Biochem. J. (1986) 235, 769–773 (Printed in Great Britain)



# Multiple forms of histone H2B from the nematode *Caenorhabditis elegans*

Jacques R. VANFLETEREN,*§ Stefaan M. VAN BUN,† Lucien L. DELCAMBE‡ and Jozef J. VAN BEEUMEN†

*Laboratorium voor Morfologie en Systematiek der Dieren, Rijksuniversiteit Gent, Ledeganckstraat 35, 9000 Gent, Belgium, †Laboratorium voor Microbiologie en Microbiële Genetica, Rijksuniversiteit Gent, Ledeganckstraat 35, 9000 Gent, Belgium, and ‡Centre National pour la Production et l'Étude de Substances d'Origine Microbienne, C.N.P.E.M., Boulevard du Rectorat 29, Bâtiment 40, Sart Tilman, 4000 Liège, Belgium

The complete amino acid sequence of histone H2B from the nematode *Caenorhabditis elegans* was determined. The protein as obtained by us is a mixture of multiple forms. Approx. 90% of the molecules consist of a polypeptide chain of 122 amino acids with alanine as *N*-terminal residue and proline at the second position. In the remaining 10% alanine is lacking and the chain starts with proline. In addition to the heterogeneity of chain length, polymorphism occurs at the positions 7 (Ala/Lys), 14 (Ala/Lys) and 72 (Ala/Ser) of the major chain and at position 6 (Ala/Lys) of the shorter chain. In the *N*-terminal third of the molecule there is a high degree of sequence homology to the corresponding region in H2B from *Drosophila* (insect), *Patella* (mollusc) and *Asterias* (starfish). In contrast, this part of the molecule differs considerably from mammalian histone H2B.

## INTRODUCTION

It has now been well established that histones H1, H2A, H2B and H3 consist of a small number of polypeptides that differ slightly in their primary structure. Many of these variants exhibit tissue-specific variation or follow a developmental pattern (Cohen *et al.*, 1975; Poccia & Hinegardner, 1975; Franklin & Zweidler, 1977; Grunstein, 1978; Brandt *et al.*, 1979; Childs *et al.*, 1979; Hieter *et al.*, 1979; Poccia *et al.*, 1981; Newrock *et al.*, 1982). We now report that histone H2B from the free-living nematode *Caenorhabditis elegans* consists of at least five different polypeptides. Attention has been focused on this nematode as a possible organism for the study of the genetic regulation of development, aging and behaviour in eukaryotic organisms. It offers the advantages of small size, short life-cycle, ease of cultivation in both axenic and monoxenic media, a genetically manipulatable system and a strictly determined development with well-characterized embryonic and post-embryonic cell lineages (Brenner, 1974; Zuckerman, 1980).

## EXPERIMENTAL

### Nematode growth

A dauer defective strain DR27 [*daf-17(m27)* of *C. elegans* was used in this study; it was provided by the *Caenorhabditis* Genetics Center, which is supported by Contract no. N01-AG-9-2113 by the U.S. National Institutes of Health and the Curators of the University of Missouri. The worms were grown on 3 kg of *Escherichia coli* cells suspended in 80 litres of S buffer (0.1 M-NaCl/0.05 M-potassium phosphate buffer, pH 6) containing 0.1% antifoam solution (M-30; Serva, Heidelberg, Federal Republic of Germany) and 50 $\mu$g of cholesterol/ml. About 100 g of axenically cultured nematodes was used as inoculum for the batch culture, which was vigorously aerated for 6 days. The yield was 1.3 kg (wet wt.) of nematodes. They were freed of bacteria and debris by flotation on 30% (w/v) sucrose/20% (v/v) silica gel (Ludox; Du Pont de Nemours, Wilmington, DE, U.S.A.) in S buffer. Repeated sedimentation in S buffer alone removed excess sucrose and Ludox. The pellets from the last wash were suspended in an equal volume of S buffer, dripped into liquid $N_2$ and stored at −196 °C.

### Preparation of histone H2B

For the preparation of histones and the purification of histone H2B we developed a considerably improved procedure compared with the one published earlier (Vanfleteren, 1982). After addition of sucrose to a final concentration of 1.7 M and an equal amount of acid-washed glass beads (0.2 mm in diameter), the nematodes were ground in a Waring Blendor for 2 min. To the homogenate phenylmethanesulphonyl fluoride, $CaCl_2$ and Triton X-100 were added at final concentrations of 2 mM, 5 mM and 1% (v/v) respectively. Nuclei and nematode fragments were sedimented at 10000 *g* for 30 min. The sediment was washed three times with 0.14 M-NaCl/10 mM-EDTA/10 mM-Tris/HCl buffer, pH 7.4, suspended in 1 M-NaCl/10 mM-EDTA/10 mM-Tris/HCl buffer, pH 7.4, and sheared in a VirTis homogenizer at position 40 for 2 min. The sediment obtained by centrifugation at 8000 *g* for 10 min was extracted once more. The supernatants were combined and diluted to 0.4 M-NaCl with distilled water, and nucleohistone was precipitated by adding 1.7 vol. of 96% (v/v) ethanol/1% (v/v) 2-mercaptoethanol. Histones were extracted from the precipitated nucleohistone with 0.2 M-$H_2SO_4$ and precipitated with 10 vol. of acetone. After redissolution in 10 mM-HCl, histone H1 was extracted selectively with 5% (v/v) $HClO_4$. The insoluble

§ To whom all correspondence should be addressed.

residue, which consisted of core histones and contaminating non-histone proteins, was washed once with acetone containing 0.1 M-HCl and then three times with pure acetone, and dried under reduced pressure. The proteins were redissolved in 0.35 M-NaCl/5 M-urea/10 mM-Tris/HCl buffer, pH 7.2, and purified by cation-exchange column chromatography on Bio-Rex 70. Pure histones were retained by the resin under these conditions and eluted with 40% (w/v) guanidinium chloride. After removal of salt by gel filtration (Bio-Gel P-6), the preparation was made 0.18 M with respect to $H_2SO_4$, and the histones were precipitated by adding 10 vol. of acetone. The proteins thus prepared are identical with the histones extracted from pure nuclei or chromatin by both chemical and electrophoretic criteria.

The slightly lysine-rich histones H2A and H2B were separated from the arginine-rich histones H3 and H4 by gel filtration (Bio-Gel P-100) in acetate/bisulphite buffer, pH 5.6, as described by Van der Westhuyzen & Von Holt (1971). The histones H2A and H4 (only partially removed by the preceding step) were removed by selective extraction by 80% (v/v) ethanol in 0.25 M-HCl. The remaining histone H2B was then finally purified by gel-exclusion chromatography on Bio-Gel P-60 in 0.1 M-guanidium chloride/0.02 M-HCl. The protein composition of the peak fractions was examined by SDS/polyacrylamide-gel electrophoresis (Thomas & Kornberg, 1975). Fractions over 90% pure were pooled and used for sequence analysis.

### Production and separation of peptides

About 5 mg of histone H2B was digested with *Staphylococcus aureus* V8 proteinase (Miles Laboratories, Slough, Berks., U.K.) for 20 h at pH 4, at an enzyme/substrate ratio of 1:40 (w/w) (Houmard & Drapeau, 1972). Two insoluble (Sa 9 and Sa 10) and eight soluble (Sa 1–Sa 8) peptides were obtained, which could be separated by preparative t.l.c. on 0.5 mm-thick cellulose plates (Merck, Darmstadt, Germany) developed in pyridine/butan-1-ol/acetic acid/water (10:15:3:12, by vol.). The position of the peptides was revealed by staining two side lanes and a central lane with 1% ninhydrin or 0.05% fluorescamine. Intact peptide material was then eluted with 50% (v/v) acetic acid as described by Chen (1976).

Peptide fraction Sa 1 was subdigested with proline-specific endopeptidase (Miles Laboratories) for 2 h at 32 °C and at an enzyme/substrate ratio of 1:40 (w/w). The resulting peptides were separated by reversed-phase h.p.l.c. on a Vydac C4 column (4 mm × 250 mm) equilibrated with 0.1% trifluoroacetic acid. Gradient elution with a linearly increasing concentration of acetonitrile was carried out by using a Du Pont 8800 liquid chromatograph.

Cleavage at methionine residues was carried out on 1.8 mg of the native protein by following the procedure of Alfageme *et al.* (1974). Three major peptides were obtained, also separated by t.l.c., and designated CB 1–CB 3.

### Automated sequence analysis

Automated Edman degradation chemistry was carried out by using the Applied Biosystems 470 gas-phase Sequenator. The phenylthiohydantoin derivatives of amino acids were quantitatively analysed by reversed-phase h.p.l.c. on a 5 μm-pore-size IBM cyanopropyl-silica column (4.6 mm × 250 mm) according to Hunkapiller & Hood (1983).

## RESULTS AND DISCUSSION

The complete amino acid sequence of the protein is given in Fig. 1. Sequence analysis on 6.5 nmol of the native protein allowed the identification of 40 out of the first 49 residues. At several positions two residues were found in a ratio of 5–10:1. The result is in agreement with the presence of two histone H2B species that differ in length by one residue, the *N*-terminal alanine. As such, each Edman cycle revealed a major amino acid and a minor component, occurring as the major amino acid in the next cycle. At positions where an amino acid is followed by an identical one at the next position, as for example for Pro-2, Lys-11 and Lys-24, a pre-indicative increase of this next residue could not be seen. Beyond residue 25 the amounts of residual protein had become so small that a minor amino acid phenylthiohydantoin derivative could no longer be detected. Aside from the heterogeneity due to the difference in chain length, we also obtained quantitative evidence for the heterogeneity at positions 7 and 14 of the main chain and at positions 6 and 13 of the minor shorter chain. At each of these positions both alanine and lysine were detected in an Ala/Lys ratio of approx. 4:1.

Most of the peptides produced by digestion with *Staphylococcus aureus* V8 proteinase could be sequenced up to the *C*-terminal residue. Peptide Sa 7 appeared to be the result of incomplete cleavage of the Glu-102–Leu-103 bond. The insoluble peptides Sa 9 and Sa 10 have the same sequence; their difference in chromatographic mobility can be interpreted as the result of a different degree of oxidation at their methionine residues. The heterogeneities at the *N*-terminus and at positions 7 and 14 of the major chain found during the sequence run of the native protein could also be detected during Edman degradation of Sa 1, a peptide without mobility on the thin-layer plate. The *S. aureus* V8 proteinase digest revealed another heterogeneity in the polypeptide chain of histone H2B, at position 72, by the occurrence of two pentapeptides (Sa 5 and Sa 6) with slightly different chromatographic mobilities and identical amino acid sequences except for residue 4, where either serine or alanine was found. Both peptides were obtained in nearly equal amounts.

In order to clearly identify some of the residues in the region 66–68 and to prove the alignment of peptides Sa 10, Sa 5(6) and Sa 3, we carried out a CNBr digest on 1.8 mg of the native protein. The peptide with the highest chromatographic mobility did not stain with either ninhydrin or Coomassie Blue. It was detected after staining with fluorescamine and found to be the tripeptide Ser-57–Met-59 (peptide CB 3). Peptide CB 1 remained at the origin and contained the *N*-terminal peptides Ala-1–Met-56 of the major protein and Pro-1–Met-55 of the minor protein, and again revealed the Ala/Lys heterogeneity at positions 7 and 14 (respectively 6 and 13). Peptide CB 2 starts off at Asn-60 and allowed the identification of Val-66, Phe-67 and Glu-68. At Edman cycle 13 we observed the phenylthiohydantoin derivatives of both serine and alanine in about equal amounts, taking the well-known breakdown of the serine derivative into consideration. This finding definitely proves that one of peptides Sa 5 and Sa 6 does not occur

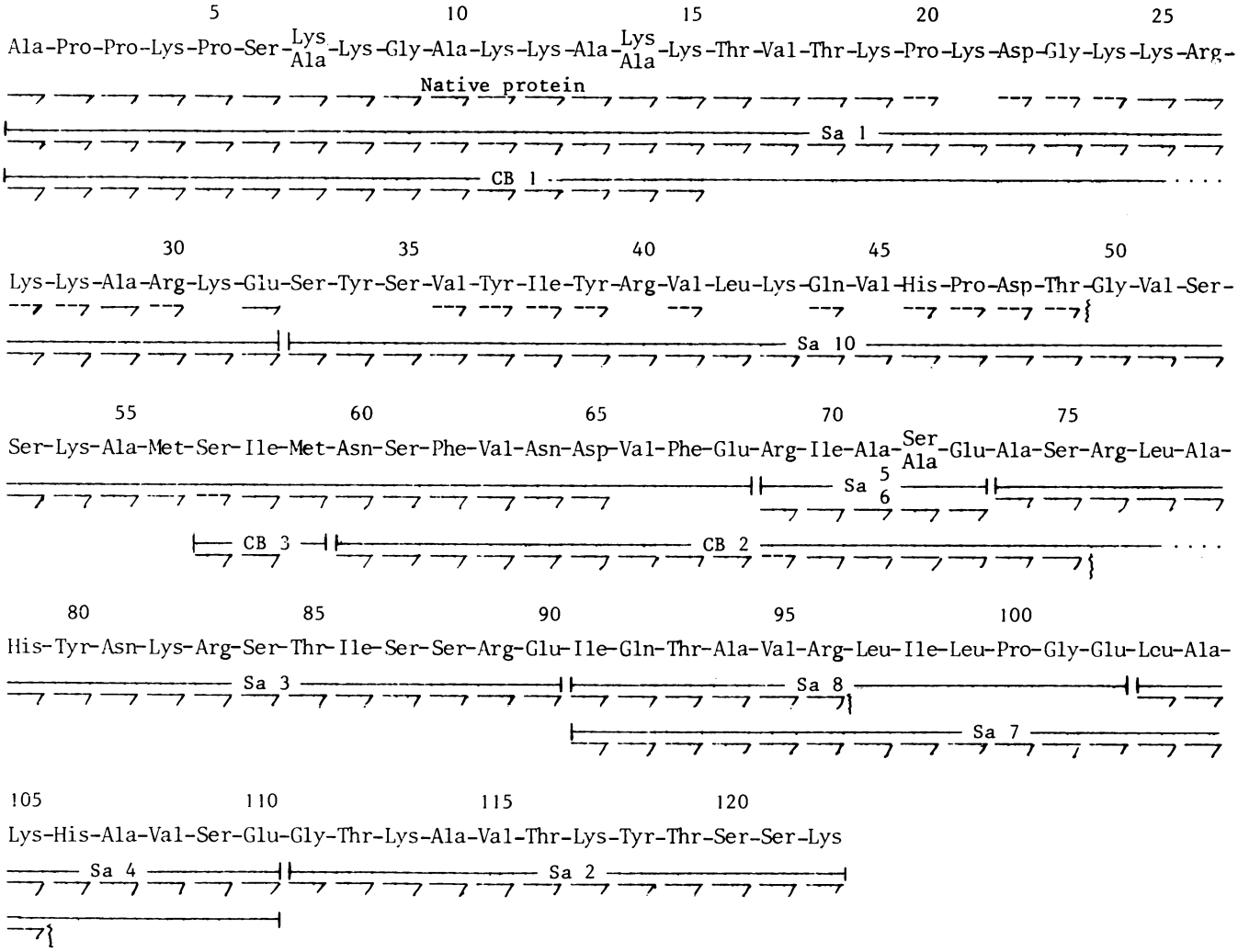


**Fig. 1. Primary structure of histone H2B from *C. elegans***

Automated sequence analysis was carried out on the native protein and on peptides obtained from a digest with *S. aureus* V8 proteinase (Sa) and with CNBr (CB). The peptides are numbered according to their increasing mobility on the preparative thin-layer plates; peptides Sa 1 and CB 1 remained at the origin. Heterogeneity was observed at the positions 7, 14 and 72. Peptide Sa 1 was subdigested with proline-specific enzyme in order to confirm the heterogeneity at the positions 7 and 14 (see the text). Unambiguous identification is shown as ⁻ɔ. A broken arrow indicates that the phenylthiohydantoin signal was weak but that there was no evidence for another residue than the one given. For some peptides the sequence analysis has been stopped before reaching the *C*-terminus ({). A minor fraction of the molecules lacked the *N*-terminal residue alanine (see the text).

at the end of the polypeptide chain of histone H2B and is effectively the result of heterogeneity at position 72.

In order to confirm the presence of both alanine and lysine at positions 7 and 14, we also attempted a subdigest of peptide fraction Sa 1 with proline-specific endopeptidase. This enzyme is known also to produce non-specific cleavage at Ala–Xaa bonds (Yoshimoto *et al.*, 1980). Three main fractions were obtained, of which the first contained the *N*-terminal peptides Ala-1–Pro-20 of the main Sa 1 peptide and Pro-1–Pro-19 of the minor Sa 1 peptide. Heterogeneity of alanine and lysine at positions 7 and 6 respectively could again be demonstrated, with an Ala/Lys ratio of 3:1. The second fraction was a mixture of the peptides Pro-1–Pro-19 and of Ala/Lys-14–Glu-32, resulting from a non-specific cleavage of the peptide bond after Ala-13 (major chain). The most interesting peptide, however, was a hexapeptide of which the first three residues displayed the sequence Lys-Thr-Val. This result can only be explained if position 14 of the major Sa 1 peptide, apart from lysine, is also occupied by alanine.

In an attempt to determine which of the residues at position 7 corresponds to either of the residues at position 14 (major chain) in the same polypeptide chain, we submitted the peptide fraction Sa 1 to thin-layer electrophoresis at pH 3.6 for 30 min. Clearly resolved peptide fractions were not obtained by this procedure; instead, substantial streaking was observed. A peptide fraction with the highest mobility (approx. 7.5 cm from the origin) and a second fraction with lower mobility (6 cm from the origin) were then subjected to automated sequence analysis. For the former peptide we found the

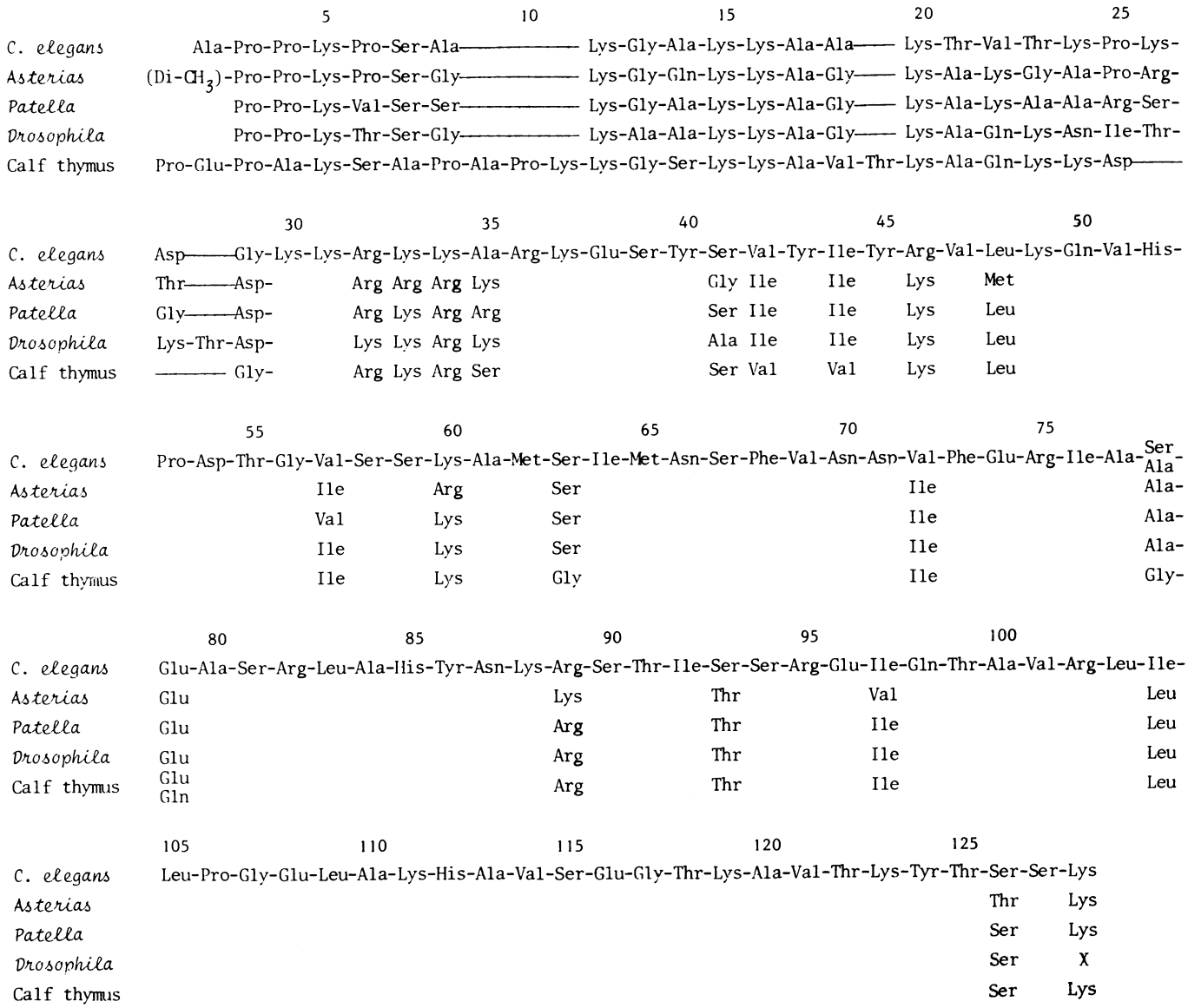


**Fig. 2. Alignment of the sequence of the major histone H2B variant of *C. elegans* with those of histones H2B from two other protostomes (the mollusc *Patella* and the insect *Drosophila*) and two deuterostomes (the starfish *Asterias* and the cow)**

Numbering is according to the cow histone H2B sequence. From position 30 onwards, residues shared by all five species are explicitly given for the nematode sequence only. Sequence data: *Patella granatina* (Van Helden *et al.*, 1979); *Asterias rubens* (Martinage *et al.*, 1985); calf thymus (Iwai *et al.*, 1972); the primary structure of *Drosophila melanogaster* H2B was taken from Isenberg (1979).

combination Ala-7 and Lys-14, whereas for the latter we detected no increase of lysine at either position. In view of the Ala/Lys ratio (4:1) at both positions 7 and 14 found during Edman degradation of the native protein, we conclude that the preparation of histone H2B as obtained by us contained at least five molecular species. Three of these have a polypeptide chain of 122 residues. The most abundant of these three subtypes contains alanine at both positions 7 and 14. The two other species should exhibit the combinations Ala-7 and Lys-14 and Lys-7 and Ala-14, but their relative amounts could not be derived from the sequence data. We have found no evidence for the combination Lys-7 and Lys-14. It cannot be concluded from the experimental findings which of these three polypeptide chains contain serine or alanine at position 72. Besides these three subtypes consisting of 122 residues, our preparation also contained at least two molecular species of 121 residues with *N*-terminal proline instead of alanine and either lysine or alanine at position 6. It is likely that the same heterogeneity exists at position 13, but this could not be definitely proven.

In Fig. 2 we have compared the primary structure of *C. elegans* histone H2B with those determined for three other invertebrate species (a mollusc, an insect and a starfish). There is a remarkable extent of homology even in the *N*-terminal part of the molecule, where the conservative nature is usually less pronounced. With the exception of one insertion in the case of the histone from

*Drosophila*, no further assumptions are required to obtain a very good alignment of these four sequences. More difficulties are encountered when calf thymus histone H2B is included in this comparison. Several insertions and deletions are needed to put regions of putative homology in register. For that reason the latter alignment is somewhat arbitrary, but this merely emphasizes the poor conservation in that region of the molecule.

The strong homology that exists at the *N*-terminal part of the histones H2B from organisms belonging to widely divergent phyla, including protostomes (nematodes, molluscs and insects) and deuterostomes (echinoderms) as well, is rather intriguing, especially when contrasted against the highly different *N*-terminal region of vertebrate (deuterostomes) histones H2B. We presume that the sequences found in the invertebrate phyla mentioned may recall a common ancestral configuration and that the *N*-terminal region in vertebrate histones H2B is a more recent feature. It is known that considerable variation has also occurred in the *N*-terminal part of sea-urchin sperm histone H2B by insertion of a reiterated peptapeptide sequence, a structural feature that is unique to sea-urchin sperm. It is absent from embryonic sea-urchin histone H2B and also from starfish sperm (Strickland *et al.*, 1977*a*,*b*, 1978, 1980; Martinage *et al.*, 1985).

Apart from the *N*-terminal area, a few substitutions occurring in the *C*-terminal two-thirds of the molecule are worth mentioning. First, most substitutions are of a conservative character, e.g. Ile→Val at the positions 42, 44, 57, 72 and 97, Ile→Leu at position 104, Ser→Thr at position 93 and Arg→Lys at positions 32–34. Non-conservative substitutions are seen at the positions 35, 41, 63 and 78. It is tempting to conclude that isoleucine at positions 44 and 72, serine at position 63, leucine at position 104 and lysine at position 60 are the ancestral residues. Alanine at position 78 might also represent the ancestral residue, which has been partly replaced by serine in nematodes. Serine at this position has also been identified in a histone H2B variant present in mouse only (Franklin & Zweidler, 1977); in the other vertebrates glycine is found at this position.

It may be concluded from our work that at least five different histone H2B genes are expressed in *C. elegans*. Sperm-specific variants, if they occur at all, would probably not be detected in this study. Further research will be needed to ascertain whether or not the expression of the subtypes reported here is developmentally regulated and to find whether other subtypes not detected in this study might be expressed at distinct periods of the life-cycle of *C. elegans*.

This work was supported by Grant no. 3.0071.84 from the Fund for Medical Scientific Research awarded to J.R.V. and J.J.V.B. J.R.V. acknowledges a Senior Research Associate Career Award with the National Fund for Scientific Research (Belgium). J.V.B. is indebted to the Fund for Joint Basic Research for providing the gas-phase Sequenator (Grant no. 2.0007.82).

## REFERENCES

Alfageme, C. R., Zweidler, A., Mahowald, A. & Cohen, L. H. (1974) J. Biol. Chem. **249**, 3729–3736
Brandt, W. F., Strickland, W. N., Strickland, M., Carlisle, L., Woods, D. & Von Holt, C. (1979) Eur. J. Biochem. **94**, 1–10
Brenner, S. (1974) Genetics **77**, 71–94
Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. **357**, 873–886
Childs, G., Maxson, R. & Kedes, L. H. (1979) Dev. Biol. **73**, 153–173
Cohen, L. H., Newrock, K. M. & Zweidler, A. (1975) Science **190**, 994–997
Franklin, S. G. & Zweidler, A. (1977) Nature (London) **266**, 273–275
Grunstein, M. (1978) Proc. Natl. Acad. Sci. U.S.A. **75**, 4135–4139
Hieter, P. A., Henricks, M. B., Hemminki, K. & Weinberg, E. S. (1979) Biochemistry **18**, 2707–2716
Houmard, J. & Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 3506–3509
Hunkapiller, M. W. & Hood, L. E. (1983) Methods Enzymol. **91**, 486–493
Isenberg, I. (1979) Annu. Rev. Biochem. **48**, 159–191
Iwai, K., Hayashi, H. & Ishikawa, K. (1972) J. Biochem. (Tokyo) **72**, 357–367
Martinage, A., Briand, G., Van Dorsselaer, A., Turner, C. H. & Sautiere, P. (1985) Eur. J. Biochem. **147**, 351–359
Newrock, K. M., Freedman, N., Alfageme, C. R. & Cohen, L. H. (1982) Dev. Biol. **89**, 248–253
Poccia, D. L. & Hinegardner, R. T. (1975) Dev. Biol. **45**, 81–89
Poccia, D. L., Salik, J. & Krystal, G. (1981) Dev. Biol. **82**, 287–296
Strickland, M., Strickland, W. N., Brandt, W. F. & Van Holt, C. (1977*a*) Eur. J. Biochem. **77**, 263–275
Strickland, W. N., Strickland, M., Brandt, W. F. & Von Holt, C. (1977*b*) Eur. J. Biochem. **77**, 277–286
Strickland, M., Strickland, W. N., Brandt, W. F., Von Holt, C., Wittman-Liebold, B. & Lehmann, A. (1978) Eur. J. Biochem. **89**, 443–452
Strickland, M. S., Strickland, W. N. & Von Holt, C. (1980) Eur. J. Biochem. **106**, 541–548
Thomas, J. O. & Kornberg, R. D. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 2626–2630
Van der Westhuyzen, D. & Von Holt, C. (1971) FEBS Lett. **14**, 333–337
Vanfleteren, J. R. (1982) Comp. Biochem. Physiol. B: Comp. Biochem. **73**, 709–718
Van Helden, P. D., Strickland, W. N., Brandt, W. F. & Von Holt, C. (1979) Eur. J. Biochem. **39**, 71–78
Yoshimoto, T., Walter, R. & Tsuru, D. (1980) J. Biol. Chem. **255**, 4786–4792
Zuckerman, B. M. (1980) Nematodes as Biological Models, vol. 1 (Behavioral and Developmental Models) and vol. 2 (Ageing and Other Model Systems), Academic Press, New York

Received 23 August 1985/10 December 1985; accepted 24 December 1985