Volume 10 Number 15 1982 Nucleic Acids Research

Nucleotide sequence of a protamine component C_{II} gene of *Salmo gairdnerii*

J.C.States, W.Connor, M.A.Wosnick*, J.M.Aiken, L.Gedamu+ and G.H.Dixon

Department of Medical Biochemistry, Faculty of Medicine, Health Sciences Center, The University of Calgary, Calgary, Alberta, Canada T2N 4N1

Received 5 May 1982; Accepted 16 June 1982

ABSTRACT

We have isolated, using nick-translated cloned protamine cDNA's as probes, several genomic clones containing protamine gene sequences from a Charon 4A library of Eco R1 digested rainbow trout (Salmo gairdnerii) DNA. One clone was chosen for detailed study and the 2.5 kbp Bam HI-Eco R1 restriction fragment containing the gene was subcloned in the plasmid pBR322. A 920 bp Bgl II - Bam HI restriction fragment contains a sequence coding for protamine component C_{II} as well as regions 5' and 3' to the mRNA coding portion. Present in the region 5' to the mRNA coding sequence are the promoter associated signals "TATA" box and "CAAT" box. The 5' untranslated region of the mRNA whose length and sequence were not established from the cDNA clones (1) was determined by nuclease mapping and starts within a sequence similar to the "capping signal" found in other genes. The protamine gene for C_{II} contains no introns, a situation common to most histone genes, but, unlike the histone genes does not occur close to other protamine genes in a "cluster".

INTRODUCTION

The protamines are a family of small, arginine rich proteins that are the major structural proteins of the sperm chromatin in rainbow trout (Salmo gairdnerii) (2). They are encoded by a multi-gene family (1,3) and the expression of these genes is under developmental control as part of the terminal differentiation of the sperm (4). The protamine polypeptides of Salmo gairdnerii can be resolved into 3-4 components by ion-exchange chromatography (1,5). Characterization of protamine mRNA has revealed four size classes, each of which, when translated in vitro, yields varying proportions of all the protamine polypeptides (5). As a result of cloning and sequencing protamine cDNA's by our laboratory (1) and others (6,7), the existence of two families of genes corresponding to protamine components C_{II} and C_{III} was deduced.

In order to investigate the protamine multi-gene family further, we have constructed a partial Eco R1 digest library of rainbow trout DNA carried in the λ Charon 4A vector. We have isolated several genomic clones (8) using

© IRL Press Limited, Oxford, England.

0305-1048/82/1015-4551$2.00/0

the previously sequenced protamine cDNA clones (1) as probes. These genomic clones were then analyzed by restriction mapping and Southern transfer and hybridization to nick-translated protamine cDNA pRTP242. The region containing the gene in each of these clones was subcloned for further analysis (8). This report deals with the detailed analysis and complete nucleotide sequence of the protamine gene region of one of these clones.

MATERIALS AND METHODS

T4 polynucleotide kinase, T4 DNA ligase and calf intestinal alkaline phosphatase were gifts of J.H. van de Sande and B.W. Kalisch. Restriction enzymes were purchased from Bethesda Research Laboratories, New England Bio-Labs and Boehringer-Mannheim; exonuclease VII from Bethesda Research Laboratories; DNase I from Worthington; tRNA, S1 nuclease, proteinase K and *E. coli* DNA polymerase I from Boehringer-Mannheim; ampicillin, calf thymus DNA and yeast RNA from Sigma; formamide from Fluka; nitrocellulose filters from Schleicher and Schuell; radioactive nucleotides from Amersham, New England Nuclear and ICN. λ Charon 4A arms and computer programs were supplied by K. Iatrou; *E. coli* strain RRI by R. Palmiter (Department of Biochemistry, University of Washington); *E. coli* JF1968 by J. Friesen (Department of Medical Genetics, University of Toronto); *E. coli* DP50 supF and protein A by J. Bell (Department of Biochemistry, San Francisco Medical Centre).

Isolation of genomic protamine clones: A partial Eco R1 digest of rainbow trout DNA was size fractionated on sucrose gradients (10-40%) and the 15 ± 2 kbp fraction was ligated to λ Charon 4A arms and packaged *in vitro* into phage particles. The phages were amplified by infection of *E. coli* DP50 supF and screened by the *in situ* hybridization technique (9). Cloned protamine cDNA pRTP242 (1), ^{32}P-labelled by nick-translation (10) served as probe.

Restriction mapping: Phage and plasmid DNA's containing trout genomic inserts hybridizable to protamine cDNA were restriction mapped using single and double restriction digests. Localization of protamine sequences was performed by identification of restriction fragments that hybridized to protamine cDNA by Southern transfer (11) and hybridization at 45°C in a buffer containing 50% formamide, 0.6 M NaCl, 0.12 M tris-HCl (pH 8), 4 mM EDTA-Na_2, 0.1% SDS and Na_4-Pyrophosphate, 100 g/ml calf thymus DNA or yeast RNA, and 0.2% Ficoll, PVP and BSA.

Subcloning: A 2.5 kbp Eco R1 - Bam HI restriction fragment from clone λTP101, which contained the protamine gene, was isolated from an agarose gel

and ligated to pBR322 cut with Eco R1 and Bam HI. The resulting chimeric plasmid, p101, was used to transform E. coli strain RRI by temperature shock and the transformed bacteria were selected by drug sensitivity (ampicillin resistant, tetracycline sensitive), grown in 1 ml cultures, and their plasmids screened by restriction analysis.

DNA Sequencing: Plasmid DNA was isolated by amplification with chloramphenicol, lysis with detergent and purification by banding cleared lysates on CsCl equilibrium gradients (12). After digestion with the appropriate restriction endonuclease, recombinant plasmid DNA was incubated with calf intestinal alkaline phosphatase, then proteinase K, extracted with chloroform-isoamyl alcohol-phenol (24:1:25) and precipitated with ethanol. The 5' ends were ^{32}P-labelled by incubating with T4 polynucleotide kinase and ATP-γ-^{32}P. The mixture was precipitated with ethanol and redissolved in a buffer suitable for the second restriction endonuclease digestion, after which, the entire mixture was electrophoresed on an agarose gel (1-2%). The fragment of interest was cut out of the gel and electroeluted. The DNA was then subjected to the G>A, G+A, T+C, C>T, A>C partial degradation reactions of Maxam and Gilbert (13) and the products were electrophoresed on 6% polyacrylamide gels containing 8 M urea (850 x 200 x 0.4 mm) for 5.5 to 25 h. The gels were adsorbed onto Whatman 3MM paper, covered with PVC film and autoradiographed at -80°C with the use of intensifying screens and preflashed Kodak XAR X-ray film. Up to 400 nucleotides could be read from a single labelled end by this method.

Nuclease mapping: The Bgl II - Bst NI restriction fragment of p101, ^{32}P-labelled at the 5' Bst NI end, was isolated as for sequencing. Protamine mRNA (poly A+), isolated as described (14), was hybridized to the labelled DNA for 1.5 h in 80% formamide, 50 mM PIPES (pH 6.5), 0.4 M NaCl, 1 mM EDTA-Na_2 at 65°C and then diluted into cold nuclease buffer (15). The mix was then incubated with nuclease S1 or exonuclease VII as per manufacturer's instructions and electrophoresed on sequencing gels with chemical cleavage products of the original DNA as markers.

RESULTS AND DISCUSSION

Analysis of genomic clone λTP101 and subclone p101: Genomic clone λTP101 was isolated and restriction mapped as described in Materials and Methods. The restriction map of the trout DNA insert of this clone is presented in Figure 1. The length of the single Eco R1 restriction fragment is 16 kbp. There are single sites for Bam HI and Kpn I and two sites for Hind III within this

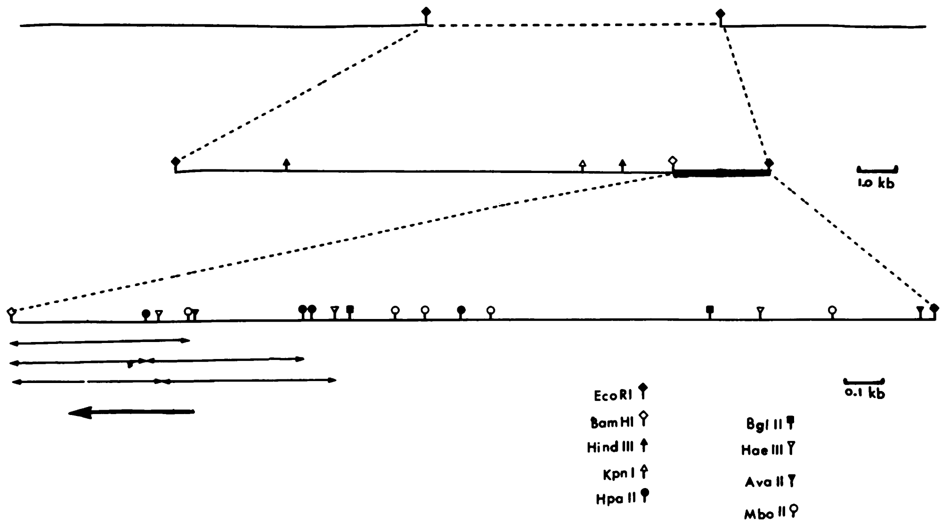

Figure 1: Restriction maps of the trout DNA inserts of clones λTP101 and p101 and hybridization analysis of p101. The top line is a representation of the original clone λTP101 and below it is an expanded view of the 16 kbp insert with the restriction sites marked. The thickened part represents the Bam HI - Eco R1 fragment that was subcloned in plasmid pBR322. An enlarged map of this 2.5 kbp fragment is shown. Beneath this, are marked the restriction fragments from single digests that hybridized to protamine cDNA after Southern transfer (light double headed arrows). The dark arrow represents the position and orientation (5' - 3') of the gene.

fragment. The 2.5 kbp Eco R1 - Bam HI restriction fragment was the only one resulting from a multiple digest to hybridize to nick-translated cloned protamine cDNA pRTP242 (data not shown). Therefore this fragment was chosen for subcloning in plasmid pBR322 for further analysis. The recombinant plasmid was named p101. The restriction map of the insert in p101 is shown in the lower part of Figure 1, along with the results of Southern transfer and hybridization to cDNA, of three restriction digest products of this insert.

Restriction sites for Mbo II, Hpa II and Hae III are useful protamine gene markers (1) and were chosen for preliminary restriction mapping, and localization and orientation of possible protamine gene(s) by Southern transfer and hybridization. An Mbo II site is near the 5' end of the mRNA and Hpa II and Hae III sites are 3' to the Mbo II site (1). Considering the localization of the marker restriction sites and the hybridization pattern, it was deduced that there is only one protamine gene present in this

subclone. It is located in the end of the insert and oriented with its 3' end toward the Bam HI site as noted in the lower part of Figure 1. Since the subcloned fragment was the only piece of the original clone to hybridize to protamine cDNA, and there is only one protamine gene in the subclone, there is only one protamine gene present in the original genomic clone of 16 kbp. This single protamine gene is located 2 kbp from one end and 14 kbp from the other end of the clone. Other protamine clones, all of 15 ± 2 kbp, appear to have only one gene in them as well (J. Aiken, unpublished observations; 8). From this information, it is apparent that protamine genes are not present in clusters as are the histone genes of trout (23), sea urchins(18) and Xenopus (24) as well as globin genes in vertebrates (27), but rather are widely separated single genes.

Sequencing of the gene containing region: Restriction endonucleases Bgl II and Ava II are more useful for sequencing utilizing 5' end labelling due to the exposure and ready labelling of the 5' position, so their cut sites were mapped on the subcloned insert DNA as well. These are also shown in Figure 1. In Figure 2 the Bgl II - Bam HI restriction fragment is shown in detail along with the sequencing strategy. Since putative control sequences should be contained in the regions 5' or 3' to the RNA coding region, the entire 920 bp Bgl II - Bam HI fragment was chosen for sequencing in order to provide flanking sequence information. The appropriate restriction fragments were 5' end labelled, isolated and analyzed by the method of Maxam and Gilbert (13) as described in Materials and Methods. Each labelling reaction, fragment isolation and sequence analysis was performed at least twice. Almost the entire 920 bp were sequenced on both strands of the DNA. The Bst NI site was determined from sequence data. The confirmatory information obtained by

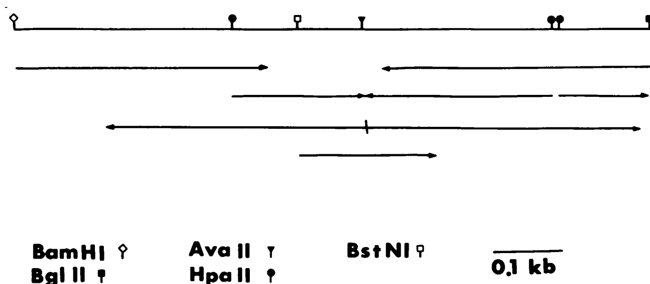

Figure 2: Strategy for sequencing the Bgl II - Bam HI restriction fragment of pl01. The restriction sites are as noted. The arrows represent the direction and range sequenced from a single 5' labelled end.

sequencing from this site was the result of nuclease mapping experiments described below. The total sequence of the 920 bp Bgl II - Bam HI restriction fragment is presented in Figure 3. The promoter associated signals the "CAAT" box (16) at -100 bp from the mRNA start the "TATA" box (17) at -28 and the "capping signal" (pyCATTCPu) (18) are denoted. The first two of these sites in the protamine gene were located by computer search for the best fit to consensus sequences for these sites at the appropriate location 5' to the transcription initiation site, assumed to be the 5' end of

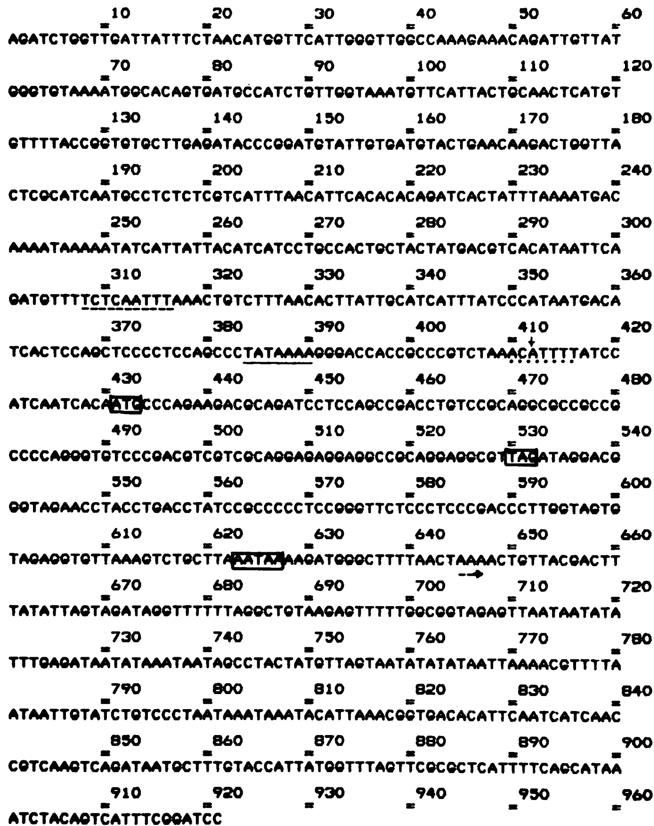

Figure 3: Sequence of the Bgl II - Bam HI restriction fragment of p101. The sequence of the DNA strand having the same sense as the mRNA is presented in a 5' to 3' direction. Promoter associated signals are underlined ("CAAT" box, ---; "TATA" box, ___). The capping signal ACATTT is underlined as well (...), and the initiation of transcription (at 412) is marked with an arrow (↓). The initiation (ATG) and termination (TAG) codons are boxed, as is the polyadenylation signal (AATAA). The site of poly(A) addition is denoted by an arrow (--→).

the mRNA as determined by nuclease mapping described below. The capping signal is that surrounding the 5' end of the mRNA and gives a reasonable match to the consensus sequence. The polyadenylation signal AATAA (19) is located approximately 20 nucleotides upstream of a region of four A's marked with a dashed arrow (--->), which is the beginning of the poly(A) tail as inferred from cDNA sequencing (1). In view of the presence of these A's in the genomic sequence, the termination of transcription and/or point of poly(A) addition is indefinite within four bases. It is possible that the coded sequence of four A's provides a short primer for poly(A) addition. The initiation (ATG) and termination (TAG) codons for translation are also noted in Figure 3.

Nuclease mapping of the 5' end of the mRNA transcribed from this gene: The 5' ends of the protamine mRNA's were not obtained in the cDNA's that were sequenced previously (1,6,7). This may have been due to obstructive secondary structures in the protamine mRNA's which hindered the reverse transcriptase, as mentioned by these authors. Therefore, it was necessary to map the 5' end of the mRNA encoded by this gene, using nuclease mapping techniques (14,20). The Bgl II - Bst NI restriction fragment, 5' end labelled on the mRNA coding strand (Bst NI site) with ^{32}P, was hybridized to total protamine mRNA (poly A^+) (5). The resulting hybrids were then digested with nuclease S1 or exonuclease VII and electrophoresed on sequencing gels along with Maxam and Gilbert (13) sequencing reaction products of the original DNA fragment as standards. The relevant portion of the resulting autoradiogram is shown in Figure 4. The nuclease digested material migrates 1½ bases faster than the corresponding sequencing standards (20). Therefore, the top band of the cluster of intense bands in the nuclease S1 track corresponds to the A at nucleotide 412 of Figure 3. The bands corresponding to the following T's are most likely due to the exonuclease activity of the S1 nibbling at the A-T rich end of the hybrid. This is supported by the sharpness of the pattern of the exonuclease VII digested material at this position. This enzyme tends to leave 1-2 bases on the end of the duplex (15) and therefore, one reads 1-2 bases further 3' than for nuclease S1. In addition, the sequence of the entire Bgl II - Bam HI fragment was computer searched for matches to the consensus sequences for intron-exon junctions (28) and there were no good matches found that would allow for an intron in the 5' leader sequence of this gene. This further supports the positioning of the mRNA start site at nucleotide 412.

The minor bands at lower positions in the gel are probably due to the

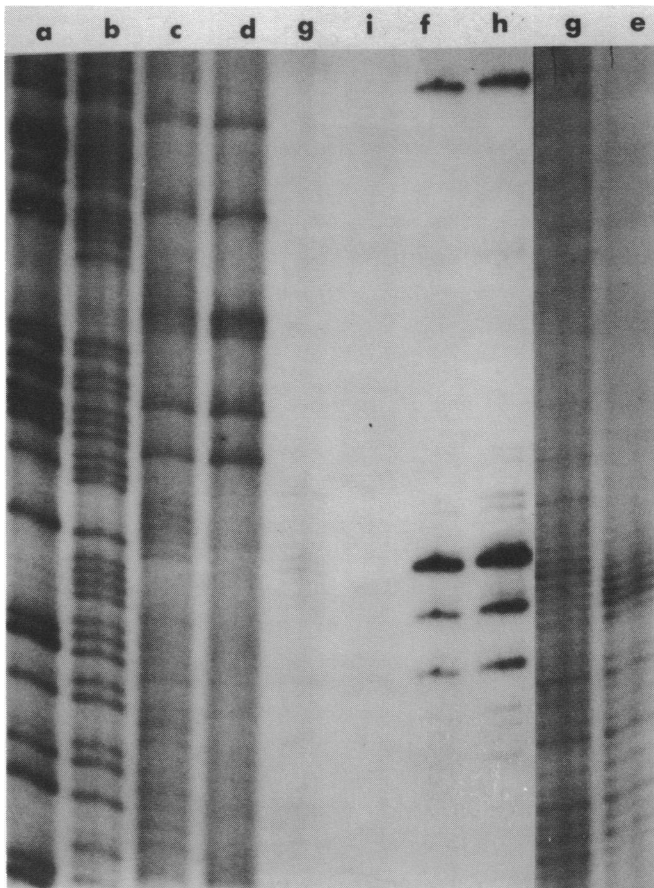

Figure 4: Nuclease mapping of the 5' end of the mRNA encoded by p101. Lanes a,b,c, and d are respectively G, G+A, T+C, and C sequencing reaction products of the DNA used for hybridization to total protamine mRNA. Lanes e and g are the results of treatment of the DNA-RNA hybrids with 100 and 10 U nuclease S1. Lanes f and h are the results of treating the hybrids with 1.0 and 0.5 U exonuclease VII. Lane i is a control using no nuclease.

heterogeneity of the protamine message population which although possessing appreciable diversity will retain a great deal of common structure. Any mismatches resulting from heterogeneity in mRNA structure would be nicked by nuclease S1 and large 5' non-homologies would be removed by exonuclease VII. This would result in bands corresponding to positions 3' to the mRNA start site. The band at a high position in both the exonuclease VII and low S1 tracks corresponds to nucleotide 363 which lies at the A in the sequence

TCACTCC, which is a reasonable capping signal (18). Starting at nucleotide 342 there is a sequence CATTTAT which is a fairly good match to the consensus for a TATA box (17). This is suggestive of another site for initiation of mRNA transcription giving an mRNA of 281 bases, but, this band disappears with more extensive digestion with S1 nuclease. Therefore, one can only assume that there is another gene which shows strong homology to the sequence present in p101, but codes for a longer mRNA. We conclude that the gene isolated in p101 codes for an mRNA of 233-237 bases starting at nucleotide 412 and terminating at nucleotides 644-648. Its coding sequence is uninterrupted by introns, unlike the majority of eucaryotic genes that have been analyzed to date. The most notable exceptions are the histone genes in trout (23), sea urchin (18) and Xenopus (24). It is tempting to suggest that only chromatin structural protein genes have no introns, but other unrelated genes such as the herpes simplex virus thymidine kinase gene (25) and human interferon B gene (26) have been shown to be uninterrupted as well. Therefore, there appears to be no simple rule determining which genes will or will not have introns.

Comparison of genomic sequence from p101 with cDNA sequences: In Figure 5 is displayed the sequence of two protamine cDNA's in comparison with the gene region of p101. pRTP43 is a representative of a family I sequence, and pRTP59, a family II sequence (1). The dashes represent nucleotides identical to those present in p101. The asterisks show gaps introduced in order to retain homology. The amino acid sequences encoded are written beneath the p101 sequence and the single change in pRTP43, at the corresponding place in its sequence. As is readily observable, pRTP59 is nearly identical to the corresponding region of p101. Indeed, there are only two differences, a G $\rightarrow$ T change at 605 and the insertion of a G between 636 and 637 when p101 and pRTP59 are compared. This amounts to 99% homology and establishes the p101 sequence as a member of family II. p101 codes for protamine component C_{II} as does pRTP59 (1). On the other hand, cDNA pRTP43 differs from p101 by the deletion of two codons as well as a number of substitutions and insertions. pRTP43 also possesses a longer 5' non-coding region that makes it larger than the transcribed region of p101. pRTP43 is most likely to be a member of either the 245 or 274 base long size classes of protamine mRNA (following removal of poly A) which have been characterized previously in this laboratory (5). The transcript from p101 would be 233 bases long after removal of poly(A). This corresponds to the 235 base long size class of protamine mRNA (5). This size class was shown to be translated *in vitro*,

```
pRTP43          - A T - C - T A G - C - - - - - T - - - - - - - - - T - - - - - - - - - - - - - - - - - - * * * G - -
                                                                                                              ala
p101    G C C C G T C T A A A C A T T T T A T C C A T C A A T C A C A A T G C C C A G A A G A C G C A G A T C C T C C
                                                                  init  pro   arg   arg   arg   arg   ser   ser
pRTP59                                                                                                  - - - - - -

        - - - - - C - - - - - - - - - - - - - - - - - T - - - - - - - - - - - - - - - - - G * * * - - - - - - - - - - - - - - -
        A G C C G A C C T G T C C G C A G G C G C C G C C G C C C C A G G G T G T C C C G A C G T C G T C G T C G C A G G A G A
         ser   arg   pro   val   arg   arg   arg   arg   arg   pro   arg   val   ser   arg   arg   arg   arg   arg   arg   arg
        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

        - - - - - - - - - - - - - - - - - - - - - - C - - - C - - - - - * * - - - - - - - - - - - A - - - A - - - - - - - - - A
        G G A G G C C G C A G G A G G C G T T A G A T A G G A C G G G T A G A A C C T A C C T G A C C T A T C C G C C C C C T *
         gly   gly   arg   arg   arg   arg   term
        - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - *

        - - - - * - - - - - - - - - - A G A C T C - - - - A C - - - - - - - - C - - - - A - - - - A - - - - - - - - - - - - - -
        C C G G G T T C T C C C T C C C * * * * * G A C C C T T G G T A G T G T A G A G G T G T T * A A A G T C T G C T T A A A
        - - - - - - - - - - - - - - - - * * * * * - - - - - - - - - - - - - - - - - - T - - - - - * - - - - - - - - - - - - - -

        - - - - - - - - - - - - G - - - - - - - - - POLY A
        T A A A A G A T G G G C * T T T T A A C T A A A A C T
        - - - - - - - - - - - - G - - - - - - - - - POLY A
```

Figure 5: Comparison of p101 coding sequence with cDNA's. The dashes represent the same base in the cDNA's as in the genomic sequence, p101. The stars denote gaps introduced in order to maintain homology. cDNA sequences are taken from Gedamu _et al_ (1981). The amino acids encoded are written below their respective codons.

using the wheat germ cell free translation system, predominantly to protamine component C_{II} (5).

The sequence of the 5' end of the mRNA encoded by p101 does not match exactly any of the RNase T1 resistant oligoribonucleotides sequenced by Davies _et al_. (21). This is probably due to the fact that the 235 base long class of protamine mRNA comprises only about 10-15% of the total mRNA population and probably only about half of this size class or 5-8% of the total protamine mRNA codes for component C_{II}. As a result, the mRNA from p101 is most likely a minor component of the total mRNA and, as such, was not in sufficient quantity to be observed in the RNase T1 resistant fragments. In Figure 6, it may be seen that there are four mismatches when the 5' leader of p101 is compared to the "21" fragment from Davies _et al_. (21). This is a homology of 85%. This raises the question of whether p101 might be a derivative of a longer gene of either the 245 or 275 base long classes (5) that has suffered a deletion and/or several point mutations resulting in a shorter 5' untranslated region. The shorter leader on this new mRNA may lead to an increased efficiency of translation _in vivo_ (22). This would explain

```
A T T T T A T C C A T C A A T C A C A A T G
U   C U U A U C U A U C A A U C A C U A U G
```

Figure 6: Comparison of the 5' leader sequence of p101 with the "21" fragment of an RNase T1 digest of total protamine mRNA. The p101 sequence is on top. The "21" sequence is on the bottom and is from Davies et al. (1979). The areas of homology are boxed.

the discrepancy between the abundance of component C_{II} polypeptides (1,5) and the paucity of mRNA for this component in the developing testis (5).

Several other protamine genomic clones are in the process of being analyzed in detail in our laboratory and should provide sequence information which will allow the evolutionary history of the protamine gene family to be traced in more detail. Furthermore, comparison of the sequences 5' to the coding regions may also shed some light on how these apparently widely scattered genes are coordinately controlled during testis development.

A final comparison of 5' regulatory sequences is made in Figure 7, where the regulatory sequences from sea urchin histone genes (18) are presented along with those of p101. In general, the homology between these sets of sequences is quite marked, although the spacing between the "capping signal" and the initiation codon is significantly shorter at 10 bp in p101 than the range, 46-69, seen in the sea urchin histone gene 5' leader sequences. The extra sequences in the histone genes at this point may correspond to that which might have been present in progenitor protamine genes as discussed above. Of special note is the presence of a pentanucleotide sequence GATCC about 9 nucleotides 5' to the TATA box. This is present in the sea urchin histone genes that have been sequenced (18) and there is a similar pentanucleotide (A → C at position 2 of the sequence) at the same place in the p101 sequence. Busslinger et al. (18) consider that this may be a trancription regulator for the histone genes in sea urchins. It would be particularly interesting if this pentanucleotide sequence were found to be characteristic of the genes for basic structural proteins of chromatin and

```
GCTCC --- 9 --- CTATAAAAGG --- 17 --- ACATTTT --- 10 --- CACAATG

GATCC -- 8-10 -- GTATAAATAG -- 16-24 -- (C/T) CATTC (A/G) -- 46-69 -- (C/A) A (C/T) CATG
```

Figure 7: Comparison of 5' p101 sequences with those of sea urchin histone genes. The p101 sequences are on top. The consensus sea urchin histone gene 5' sequences are on the bottom. The numbers represent the variable nucleotides between these sequences.

thus imply that these genes are under some special co-ordinate control. Further investigation of these types of genes will help determine if this is the case.

ACKNOWLEDGEMENT

We appreciate the many helpful discussions with Dr. Kostas Iatrou. This work was supported by the Medical Research Council of Canada through an operating grant to G.H.D.. Through a post-doctoral fellowship to M.A.W., the Alberta Heritage Foundation for Medical Research through a post-doctoral fellowship to J.C.S. and the W.H. Davies Foundation through a pre-doctoral fellowship to J.M.A.

*Present address: Connaught Research Institute, Building 17, 1755 Steeles Avenue West, Willowdale, Ontario, Canada.

+Present address: Biochemistry Division, Department of Chemistry, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.

REFERENCES

1. Gedamu, L., Wosnick, M.A., Connor, W., Watson, D.C., Dixon, G.H. and Iatrou, K. (1981) Nuc. Acids Res. 9, 1463-1482.
2. Marushige, K. and Dixon, G.H. (1969) Dev. Biol. 19, 397-414.
3. Sakai, M., Kurizama, Y.F. and Muramatsu, M. (1978) Biochem. 17, 5510-5515.
4. Louie, A.J. and Dixon, G.H. (1972) J. Biol. Chem. 247, 5498-5505.
5. Gedamu, L. and Dixon, G.H. (1979) Nuc. Acids Res. 6, 3661-3672.
6. Sakai, M., Fujii-Kuriyama, Y., Saito, T. and Muramatsu, M. (1981) J. Biochem. 89, 1863-1868.
7. Jenkins, J.R. (1979) Nature 279, 809-811.
8. Wosnick, M.A., Connor, W., Aiken, J., Gedamu, L. and Dixon, G.H. (1981) Proc. Can. Fed. Biol. Soc. 24, 285.
9. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
10. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) Proc. Nat. Acad. Sci. USA 72, 1184-1188.
11. Southern, E. (1975) J. Molec. Biol. 98, 503-517.
12. Katz, L., Kinsburg, D. and Helsinki, D.R. (1973) J. Bacteriol. 114, 577-591.
13. Maxam, A. and Gilbert, W. (1980) Methods in Enzymol. Vol. 65, Grossman, L. and Moldave, K. (eds), Academic Pres, N.Y., pp 499-560.
14. Gedamu, L., Iatrou, K. and Dixon, G.H. (1978) Biochem. J. 171, 589-599.
15. Sharp, P.A., Berk, A.J. and Berget, S.M. (1980) Methods in Enzymol. Vol. 65, Grossman, L. and Moldave, K (eds), Academic Press, N.Y., pp 750-768.
16. Benoist, C., O'Hare, K., Breathnach, A. and Chambon, P. (1980) Nuc. Acids Res. 8, 127-142.
17. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
18. Busslinger, M., Portman, R., Irminger, J.C. and Birnstiel, M. (1980) Nuc. Acids Res. 8, 957-977.
19. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-250.

20. Sollner-Webb, B. and Reeder, R.H. (1979) Cell 18, 485-499.
21. Davies, P.L., Dixon, G.H, Simoncsits, A. and Brownlee, G.G. (1979) Nuc. Acids Res. 7, 2323-2345.
22. Kozak, M. (1981) Curr. T. Micr. 1981, 81-123.
23. Connor, W. and Dixon, G.H., manuscript in preparation.
24. van Dongen, W., de Laaf, L., Zaal, R., Moorman, A. and Destrée, O. (1981) Nuc. Acids Res. 9, 2297-2311.
25. Wagner, M.J., Sharp, J.A. and Summers, W.C. (1981) Proc. Nat. Acad. Sci. USA 78, 1441-1445.
26. Ohno, S. and Taniguchi, T. (1981) Proc. Nat. Acad. Sic. USA 78, 5305-5309.
27. Dolan, M., Sugarman, B.J., Dodgson, J.B. and Engel, J.D. (1981) Cell 24, 669-677.
28. Sharp, P.A. (1980) Cell 23, 643-646.