The High Molecular Weight Chromatin Proteins of Winter Flounder Sperm Are Related to an Extreme Histone H1 Variant*

(Received for publication, November 7, 1997, and in revised form, December 26, 1997)

Catherine E. Watson‡ and Peter L. Davies§
From the Department of Biochemistry, Queen’s University, Kingston, Ontario K7L 3N6, Canada

Unlike mammals, birds, and most other fishes, winter flounder completes spermatogenesis without replacing its germ cell histones with protamines. Instead, during spermiogenesis, these fish produce a family of high molecular weight (60,000–200,000) basic nuclear proteins (HM,BNPs) that bind to sperm chromatin containing the normal complement of histones. These large, basic proteins are built up of tandem iterations of oligopeptide repeats that contain phosphorylatable DNA-binding motifs. Although the HM,BNPs have no obvious homology to histones, protamines, or other sperm-specific chromatin proteins, we report here the isolation of a clone to histones, protamines, or other sperm-specific chromatin proteins. One strategy used by some vertebrates and many invertebrates is to retain histones in a nucleosomal arrangement but to incorporate sperm-specific histone variants and/or other specialized basic proteins into the condensing chromatin (11, 12). In mammals, birds, and most fishes, DNA condensation is ultimately accomplished using protamines. These small arginine-rich proteins replace the histones and by doing so eradicate the nucleosomal organization established by the histones.

The winter flounder is one of the minority of bony fishes that retains its histones throughout spermatogenesis and does not replace them with protamines. Moreover, it does not synthesize significant quantities of sperm-specific histone variants (13). The winter flounder does, however, produce a group of high molecular weight basic nuclear proteins (HM,BNPs)1 in mid-to-late spermatids. These unique proteins are retained in the mature sperm, where they comprise >25% of the total acid-soluble proteins. As judged by SDS-polyacrylamide gel electrophoresis, there are at least 15 HM,BNPs that range in apparent molecular weight from 80,000 to 150,000, with a major band at ~110,000 and trace quantities of larger proteins up to 200,000. Amino acid analysis revealed that this group of proteins is constructed primarily from four amino acids: Arg (24%), Ser (23%), Lys (15%), and Pro (14%), which reflects their underlying simple repetitive sequences of dodecapeptides, with the consensus sequence SPMRSRSPSRSK, and heptapeptides, with the sequence RRVXXPK (where XX is QT or PS) (14). This simple composition and intermediate basicity suggests that the HM,BNPs might best fit in the class of chromatin proteins intermediate between histones and protamines (15).

Although the extreme repetitiveness of the HM,BNPs has precluded us from directly sequencing the proteins and from isolating full-length cDNA and genomic clones, we have obtained partial nucleotide sequences including the proximal promoter, 5’ and 3’ UTRs and about 1.5 kb of the coding region

* This work was supported by a grant from the Medical Research Council of Canada and by an NSERC studentship and Queen’s University Dean’s award (to C. E. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

‡Present address: Div. of Endocrinology, Diabetes & Metabolic Disease, Thomas Jefferson University, 1020 Locust St., Suite 348, Philadelphia, PA 19107-6799.
§To whom correspondence should be addressed. Tel.: 613-545-2983; Fax: 613-545-2497; E-mail:DaviesP@post.queensu.ca.

1The abbreviations used are: HM,BNP, high molecular weight basic nuclear protein; ORF, open reading frame; UTR, untranslated region; kb, kilobase pair(s); bp, base pair(s).
RESULTS

Isolation of Genomic Clones—The HM\_BNPs are a family of abundant proteins produced only in the testis during spermatogenesis. Their abundance was reflected in the frequency with which HM\_BNP clones were isolated from a mid-spermatid stage testis cDNA library (17). However, numerous attempts to recover a full-length cDNA have been hampered by the extreme repetitiveness of the sequence. As a result, we have had to piece together information about these proteins and their sequence using a number of different strategies. In an attempt to obtain information about HM\_BNP gene structure and regulation, we screened a partial Sau3A-digested winter flounder genomic DNA library using an incomplete cDNA clone. Approximately 200,000 plaques containing units >3 genomic equivalents were screened using radiolabeled HM\_BNP cDNA clone 3\'-5. Over 50 hybridization signals of varying intensities were detected. 23 of the phage producing strong signals were plaque-purified, and restriction enzyme maps for 10 of these clones were constructed.

The clones were grouped into two main classes (Fig. 1) based on the similarity and overlap of their restriction maps. Phage in class A formed a group of five overlapping clones (2B, 2D, 2E, 2F, and 1D) that spanned more than 20 kb. Each clone possessed a discrete region that hybridized strongly to the HM\_BNP cDNA probe. With the exception of clone 4B (which was truncated in the region of hybridization), the match to the HM\_BNP cDNA probe was restricted to a 1.6-kb Sac\_I fragment. Class B comprised three clones (3B, 3C, and 3F) representing ~18 kb of genomic DNA. These phage also contained a single region of hybridization that was localized around a common Sac\_I site. The remaining two clones (1C and 2A) appear to have unique restriction maps and showed hybridization only at one end of the phage insert.

Based on SDS-polyacrylamide gel electrophoresis experiments, the HM\_BNPs range in size from 80 to 200 kDa, with the most abundant protein being approximately 110 kDa (13). Such size estimates predict a mRNA of at least 2 kb and encode a full-length HM\_BNP sequence. However, because these clones could possibly represent HM\_BNP exons, pseudogenes, or related sequences and might therefore provide important information about these proteins and their genes, we chose to investigate one of the clones (2B) from class A.

The Region of Clone 2B That Hybridized to HM\_BNP cDNA Contains a Putative Gene—The 1.6-kb HindIII/Sac\_I fragment (which hybridized to the HM\_BNP cDNA) of genomic clone 2B and its flanking regions were subcloned and sequenced. The

FIG. 1. Restriction enzyme maps of genomic library clones. Restriction enzyme maps for the inserts of λ FixII clones were constructed based on single and double digestion patterns and are drawn to the scale indicated. Clones are grouped into classes A, B, and C. The restriction enzymes are abbreviated as follows: B, Bam\_HI; E, Eco\_R\_I; H, Hind\_III; K, Kp\_I; L, Sac\_I; S, Sac\_I. The regions of the clones that hybridized to the HM\_BNP cDNA are indicated by black rectangles.
nucleotide sequence obtained contains a long ORF from bp 498–1324 (Fig. 2), which when conceptually translated (beginning at the first in-frame methionine codon) would encode a 265-amino acid-long, 30-kDa protein. This ORF is followed by two potential polyadenylation signals (AATAAA), 107 (bp 1431) and 225 bp (bp 1549) downstream from the predicted translation termination codon. The sequence reported here also extends over 500 bp on the 5'-side of the coding region.

The 2B Genomic Sequence Is a HMrBNP Homolog—Portions of the nucleotide sequence of 2B showed a high degree of identity to a representative HMrBNP gene sequence. Dot matrix analysis of the two sequences (Fig. 3) using a stringency of 80% identity (16 out of 20) showed extensive areas of homology. These regions included portions of the proximal promoter, the 5'-UTR, and the coding sequence. Curiously, the putative CCAAT and TATA boxes identified in the proximal promoter region of the HMrBNP sequence were not conserved in 2B.

DNA coding for the predicted N-terminal region of 2B (bp 557–626) matching the HMrBNP 5'-UTR has been translated in Fig. 2, it is not known at this time if this sequence is 5'-UTR or coding sequence because it is only after the second in-frame ATG (bp 596–598) that HMrBNP-like sequences begin (see below). Dot matrix analysis comparing the coding regions of 2B and HMrBNP showed a high degree of identity and tandem repetition between these two sequences as indicated by the multiple lines parallel to the diagonal (Fig. 3). These regions of identity occurred over discontinuous stretches of about 100 bp in length, suggesting a closely related but distinct protein sequence. The similarity to the HMrBNP sequence ended abruptly around nucleotide 1075 of clone 2B (codon 184).

**Genomic Clone 2B Encodes a HMrBNP/Histone H1 Hybrid**—The predicted amino acid sequence of the N-terminal region of 2B (residues 14–188) had an amino acid composition similar to that of the HMrBNPs. The same four amino acids Arg (21%), Ser (22%), Lys (17%), and Pro (11%) were by far the most abundant and together made up 71% of the composition. In addition, 2B contained numerous sequences (Figs. 2 and 4) similar to the heptapeptide and dodecapeptide repeats obtained by endoproteinase Lys-C digestion (14). Interestingly, these repeats tend to occur in a defined order, interspersed with the sequences SPK and MRAKSPRRSK, such that they form a 32-amino acid sequence (Fig. 4). These
larger repeats, with the consensus sequence KSPMRSSPSRSKSPKRRVKTPKMRAKSPRRS, occurred four times (three linked in tandem) within the first 170 residues of the ORF and showed 71–90% identity to each other. Such repeats were also detected in the HM,BNP coding region (GenBank™ accession number 39735). The three HM,BNP repeats selected for comparison (x, y, and z) showed 80–90% similarity to the repeats in 2B. In addition, the predicted amino acid sequence of 2B contained four tandemly arrayed 7-amino acid repeats between residues 156–183 that overlap with the most C-terminal 32-amino acid repeat. These repeats (consensus sequence SPKMRAK) were distinct from both the heptapeptide repeats determined by peptide sequencing (RRVQTPK) and those present in the 32-amino acid repeats (RRVKTPK). They are more like the sequences corresponding to the 32- and 7-amino acid repeats are shaded consensus are (* above.

The alignment shown in Fig. 5. There are a number of “testis-specific” amino acid substitutions in the well conserved core. These include (using the human H1T numbering) K50E, V52L, K56Q, and V61M. Of the three, the last change was seen in 2B (residue 210).

The Isolated Genomic Clones Are Not the Product of in Vitro Recombination—Because the AFixII library was propagated in a Rec+ host, it was possible that the clones obtained from this library had been subject to internal recombination/deletion. To address this question, representative clones 1D (the longest of the class A clones) and 3B (a class B clone) were examined by Southern blot hybridization (Fig. 6). The hybridizing fragments from the λ clones (digested with three different restriction enzymes) all aligned perfectly with bands in the genomic DNA lanes. Clone 1D gave rise to single bands of hybridization at 10, 3, and 10 kb with EcoRI, HindIII, and SacI, respectively, whereas clone 3B produced 6-, 6-, and 5-kb bands of hybridization after digestion with the same three enzymes, respectively. This suggested that the clones were not the result of internal recombination but were continuous genomic fragments. However, as the corresponding genomic bands were far less intense on the autoradiograph than other HM,BNP gene signals, they are presumed to encode minor variants of the HM,BNP gene family or possibly pseudogenes.

**DISCUSSION**

The main purpose of spermiogenesis is to produce a streamlined, motile cell that can efficiently transfer the male’s genetic
mouse (36), mouse T (37), human T (38), and pig T (39). (winter flounder genomic clone 2B; this paper), Chicken (34), frog (J. G. Schilthuis, M. Hagenaar, and O. H. J. Destre, unpublished data), trout (35), are indicated on the right.

The blot was probed with H\textsubscript{\textit{m}} electrophoresed through a 0.8% agarose gel and transferred onto a nylon membrane. The blot was digested with the restriction enzymes indicated. Digests were purified phage DNA (100 ng combined with 10 \textmu g of calf thymus DNA) and then washed twice with 0.5 \times SSC/0.1% SDS for 20 min at 68 °C. X-ray film was exposed to the membrane for 21 h. The representative genomic DNA size markers (see Fig. 5). Although one of the 10 clones isolated (1C) contained 1.5 kb from the 5'-end of a H\textsubscript{\textit{m}} gene, this identity occurred at one end of the phage insert where the bulk of the repeat had been removed by the original cloning. It is important to emphasize that this linkage is not a recombination event. When the \lambda FxiII genomic library was screened with H\textsubscript{\textit{m}}, BNP cDNA, many hybridization signals were detected. Restriction enzyme analysis of 10 of these clones picked at random did not identify a full-length H\textsubscript{\textit{m}}. The sequences are identified as clone 2B.

Prior to this study, it was thought that winter flounder used an extreme variation of the second approach. In addition to retaining their histones, they synthesize a novel group of sperm chromatin proteins, the H\textsubscript{\textit{m}}, BNP\textsubscript{s}, that appear to be involved in binding and condensing DNA and are thus functionally related to protamines and linker histones (13, 22). However, characterization of the H\textsubscript{\textit{m}}, BNP\textsubscript{s} in isolation failed to identify a link to either group or indeed to any other proteins. Their amino acid composition is intermediate between those of histones and protamines, but their size and amino acid sequence resembles neither one. Recent database searches of Genbank\textsuperscript{39} and Swiss-prot with the H\textsubscript{\textit{m}}, BNP peptide repeats and partial nucleotide sequences did not reveal any similar proteins.

The isolation and sequencing of genomic clone 2B has at long last shed light on the origin of the H\textsubscript{\textit{m}}, BNP\textsubscript{s}. The H\textsubscript{\textit{m}}, BNP\textsubscript{s} are clearly related to linker histones because the putative coding region of 2B is homologous to the H\textsubscript{\textit{m}}, BNP\textsubscript{s} at its N terminus and to the globular region of histone H1 at its C terminus. Also, the 5'-flanking region of the 2B ORF has very high sequence identity to the proximal promoter of the H\textsubscript{\textit{m}}, BNP genes.

Indeed, it is notable that the intransigence of the H\textsubscript{\textit{m}} genes to cloning facilitated the isolation of clone 2B. Prior to this study, it was thought that winter flounder used an extreme variation of the second approach. In addition to retaining their histones, they synthesize a novel group of sperm chromatin proteins, the H\textsubscript{\textit{m}}, BNP\textsubscript{s}, that appear to be involved in binding and condensing DNA and are thus functionally related to protamines and linker histones (13, 22). However, characterization of the H\textsubscript{\textit{m}}, BNP\textsubscript{s} in isolation failed to identify a link to either group or indeed to any other proteins. Their amino acid composition is intermediate between those of histones and protamines, but their size and amino acid sequence resembles neither one. Recent data base searches of Genbank\textsuperscript{39} and Swiss-prot with the H\textsubscript{\textit{m}}, BNP peptide repeats and partial nucleotide sequences did not reveal any similar proteins.

The isolation and sequencing of genomic clone 2B has at long last shed light on the origin of the H\textsubscript{\textit{m}}, BNP\textsubscript{s}. The H\textsubscript{\textit{m}}, BNP\textsubscript{s} are clearly related to linker histones because the putative coding region of 2B is homologous to the H\textsubscript{\textit{m}}, BNP\textsubscript{s} at its N terminus and to the globular region of histone H1 at its C terminus. Also, the 5'-flanking region of the 2B ORF has very high sequence identity to the proximal promoter of the H\textsubscript{\textit{m}}, BNP genes.

It is important to emphasize that this linkage is not a recombination event. When the \lambda FxiII genomic library was screened with H\textsubscript{\textit{m}}, BNP cDNA, many hybridization signals were detected. Restriction enzyme analysis of 10 of these clones picked at random did not identify a full-length H\textsubscript{\textit{m}}, BNP gene, despite their strong hybridization signal on phage DNA blots (see Fig. 5). Although one of the 10 clones isolated (1C) contained 1.5 kb from the 5'-end of a H\textsubscript{\textit{m}}, BNP gene, this identity occurred at one end of the phage insert where the bulk of the repeat had been removed by the original Sau3A digestion (data not shown). The failure to isolate a full-length H\textsubscript{\textit{m}}, BNP gene is entirely consistent with the extraordinary difficulties experienced in trying to clone these highly repetitive sequences at the cDNA level. However, because the host cells used to screen the genomic library (E. coli NM532) were Rec+, the possibility that the isolated clones had undergone recombination/deletion was investigated. Southern blot analysis of these clones alongside serotonin genomic DNA indicates that they contain bona fide DNA fragments that have not been rearranged. It also suggests that gene 2B represents a minor constituent, perhaps a single copy gene, among the H\textsubscript{\textit{m}}, BNP multi-gene family. Indeed, it is notable that the intraspecificity of the H\textsubscript{\textit{m}}, BNP genes to cloning facilitated the isolation of clone 2B.

It is quite likely that the progenitor of 2B and the H\textsubscript{\textit{m}}, BNP\textsubscript{s} was a testis-specific variant of histone H1, which would account for the tissue and stage specificity of H\textsubscript{\textit{m}}, BNP expres-
tion. Some testis-specific histones are notably more basic (and arginine-rich) than their somatic counterparts and have N- or C-terminal regions that contain short simple repeats of the SPKK motif (6, 10, 23, 24) that are so abundant in the HM$_r$B-NPs. Amplification of similar repeats could have produced an extreme H1 variant like 2B to assist in regulated and reversible sperm chromatin condensation. Gene duplication, rearrangement, and loss of the globular H1-like domain in the HM$_r$B-NP progenitor have apparently given rise to this unique auxiliary sperm chromatin protein. Moreover, the demand for this protein seems to have led to the extensive expansion and amplification of its gene to the point where there are now at the very least 15 HM$_r$B-NP isoforms (13).

This recognition of HM$_r$B-NP origins is particularly interesting in relation to recent work on the “protamine-like” sperm proteins of the mollusk, Mytilus californianus. Carlos et al. (8, 25) have demonstrated that two of the three major protamine-like proteins in this invertebrate are in fact post-translational 25) have demonstrated that two of the three major protamine-specific proteins originated from histones was originally proposed to be their stage of evolution. The idea that sperm basic chromatin proteins are evolutionary derivatives of histones is beginning to appear that the distinction among protamines, protamine-like proteins, and linker histones may simply be a stage of evolution. The idea that sperm basic chromatin proteins originated from histones was originally proposed by Subirana et al. (30). This hypothesis has since been refined by Ausio and co-workers to suggest that such proteins arose from a primitive histone H1, and it is supported by their extensive biochemical analysis of sperm nuclear proteins from a wide variety of lower eukaryotes (31–33). The isolation and characterization of winter flounder clone 2B provides evidence in vertebrates that specialized sperm chromatin proteins have evolved from the N-terminal tail of a progenitor linker histone.

Acknowledgments—We thank Dr. Garth Fletcher and co-workers for supplying tissue samples and Sherry Gauthier for expert technical assistance.

REFERENCES

1. Wolfe, A. (1995) Chromatin Structure and Function, 2nd Ed., Academic Press, Inc., San Diego, CA
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Nature 389, 251–260
3. Crance-Robinson, C. (1997) Trends Biochem. Sci. 22, 75–77
4. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) Nature 362, 219–223
5. Gabrielli, F. (1989) in Histones and Other Basic Nuclear Proteins (Hnilica, L. S., Stein, G. S., and Stein, J. L., eds) pp. 3–16, CRC Press, Boca Raton, FL
6. Kistler, W. S. (1989) in Histones and Other Basic Nuclear Proteins (Hnilica, L. S., Stein, G. S., and Stein, J. L., eds) pp. 331–345, CRC Press, Boca Raton, FL
7. Kasinsky, H. E., Mann, M., Huang, S. Y., Fabre, L., Coyle, B., and Byrd, E. W., Jr. (1987) J. Exp. Zool. 243, 137–151
8. Carlos, S., Jutglar, L., Borrell, I., Hunt, D. F., and Ausio, J. (1993) J. Biol. Chem. 268, 185–194
9. Schuel, E., Trieschnigg, L., Schulze, B., Schmidt, E. R., Pitzel, S., Zeckel, K., and Grossbach, U. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 2481–2485
10. Romano, G. (1992) Cell Biol. Intl. Reports 16, 197–206
11. Jutglar, L., Borrell, I., and Ausio, J. (1991) J. Biol. Chem. 266, 8184–8191
12. Kasinsky, H. E. (1989) in Histones and Other Basic Nuclear Proteins (Hnilica, L. S., Stein, G. S., and Stein, J. L., eds) pp. 73–163, CRC Press, Boca Raton, FL
13. Kennedy, B. P., and Davies, P. L. (1988) J. Biol. Chem. 255, 2533–2539
14. Young, D. C., and Davies, P. L. (1987) Biochem. Cell Biol. 65, 909–916
15. Bloch, D. P. (1969) Genetics 61, (suppl.) 93–111
16. Scott, G. K., Hew, C. L., and Davies, P. L. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 2613–2617
17. Watson, C. E. (1995) Structural Analysis of Flounder High Molecular Weight Sperm Chromatin Proteins and Their Genes. Ph. D. Dissertation, Queen’s University, Kingston, ON, Canada
18. Cerf, C., Lippens, G., Muyldermans, S., Segers, A., Ramakrishnan, V., Wodak, S. J., Hallenga, K., and Wyns, L. (1993) Biochemistry 32, 11345–11351
19. Oliva, R., and Dixon, G. H. (1991) in Progress in Nuclear Acid Research and Molecular Biology (Cohn, W. E., and Moldave, K. v.) pp. 26–94, Academic Press Inc., San Diego, CA
20. Daisley, S. L., and Davies, P. L. (1982) Biochim. Biophys. Acta 698, 271–279
21. Kadura, S. N., Khrapunov, S. N., Chabanny, V. N., and Berdychev, G. D. (1983) Comp. Biochem. Physiol. 74B, 343–350
22. Kennedy, B. P., and Davies, P. L. (1982) J. Biol. Chem. 257, 11160–11165
23. Puccia, D. L. (1987) in Molecular Regulation of Nuclear Events in Mitosis and Meiosis (Schlegel, R. A., Halleck, M. S., and Rao, P. N., eds) pp. 149–147, Academic Press, New York
24. Massey, C. B., Jr., and Watts, S. A. (1992) J. Exp. Zool. 262, 9–15
25. Carlos, S., Hunt, D. F., Rocchini, C., Arnott, D. P., and Ausio, J. (1993) J. Biol. Chem. 268, 195–199
26. Rocchini, C., Rice, P., and Ausio, J. (1995) FEBS Lett. 363, 37–40
27. Giancotti, V., Buratti, E., Santucci, A., Neri, P., and Crane-Robinson, C. (1992) Biochim. Biophys. Acta 1119, 296–302
28. Kennedy, B. P., and Davies, P. L. (1985) J. Biol. Chem. 260, 4338–4344
29. Saperas, N., Ausio, J., Lloris, D., and Chiva, M. (1994) J. Mol. Biol. 239, 292–326
30. Subirana, J. A., Cox colluela, C., Plau, J., and Unzeta, M. (1973) Biochim. Biophys. Acta 317, 364–379
31. Rocchini, C., Zhang, F., and Ausio, J. (1995) Biochemistry 34, 15704–15712
32. Rocchini, C., Marx, R. M., von Carosfeld, J. S., Kasinsky, H. E., Rosenberg, E., Sommer, F., and Ausio, J. (1996) J. Mol. Biol. 42, 240–246
33. Ausio, J., Van Veghel, M. L. J., Gomez, R., and Barreda, D. (1997) J. Mol. Biol. 265, 91–96
34. Coles, L. S., Robins, A. J., Madley, K. L., and Wels, J. R. (1987) J. Biol. Chem. 262, 9656–9663
35. Mequita, J., Connor, W., Winkfein, R. F., and Dixon, G. H. (1984) J. Mol. Biol. 21, 209–219
36. Cheng, G. H., Nandi, A., Clerk, S., and Skoultchi, A. I. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 7002–7006
37. Drabent, B., Bode, C., and Doenecke, D. (1993) Biochim. Biophys. Acta 1216, 311–313
38. Drabent, B., Kardaliou, E., and Doenecke, D. (1991) Gene (Amst.) 103, 263–268
39. Cole, K. D., York, R. G., and Kistler, W. S. (1984) J. Biol. Chem. 259, 13695–13702