Biochem. J. (1991) 277, 39–45 (Printed in Great Britain)

Processing of the precursor of protamine P2 in mouse

Peptide mapping and *N*-terminal sequence analysis of intermediates

Danièle CARRÉ-EUSÈBE,*§ Florence LEDERER,† Kim Ho Diêp LÊ† and Susan M. ELSEVIER‡
* Laboratoire de Biochimie du Développement, Institut Jacques Monod et Université Paris VII, 2 Place Jussieu, 75251 Paris Cedex 05, France, † Centre National de la Recherche Scientifique, UA 122, Hôpital Necker, 161 Rue de Sèvres, 75730 Paris Cedex 15, France, and ‡ Laboratoire de Biologie Moléculaire de l'Ecole Pratique des Hautes Etudes, 3ème Section, Institut Jacques Monod, 2 Place Jussieu, 75251 Paris Cedex 05, France

Protamine P2, the major basic chromosomal protein of mouse spermatozoa, is synthesized as a precursor almost twice as long as the mature protein, its extra length arising from an *N*-terminal extension of 44 amino acid residues. This precursor is integrated into chromatin of spermatids, and the extension is processed during chromatin condensation in the haploid cells. We have studied processing in the mouse and have identified two intermediates generated by proteolytic cleavage of the precursor. H.p.l.c. separated protamine P2 from four other spermatid proteins, including the precursor and three proteins known to possess physiological characteristics expected of processing intermediates. Peptide mapping indicated that all of these proteins were structurally similar. Two major proteins were further purified by PAGE, transferred to poly(vinylidene difluoride) membranes and submitted to automated *N*-terminal sequence analysis. Both sequences were found within the deduced sequence of the precursor extension. The *N*-terminus of the larger intermediate, PP2C, was Gly-12, whereas the *N*-terminus of the smaller, PP2D, was His-21. Both processing sites involved a peptide bond in which the carbonyl function was contributed by an acidic amino acid.

INTRODUCTION

Protamines, the basic chromosomal proteins of mammalian spermatozoa, are synthesized and integrated into chromatin during spermatid differentiation and play an essential role in the condensation of DNA into its compact transcriptionally inactive form. In mouse, two protamines are present, P1 and P2. Both proteins are small, highly basic and rich in cysteine residues [1–4], and they are encoded by closely linked single-copy genes [5]. Protamine P1, the less abundant form in mouse, is found in most other mammals [6–14], and its synthesis follows a conventional route [15,16]. Protamine P2, in contrast, has been identified in only a few other mammals, including man [12,17–19], and it is synthesized as a precursor [16,20,21]. The mouse precursor of protamine P2 is nearly twice as long as the mature protein, its extra length arising from an *N*-terminal extension of 44 amino acid residues (methionine included). It is integrated into spermatid chromatin, and the extension is processed during chromatin condensation.

The mechanism of processing of the precursor of protamine P2 is not yet understood. In previous work [22], we have described six basic proteins from mouse spermatids, proteins A–F, which show physiological characteristics predicted for the precursor of protamine P2 and intermediates generated by processing. Proteins A–F were soluble at the same concentration of trichloroacetic acid as protamine P2, and were insoluble, like protamine P2, in the presence of SDS. Translation *in vitro* of testis mRNAs confirmed that protein A represented the precursor of protamine P2 [16,20] and suggested that the other proteins might represent processing intermediates. Pulse labelling *in vivo* indicated that a precursor–product relationship existed between the proteins and protamine P2.

In the present paper we show by peptide mapping that structural similarities exist between mouse proteins A–D and protamine P2. *N*-Terminal sequence analysis has further indicated that two of the proteins arise by proteolytic cleavage of the *N*-terminal extension of the precursor.

MATERIALS AND METHODS

Protein extraction from testis or epididymis

Epididymides and testes were obtained from mice of inbred strain C57BL/6 (Iffa Credo) and were stored at −70 °C. Acid-soluble proteins were extracted from sonication-resistant nuclei prepared from these organs. Throughout the extractions, buffers contained 2 mM-phenylmethanesulphonyl fluoride, glassware was siliconized, centrifugation was at 11200 *g* for 10 min and all operations were carried out at 0–4 °C, unless otherwise noted.

Epididymides or testes from five mice were homogenized by hand in 5 ml of 75 mM-NaCl/24 mM-EDTA, pH 8.3. Tissue fragments were allowed to settle, and cells in the supernatant were collected by centrifugation. They were resuspended in 5 ml of 10 mM-Tris/HCl buffer, pH 8.0, and submitted to sonication in four bursts of 20 s each with a Prosciences sonicator at dial setting 2.2. Sonication-resistant nuclei were collected by centrifugation and, in the case of epididymides, further purified by passage through a 10 ml gradient of 20–55% (w/v) sucrose in 10 mM-Tris/HCl buffer, pH 8.0, centrifuged at 2600 *g* for 15 min in a swinging-bucket rotor. The nuclear pellet was resuspended in 0.5–1.0 ml of 1.0% (v/v) Triton X-100 in 10 mM-Tris/HCl buffer, pH 8.0, incubated for 15 min, centrifuged and washed with 4 ml of 10 mM-Tris/HCl buffer, pH 8.0. The resulting chromatin was dissociated in 0.5–1.0 ml of 150 mM-Tris/HCl buffer, pH 8.0, containing 1.1 M-NaCl, 6.0 M-urea and 0.1 M-2-mercaptoethanol and incubated at 37 °C for 60 min. This suspension was adjusted to pH 1.0 by addition of 4 M-HCl to approx. 0.2 M-HCl. It was incubated for 15 min, and the precipitate was removed by centrifugation. The supernatant was adjusted to 1% trichloroacetic acid by addition of 5% (w/v)

Abbreviations used: PP2A, protein A; PP2C, protein C; PP2D, protein D.

§ To whom correspondence should be addressed.

trichloroacetic acid, then incubated for 18 h, and any further precipitate was also removed by centrifugation. Finally, the supernatant was adjusted to 20% trichloroacetic acid by addition of 60% (w/v) trichloroacetic acid and incubated for 18 h. The 1–20% trichloroacetic acid precipitate was collected by centrifugation and the proteins it contained were solubilized with serial 100 μl portions of 0.2 M-2-mercaptoethanol. Each successive portion was agitated for 2 min, collected by centrifugation and analysed by electrophoresis on 10% (w/v) polyacrylamide gels. Eluates containing proteins were pooled and frozen at −20 °C. Protein concentrations were determined as described previously [23], with salmine sulphate (Mann) as reference standard. Approx. 50 μg of proteins was extracted from each epididymis, and 20 μg from each testis.

Carboxyamidomethylation

Cysteine residues in the extracted proteins were blocked and labelled by reaction with iodo[^{14}C]acetamide [24]. Proteins were incubated at 10 mg/ml in 0.5 M-Tris/HCl buffer, pH 8.5, containing 6 M-guanidinium chloride and 7 mM-dithiothreitol for 90 min at 37 °C. Iodo[^{14}C]acetamide (1.96 GBq/mmol; Amersham) was then added at a ratio of 1.85 MBq/mg of protein, and the reaction was allowed to proceed for 45 min at 0 °C in the dark. Unlabelled 0.2 M-iodoacetamide was then added to a final concentration of 23 mM, and the reaction was continued for 90 min. A final addition of 2 mM-dithiothreitol/7 mM-iodoacetamide was made, and the reaction was stopped after 30 min. Proteins were purified by filtration through Sephadex G-15 equilibrated and eluted with 20 mM-ammonium bicarbonate buffer, pH 8.5. Proteins for use as carrier were similarly carboxyamidomethylated, with no added iodo[^{14}C]acetamide.

H.p.l.c.

Gel-permeation chromatography was carried out in a Beckman model 332 h.p.l.c. apparatus with protein samples dissolved in 20 μl of 0.2% trifluoroacetic acid (Sequanal grade; Pierce Chemical Co.) at 0.3–8 mg/ml. Each sample was injected on to a Beckman TSK/SW 2000 analytical column (7.5 mm × 300 mm) equilibrated with 0.2% trifluoroacetic acid and eluted (0.25 ml/min) with the same solvent. Absorbance at 214 nm was recorded with an on-line detector, and 0.1 ml fractions were collected. When required, selected fractions were pooled, vacuum-dried and rechromatographed with the use of the same procedure.

Endoproteinase digestion

Hydrolysis by endoproteinase Lys-C (EC 3.4.21.50) was carried out on 11 μg of protein, with 0.8 μg of the enzyme (30 units/mg of protein; Boehringer) in 15 μl of 20 mM-ammonium bicarbonate buffer, pH 8.5. Digestion was for 165 min at 20 °C. Digests were then freeze-dried, dissolved in 10 μl of 10 mM-HCl and vacuum-dried. Treatment of selected samples with alkaline phosphatase (EC 3.1.3.1) before endoproteinase digestion was performed with 10 μg of protein and 2.5 units of the enzyme (RNAase-free; Boehringer) in 12.5 μl of 0.1 M-ammonium bicarbonate buffer, pH 8.5. Incubation was for 1 h at 45 °C.

Digestion by endoproteinase Glu-C (*Staphylococcus aureus* V8 proteinase, EC 3.4.21.19) was carried out on 12 μg of protein with 0.33 μg of the enzyme (20 units/mg of protein; Boehringer) in 15 μl of 100 mM-ammonium bicarbonate buffer, pH 8.5, for 5 or 105 min at 37 °C.

PAGE

Proteins were analysed by electrophoresis on 10% (w/v) polyacrylamide gels in vertical slabs (14 cm × 10 cm × 1 mm) containing 0.9 M-acetic acid/2.5 M-urea [23]. The electrolyte was 0.9 M-acetic acid, and pre-electrophoresis was for 90 min at 15 V/cm. Before application to the gel, samples were dried, dissolved in 5 μl of 0.9 M-acetic acid/8.0 M-urea/0.5 M-2-mercaptoethanol/0.001% Pyronine G and incubated for 1 h at 37 °C. Migration markers were calf thymus histones (Sigma Chemical Co.), salmine sulphate (Mann) and polymyxin B sulphate (Sigma Chemical Co.). After electrophoresis for 90 min at 15 V/cm, proteins were stained for 1 h with 0.2% Amido Black in acetic acid/methanol/water (7:40:53, by vol.). For fluorography, destained gels were incubated for 1 h in acetic acid/methanol/water (1:3:6, by vol.), for 30 min in En^{3}Hance (NEN) and for 5 min in cold distilled water. Dried gels were exposed to Kodak or Fuji X-ray film at −70 °C.

Peptides were analysed by electrophoresis on 15% polyacrylamide gels with 0.9 M-acetic acid as electrolyte [25]. The separating gel contained 15% (w/v) acrylamide/0.5% (w/v) bisacrylamide/6.25 M-urea/0.9 M-acetic acid and was subjected to pre-electrophoresis for 90 min at 32 V/cm. A 5 mm stacking gel of 3% (w/v) acrylamide/0.6% (w/v) bisacrylamide/1.25 M-urea/0.9 M-acetic acid was added, and pre-electrophoresis was continued for 7 min at 24 V/cm. Samples were then applied, and electrophoresis was performed for 90 min at 24 V/cm.

N-Terminal sequence analysis

Proteins from testis were submitted to electrophoresis on 10% polyacrylamide gels as described above, in a single 12.5 cm-wide sample lane. For purification of protein C, 80 μg of protein from testis was applied to the gel, and for protein D, 300 μg was applied. To avoid protein oxidation and blocking of *N*-terminal functions, cyanate-free urea (BRL) was used in the gel, and pre-electrophoresis (1.5 h) and electrophoresis (4 h) were performed at 4 °C with 0.9 M-acetic acid (h.p.l.c. grade) in degassed water containing 0.1 mM-sodium thioglycollate [26]. The concentration of Pyronine G in the sample solution was lowered to 0.0001%.

Transfer of proteins to poly(vinylidene difluoride) membranes (Immobilon; 0.45 μm pore size; Millipore) was performed after electrophoresis with a Trans-Blot apparatus (Bio-Rad Laboratories) [27]. The membranes were equilibrated successively in methanol, distilled water and 0.7% (v/v) acetic acid. Equilibration of the gel in 0.7% acetic acid was found to be unnecessary. Transfer was then carried out in 0.7% acetic acid at 4 °C with three membranes. Proteins migrated toward the cathode for 40 min at 170 mA. The membranes were then washed with distilled water, stained with 0.1% Amido Black in methanol/water (1:1, v/v), destained briefly in methanol/acetic acid/water (5:1:4, by vol.) and air-dried. Filters were finally wetted with distilled water, and a strip corresponding to the protein of interest was cut out.

Sequence analysis was performed in an Applied Biosystems 470A gas-phase protein sequencer, its sample cartridge loaded to capacity with the protein-charged membrane. Amino acid phenylthiohydantoin derivatives were identified by using an on-line 120 A analyser. The [^{14}C]carboxyamidomethylcysteine derivative, which was eluted with the glutamate derivative, was identified by its radioactive label in a liquid-scintillation spectrometer. Protamine P2 and protein C, purified by h.p.l.c., were submitted directly to analysis in the same sequencer.

RESULTS

Fractionation of testicular and epididymal proteins by gel-permeation h.p.l.c.

Proteins were extracted from sonication-resistant nuclei prepared from either testis or epididymis. Sonication left only the

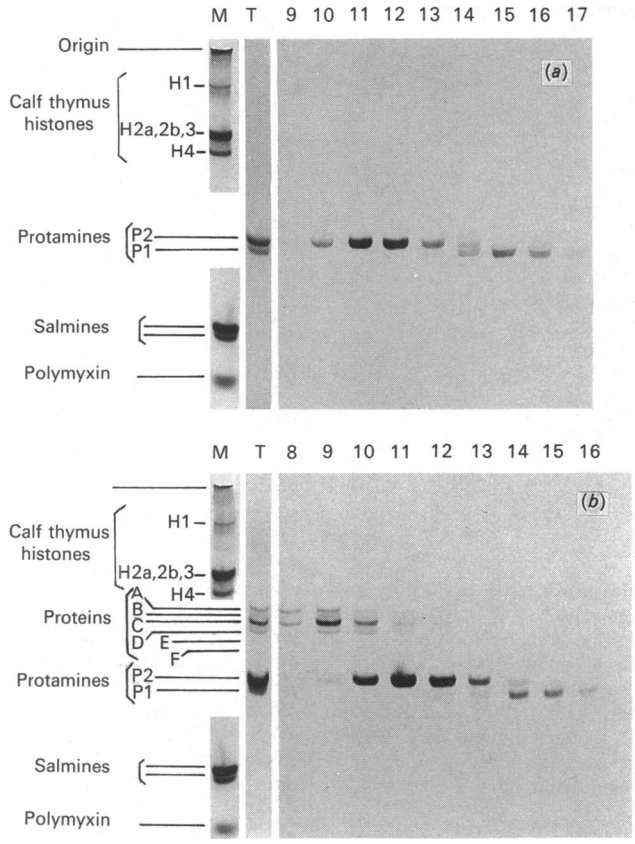

Fig. 1. Electrophoretic analysis of proteins fractionated by h.p.l.c.

[^{14}C]Carboxyamidomethylated proteins (80 μg) from epididymis (*a*) or testis (*b*) were fractionated by gel-permeation chromatography on a Beckman TSK/SW 2000 column with 0.2% trifluoroacetic acid as elution solution. Fractions were analysed by electrophoresis on 10% polyacrylamide gels. Migration markers (4–5 μg), in lane M, were stained with Amido Black. On each fluorograph, the relevant unfractionated proteins are shown in lane T. Following lanes show successive h.p.l.c. fractions, each fraction representing a total volume of 0.1 ml. A 3 μl portion was mixed with unfractionated carrier (4 μg) before application to the gel. Scanning at 600 nm [28] of similar gels stained with Amido Black revealed that unfractionated proteins from testis contained approx. 45% protamine P2, 20% protamine P1, 11% protein C, 8% protein A, 4% protein D, 2% protein B and 10% others.

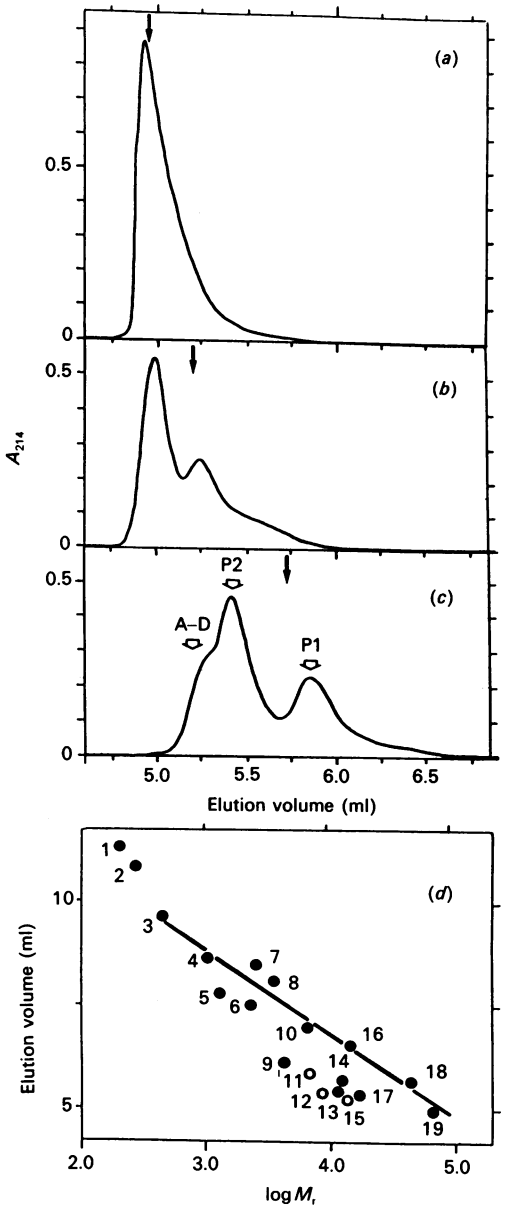

Fig. 2. Fractionation of testicular proteins by gel-permeation h.p.l.c.

Testicular proteins were chromatographed on a Beckman TSK/SW 2000 column with trifluoroacetic acid at 0.05% (*a*), 0.1% (*b*) or 0.2% (*c*) as the elution solution. On each profile, the elution volume of cytochrome *c* is indicated by an arrow. Positions of protamine P1, protamine P2 and proteins A–D are indicated in (*c*). In (*d*), elution volumes in 0.2% trifluoroacetic acid were plotted versus log M_r for reference proteins (●) and three proteins from testis (○). Amino acids, peptides and proteins and their M_r values were: 1, tryptophan, 204; 2, glycylphenylalanine, 293; 3, glycylhistidylarginylproline, 465; 4, bradykinin, 1060; 5, angiotensin-I, 1296; 6, vasointestinal peptide fragment '10–28', 2339; 7, insulin A-chain (oxidized), 2530; 8, insulin B-chain (oxidized), 3496; 9, salmine, 4250; 10, aprotinin, 6511; 11, mouse protamine P1, 6825; 12, mouse protamine P2, 8457; 13, calf histone H4, 11282; 14, cytochrome *c*, 12384; 15, mouse PP2A, 13371; 16, egg-white lysozyme, 14300; 17, myoglobin, 16900; 18, ovalbumin, 42700; 19, BSA, 66000.

highly condensed nuclei of spermatozoa in epididymal homogenates and condensing nuclei of spermatids plus nuclei of spermatozoa in testis homogenates [28,29]. Acid-soluble proteins were then extracted from chromatin of these nuclei and analysed by electrophoresis. As shown in Fig. 1(*a*) (lane T), extracts from epididymal spermatozoa contained protamines P1 and P2. Extracts from testis (Fig. 1*b*, lane T) contained protamines P1 and P2 plus spermatid proteins A–F. Proteins A–D were major species, whereas proteins E and F were present in such limited concentrations that they could not always be monitored here. Histones, characteristic of nuclei destroyed by sonication, were present in only trace amount.

Gel-permeation h.p.l.c. was used to purify these proteins, and, as shown in Fig. 2, the best resolution was obtained with 0.2% trifluoroacetic acid as the elution solution. All of these proteins were eluted with a total volume of 1 ml, and all were eluted much earlier than less basic reference proteins of similar M_r (Fig. 2*d*) [30]. Electrophoretic analysis of the h.p.l.c. fractions (Fig. 1) revealed that protamine P1 and protamine P2 from either testis or epididymis were separated on the column. Proteins A, B, C and D were eluted ahead of protamine P2, with protein A being eluted somewhat earlier than proteins C and D. A second cycle of gel-permeation h.p.l.c. with the appropriate fractions was used to purify protamine P1 and protamine P2. A mixture of proteins A–D was also purified and contained protein C with smaller amounts of proteins A, B and D.

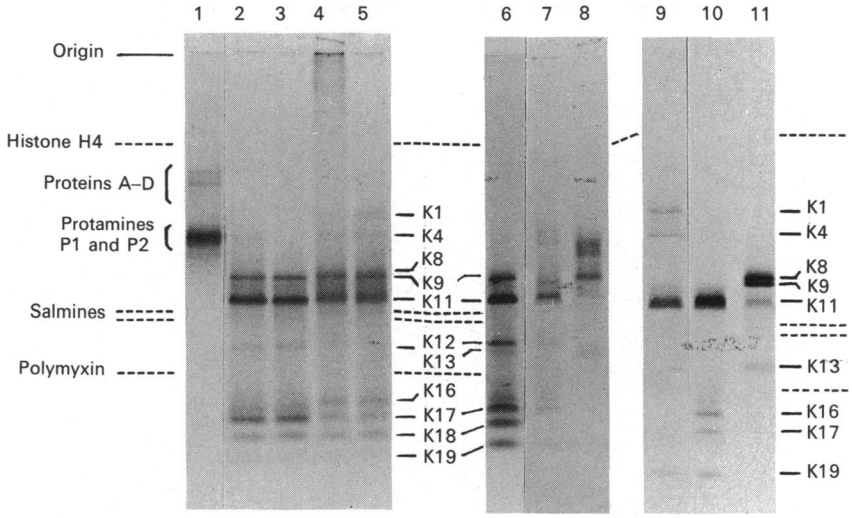

Fig. 3. Peptides produced by endoproteinase Lys-C

[^{14}C]Carboxyamidomethylated proteins (1 μg; approx. 10^4 c.p.m.) from epididymis or testis, purified by h.p.l.c. when required, were mixed with unfractionated carrier (10 μg) from the same source and digested with endoproteinase Lys-C. In two cases, noted below, proteins were treated with alkaline phosphatase before digestion. Peptides were then submitted to electrophoresis on 15% polyacrylamide gels. As described in the text, peptides K11, K17, K18 and K19 were derived from protamine P2, and peptides K9 and K13 were derived from protamine P1. Recovery of peptide K19 was somewhat variable from one gel to another since this small peptide was partly lost during staining and destaining of the gel. Dotted lines indicate positions of migration markers on the gels. Lanes 1–6 show Amido Black-stained gels, and lanes 7–11 show fluorographs. Lane 1, unfractionated proteins from testis, undigested; lane 2, protamines P1 and P2 from epididymis; lane 3, protamines P1 and P2 from epididymis, alkaline-phosphatase-treated; lane 4, unfractionated proteins from testis; lane 5, unfractionated proteins from testis, alkaline-phosphatase-treated; lane 6, protamines P1 and P2 from epididymis; lane 7, protamine P2 from epididymis; lane 8, protamine P1 from epididymis; lane 9, proteins A–D; lane 10, protamine P2 from testis; lane 11, protamine P1 from testis.

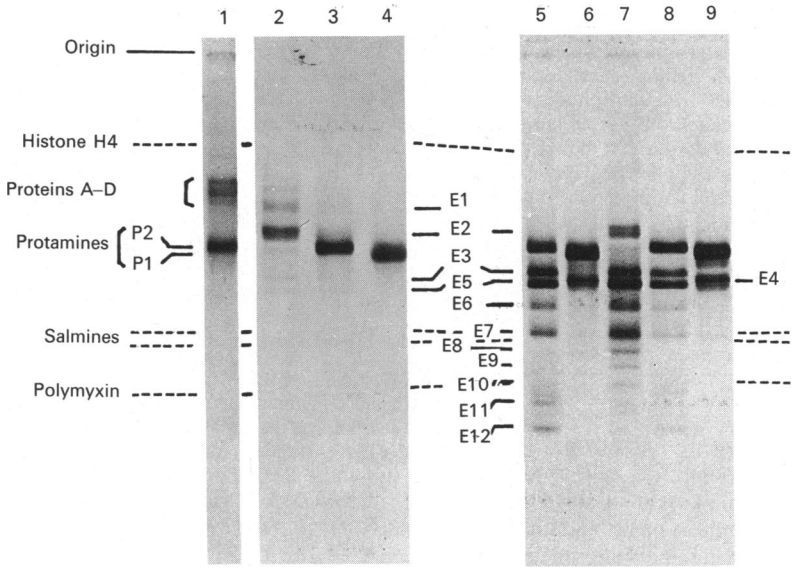

Fig. 4. Peptides produced by endoproteinase Glu-C

[^{14}C]Carboxyamidomethylated proteins (2 μg; approx. 5×10^3 c.p.m.) from epididymis or testis, purified by h.p.l.c. when required, were mixed with unfractionated carrier (10 μg) from the same source and digested with endoproteinase Glu-C for either 5 or 105 min. Digests were analysed by electrophoresis on 15% polyacrylamide gels. Fluorographs are shown with positions of migration markers indicated by dotted lines. Lane 1, unfractioned proteins from testis, undigested; lane 2, proteins A–D, 5 min; lane 3, protamine P2 from testis, 5 min; lane 4, protamine P1 from testis, 5 min; lane 5, protamine P2 from epididymis, 105 min; lane 6, protamine P1 from epididymis, 105 min; lane 7, proteins A–D, 105 min; lane 8, protamine P2 from testis, 105 min; lane 9, protamine P1 from testis, 105 min.

Endoproteinase digestion of protamine P1, protamine P2 and proteins A–D

To compare the primary structures of proteins A–D with those of protamines P1 or P2, [^{14}C]carboxyamidomethylated proteins were purified by two cycles of h.p.l.c. and digested with either endoproteinase Lys-C or endoproteinase Glu-C, which cleave peptide bonds involving the carbonyl function of lysine or glutamic acid residues respectively. The resulting peptides were then submitted to electrophoresis on acid/urea/15%-polyacrylamide gels. Since basic proteins migrate with rates approximately proportional to their M_r values on such gels, the peptides could be identified with reference to the known sequences of

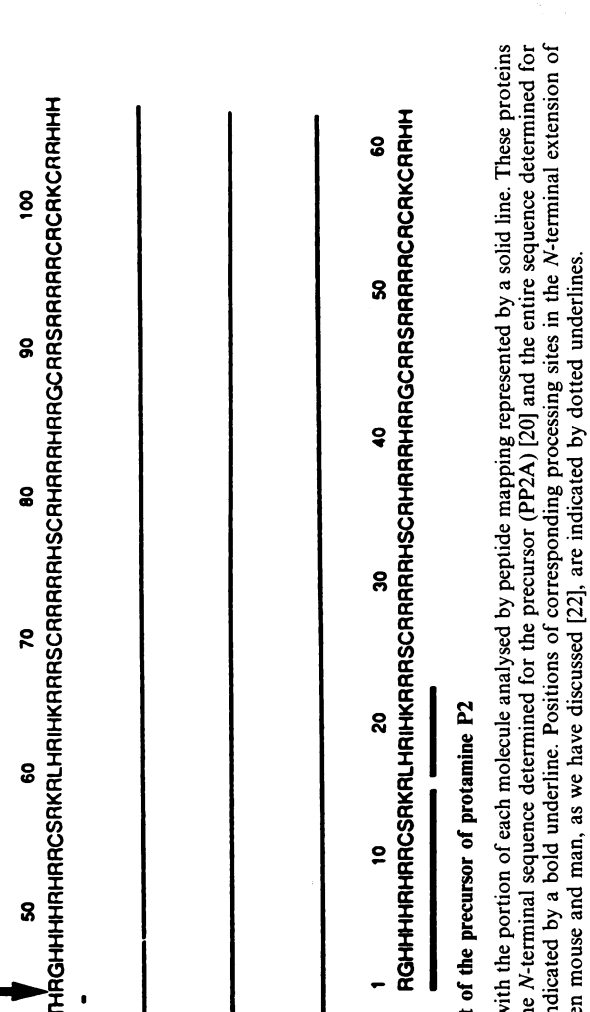

Fig. 5. *N*-Terminal sequences for protein C (PP2C) and protein D (PP2D) compared with that of the precursor of protamine P2

N-Terminal sequences determined for protein C (PP2C) and protein D (PP2D) are shown, with the portion of each molecule analysed by peptide mapping represented by a solid line. These proteins are aligned relative to the sequence for the entire precursor as predicted from DNA [20], the *N*-terminal sequence determined for the precursor (PP2A) [20] and the entire sequence determined for protamine P2 [17]. The portion of the protamine P2 sequence that we have confirmed is indicated by a bold underline. Positions of corresponding processing sites in the *N*-terminal extension of the precursor are indicated by arrows. Regions of the extension that are conserved between mouse and man, as we have discussed [22], are indicated by dotted underlines.

protamine P1 [15], protamine P2 [17] and the precursor of protamine P2 [20].

Endoproteinase Lys-C. Digests of protamine P2 from epididymis (Fig. 3, lane 7) showed three major radiolabelled peptides, K11, K17 and K19. Two minor peptides, K12 and a peptide below K9, varied widely in intensity in different experiments and probably resulted from a contaminating activity of the enzyme. Only one additional unlabelled peptide, K18, from protamine P2 was revealed in digests stained with Amido Black (lanes 2 and 6). The origin of the four peptides from protamine P2, deduced from its sequence seen in Fig. 5, are the following: K17 represents the *N*-terminus (residues 1–14), K18 the peptide containing no labelled cysteine (residues 15–21), K11 the core (residues 22–57) and K19 the *C*-terminus (residues 58–62). Protamine P2 from testis (lane 10) consistently showed an additional peptide, K16. Peptides K16 and K17 were purified by reverse-phase h.p.l.c. and submitted to *N*-terminal sequence analysis. Both were found to represent the *N*-terminus of protamine P2 (results not shown), indicating that a modification of the *N*-terminus is present in testis. Treatment of the protein with alkaline phosphatase before digestion did not change the appearance of any peptide (lanes 3 and 5).

Digests of protamine P1 from epididymis (lane 8) showed two major peptides, K9 and K13, in addition to undigested protamine P1. Since the sequence of mouse protamine P1 includes three lysine residues (positions 9, 48 and 49) in a total of 50 residues [15], it may be deduced that K13 represents the *N*-terminus (residues 1–9), and K9 most of the remainder of the molecule (residues 10–48). Protamine P1 from testis (lane 11) showed an additional peptide, K8, probably a modified form of K9. Minor peptides arose from traces of protamine P2 in this protein preparation. As with protamine P2, none of the peptides (lanes 3 and 5) were modified by phosphatase treatment.

Digests of proteins A–D (lane 9) showed peptides K11 and K19, in common with protamine P2 (lanes 7 and 10), and an additional cluster of peptides K1–K4. This cluster may be interpreted from the amino acid sequence of the precursor. Since no lysine residue is present in the *N*-terminal extension of the precursor (Fig. 5), digestion should yield K19, K11, unlabelled K18 and an additional peptide corresponding to K17 (the *N*-terminus of protamine P2) joined to the *N*-terminal extension. K1, in all probability from its high M_r and weak label, corresponds to this additional peptide, whereas K17 is notably absent from the protein A–D digest. Digestion of a series of molecules derived from the precursor by modification of the extension, in particular intermediates generated by processing, should yield a series of discrete peptides slightly smaller than peptide K1. The peptide cluster K1–K4 appears to correspond to such molecules. Proteins A–D thus contain within their sequences K11 and K19, representing the *C*-terminus and core of protamine P2, plus K1–K4, representing the *N*-terminus of protamine P2 joined to the extension of its derivatives found *in vivo*. Peptides derived from protamine P1 do not appear in this digest.

Endoproteinase Glu-C. Peptides obtained from endproteinase Glu-C digestion are shown in Fig. 4. Protamines P1 (lane 4) and P2 (lane 3) remain essentially intact after 5 min of digestion, since they contain no glutamic acid residues. In contrast, proteins A–D (lane 2) yield three bands, a minor band co-migrating with protein C, a stronger band E1 and a major band E2. The precursor contains six glutamic acid residues (positions 11, 22, 24, 25, 35 and 38) in its *N*-terminal extension (Fig. 5). The peptide co-migrating with protein C is therefore likely to represent a partial digestion product in which the entire protamine P2 sequence is joined to 33 amino acid residues of the extension, whereas E1 represents protamine P2 joined to 22, 20 or 19 residues and E2 represents protamine P2 joined to nine or

Table 1. Automated *N*-terminal sequence analysis of protein C purified by electrophoresis and electroblotted on to poly(vinylidene difluoride) membranes

Protein C, purified by electrophoresis on 10% polyacrylamide gels and transferred in acetic acid solution to poly(vinylidene difluoride) membranes, was submitted to Edman degradation in an Applied Biosystems 470 A gas-phase sequencer. Reagent penetration was somewhat impeded since, in order to introduce an adequate quantity of protein (approx. 0.5 nmol), the sample cartridge was loaded to capacity with membranes.

	Amino acid residue (pmol over background)						
Cycle no.	G	P	H	Q	G	P	G
1	39.6	–	–	–	–	–	–
2	38.7	21.5	–	–	–	–	–
3	–	18.8	4.8	–	–	–	–
4	–	6.4	9.6	12.7	–	–	–
5	–	–	7.0	11.5	7.6	–	–
6	–	–	–	–	10.6	6.1	–
7	–	–	–	–	–*	9.7	10.9*

* Glycine recovered in cycle 7 included probable carry-over from glycine in position 5.

six residues. Since no additional peptides are present, it appears that proteins B–D yield digestion products that co-migrate with those of protein A, the precursor. This result implies that proteins B–D also contain the entire protamine P2 sequence plus the protein region rich in glutamic acid residues, the *N*-terminal extension of the precursor or portions thereof.

For longer incubation times, namely 105 min, peptide bonds in addition to those involving glutamic acid residues were hydrolysed. The specificity of this additional activity is not known, but the peptides produced are nevertheless informative. Protamine P2 (lanes 5 and 8) was partially hydrolysed into E3, E5, E6, E7, E11 and E12. In contrast, protamine P1 (lanes 6 and 9) remained essentially intact with only one additional peptide, E4, appearing. In digests of proteins A–D (lane 7), E2 remained and several additional peptides appeared, E3, E5, E6, E7, E8 and E9, four of which correspond to peptides in the protamine P2 digest. E4, which was characteristic of protamine P1, is not found in the protein A–D digest. The production of a limited number of identical peptides from proteins A–D and protamine P2, by this additional cleavage activity, again demonstrates the presence of the protamine P2 sequence within those of proteins A–D.

Determination of *N*-terminal sequences

The *N*-terminal sequence of protamine P2 was detemined on protein extracted from epididymis and purified by two cycles of h.p.l.c. It was free from contamination by protamine P1 and other proteins, as judged by electrophoresis on 10%-polyacrylamide gels. The *N*-terminal sequence determined by automated Edman degradation was RGHHHHRHRRCSRKXLHRIHKR, and this sequence is in complete agreement with sequences published previously for mouse protamine P2 (Fig. 5) [17,20]. As noted above, corroborating partial *N*-terminal sequences were also obtained for two peptides representing the *N*-terminus of protamine P2.

The *N*-terminal sequence of protein C was analysed, first, with the use of protein extracted from testis and purified by two cycles of h.p.l.c. This preparation still contained traces of proteins A and D. The following sequence was obtained by automated Edman degradation: (G/S)PH(Q/E)GPGQ. Minor derivatives in several cycles represented residues from the recognizable sequences of proteins A and D. A second sequence analysis was carried out on protein C purified by electrophoresis on a 10%-polyacrylamide gel with subsequent electroblotting of the proteins on to poly(vinylidene difluoride) membranes in acetic acid solution. A membrane strip corresponding to protein C was charged directly into the sequencer. As shown in Table 1, the first six residues were identified unambiguously as GPHQGP, with no evidence of contamination by other proteins. This result therefore clarifies both ambiguities present in the sequence determined on h.p.l.c.-purified protein and otherwise confirms the sequence. As shown in Fig. 5, this *N*-terminal sequence for protein C is present within the sequence of the precursor of protamine P2, beginning with Gly-12.

Protein D was also purified by electrophoresis of proteins extracted from testis with subsequent electroblotting of the proteins on to poly(vinylidene difluoride) membranes. Automated sequence analysis was carried out on two such preparations, giving corroborating results. In both cases, the early cycles of Edman degradation showed high background due to the large quantity of membrane necessarily loaded into the sample cartridge. Consequently identification of amino acid residues in four cycles was uncertain. The sequence obtained for protein D was the following: XERXXQGQGQXLS. As shown in Fig. 5, this sequence is present unambiguously within the sequence of the precursor, beginning with His-21.

DISCUSSION

Processing of the *N*-terminal extension of the precursor of protamine P2, protein A, takes place in spermatid chromatin, and our previous physiological experiments have suggested that at least five intermediates, proteins B–F, are generated during this process [22]. In the work reported here, we have studied the structure of two of these proteins, proteins C and D. Sequence analysis has located the *N*-termini of each protein within the *N*-terminal extension of the precursor of protamine P2, and peptide mapping has demonstrated that these proteins continue through the protamine P2 sequence to its *C*-terminus. These structural studies therefore indicate that proteins C and D are in fact intermediates generated by proteolytic processing *in vivo* of the *N*-terminal extension. M_r values calculated for the proteins from their sequences are 12109 for protein C and 11235 for protein D. Since the precursor of protamine P2, protein A, has been designated previously as PP2 [16], we propose that proteins A, C and D may now be designated as PP2A, PP2C and PP2D respectively.

Processing takes place during chromatin condensation and cleaves portions of the presursor that presumably have served their purpose and are no longer necessary for, or perhaps even compatible with, the compact structure of chromatin in spermatozoa. We have pointed out regions of the *N*-terminal extension that are conserved between the precursors in mouse and man, their conservation suggesting that these regions may be important for precursor function [22]. In Fig. 5 it may be noted that cleavage giving rise to PP2C removes a conserved region of 11 residues from the *N*-terminus of the precursor. Cleavage giving rise to PP2D also occurs adjacent to a similarly conserved region. Spermatid proteins B, E and F, which were present in concentrations too low to be sequenced here, appear from their estimated M_r values [22] to originate possibly from cleavages within conserved regions.

It may also be seen that PP2C and PP2D both result from cleavage of a peptide bond whose carbonyl function is contributed by an acidic amino acid (Fig. 5). The final cleavage site giving rise to mature protamine P2, however, does not involve an

acidic residue. Two proteins with M_r values approximating those of mouse spermatid proteins E and F have been purified from human spermatozoa, and sequence analysis has indicated that they might have arisen from processing of a human precursor of protamine P2 [31,32]. Only one of the respective cleavage sites involves an acidic residue, as does the cleavage site producing the larger of the two forms of human protamine P2 [12,18,33]. It therefore appears that processing of the precursor of protamine P2 involves proteolytic activities of different or multiple specificity. Nuclear proteinases that might carry out processing during chromatin remodelling have not yet been characterized in testis and promise to be a challenging subject for future research.

This work was supported by grants from the Centre National de la Recherche Scientifique and the Université Paris VII, Paris, France. The Institut Jacques Monod is an 'Institut Mixte', C.N.R.S.–Université Paris VII. We are indebted to G. Desvages for valuable advice on h.p.l.c. techniques and to P. Y. Haumont and S. Cortial for assistance with amino acid sequence determinations.

REFERENCES

1. Bellvé, A. R. & Carraway, R. (1978) J. Cell Biol. **79**, A177
2. Bellvé, A. R., Anderson, E. & Hanley-Bowdoin, L. (1975) Dev. Biol. **47**, 349–365
3. Calvin, H. I. (1976) Biochim. Biophys. Acta **434**, 377–389
4. Balhorn, R., Gledhill, B. L. & Wyrobek, A. J. (1977) Biochemistry **16**, 4074–4080
5. Hecht, N. B., Kleene, K. C., Yelick, P. C., Johnson, P. A., Pratcheva, D. D. & Ruddle, F. H. (1986) Somatic Cell Mol. Genet. **12**, 203–208
6. Corzett, M., Blacher, R., Mazrimas, J. A. & Balhorn, R. (1987) J. Cell Biol. **105**, 152a
7. Bélaïche, D., Loir, M., Kruggle, W. & Sautière, P. (1987) Biochim. Biophys. Acta **913**, 145–149
8. Mazrimas, J. A., Corzett, M., Campos, C. & Balhorn, R. (1986) Biochim. Biophys. Acta **872**, 11–15
9. Tobita, T., Tsutsumi, H., Kato, A., Suzuki, H., Nomoto, M., Nakano, M. & Ando, T. (1983) Biochim. Biophys. Acta **744**, 141–146
10. Sautière, P. Bélaïche, D., Martinage, A. & Loir, M. (1984) Eur. J. Biochem. **144**, 121–125
11. Ammer, H. & Henschen, A. (1988) Biol. Chem. Hoppe-Seyler **369**, 1301–1306
12. Ammer, H., Henschen, A. & Lee, C. H. (1986) Biol. Chem. Hoppe-Seyler **367**, 515–522
13. Ammer, H. & Henschen, A. (1987) Biol. Chem. Hoppe-Seyler **368**, 1619–1626
14. Ammer, H. & Henschen, A. (1988) FEBS Lett. **242**, 111–116
15. Kleene, K. C., Distel, R. J. & Hecht, N. B. (1985) Biochemistry **24**, 719–722
16. Kleene, K. C. & Flynn, J. (1987) Dev. Biol. **123**, 125–135
17. Bellvé, A. R., McKay, D. J., Renaux, B. S. & Dixon, G. H. (1988) Biochemistry **27**, 2890–2897
18. McKay, D. J., Renaux, B. S. & Dixon, G. H. (1986) Eur. J. Biochem. **156**, 5–8
19. Pirhonen, A., Valtonen, P., Linnala-Kankkunen, A., Heiskanen, M. L. & Mäenpää, P. H. (1990) Biochim. Biophys. Acta **1039**, 177–180
20. Yelick, P. C., Balhorn, R., Johnson, P., Corzett, M., Mazrimas, J. A., Kleene, K. C. & Hecht, N. B. (1987) Mol. Cell. Biol. **7**, 2173–2179
21. Johnson, P. A., Peschon, J. J., Yelick, P. C., Palmiter, R. D. & Hecht, N. B. (1988) Biochim. Biophys. Acta **950**, 45–53
22. Elsevier, S. M., Noiran, J. & Carré-Eusèbe, D. (1991) Eur. J. Biochem. **196**, 167–175
23. Elsevier, S. M. (1982) Dev. Biol. **90**, 1–12
24. Mayes, E. L. V. (1984) Methods Mol. Biol. **1**, 33–39
25. Zirkin, B. R., Chang, T. S. K. & Heaps, J. (1980) J. Cell Biol. **85**, 116–121
26. Moos, M. Jr., Nguyen, N. Y. & Liu, T. Y. (1988) J. Biol. Chem. **263**, 6005–6008
27. Matsudaira, P. (1987) J. Biol. Chem. **262**, 10035–10038
28. Balhorn, R., Weston, S., Thomas, C. & Wyrobek, A. J. (1984) Exp. Cell Res. **150**, 298–308
29. Meistrich, M. L., Reid, B. O. & Barcellona, W. J. (1976) Exp. Cell. Res. **99**, 72–78
30. Irvine, G. B. & Shaw, C. (1986) Anal. Biochem. **155**, 141–148
31. Sautière, P., Martinage, A., Bélaïche, D., Arkhis, A. & Chevaillier, P. (1988) J. Biol. Chem. **263**, 11059–11062
32. Martinage, A., Arkhis, A., Alimi, E., Sautière, P. & Chevaillier, P. (1990) Eur. J. Biochem. **191**, 449–451
33. Domenjoud, L., Nussbaum, G., Adham, I. M., Greeske, G. & Engel, W. (1990) Genomics **8**, 127–133

Received 29 November 1990; accepted 8 January 1991