

- Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., & Rich, A. (1981) *Science (Washington, D.C.)* 211, 171-176.
- Widom, J., & Baldwin, R. L. (1980) *J. Mol. Biol.* 144, 431-453.
- Widom, J., & Baldwin, R. L. (1983a) *Biopolymers* 22, 1595-1620.
- Widom, J., & Baldwin, R. L. (1983b) *Biopolymers* 22, 1621-1632.
- Wilson, R. W., & Bloomfield, V. A. (1979) *Biochemistry* 18, 2192-2196.
- Zacharias, W., Larson, J. E., Klysik, J., Stirdivant, S. M., & Wells, R. D. (1982) *J. Biol. Chem.* 257, 2775-2782.
- Zacharias, W., Martin, J. C., & Wells, R. D. (1983) *Biochemistry* 22, 2398-2405.
- Zimmerman, S. B. (1982) *Annu. Rev. Biochem.* 51, 395-427.

## Nucleotide Sequence of a cDNA Clone Encoding Mouse Protamine 1<sup>†</sup>

Kenneth C. Kleene, Robert J. Distel, and Norman B. Hecht\*

Department of Biology, Tufts University, Medford, Massachusetts 02155

Received July 5, 1984

**ABSTRACT:** The nucleotide sequence of a 404-base cDNA encoding the cysteine-rich, tyrosine-containing mouse protamine has been determined. This insert, isolated from a mouse testis cDNA library, encodes a polypeptide of 50 amino acids of which 28 are arginine, 9 are cysteine, and 3 are tyrosine. The insert contains the complete 3' noncoding region of 151 bases and most of the 5' noncoding region. The predicted amino acid sequence of mouse protamine 1 is about 80% homologous to boar protamine and 67% homologous to bull protamine and contains the central, highly basic domain of four arginine clusters found in the trout protamines. The identification of a cDNA clone for a mouse protamine will facilitate studies of the evolution, regulation, and protein-DNA interaction of this nuclear protein unique to haploid spermatogenic cells.

**D**uring spermatogenesis in vertebrates, the histones on chromatin are replaced by protamines, basic proteins 27-65 amino acids long containing 40-70% arginine and relatively little lysine (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978; Nakano et al., 1976; Sautière et al., 1981). This change in chromatin, occurring late in the haploid phase of sperm differentiation, produces a DNA-protamine complex that is extremely compact, devoid of nucleosomal structure, and inactive in RNA synthesis (Kierszenbaum & Tres, 1975). The protamines of placental mammals (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978) are distinct from those of fish and birds (Nakano et al., 1976; Sautière et al., 1981) because they contain 8-18% cysteine, which serves to further condense and stabilize the mammalian sperm nucleus by forming disulfide bridges. Although the expression and evolution of the protamine gene family in the trout have been extensively studied by Dixon and his colleagues (Iatrou & Dixon, 1977; Aiken et al., 1983), studies of the protamines in mammals have been restricted to the synthesis, structure, and function of these unique proteins (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978; Rodman et al., 1984; Bellvé et al., 1975; Goldberg et al., 1977; Mayer et al., 1981; Balhorn et al., 1984).

We have previously isolated a cDNA clone from a mouse testis cDNA library and shown by RNA blots that it is complementary to a small, highly abundant, testis-specific mRNA (Kleene et al., 1983). This mRNA is translationally regulated: it is first present at high levels in round spermatids but is first translated in elongating spermatids (Kleene et al., 1983, 1984), the stage when protamines are synthesized in mice (Bellvé et

al., 1975; Goldberg et al., 1977; Mayer et al., 1977; Balhorn et al., 1984). The properties of this mRNA led us to suspect that it encoded a protamine. This suspicion has been confirmed in the present study by DNA sequencing, which demonstrates that this cDNA clone (pMP1-1) encodes mouse protamine 1, the cysteine-rich, tyrosine-containing protamine.

### MATERIALS AND METHODS

Clone pMP1-1 was isolated from a mouse testis cDNA library in pUC8 by colony hybridization using a nick-translated insert from the plasmid pIC3 as a probe (Kleene et al., 1983, 1984). Plasmid pIC3 was 1 of 17 of 750 clones from a testis cDNA library that hybridized much more strongly to <sup>32</sup>P-labeled cDNA copied from cytoplasmic poly(A<sup>+</sup>) RNA of round spermatids than pachytene spermatocytes (Kleene et al., 1983). The nucleotide sequence of the cDNA inserts was determined on both strands by the method of Maxam & Gilbert (1978). The 3' and 5' termini of restriction fragments were labeled by using the Klenow fragment of *Escherichia coli* DNA polymerase I and T4 polynucleotide kinase, respectively (Maxam & Gilbert, 1978). In some cases, the complementary strands were separated by denaturation with methylmercury and electrophoresis on 5-8% acrylamide gels (Schwarzbauer et al., 1983).

### RESULTS AND DISCUSSION

The nucleotide sequence of both strands of the 404-base insert of pMP1-1 was determined by using the strategy depicted in Figure 1. There was complete agreement in the sequences determined from each strand. The first initiation codon (AUG) is located 78 bases from the 5' end of pMP1-1 and appears in the context of a sequence (GCACCAUGG) that closely matches the consensus sequence (CC<sub>6</sub>CCAUGG) for initiation of protein synthesis in eukaryotic cells (Kozak, 1984). The amino acid sequence deduced from the following nucleotide sequence encodes a polypeptide of 50 amino acids

<sup>†</sup> This work was supported by USPHS Grant GM 29224 to N.B.H.

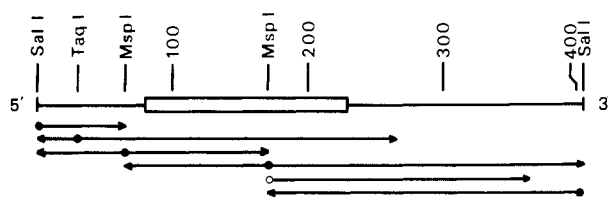


FIGURE 1: Partial restriction map and DNA sequencing strategy used for pMP1-1. DNA fragments labeled at the 3' (●) or 5' (○) ends were sequenced by the method of Maxam & Gilbert (1978). The rectangle indicates the translated portion of the protamine mRNA.

with a calculated molecular weight of 6825 (Figure 2). Although the sperm of most mammals contain a single species of protamine (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978), there are two protamines in the mouse (Bellvé & Carraway, 1978; Bellvé, 1979; Mayer et al., 1981; Balhorn et al., 1984): protamine 1, a polypeptide estimated to have 55 amino acids including 4 tyrosines, 9 cysteines, and no histidines, and protamine 2, a polypeptide of about 65 amino acids including 12 histidines, 3 cysteines, and no tyrosines (Bellvé, 1978). The predicted amino acid composition of the protamine encoded by pMP1-1 in Figure 2 agrees almost perfectly with the composition of protamine 1 (Bellvé, 1979; Bellvé & Carraway, 1978) in the numbers of arginine (28 vs. 32–33), cysteine (9 vs. 9), tyrosine (3 vs. 4), serine (4 vs. 4), lysine (3 vs. 3), and alanine, threonine, and isoleucine (1 vs. 1). The 3' untranslated sequence of MP1-mRNA is 151 bases excluding the poly(A) tail. The total length of the 5' un-

translated sequence of the mRNA for protamine 1 can be deduced to be about 113 bases from the combined length of the coding and 3' untranslated region of pMP1-1 excluding the poly(A) and the size of protamine mRNA after deadenylation with RNase H, 420–307 bases (Kleene et al., 1984). Hence, pMP1-1 lacks about 30 bases at the 5' end.

One ambiguity in the coding sequence should be noted. When the complementary strands of the internal 105-base Msp I fragment of pMP1-1 were denatured with methylmercury and separated on an 8% polyacrylamide gel, we unexpectedly obtained four fragments corresponding to two variants that differed at two bases. One contained AGG CGA (Arg Arg) for amino acids 26 and 27 as shown in Figure 2, whereas the second contained TGG TGA (Trp Stop) at the same positions. The two remaining fragments were the appropriate complementary strands of each variant. To resolve this ambiguity, we also sequenced the 310 bases from the first Msp I site to the start of the poly(A) tail of another insert, pMP1-2. This 310-base section of the insert of pMP1-2 was identical at every base to the sequence shown in Figure 2, and there was no trace of the TGG TGA variant in separated strands of the internal Msp I fragment. Since the sequences shown in Figure 2 appear in both pMP1-1 and pMP1-2 and the predicted size and amino acid composition shown in Figure 2 agree with analyses of the protamine 1 protein (Bellvé, 1979; Bellvé & Carraway, 1978), we believe that the TGG TGA variant is a recombinant DNA artifact.

The complete amino acid sequences of mouse 1, bull (Coelingh et al., 1972), boar (Tobita et al., 1983), and trout

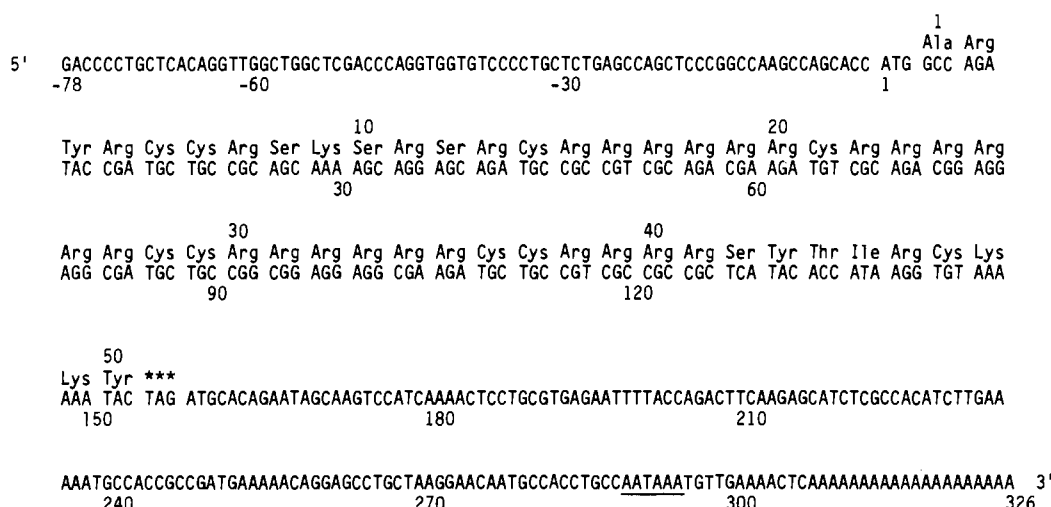


FIGURE 2: Nucleotide sequence and corresponding amino acid sequence of the cDNA insert of pMP1-1. The deduced amino acid sequence is presented above the nucleotide triplets. Amino acids are numbered beginning with alanine because methionine does not appear at the amino terminus of protamines (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978). The termination codon, TAG, is denoted by asterisks. The canonical hexanucleotide, AATAAA, is underlined.

	10	20	30	40	50
Mouse I	A R Y R C C R S K S R S R C R R R R R	R C R R R R R R R	C C R R R R R R R	C C R R R R S Y T	I R C K K Y
Rat	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -
Boar	- - - - - H - - - - - P - - - - -	- - - - -	- - P - - - - - A V - - - - -	- - V - - - - R R C	
Bull	- - - - - L T H - G - - - - -	- R C - - - - -	R F G - - - -	- - - V C - - V - - - - T R Q	
Human I	- - - - - ? A - - - Q - - - - Y Y ? Q ? Q				M Y H R R R C
Trout	M P - - - S * * * P V - - - - - * R V S - - - - -			G G - - - - -	

FIGURE 3: Amino acid sequence homologies between the tyrosine-rich mouse, bull (Coelingh et al., 1972), boar (Tobita et al., 1983), rat (Kistler et al., 1976), human I (Gaastra et al., 1978), and trout (Aiken et al., 1983) protamines. Gaps were introduced to maximize the homologies between the mouse, bull, boar, and trout protamines. Regions of nucleotide heterogeneity in six different trout protamine genes are represented by asterisks. Amino acids that are the same as the mouse are denoted by a dash. The amino acids of the mouse protamine have been numbered. A, alanine; C, cysteine; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; Y, tyrosine.

(Aiken et al., 1983) protamines and partial amino acid sequences for rat (Kistler et al., 1976) and human I (Gaastra et al., 1978) protamines are compared in Figure 3. The protamines of mouse, bull, and boar are organized into three domains, a central highly basic core consisting of clusters of arginines and two less basic regions at the amino and carboxy termini. The central basic domain contains three to four clusters of four to seven arginines that are separated by neutral amino acids, cysteine, proline, phenylalanine, or glycine. These arginine clusters are involved in the binding of protamines to DNA. The protamines of mouse, boar, and bull contain six to nine cysteines distributed throughout each polypeptide. This high content of cysteines is typical of mammalian protamines (Coelingh et al., 1972; Bellvé, 1979; Bellvé & Carraway, 1978; Kistler et al., 1976; Tobita et al., 1983; Gaastra et al., 1978) and results in the cross-linking of DNA-protamine complexes by S-S bonds in the epididymis (Marushige & Marushige, 1975).

The sequences at the amino termini of the five mammalian protamines show considerable similarity, being homologous at the first four amino acids, and at a minimum of 10 of the first 15 amino acids. The conserved amino acids at the amino terminus include three serines at amino acids 8, 10, and 12 in four of five species, the exception being a threonine at amino acid 8 in the bull. These serines are thought to be important sites for phosphorylation involved in regulating the binding of protamines to DNA (Marushige & Marushige, 1975). The threonine at amino acid 8 in the bull should be regarded as a potential site for phosphorylation too, because threonine is phosphorylated in histones (Mathews & Huebner, 1984). As pointed out previously, the amino acid sequence of the carboxy terminus of protamines is more variable than that of the amino terminus (Monfoort et al., 1973). Nevertheless, five of the last nine amino acids are common to bull, boar, and mouse I. Overall, the homology of the bull and mouse protamines is somewhat lower (about 67%) than that of the bull and boar and boar and mouse protamines (about 80%). Since the sequence of the amino terminus of human protamine 2 (Gaastra et al., 1978) and the amino acid composition of mouse protamine 2 (Bellvé, 1979; Bellvé & Carraway, 1978) differ markedly from the protamines shown in Figure 3, their structure and evolution should be especially interesting.

Several unidentified chromatin proteins that copurify with protamines, but are of larger size, are synthesized in late spermatids *in vivo* (Balhorn et al., 1984) and *in vitro* by using RNA purified from mouse testes (Elsevier, 1982). These proteins, which disappear from chromatin before the spermatozoon enters the epididymis, have been speculated to play a role in initiating chromatin condensation or represent "precursor protamines" (Balhorn et al., 1984; Elsevier, 1982). It is unlikely that one of these proteins is a precursor to protamine 1 because the size and amino acid composition of the protein encoded by protamine 1 mRNA agree with analyses of the mature protein purified from spermatozoa (Bellvé, 1979; Bellvé & Carraway, 1978).

The comparison of the amino acid sequence of mouse protamine 1 and the trout protamines reveals several similarities and differences. Although the protamines from trout and mouse protamine 1 differ significantly in length (33 vs. 50 amino acids), the central, highly basic region containing four arginine clusters has been conserved (Figure 3). The trout protamines lack much of the N-terminal and C-terminal regions present in mammalian protamines, suggesting much of protamine function can be accomplished by this domain. Sequence analysis of the family of trout protamine DNA

sequences has demonstrated that although all six codons encoding arginine are used, CGC and AGG are the most frequently favored codon assignments for this amino acid (Gedamu et al., 1981). Similar results are found for mouse protamine 1 (Figure 2). In contrast, the codon heavily favored for serine in trout protamines, UCC, does not encode any serine in the mouse protamine 1.

Despite the fact that expression of protamine mRNAs in both species is regulated at the level of translation (Iatrou & Dixon, 1977; Kleene et al., 1984), the 5' untranslated sequences differ markedly in length, 14–19 bases in the trout (Aiken et al., 1983) vs. about 113 bases in the mouse. The 3' untranslated regions of mouse and trout protamine mRNAs (Aiken et al., 1983) both contain the canonical signal (AAUAAA) believed necessary for polyadenylation and/or cleavage of the transcript (Fitzgerald & Shenk, 1981; Montell et al., 1983). This is relatively unusual in mRNAs for basic nuclear proteins (Birchmeir et al., 1983) and implies that processing of the 3' ends of protamine mRNAs will occur by a different pathway than histone mRNAs (Fitzgerald & Shenk, 1981; Montell et al., 1983; Birchmeir et al., 1983). It is interesting to note that although trout and mouse protamine mRNAs are polyadenylated when stored as ribonucleoproteins, both protamine mRNAs are substantially deadenylated when translated (Iatrou & Dixon, 1977; Kleene et al., 1984).

The cDNA clones for the mouse protamine 1 should prove useful in future studies of the evolution and regulation of the mammalian protamines. At the present time, we are using these clones to identify clones encoding mouse protamine 2 and genomic protamine sequences. Our previous work using pMP1-1 and -2 has demonstrated that the expression of the protamine 1 mRNA is controlled at the level of translation: it is first present at maximal levels in early haploid cells where it is stored as a nontranslated RNP, and it is translated in late haploid cells after RNA synthesis has ceased due to nuclear condensation (Kleene et al., 1983, 1984). The protamine mRNAs in the mouse are especially well-suited for studies of translational regulation, because they are small and highly abundant and their proteins can be readily distinguished by electrophoresis.

The mRNAs for protamine 1, a novel  $\alpha$  tubulin and several unidentified proteins all appear to be highly enriched in (or specific to) the haploid stages of spermatogenesis (Erickson et al., 1980; Kleene et al., 1983; Distel et al., 1983). The existence of these putative haploid-specific mRNAs implies that new mRNAs are transcribed in haploid spermatogenic cells, although the stage when transcription of these mRNAs starts remains to be established directly. Transcription of mRNAs in haploid spermatogenic cells might be expected to have profound effects on the differentiation of sperm because it would lead to the expression of recessive alleles resulting in the unequal transmission of alleles to progeny. Significantly, the effects of haploid gene expression in mammals appear to be rare, being restricted to distortions in the transmission ratios of alleles at the t locus by male mice (Erickson et al., 1981; Bennett, 1975) and the failure of male transgenic mice to transmit a metallothionein-thymidine kinase fusion gene to progeny (Palmiter et al., 1984). One reason why the consequences of haploid gene transcription on sperm differentiation are rare is that haploid spermatogenic cells are interconnected by intercellular bridges that could allow the exchange of haploid gene products. The identification of cDNA clones for these putative haploid-specific mRNAs should make it possible to determine when these mRNAs are initially transcribed and whether the products of genes that are transcribed during the

haploid stages of spermatogenesis are distributed randomly through intercellular bridges.

#### ACKNOWLEDGMENTS

We thank Dr. William Crain for advice on DNA sequencing and Dr. Lee Silver for providing the mouse testis cDNA library.

#### REFERENCES

- Aiken, J. M., McKenzie, D., Zhao, Z., States, J. C., & Dixon, G. H. (1983) *Nucleic Acids Res.* 11, 4907-4922.
- Balhorn, R., Weston, S., Thomas, C., & WYROBEK, A. J. (1984) *Exp. Cell Res.* 150, 298-308.
- Bellvé, A. R. (1979) in *Oxford Review of Reproductive Biology* (Finn, C. A., Ed.) Vol. 1, pp 159-260, Oxford University Press, Oxford.
- Bellvé, A. R., & Carraway, T. (1978) *J. Cell Biol.* 79, 177A.
- Bellvé, A. R., Anderson, E., & Hanley-Bowdoin, L. (1975) *Dev. Biol.* 47, 349-365.
- Bennett, D. M. (1975) *Cell (Cambridge, Mass.)* 6, 441.
- Birchmeir, C., Folk, W., & Birnstiel, M. L. (1983) *Cell (Cambridge, Mass.)* 35, 433-440.
- Coelingh, J. P., Monfoort, C. H., Rozijn, T. H., Gevers-Leuven, J. A., Shiphof, R., Steyn-Parve, E. P., Braunitzer, G., Shrank, B., & Ruhfus, A. (1972) *Biochim. Biophys. Acta* 285, 1-14.
- Distel, R. J., Kleene, K. C., & Hecht, N. B. (1984) *Science (Washington, D.C.)* 224, 68-70.
- Elsevier, S. M. (1982) *Dev. Biol.* 90, 1-12.
- Erickson, R. P., Kramer, J. M., Rittenhouse, J., & Salkeld, A. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77, 6086-6090.
- Erickson, R. P., Lewis, S. E., & Bentley, M. (1981) *J. Reprod. Immunol.* 3, 195-217.
- Fitzgerald, M., & Shenk, T. (1981) *Cell (Cambridge, Mass.)* 24, 251-260.
- Gaastra, W., Lukkes-Hofstra, J., & Kolk, A. H. J. (1978) *Biochem. Genet.* 16, 525-529.
- Gedamu, L., Wosnick, M. A., Connor, W., Watson, D. C., & Dixon, G. H. (1981) *Nucleic Acids Res.* 9, 1463-1482.
- Goldberg, R. B., Geremia, R., & Bruce, W. R. (1977) *Differentiation (Berlin)* 7, 167-180.
- Iatrou, K., & Dixon, G. H. (1977) *Cell (Cambridge, Mass.)* 10, 433-441.
- Kierszenbaum, A. L., & Tres, L. L. (1975) *J. Cell Biol.* 65, 258-270.
- Kistler, W. S., Keim, P. S., & Heinrickson, R. L. (1976) *Biochim. Biophys. Acta* 427, 752-757.
- Kleene, K. C., Distel, R. D., & Hecht, N. B. (1983) *Dev. Biol.* 98, 455-464.
- Kleene, K. C., Distel, R. D., & Hecht, N. B. (1984) *Dev. Biol.* 105, 71-79.
- Kozak, M. (1984) *Nucleic Acids Res.* 12, 857-872.
- Marushige, Y., & Marushige, K. (1975) *J. Biol. Chem.* 250, 39-45.
- Mathews, H. R., & Huebner, V. D. (1984) *Mol. Cell. Biochem.* 59, 81-101.
- Maxam, A. M., & Gilbert, W. (1978) *Methods Enzymol.* 65, 499-560.
- Mayer, J. F., Chang, T. S. K., & Zirkin, B. R. (1981) *Biol. Reprod.* 25, 1041-1051.
- Monfoort, C. H., Shiphof, R., Rozijn, T. H., & Steyn-Parve, E. D. (1973) *Biochim. Biophys. Acta* 427, 752-757.
- Montell, C., Fisher, E. F., Caruthers, M. H., & Berk, A. J. (1983) *Nature (London)* 305, 600-605.
- Nakano, M., Tobita, T., & Ando, T. (1976) *Int. J. Pept. Protein Res.* 8, 565-578.
- Palmiter, R. D., Wilkie, T. M., Chen, H. Y., & Brinster, R. L. (1984) *Cell (Cambridge, Mass.)* 36, 869-877.
- Rodman, T. C., Pruslin, F. H., & Allfrey, V. G. (1984) *Exp. Cell Res.* 150, 269-281.
- Sautière, P., Briand, G., Gusse, M., & Chevallier, P. (1981) *Eur. J. Biochem.* 119, 251-255.
- Schwarzbauer, J. E., Tamkur, J. W., Lemischka, I., & Hynes, R. O. (1983) *Cell (Cambridge, Mass.)* 35, 421-431.
- Tobita, T., Tsutsumi, H., Kato, A., Suzuki, H., Nomoto, M., Nakano, M., & Ando, T. (1983) *Biochim. Biophys. Acta* 744, 141-146.