The nucleotide sequence of the gene encoding the sperm specific histone subtype H2B-1 from the sea urchin *Strongylocentrotus purpuratus*

Zhichun Lai, Toby Lieber and Geoffrey Childs*

Departments of Genetics and Molecular Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Submitted 7 October 1986

The sperm H2B-1 gene is typical of many cell-cycle regulated histone genes. It contains no intervening sequences but retains conserved sequence elements at both the 5' and 3' ends of the gene including the 3' dyad sequence element (indicated with inverted arrows) and CAGA box (1) and three copies of the sequence CTNATTGCGATA (positions -46; -16; -205) (2,3) whose core is identical with the SV40 and Ig enhancers (4-6). Between two of these three repeated sequences are two inverted copies of the CCAAT sequence (positions -100 and -132). This gene is transcribed in a sex and tissue specific manner only in testis (see ref. 4). The mRNA begins at base 321 (+1; indicated with an arrow) and ends at base 836 (+516) encoding an H2B-1 protein of 139 amino acids with a basic repeat motif at the N-terminus (indicated with arrows). Comparison of the sequences upstream of testis, somatic and embryonic H2B genes does not reveal an obvious explanation for the observed sex and tissue specific expression of the testis genes.

```

-320   GGATGTATAAATATGTAGTCTACACATAAAGGAACATCCGCCATTATTCACCTTCTCAGTGA
-253   CGTACTCAATCGACGTCGACGCGATCCTCGAACTCGCTAATTGCGATAAGCGCGCGTTTTTTATTGACA
-182   CTATCTAATTTCATATGCGCGCGTTTTTTATTTACACACTTCTGATTCGCTACAAATATTCGGTAGCTTA
-111   GCTTTGATTGGTAGGTTCTCAGATGTTGCCGTCGCGCTCGTATCCCGAATCCTCATTTGCGATCATCA
-040   TCTATTATAAATAGGTCGATCACCAGGTTGTACATCATTCTAGATTGATTGAAATCCTAAGCAATCATC
+032   ATG CCG TCT CAG AGG AGT CCT ACC AAG CGA AGT CCG ACC AAA CGT AGC CCC CAG
      MET pro ser gln arg ser pro thr lys arg ser pro thr lys arg ser pro gln
+086   AAG GGA GCT GGG AAG GGA GGC AAG-GGA TCC AAA CGA GGA GGA AAG GCC AGC CGT
      lys gly ala gly lys gly gly lys gly ser lys arg gly glu ser tyr gly ile
+140   CGT GGT GGA GCC GCT GTA AGG AGA CGT CGA CGA AGA CSC GAG ACC TAT GGA ATC
      arg gly gly ala ala val arg arg arg arg arg arg arg glu ser tyr gly ile
+194   TAC ATC TAC AAA GTG TTG AAG CAA GTT CAT CCC GAC ACT GGC ATT TCT AGC CGC
      tyr ile tyr lys val leu lys gln val his pro asp thr gly ile ser ser arg
+248   GGC ATG TCC ATC ATG AAC AGT TTT GTC AAC GAT GTG TTT GAG CSC GTC GCT GCC
      gly MET ser ile MET asn ser phe val asn asp val phe glu arg val ala ala
+302   GAA GCT TCC CBT CTT ACC AAG TAC AAC CSC AGG AGC ACC GTT TCA AGC CSC GAG
      glu ala ser arg leu thr lys tyr asn arg arg ser thr val ser ser arg glu
+356   ATC CAG ACC GCT GTC CSC CTT CTT CTC OCT GGA GAG CTG GCC AAG CAC GCC GTT
      ile gln thr ala val arg leu leu leu pro gly glu leu ala lys his ala val
+410   TCT GAA GGC ACC AAA GCT GTC ACC AAA TAC ACC ACG TCT CGA TAG ATCCGAGACCT
      ser glu gly thr lys ala val thr lys tyr thr thr ser arg
+466   GGCTGGCCTGGAGACCATCTCCAAAACCAACGGCTCTTTTCAGAGCCACCAATTTTCAAGAAAGATCGA
+537   GATAGATTATTTCCC +549

```

*To whom correspondence should be addressed

References: 1. Hentschel, C.C. and Birnstiel, M.L. (1981). *Cell* **25**, 301-303. 2. Harvey, R., Robins, J., and Wells, J. (1982) *Nucleic Acids Res.* **10**, 7851-7863. 3. Perry, M., Thompson, G. and Roeder, R. (1985). *J. Mol. Biol.* **185**, 479-499. 4. Lai, Z.C. and Childs, G. (1986). *Nucleic Acids Res.* **14**, 6845-6856. 5. Sive, H.L. and Roeder, R.G. (1986). *Proc. Nat. Acad. Sci.* **83**, 6382-6386. 6. Sen, R. and Baltimore, D. (1986). *Cell* **46**, 705-716.

This work was supported by NIH grant GM-30333. T.L. was supported by an NIH predoctoral training grant.