Review Letter

MORE HISTONE STRUCTURES

Claus VON HOLT, Walter N. STRICKLAND, Wolf F. BRANDT and Marie S. STRICKLAND
Department of Biochemistry, Chromatin Research Unit, Council for Scientific and Industrial Research, University of Cape Town, Private Bag Rondebosch, 7700 South Africa

Received 3 January 1979

1. The primary structure of histones

Since the elucidation of the amino acid sequence of the prototype histones H3, H4, H2A and H2B, a considerable amount of information on the occurrence of variants of these structures has become available. (Bibliographic references to details of the histone primary structures are given in fig.1–3; for review see [34].)

It appears that the histones H3 and H4, over a wide range of evolution, vary only slightly whereas the histones H2A and H2B undergo more extensive evolutionary changes (fig.1–3).

The variations of the structures can be classified into two types. The first group comprises simple point mutations or deletions of one or several residues. The second type of structural variations consists in extensive modifications of major areas of the molecules through reiteration, insertion, deletion and point mutations. Such changes have led, in some variants, to the establishment of new domains in the structure.

Whereas the former type of structure variation occurs in all four histones, the second group of more extensive variations up to now has been identified only in the histones of the H2A and H2B type. The first group of alterations does not change the general structure of the molecules drastically. Such changes occur infrequently in the histones H3 and H4. In histone H3, four positions are variable. In the histone H4 only three variable positions have been identified. In both these histones there appears to be no regional preference for the conservative point mutations.

The mutations involving cysteine in position 96 of histone H3 and 73 of histone H4 may assign to the molecules new properties. The potential cysteine–cystine interconversion between neighbouring histones may affect the reversibility of the pairing of histones H3 and H4 in chromatin subunits. In fact, it appears that the two histone H3 chains in the nucleosomes approach each other very closely in the corresponding regions of the molecules (see section 3).

In addition to major changes in the structure (see below) some largely conservative point mutations also occur in the histones H2B (fig.2) but more frequently than in the histones H3 and H4. A similar situation pertains to histone H2A (fig.1).

The occasional occurrence of these limited changes in the molecules of the four histones in organisms separated widely from each other on the evolutionary scale, does not alter substantially the generally accepted view on the extremely conservative nature of the structures of these molecules throughout evolution (for review see [34]). Certainly this notion appears justified for the histones H3 and H4 for which complete or partial structures representing the situation in various divisions of the animal and plant kingdom are available (fig.1).

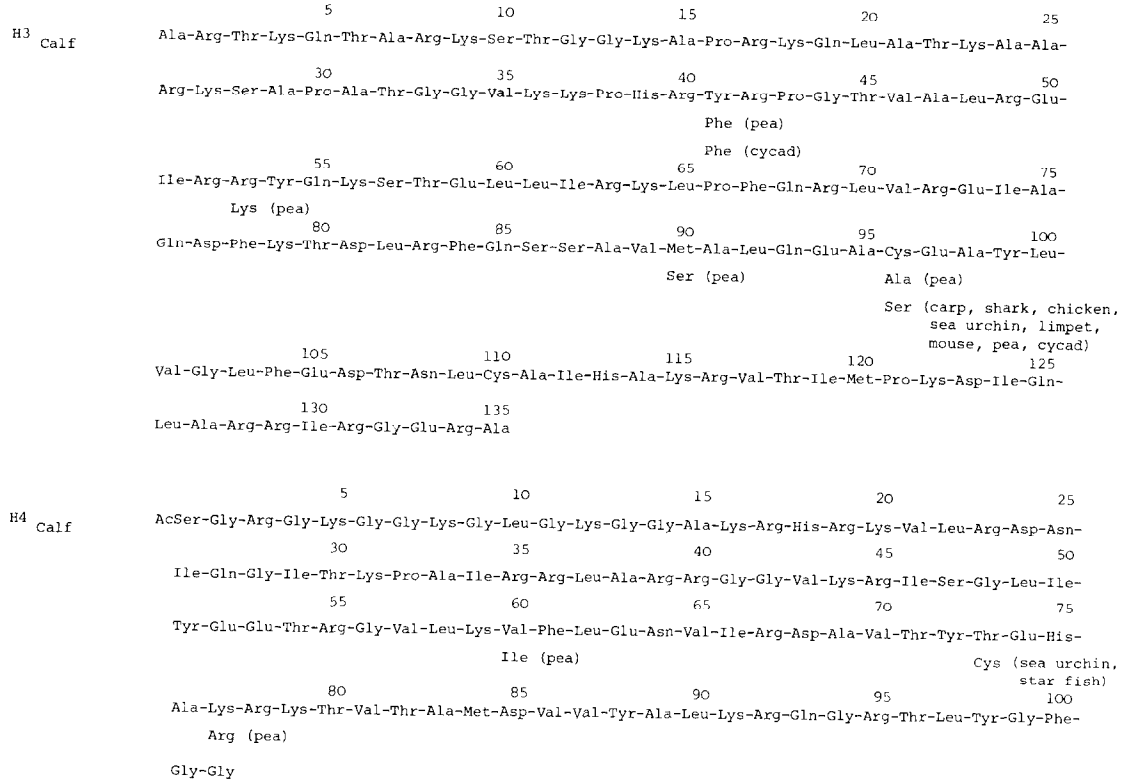
For the histones H2A and H2B, however, such a generalization may be misleading. In these two histones the second group of structural variations characterized by multiple insertions, deletions of several residues together with point mutations has resulted in substantial changes of structures (fig.1,2).

In the histone H2B group, taking into account the extensive structural variations, the following picture of the general structure of these molecules has emerged. The sequence of the 89 C-terminal amino acids consists mainly of hydrophobic residues

and has been conserved during evolution in its general features, being common to plant histones and animal histones alike. Differences between the various members of this histone group are characterized by mainly conservative point mutations. In the 17 variants for which partial or complete sequences in this region are available, a total of 35 such mutations has been found. Towards the N-terminal, this region is preceded by a highly basic area of 9 residues out of which in animal histones 7 or 8 are either arginine or lysine. From here to the N-terminus extends a region, differing from variant to variant, of ~20–50 highly variable residues. In this region the degree of variability increases with the distance from the block of basic amino acids (fig.2).

The variability of the N-terminal part of the histones H2B may be one of the expressions of increasing chromatin complexity at different levels of evolution (phylogenetic histone programme) and/or an expression of differential genetic activity of cells/tissue during the life cycle of one organism at any particular evolutionary level (ontogenetic histone programme).

The comparison of the structures of the histones H2B lends support to both these views. A corollary of the possible existence of an ontogenic histone programme is the existence of tissue specific histones. Therefore comparison of histone structures, e.g., to determine evolutionary rates should strictly be made only of histones from comparable tissues. The histones isolated from the same tissue (testes) of a limpet (mollusc) and of trout (early vertebrate) differ considerably from each other in the N-terminal region, possibly as an expression of the long evolutionary distance between the two organisms (fig.2). In the same vein, the histones isolated from the terminally differentiated nucleated erythrocytes of evolutionary closely related classes (aves, amphibia and reptiles) differ only very little from each other (three sites of point mutations) and from the histone H2B isolated from the also terminally differentiated thymus tissue of another vertebrate (calf) (fig.2). In this way the



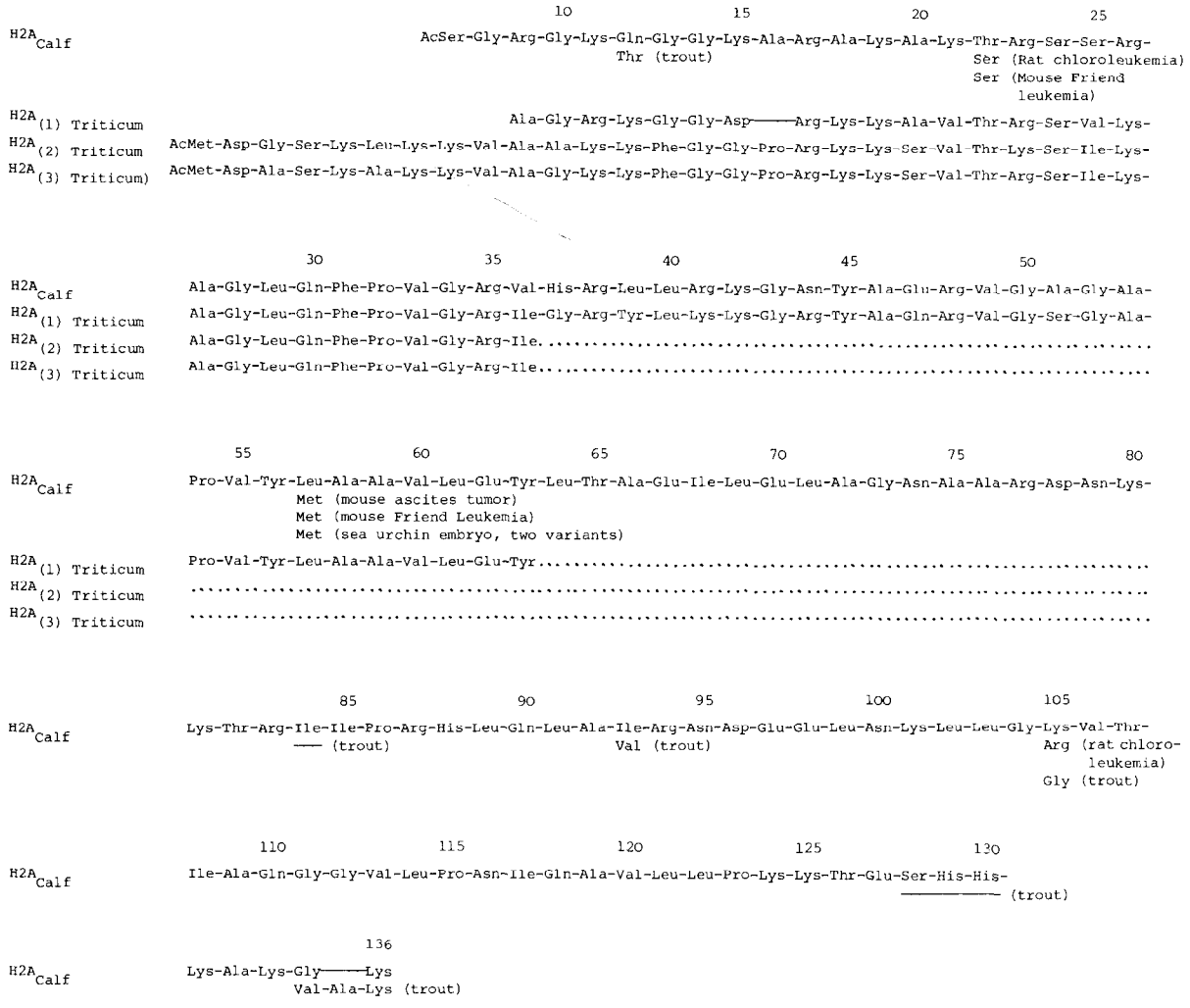
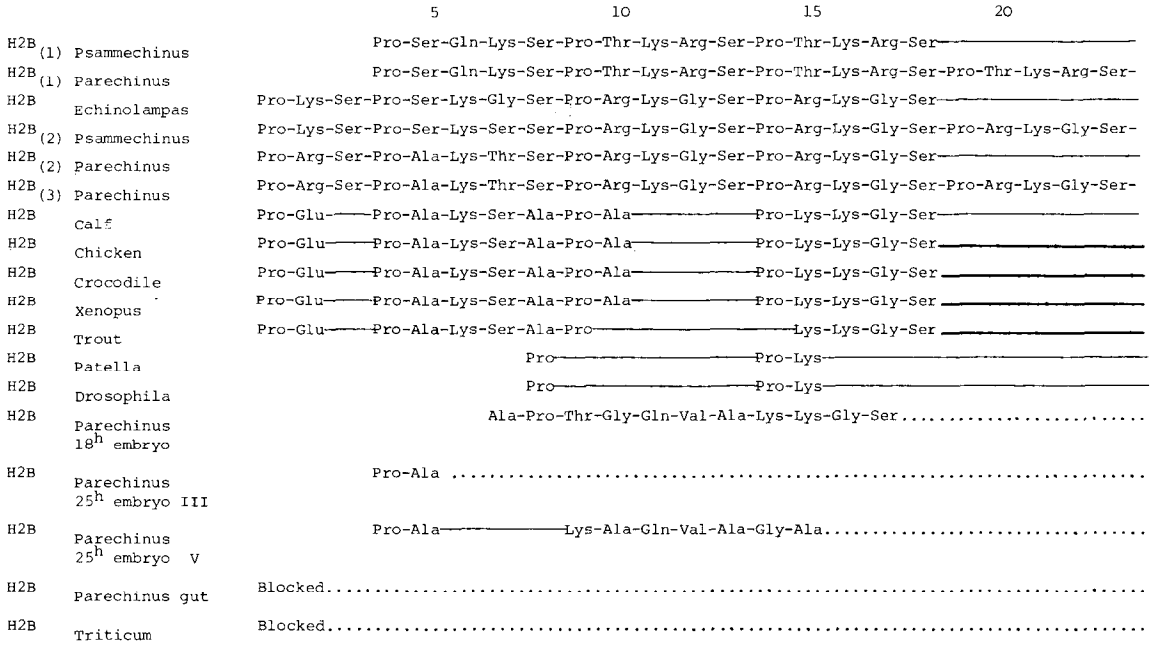
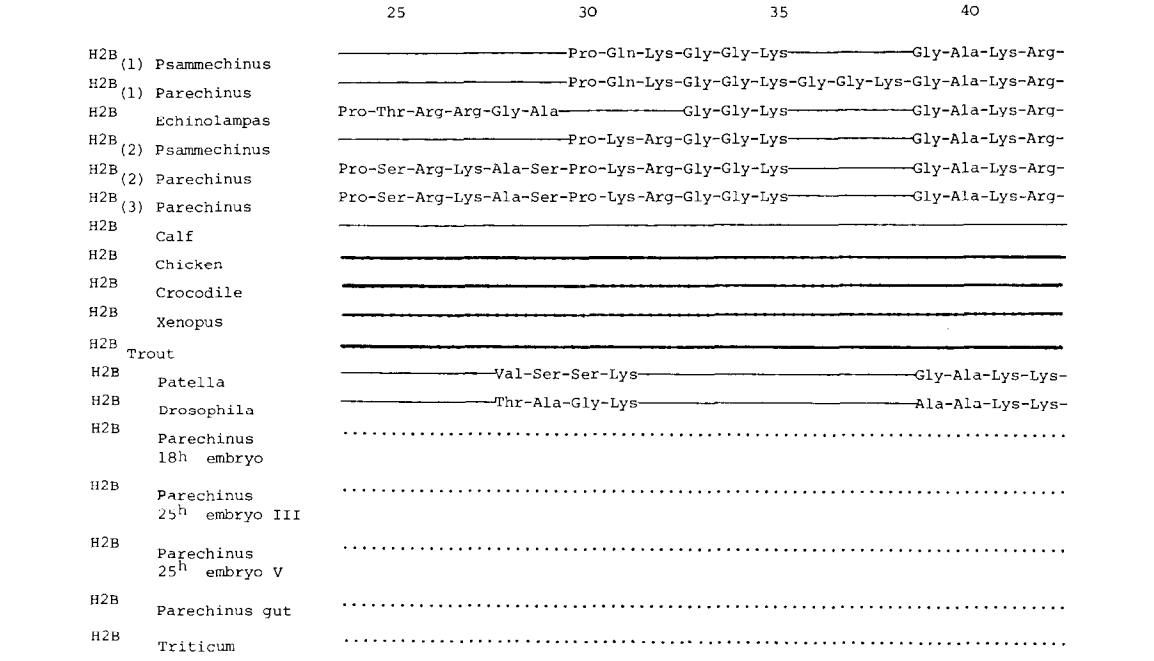


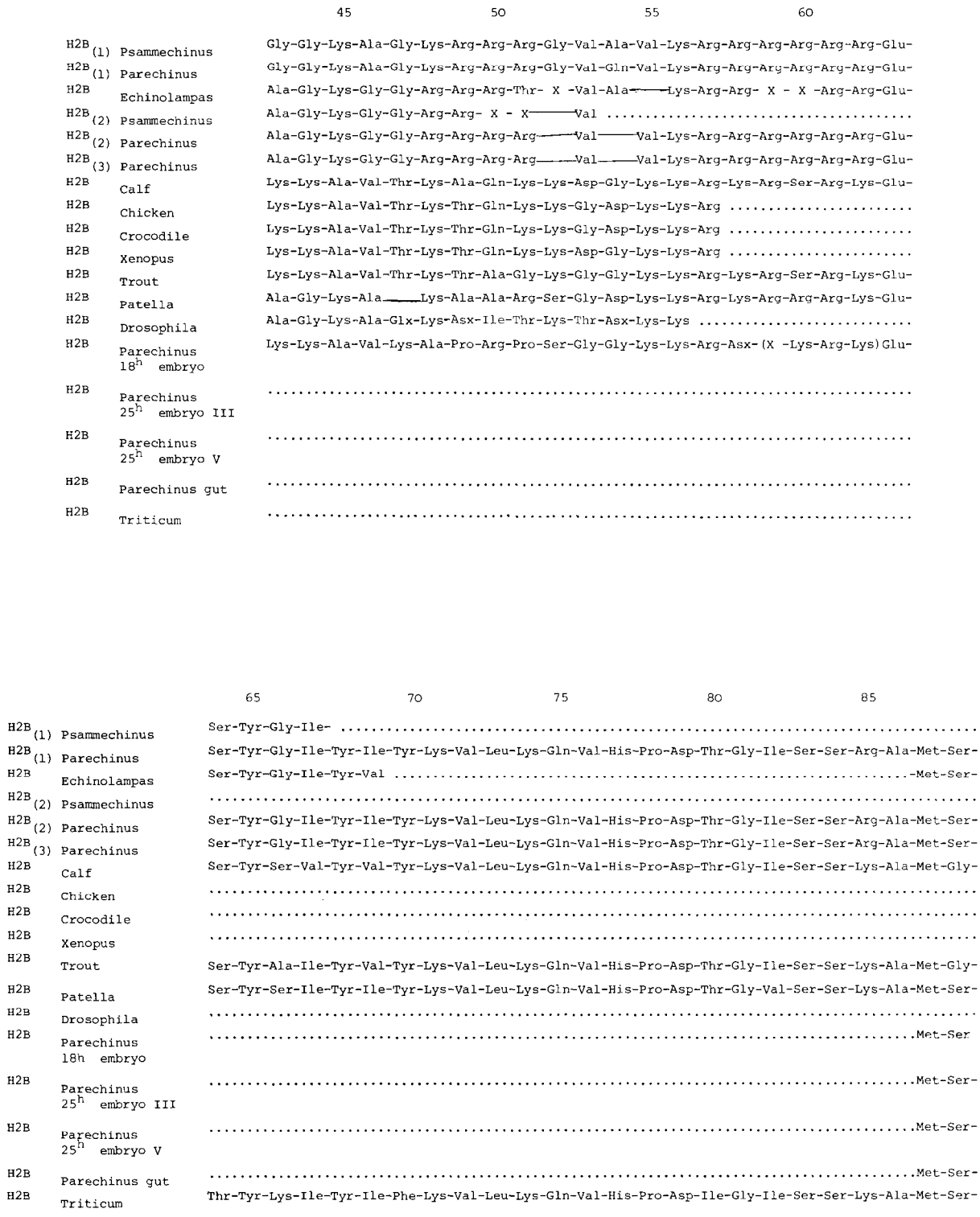
Fig.1. Sequences of histone H3. Mammal: *Bos taurus* (calf) complete sequence [1]; *Mus musculus* (mouse) peptide mapping [2]. Bird: *Gallus domesticus* (chicken) complete sequence [3,4]. Fish: *Letiobus bubalus* (carp) complete sequence [5]; *Salmo trutto* (trout) partial sequence [6]; *Poroderma africanus* (shark) complete sequence [8]. Mollusc: *Patella granatina* (limpet) partial sequence [7]. Fungus: *Saccharomyces cerevisiae* (baker's yeast) partial sequence [9]; *Neurospora crassa* three N-terminal residues [10]. Plant: *Pisum sativum* (pea) complete sequence [11]; *Encephalartos caffer* (cycad) partial sequence [7].

Sequences of histone H4. Mammal: *Bos taurus*, calf thymus complete sequence [12–14], lymphosarcoma peptide mapping [15], fetal calf thymus peptide mapping [15]; *Rattus norwegicus* (laboratory rat), Novikoff hepatoma peptide mapping [15]. Sea urchin: *Parechinus angulosus* partial sequence [16]; *Enchinolampas crassa* amino acid composition [16]; *Psammechinus miliaris* peptide mapping [17]. Starfish: *Marthasterias glacialis* amino acid composition [16]. Plant: *Pisum sativum* complete sequence [18].

Sequences of histones H2A, aligned for homologies. Numbering refers to alignment positions, and not to sequence positions. Deletions in the alignment (——); comparable regions not sequenced (· · ·); the total extent of the unsequenced positions cannot be indicated at this stage. Mammal: *Bos taurus* (calf) complete sequence [19]; *Mus musculus* (mouse), ascites tumor cells peptide mapping [2], Friend leukemia cells peptide mapping [20]; *Rattus norwegicus* (rat) chloroleukemia cells complete sequence [21]. Fish: *Salmo irideus* (rainbow trout) complete sequence [22]. Sea urchin: *Parechinus angulosus* partial sequence [23]. Wheat germ: *Triticum aestivum*, three variants partial sequence [48].


5
10
15
20
H2B (1) Psammechinus
Pro-Ser-Gln-Lys-Ser-Pro-Thr-Lys-Arg-Ser-Pro-Thr-Lys-Arg-Ser
H2B (1) Parechinus
Pro-Ser-Gln-Lys-Ser-Pro-Thr-Lys-Arg-Ser-Pro-Thr-Lys-Arg-Ser-Pro-Thr-Lys-Arg-Ser-
H2B Echinolampas
Pro-Lys-Ser-Pro-Ser-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser
H2B (2) Psammechinus
Pro-Lys-Ser-Pro-Ser-Lys-Ser-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser-
H2B (2) Parechinus
Pro-Arg-Ser-Pro-Ala-Lys-Thr-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser
H2B (3) Parechinus
Pro-Arg-Ser-Pro-Ala-Lys-Thr-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser-Pro-Arg-Lys-Gly-Ser-
H2B Calf
Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Ala-Pro-Lys-Lys-Gly-Ser
H2B Chicken
Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Ala-Pro-Lys-Lys-Gly-Ser
H2B Crocodile
Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Ala-Pro-Lys-Lys-Gly-Ser
H2B Xenopus
Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Ala-Pro-Lys-Lys-Gly-Ser
H2B Trout
Pro-Glu-Pro-Ala-Lys-Ser-Ala-Pro-Lys-Lys-Gly-Ser
H2B Patella
Pro-Pro-Lys
H2B Drosophila
Pro-Pro-Lys
H2B Parechinus 18h embryo
Ala-Pro-Thr-Gly-Gln-Val-Ala-Lys-Lys-Gly-Ser
H2B Parechinus 25h embryo III
Pro-Ala
H2B Parechinus 25h embryo V
Pro-Ala-Lys-Ala-Gln-Val-Ala-Gly-Ala
H2B Parechinus gut
Blocked
H2B Triticum
Blocked


25
30
35
40
H2B (1) Psammechinus
Pro-Gln-Lys-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B (1) Parechinus
Pro-Gln-Lys-Gly-Gly-Lys-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B Echinolampas
Pro-Thr-Arg-Arg-Gly-Ala-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B (2) Psammechinus
Pro-Lys-Arg-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B (2) Parechinus
Pro-Ser-Arg-Lys-Ala-Ser-Pro-Lys-Arg-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B (3) Parechinus
Pro-Ser-Arg-Lys-Ala-Ser-Pro-Lys-Arg-Gly-Gly-Lys-Gly-Ala-Lys-Arg-
H2B Calf
H2B Chicken
H2B Crocodile
H2B Xenopus
H2B Trout
H2B Patella
Val-Ser-Ser-Lys-Gly-Ala-Lys-Lys-
H2B Drosophila
Thr-Ala-Gly-Lys-Ala-Ala-Lys-Lys-
H2B Parechinus 18h embryo
H2B Parechinus 25h embryo III
H2B Parechinus 25h embryo V
H2B Parechinus gut
H2B Triticum


45 50 55 60
H2B (1) Psammechinus Gly-Gly-Lys-Ala-Gly-Lys-Arg-Arg-Arg-Gly-Val-Ala-Val-Lys-Arg-Arg-Arg-Arg-Arg-Arg-Glu-
H2B (1) Parechinus Gly-Gly-Lys-Ala-Gly-Lys-Arg-Arg-Arg-Gly-Val-Gln-Val-Lys-Arg-Arg-Arg-Arg-Arg-Arg-Glu-
H2B Echinolampas Ala-Gly-Lys-Gly-Gly-Arg-Arg-Arg-Thr- X -Val-Ala———Lys-Arg-Arg- X - X -Arg-Arg-Glu-
H2B (2) Psammechinus Ala-Gly-Lys-Gly-Gly-Arg-Arg- X - X———Val .....
H2B (2) Parechinus Ala-Gly-Lys-Gly-Gly-Arg-Arg-Arg-Arg———Val———Val-Lys-Arg-Arg-Arg-Arg-Arg-Arg-Glu-
H2B (3) Parechinus Ala-Gly-Lys-Gly-Gly-Arg-Arg-Arg-Arg———Val———Val-Lys-Arg-Arg-Arg-Arg-Arg-Arg-Glu-
H2B Calf Lys-Lys-Ala-Val-Thr-Lys-Ala-Gln-Lys-Lys-Asp-Gly-Lys-Lys-Arg-Lys-Arg-Ser-Arg-Lys-Glu-
H2B Chicken Lys-Lys-Ala-Val-Thr-Lys-Thr-Gln-Lys-Lys-Gly-Asp-Lys-Lys-Arg .....
H2B Crocodile Lys-Lys-Ala-Val-Thr-Lys-Thr-Gln-Lys-Lys-Gly-Asp-Lys-Lys-Arg .....
H2B Xenopus Lys-Lys-Ala-Val-Thr-Lys-Thr-Gln-Lys-Lys-Asp-Gly-Lys-Lys-Arg .....
H2B Trout Lys-Lys-Ala-Val-Thr-Lys-Thr-Ala-Gly-Lys-Gly-Gly-Lys-Lys-Arg-Lys-Arg-Ser-Arg-Lys-Glu-
H2B Patella Ala-Gly-Lys-Ala———Lys-Ala-Ala-Arg-Ser-Gly-Asp-Lys-Lys-Arg-Lys-Arg-Arg-Arg-Lys-Glu-
H2B Drosophila Ala-Gly-Lys-Ala-Glx-Lys-Asx-Ile-Thr-Lys-Thr-Asx-Lys-Lys .....
H2B Parechinus 18h embryo Lys-Lys-Ala-Val-Lys-Ala-Pro-Arg-Pro-Ser-Gly-Gly-Lys-Lys-Arg-Asx-(X -Lys-Arg-Lys)Glu-
H2B Parechinus 25h embryo III .....
H2B Parechinus 25h embryo V .....
H2B Parechinus gut .....
H2B Triticum .....
65 70 75 80 85
H2B (1) Psammechinus Ser-Tyr-Gly-Ile- .....
H2B (1) Parechinus Ser-Tyr-Gly-Ile-Tyr-Ile-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Ile-Ser-Ser-Arg-Ala-Met-Ser-
H2B Echinolampas Ser-Tyr-Gly-Ile-Tyr-Val .....-Met-Ser-
H2B (2) Psammechinus .....
H2B (2) Parechinus Ser-Tyr-Gly-Ile-Tyr-Ile-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Ile-Ser-Ser-Arg-Ala-Met-Ser-
H2B (3) Parechinus Ser-Tyr-Gly-Ile-Tyr-Ile-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Ile-Ser-Ser-Arg-Ala-Met-Ser-
H2B Calf Ser-Tyr-Ser-Val-Tyr-Val-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Ile-Ser-Ser-Lys-Ala-Met-Gly-
H2B Chicken .....
H2B Crocodile .....
H2B Xenopus .....
H2B Trout Ser-Tyr-Ala-Ile-Tyr-Val-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Ile-Ser-Ser-Lys-Ala-Met-Gly-
H2B Patella Ser-Tyr-Ser-Ile-Tyr-Ile-Tyr-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Thr-Gly-Val-Ser-Ser-Lys-Ala-Met-Ser-
H2B Drosophila .....
H2B Parechinus 18h embryo .....Met-Ser
H2B Parechinus 25h embryo III .....Met-Ser-
H2B Parechinus 25h embryo V .....Met-Ser-
H2B Parechinus gut .....Met-Ser-
H2B Triticum Thr-Tyr-Lys-Ile-Tyr-Ile-Phe-Lys-Val-Leu-Lys-Gln-Val-His-Pro-Asp-Ile-Gly-Ile-Ser-Ser-Lys-Ala-Met-Ser-

	90	95	100	105	110
H2B(1) Psammechinus	.....Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu- X - Ile-Ala-.....				
H2B(1) Parechinus	Val-Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Ala-Glu-Ala-Gly-Arg-Leu-Thr-Thr-Tyr-Asn-Arg-				
H2B Echinolampas	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Ala-Glu-Ala-Ser-Arg-Leu-Ala/Ser-His-Ala/Tyr-Asn-Arg-				
H2B(2) Psammechinus	.....Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Thr-Ser-Ala-Asn-Arg-				
H2B(2) Parechinus	Val-Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Thr-Ser-Ala-Asn-Arg-				
H2B(3) Parechinus	Val-Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Ser-Glu-Ala-Ser-Arg-Leu-Thr-Ser-Ala-Asn-Arg				
H2B Calf	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Ala-His-Tyr-Asn-Lys-				
H2B Chicken	.....Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Ala-His-Tyr-Asn-Lys-				
H2B Crocodile	.....Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Ala-His-Tyr-Asn-Lys-				
H2B Xenopus	.....Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-Ala-His-Tyr-Asn-				
H2B Trout	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ser-Ser-Arg-Leu-Ala-His-Tyr-Asn-Lys-				
H2B Patella	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe-Glu-Arg-Ile-Ala-Ala-Glu-Ala-Ser-Arg-Leu-Ala-His-Tyr-Asn-Lys-				
H2B Drosophila					
H2B Parechinus 18h embryo	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp Ile/Val Phe-Glu-Arg-Ile-Ala-Gly-Glu-Ala-Ser-Arg-Leu-				
H2B Parechinus 25h embryo III	Ile-Met-Asn-Ser-Phe-Val-Asn-Asp-Ile-Phe				
H2B Parechinus 25h embryo V	Val/Ile Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu				
H2B Parechinus gut	Leu-Met-Asn-Ser-Phe-Val-Asn-Asp-Val-Phe-Glu-Arg-Ile-Ala-Ala-Glu-Ala-Ser-Arg-Leu				
H2B Triticum	Ile-Met-Asn-Ser-Phe-Ile-Asn-Asp-Ile-Phe-Glu-Lys-Leu-Ala-Gly-Glu-Ser-Ala-Lys-Leu-Ala-Arg-Tyr-Asn-Lys-				

	115	120	125	130	135
H2B(1) Psammechinus	.....				
H2B(1) Parechinus	Arg-Ser-Thr-Val-Ser-Ser-Arg-Glu-Val-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B Echinolampas	Arg-Ser-Thr-Ile-Ser-Ser-Arg- X -Ile-				
H2B(2) Psammechinus	Arg-Ser-Thr-Ile-Ser-Ser-Arg- X -Ile-				
H2B(2) Parechinus	Arg-Ser-Thr-Val-Ser-Ser-Arg-Glu-Ile-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B(3) Parechinus	Arg-Ser-Thr-Val-Ser-Ser-Arg-Glu-Ile-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B Calf	Arg-Ser-Thr-Ile-Thr-Ser-Arg-Glu-Ile-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B Chicken	Arg-Ser-Thr-Ile-				
H2B Crocodile	Arg-Ser-Thr-Ile-Thr-Ser-Arg-				
H2B Xenopus	.....				
H2B Trout	Arg-Ser-Thr-Ile-Thr-Ser-Arg-Glu-Ile-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B Patella	Arg-Ser-Thr-Ile-Thr-Ser-Arg-Glu-Ile-Gln-Thr-Ala-Val-Arg-Leu-Leu-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				
H2B Parechinus 18h embryo	.....				
H2B Parechinus 25h embryo III	.....				
H2B Parechinus 25h embryo V	.....				
H2B Parechinus gut	.....				
H2B Triticum	Lys-Pro-Thr-Ile-Thr-Ser-Arg-Glu-Ile-Gln-Thr-Ser-Val-Arg-Leu-Val-Leu-Pro-Gly-Glu-Leu-Ala-Lys-His-Ala-				

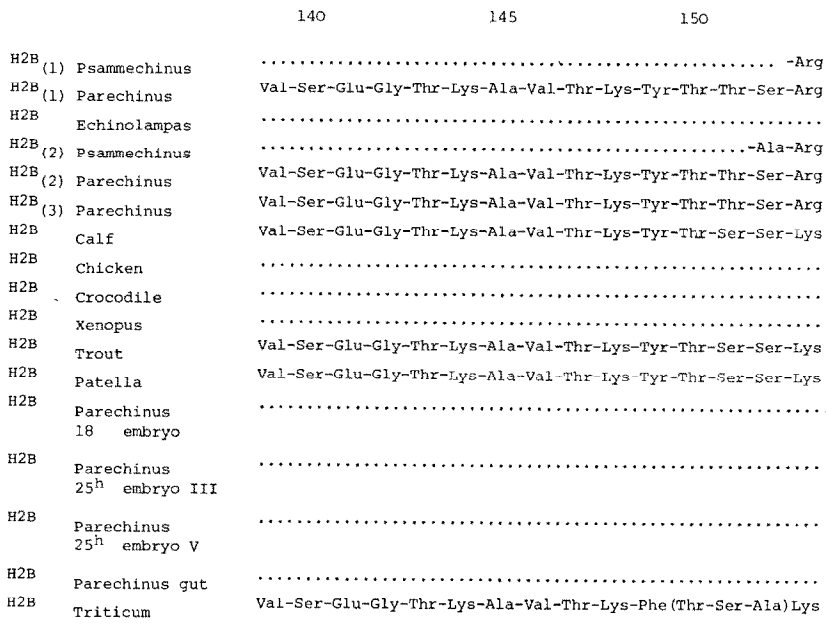


Fig.2. Sequences of histone H2B, aligned for homologies. Numbering refers to alignment positions and not to sequence positions. Deletions in the alignment (——); regions not sequenced (· · ·). Mammal: *Bos taurus* (calf) complete sequence [24]. Bird: *Gallus domesticus* (chicken) partial sequence [25]. Reptile: *Crocodilus niloticus* partial sequence [25]. Amphibian: *Xenopus laevis* partial sequence [25]. Fish: *Salmo trutta* (trout) partial sequence [26]. Sea urchins: *Parechinus angulosus,* sperm, three variants complete sequence [27–29], embryo, two variants partial sequence [23,30], diploid gut cell, one variant partial sequence [23]; *Psammechinus miliaris,* sperm, two variants partial sequence [31]; *Echinolampas crassa,* sperm, two variants partial sequence [32]. Mollusc: *Patella granatina* (limpet) complete sequence [33]. Insect: *Drosophila melanogaster* partial sequence [34]. Plant: *Triticum aestivum* (wheat) partial sequence [35].

variability of the histone H2B structure could be seen as an expression of a phylogenetic histone programme.

At this stage there are not enough comparative data available to draw final conclusions as to the general occurrence of tissue specificity in the histone H2B group. A definite case however, can be made for the histones H2B from terminally differentiated haploid sperm cells of different sea urchin species. In these proteins entirely new structural domains have become established. The latter proteins appear to be typical for the cell type as well as for the species. The sperm cells of the sea urchin *Parechinus angulosus* contain three structural variants of histone H2B, characterized by a repeating pentapeptide structure in the N-terminal part of the molecule (fig.2). Similar histones have now also been found in the sperm of two other sea urchin species, *Psammechinus miliaris* and *Echinolampas crassa* (fig.2). Six different sperm histones H2B have been characterized from only three species investigated (fig.2). The high abundance of different sperm histone H2B variants in sea urchins indicates a very rapid evolution of this particular histone to meet the special needs for survival of that cell type which in very large numbers for a very short time (between spawning and fertilization) is exposed to high selection pressure. Such a scenario is very suitable to select any favourable histone mutation because of its immediate confirmation by successful fertilisation.

The most distinctive feature of these histones H2B consists in the repeating pentapeptides in the N-terminal part of the molecule comprising proline at the N-terminus and serine at the C-terminus. The pentapeptide region leads into a lysine, glycine-rich region in which again repeat structures with a certain amount of variation can be recognized. Gradually,

from alignment position 55 onwards, homologies to calf thymus histone H2B become more evident. The C-terminal two-thirds of the molecules are highly homologous to each other and to the somatic bovine H2B histone with only occasional amino acid substitutions.

The general structure of a pentapeptide with N-terminal proline, a C-terminal hydroxy amino acid and 1–3 central basic amino acids is also present, usually however, not as a repeat structure, in a variety of other proteins found in association with DNA, e.g., trout histone T [41], HMG protein from trout [42], HMG1 [43], HMG 17 [44], histone H1 from rabbit [45], histone H1 from trout [39], histone H2B from calf [24], chicken, crocodile and *Xenopus* [25] and trout [26]. In the protamines this structure occurs repeatedly in four molecules of the clupein group [46] and two of the iridine group [46].

The pentapeptide structure due to its basicity and the potential phosphorylation site of the C-terminal end, may be particularly suitable to provide centers for reversible association between DNA and nuclear proteins. The structural homologies between the pentapeptide region of the sperm histones and the protamines suggest an evolutionary relationship between these two types of sperm proteins. The early presence of histones and the late arrival of protamines on the evolutionary scene make the sea urchin sperm histones a likely ancestor of the protamines. Terminally differentiated diploid cells at various levels of organisation (testes tissue, erythropoietic tissue and lymphatic tissue) appear to contain only one typical histone H2B whereas the haploid, terminally differentiated sea urchin sperm cell contains several different types.

In addition to the four prototype histones (H3, H4, H2A and H2B) which form the nucleosome, a fifth group of histones characterised by a high lysine and alanine content, the histones H1, occur in chromatin. It has been suggested that they play a role crosslinking the nucleosomes [47]. In nucleated erythrocytes a group of histones also rich in alanine and lysine and moderately rich in arginine occurs, the histones H5. Comparison of the amino acid composition and the known primary structures of the histones H1 and H5 (fig.3) reveals that there is at this stage no real justification to put these histones into two different categories on structural grounds. It is well established that the histone H1 structures exhibit a considerable degree of variation from species to species and within one species from tissue to tissue [34]. In view of the structural variability of the histones H1 on the one side and the extensive sequence identities with, or homologies to, the histones H5, the latter can be considered as a series of tissue specific members of the histone H1 family, exhibiting a higher arginine content similar to the histone H1 from the sperm cells of the sea urchin *Parechinus angulosus* (fig.3).

The proteins in this histone H1 family are characterised by a core of largely hydrophobic amino acids extending from alignment position 45–117 (fig.3). In this region several stretches of the respective sequences are either virtually identical or highly homologous with most of the amino acid replacements being of the conservative type. Whereas the sea urchin sperm H1 in this region contains four methionine residues (fig.3) the sea urchin embryo histone has only two methionines [23], the erythrocyte specific protein only one and the other histones H1 none (fig.3).

Towards the N-terminus of the molecules (alignment position 44–1) a more variable region follows, rich in basic amino acids, alanine, proline and serine. In this area in the sperm histone, a repeat structure

```
                                    5                   10                  15                  20
Sea Urchin          Pro-Gly-Ser-Pro-Gln-Lys-Arg-Ala-Ala-Ser-Pro-Arg-Lys-Ser-Pro-Arg-Lys-Ser-Pro-Lys-
Trout testis                                                                  AcAla-Glu-Val-Ala-Pro-Ala-Pro-
Chicken H5                                                                                Thr-Glu-Ser-
Rabbit thymus (RTL-2)                  AcSer-Glu-Thr-Ala-Pro-Val-Ala-Pro-Ala-Ala-Pro-Ala-Pro-Ala-
Rabbit thymus (RTL-3)            AcSer-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Glu-Lys-Ser-
Rabbit thymus (RTL-4)            AcSer-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Lys-Ser-Pro-
Calf thumus (CTL-1)         AcSer-Glu-Ala-Pro-Ala-Glu-Thr-Ala-Ala-Pro-Ala-Pro-Ala-Pro-———Lys-Ser-Pro-
Chicken H1 (CEL-5)               AcSer-Glu-Ala-Pro-Thr-Val-Ala-Ala-Pro-Ala-Val-Ser-Ala-Pro-Gly-
Goose H5                                                                      Thr-Asp-Ser-Pro-
Pigeon H5                                                                  Thr-Glu-Ser-Pro-Ile-Pro-
```

	25	30	35	40
Sea Urchin	Lys-Ser-Pro-Arg-Lys-Ala-Ser-Ala-Ser-Pro-Arg-Arg-Lys-Ala-Lys-Arg-Ala-Arg-Ala-Ser-			
Trout testis	Ala-Ala-Ala-Ala-Pro-Ala-Lys-Ala-Pro-Lys-Lys-Lys-Ala-Ala-Ala-Lys-Pro-Lys-Lys-Ser-			
Chicken H5	Leu-Val-Leu-Ser-Pro-Ala-Pro-Ala-Lys-Pro-Lys-Gln/Arg-Val-Lys-Ala-Ser-Arg-Arg-Ser-Ala-			
Rabbit thymus (RTL-2)	Glu-Lys-Thr-Pro-Val-Lys-Ala-Lys-Lys-Lys-Pro-Gly-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-			
Rabbit thymus (RTL-3)	Pro-Ala-Lys-Lys-Lys-Lys-Ala-Ala-Lys-Lys-Pro-Gly-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-			
Rabbit thymus (RTL-4)	Ala-Lys-Thr-Pro-Val-Lys-Ala-Arg-Lys-Lys-Lys-Ser-Ala-Gly-Ala-Ala-Lys-Arg-Lys-Ala-			
Calf thymus (CTL-1)	Ala-Lys-Thr-Pro-Val-Lys-Ala-Ala-Lys-Lys-Lys-Lys-Pro-Ala-Gly-Ala-Arg-Arg-Lys-Ala-			
Chicken H1 (CEL-5)	Ala-Lys-Lys-Pro-Lys-Ala-Ala-Lys-Lys-Lys-Ala-Ala-Gly-Gly-Ala-Lys-Ala-Arg-Lys-Pro-			
Goose H5	Ile-Pro-Ala-Pro-Ala-Pro-Ala-Ala-Lys-Pro-Lys-Arg-Ala-Arg-Ala-Pro-Arg-Lys-Pro-Ala-			
Pigeon H5	Val-Pro-Ala-Pro-Ala-Pro-Ala-Ala-Lys-Pro-Lys-Pro-Lys-Arg-Val-Ser-Lys-Arg-Pro-Ala-			

	45	50	55	60
Sea Urchin	Thr-His-Pro-Pro-Val-Leu-Glu-Met-Val-Gln-Ala-Ala-Ile———Thr-Ala-Met-Lys-Glu-Arg-Lys-			
Trout testis	Gly———Pro-Ala-Val-Gly-Glu-Leu-Ala-Gly-Lys-Ala-Val———Ala-Ala-Ser-Lys-Glu-Arg-Ser-			
Chicken H5	Ser-His-Pro-Thr-Tyr-Ser-Glu-Met-Ile-Ala-Ala-Ala-Ile———Arg-Ala-Glu-Lys-Ser-Arg-Gly-			
Rabbit thymus (RTL-2)	Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala———Ala-Ser-Lys-Glu-Arg-Ser-			
Rabbit thymus (RTL-3)	Ala-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala———Ala-Ser-Lys-Glu-Arg-Asn-			
Rabbit thymus (RTL-4)	Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala———Ala-Ser-Lys-Glu-Arg-Ser-			
Calf thymus (CTL-1)	Ser-Gly-Pro-Pro-Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ala———Ala-Ser-Lys-Glu-Arg-Ser-			
Chicken H1 (CEL-5)	Ala-Gly-Pro———Val-Ser-Glu-Leu-Ile-Thr-Lys-Ala-Val-Ser———Ala-Ser-Lys- region unknown			
Goose H5	Ser-His-Pro-Thr-Tyr-Ser-Glu-Met-Ile-Ala-Ala-Ala-Ile———Arg-Ala-Asp-Lys-Ser-Arg-Gly-			
Pigeon H5	Ser-His-Pro-Pro-Tyr-Ser-Asp-Met-Ile-Ala-Ala-Ala			

	65	70	75	80
Sea Urchin	Gly———Ser-Ser-Ala-Ala-Lys-Ile-Lys-Ser-Tyr-Met-Ala-Ala-Asn-Tyr-Arg-Val-Asp-Met-Asn-			
Trout testis	Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ser-Leu-Ala-Ala-Gly-Gly-Tyr————Asp-Val-Glu-			
Chicken H5	Gly———Ser-Ser-Arg-Gln-Ser-Ile-Gln-Lys-Tyr-Ile-Lys-Ser-His-Tyr-Lys-Val-Gly-His-Asn-			
Rabbit thymus (RTL-2)	Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr-Asp-Val-Glu-Lys-Asn-			
Rabbit thymus (RTL-3)	Gly-Leu-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Gly-Gly-Tyr-Asp-Val-Glu————			
Rabbit thymus (RTL-4)	Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr			
Calf thymus (CTL-1)	Gly-Val-Ser-Leu-Ala-Ala-Leu-Lys-Lys-Ala-Leu-Ala-Ala-Ala-Gly-Tyr			
Chicken H1 (CEL-5)	region unknown Ala-Leu-Ala-Ala-Gly-Gly-Tyr-Asp-Val-Glu-Lys-Gly-			
Goose H5	Gly———Ser-Ser-Arg-Gln-Ser-Ile-Gln-Lys-Tyr-Val-Lys-Ser-His-Tyr-Lys-Val-Gly-Gln-His-			

	85	90	95	100
Sea Urchin	Val-Leu-Ala-Pro-His-Val-Arg-Arg-Ala-Leu-Arg-Asn-Gly-Val-Ala-Ser-Gly-Ala-Leu-Lys-			
Trout testis	Lys-Asn-Asn-Ser-Arg-Val-Lys-Ile-Ala-Val-Lys-Ser-Leu-Val-Thr-Lys-Gly-Thr-Leu-Val-			
Chicken H5	Ala-Asp-Leu-Gln———Ile-Lys-Leu-Ser-Ile-Arg-Arg-Leu-Leu-Ala-Ala-Gly-Val-Leu-Lys-			
Rabbit thymus (RTL-2)	Asn-Ser-Arg			
Rabbit thymus (RTL-3)	Lys-Asn-Asn-Ser-Arg-Ile-Lys-Leu-Gly-Leu-Lys-Ser-Leu-Val-Ser-Lys-Gly-Thr-Leu-Val-			
Chicken H1 (CEL-5)	Asn-Ser-Arg			
Goose H5	Ala-Asp-Leu-Gln-Ile-Lys-Leu———Ala-Ile-Arg-Arg-Leu-Leu-Thr-Thr-Gly-Val-Leu-Lys-			

	105	110	115	120
Sea Urchin	Gln-Val-Thr-Gly-Thr-Gly-Ala-Ser-Gly-Arg-Phe-Arg-Val-Gly-Ala-Val-Ala-Lys-Pro-Lys-			
Trout testis	Glu-Thr-Lys-Gly-Thr-Gly-Ala-Ser-Gly-Ser-Phe-Lys-Leu-Asn-Lys-Lys-Ala-Val-Glu-Ala-			
Chicken H5	Gln-Thr-Lys-Gly-Val-Gly-Ala-Gly-Ser-Ser-Phe-Arg-Leu-Ala-Lys-Ser-Asp-Lys-Ala-Lys-			
Rabbit thymus (RTL-3)	Glu-Thr-Lys-Gly-Thr-Gly-Ala-Ser-Gly-Ser-Phe-Lys-Leu-Asp-Lys-Lys-Ala-Ala-Ser-Gly-			
Goose H5	Gln-Thr-Lys-Gly-Val-Gly-Ala-Ser-Gly-Ser-Phe-Arg-Leu-Ala-Lys-Gly-Asp-Lys-Ala-Lys-			

```
                        125                 130                 135                 140
Sea Urchin              Lys-Ala-Lys-Lys-Thr-Ser-Ala-Ala-Ala-------Lys-Ala-Lys-Lys-Ala-Lys-Ala-Ala-Ala-Ala-Lys-
Trout testis            Lys-Lys-Pro-Ala-Lys-Lys-Ala-Ala-Ala-Pro-Lys-Ala-Lys-Lys----------Val-Ala-Ala-Lys-
Chicken H5              Arg-Ser-Pro-Gly-Lys-Lys-Lys-Ala-Lys(Thr 3; Ser 8; Pro 10; Gly 2; Ala 16; Val 3; Lys 32; Arg 13)
Rabbit thymus (RTL-3)   Glu-Ala-Lys-Pro-Lys-Pro-Lys-Lys-Ala-Gly-Ala-Ala-Lys-Pro-Lys-Lys-Pro-Ala-Gly-Ala-Thr-
Goose H5                Arg-Ser-Pro-Ala-Gly-Arg-Lys-Lys-Lys

                    145                 150                 155                 160
Sea Urchin          Lys-Ala-------Arg-Lys/Arg-Ala-Lys-Ala-Ala-Ala-Lys-Arg-Lys-Ala-Ala-Leu-Ala-Lys-Lys-Lys-Ala-
Trout testis        Lys-------Pro-Ala-Ala-----------------------------------------------------------------------
Rabbit thymus (RTL-3) Pro-Lys-Lys-Pro-Lys-Lys-Ala-Ala-Gly-Ala-Lys-Lys-Ala-Val-Lys-Lys-Thr-Pro-Lys-Lys-Ala-

                    165                 170                 175                 180
Sea Urchin          Ala-Ala-Ala-Lys-Arg-Lys-Ala-Ala-Ala-Lys-Ala-Lys-Lys-Ala-Lys-Lys-Pro-Lys-Lys-Lys-
Trout testis        ----------------------------------------------------------Ala-Lys-Lys-Pro-Lys-Lys-Val-
Rabbit thymus (RTL-3) Pro-Lys-Pro-Lys-Ala-Ala-Ala-Lys-Pro-Lys-Val-Ala-Lys-Pro-Lys-Ser-Pro-Ala-Lys-Val-

                    185                 190                 195                 200
Sea Urchin          Ala-Ala-Lys-Lys-Ala-Lys-Lys-Pro-Ala-Lys-Lys-Ser-Pro-Lys-Lys-Ala-Lys-Lys-Pro-Ala-
Trout testis        Ala-Ala-Lys-Lys-Ala-Val-Ala-------Ala-Lys-Lys-Ser-Pro-Lys-Lys-Ala-Lys-Lys-Pro-Ala-
Rabbit thymus (RTL-3) Ala-Lys-Ser-Pro-Lys-Lys-Ala-Lys-Ala-Val-Lys-Pro-Lys-Ala-Ala-Lys-Pro-Lys-Ala-Pro-

                    205                 210                 215                 220
Sea Urchin          Lys-Lys-Ser-Pro-Lys-Lys-Lys-Lys(Ala-Ala-Gly-Lys-Lys)Arg-Ser-Pro-Lys-Lys-Ala-
Trout testis        ----------Thr-Pro-Lys-Lys----------Ala-Ala-------Lys----------Ser-Pro-Lys-Lys-Ala-
Rabbit thymus (RTL-3) Lys-Pro-Lys-Ala-Ala-Lys-Ala-Lys-Lys-Thr-Ala-Ala-Lys-Lys-Lys-Lys

                        225                 230                 235                 240                 245
Sea Urchin          Lys-Lys-Ala-Ala-Gly-Lys-Arg-Lys-Pro----------------------------Ala-Ala-Lys-Lys-Ala-Arg-Arg-
Trout testis        Thr-Lys-Ala-Ala----------Lys-Pro-Lys-Ala-Ala-Lys-Pro-Lys-Lys-Ala-Ala-Lys----------

                            250                 255                 260
Sea Urchin          Ser-Pro-Arg-Lys-Ala-Gly-Lys-Arg-Arg-Ser-Pro-Lys-Lys(Ala-Arg-Lys)
Trout testis        Ser-Pro-Lys-Lys----------Val-Lys-Lys-Pro-Ala-Ala-Ala-Lys-Lys
```

Fig.3. Sequences of histones H1, aligned for homologies. Numbering refers to alignment positions and not to sequence positions. Deletion in the alignment: (——). The total number of residues is only known for the sea urchin sperm, rabbit thymus RTL3 and trout testis variants. The extent of unsequenced regions therefore cannot be indicated. Mammal: *Bos taurus* calf thymus CTL-1 partial sequence [36]; *Oryctolagus*, thymus RTL-2 and RTL-4 partial sequence [36], thymus RTL-3 complete sequence [36]. Birds: *Gallus domesticus* (chicken), erythrocyte H5 partial sequence [37], erythrocyte CEL-5 partial sequence [36]; *Anser anser* (goose) partial sequence [38]; *Columba spp.* (pigeon) erythrocyte partial sequence [38]. Fish: *Salmo spp.* (trout) testis complete sequence [39]. Sea urchin: *Parechinus angulous* sperm cells complete sequence [40].

Ser–Pro (basic residue)$_2$ can be recognised (alignment positions: 10–13; 14–17; 18–21; 22–25; 29–32). This regular structure is absent from the other members of the family. The length of this region varies, being shorter by 15–18 residues in the erythrocyte specific histones and longest in the sperm histone. There is little homology in this region between the proteins originating from different cell types though variants isolated from comparable cells exhibit a high degree of homology.

Towards the C-terminus from the central core a second basic region extends from alignment position 120–263. The complete primary structure in this domain has only been established for the variants from trout testis, rabbit thymus RTL3 and sea urchin sperm, for the remainder of the variants the information is fragmentary. The region is rich in lysine and alanine with interspaced proline residues. In sea urchin sperm histone the proline often occurs next to serine and both are usually flanked by basic residues, for example:

$(Lys)_2Ser–Pro(Lys)_2$ (Position 190–195)

$(Lys)_2Ser–Pro(Lys)_4$ (Position 201–208)

$(Lys)–(Arg)_2Ser–Pro(Lys)_2$ (Position 212–218)

$(Arg)_2Ser–Pro–Arg$ (Position 234–238)

$Lys–Ser–Pro–(Arg)_2$ (Position 239–243)

Such an association of the serine–proline sequence with basic residues also occurs in the trout testis variant (fig.3).

In both proteins most of those proline residues not associated with serine are also flanked in a similar fashion by basic amino acids (fig.3).

If one accepts a number of deletions (fig.3) then the carboxy terminal domain of trout testis and sea urchin sperm histone H1 are highly homologous. Both can be envisaged to have evolved via a number of duplications (fig.4).

In both histone groups, H1 and H2B, it becomes apparent that in the course of evolution selection pressure has led to the repeated duplication of amino acid sequences suitable to interact with DNA until an optimal degree of interaction is reached. In this fashion the pentapeptide repeat structures in the N-terminal domains of the histones H2B became established. Similarly the more extended repeat structures in the C-terminal domain of the histones H1 from sea urchin sperm and trout testis may have arisen in this fashion (fig.4). Though this construction principle is also

Fig.4. The carboxyl terminal region of histones H1 from sea urchin sperm cells, trout testis and rabbit thymus. The sequences have been arranged to indicate possible evolutionary pathways by a series of variably sized duplications. The numbers shown correspond to alignment position as given in fig.3.

recognisable in the rabbit thymus histone H1 variant the repeat sequences in the C-terminal part are not as clearly preserved as in the trout and sea urchin histone H1 variants (fig.4).

2. Histone variants and differentiation

Cells involved in differentiation processes appear to contain a number of variants not only of histone H2B but also of histone H2A (fig.1,2) and H1 [23]. This leads to the previously mentioned notion that the variability of histone structures is an expression of qualitatively different genetic activity of cells during the life cycle of one organism at any particular evolutionary level, in short, the existence of histone programmes during differentiation.

The evidence for this is compelling. Whereas early sea urchin gastrula contains only one predominant type of histone H2B, the late gastrula contains several distinct variants (fig.2). Once differentiation into various tissues is completed, a typical diploid cell (sea urchin gut) contains again only a single predominant histone variant, previously not present at the embryonic stages (fig.2). The haploid sperm cells contain three entirely different tissue specific histone H2B variants not detectable at the embryonic stage. The histone H2B from wheat germ, also an embryonic tissue, (fig.2) is similar in structure to the animal-type H2B in that the C-terminal largely hydrophobic part is highly homologous to the animal histones. The N-terminal is, compared to animal histones H2B, variable and considerably longer than the somatic animal H2B histones. Whereas the corresponding region in animal histones comprises 30–40 residues the wheat germ H2B has in this region close to 60 residues. Several structural variants, not yet fully characterised, of the H2B variety occur in the embryonic tissue of this plant [35].

From these data on the primary structures of histone H2B variants, the general picture of a protein molecule emerges which has been subjected in the course of evolution to two entirely different selection pressures focussed onto different parts of the molecule. Histone–histone interaction, necessary for the reproducible assembly of the chromatin structure, has favoured a relative constancy in the C-terminal hydrophobic part of the molecule whereas the changing demands of histone–DNA interaction have selected a large variety of N-terminal sequences. The latter can thus provide a variety of chromatin structures suitable for specific regulatory processes.

Though only limited sequence data are available for the histone H2A variants from animals, a similar situation appears to pertain to those histones (fig.1). Several variants characterised at this stage only by different amino acid compositions and partial sequences have been isolated from embryonic tissue, fully differentiated diploid cells and haploid sperm cells of the same sea urchin species [23].

The characterization of histones H2A from plants is even more incomplete than that from animal histones. Only partial structures of these histones isolated from wheat germ, an embryonic tissue, became recently available (fig.1). Several variants of histone H2A are present [48]. One type is very similar in overall structure to the prototype histone from fully differentiated diploid animal cells (calf thymus) but with 17 point mutations in the first 55 residues which have been positioned. The other two types of variants are characterized by a similar sequence in that region but have additional N-terminal extensions (fig.1).

These recent developments in the identification and characterization of histone variants by sequence analysis and the demonstration that in the sea urchin these variants appear in a programmed fashion, provide the structural proof for the notion of a specific role of histones in the process of differentiation. Such a role was suggested earlier by the results of the investigation of Asao [49] on differences in the regional distribution of histone subclasses in newt blastula and gastrula. In a similar vein, Benttinen and Comb [50] found stage-specific histone fractions on gel electrophoresis of sea urchin embryo histones; similar results were reported by a number of other groups [51–53]. Cohen, Newrock and Zweidler [54] supplied additional electrophoretic evidence showing a programmed appearance of pulse-labelled histone variants in the developing sea urchin embryo. Arceci et al. [55] and Newrock et al. [56] have reported stage-specific histone mRNA species in sea urchin embryos.

Generally the evaluation of gel electrophoretic results is complicated by the fact that histones with completely different primary structure may have under certain conditions the same electrophoretic

mobility. For example, histone $H2B_{(1)\ Parechinus}$ and $H2B_{(3)\ Parechinus}$ on the one side and $H2B_{(2)\ Parechinus}$ and H3 on the other have identical electrophoretic mobility [27–29], as have the sea urchin embryo histone H2A and H2B [23]. Some of the plant histones of the H2A type have a mobility very similar to that of histone H3 and are virtually indistinguishable from each other [48]. On the other hand it is well established that amino acid side chain modification leading to a change of the size–charge ratio of the molecule can have a pronounced influence on the electrophoretic mobility resulting in distinct fractions with however identical amino acid sequences (for review see [57]). The unequivocal proof of the existence of histone variants obviously lies in the elucidation of their amino acid sequence.

The programmed appearance of histone variants during embryonic development, the occurrence of certain variants in highly specialized cells and the occurrence of histones with particular structures at specific levels of evolutionary complexity of organisms requires a more intricate explanation which goes beyond that of the provision of a uniform structural framework to accommodate the DNA.

3. Histone variants and chromatin structure

The periodicity of nuclease-sensitive sites in chromatin [58], the existence of specific histone–histone complexes in solution [59], the presence of these specific, histone pairs in chromatin [60,61], together with electron microscopic evidence [62], led to the proposal by Kornberg [63] of the universal, uniform chromatin subunit, the nucleosome [63]. Models of this subunit are based on data derived from chemical crosslinking of histones in situ [61,64], neutron scattering [47,65–69], high resolution electron microscopy [70], limit digestion of chromatin with nucleases [71–73], limit digestion of subunits with proteolytic enzymes [74,75] and extrapolation from the conformation of individual histones in solution as revealed by NMR spectroscopy [76–78].

Most of the experimental data are best met by a model of core particles consisting of a central protein body associated on its circumference with a supercoiled DNA dyad of ~140 base pairs. The X-ray diffraction pattern of core particle crystals is consistent with this model. The wedge-shaped, disc-like particle having the dimensions of 110 × 110 × 57 Å is composed of histones, associated with 1¾ turns of a superhelical DNA dyad with a 28 Å pitch and a 90 Å diam. [79]. These crystallographic dimensions are in good agreement with those determined by electron microscopic techniques [70] and with those of the model proposed on the basis of neutron-scattering and X-ray scattering data [69].

The present X-ray crystallographic data with a 25 Å resolution [79], do not allow conclusions concerning the organisation of the histone core. The latter, on the basis of crosslinking with di-imido esters is seen as an octamer, consisting of a pair of each of the four histone H2A, H2B, H3 and H4 [64]. In this octamer close contacts between histone H2A and H4 as well as H2A and H2B have been revealed by crosslinking with ultraviolet light or with tetranitromethane [80,81]. Similarly, close H3–H3 contacts in the region of the cysteine residue 110 exist [72].

It appears that the histones H3 and H4 largely determine the organisation of DNA into a subunit structure, judged by the periodicity of nuclease-sensitive sites in DNA-histone complexes assembled in vitro [82,83].

The subunit model with its central protein core interacting on its surface with a supercoiled DNA dyad is in general agreement with the conformation one would expect the histones to assume, even in the absence of DNA, given their primary structures characterized by a clustering of basic amino acids and hydrophobic residues, respectively.

The basic regions lying on the surface of the histone complex determine the protein–DNA interaction. It is the latter which is probed with the various nucleases. Transcriptionally active regions of chromatin are particularly susceptible to DNase 1 [75,84] but are not preferentially nicked by micrococcal nuclease. The size of the DNA segment attached to the chromatin subunit in such regions appears to be shorter [85].

These observations correlate well with the occurrence of histone variants specific for the particular functional state of chromatin. This strongly suggests that histone variants may be one of the determinants responsible for the different sizes of DNA fragments produced by nuclease nicks from chromatin of various

tissues. The histone variants, presumably, lend to the DNA a conformation which not only facilitates nuclease cleavage but also transcriptional or not-yet-defined essential pre-transcriptional events. Post-synthetic modification, like acetylation, may well fulfil a similar function [86]. On the other hand, N-terminal extension of the histones H2B and H1 variants in sea urchin sperm (fig.2,3) as well as in the histone H2A and H2B in the dormant embryonic tissue of wheat germ (fig.1,2) appears to be one of the structural pre-requisites for the packing of DNA during temporary suppression of transcriptional activity.

The fifth histone family, the H1 histones, is probably involved in the close packing of the elementary nucleosome repeat structure into higher orders of chromatin condensation. Their likely tertiary structure as proposed by Hartman et al. [87] entailing an N-terminal basic random coil extension, the 'nose', a globular structure of the hydrophobic core and a C-terminal 'tail' (fig.3) can easily be envisaged to fulfil a crosslinking function between nucleosomes [47] otherwise widely separated. On the other hand, by attaching itself to the stretch of internucleosomal linker-DNA [88] via the 'nose' and the 'tail', one could envisage the histone H1 providing centres for protein–protein interaction in between adjacent nucleosomes through the protruding globular core of the histone H1. Such protein–protein interaction centres could play a role in the formation of super-coil structures.

The construction of the N-terminal extension of the histones H1 appears to vary considerably from molecule to molecule not only in total length but also in the degree of potential secondary structure, e.g., the occurrence of several proline-containing tetrapeptide repeats (fig.3) in the sea urchin sperm variant. Such variations will allow the chromatin to take on highly specific conformations in certain areas.

The nuclease-produced 140–180 base pair pieces [71] are usually taken as evidence for the existence of the nucleosome. Such regularly spaced nicks can, however, be produced equally well with micrococcal nuclease in the presence of 5 M urea [89], though at this urea concentration, the typical X-ray diffraction pattern of chromatin with the 110 Å deflection has disappeared indicating the transition to an extended structure in which the nucleosome has vanished [90]. Also, an ionic strength-dependent transition between extended fiber structure and a nucleosome pattern becomes evident in electron microscopic investigations [91]. Since under both these conditions DNA–histone salt linkages remain intact, the usefulness of micrococcal nuclease as a nucleosome probe is limited, particularly in transcriptionally active chromatin like lampbrush chromosomes and nucleoli in which electron microscopy has revealed the absence of nucleosomes [92], though the nuclease digestion pattern of DNA from nucleolar chromatin is indistinguishable from the nucleosomal DNA pattern of quiescent tissue [93]. It thus appears that differential susceptibility to either micrococcal nuclease or DNase is a probe only for the extent of interaction between the basic regions of the histones and DNA reflecting a particular DNA conformation, and that the digestion pattern is largely independent of the presence of possible quarternary structures of the histone protein within the histone–DNA complexes, one of which gives rise to the nucleosome structure.

The quaternary structure of the histone protein, its potential for dissociation into subunits and into its component polypeptide chains as well as the degree of its interaction with DNA will be determined primarily by the structure of the component polypeptide chains, histone H3, H4, H2A and H2B. In the course of evolution, the component polypeptide chains have been selected to enable the chromatin to fold and unfold reproducibly for the replication cycle, a function which is based on protein–protein recognition. The fundamental identity of that process in all eucaryotic cells probably accounts for the conservative structure of the polypeptide chains of the histones H3 and H4 and is reflected in the high degree of homology in the largely hydrophobic C-terminal half of the many histone H2B variants, and also in the homologies of the globular cores of the histones H1. The demands for suitable protein dictated DNA conformations on which regulatory processes can manifest themselves have shaped the variable N-terminal regions in the histones H2B, H2A and H1.

Further work on the comparison of histone structures at different levels of evolutionary complexity as well as during the process of differentiation may well provide the answer to the tantalizing question of the validity also at the molecular level of Haeckel's 'fundamental biogenetic law', formulated in 1866,

on morphological evidence, that 'ontogeny is a short recapitulation of phylogeny'.

Acknowledgements

We thank Miss P. Angear for typing the manuscript and Mrs M. Behrens for assistance in compiling the bibliography. With Paul van Helden, Jerry Rodrigues, Dennis Maeder and Peter Sheppard, postgraduate students in our department, we had many stimulating discussions. The experimental work done in this laboratory has been sponsored by grants of the Council for Scientific and Industrial Research, Republic of South Africa, and the University of Cape Town Research Committee to C.v.H.

References

[1] De Lange, R. J., Hooper, J. A. and Smith, E. L. (1973) J. Biol. Chem. 248, 3261–3274.
[2] Franklin, S. G. and Zweidler, A. (1977) Nature 266, 273–275.
[3] Brandt, W. F. and Von Holt, C. (1974a) Eur. J. Biochem. 46, 407–417.
[4] Brandt, W. F. and Von Holt, C. (1974b) Eur. J. Biochem. 46, 419–429.
[5] Hooper, J. A., Smith, E. L., Sommer, K. R. and Chalkley, R. (1973) J. Biol. Chem. 248, 3275–3279.
[6] Candido, E. P. M. and Dixon, G. H. (1972) Proc. Natl. Acad. Sci. USA 69, 2015–2019.
[7] Brandt, W. F., Strickland, W. N., Morgan, M. and Von Holt, C. (1974a) FEBS Lett. 40, 167–172.
[8] Brandt, W. F., Strickland, W. N. and Von Holt, C. (1974b) FEBS Lett. 40, 349–352.
[9] Brandt, W. F. and Von Holt, C. (1976) FEBS Lett. 65, 386–390.
[10] Goff, C. G. (1976) J. Biol. Chem. 251, 4131–4138.
[11] Patthy, L. and Smith, E. L. (1973) J. Biol. Chem. 248, 6834–6840.
[12] De Lange, R. J., Fambrough, D. M., Smith, E. L. and Bonner, J. (1969a) J. Biol. Chem. 244, 319–334.
[13] Quagliarotti, G., Ogawa, Y., Taylor, C. W., Sautiere, P., Jordan, J., Starbuck, W. C. and Busch, H. (1969) J. Biol. Chem. 244, 1796–1802.
[14] Ogawa, Y., Quagliarotti, G., Jordan, J., Taylor, C. W., Starbuck, W. C. and Busch, H. (1969) J. Biol. Chem. 244, 4387–4392.
[15] Desai, L., Ogawa, Y., Mauritzen, C. M., Taylor, C. W. and Starbuck, W. C. (1969) Biochim. Biophys. Acta 181, 146–153.
[16] Strickland, M., Strickland, W. N., Brandt, W. F. and Von Holt, C. (1974) FEBS Lett. 40, 346–348.
[17] Wouters-Tyrou, D., Sautiere, P. and Biserte, G. (1976) FEBS Lett. 65, 225–228.
[18] De Lange, R. J., Fambrough, D. M., Smith, E. L. and Bonner, J. (1969b) J. Biol. Chem. 244, 5669–5679.
[19] 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.
[20] Blankstein, L. A., Stollar, B. D., Franklin, S. G., Zweidler, A. and Levy, S. B. (1977) Biochemistry 16, 4557–4562.
[21] Laine, B., Sautiere, P. and Biserte, G. (1976) Biochemistry 15, 1640–1645.
[22] Bailey, G. S. and Dixon, G. H. (1973) J. Biol. Chem. 248, 5463–5472.
[23] Brandt, W. F., Strickland, W. N., Strickland, M., Carlisle, L., Woods, D. and Von Holt, C. (1979) Eur. J. Biochem. in press.
[24] Iwai, K., Hayashi, H. and Ishikawa, K. (1972) J. Biochem. (Tokyo) 72, 357–367.
[25] Van Helden, P., Strickland, W. N., Brand, W. F. and Von Holt, C. (1978) Biochim. Biophys. Acta 533, 278–281.
[26] Kootstra, A. and Bailey, G. S. (1976) FEBS Lett. 68, 76–78.
[27] Strickland, M., Strickland, W. N., Brandt, W. F. and Von Holt, C. (1977a) Eur. J. Biochem. 77, 263–275.
[28] Strickland, W. N., Strickland, M., Brandt, W. F. and Von Holt, C. (1977b) Eur. J. Biochem. 77, 277–286.
[29] Strickland, M., Strickland, W. N., Brandt, W. F., Von Holt, C., Lehmann, A. and Wittmann-Liebold, B. (1978a) Eur. J. Biochem. 89, 443–452.
[30] Brandt, W. F. and Von Holt, C. (1978b) Biochim. Biophys. Acta 537, 177–181.
[31] Strickland, M., Strickland, W. N., Brandt, W. F. and Von Holt, C. (1978b) Biochim. Biophys. Acta 536, 289–297.
[32] Strickland, W. N., Strickland, M. and Von Holt, C. (1978c) unpublished.
[33] Van Helden, P. D., Strickland, W. N., Brandt, W. F. and Von Holt, C. (1978) Eur. J. Biochem. in press.
[34] Elgin, S. C. R. and Weintraub, H. (1975) Annu. Rev. Biochem. 44, 725–774.
[35] Brandt, W. F., Rodrigues, J. de A. and Von Holt, C. (1978b) unpublished.
[36] Cole, R. D. (1977) in: The Molecular Biology of the Mammalian Genetic Apparatus (Ts'o, P. O. P. ed) vol. 1, pp. 93–104, Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
[37] Sautiere, P., Briand, G., Kmiecik, D., Loy, O., Biserte, G., Garel, A. and Champagne, N. (1976) FEBS Lett. 63, 164–166.
[38] Yaguchi, M., Roy, C., Dove, M. and Seligy, V. (1977) Biochem. Biophys. Res. Commun. 76, 100–106.
[39] Macleod, A. R., Wang, N. C. W. and Dixon, G. H. (1977) Eur. J. Biochem. 78, 281–291.

[40] Strickland, W. N., Strickland, M., Brandt, W. F., Von Holt, C., Lehmann, A. and Wittmann-Liebold, B. (1978) unpublished.
[41] Huntley, G. H. and Dixon, G. H. (1972) J. Biol. Chem. 247, 4916–4919.
[42] Watson, D. C., Peters, E. H. and Dixon, G. H. (1977) Eur. J. Biochem. 74, 53–60.
[43] Walker, J. M., Goodwin, G. H., Johns, E. W., Wietzes, P. and Gaastra, W. (1977) Int. J. Prot. Pept. Res. 9, 220–223.
[44] Walker, J. M., Hastings, J. R. B. and Johns, E. W. (1977) Biochem. Biophys. Res. Commun. 73, 72–78.
[45] Rall, S. C. and Cole, R. D. (1971) J. Biol. Chem. 246, 7175–7190.
[46] Ando, T., Yamasaki, M. and Suzuki, K. (1973) in: Molecular Biology Biochemistry and Biophysics, vol. 12, pp. 58–80, Springer, Berlin, Heidelberg.
[47] Baldwin, J. P., Boseley, P. G., Bradbury, E. M. and Ibel, K. (1975) Nature 253, 245–249.
[48] Rodrigues, J. de A., Brandt, W. F. and Von Holt, C. (1978) Biochim. Biophys. Acta, in press.
[49] Asao, T. (1970) Exp. Cell Res. 61, 255–265.
[50] Benttinen, L. C. and Comb, D. G. (1971) J. Mol. Biol. 57, 355–358.
[51] Easton, D. and Chalkley, R. (1972) Exp. Cell Res. 72, 502–508.
[52] Ruderman, J. V., Baglioni, C. and Gross, P. R. (1974) Nature 247, 36–38.
[53] Gineitis, A. A., Stankeviciute, J. V. and Vorob'ev, V. I. (1976) Dev. Biol. 52, 181–192.
[54] Cohen, L. H., Newrock, K. M. and Zweidler, A. (1975) Science 190, 994–997.
[55] Arceci, R. J., Senger, D. R. and Gross, P. R. (1976) Cell 9, 171–178.
[56] Newrock, K. M., Cohen, L. H., Hendricks, M. B., Donnelly, R. J. and Weinberg, E. S. (1978) Cell 14, 327–336.
[57] Dixon, G. H., Candido, E. P. M., Honda, B. M., Louie, A. J., Macleod, A. R. and Sung, M. T. (1975) in: The Structure and Function of the Chromatin, Ciba Found. Symp. vol. 28, pp. 229–250, Elsevier/Excerpta Medica/North-Holland, Amsterdam, New York, Oxford.
[58] Hewish, D. R. and Burgoyne, L. A. (1973) Biochem. Biophys. Res. Commun. 52, 504–510.
[59] Van der Westhuyzen, D. R. and Von Holt, C. (1971) FEBS Lett. 14, 333–337.
[60] Kornberg, R. D. (1974) Science 184, 868–871.
[61] Kornberg, R. D. and Thomas, J. O. (1974) Science 184, 865–868.
[62] Olins, A. L. and Olins, D. E. (1974) Science 183, 330–331.
[63] Kornberg, R. D. (1977) Annu. Rev. Biochem. 46, 931–954.
[64] Thomas, J. O. and Kornberg, R. D. (1975) Proc. Natl. Acad. Sci. USA 72, 2626–2630.
[65] Bram, S., Butler-Browne, G., Bradbury, E. M., Baldwin, J., Reiss, C. and Ibel, K. (1974) Biochimie 56, 987–994.
[66] Bram, S., Butler-Browne, G., Baudy, P. and Ibel, K. (1975) Proc. Natl. Acad. Sci. USA 72, 1043–1045.
[67] Pardon, J. F., Worcester, D. L., Wooley, J. C., Thatchell, K., Van Holde, K. E. and Richards, B. M. (1975) Nucl. Acids Res 2, 2163–2176.
[68] Baudy, P., Bram, S., Vastel, D., Lepault, J. and Kitzis, A. (1976) Biochem. Biophys. Res. Commun. 72, 176–183.
[69] Pardon, J. F., Worcester, D. L., Wooley, J. C., Cotter, R. I., Lilley, D. M. J. and Richards, B. M. (1977) Nucl. Acids Res. 4, 3199–3214.
[70] Langmore, J. P. and Wooley, J. C. (1975) Proc. Natl. Acad. Sci. USA 72, 2691–2695.
[71] Axel, R., Melchior, W., Sollner-Webb, B. and Felsenfeld, G. (1974) Proc. Natl. Acad. Sci. USA 71, 4101–4105.
[72] Camerini-Otero, R. D. and Felsenfeld, G. (1977) Proc. Natl. Acad. Sci. USA 74, 5519–5523.
[73] Felsenfeld, G., Sollner-Webb, B. and Camerini-Otero, R. D. (1977) in: The Molecular Biology of the Mammalian Genetic Apparatus (T'so, P. O. P. ed) vol. 1, pp. 255–263, Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
[74] Weintraub, H., Palter, K. and Van Lente, F. (1975) Cell 6, 85–110.
[75] Weintraub, H. and Groudine, M. (1976) Science 193, 848–856.
[76] Bradbury, E. M. and Rattle, W. E. (1972) Eur. J. Biochem. 27, 270–281.
[77] Bradbury, E. M. et al. (1975) in: The Structure and Function of Chromatin, Ciba Found. Symp. vol. 28, pp. 131–148, Elsevier/Excerpta Medica/North-Holland, Amsterdam, New York, Oxford.
[78] Bradbury, E. M., Hjelm, R. P., Carpenter, B. G., Baldwin, J. P., Kneale, G. G. and Hancock, G. G. (1977) in: The Molecular Biology of the Mammalian Genetic Apparatus (T'so, P. O. P. ed) vol. 1, pp. 53–70, Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
[79] Finch, J. T., Lutter, L. C., Rhodes, D., Brown, R. S., Rushton, B., Lewitt, M. and Klug, A. (1977) Nature 269, 29–36.
[80] Martinson, H. G. and McCarthy, B. J. (1976) Biochemistry 15, 4126–4131.
[81] Martinson, H. G., Shetlar, M. D. and McCarthy, B. J. (1976) Biochemistry 15, 2002–2007.
[82] Camerini-Otero, R. D., Sollner-Webb, B. and Felsenfeld, G. (1976) Cell 8, 333–347.
[83] Moss, T., Stephens, R. M., Crane-Robinson, C. and Bradbury, E. M. (1977) Nucl. Acids Res. 4, 2477–2485.
[84] Gottesfeld, J. M. and Partington, G. A. (1977) Cell 12, 953–962.
[85] Keichline, L. D. and Wassarman, P. M. (1977) Biochim. Biophys. Acta 475, 139–151.
[86] Sealy, L. and Chalkley, R. (1978) Nucl. Acids Res. 5, 1863–1876.
[87] Hartman, P. G., Chapman, G. E., Moss, T. and Bradbury, E. M. (1977) Eur. J. Biochem. 77, 45–51.

[88] Whitlock, J. P. and Simpson, R. T. (1977) J. Biol. Chem. 252, 6516–6520.
[89] Jackson, V. and Chalkley, R. (1975) Biochem. Biophys. Res. Commun. 67, 1391–1400.
[90] Carlson, R. D., Olins, A. L. and Olins, D. E. (1975) Biochemistry 14, 3122–3125.
[91] Tsanev, R. and Petrov, P. (1976) J. Micro. Biol. Cell. 27, 11–18.
[92] Franke, W. W., Scheer, U., Trendelenburg, M. F., Spring, H. and Zentgraf, H. (1976) Cytobiologie 13, 401–434.
[93] Piper, P. W., Celis, J., Kaltoft, K., Leer, J. C., Nielsen, O. F. and Westergaard, O. (1976) Nucl. Acids Res. 3, 493–505.
[94] Haeckl, E. (1866) Generelle Morphologie der Organismen Reimer, Berlin.