The primary structure of histone H2A from the nematode *Caenorhabditis elegans*

Jacques R. VANFLETEREN,* Stefaan M. VAN BUN† and Jozef J. VAN BEEUMEN†

*Laboratorium voor Morfologie en Systematiek der Dieren and †Laboratorium voor Microbiologie en Microbiële Genetica, Rijksuniversiteit Gent, Ledeganckstraat 35, B-9000 Gent, Belgium

The complete primary structure of histone H2A from the nematode *Caenorhabditis elegans* was determined. The amino acid chain consists of 126 amino acid residues and has a blocked *N*-terminus. By comparison with calf thymus histone H2A, the nematode protein shows five deletions, two insertions and 16 substitutions. Most of the changes occur in the *N*- and *C*-terminal regions of the molecule, whereas the central part covering the residues 21–120 is quite well conserved. The lysine residues 5, 8 and 10 were found to be partially acetylated.

INTRODUCTION

Histone H2A has diverged the most of the core histones during evolution. From the sequences that have been established it is clear that this variability is mainly restricted to short *N*- and *C*-terminal regions and that the large central part of the molecule is fairly well conserved. The latter part is involved in inter-histone interactions in the nucleosome, whereas the more hydrophilic *N*- and *C*-terminal regions are thought to interact with DNA (Isenberg, 1979).

Over the last decade it has also been found that histone H2A actually may be composed of quite a number of isoprotein species or variants. Mammalian cells may contain up to six isoforms. Any of these can be covalently linked to ubiquitin (Grove & Zweidler, 1984; West & Bonner, 1980; Wu *et al.*, 1986). Microheterogeneity of histone H2A was also demonstrated in several invertebrate species, particularly in echinoderms (Brandt *et al.*, 1980) and in plants (Rodrigues *et al.*, 1985). The H2A isoforms may show a cell-cycle-dependent (Wu & Bonner, 1981; Wu *et al.*, 1982) or stage-specific expression (Brandt *et al.*, 1980; Newrock *et al.*, 1982).

We have studied the primary structure of histone H2A extracted from the free-living nematode *Caenorhabditis elegans*. Nematodes constitute one of the oldest invertebrate phyla, and we were interested to learn more about the sequence diversity and the extent of microheterogeneity in nematode histone H2A. In the present paper we present the complete amino acid sequence of histone H2A from *C. elegans*. No evidence was found for the existence of histone H2A isoforms in this worm.

EXPERIMENTAL

Preparation of histone H2A

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 procedures for nematode growth and histone preparation have been described previously (Vanfleteren *et al.*, 1986). The slightly lysine-rich histones H2A and H2B were separated from the arginine-rich histones H3

and H4 by size-exclusion chromatography in acetate/bisulphite buffer, pH 5.6, as recommended by Van der Westhuyzen & Von Holt (1971). Histone H2A was then finally purified in a preparative acetic acid/urea/Triton X-100/polyacrylamide gel, prepared and run as described by Bonner *et al.* (1980) except for a lower concentration of urea (6 M) in the separating gel. Protein bands were detected by short staining with Coomassie Blue R-250. The slowest-moving band, which contained histone H2A, was excised and cut into small pieces, and the protein was eluted with 70% (v/v) formic acid for 24 h. Triton X-100 and Coomassie dye were extracted from the dried (Savant Speed Vac Concentrator) sample with ethanol/methanol mixture (1:1, v/v) containing 5 mM-H₂SO₄ final concentration. The protein pellet was then washed twice with the ethanol/methanol mixture to remove the acid and dried. The final yield of histone H2A was approx. 100 nmol. It was purified from 1.3 kg wet wt. of nematode tissue and about 325 mg of crude histone protein.

Production and separation of peptides

About 80 nmol of histone H2A was dissolved in 1 ml of 2% (v/v) formic acid and heated at 106 °C for 4 h (Inglis, 1983). Solid trichloroacetic acid was added at a final concentration of 20% (w/v) to precipitate medium- and large-sized peptides. The supernatant (10000 g for 5 min) was dried in the Savant Speed Vac Concentrator connected to two refrigerated (–196 °C) traps and a high-vacuum pump. This fraction contained mainly a dipeptide which was further purified by t.l.c. on high-performance t.l.c. cellulose developed in pyridine/butan-1-ol/acetic acid/water (10:15:3:12, by vol).

The trichloroacetic acid-insoluble pellet was washed once with 20% (w/v) trichloroacetic acid and dried. The peptides were dissolved in 0.3% (v/v) trifluoroacetic acid and separated by reversed-phase h.p.l.c. on a Radial-Pak C₁₈ column (Waters) with a linear gradient of 0–60% (v/v) acetonitrile in 0.3% (v/v) trifluoroacetic acid. A small amount of the major peptide fractions was subjected to a few cycles of Edman degradation in the Sequenator. Putative *N*-terminal fragments of the molecule, yielding no sequence information, were pooled and subdigested with endoproteinase Arg-C (from mouse submaxillary gland; Boehringer, Mannheim,

Fig. 1. Primary structure of histone H2A from *C. elegans*

Automated sequence analysis was carried out on peptides obtained from mild acid hydrolysis (AH) and from digests with endoproteinase Arg-C (Arg-C) and *Staphylococcus aureus* V8 proteinase (Sa). Medium- and large-sized peptides produced by the acid hydrolysis were precipitated with trichloroacetic acid. AH(sup) refers to the supernatant portion. The pelleted peptides are numbered according to their elution during subsequent h.p.l.c. separation (e.g. AH 3). N-Terminal fragments were subdigested with Arg-C proteinase. The soluble peptides were likewise separated and numbered; the precipitated material was called Arg-C core. The peptides originating from the *S. aureus* V8 proteinase digest of native histone H2A are numbered according to their increasing mobility on thin-layer plates; peptide Sa 1 remained at the origin. For some peptides the sequence analysis has been stopped before reaching the C-terminus (—|—). Sequence positions that differ from calf thymus histone H2A (Sautière *et al.*, 1974) are indicated in script below the amino acid chain. Ac, acetylation; A, insertion; —, deletion.

Germany) in 50 mM-NH₄HCO₃ for 3 h at 37 °C and at an enzyme/substrate ratio of 1:40 (w/w). When the digestion was completed, 2 vol. of 0.1% (v/v) trifluoroacetic acid was added and the sample was centrifuged in an Eppendorf centrifuge (10000 *g* for 5 min). A small pellet was retained and is referred to below as 'Arg-C core'. The supernatant was loaded on a μ Bondapak C₁₈ reversed-phase column (4 mm $\times$ 150 mm) (Waters) and eluted with a linear gradient of acetonitrile in 0.1% (v/v) trifluoroacetic acid.

A separate sample containing approx. 20 nmol of protein was dissolved in ammonium acetate buffer, pH 4, and digested with *Staphylococcus aureus* V8 proteinase (EC 3.4.21.19) (Miles Laboratories, Slough, Berks., U.K.) overnight at 37 °C and at an enzyme/substrate ratio of 1:60 (w/v). A second amount of fresh enzyme was added and digestion was continued for another 20 h. The resulting peptides were separated by t.l.c. on

high-performance t.l.c. cellulose on aluminium sheets (Merck, Darmstadt, Germany) developed in pyridine/butan-1-ol/acetic acid/water (10:15:3:12, by vol.). Two side lanes and a central lane were cut with a guillotine and stained with Coomassie Blue R-250, exactly as if they were polyacrylamide gels, to reveal the position of large peptides. Intact peptide material was then eluted with 50% (v/v) acetic acid and dried.

Automated sequence analysis

Automated Edman degradation chemistry was carried out by using the Applied Biosystems 470 gas-phase Sequenator. The amino acid phenylthiohydantoin derivatives were analysed by reversed-phase h.p.l.c. on a 5 μ m-particle-size IBM cyanopropyl column (4.6 mm $\times$ 250 mm) essentially as described by Hunkapiller & Hood (1983) but with 3.75% (v/v) tetrahydrofuran included in the acetate buffer in order to

	10	20	30	40	50
Calf	Ac - SGRGKQGGA	RA--KAKTRSSR	AGLQFPVGRV	HRLLRKGNYA	ERVGAGAPVY
Trout	Ac - SGRGKTGGKA	RA--KAKTRSSR	AGLQFPVGRV	HRLLRKGNYA	ERVGAGAPVY
<i>C. elegans</i>	(Ac?) - SGRGK-GGKA	KTGG KAKSRSSR	AGLQFPVGR L	H R ILRKGNYA	Q RVGAGAPVY
<i>Drosophila</i>	? - SGRGK-GGK V	KG --KAKSR S NR	AGLQFPVGR I	HRLLRKGNYA	ERVGAGAPVY
Starfish	Ac - SGRGK-GGKA	RA--KAKSR S A	AGLQFPVGRV	H R FLRKGNYA	E RVGAGAPVY
Sea-urchin	Ac - SGRGK-G A KA	KG --KAKSRSSR	AGLQFPVGRV	H R FLRKGNYA	N RVGAGAPVY
Cuttlefish	Ac - SGRGK-GGK V	KG --K S KTRSSR	AGLQFPVGR I	HRLLRKGNYA	Q RVGAGAPVY
<i>Sipunculus</i>	Ac - SGRGK-GGKA	KG --K S KSRSSR	AGLQFPVGR I	HRLLRKGNYA	ER I GAGAPVY
Wheat	NH ₂ - A GR-K-GG D -	R K --KAKTR S V K	AGLQFPVGR I	G RYL K KGRYA	Q RVG S GAPVY
	60	70	80	90	100
Calf	LAAVLEYLTA	EILELAGNAA	RDNKKTRIIP	RHLQLAIRND	EELNKL L GKV
Trout	LAAVLE-LTA	EILELAGNAA	RDNKKTR-I P	RHLQLA V RND	EELNKL L G G V
<i>C. elegans</i>	LAAVLEYL A A	E V L ELAGNAA	RDNKKTRI A P	RHLQLA V RND	EELNKL L A G V
<i>Drosophila</i>	LAAV M EY LTA	E V L ELAGNAA	RDNKKTRIIP	RHLQLAIRND	EELNKL L S G V
Starfish	L AAV M EY L A A	EILELAGNAA	RDNKKTRIN P	RHLQLAIRND	EELNKL L S G V
Sea-urchin	LAAVLEYL A A	EILELAGNAA	RDNKKTRIIP	RHLQLAIRND	EELNKL L G G V
Cuttlefish	LAAV M EY L A A	E V L ELAGNAA	RDNKK S RIIP	RHLQLAIRND	EELNKL L S G V
<i>Sipunculus</i>	LAAV M EY L A A	E V L ELAGNAA	RDNKKTRIIP	RHLQLAIRND	EELNKL L S G V
Wheat	LAAVLEYL A A	E V L ELAGNAA	K DNKKTRI I V P	RHL L LA I V R ND	Q ELG R LL A G V
	110	120			
Calf	TIAQGGVLPN	IQAVLLPKKT	ESHKKAK-GK		
Trout	TIAQGGVLPN	IQAVLLPKKT	E---KAK V AK		
<i>C. elegans</i>	TIAQGGVLPN	IQAVLLPKKT	GGDKE		
<i>Drosophila</i>	TIAQGGVLPN	IQAVLLPKKT	EK		
Starfish	TIAQGGVLPN	IQAVLLPKKT	AKAAK		
Sea-urchin	TIAQGGVLPN	IQAVLLPKKT	GSKSSK		
Cuttlefish	TIAQGGVLPN	IQAVLLPKKT	QKAAK		
<i>Sipunculus</i>	TIAQGGVLPN	IQAVLLPKKT	QKSK		
Wheat	TIA H GGV I PN	I N S V LLPK S	PAAAEKEAKS P KKKT S T A A K S P K K T A V A A T K E		

Fig. 2. Comparison of the sequences of histone H2A from two vertebrate species, one plant and species belonging to several invertebrate phyla

The sequences are aligned and numbered up to residue 120 of the calf histone H2A sequence. Sequence data: calf thymus histone H2A (Sautière *et al.*, 1974); rainbow-trout testis histone H2A (Conner *et al.*, 1984); *Drosophila* histone H2A (Wells, 1986); starfish (*Asterias rubens*) gonad histone H2A (Martinage *et al.*, 1983); sea-urchin (*Parechinus angulosus*) sperm histone H2A (Strickland *et al.*, 1980); cuttlefish (*Sepia officinalis*) testis histone H2A (Wouters-Tyrou *et al.*, 1982); *Sipunculus nudus* erythrocyte histone H2A (Kmieciak *et al.*, 1983); wheat-germ histone H2A₁ (Rodrigues *et al.*, 1985).

differentiate diphenylthiourea from the phenylhydantoin derivative of methionine.

RESULTS AND DISCUSSION

The complete amino acid sequence of the protein and the alignment of the peptides that have been used for elucidation of this sequence are given in Fig. 1. The alignment of the peptides is based on the highly conservative nature of the protein and strengthened by peptide overlaps covering two-thirds of the molecule.

Sequencing of native histone H2A showed that the protein was blocked at the *N*-terminal. Prolonged digestion of a separate sample of histone H2A with *S. aureus* V8 proteinase yielded a partially unblocked *N*-terminal peptide, which enabled the determination of the *N*-terminal sequence 1–5. The ability of *S. aureus* V8 proteinase preparations to unblock *N*-terminal acetylserine has been reported previously (Martinage *et al.*, 1983). Therefore, and since α -*N*-acetylserine is the most common *N*-terminal residue of histone H2A (Isenberg, 1979; Wu *et al.*, 1986), we assume that nematode histone H2A is most probably blocked by acetylation.

The amino acid sequence of *C. elegans* histone H2A shows in its *N*-terminal part 1–20 one deletion, two insertions and three substitutions with respect to calf thymus histone H2A. The *C*-terminal region of the protein (residues 121–126) is so variable that any homology with the calf histone in this region no longer exists. Considerable variability is common in this region among histone H2A sequences, but the net charge essentially remains positive except in *Drosophila* histone H2A (Fig. 2). In the nematode histone H2A also the overall charge density is negative in this region, rising to zero when the sequence 119–126 is taken into consideration. This may severely weaken, if not preclude, protein–DNA interaction at the *C*-terminal region of the molecule. The central part is very well conserved, with only nine substitutions with respect to calf thymus histone H2A. Five of these are of a conservative nature, and four are non-conservative (Glu-42→Gln; Thr-60→Ala; Gly-99→Ala; Lys-100→Gly).

The deletion of the glutamine residue present at position 6 of the calf histone H2A sequence appears to be characteristic for invertebrate species. Deletion of two histidine residues near the *C*-terminus (positions 123–124

of the calf histone) has been observed in invertebrates and lower vertebrates (Fig. 2) (Wells, 1986).

The replacement of leucine by methionine at position 51 of the calf thymus histone differentiates two isoforms of histone H2A in mammals (Blankstein *et al.*, 1977). Leucine is also found at this position in sea-urchin sperm histone H2A, but is replaced by methionine in embryonic tissue. We found no evidence for the occurrence of methionine in the corresponding position in *C. elegans* histone H2A, nor did we find firm evidence of any other heterogeneities.

Two-dimensional electrophoresis in acetic acid/urea/Triton X-100/polyacrylamide and acetic acid/urea/cetyltrimethylammonium bromide/polyacrylamide gels showed that histone H2A prepared from non-synchronized nematode populations is not extensively secondarily modified (results not shown). We found that the lysine residues 5, 8 and 10 were acetylated to the extent of some 25%, 10% and 15% respectively.

Very recently, several histone genes of *C. elegans* have been cloned and sequenced (S. Roberts, M. Sanicola, S. W. Emmons & G. Childs, personal communication). The amino acid sequence predicted from a cloned histone H2A gene and the one proposed here match each other completely.

We thank Dr. S. Roberts and her colleagues at the Department of Molecular Biology, Albert Einstein College of Medicine, New York, NY, U.S.A., for providing the amino acid sequence as predicted from the nucleotide sequence of a cloned histone H2A gene before its publication. We are also grateful to Dr. P. Sautière (Lille, France) for providing the phenylthiohydantoin derivatives of the most common modified amino acids. 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.

REFERENCES

- Blankstein, L. A., Stoller, B. D., Franklin, S. G., Zweidler, A. & Levy, S. B. (1977) *Biochemistry* **16**, 4557–4562
- Bonner, W. M., West, M. H. P. & Stedman, J. D. (1980) *Eur. J. Biochem.* **109**, 17–23
- Brandt, W. F., Strickland, W. N., Strickland, M., Carlisle, L., Woods, D. & Von Holt, C. (1980) *Eur. J. Biochem.* **94**, 1–10
- Conner, W., Mezquita, J., Winkfein, R. J., States, J. C. & Dixon, G. H. (1984) *J. Mol. Evol.* **30**, 227–235
- Grove, G. W. & Zweidler, A. (1984) *Biochemistry* **23**, 4436–4443
- Hunkapiller, M. W. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 486–493
- Inglis, A. S. (1983) *Methods Enzymol.* **91**, 324–332
- Isenberg, I. (1979) *Annu. Rev. Biochem.* **48**, 159–191
- Kmiecik, D., Couppez, M., Belaiche, D. & Sautière, P. (1983) *Eur. J. Biochem.* **135**, 113–121
- Martinage, A., Belaiche, D., Dupressoir, T. & Sautière, P. (1983) *Eur. J. Biochem.* **130**, 465–474
- Newrock, K. M., Freedman, N., Alfageme, C. R. & Cohen, L. H. (1982) *Dev. Biol.* **89**, 248–253
- Rodrigues, J. de A., Brandt, W. F. & Von Holt, C. (1985) *Eur. J. Biochem.* **150**, 499–596
- Sautière, P., Tyrrou, D., Laine, B., Mizon, J., Ruffin, P. & Biserte, G. (1974) *Eur. J. Biochem.* **41**, 563–576
- Strickland, W. N., Strickland, M. S., De Groot, P. C. & Von Holt, C. (1980) *Eur. J. Biochem.* **109**, 151–158
- Van der Westhuyzen, D. & Von Holt, C. (1971) *FEBS Lett.* **14**, 333–337
- Vanfleteren, J. R., Van Bun, S. M., Delcambe, L. L. & Van Beeumen, J. J. (1986) *Biochem. J.* **235**, 769–773
- Wells, D. E. (1986) *Nucleic Acids Res.* **14**, r119–r149
- West, M. H. P. & Bonner, W. M. (1980) *Biochemistry* **19**, 3238–3245
- Wouters-Tyrrou, D., Martin-Ponthieu, A., Briand, G., Sautière, P. & Biserte, G. (1982) *Eur. J. Biochem.* **124**, 489–498
- Wu, R. S. & Bonner, W. M. (1981) *Cell (Cambridge, Mass.)* **27**, 321–330
- Wu, R. S., Tsai, S. & Bonner, W. N. (1982) *Cell (Cambridge, Mass.)* **31**, 367–374
- Wu, R. S., Panusz, H. T., Hatch, C. L. & Bonner, W. M. (1986) *CRC Crit. Rev. Biochem.* **20**, 201–263

Received 20 November 1986/16 January 1987; accepted 5 February 1987