Proc. Nat. Acad. Sci. USA
Vol. 69, No. 4, pp. 882–884, April 1972

Complete Amino-Acid Sequence of Calf-Thymus Histone III

(acetyllysine/methyllysine/acetylation/methylation)

ROBERT J. DELANGE, JOHN A. HOOPER, AND EMIL L. SMITH

Department of Biological Chemistry, School of Medicine, and the Molecular Biology Institute, University of California, Los Angeles, Calif. 90024

Contributed by Emil L. Smith, February 2, 1972

ABSTRACT **Calf-thymus histone III is a single polypeptide chain of 135 residues (combined molecular weight of 15,324) with alanine at both the amino and carboxyl ends. The NH_2-terminal region (Residues 1–53) of histone III is strongly basic (net charge of +18) and contains lysines-14 and -23, which are ϵ-*N*-acetylated in a fraction of the molecules, as well as lysines-9 and -27, which are partially ϵ-*N*-methylated. The COOH-terminal region (Residues 54–135) is only slightly basic (net charge of +4), contains most of the hydrophobic residues, and has a 29-residue sequence that lacks a basic residue. The two cysteines are in the nonbasic region at positions 96 and 110. A few sequence similarities of calf-thymus histone III with other histones have been noted.**

Histones are basic proteins complexed with DNA in the nuclei of all eukaryotic organisms. Although the functions of histones have not been precisely delineated, it seems likely that they play an important role in the regulation of transcription and replication, in addition to serving as structural components of chromosomes (1).

With the availability of improved methods for the separation of histones from each other and from other nuclear components, structural studies were begun on several histones from various organisms. The sequence of calf-thymus histone IV (f2a1, GAR histone), one of the two arginine-rich histones, was completed first (2, 3), and this was soon followed by the sequence of pea-seedling histone IV (4). Histone IV shows a remarkable preservation of sequence throughout the evolution of eukaryotes and, in addition, has a nonrandom distribution of residues and specific sites of ϵ-*N*-acetylation and ϵ-*N*-methylation (4). Some of these findings were independently substantiated in another laboratory (5), and the sequence studies have now been extended to include Novikoff hepatoma (6), rat (7), and porcine (8) histones IV.

Although partial sequences of other histones have been reported (1, 9–13), the only other complete sequence reported to date is that of calf-thymus histone IIb2 (f2b) (14, 15). We now present the complete amino-acid sequence of calf-thymus histone III (f3), the other arginine-rich histone, and the only histone that contains cysteinyl residues. A preliminary account of some of these studies has been given (16).

METHODS

Calf-thymus histone III was prepared by method 2 of Johns (17) and further purified by column chromatography on Bio-Rex 70 (18). The *S*-carboxymethyl derivative was prepared by reduction and alkylation with iodoacetate (19), and the oxidized derivative was prepared by oxidation with performic acid (20). Tryptic, chymotryptic, and cyanogen bromide peptides were obtained from the derivatized proteins by procedures similar to those described for histone IV (3, 21). Larger peptides were further hydrolyzed with trypsin, chymotrypsin, pepsin, or thermolysin to give smaller fragments. The procedures for sequential Edman degradation, hydrolysis with carboxypeptidases or aminopeptidases, and all other methods have been described (3, 19, 21).

Abbreviations: Lys(Ac), ϵ-*N*-acetyllysine; Lys$(CH_3)_{1,2}$, ϵ-*N*-monomethyllysine and ϵ-*N*-dimethyllysine.

RESULTS AND DISCUSSION

The complete amino-acid sequence of calf-thymus histone III is shown in Fig. 1. On the basis of this sequence, the molecular weight of the isoelectric protein is 15,324, assuming one *N*-acetyl at either Residue 14 or 23 and two monomethyllysines, one each at Residues 9 and 27. Cyanogen bromide peptides were obtained by cleavage COOH-terminal to the two methionines (Residues 90 and 120). Tryptic peptides were obtained from the maleylated histone by cleavage COOH-terminal to all arginyl residues except Residue 42, which is present in an Arg–Pro sequence. The order of the tryptic peptides within the cyanogen bromide peptides was deduced from the sequence studies on chymotryptic peptides from the protein and on thermolysin, chymotryptic, peptic, and tryptic peptides from the cyanogen bromide peptides.

Although histone III is 33 residues longer than histone IV, certain features of these two arginine-rich histones from calf thymus are similar. The NH_2-terminal regions of both histones are rich in basic residues and contain the sites of ϵ-*N*-acetylation and ϵ-*N*-methylation. Calf-thymus histone IV (3) contains only one site of methylation (Residue 20) and one site of acetylation (Residue 16), whereas calf-thymus histone III (Fig. 1) contains two sites of each lysine modification: methylation of Residues 9 and 27 and acetylation of Residues 14 and 23. The COOH-terminal regions of both histones are rich in hydrophobic residues, and include most of the aromatic amino acids. These features are summarized in Tables 1 and 2. A few sequence similarities of histones IIb1, III, and IV are given in Table 3.

One preparation of histone III indicated that acetylation was present on 25% of the molecules at Residue 14 and 29% at Residue 23. At present, we cannot determine whether any molecules are acetylated at both positions or if acetylation at one site precludes acetylation at the other site. In another preparation of histone III, Residue 23 was acetylated in 43% of the molecules. The unacetylated form of the peptide containing Residue 14 was not isolated in this preparation. In this latter preparation, 40% of the molecules contained dimethyllysine, 34% contained monomethyllysine, and 26% contained lysine at Residue 27. Similar calculations have not yet been made for Residue 9, although both monomethyl- and dimethyllysine are present at this site.

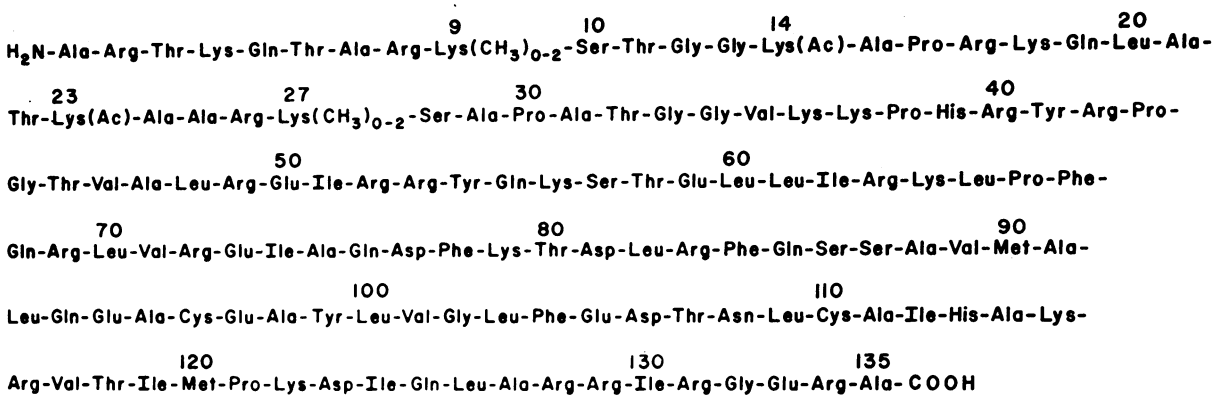

FIG. 1. Sequence of calf-thymus histone III. See the text for a description of the peptides used to establish this sequence.

From an examination of the sequences around the methylated residues in both histones (Table 2), the only feature in common is that the methylated lysine is adjacent and COOH-terminal to an arginyl residue. However, other lysyl residues (Residue 79 in histone IV, and Residues 18 and 64 in histone III) are also in this arrangement, and no evidence for their methylation was obtained. It seems probable that separate enzymes are involved in the methylation of histones III and IV, since both histones are methylated in calf thymus, but only histone III is methylated in pea seedling (4, 22). If this is so, then one might expect the sequences around the methylated sites in histone III to resemble each other more than they do the sequence around the site methylated in histone IV. Indeed, the identical sequence Ala-Arg-Lys-$(CH_3)_{1,2}$-Ser is found at both sites of methylation in histone III. The other two Arg–Lys sequences in histone III (Residues 18, 19 and 64, 65) do not have an alanyl residue NH_2-terminal to the arginyl residue, nor a seryl residue COOH-terminal to the lysyl residue; these differences may explain why these lysyl residues are not recognized by the methylating enzyme.

Similarly, the only feature common to the three sites of acetylation in the two histones (Table 2) is that the lysyl residue that is acetylated is NH_2-terminal to, and separated by two residues from, an Arg–Lys sequence. In two of the three instances this lysyl residue (in the Arg–Lys sequence) is also methylated. Therefore, a recognition site for acetylation of lysyl residues in histones might be

-Lys-X-Y-Arg-Lys.

However, since pea-seedling (4) and trout-testis histones IV (23) are acetylated at other sites that do not conform to this type of sequence, either there are additional acetylating enzymes with different specificities, or the sequence indicated is fortuitous in these three cases. Although the arginine-rich histones (III and IV) appear to be the major (if not the only) substrates for the methylating and acetylating enzymes, histone IIb2 contains the sequence *Lys*-Arg-Ser-Arg-Lys (Residues 30–34) (14), which conforms to the structure of the sites acetylated in histones III and IV. If acetylation of histone IIb2 does occur, it could be conjectured that lysyl Residue 30 is the one modified, although a similar sequence *Lys*-Lys-Arg-Lys-Arg (Residues 27–31) might also be considered. No other sequences of this type or of the type found around the methylated lysines in histone III have been observed in the sequences of other histones reported to date.

Histone III differs markedly from histone IV in having 33 more residues, including two cysteinyl residues, and in having a region of 29 residues lacking a basic residue (compared to 11 residues in histone IV). The two cysteinyl residues in this nonbasic region are separated by 13 residues, and can form

TABLE 1. *Distribution of residues in calf-thymus histones III and IV*

	Histone III		Histone IV (3)	
	Residues 1–53	Residues 54–135	Residues 1–45	Residues 46–102
Basic residues*	19	15	17	10
Acidic residues†	1	11	1	7
Net charge‡	+18	+4	+16	+3
Hydrophobic residues§	6	28	8	22
Aromatic residues	1	6	0	6
Prolyl residues	4	2	1	0
Longest sequence lacking a basic residue	8 residues (Res. 28–35)	29 residues (Res. 84–112)	7 residues (Res. 24–30)	11 residues (Res. 80–90)

* Includes lysine, arginine, histidine, and α-amino groups.

† Includes aspartic acid, glutamic acid, and α-COOH groups.

‡ Assumes that all basic and acidic groups are in charged form.

§ Includes only leucine, isoleucine, valine, methionine, tyrosine, and phenylalanine.

TABLE 2. *Sites of ϵ-N-acetylation and ϵ-N-methylation in calf-thymus histones III and IV*

	Sites of ϵ-*N*-Methylation*	Sites of ϵ-*N*-Acetylation
Histone III	9 Ala-Arg-Lys$(CH_3)_{1,2}$-Ser-Thr	14 Gly-Gly-Lys(Ac)-Ala-Pro-Arg-Lys
	27 Ala-Arg-Lys$(CH_3)_{1,2}$-Ser-Ala	23 Ala-Thr-Lys(Ac)-Ala-Ala-Arg-Lys†
Histone IV (3)	20 His-Arg-Lys$(CH_3)_{1,2}$-Val-Leu	16 Gly-Ala-Lys(Ac)-Arg-His-Arg-Lys†

* Numbers above lysyl residues indicate residue numbers in the complete sequences.

† These lysyl residues are ϵ-*N*-methylated.

TABLE 3. *Sequence similarities of histone III with other histones**

	30 40
IV (3)	**Gln-Gly-Ile-Thr-Lys-Pro-Ala-Ile-Arg-***Arg-Leu-***Ala-***Arg***	***Arg-*Gly-Gly-Val-Lys-***Arg***
	22 † ‡ 32
III	**Gln-Leu-Ala-Thr-Lys-** **Ala-Ala-Arg-***Lys-Ser-***Ala-***Pro-*Ala-***Thr-*Gly-Gly-Val-Lys-***Lys***
	75 84
IV (3)	**His-Ala-Lys-Arg-**Lys-Thr-**Val-Thr-Ala-Met**
	113 120
III	**His-Ala-Lys-Arg-** **Val-Thr-Ile-Met**
IIbl (see ref. 1)§	***Glu-*Leu-Ala-Gly-***Asn-*Ala-Ala-Arg**	**Lys-Thr-Asx-Leu-Arg**
	20 † 80
III	***Gln-*Leu-Ala-Thr-***Lys-*Ala-Ala-Arg**	**Lys-Thr-Asp-Leu-Arg**
	8 ‡ † 18
III	**Arg-Lys-**Ser-Thr-***Gly-*Gly-**Lys-Ala-***Pro-*Arg-Lys**
	17 † 26 ‡
III	**Arg-Lys-**Gln-Leu-***Ala-*Thr-**Lys-Ala-***Ala-*Arg-Lys**

* **Bold face** type indicates identity; *italics* indicate residues whose nucleotide triplet codes differ by a single base change.
† These lysyl residues are ϵ-*N*-acetylated.
‡ These lysyl residues are ϵ-*N*-methylated.
§ The complete sequence of histone IIb1 is not known.

either intramolecular or intermolecular disulfide bonds (24). This feature permits the formation of dimers and higher polymers of calf-thymus histone III; these have been observed in a number of studies (see Ref. 22, for example). The molecules with intramolecular disulfide bonds migrate more rapidly in in acrylamide gel electrophoresis than do the reduced molecules, whereas all of the multimers migrate more slowly (24). It seems reasonable to suppose that the cysteinyl residues play important roles in the structural and functional changes that occur during processes such as transcription, replication, mitosis, etc. Indeed, Sadgopal and Bonner (25) have obtained evidence that histone III is primarily present in interphase as the reduced molecule, but is largely in the disulfide form in metaphase chromatin. It is of interest that the histones III of rodents and all lower animals tested have only one cysteinyl residue and can, therefore, only form dimers, whereas histones III of the rabbit and higher animals have two cysteinyl residues (24).

From the preceding comments it can be inferred that histone III (like histone IV) probably contains primary binding sites for DNA in the NH_2-terminal region of the molecule, where modification reactions (acetylation, methylation, and phosphorylation) might have effects on the degree of repression of DNA. The COOH-terminal region of histone III might have fewer interactions with DNA and be free to assume its own specific conformations or to interact with other components of chromatin. The cysteinyl residues might be expected to play an important role in any such interactions. Since Fambrough and Bonner observed very similar peptide maps for histone III from pea and calf (22), it seems probable that a considerable amount of the sequence has been preserved throughout evolution, as in the case of histone IV. This can best be interpreted as indicating that the function of this molecule is so closely linked to its entire structure that only a few minor changes can be tolerated.

We express sincere appreciation to Miss Dorothy McNall and Linda Anderson for technical assistance. Special thanks go to Dr. Douglas M. Fambrough, who prepared some of the histone III in Dr. James Bonner's laboratory. This work was supported by Grant GM 11061 of the National Institute of General Medical Sciences of the United States Public Health Service. Dr. Hooper is a recipient of a postdoctoral fellowship from the National Institutes of Health, no. 5 FO2 AM 44230-02.

1. DeLange, R. J. & Smith, E. L. (1971) *Annu. Rev. Biochem.* **40**, 279–314.
2. DeLange, R. J., Smith, E. L., Fambrough, D. M. & Bonner, J. (1968) *Proc. Nat. Acad. Sci. USA* **61**, 1145–1146.
3. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem.* **244**, 319–334.
4. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1969) *J. Biol. Chem.* **244**, 5669–5679.
5. Ogawa, Y., Quagliarotti, G., Jordan, J., Taylor, C. W., Starbuck, W. C. & Busch, H. (1969) *J. Biol. Chem.* **244**, 4387–4392.
6. Wilson, R. K., Starbuck, W. C., Taylor, C. W., Jordan, J. & Busch, H. (1970) *Cancer Res.* **30**, 2942–2951.
7. Sautière, P., Tyrou, D., Moschetto, Y. & Biserte, G. (1971) *Biochimie* **53**, 479–483.
8. Sautière, P., Breynaert, M.-D., Moschetto, Y. & Biserte, G. (1970) *C. R. Acad. Sci. Paris*, **271**, 364–365.
9. Olson, M. O. J., Sugano, N., Yeoman, L. C., Johnson, B. R. & Starbuck, W. C. (1971) *Fed. Proc.*, **30**, 1295.
10. Hayashi, H. & Iwai, K. (1971) *J. Biochem.* **70**, 543–547.
11. Rall, S. C. & Cole, R. D. (1971) *J. Biol. Chem.* **246**, 7175–7190.
12. Greenaway, P. J. (1971) *Biochem. J.* **124**, 319–325.
13. Greenaway, P. J. & Murray, K. (1971) *Nature New Biol.* **229**, 233–238.
14. Iwai, K., Ishikawa, K. & Hayashi, H. (1970) *Nature* **226**, 1056–1058.
15. Hnilica, L. S., Kappler, H. A. & Jordan, J. J. (1970) *Experientia* **26**, 353.
16. DeLange, R. J., Smith, E. L. & Bonner, J. (1970) *Biochem. Biophys. Res. Commun.* **40**, 989–993.
17. Johns, E. W. (1964) *Biochem. J.* **92**, 55–59.
18. Murray, K. (1966) *J. Mol. Biol.* **15**, 409–419.
19. DeLange, R. J. (1970) *J. Biol. Chem.* **245**, 907–916.
20. Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611–621.
21. DeLange, R. J., Fambrough, D. M., Smith, E. L. & Bonner, J. (1968) *J. Biol. Chem.* **243**, 5906–5913.
22. Fambrough, D. M. & Bonner, J. (1968) *J. Biol. Chem.* **243**, 4434–4439.
23. Sung, M. & Dixon, G. H. (1970) *Proc. Nat. Acad. Sci. USA* **67**, 1616–1623.
24. Panyim, S., Sommer, K. R. & Chalkley, R. (1971) *Biochemistry* **10**, 3911–3917.
25. Sadgopal, A. & Bonner, J. (1970) *Biochim. Biophys. Acta* **207**, 227–239.