure liquid chromatography using a Waters Associates Radial-PAK C18 and a Radial-PAK Nitrite CN cartridge independently for each identification [19].

3. RESULTS AND DISCUSSION

Histone H2A.Z constitutes only ~2% of the total H2A protein of calf thymus chromatin [4]. We isolated 0.5 mg calf thymus H2A.Z from 320 mg of total core histones by employing preparative gel electrophoresis and electroelution. Thus, about a 200-fold purification of H2A.Z was achieved with ~30% yield. The isolated protein was electrophoretically homogeneous in three different gel systems (fig.1), and possessed an amino acid composition similar to the major calf thymus H2A subtype, H2A.1 (table 1).

Peptide mapping studies have provided only

Table 1

Amino acid composition of calf thymus histone H2A.Z

Amino acid	H2A.Z	H2A.1
Asp	4.2	6.2
Thr	6.0	3.9
Ser	6.3	3.1
Glu	8.7	9.3
Pro	–	3.9
Gly	11.8	10.9
Ala	14.3	13.2
Cys	–	0
Val	5.9	6.2
Met	–	0
Ile	7.3	4.7
Leu	10.9	12.4
Tyr	1.6	2.3
Phe	1.0	0.8
His	4.9	3.1
Lys	11.2	10.9
Arg	7.9	9.3

The values presented represent mol % and are the average of 4 analyses on H2A.Z. The values presented for H2A.1 are calculated from the known amino acid sequence [20]

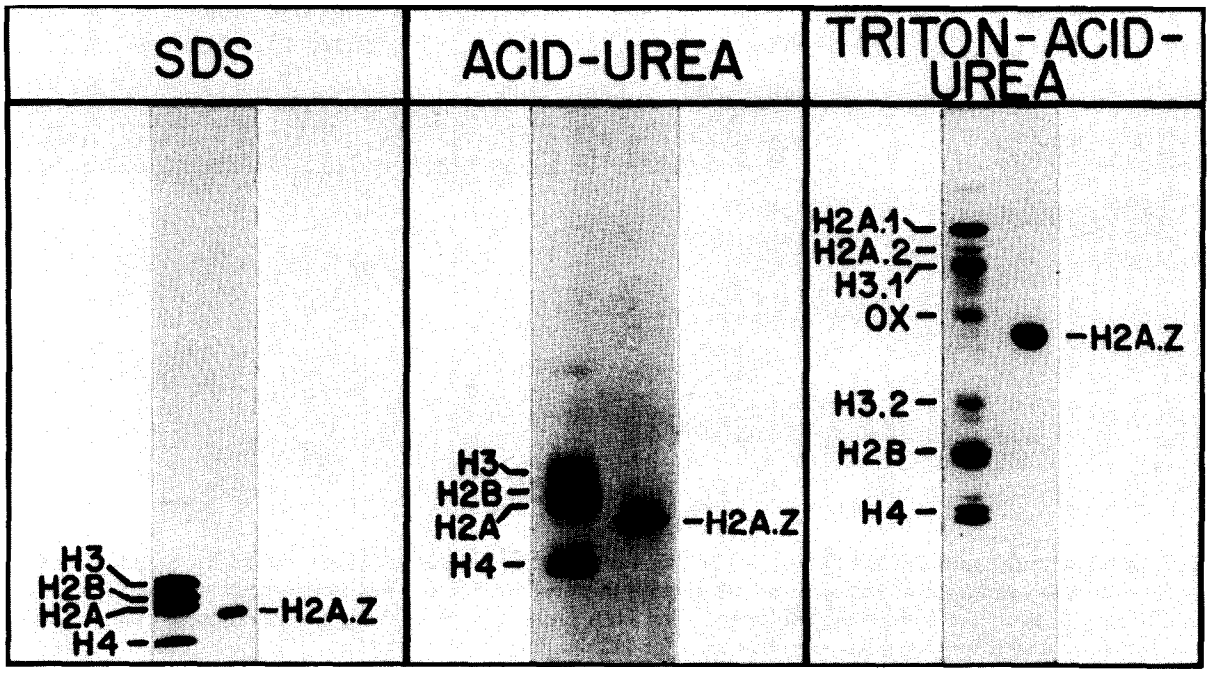

Fig.1. Electrophoretic profiles of calf thymus core histones and purified histone H2A.Z.

suggestive evidence that H2A.Z is a member of the H2A protein family. One-dimensional resolution of tryptic peptides has shown that only 2 out of 8 arginine and 0 out of 7 lysine containing peptides are shared between H2A.Z and major H2A variants [5,7]. We therefore performed an automated sequence analysis which proved to be successful because unlike other H2A species, H2A.Z does not have a blocked N-terminus. As indicated by the boxed regions in fig.2, the N-terminal domain of H2A.Z possesses 60% sequence homology with calf thymus H2A.1. Relative to H2A.1, the N-terminal region of H2A.Z has a 2 residue extension and 10 amino acid substitutions. These differences all occur within the first 21 amino acids of H2A.Z and can be attributed to 5 single and 5 double base changes in the respective codons. Because of 4 radical amino acid substitutions at positions 4, 5, 8 and 17 in H2A.Z, the net charge of the N-terminal domain of H2A.Z is +6, as opposed to +8 for the corresponding region of H2A.1. Therefore, the N-terminal region of H2A.Z is less basic than that of H2A.1.

To evaluate the evolutionary significance of the sequence differences between calf thymus histones H2A.Z and H2A.1, we have further compared the N-terminal sequence of calf thymus H2A.1 with those published for other H2A proteins. Evolutionarily distant H2A proteins possess more homology with calf thymus H2A.1, within their N-terminal domains, than does calf thymus H2A.Z (with the exception of H2As from wheat and yeast) (fig.2). However, from the H2A N-terminal sequences shown in fig.2, it is possible to generate a hypothetical peptide with 87% sequence homology to H2A.Z (not shown). Furthermore, the most highly conserved region among H2A subtypes, which yields a phenylalanine containing tryptic peptide, is present in H2A.Z, in agreement with [5]. Thus, we conclude that H2A.Z is indeed a member of the H2A protein family. Apparently the gene(s) encoding H2A.Z arose from a duplication event(s) which must have occurred early in the evolution of eukaryotes, because histone H2A.Z is found in organisms as primitive as sea urchin [7].

Do differences in primary structure between calf thymus histones H2A.Z and H2A.1 result in potential alterations in secondary structure? To address this question we employed an algorithm which predicts secondary structures [33]. Fig.3 shows that H2A.Z and H2A.1 possess essentially identical predicted secondary structures within their N-terminal domains. It is noteworthy that even a region which exhibits a 50% difference in amino acid sequence (residues 12–17 in H2A.Z) still exhibits the same high probability for α-helix formation. Therefore, the predicted secondary

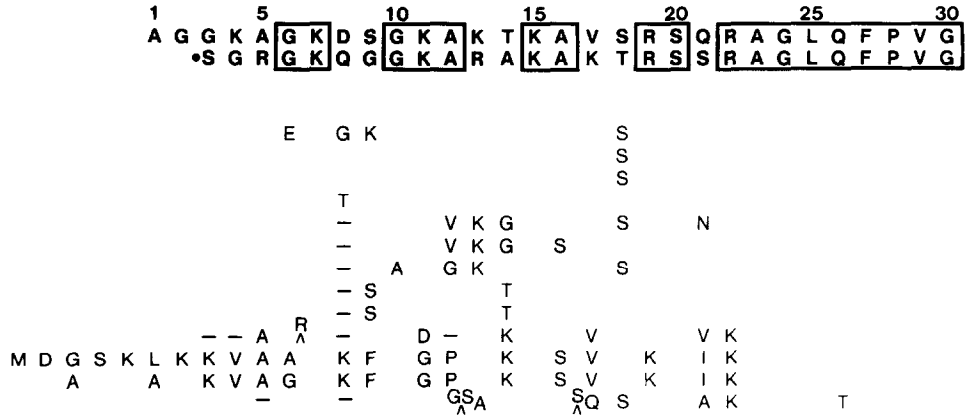

Fig.2. Amino acid sequences of the N-terminal domains of calf thymus histones H2A.Z and H2A.1 and other H2A proteins. Common amino acids between the calf thymus proteins are boxed. Only amino acids different from those found in calf thymus H2A.1 are indicated for the other H2A proteins: (–) deletion; (∧) insertion; (*) deduced from DNA sequence. The following references were used for H2A sequences. Calf thymus [18]; mouse leukemia [21]; rat leukemia [22]; chicken erythrocyte [23,24]; trout [25]; *Drosophila* [26]; cuttlefish [27]; sea urchin 1 (*P. miliaris* gonad [28]; sea urchin 2 (*P. miliaris*) [29]; sea urchin 3 (*S. purpuratus*) [30]; wheat 1, 2 and 3 [31]; and yeast 1 and 2 [32].

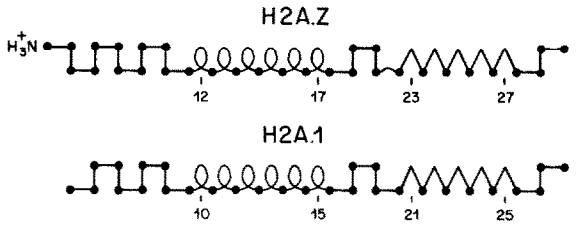

Fig.3. Comparison of the predicted secondary structures of the N-terminal domains of calf thymus histones H2A.Z and H2A.1. The probability of α-helix, β-sheet, β-turn, or coil as a function of amino acid sequence was calculated according to [33]. Residues are represented schematically in helical (α), β-sheet (β), β-turn (β-turn) and coil (coil) conformational states.

structure of the N-terminal domain of H2A.Z is consistent with the conclusion that the protein is a member of the histone H2A class.

The sequence of the N-terminal domain of H2A.Z exhibits two features that are consistent with the finding that H2A.Z is enriched in transcriptionally active chromatin [8]:

(1) N-Terminal regions of the core histones may function in stabilizing internucleosomal contacts for the packaging of polynucleosomal chains into higher order structures [34]. Thus, the less basic N-terminal domain of H2A.Z might render polynucleosomal chains containing this protein more accessible for transcription.

(2) H2A.Z is not phosphorylated and lacks the sequence *N*-AcSerGlyArg which contains the apparent exclusive site of phosphorylation among H2A subtypes [35]; the level of phosphorylated H2A has been directly correlated with the degree of heterochromatization [36,37].

The sequence data provided here should permit identification of recombinant DNA clones carrying H2A.Z genes. The complete sequence of the protein can then be determined more readily than by classical techniques of protein chemistry. An analysis of flanking sequences and gene organization relative to those of the other histones may prove revealing because histone H2A.Z synthesis is not co-ordinated with S phase [9]. The sequence data provided here should also permit elicitation of antibodies to chemically synthesized H2A.Z peptides. These antibodies could be used to isolate nucleosomes containing stoichiometric amounts of H2A.Z. The protein components, DNA sequence composition, and arrangement along polynucleosomal chains of such nucleosomes could then be studied.

ACKNOWLEDGEMENTS

We thank Dr Brent Reed and Messrs Lonnie Sorrells and Melvin Dews for performing the amino acid analysis. We are indebted to Dr J. Donald Capra in whose laboratory the automated sequencing was performed and to Dr Michael Karels for making available to us the secondary structure prediction routine. This research was supported by grants from NIH (GM22201 and GM29935) and The Robert A. Welch Foundation (I-823) to W.G., by NIH grant AI12127 to J. Donald Capra, and by NIH grant GM28731 to R.H.; C.S. was supported by an NIH postdoctoral fellowship (GM07710).

REFERENCES

[1] McGhee, J.D. and Felsenfeld, G. (1980) Annu. Rev. Biochem. 49, 1115–1156.
[2] Franklin, S.G. and Zweidler, A. (1977) Nature 266, 273–275.
[3] Bafus, N.L., Albright, S.C., Todd, R.D. and Garrard, W.T. (1978) J. Biol. Chem. 253, 2568–2574.
[4] Zweidler, A. (1976) in: Organization and Expression of Chromosomes (Allfrey, V.G. et al. eds) pp.187–196, Dahlem Konferenzen, Berlin.
[5] West, M.H.P. and Bonner, W.M. (1980) Biochemistry 19, 3238–3245.
[6] Albright, S.C., Nelson, P.P. and Garrard, W.T. (1979) J. Biol. Chem. 254, 1065–1073.
[7] Wu, R.S., Nishioka, D. and Bonner, W.M. (1982) J. Cell Biol. 93, 426–431.
[8] Gabrielli, F., Hancock, R. and Faber (1981) Eur. J. Biochem. 120, 363–369.
[9] Wu, R.S. and Bonner, W.M. (1981) Cell 27, 321–330.
[10] Johns, E.W. (1964) Biochem. J. 92, 55–59.
[11] Oliver, D., Sommer, K.R., Panyim, S., Spiker, S. and Chalkley, R. (1972) Biochem. J. 129, 349–353.
[12] Albright, S.C., Wiseman, J.M., Lange, R.A. and Garrard, W.T. (1980) J. Biol. Chem. 255, 3673–3684.
[13] Mardian, J.K.W. and Isenberg, I. (1978) Anal. Biochem. 91, 1–12.

[14] Panyim, S. and Chalkley, R. (1969) Arch. Biochem. Biophys. 130, 337–346.
[15] Thomas, J.O. and Kornberg, R.D. (1975) Proc. Natl. Acad. Sci. USA 72, 2626–2630.
[16] Zweidler, A. (1978) Methods Cell Biol. 17, 223–233.
[17] Drescher, D.G. and Lee, K.S. (1978) Anal. Biochem. 84, 559–569.
[18] McCumber, L.J., Qadeer, M. and Capra, J.D. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed) pp.165–172, Elsevier Biomedical, Amsterdam, New York.
[19] Klapper, D.G., Wilde, C.W. iii and Capra, J.D. (1978) Anal. Biochem. 60, 147–150.
[20] Yeoman, L.C., Olson, M.O.J., Sugano, N., Jordan, J.J., Taylor, C.W., Starbuck, W.C. and Busch, H. (1972) J. Biol. Chem. 247, 6018–6023.
[21] Blankstein, L.A., Stollar, B.D., Franklin, S.G., Zweidler, A. and Levy, S.B. (1977) Biochemistry 16, 4557–4562.
[22] Laine, B., Kmiecik, D., Sautiere, P. and Biserte, G. (1978) Biochimie 60, 147–150.
[23] Laine, B., Sautiere, P. and Biserte, G. (1981) Biochemistry 15, 1640–1645.
[23] D'Andrea, R., Harvey, R. and Wells, J.R.E. (1981) Nucleic Acid Res. 9, 3119–3128.
[25] Bailey, G.S. and Dixon, G.H. (1973) J. Biol. Chem. 248, 5463–5472.
[26] Goldberg, M. (1979) Dissertation (Stanford University, Palo Alto CA).
[27] Wouters-Tyrou, D., Martin-Ponthieu, A., Briand, G., Sautiere, P. and Biserte, G. (1982) Eur. J. Biochem. 124, 489–498.
[28] Wouters, D., Sautiere, P. and Biserte, G. (1978) Eur. J. Biochem. 90, 231–239.
[29] Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucleic Acids Res. 8, 957–977.
[30] Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033–1044.
[31] Rodrigues, J.D.A., Brandt, W.F. and Von Holt, C. (1979) Biochim. Biophys. Acta 578, 196–206.
[32] Choe, J., Kolodrubetz, D. and Grunstein, M. (1982) Proc. Natl. Acad. Sci. USA 79, 1484–1487.
[33] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[34] Allan, J., Harborne, N., Rau, D.C. and Gould, H. (1982) J. Cell Biol. 93, 285–297.
[35] Pantazis, P. and Bonner, W.M. (1981) J. Biol. Chem. 256, 4669–4675.
[36] Gurley, L.R., Walters, R.A., Barham, S.S. and Deaven, L.L. (1978) Exp. Cell Res. 111, 373–383.
[37] Halleck, M.S. and Gurley, L.R. (1980) Exp. Cell Res. 125, 377–388.