The Structures of Histones

Robert J. DeLange and Emil L. Smith*

Department of Biological Chemistry, University of California School of Medicine, Los Angeles, California 90024

Received June 8, 1972

Histones are basic proteins that are present in the chromosomes of eukaryotic organisms, *i.e.*, those with true nuclei.[1] By virtue of their basicity the histones are at least partially responsible for neutralizing the anionic phosphate groups of DNA. Although the evidence to date strongly suggests that histones may be involved in the regulation of replication or transcription,[1,2] the precise roles of these proteins are still unknown. In attempts to understand histone functions, considerable activity has developed in the past few years toward defining precisely the structures of these unusual proteins.[1,2]

In contrast to an earlier view that there might be a large variety of histones, it now appears that there are only a relatively limited number of different kinds of histones as defined by their primary amino acid sequences; however, each type of histone may exist in a variety of forms by virtue of posttranscriptional modifications (those occurring after synthesis of the polypeptide chains) of certain amino acid side chains by methylation, acetylation, and phosphorylation.[1,2] Although there is, as yet, no definitive evidence regarding histone function, the discovery of the extreme conservatism in evolution of the amino acid sequence of at least one type of histone,[3] and undoubtedly of others,[2] has generated great interest in histone structure, largely because it is recognized that such conservatism in macromolecules must reflect very precisely defined roles for these proteins.

Some of the characteristics of each of the major histone fractions from calf thymus are given in Table I. Each individual type of histone is usually recognized by its relative content of arginine and lysine. The classification system of Rasmussen, Murray, and Luck[4] will be used in this article, but Table I indicates the designations in the other major nomenclature system.

Calf thymus contains three histone fractions (Table I) in which there is more lysine than arginine and two in which there is more arginine than lysine. Subfractions have been observed for each type. Some of these subfractions are the result of chemical modification of the amino acid side chains, whereas in other instances subfractions manifest variations in the amino acid sequences (see below).

Although Table I describes calf thymus histones, the same characteristics are generally valid for histones from all other animal or plant tissues. The major differences in the histones from different organisms are: (a) differing numbers of subfractions (*e.g.*, at least five subfractions of histone I in rabbit thymus[5] and nine subfractions of histone IV from trout testis[6]); (b) quantitative differences, particularly in the distribution among subfractions; (c) differences in electrophoretic mobility; and (d) occasionally, the presence of a unique histone fraction in certain cells or tissues such as in nucleated erythrocytes,[7] trout tissues,[8] or sea cucumber gonads.[9] Arginine-rich histones are the least variable in structure from species to species,[2,3] whereas the very lysine-rich histones (I) seem to be the most variable.[2,10]

Primary Structures and Properties

The complete amino acid sequences of several histones are now known, and partial structures have been reported for others (see below). Each of these various types will be discussed in turn.

Histone IV. Calf thymus histone IV was the first histone sequenced as reported by DeLange, *et al.*[11] (Figure 1), and later by Ogawa, *et al.*[12] These studies convincingly demonstrated that this histone possesses a unique sequence and is present in substantial amount in

Since receiving his B.S. degree from Brigham Young University in 1961, Dr. DeLange's research has principally involved studies on protein structure and function with special emphasis on amino acid sequences. This interest was generated during study for the Ph.D. under Professor E. G. Krebs at the University of Washington and was continued during postdoctoral work with Professor E. L. Smith at UCLA. Since 1967 he has been a member of the faculty in the Department of Biological Chemistry at UCLA and is presently Associate Professor.

Emil L. Smith received his B.S. and Ph.D. degrees from Columbia University. He was an instructor in Biophysics at Columbia from 1936 to 1938, a Guggenheim Fellow at Cambridge and Yale, and then Fellow at the Rockefeller Institute. After 4 years in industry, he joined the faculty of the University of Utah School of Medicine in 1946. Since 1963, he has been Professor and Chairman of the Department of Biological Chemistry at UCLA School of Medicine. His main research interests are in enzyme and protein chemistry.

(1) For comprehensive background information on histones the reader is referred to "Histones and Nucleohistones," D. M. P. Phillips, Ed., Plenum Press, New York, N. Y., 1971, for the more recent work, and to "The Nucleohistones," J. Bonner and P. Ts'o, Ed., Holden-Day, San Francisco, Calif., 1964, for the earlier work.

(2) R. J. DeLange and E. L. Smith, *Annu. Rev. Biochem.*, **40**, 279 (1971).

(3) R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, **244**, 5669 (1969).

(4) P. S. Rasmussen, K. Murray, and J. M. Luck, *Biochemistry*, **1**, 79 (1962).

(5) M. Bustin and R. D. Cole, *J. Biol. Chem.*, **244**, 5286 (1969).

(6) M. Sung and G. H. Dixon, *Proc. Nat. Acad. Sci. U. S.*, **67**, 1616 (1970).

(7) J. M. Neelin in "The Nucleohistones," J. Bonner and P. Ts'o, Ed., Holden-Day, San Francisco, Calif., 1964, pp 66–71.

(8) D. T. Wigle and G. H. Dixon, *J. Biol. Chem.*, **246**, 5636 (1971).

(9) J. Phelan, J. Subirana, and R. D. Cole, personal communication.

(10) S. C. Rall and R. D. Cole, *J. Biol. Chem.*, **246**, 7175 (1971). The results reported in this paper have been updated in a personal communication from Professor Cole as indicated in Figure 5.

(11) R. J. DeLange, E. L. Smith, D. M. Fambrough, and J. Bonner, *Proc. Nat. Acad. Sci. U. S.*, **61**, 1145 (1968); R. J. DeLange, D. M. Fambrough, E. L. Smith, and J. Bonner, *J. Biol. Chem.*, **244**, 319 (1969).

(12) Y. Ogawa, G. Quagliarotti, J. Jordan, C. W. Taylor, W. C. Starbuck, and H. Busch, *ibid.*, **244**, 4387 (1969).

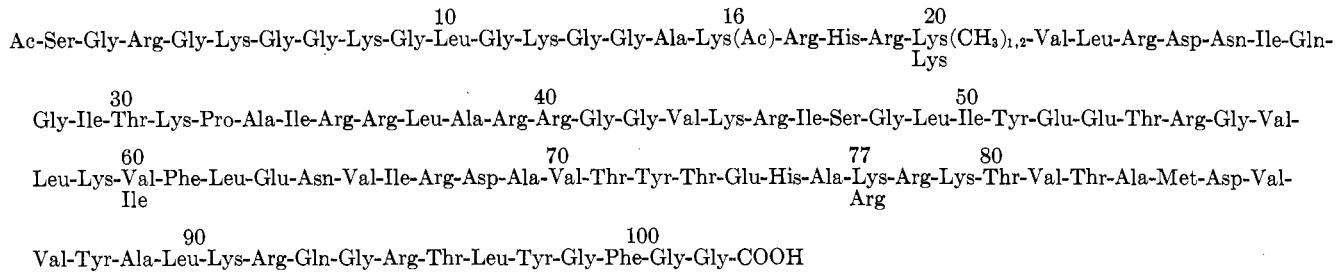

Figure 1. Amino acid sequences of calf thymus (continuous sequence) and pea seedling histones IV (f2a1). Residues which differ in the pea seedling histone are shown below the continuous sequence. See ref 3 and 11.

Table I
Characterization of Calf Thymus Histones[a]

Class	Fraction	Minimum number of subfractions	Ratio of lysine to arginine	Total residues	Mol wt	NH_2 terminal	COOH terminal
Very lysine rich	I (f1)	3–4	20	~215	~21,000	Ac-Ser	Lys
Lysine rich	IIb1 (f2a2)	2	1.25	129	~14,500	Ac-Ser	Lys
	IIb2 (f2b)	0–2	2.5	125	13,774	Pro	Lys
Arginine rich	III (f3)	3	0.72	135	15,324	Ala	Ala
	IV (f2a1)	4	0.79	102	11,282	Ac-Ser	Gly

[a] Information described in this table has been compiled from the references cited for each histone in this article and from ref 2.

chromatin. These results are to be contrasted with earlier views postulating a great multiplicity of histones (even one per gene) which would have precluded the possibility that a histone fraction obtained in good yield could consist of a unique protein.

The studies on histone IV showed that the basic residues are present in clusters and are not evenly distributed throughout the molecule. As summarized in Table II, the NH_2-terminal region of histone IV is highly basic (net charge of +16), whereas the COOH-terminal region is only slightly basic (net charge of +3) and contains most of the hydrophobic (including all of the aromatics) and hydroxyl-containing residues. These features were interpreted[11] as indicating that the primary binding sites of histone IV for DNA are located in the NH_2-terminal region of the protein, whereas the COOH-terminal region probably has structural features of its own (not superimposed by DNA) which might allow it to interact with other components of the nucleus.

Murray[13] reported in 1964 the presence of N^{ϵ}-methyllysine in histones and Paik and Kim[14] demonstrated that both the monomethyl and dimethyl derivatives were present. It remained for the sequence studies[11] to show that only one of the 11 lysyl residues in histone IV was methylated (Figure 1). It is noteworthy that, when the first 20 residues of histone IV are arranged as a helical cylinder,[15] the methyllysine residue is positioned along a "face" formed by four glycyl residues. Since glycyl residues are evidently often required at positions where bulkier residues would create steric problems,[16] it seems plausible that the bulky methyl groups on the long side chain of lysine could create a special site of interaction or could produce a steric disturbance in the structure.

N^{ϵ}-Acetyllysine, which had never been found previously in proteins, was discovered in histone IV by ourselves[11] and by Gershey, *et al.*[17] In the sequence studies only one, residue 16, of the 11 lysines was found to be acetylated in calf thymus histone IV, and in only about half of the molecules.[11] Whereas the N^{α}-acetyl group of histone IV is metabolically stable, the N^{ϵ}-acetyl group turns over at an appreciable rate.[17,18] A number of studies have implicated acetylation and deacetylation of histones with processes of genetic regulation.[2,18] Although only one lysyl residue was found to be acetylated in the sequence studies, Wangh, Ruiz-Carrillo, and Allfrey[19] have recently reported the presence of di- and triacetylated molecules, both being present in very small amounts. It is probable that these correspond to similar fractions of trout testis histone IV where as many as four lysyl residues near the NH_2 terminus may be acetylated.[6] In the helical wheel arrangement[15] the four lysyl residues that may be acetylated form a highly basic "face" which might serve as a primary DNA binding site. It is evident that acetylation of these lysyl residues could significantly alter the affinity of the histone for DNA.

Trout testis histone IV may also be phosphorylated at the NH_2 terminus, yielding the unusual *N*-acetyl-*O*-phosphorylseryl residue. In the helical wheel arrangement the seryl phosphate group is also part of the same "face" formed by the four lysyl residues which may be

(13) K. Murray, *Biochemistry*, **3**, 10 (1964).
(14) W. K. Paik, and S. Kim, *Biochem. Biophys. Res. Commun.*, **27**, 479 (1967).
(15) M. Schiffer and A. B. Edmundson, *Biophys. J.*, **7**, 121 (1967).
(16) T. Y. Shih and J. Bonner, *J. Mol. Biol.* **50**, 333 (1970).
(17) E. L. Gershey, G. Vidali, and V. G. Allfrey, *J. Biol. Chem.*, **243**, 5018 (1968).
(18) V. G. Allfrey in "Histones and Nucleohistones," D. M. P. Phillips, Ed., Plenum Press, New York, N. Y., 1971, Chapter 6.
(19) L. Wangh, A. Ruiz-Carrillo, and V. G. Allfrey, in press.

Table II

Distribution of Residues in Histones[a]

	Histone IV (102 residues)		Histone III (135 residues)		Histone IIb2 (125 residues)		Histone IIb1 (131 residues)		Histone I (~212 residues)		
	1–45	46–102	1–53	54–135	1–50	51–125	1–36	37–131	1–41	42–107	108–end
Charged residues[b]											
Basic	17	10	19	15	19	13	13	17	10	11	43
Acidic	1	7	1	11	2	8	0	10	3	5	3
Net charge	+16	+3	+18	+4	+17	+5	+13	+7	+7	+6	+40
Longest basic cluster	5 (16–20)[c]	3 (77–79)	2	2	5 (27–31)	2	2	3 (123–125)	4 (19–23)	2	?
Longest nonbasic region	7 (24–30)	11 (80–90)	8 (28–35)	29 (84–112)	8 (35–42)	14 (58–71)	8 (21–28)	28 (43–70)	14 (1–14)	10 (67–76)	6–7
Hydrophobics[d]											
Aliphatic	8	16	5	22	6	17	5	25	0	16	4
Aromatic	0	6	1	6	3	4	1	3	0	2	0
Hydroxyl residues[e]	2	11	9	9	9	18	4	8	3	13	6
Prolyl residues	1	0	4	2	5	1	1	4	7	0	16

[a] See Figures 1–5. [b] Includes all α-amino, α-carboxyl, aspartyl, glutamyl, histidyl, arginyl, and lysyl (unsubstituted form) residues. [c] Numbers in parentheses refer to residue numbers. [d] Includes valyl, methionyl, leucyl, isoleucyl, tyrosyl, and phenylalanyl residues only. [e] Includes seryl, threonyl, and tyrosyl residues only.

acetylated (see above) and might be pictured as contributing to the neutralization of this basic region and subsequent changes in DNA–histone interactions. Phosphorylation of histones may be correlated with RNA synthesis,[2,18] but if histone IV is phosphorylated (other than in trout testis) the level is so low that it has not been conclusively detected to date. If only a few genes are derepressed, the total histone IV pool would be expected to contain only a small amount of the phosphorylated derivative, assuming a correlation between these processes.

Comparison of the sequences of pea seedling[3] and calf thymus[11] histone IV reveals a remarkable conservation of sequence throughout the long period since the evolutionary divergence of these organisms from their common ancestral forms. As shown in Figure 1, there are only two differences in sequence. Such conservation of sequence indicates that the function of histone IV is so dependent on the entire structure of the molecule that little change could be tolerated. The same histone from two additional mammals, rat[20] and pig,[21] is identical in sequence to calf thymus histone IV. Pea seedling histone IV is not methylated,[3] which indicates that different methylating enzyme are involved in the methylation of histones III and IV, since pea seedling histone III is methylated (both histones are methylated in calf thymus). Pea seedling histone IV is acetylated at two sites, lysine-16 being the major one as in calf thymus histone IV and the other site probably being lysine-8.[3]

In addition to the above derivatives, histone IV may contain methylated forms of arginine since Paik and Kim[22] have found this derivative in total histones.

Histone III. The complete sequence of calf thymus histone III (Figure 2), recently established in our laboratory,[23] indicates certain features in common with histone IV. Both are arginine rich, with the NH_2-terminal regions being highly basic and containing the N^{ϵ}-acetyl and N^{ϵ}-methyl derivatives of lysine, whereas the COOH-terminal regions are only slightly basic and contain most of the hydrophobic residues (Table II).

Histone III differs from histone IV in having more residues (135 *vs.* 102), a free α-amino group, cysteine (residues 96 and 110), no sequential clusters of more than two basic residues, two sites of N^{ϵ}-acetylation (residues 14 and 23), two sites of N^{ϵ}-methylation (residues 9 and 27), and an acidic region of 29 residues (84–112) lacking a basic residue (Table II). Although the sequences of the two histones are generally different, there are a few similar regions.[23]

Another difference between histones III and IV is the presence of N^{ϵ}-trimethyllysine in the former. This betaine derivative of lysine was first discovered in the

(20) P. Sautiere, D. Tyrou, Y. Moschetto, and G. Biserte, *Biochimie*, **53**, 479 (1971).
(21) P. Sautiere, M.-D. Lambelin-Breynaert, Y. Moschetto, and G. Biserte, *ibid.*, **53**, 711 (1971).
(22) W. K. Paik and S. Kim, *Arch. Biochem. Biophys.*, **134**, 632 (1969).
(23) R. J. DeLange, J. A. Hooper, and E. L. Smith, *Proc. Nat. Acad. Sci. U. S.*, **69**, 882 (1972).

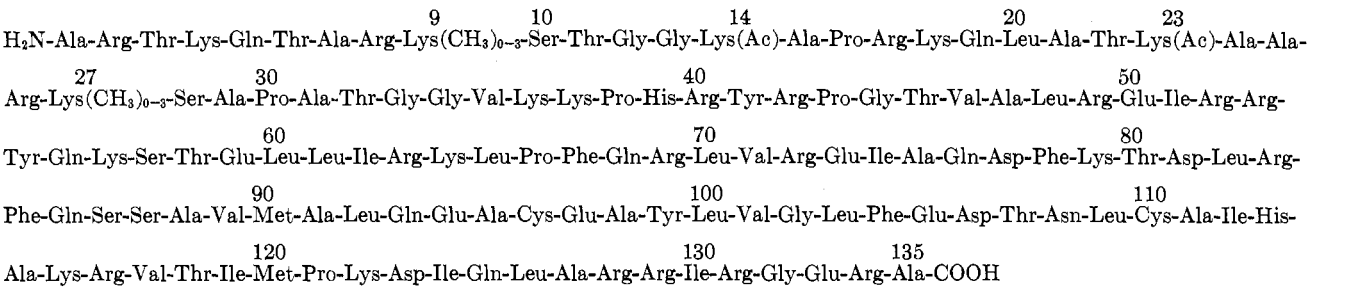

Figure 2. Amino acid sequence of calf thymus histone III (f3). See ref 23.

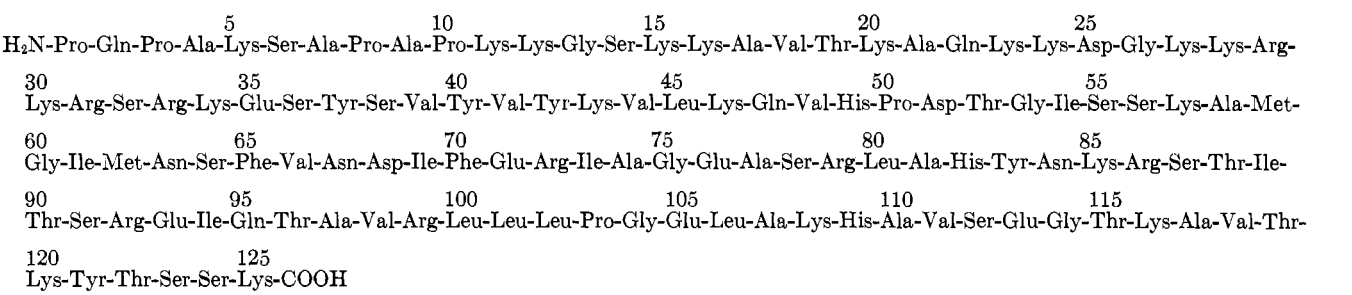

Figure 3. Amino acid sequence of calf thymus histone IIb2 (f2b). Lysyl residue 19 in ref 33 has been transposed to position 16 as shown (personal communication from K. Iwai).

cytochromes *c* of ascomycetes and plants[24,25] and in unfractionated histones,[26] but it is absent from histone IV. Analysis showed that all three methyl derivatives are present at both sites of methylation.[27]

In histones III and IV the methylated lysyl residue is in each instance on the COOH-terminal side of an arginyl residue (Arg-Lys$(CH_3)_{0-3}$). However, since other Arg-Lys sequences in these histones are not methylated, additional features must account for the specificity of methylation. For histone III, both sites of methylation are in the sequence Ala-Arg-Lys$(CH_3)_{0-3}$-Ser, the only sequences of this type in the protein. Since the methylase for histone III is probably distinctive from the one for histone IV (see above), this sequence, Ala-Arg-Lys-Ser, might represent the recognition site for the histone III methylase.

The three sites of acetylation in calf thymus histones III and IV show the common sequence, Lys(Ac)-X-Y-Arg-Lys. In two of the three sites the lysyl residue in the Arg-Lys sequence is also methylated. However, other lysyl residues, present in sites which do not conform to the above sequence, are also acetylated in calf,[19] pea,[3] and trout testis.[6] Therefore, either this sequence is fortuitous in these three instances, or there are other acetylases which recognize different sequences.

Since histone III is unique in containing cysteine, it might be conjectured that any unusual function of this histone might be correlated with the presence of this amino acid. Indeed, it has been reported that the ratio of thiol to thiol + disulfide alters in parallel with changes in the structure and genetic activity of chromatin.[2,28] In general the disulfide content increases as the chromatin becomes more condensed and inactive. Thus, processes such as mitosis, transcription, and replication might all include prominent changes in the cysteinyl–cystinyl content of histone III and other chromosomal proteins. It is of interest that histone III of rabbit and higher animals contains two cysteinyl residues, whereas those of rodents, lower animals, and plants contain only one cysteinyl residue.[29] This would allow polymer formation from histone III of the higher animals but only dimer formation in other species.

Olson, *et al.*,[30] have determined the sequence of the first 26 residues of histone III with the sequenator. Their data agree with ours except that residue 19 is glutamic acid in their sequence, glutamine in ours. They did not determine the positions of the methylated and acetylated lysyl residues. The sequences of most of the tryptic peptides from this region of the molecule had previously been reported by our laboratory.[31] Yokotsuka, Kikuchi, and Shimura[32] have also reported sequence studies on this histone.

Histone IIb2. This was the first of the lysine-rich histones to be sequenced[33] (Figure 3). Another version

(24) R. J. DeLange, A. N. Glazer, and E. L. Smith, *J. Biol. Chem.*, **244**, 1385 (1969).
(25) R. J. DeLange, A. N. Glazer, and E. L. Smith, *ibid.*, **245**, 3325 (1970).
(26) K. Hempel, H. W. Lange, and L. Birkofer, *Naturwissenschaften*, **55**, 37 (1968).
(27) The first preparation analyzed (see ref 13) showed no N^{ϵ}-trimethyllysine; a more recent preparation did contain this derivative (unpublished results by R. J. DeLange).
(28) A. Sadgopal and J. Bonner, *Biochim. Biophys. Acta*, **207**, 206, 227 (1970).
(29) S. Panyim, K. R. Sommer, and R. Chalkley, *Biochemistry*, **10**, 3911 (1971).
(30) M. O. J. Olson, J. Jordan, and H. Busch, *Biochem. Biophys. Res. Commun.*, **46**, 50 (1972).
(31) R. J. DeLange, E. L. Smith, and J. Bonner, *ibid.*, **40**, 989 (1970).
(32) K. Yokotsuka, A. Kikuchi, and K. Shimura, *J. Biochem. (Tokyo)*, **71**, 133 (1972).
(33) K. Iwai, K. Ishikawa, and H. Hayashi, *Nature (London)*, **226**, 1056 (1970).

```
1                                   10                  15                  20                  25
Ac-Ser-Gly-Arg-Gly-Lys-Gln-Gly-Gly-Lys-Ala-Arg-Ala-Lys-Ala-Lys-Thr-Arg-Ser-Ser-Arg-Ala-Gly-Leu-Gln-Phe-Pro-Val-Gly-Arg-

30                  35                  40                  45                  50                  55
Val-His-Arg-Leu-Leu-Arg-Lys-Gly-Asn-Tyr-Ala-Glu-Arg-Val-Gly-Ala-Gly-Ala-Pro-Val-Tyr-Leu-Ala-Ala-Val-Leu-Glu-Tyr-Leu-

60                  65                  70                  75                  80                  85
Thr-Ala-Glu-Ile-Leu-Glu-Leu-Ala-Gly-Asn-Ala-Ala-Arg-Asp-Asn-Lys-Lys-Thr-Arg-Ile-Ile-Pro-Arg-His-Leu-Gln-Leu-Ala-Ile-Arg-

90                  95                  100                 105                 110                 115
Asn-Asp-Glu-Glu-Leu-Asn-Lys-Leu-Leu-Gly-Lys-Val-Thr-Ile-Ala-Gln-Gly-Gly-Val-Leu-Pro-Asn-Ile-Gln-Ala-Val-Leu-Leu-Pro-

120                 125                 129
Lys-Lys-Thr-Glu-Ser-His-His-Lys-Ala-Lys-Gly-Lys-COOH
```

Figure 4. Amino acid sequence of calf thymus histone IIb1 (f2a2). See ref 35.

```
                                                    10                  20
RT-3    Ac-Ser-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Glu-Lys-Ser-Pro-Ala-Lys-Lys-Lys-— -Lys-Ala-Ala-Lys-Lys-
RT-4    Ac-Ser-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-— -Lys-Ser-Pro-Ala-Lys-Thr-Pro-Val-Lys-Ala-Arg-Lys-Lys-
CT-1    Ac-Ser-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Pro-Lys-Ser-Pro-Ala-Lys-Thr-Pro-Val-Lys-Ala-Ala-Lys-Lys-

        30                  40                  50
Pro-Gly-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-Ala-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala-Ala-Ser-Lys-Glu-Arg-Asn-
Lys-Ser-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala-Ala-Ser-Lys-Glu-Arg-Ser-
Lys-Lys-Pro-Ala-Gly-Ala-Arg-Arg-Lys-Ala-Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala-Ala-Ser-Lys-Glu-Arg-Ser-

        60                  70                  80
Gly-Leu-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Gly-Gly-Tyr-Asp-Val-Glu-Lys-Asn-Asn-Ser-Arg-Ile-Lys-Leu-Gly-Leu-Lys
Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr
Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr

        90                  100
Ser-Leu-Val-Ser-Lys-Gly-Thr-Leu-Val-Glu-Thr-Lys-Gly-Thr-Gly-Ala-Ser-Gly-Ser-Phe-Lys(Lys42,Asp,Thr3,Ser3,Glu,Pro16,Gly4-5,

Ala34,Val3,Leu)
```

Figure 5. Amino acid sequences of histones I (fI) from rabbit thymus (RT) and calf thymus (CT). See ref 10. The numbering system includes proposed deletions in the sequences.

of the sequence, which is not complete and varies slightly from that shown in Figure 3, has also been reported.[34]

The major similarity between histone IIb2 and histones III and IV is in the distribution of residues. Once again, the NH_2-terminal region is highly basic, whereas the COOH-terminal region is less basic and contains most of the hydrophobic residues (Table II). Histone IIb2 has a sequence of five consecutive basic residues (residues 27–31), a feature in common with histone IV. However, N^{ϵ}-acetyllysine and N^{ϵ}-methyllysine were not reported in this histone.

Histone IIb1. Yeoman, *et al.*,[35] recently completed the sequence of this histone (Figure 4). This sequence contained a few points of discrepancy with the partial sequence reported by Sautière, *et al.*,[36] but the latter group now agrees that the sequence is as shown in Figure 4.[36]

From the sequence it is apparent that the first nine residues of histone IIb1 correspond to those of histone IV, if it is assumed that there has been a deletion at residue 6 in histone IV (or an insertion in histone IIb1). Similar side-chain modification reactions, particularly in the two trout testis histones,[6] have been noted in these regions of the molecules.

The NH_2-terminal region of histone IIb1 is again rather basic (Table II), and there are two large hydrophobic regions in the COOH-terminal region which lack basic residues: residues 43–70 (−3 charge) and residues 100–117 (no charged residues). The longest sequence of basic residues is near the COOH terminus (His-His-Lys) and there is a sequence of three acidic residues in a row (residues 90–92).

Histone I. This is the longest of the histones (Table I). It shows heterogeneity not only in the sequences of the subfractions isolated from the same organ (Figure 5) but also when different species are compared. Although histone I has not been reported to be methylated or N^{ϵ}-acetylated (even in trout testis), it is phosphorylated (see below).

The amino acid sequence work on histone I, performed mainly by Cole and his collaborators,[10] has been complicated by the necessity of first purifying the individual subfractions, but sequence comparisons of these subfractions are of considerable interest. As shown in Figure 5, the sequence of 107 residues (the numbering system includes deletions) from the NH_2 terminus has been established for subfraction 3 of rabbit thymus. Two other subfractions, one from rabbit thymus and one from calf thymus, have been sequenced to residue 73. In addition, the sequences of 22 tryptic peptides from the COOH-terminal half of the rabbit thymus

(34) L. S. Hnilica, H. A. Kappler, and J. J. Jordan, *Experientia*, **26**, 353 (1970).
(35) L. C. Yeoman, M. O. J. Olson, N. Sugano, J. J. Jordan, C. W. Taylor, W. C. Starbuck, and H. Busch, *J. Biol. Chem.*, in press.
(36) P. Sautière, D. Tyrou, B. Laine, J. Mizon, M.-D. Lambelin-Breynaert, P. Ruffin, and G. Biserte, *C. R. Acad. Sci.*, **274**, 1422 (1972) (personal communication from P. Sautière).

subfraction 3 are nearly complete, but ordering of these peptides is complicated by the great preponderance of alanyl, lysyl, and prolyl residues (approximately 93 of the 109 total residues). Arginine, isoleucine, and aromatic amino acids are absent from the COOH-terminal half of the molecule. This region of histone I is similar to the protamines except for its larger size and the presence of lysine in place of arginine. It is of interest that Phelan, *et al.*,[37] have isolated a lysine-rich protamine from ripe gonads of mussels.

Gottesfeld, *et al.*,[38] have found that the subfractions of histone I confer differing stabilities to DNA as measured by resistance to DNase. Therefore, differences in the structural roles of the subfractions are indicated. In addition residue 38 (serine), which is phosphorylated in some subfractions, is replaced by an alanyl residue in rabbit thymus subfraction 3. Since phosphorylation of histone I appears to be correlated with genetic regulation,[2,18] this change of alanine for serine suggests differences in the functions of the subfractions as well.

Histone V. This histone has been found only in nucleated erythrocytes.[7] The partial sequences of this histone reported by Greenaway and Murray[39] and Greenaway[40] show several interesting characteristics. (a) Like histone I, subfractions of histone V appear to vary in sequence.[39] (b) There is clustering of basic residues, possibly as many as four in a row.[40] (c) A peptide containing phosphorylserine in the sequence Ser(P)-Asp-Lys was isolated, and the same peptide was isolated in nonphosphorylated form.[40]

It is thought that histone V is synthesized in place of a portion of the histone I fraction in nucleated erythrocytes, since the two fractions together equal the usual amount of the whole histone complement represented by histone I in other tissues. Since the genetic material in nucleated erythrocytes is completely repressed, it is thought that the function of histone V is to combine with DNA to prevent genetic expression. If this is so, the finding of phosphorylserine suggests that, at least in histone V, phosphorylation may not lead to derepression of the genetic material or dissociation of histones as suggested for other systems.

Other than histone V, few examples of special histones have been reported; however, Wigle and Dixon[8] have described a new histone in trout tissues and Phelan, Subirana, and Cole[9] have recently described a new fraction containing lysine-rich histones in gonads of the male *Holothuria tubulosa*.

(37) J. Phelan, J. Colom, J. Subirana, and R. D. Cole, personal communication.

(38) J. M. Gottesfeld, M. Calvin, R. D. Cole, D. M. Igdaloff, V. Moses, and W. Vaughn, *Biochemistry*, **11**, 1422 (1972).

(39) P. J. Greenaway and K. Murray, *Nature, New Biol.*, **229**, 233 (1971).

Summary

Present evidence indicates that in calf thymus and most other eukaryotic tissues there are five major histone fractions, all of which can be modified after synthesis to give rise to subfractions. In the case of the very-lysine-rich histone I, subfractions are also the result of heterogeneity in amino acid sequences. To what extent this sequence heterogeneity and the side-chain modification reactions are able to influence the functions of histones is not clear, but a considerable number of molecular species are obviously present and may be required to provide the necessary fine adjustments during changes in chromatin structure and function. On occasion a special requirement leads to the production of a new type of histone.

Since all somatic cells in an organism have the same complement of DNA, each cell would be like every other cell if either all or the same part of the DNA in all cells were genetically expressed. Obviously, this is not the case, *i.e.*, a liver cell is different than a nerve cell, a muscle cell, etc., and some mechanism for differential expression of genes (DNA) in order to permit differentiation of cells must exist both in the developing organism as well as in the adult form. Histones appear to be a part of this mechanism by functioning as repressors of all genes except those which must be active in order to make and maintain that particular type of cell.[2]

At first it was thought that the histones themselves would be specific in determining which genes are expressed and which ones are "turned off" in the cell. However, the limited kinds of histones, the failure to demonstrate specificity in the histone fraction, and the finding that components in the so-called "acidic protein fraction" of nuclei do supply specificity to the genetic expression of each cell have all led to the view that histones are rather nonspecific repressors.[2] This is not to imply that they do not bind to DNA and other components of chromatin in a highly specific manner.

Besides acting as nonspecific repressors, histones are important in the "packaging" of the genetic material into a compact state by neutralizing the negative charges of the DNA and, probably, in the regulation of new DNA synthesis.[2] Whatever the precise roles of histones may turn out to be, the fact that these have changed relatively little during the entire course of the evolution of higher plants and animals, as evidenced by their remarkable conservation of sequence, implies a high degree of importance in nuclear function.

(40) P. J. Greenaway, *Biochem. J.*, **124**, 319 (1971).