MOLECULAR AND CELLULAR BIOLOGY, June 1987, p. 2173–2179
0270-7306/87/062173-07$02.00/0

Vol. 7, No. 6

Mouse Protamine 2 Is Synthesized as a Precursor whereas Mouse Protamine 1 Is Not

PAMELA C. YELICK,[1] ROD BALHORN,[2] PAULA A. JOHNSON,[1] MICHELE CORZETT,[2] JOE A. MAZRIMAS,[2] KENNETH C. KLEENE,[1] AND NORMAN B. HECHT[1]*

Department of Biology, Tufts University, Medford, Massachusetts 02155,[1] and Lawrence Livermore National Laboratory, Biomedical Sciences Division, Livermore, California 94550[2]

Received 3 November 1986/Accepted 18 March 1987

The nuclei of mouse spermatozoa contain two protamine variants, mouse protamine 1 (mP1) and mouse protamine 2 (mP2). The amino acid sequence predicted from mP1 cDNAs demonstrates that mP1 is a 50-amino-acid protein with strong homology to other mammalian P1 protamines. Nucleotide sequence analysis of independently isolated, overlapping cDNA clones indicated that mP2 is initially synthesized as a precursor protein which is subsequently processed into the spermatozoan form of mP2. The existence of the mP2 precursor was confirmed by amino acid composition and sequence analysis of the largest of a set of four basic proteins isolated from late-step spermatids whose synthesis is coincident with that of mP1. The sequence of the first 10 amino acids of this protein, mP2 precursor 1, exactly matches that predicted from the nucleotide sequence of cDNA and genomic mP2 clones. The amino acid composition of isolated mP2 precursor 1 very closely matches that predicted from the mP2 cDNA nucleotide sequence. Sequence analysis of the amino terminus of isolated mature mP2 identified the final processing point within the mP2 precursor. These studies demonstrated that mP2 is synthesized as a precursor containing 106 amino acids which is processed into the mature, 63-amino-acid form found in spermatozoa.

Protamines are small, basic, arginine-rich proteins that replace histones in the later stages of spermatogenesis in many vertebrates (N. B. Hecht, *in* G. Stein and J. Stein, ed., *Basic Chromosomal Proteins: Structure, Organization and Regulation of the Gene*, in press). During spermatogenesis in mammals, the histones are not directly replaced by protamines but by a population of basic proteins that are transiently associated with the spermatid nucleus (14, 35; Hecht, in press). These testis-specific proteins are removed from the condensing chromatin at later stages of spermiogenesis and are replaced by protamines, the predominant class of proteins found complexed with DNA in the mature spermatozoon. Mammalian protamines contain the amino acid cysteine, which is believed to function in the stabilization and compaction of the sperm nucleus through disulfide bond cross-linking (7).

Protamines have been isolated from a large number of vertebrates including fish (3, 28), domestic fowl (38), mice (6, 27, 40; A. R. Bellvé and R. Carraway, J. Cell. Biol. **79**:177a, 1978), rats (20, 21), bulls (10, 32, 36), boars (36, 43), rams (29, 36, 41), stallions (36), guinea pigs (8, 9), rabbits (8, 9), and humans (2, 12, 25, 33, 39). Complete sequence analysis of the P1 protamines from bulls (10, 32), boars (43), rams (41), and humans (2, 33) and the predicted amino acid sequence derived from a cDNA sequence for mouse protamine 1, (mP1) (24) have revealed identical lengths of 50 amino acids and strong sequence homologies among these mammalian protamines. P1 protamine is organized into three domains consisting of a central basic core with clusters of arginine and two less basic amino- and carboxyl-terminal regions. The conservation of the highly basic central core region appears essential to protamine function since this region is also present in all the trout protamines examined (1).

The nuclei of all mammalian sperm examined contain P1 protamine. Certain species, including mice (Bellvé and Carraway, J. Cell Biol. 1978), humans (2, 25, 34, 39), and hamsters (P. A. Bower, P. C. Yelick, and N. B. Hecht, Biol. Reprod., in press; R. Balhorn unpublished observation), contain a second protamine variant which is surprisingly different in amino acid composition from P1 protamines. Although protamine 2 variants contain high levels of arginine and cysteine, as do P1 protamines, P2 protamine is distinguished from P1 in that it has a high histidine content (greater than 20% in mice), whereas protamine 1 contains no histidine residues.

We have previously reported the isolation and characterization of a cDNA encoding mP1 (24). Here we present the nucleotide sequence for mouse protamine 2 (mP2) obtained from two overlapping mouse testicular cDNAs. The predicted amino acid sequences of the proteins encoded by these two cDNA clones indicate that mP2 is first synthesized as a precursor molecule containing 106 amino acids. In contrast to mP1, which is synthesized with a size the same as that of the mP1 molecule found in mouse spermatozoa, mP2 is initially synthesized as a precursor protein substantially larger, and of very different amino acid composition, than the protamine 2 present in mouse sperm. This was confirmed by the isolation and analysis of the largest of a set of four basic proteins isolated from sonication-resistant late-step spermatids.

MATERIALS AND METHODS

Isolation of cDNA mP2 clones. Two clones, pmP2-1 and pmP2-2, were isolated from a mouse testis cDNA library by colony hybridization by using a ^{32}P-labeled nick-translated insert from a plasmid previously designated HSAR 900 (22). HSAR 900 represents an abundant poly(A)$^+$ mRNA present in round and elongating spermatids. After preliminary characterization by DNA restriction mapping, the nucleotide sequences of the cDNA inserts of clones pmP2-1 and

* Corresponding author.

pmP2-2, containing 530 and 437 base pairs, respectively, were determined for both strands by the method of Maxam and Gilbert (30). Fragments for sequencing were labeled at the 3′ termini by using the Klenow fragment of *Escherichia coli* DNA polymerase I (30). Separation of cDNA strands was achieved by denaturation in methyl mercury and electrophoresis in 10% acrylamide gels (42).

Isolation of sperm-specific and late-step spermatid-specific proteins. Late-step sonication-resistant spermatids were isolated from 200 frozen (−20°C) testes, as described previously (6). Sperm were isolated from the caput, corpus, and caudal regions of 100 epididymides by teasing the organ sections into saline containing 0.01 M Tris buffer (pH 8.0) (Tris-saline) and filtering out the connective tissue and debris by using silk gauze (pore size, 80 μm). The sperm were subsequently washed two times in Tris-saline with brief sonication (Sonifier; Branson Sonic Power Co., Danbury, Conn.; setting 1 for 5 to 10 s) and pelleted by centrifugation at 3,000 × *g* for 3 min. The spermatids and sperm were reduced with 10 mM dithiothreitol–0.01 M Tris (pH 8.0), and the tails, acrosomes, and membranes were removed by treatment with 1% mixed alkyltrimethylammonium bromide (5). The resulting amembranous nuclei were washed in Tris-saline and dissolved in 5 M guanidine hydrochloride, and the basic proteins were extracted as described previously (5).

HPLC separation of sperm and spermatid chromatin proteins. Total sperm or spermatid basic protein was dissolved at a concentration of 1 to 5 mg/ml in 5 M guanidine hydrochloride–0.01 M Tris (pH 8)–0.002 M EDTA containing a 25-fold molar excess of dithiothreitol over protamine cysteine and reduced at 21°C for approximately 16 h. Immediately before chromatography, the sample was dialyzed against 2 liters of 10 mM hydrochloric acid containing 300 mg of dithiothreitol.

Both the sperm and spermatid proteins were separated by reverse-phase high-pressure liquid chromatography (HPLC) on a Nucleosil RP-C18 column (7.5 mm [inside diameter] by 300 mm; Machery-Nagel, Duren, Federal Republic of Germany) by using a Beckman 332 Gradient System HPLC (Beckman Instruments, Inc., Mountain View, Calif.) with two 110A pumps and a controller. Trifluoroacetic acid was added to the protein solution to a final concentration of 0.1%, the sample was injected, and a multistep linear gradient was used to fractionate the proteins. The mobile phase used to elute the proteins (buffer A) was aqueous 0.1% trifluoroacetic acid, and the mobile-phase modifier (buffer B) was 30% (vol/vol) acetonitrile (HPLC grade; J. T. Baker Chemical Co., Phillipsburg, N.J.)–0.1% trifluoroacetic acid. The flow rate was maintained at 1.0 ml/min. The gradient was generated in steps by increasing the percent buffer B as follows: increase from 40 to 65% over 20 min, hold at 65% for 5 min, increase from 65 to 80% over 30 min, increase from 80 to 100% over 10 min, and decrease from 100 to 40% in 1 min. Protein elution was monitored by using an ISCO V4 variable-wavelength UV detector (ISCO, Lincoln, Nebr.), and each peak was collected and lyophilized.

Amino acid analysis and protein sequencing. The dried protein samples were dissolved in 6 M guanidine hydrochloride–0.002 M EDTA–0.01 M Tris (pH 8) and carboxymethylated as described for bull protamine (32). After desalting by chromatography on a Sephadex G-25 column (1 by 100 cm) and lyophilization, 100 μg of protein was dissolved in 6 M hydrochloric acid, sealed under vacuum, and hydrolyzed at 110°C for 24 h. The hydrolysates were lyophilized and analyzed on a Beckman 6300 Amino Acid Analyzer.

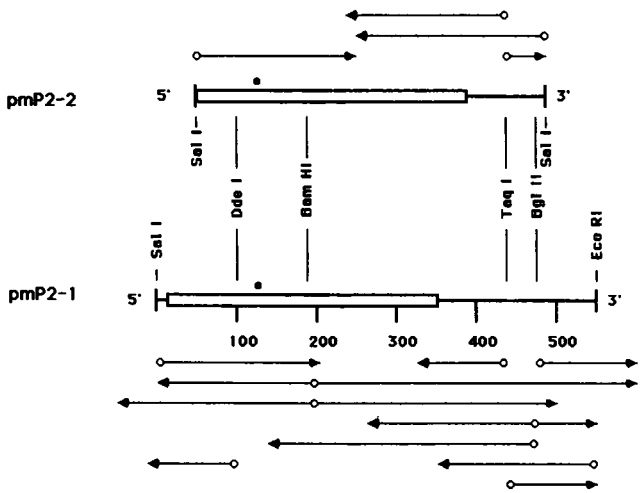

FIG. 1. Partial restriction map and DNA sequencing strategy for mP2. DNA fragments labeled at the 3′ ends were sequenced by the procedure of Maxam and Gilbert (30). Symbols: ▭, open reading frames in the cDNAs; ——, untranslated regions; ○, the labeled ends of the fragments which were sequenced. Arrows extending past the cDNA inserts indicate sequence which was read through the cloning site into the pUC8 vector sequence. The solid squares above the coding portion of the cDNAs indicate the amino-terminal end of mature mP2.

The amino-terminal sequences of carboxymethylated precursor 1 (5 μg) and mP2 (75 μg) were determined at the Protein Structure Laboratory, University of California, Davis, by the same methods described for bull protamine (32).

RESULTS

Nucleotide sequence of the mP2 precursor. The restriction sites and strategy used to sequence clones pmP2-1 and pmP2-2 by the sequencing method of Maxam and Gilbert (30) are shown in Fig. 1. The nucleotide sequences of and predicted amino acid sequences encoded by mP1 and mP2 cDNAs are shown in Fig. 2. DNA sequence analysis of the isolated insert from pmP2-1 revealed two possible initiation codons (AUG) located 3 and 18 bases from the 5′ end of the sequence (Fig. 2B). An entire consensus sequence for initiation of protein synthesis in eucaryotic cells (26) is not present in pmP2-1. However, analysis of a genomic clone for mP2 (data not shown) indicated a consensus sequence (GCACCATGG) surrounding the first AUG codon of pmP2-1. This sequence exactly matches the translational initiation sequence for mP1 (24).

The nucleotide sequences of the mP2 cDNAs indicate a coding region of 324 bases and a 3′ untranslated region of 189 bases excluding the poly(A) tail. The full-length mP2 mRNA was estimated to be about 830 bases long by its mobility in agarose denaturing gels (16), and the poly(A)$^+$ tail was estimated to be 160 bases long by RNase H analysis (data not shown). This leaves approximately 157 bases for the 5′ untranslated region of the mP2 mRNA. The lengths of the untranslated regions of mP2 mRNA are similar to those of mP1 mRNA, which contains a 5′ untranslated region of about 150 bases, a 3′ untranslated region of 151 bases, and a poly(A) tail of approximately 160 bases (24). The predicted amino acid sequence encoded by the mP1 cDNA agrees well with the amino acid composition of P1 protamine isolated from mouse spermatozoa, confirming that no additional

```
A.  mP1

5'
 GACCCCTGCTCACAGGTTGGCTGGCTCGACCCAGGTGGTGTCCCCTGCTCTGAGCCAGCTCCCGGCCAA
-78                -60                                   -30

                  1                                      10
                  Ala Arg Tyr Arg Cys Cys Arg Ser Lys Ser Arg Ser Arg
GCCAGCACC ATG GCC AGA TAC CGA TGC TGC CGC AGC AAA AGC AGG AGC AGA
           1                                     30

                             20
Cys  Arg Arg Arg Arg Arg Arg Cys Arg Arg Arg Arg Arg Arg Cys Cys
TGC  CGC CGT CGC AGA CGA AGA TGT CGC AGA CGG AGG AGG CGA TGC TGC
                          60                                    90

30                                           40
Arg  Arg Arg Arg Arg Arg Cys Cys Arg Arg Arg Arg Ser Tyr Thr  Ile
CGG  CGG AGG AGG CGA AGA TGC TGC CGT CGC CGC CGC TCA TAC ACC  ATA
                                          120

                    50
Arg Cys Lys Lys Tyr ***
AGG TGT AAA AAA TAC TAG ATGCACAGAATAGCAAGTCCATCAAAACTCCTGCGTGAGAATT
                 150                                  180

TTACCAGACTTCAAGAGCATCTCGCCACATCTTGAAAAATGCCACCGCCGATGAAAAACAGGAGCC
          210                                  240

TGCTAAGGAACAATGCCACCTGCCAATAAATGTTGAAAACTCAAAAAAAAAAAAAAAAAAA      3'
    270                                300                      326

B.  mP2
          1                                      10
      Met Val Arg Tyr Arg Met Arg Ser Pro Ser Glu Gly Pro His Gln
   CC ATG GTT CGC TAC CGA ATG AGG AGC CCC AGT GAG GGT CCG CAC CAG
      1                                      30
                          20                                      30
Gly  Pro Gly Gln Asp His Glu Arg Glu Glu Gln Gly Gln Gly Gln Gly
GGG  CCT GGA CAA GAC CAT GAA CGC GAG GAG CAG GGG CAG GGG CAA GGG
                      60                                ▼   90
                                       40
Leu  Ser Pro Glu Arg Val Glu Asp Tyr Gly Arg Thr His Arg Gly His
CTG  AGC CCA GAG CGC GTA GAG GAC TAT GGG AGG ACA CAC AGG GGC CAC
                                         120

                50                                         60
His  His His Arg His Arg Arg Cys Ser Arg Lys Arg Leu His Arg Ile
CAC  CAC CAC AGA CAC AGG CGC TGC TCT CGT AAG AGG CTA CAT AGG ATC
              150                                        180
                               70
His Lys Arg Arg Arg Ser Cys Arg Arg Arg Arg Arg His Ser Cys Arg
CAC AAG AGG CGT CGG TCA TGC AGA AGG CGG AGG AGA CAC TCC TGC CGC
                             210
     80                                        90
His  Arg Arg Arg His Arg Arg Gly Cys Arg Arg Ser Arg Arg Arg  Arg
CAC  AGG AGG CGG CAT CGC AGA GGC TGC AGA AGA TCC CGA AGG AGG  AGG
 240                                         270
                        100
Arg Cys Arg Cys Arg Lys Cys Arg Arg His His His ***
AGA TGC AGG TGC AGG AAA TGT AGG AGG CAC CAT CAC TAA GCCTCCCCAGGCC
                     300                                    330

TGTCCATTCTGCCTGGAGCCAAGGAAGTCACTTGCCCAAGGAATAGTCACCTGCCCAAGCAACAT
                           360                          390

CATGTGAGGCCACACCACCATTCCATGTCGATGTCTGAGCCCTGAGCTGCCAAGGAGCCACGAGA
                     420                                450

TCTGAGTACTGAGCAAAGCCACCTGCCAAATAAAGCTTGACACGAGAAAAAAAAAAAAAAAAA 3'
              480                          510
```

FIG. 2. Nucleotide sequences, and corresponding amino acid sequences, of the cDNAs for mP1 and mP2. Identical sequences were obtained for the shared regions of the inserts of pmP2-1 and pmP2-2. The predicted amino acid sequences of the cDNAs are presented above the nucleotide triplets. Amino acids are numbered beginning with alanine for mP1 and valine for mP2 since methionine does not appear at the amino termini of isolated mouse protamines. The mP1 sequence is from Kleene et al. (24). The arrowhead in the mP2 sequence indicates the first amino acid of the form of mP2 isolated from spermatozoa. The termination codons, TAG for mP1 and TAA for mP2, are denoted by asterisks. The canonical hexanucleotide, AATAAA, is indicated for each cDNA by a line above the nucleotide sequence. The nucleotide sequences found in the 3′ untranslated regions of both mP1 and mP2 are underlined.

amino acids are present in the mP1 primary translation product (24; Bellvé and Carraway, J. Cell. Biol. 79:177a, 1978).

Identification of the amino terminus of mP2 precursor 1. Basic proteins isolated from epididymal sperm and sonication-resistant late-step spermatids were fractionated by HPLC (Fig. 3) and separated by gel electrophoresis (Fig. 4). Confirmation of the translational start site of the mP2 cDNA was provided by sequencing the amino terminus of isolated mP2 precursor 1. The first 10 amino acids of this protein exactly match the amino acid sequence predicted from the pmP2-1 nucleotide sequence starting with the first AUG codon located at nucleotide 3 (Fig. 2 and 5). In addition, the amino acid composition of isolated mP2 precursor 1 very closely matches that predicted from pmP2-1 (Table 1). This information allowed us to determine that the length of the mP2 precursor is 106 amino acids, with a calculated molecular mass of 13,509 daltons.

Amino-terminal sequence of mature mP2. Previous amino acid composition data for mP2 isolated from mouse spermatozoa (Bellvé and Carraway, J. Cell. Biol. 79:177a, 1978) indicated that mP2 contains about 62 amino acids and that it does not contain threonine residues. This allowed us to determine that the final processing site for the mP2 precursor is located somewhere after the threonine residue at position 42 in the mP2 precursor sequence (Fig. 2). Confirmation of the amino terminus of mature mP2 was obtained by sequencing the first 20 amino acids of mP2 isolated from sperm and purified by reverse-phase HPLC (Fig. 3). This amino acid sequence exactly matches residues 44 to 63 of mP2 precursor 1 as predicted from the pmP2-1 cDNA (Fig. 2 and 5). Determination of the mature mP2 amino terminus located the final processing point of the precursor between histidine-43 and arginine-44 of the mP2 precursor. Processing and subsequent removal of the 43-amino-acid leader from the mP2 precursor results in the formation of a 63-amino-acid mature mP2 with a calculated molecular mass of 8,595 daltons. A comparison of the amino acid composition predicted from the mature portion of the pmP2-1 nucleotide sequence to that previously published for mP2 (Bellvé and Carraway, J. Cell Biol. 79:177a, 1978) yields similar results. The numbers of arginines (32 and 35 to 36), histidines (13 and 12), lysines (3 and 3), cysteines (7 and 4), serines (4 and 3), glycines (2 and 2), leucines (1 and 1), and isoleucines (1 and 1) match very closely.

Comparison of mP1 and mP2 cDNAs. In confirmation of our previous observation that the cDNAs for mP1 and mP2 (previously called HSAR 700 and HSAR 900) do not cross-

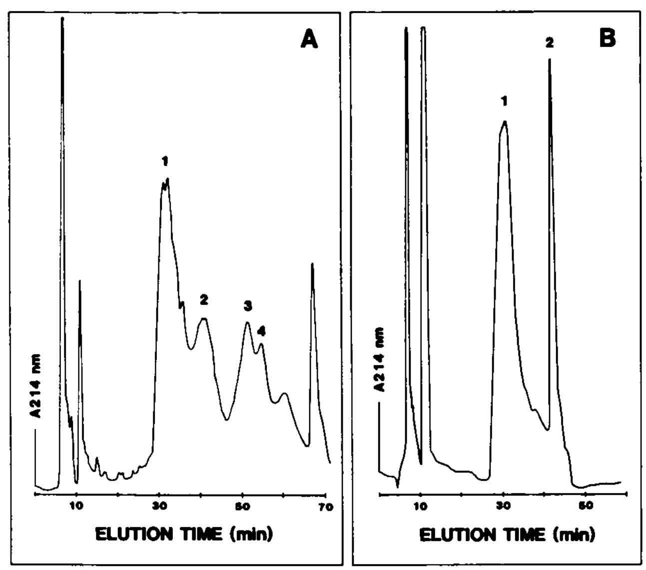

FIG. 3. HPLC separation of mouse late-step spermatid (A) and caudal epididymal sperm (B) basic proteins. After reduction with dithiothreitol, 1 to 2 mg of total protein was chromatographed on a Nucleosil RP-C18 reverse-phase column as described in Materials and Methods, and the numbered peaks were collected for analysis. Peak 1, mP2; peak 2, mP1; peak 3, mixture of precursors 2, 3, and 4; peak 4, precursor 1.

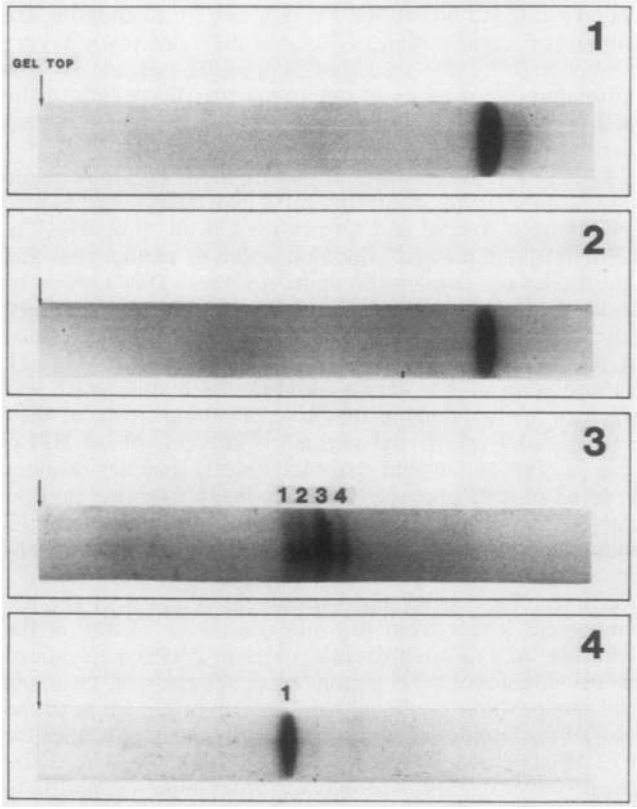

FIG. 4. Electrophoretic separation of HPLC-fractionated, spermatid-specific basic proteins. Fractions were collected (Fig. 3A), lyophilized, and dissolved in 20% sucrose–0.9% acetic acid–0.5 M mercaptoethanol. Samples were electrophoresed at 130 V for 1.5 h in acid-urea gels (5) containing 2.5 M urea, stained with naphthol blue-black, and destained electrophoretically. The direction of electrophoresis was from left (+) to right (−). Gel patterns are shown for each of the four HPLC peaks. The precursor proteins are labeled 1 through 4 based on their electrophoretic mobilities.

hybridize (22), sequence comparison of the cDNAs coding for mP1 and mP2 detected little homology at the nucleotide level. The cDNAs for mP1 and mP2 both contain the canonical signal AAUAAA believed to be necessary for polyadenylation of the nuclear transcript, a sequence absent from most histone mRNAs (19). There is one region of significant nucleotide homology located in the 3′ untranslated region of both mouse protamine cDNAs. Fifteen of seventeen nucleotides (and 10 of 10 nucleotides) directly upstream of the polyadenylation signal are identical in the two mouse protamine cDNAs, suggesting an important role for this sequence in protamine mRNA metabolism (Fig. 2). In the mP2 cDNA, part of this sequence is repeated again in the 3′ untranslated region upstream from the first (Fig. 2B).

The cDNA sequences for the two mouse protamines and the availability of nucleotide sequences for the multiple trout protamines allowed the codon usage preference for these proteins to be compared within and between species (Table 2). In mice, all six codons for arginine are utilized for both mP1 and mP2, with mP2 showing a strong preference for AGG. In trout, CGC and AGG are the most utilized arginine codons (13, 37). Both mP1 and mP2 favor the cysteine codon UGC over UGU. The mP1 gene and every trout protamine gene analyzed terminates with the codon UAG, whereas mP2 utilizes the codon UAA for termination. For serine, trout protamines utilize the codon UCC exclusively,

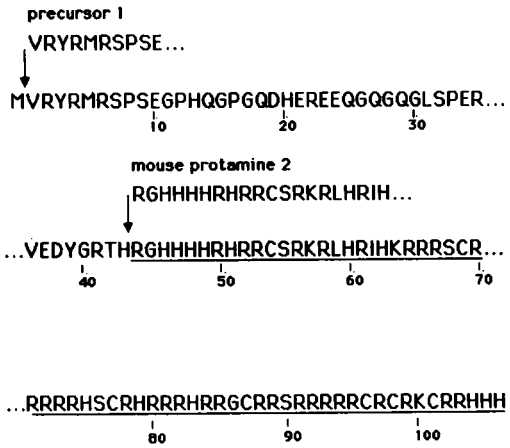

FIG. 5. Amino-terminal amino acid sequences for mP2 and its precursor. The amino-terminal sequences for the precursor and mP2 were determined as described in Materials and Methods. The complete primary sequence for the mP2 precursor was determined from the mP2 cDNA. The primary sequence corresponding to the mature (processed) mP2 molecule is underlined.

whereas four of the six possible codons that code for serine are found in the two mouse protamine sequences. The predicted amino acid sequence encoded by the mP2 cDNA depicts a precursor protein whose 43-residue leader is very different in charge and composition from the very basic mature portion of the molecule. A high glutamic and aspartic acid content and the presence of proline and valine residues distinguish this portion of the molecule from mature mP2. Although the amino acids methionine, proline, and valine present in the precursor of mP2 are absent in mouse protamines, they are present in the trout protamines.

DISCUSSION

Analysis of in vivo-labeled arginine-rich mouse testicular proteins by gel electrophoresis has demonstrated that two protamine variants and four lower-mobility proteins are synthesized during stages 12 to 15 of spermiogenesis (6). The latter group of proteins, termed transition proteins, contain

TABLE 1. Amino acid composition of mP1, mP2, and the mP2 precursor

Amino acid	Amt (mol/100 mol of amino acid) in:			
	mP1	mP2	mP2 precursor 1	mP2 precursor 1 from cDNA
Arginine	56.0	50.8	42.7	36.2
Lysine	6.0	0	3.5	2.9
Histidine	0	20.6	12.2	14.3
Tyrosine	6.0	0	2.2	1.9
Leucine	0	1.6	1.8	1.9
Isoleucine	2.0	1.6	1.0	1.0
Alanine	2.0	0	0.4	0
Glycine	0	3.2	7.4	8.6
Glutamic acid	0	4.8	8.5	10.5
Serine	8.0	6.3	5.8	6.7
Threonine	2.0	0	0.9	1.0
Valine	.0	0	1.1	1.9
Aspartic acid	0	0	2.1	1.9
Methionine	0	0	0.4	1.0
Proline	0	0	3.1	3.8
Cysteine	18.0	11.1	6.8	6.7

TABLE 2. Codon usage for mouse and trout protamines

Codon	Amino acid	No. used for:				
		Mouse			Trout protamines	
		Mature mP1[a]	Mature mP2	Precursor mP2	Source 1[b]	Source 2[c]
UUU	Phe	0	0	0	0	0
UUC	Phe	0	0	0	0	0
UUA	Leu	0	0	0	0	0
UUG	Leu	0	0	0	0	0
CUU	Leu	0	0	0	0	0
CUC	Leu	0	0	0	0	0
CAU	Leu	0	1	1	0	0
CUG	Leu	0	0	1	0	0
AUU	Ile	0	0	0	0	0
AUC	Ile	0	1	1	3	3
AUA	Ile	1	0	0	0	0
AUG	Met	0	0	1	5	3
GUU	Val	0	0	1	0	1
GUC	Val	0	0	0	7	4
GUA	Val	0	0	1	0	0
GUG	Val	0	0	0	10	8
UCU	Ser	0	1	1	0	0
UCC	Ser	0	2	2	3	16
UCA	Ser	1	1	1	0	0
UCG	Ser	0	0	0	0	0
CCU	Pro	0	0	1	4	4
CCC	Pro	0	0	1	15	11
CCA	Pro	0	0	1	0	0
CCG	Pro	0	0	1	0	0
ACU	Thr	0	0	0	0	0
ACC	Thr	1	0	0	0	0
ACA	Thr	0	0	1	0	0
ACG	Thr	0	0	0	0	0
GCU	Ala	0	0	0	0	0
GCC	Ala	1	0	0	7	4
GCA	Ala	0	0	0	0	0
GCG	Ala	0	0	0	0	0
UAU	Tyr	0	0	1	0	0
UAC	Tyr	3	0	1	0	0
UAA	Termination	0	1	1	0	0
UAG	Termination	1	0	0	10	8
CAU	His	0	3	4	0	0
CAC	His	0	10	12	0	0
CAA	Gln	0	0	2	0	0
CAG	Gln	0	0	3	0	0
AAU	Asn	0	0	0	0	0
AAC	Asn	0	0	0	0	0
AAA	Lys	3	1	1	0	0
AAG	Lys	0	2	2	0	0
GAU	Asp	0	0	0	0	0
GAC	Asp	0	0	2	0	0
GAA	Glu	0	0	1	0	0
GAG	Glu	0	0	5	0	0
UGU	Cys	2	1	1	0	0
UGC	Cys	7	6	6	0	0
UGA	Termination	0	0	0	0	0
UGG	Trp	0	0	0	0	0
CGU	Arg	2	2	2	35	27
CGC	Arg	7	3	6	66	52
CGA	Arg	4	1	2	8	8
CGG	Arg	3	3	3	7	5
AGU	Ser	0	0	1	0	0
AGC	Ser	3	0	2	8	6
AGA	Arg	6	7	7	26	18
AGG	Arg	6	16	18	50	40
GGU	Gly	0	0	1	0	0
GGC	Gly	0	2	2	10	8
GGA	Gly	0	0	1	10	8
GGG	Gly	0	0	5	0	0

[a] From Kleene et al. (24).
[b] From Mori et al. (37).
[c] From Gedamu et al. (13).

cysteine and disappear from the spermatid chromatin before the maturing gamete enters the caput epididymis. It has been proposed (6) that the transition proteins represent a form of precursor protamine or play a role in organizing protamine deposition onto DNA or both. The inability to detect cell-free translation products directed by mouse protamine mRNAs which comigrate with isolated protamine markers has led others also to propose that the mouse protamines are synthesized in a form other that that found in spermatozoa (11; E. D. Wieben, Ph.D. dissertation, Yale University, New Haven, Conn., 1979). Our recent sequence data for a cDNA of mP1 has established that the mP1 molecule is not synthesized as a precursor protein. However, the nucleotide sequence presented in this report indicates that mP2 is synthesized as an appreciably longer molecule than that found in mature sperm. Furthermore, the amino acid composition of the largest of the four transition proteins matches closely the amino acid composition predicted from the mP2 cDNAs. The first AUG in the pmP2-1 cDNA sequence is the initiation codon. A low but reproducibly detectable level of methionine was present in the protein fraction analyzed, and more importantly, sequence analysis of the first 10 amino acids of the predominant transition protein revealed the N-terminal amino acid to be valine.

Although both mouse protamines are synthesized in late spermatids, the tyrosine variant, mP1, is detectable in epididymal sperm approximately 1 day before the appearance of the histidine variant, mP2 (6). This observation, along with the fact that mP2 precursor 1 can be isolated from sonication-resistant spermatids, indicates that the precursor enters the nucleus and binds to the condensing DNA. Subsequent cleavage and removal of the 43-amino-acid leader results in the formation of mature mP2. The presence of a sizable pool of mP2 precursor and three other closely related transition proteins detectable by naphthol blue-black staining of basic proteins isolated from late-step spermatids (precursors 2, 3, and 4 in Fig. 3 and 4) suggests that mP2 precursor processing does not occur immediately after translation. Amino acid composition analysis of these four proteins (data not shown) suggested that they represent processing intermediates or different charge variants of the original mP2 precursor.

The regulatory significance of posttranslational processing for the mP2 molecule is unclear. mP2 precursor processing may facilitate the ordered deposition of the protamines onto condensing DNA. The fact that one but not both of the mouse protamine variants is synthesized as a precursor is surprising, although it may be a reflection of cellular protein turnover mechanisms. Bachmair et al. (4) have recently suggested that the half-life of a protein may be related to its amino-terminal amino acid. They reported half-lives ranging from less than 3 min for proteins whose N-terminal amino acid is arginine, the N-terminal amino acid of mature mP2, to 20 h for proteins whose N-terminal amino acid is alanine or valine, the N-terminal amino acid for mP1 and the mP2 precursor. The additional requirement for processing of mP2 may explain the apparent evolutionary preference for P1 protamine in the spermatozoa of many mammals.

Although the two mouse protamines differ greatly in nucleotide sequence and amino acid composition, they share many common properties. Unlike the protamine multigene family in trout (1, 18), both mouse protamine genes are single-copy genes located on mouse chromosome 16 (17). The protamine genes are translationally regulated (15); mP1 and mP2 mRNAs are first detected in RNA isolated from the testes of prepuberal, 22-day-old CD1 mice (16) and in RNA isolated from round spermatids (23). The protamines are first synthesized approximately 8 days later during stages 12 to 15 of spermiogenesis (6, 27, 31; Hecht, in press). The marked shortening of the mP1 mRNA by deadenylation of approximately 130 bases at the time of translation (22) was also seen with mP2 mRNA (data not shown).

It is surprising that the two mouse protamines are so different in size and sequence and that the mP2 precursor requires posttranslational processing, whereas mP1 does not. Although both contain high levels of arginine, the arginine residues of mP2 are scattered throughout the entire molecule and not clustered in the central region of the protein as in P1 vertebrate protamines (24; Hecht, in press). mP1 lacks histidine, whereas mP2 contains 13 histidine residues (greater than 20%). mP1 contains three tyrosine residues, whereas mP2 contains none. These marked differences in composition suggest that the two mouse protamines may not have evolved from a common sequence. However, since the CAC and CAU codons for histidine can be derived from arginine and tyrosine codons by a single base-pair substitution, the possibility that both mP1 and mP2 originated from a common primordial sequence cannot be discounted. One notable similarity between the two protamines is the location and spacing of cysteine residues throughout the molecules. Alignment of cysteine residues may be important for proper interaction, cross-linking, and deposition of mP1 and the mP2 precursor onto the DNA during nuclear condensation.

Southern blot analyses have indicated nucleotide homology between genomic DNA from diverse vertebrate species and mP1 and mP2 cDNAs (Bower et al., in press). As predicted from the known amino acid sequence of P1 protamines, mP1 cDNA should and does hybridize to genomic DNA restriction fragments from many different mammals including rats, bulls, boars, horses, rams, dogs, hamsters, and humans (Bower et al., in press; unpublished data). Under hybridization conditions in which the cDNAs for mP1 and mP2 do not cross-hybridize, mP2 cDNA hybridizes to genomic DNA of rats, hamsters, and chickens (unpublished data).

It is intriguing that mammalian spermatozoa contain such variable amounts of the two protamine variants. The comparison of protamine expression in rats and mice demonstrates this well. mP1 constitutes about 30% of the total protamine of mouse epididymal spermatozoa (6). Assuming equal distribution of the two variants among individual spermatozoa, mP2 is the predominant form of protamine in mouse sperm. In contrast, only P1 protamine is detectable by electrophoresis of rat spermatozoan proteins (9). Recently, however, hybridization studies have demonstrated that RNA transcripts homologous to and of the same size as mP2 mRNA are present on rat testis polysomes at a very low level (Bower et al., in press). Clearly a more complete understanding of the mechanism(s) for the expression and processing of the P2 protamine will clarify the differential phylogenetic expression of this male germ-cell-specific protein.

ACKNOWLEDGMENTS

The mouse total testis cDNA library used in our screening was generously provided by Lee Silver. We thank G. R. Green and R. McIsaac for helpful comments and critical review of the manuscript and S. E. Yelick for expertise in computer graphics.

This work was supported by Public Health Service grant GM 29224 from the National Institutes of Health.

LITERATURE CITED

1. **Aiken, J. M., D. McKenzie, H. Z. Zhao, J. C. States, and G. H. Dixon.** 1983. Sequence homologies in the protamine gene family of rainbow trout. Nucleic Acids Res. **11**:4907–4992.
2. **Ammer, H., A. Henschen, and C. Lee.** 1986. Isolation and amino acid sequence analysis of human sperm protamines P1 and P2. Occurrence of two forms of protamine P2. Biol. Chem. Hoppe-Seyler **367**:515–522.
3. **Ando, T., and S. Watanabe.** 1969. A new method for fractionation of protamines and the amino acid sequence of salmine and three components of iridine. Int. J. Protein Res. **1**:221–224.
4. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234**:179–186.
5. **Balhorn, R., B. L. Gledhill, and A. J. Wyrobek.** 1977. Mouse sperm chromatin proteins: quantitative isolation and partial characterization. Biochemistry **16**:4074–4080.
6. **Balhorn, R., S. Weston, C. Thomas, and A. J. Wyrobek.** 1984. DNA packaging in mouse spermatids. Synthesis of protamine variants and four transition proteins. Exp. Cell Res. **150**:298–308.
7. **Bedford, J. M., and H. I. Calvin.** 1974. The occurrence and possible significance of -S-S- crosslinks in sperm heads with particular reference to eutherian mammals. J. Exp. Zool. **188**:137–156.
8. **Bellvé, A. R., E. Anderson, and L. Hanley-Bowdoin.** 1975. Synthesis and amino acid composition of basic proteins in mammalian sperm nuclei. Dev. Biol. **47**:349–365.
9. **Calvin, H. I.** 1976. Comparative analysis of the nuclear basic proteins in rat, human, guinea pig, mouse and rabbit spermatozoa. Biochim. Biophys. Acta **434**:377–389.
10. **Coelingh, J. P., C. H. Monfoort, T. H. Rozijin, J. A. Gevers-Leuven, R. Schiphof, E. P. Steyn-Parve, G. Braunitzer, B. Schrank, and A. Ruhfus.** 1972. The complete amino acid sequence of the basic nuclear protein of bull spermatozoa. Biochim. Biophys. Acta **285**:1–14.
11. **Elsevier, S. M.** 1982. Messenger RNA encoding basic chromosomal proteins of mouse testis. Dev. Biol. **90**:1–12.
12. **Gaastra, W., J. Lukkes-Hofstra, and A. H. J. Kolk.** 1978. Partial covalent structure of two basic chromosomal proteins from human spermatozoa. Biochem. Genet. **16**:525–529.
13. **Gedamu, L., M. A. Wosnick, W. Connor, D. C. Watson, G. H. Dixon, and K. Iatrou.** 1981. Molecular analysis of the protamine multigene family in rainbow trout testis. Nucleic Acids Res. **9**:1463–1482.
14. **Grimes, S. R., Jr., M. L. Meistrich, R. D. Platz, and L. S. Hnilica.** 1977. Nuclear protein transitions in rat testis spermatids. Exp. Cell Res. **110**:31–39.
15. **Hecht, N. B.** 1986. Regulation of gene expression during mammalian spermatogenesis. p. 151–193. *In* J. Rossant and R. Pedersen (ed.), Experimental approaches to mammalian embryonic development. Cambridge University Press, New York.
16. **Hecht, N. B., P. A. Bower, S. H. Waters, P. C. Yelick, and R. J. Distel.** 1986. Evidence for haploid expression of mouse testicular genes. Exp. Cell Res. **164**:183–190.
17. **Hecht, N. B., K. C. Kleene, P. C. Yelick, P. A. Johnson, D. D. Pravtcheva, and F. H. Ruddle.** 1986. Mapping of haploid expressed genes: genes for both mouse protamines are located on chromosome 16. Somatic Cell Mol. Genet. **12**:203–208.
18. **Iatrou, K., and G. H. Dixon.** 1977. The distribution of poly $(A)^+$ and poly $(A)^-$ protamine messenger RNA sequences in the developing trout testis. Cell **10**:433–441.
19. **Kedes, L. H.** 1979. Histone genes and histone messengers. Annu. Rev. Biochem. **48**:837–870.
20. **Kistler, W. S., M. E. Geroch, and H. G. Williams-Ashman.** 1973. Specific basic proteins from mammalian testes. J. Biol. Chem. **248**:4532–4543.
21. **Kistler, W. S., P. S. Keim, and R. L. Heinrickson.** 1976. Partial structural analysis of the basic chromosomal protein of rat spermatozoa. Biochim. Biophys. Acta **427**:752–757.
22. **Kleene, K. C., R. J. Distel, and N. B. Hecht.** 1983. cDNA clones encoding cytoplasmic poly$(A)^+$ RNAs which first appear at detectable levels in haploid phases of spermatogenesis in the mouse. Dev. Biol. **98**:455–464.
23. **Kleene, K. C., R. J. Distel, and N. B. Hecht.** 1984. Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. Dev. Biol. **105**:71–79.
24. **Kleene, K. C., R. J. Distel, and N. B. Hecht.** 1985. Nucleotide sequence of a cDNA clone encoding mouse protamine 1. Biochemistry **24**:719–722.
25. **Kolk, A., and T. Samuel.** 1975. Isolation, chemical and immunological characterization of two strongly basic nuclear proteins from human spermatozoa. Biochim. Biophys. Acta **393**: 307–319.
26. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
27. **Lam, D. M. K., and W. R. Bruce.** 1971. The biosynthesis of protamine during spermatogenesis of the mouse: extraction, partial characterization, and site of synthesis. J. Cell. Physiol. **78**:13–24.
28. **Ling, V., B. Jergil, and G. H. Dixon.** 1971. The biosynthesis of protamine in trout testis. III. Characterization of protamine components and their synthesis during testis development. J. Biol. Chem. **246**:1168–1176.
29. **Loir, M., and M. Lanneau.** 1975. An electrophoretic analysis of the basic nuclear proteins of ram spermatids. Exp. Cell Res. **92**:509–512.
30. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
31. **Mayer, J. F., Jr., T. S. K. Chang, and B. R. Zirkin.** 1981. Spermatogenesis in the mouse. 2. Amino acid incorporation into basic nucleoproteins of mouse spermatids and spermatozoa. Biol. Reprod. **25**:1041–1051.
32. **Mazrimas, J. A., M. Corzett, C. Campos, and R. Balhorn.** 1986. A corrected primary sequence for bull protamine. Biochim. Biophys. Acta **872**:11–15.
33. **McKay, D. J., B. S. Renaux, and G. H. Dixon.** 1985. The amino acid sequence of human sperm protamine. P1. Biosci. Rep. **5**:383–391.
34. **McKay, D. J., B. S. Renaux, and G. H. Dixon.** 1986. Human sperm protamines. Amino-acid sequences of two forms of protamine P2. Eur. J. Biochem. **156**:5–8.
35. **Meistrich, M. L., P. K. Trostle, and W. A. Brock.** 1981. Association of nucleoprotein transitions with chromatin changes during rat spermatogenesis, p. 151–166. *In* G. Jagiello and H. J. Vogel (ed.), Bioregulators of reproduction. Academic Press, Inc., New York.
36. **Monfoort, C. H., R. Schiphof, T. H. Rozijn, and E. P. Steyn-Parvé.** 1973. Amino acid composition and carboxyl-terminal structure of some basic chromosomal proteins of mammalian spermatozoa. Biochim. Biophys. Acta **322**:173–177.
37. **Mori, N., Y. Funatsu, K. Hiruta, and S. Goto.** 1985. Analysis of translational fidelity of ribosomes with protamine messenger RNA as a template. Biochemistry **24**:1231–1239.
38. **Nakano, M., T. Tobita, and T. Ando.** 1973. Studies on a protamine (galline) from fowl sperm. 1. Fractionation and some characterization. Int. J. Pept. Protein Res. **5**:149–159.
39. **Puwaravutipanich, T., and S. Panyim.** 1975. The nuclear basic proteins of human testis and ejaculated spermatozoa. Exp. Cell Res. **90**:153–158.
40. **Rodman, T. C., F. H. Pruslin, and V. G. Allfrey.** 1984. Protamine-DNA association in mammalian spermatozoa. Exp. Cell Res. **150**:269–281.
41. **Sautière, P., D. Belaiche, A. Martinage, and M. Loir.** 1984. Primary structure of the ram (*Ovis aries*) protamine 1. Eur. J. Biochem. **144**:121–125.
42. **Schwarzbauer, J. E., J. W. Tankur, I. Lemischka, and R. O. Hynes.** 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. Cell **35**:421–431.
43. **Tobita, T., H. Tsutsumi, A. Kato, H. Suzuki, M. Nomoto, M. Nakano, and T. Ando.** 1983. Complete amino acid sequence of boar protamine. Biochim. Biophys. Acta **744**:141–146.