Volume 11 Number 14 1983 Nucleic Acids Research

Sequence homologies in the protamine gene family of rainbow trout

J.M.Aiken, D.McKenzie, H.-Z.Zhao, J.C.States and G.H.Dixon

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

Received 1 March 1983; Revised and Accepted 24 June 1983

ABSTRACT

We have sequenced five different rainbow trout protamine genes plus their flanking regions. The genes are not clustered and do not contain intervening sequences. There is an extremely high degree of sequence conservation in the coding and 3' untranslated regions of the gene. Downstream sequences exhibit little homology though conserved regions are found 250 base pairs 3' to the gene. There are four regions upstream of the gene that are highly conserved in the six clones, including the canonical Goldberg- Hogness box which is 45 base pairs 5' to the coding region. A second homologous region is found 90 bases upstream. Although in the same approximate location as the CAAT box found upstream of other genes, it does not contain the canonical CAAT sequence. Further upstream of the protamine genes at -115 there is an A-T rich sequence while a 25 base pair conserved sequence is located 150 bases upstream. In addition we report the presence of a potential Z-DNA region of predominantly A-C repeats approximately one kilobase downstream of one of the genes.

INTRODUCTION

Evolutionarily conserved DNA sequences have been strongly implicated in the regulation of eukaryotic gene expression. Comparison of the 5' flanking regions of various genes has resulted in the discovery of consensus sequences (reviewed by 1), such as the Goldberg-Hogness box (25 base pairs upstream of the RNA initiation site) and the CAAT box (approximately 80 base pairs 5' to the gene). Evidence of the specific function of these regions in the transcriptional process has been somewhat equivocal. *In vitro* studies indicate only the Goldberg-Hogness box is necessary for faithful transcription (2,3). *In vivo* experiments, however, demonstrate that both the Goldberg-Hogness and the CAAT box are required for efficient transcription (4,5,6). The Goldberg-Hogness box is necessary for the specificity of initiation, while the CAAT region affects transcriptional efficiency. The sea urchin histone H2A gene seems to be an exception, as the deletion of the CAAT region does not affect the rate of transcription (7). Sequences far from the mRNA initiation site have also

been shown, in vivo, to be required for transcription. Regions 200 base pairs upstream of the SV40 early gene (8,9) and 184 to 524 bases upstream of the histone H2A gene (10) also influence the transcriptional process.

Sequence analysis of the protamine gene family in the rainbow trout (Salmo gairdnerii) provides an excellent opportunity to identify conserved sequences of genes transcribed in the same tissue at the same developmental stage. Protamine mRNAs are transcribed only during the primary spermatocyte stage during the maturation of the trout testis (11). After a period of storage as inactive post-ribosomal mRNP's (11,32,33) translation begins at the middle spermatid stage when they direct the synthesis of at least three or four different protamines (11). Solution hybridization experiments indicate the presence of 4-6 genes per haploid trout cell (12). Thus comparison of the flanking regions of several protamine genes should allow the identification of those sequences important in their regulation.

We present here coding and flanking sequences of five different protamine genes. Comparison of these genes reveals excellent homology in the coding and 3' untranslated regions of the genes, with the upstream region characterized by several blocks of homology. In addition, we report the presence of an A-C rich region located one kilobase downstream of one of the protamine genes.

MATERIALS AND METHODS

Restriction endonucleases were purchased from Bethesda Research Laboratories and from New England Biolabs. DNA polymerase (Klenow fragment) was obtained from Boehringer Mannheim. Bacterial alkaline phosphatase and T4 polynucleotide kinase were products of Bethesda Research Laboratories. Radioactive nucleotides were purchased from New England Nuclear and ICN.

Library Screening

Construction of the rainbow trout Eco R1 Charon 4A library has already been reported (13). The library was screened according to the plaque hybridization procedure of Benton and Davis (14) using a ^{32}P-labelled protamine cDNA probe, pRTP 242 (15).

Subcloning

5 micrograms of recombinant phage DNA digested with an appropriate restriction endonuclease(s) were ligated to 50-100 ng of pBR 322 (or a pBR 322 derivative constructed to contain a Kpn 1 site) digested with the same restriction enzyme(s). This ligated DNA was then transformed into RR1 by

the method of Norgard et al. (16). The resulting colonies were then screened with ^{32}P labelled pRTP 242 to isolate protamine gene-containing subclones.

DNA Sequencing

The DNA's of recombinant subclones were digested with restriction endonucleases, dephosphorylated and 5'-end labelled (17). The fragments were then digested with a second endonuclease, electrophoresed on a 1.5% agarose gel and the fragments of interest isolated by electroelution. DNA sequencing reactions were performed as described by Maxam and Gilbert (17). The reaction products were separated on 8% polyacrylamide-urea, 10% polyacrylamide-formamide or 20% polyacrylamide-urea gels.

RESULTS

Isolation and Initial Characterization of Protamine Clones

To ensure a complete screening of the Eco R1 Charon 4A trout library, 1.2×10^6 clones equivalent to 1.8×10^{10} base pairs or 3.2 genomes of trout DNA were searched. Such a screening provides a 97% probability of isolating any given protamine gene (18). Forty-nine of the clones hybridized to a ^{32}P-labelled protamine cDNA probe, pRTP 242 (15), and were plaque purified. Characterization by restriction endonuclease mapping coupled with Southern blotting led to the identification of five different classes of clones. Restriction maps, location of the genes, and direction of transcription of these five clones plus TP 101, a gene previously isolated and sequenced in our laboratory (13), are provided in Figure 1.

Several important observations can be made. First, in each clone, there is only a single small region that hybridizes to the protamine cDNA probe. For example, with TP 16 and TP21, the region containing the entire protamine gene is less than one kilobase in length. Secondly, there are a number of striking similarities between some of the clones. The restriction endonuclease patterns in the vicinity of the protamine genes in TP14, TP15 and TP17 are quite similar. There is also considerable homology between TP16 and TP21. TP101 has a restriction pattern distinct from the other five.

Sequence Data

Restriction fragments containing protamine cDNA hybridizing regions were isolated from the five clones and subcloned into either pBR322 or a pBR322 derivative constructed to contain a Kpn 1 site. Restriction maps of these subclones were then determined (data not presented) and fragments

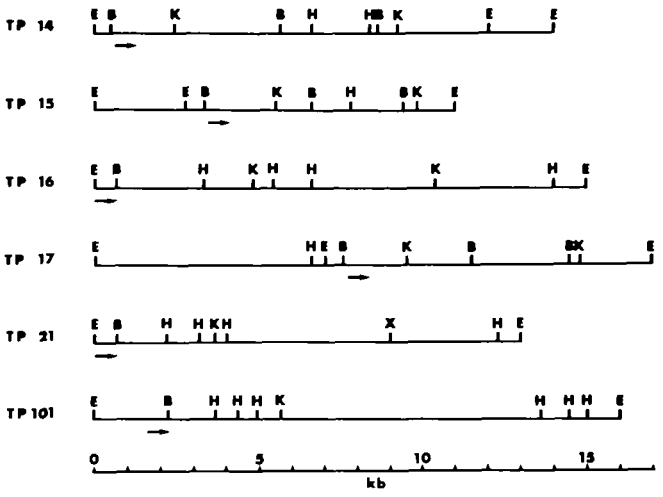

Figure 1: Restriction maps of six protamine gene-containing inserts isolated from an Eco Rl Charon 4A library. The arrows indicate the location, approximate size and direction of transcription (5' -- 3'). B = Bam Hl, E = Eco Rl, H = Hind III, K = Kpn I, X = XbaI.

of interest were 5'-end labelled and sequenced. We present here sequence data of the protamine coding region as well as 270 bases upstream and 300 bases downstream (Figures 3-5). With TP16 and TP21, the upstream sequence data is limited to 140 bases (up to the Eco Rl terminus of these two clones).

The degree of sequence homology between the different clones in the coding region and flanking sequences (Figure 2) reveals a number of general trends. The 5' flanking region of the protamine genes contain large blocks of sequences common to all of the genes. This homology extends 218 base pairs upstream of the coding region, at which point the homology between TP 101 and the other clones breaks down.

Comparisons of the coding sequence in the six protamine clones (Figure 3) shows that intervening sequences are absent. The clones are either identical to or exhibit slight changes from protamine cDNAs, pRTP 43 and pRTP 242 (15). The potential protein sequence, represented by the coding sequences of the genes is also presented in Figure 3. The four different protamines are all 32 amino acids in length and exceedingly arginine-rich (21-22 residues). The nucleotide sequence in the coding region of the six genes varies at only six positions. Three of these changes result in alterations of the protein sequence. Position 27 can either contain a

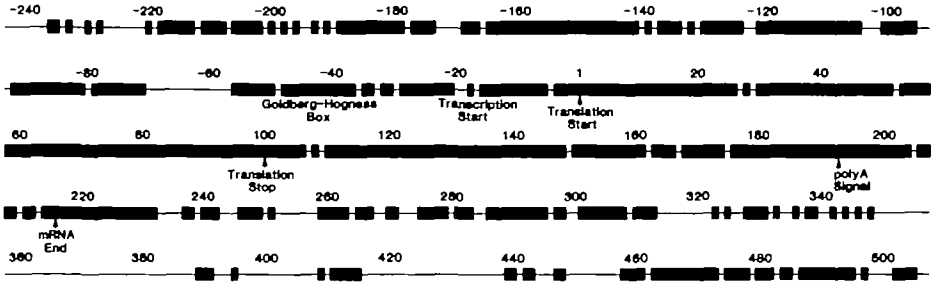

Figure 2: Comparison of sequence homology of the six protamine genes plus flanking regions. Solid bars indicate sequences completely conserved in all clones containing the indicated sequence (4 clones for -240 to -140, 6 clones for -140 to +510). Note that initiation site for translation rather than transcription is numbered +1.

```
1                   20                  40
*                   *                   *
ATG CCC AGA AGA CGC AGA TCC TCC AGA CCA CCT GTC CGC AGG CGC CGC CGC   TP  14
met pro arg arg arg arg ser ser arg pro pro val arg arg arg arg arg

ATG CCC AGA AGA CGC AGA TCC TCC AGA CGA CCT GTC CGC AGG CGC CGC CGC   TP  15
met pro arg arg arg arg ser ser arg arg pro val arg arg arg arg arg

ATG CCC AGA AGA CGC AGA TCC TCT AGC CGA CCT GTC CGC AGG CGC CGC CGC   TP  16
met pro arg arg arg arg ser ser ser arg pro val arg arg arg arg arg

ATG CCC AGA AGA CGC AGA TCC TCC AGA CGA CCT GTC CGC AGG CGC CGC CGC   TP  17
met pro arg arg arg arg ser ser arg arg pro val arg arg arg arg arg

ATG CCC AGA AGA CGC AGA TCC TCC AGC CGA CCT GTC CGC AGG CGC CGC CGC   TP  21
met pro arg arg arg arg ser ser ser arg pro val arg arg arg arg arg

ATG CCC AGA AGA CGC AGA TCC TCC AGC CGA CCT GTC CGC AGG CGC CGC CGC   TP 101
met pro arg arg arg arg ser ser ser arg pro val arg arg arg arg arg

          60                  80                  100
          *                   *                   *
CCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA GGC CGC AGG AGG CGT TAG   TP  14
pro arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter

CCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA GGC CGC AGG AGG CGT TAG   TP  15
pro arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter

GCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA CGC CGC AGG AGG CGT TAG   TP  16
ala arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter

CCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA CGC CGC AGG AGG CGT TAG   TP  17
pro arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter

GCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA GGC CGC AGG AGG CGT TAG   TP  21
ala arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter

CCC AGG GTG TCC CGA CGT CGT CGC AGG AGA GGA GGC CGC AGG AGG CGT TAG   TP 101
pro arg val ser arg arg arg arg arg arg gly gly arg arg arg arg ter
```

Figure 3: DNA sequence of the coding region of the six protamine genes. The potential amino acid residues are given beneath each codon. Regions of nucleotide heterogeneity are boxed. TP 101 data is from States et. al. (13). Met is the initiation codon and is not found in mature protamine proteins.

cytosine or an adenine. Thus either arginine (AGA) or serine (AGC) is present at the eighth amino acid residue. TP 14 differs from the other genes in having a cytosine (instead of a guanine) at position 29 ---a change that results in a proline (CCA) in place of arginine (CGA). Either guanine or cytosine is present at position 52, a shift that produces either an alanine (GCC) or proline (CCC) at the 17th amino acid residue.

The DNA sequence data of TP14, TP16 and TP21 provide the first evidence of such protamine genes in the trout. Protamine sequences thus far published (15,19) have not indicated a proline-proline sequence in the 8th and 9th amino acid residues of the protein sequence or that an alanine can be present at the 17th residue. It is probable such sequences are either minor components or intra-specific variants, as even after careful purification of protamine polypeptides by CM-52 chromatography, amino acid analysis indicates none of the 3-4 residue peaks shows integral values for isoleucine, valine, proline, or alanine (20,21).

The first 25 base pairs upstream of the coding region display a high degree of conservation between the six clones. Though protamine mRNA initiation occurs in this region, (13,22), the canonical CAP sequence, GTTGCTCCTNAC, (1) is not evident. Studies investigating the precise point of protamine mRNA initiation have produced conflicting results. States et al. (13) presented data indicating the initiation of protamine transcription of TP101 is at position -19. Gregory et al. (22) studying a gene almost identical to TP101 found protamine transcription to start five bases downstream at position -14. Our sequence data would support the latter observation as position -14 and surrounding nucleotides are conserved, whereas position -19 is variable. Regardless of which transcriptional initiation point is correct, the protamine mRNA 5' untranslated leader sequences are exceptionally short (either 14 or 19 bases).

The Goldberg-Hogness or TATA box is located 47 base pairs upstream of the protamine coding region. It has been noted (22), that this protamine Goldberg-Hogness region is identical to that of the adenovirus major late promoter (23) and conalbumin promoter (24). This eleven base pair sequence, CTATAAAAGGG, is totally conserved in all our clones.

There are three other regions upstream of the coding region that are conserved in all the protamine genes sequenced (Figure 4). Approximately 90 bases upstream (positions -92 to -71) is a 22 base pair region (CATCATTTATCC(A/C)ATAATGACA) about the same distance upstream as the CAAT

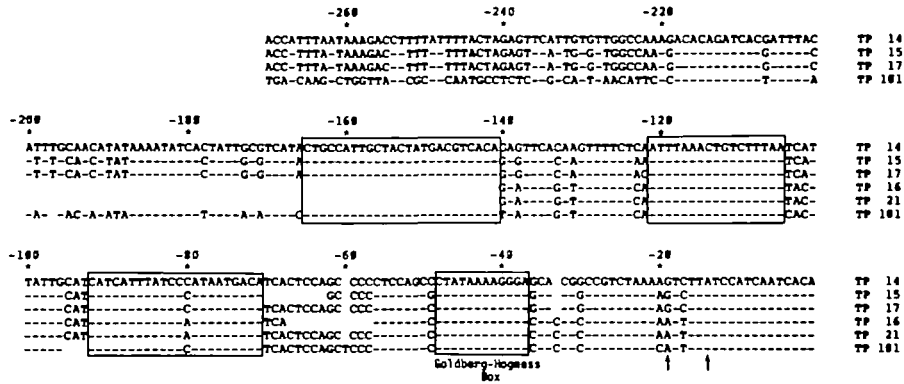

Figure 4: 5' flanking region of the protamine genes. Dashes indicate nucleotides that are completely homologous in all the clones. The arrows indicate possible CAP sites (see text). The location of the Goldberg-Hogness region and other extensive, highly conserved sequences are boxed. Blanks have been inserted to indicate gaps which have been introduced for optimal sequence alignment. Due to lack of homology between +303 to +402 the clones have been organized into two subgroups, with TP15 and TP17 being compared to TP14 while TP 21 and TP101 are compared to TP16.

or Chambon box found in many other gene systems (25). A second conserved region (ATTTAAACTGTCTTTAA) is A-T rich, 17 nucleotides in length, and is located 121 base pairs from the coding region. The third conserved sequence (CTGCCATTGCTACTATGACGTCACA) in the 5' flanking region extends for 25 nucleotides (-165 to -141).

Sequences immediately downstream of the protamine coding region exhibit an exceedingly high level of homology (positions 103 to 232). Contained within this region is the canonical polyadenylation signal AATAA and the point at which the mature protamine message ends. cDNA data indicates that the sequence of the last four nucleotides prior to the poly A tail of the protamine message is AACT (positions 211-215). Downstream from this sequence is AAAA, also conserved. Thus the first four adenines of the poly A tail could be coded in the genome and the end of the message could be up to four nucleotides further downstream at position 219. As noted by Benoist et al (26), this seems to be a general phenomenon. In the majority of eukaryotic genes studied, the first residues of the poly A tail could be transcribed from genomic DNA.

Whereas the coding region and much of the 3' untranslated region of the message is exceedingly G-C rich, the 3' end of the message (positions 179-215) is A-T rich. In spite of little sequence conservation, this trend

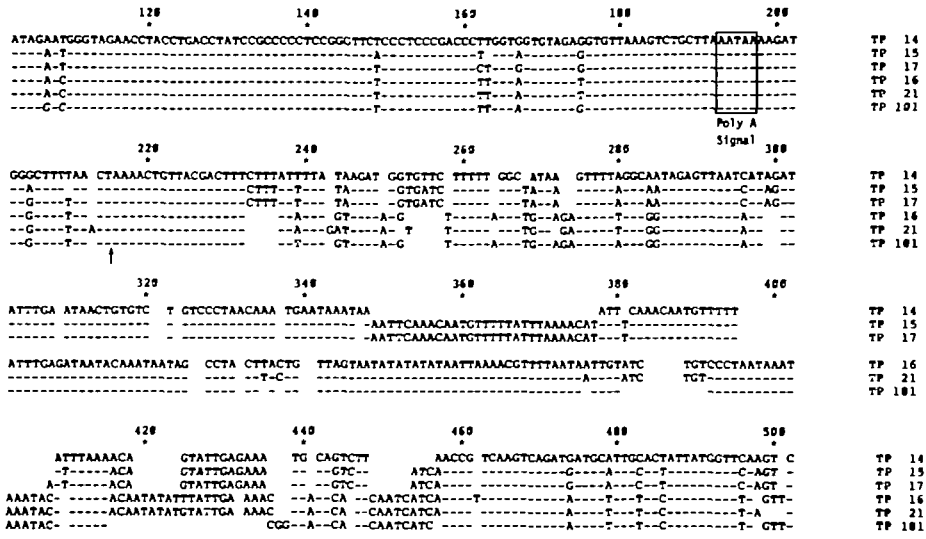

Figure 5: 3' flanking region of the protamine genes. Dashes indicate nucleotides that are completely homologous in all the clones. The poly A addition signal is boxed and the mRNA end is indicated by an arrow. Blanks have been introduced for optimal sequence alignment.

continues downstream from the message. The adjacent 76 nucleotides of TP 15 (positions 233-313) and 80 nucleotides of TP16 (positions 237-324) contain 76% of the sequence as either adenine or thymine. Little homology is evident from the end of the message for 225 bases (positions 232-457). At this point, blocks of conserved sequences predominate for the remaining 60 nucleotides.

A-C Rich Region

Sequence analysis of regions adjacent to one of the protamine gene (TP15) has resulted in the discovery of an exceedingly A-C rich region approximately 1 kilobase downstream from the gene (Figures 6 & 7). This region extends for 129 nucleotides and contains several blocks of A-C repeats (Figure 8). Approximately two-thirds of the sequence consists of a highly conserved 46 base pair internal tandem repeat. Though A-C repeats predominate, other purine-pyrimidine combinations are also common. In total, 87% of the sequence consists of alternating purine-pyrimidine dinucleotides.

Preliminary hybridization data (D. McKenzie, unpublished observations) of the other protamine clones suggests the A-C repeat region is common to TP 14 and TP 17 as well. The apparent evolutionary conservation of this region suggests it may have some functional significance.

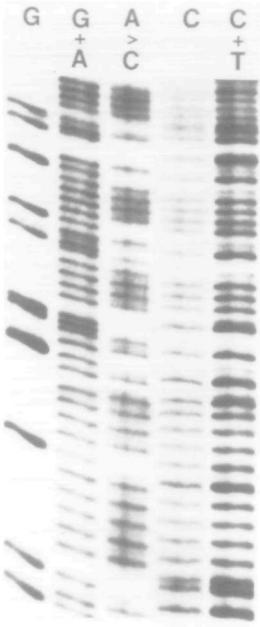

Figure 6: Sequencing gel containing a portion of the large A-C rich region located approximately one Kilobase downstream of one of the protamine genes (TP 15).

DISCUSSION

Lack of Introns

We have sequence information on six different protamine genes. A striking characteristic of these genes is a lack of introns --- a relatively uncommon feature shared by few other genes including those for the histones (27), human interferon genes (28), yeast glyceraldehyde-3-phosphate dehydrogenase (29) as compiled in (30). There is evidence indicating that a major function of intervening sequences may be to control the post-transcriptional processing of the mRNA's (31). Selective processing of introns may control the entrance of messages into

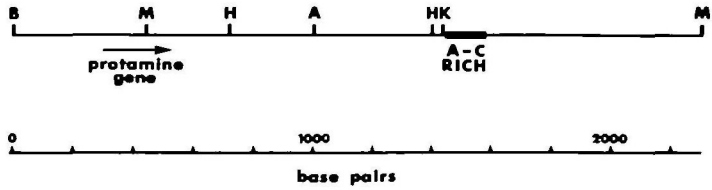

Figure 7: Restriction map of the region of TP 15 containing the A-C rich sequence (solid bar) and its proximity to the protamine gene. A = Ava II, B = Bam Hl, H = Hind III, K = Kpn 1, M = Msp I.

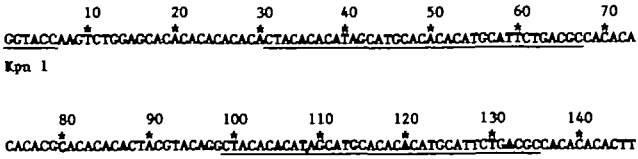

Figure 8: DNA sequence of the A-C rich region. The Kpn 1 site is indicated and the 46 base pair internal repeat is underlined.

the cytoplasm. Protamine expression is known to be under strong translational control (11,32,33). Protamine transcription occurs early in spermatogenesis (at the primary spermatocyte stage) while translation occurs late (at the spermatid stage). The message is stored, prior to translation, in the cytoplasm as an inactive mRNP particle. It is possible that such strong translational control eliminates the necessity for a regulatory step at the level of mRNA processing. Thus intervening sequences may not be necessary for genes that are solely under translational control. We are not predicting that other translationally controlled gene products (such as stored maternal mRNA in oocytes) will be intronless. Indeed the majority of such messages are likely also to be produced at other developmental stages where translational control is not evident and thus may be under control of other mechanisms including post-transcriptional processing. However it seems probable that gene products, such as protamines, that are always under strong translational regulation, would have little need for post-transcriptional processing controls.

Protamine Genes Are Not Clustered

The protamine genes we have isolated are not clustered but rather occur singly. Each of the protamine cDNA hybridizing regions in the clones (Figure 1) contains only one protamine gene. It is possible in TP16 and TP21, which have protamine genes abutting the EcoR1 terminus of the clones, that there could be other protamine genes adjacent to that EcoR1 end. However, in TP17 at least a seven kilobase region on each side of the gene does not contain another protamine gene. Thus our evidence strongly suggests that, unlike other multigene families such as the histone and globin gene clusters, protamine genes occur singly.

Upstream Homology

One of the initial objectives of this work was to isolate and compare a number of different protamine gene sequences. Through this comparison,

those DNA sequences important in the regulation of the gene might be inferred. Figure 2 graphically illustrates the degree of homology within the protamine coding regions and its flanking regions. Regions that are variable can probably be ruled out as having functional significance. Thus at this time we are not attaching any importance to a potential small reading frame located 150 nucleotides upstream from the protamine coding region or to an inverted repeat in the same region because these sequences are not conserved in all the clones.

Sequence comparisons with other gene systems (1,25) indicates the existence of three regions upstream of eukaryotic genes as having a functional role in transcription.

The first of these, the Goldberg-Hogness box, occurs approximately 25 base pairs before the CAP site. In vitro and in vivo data indicate that this sequence is essential for accurate initiation of the mRNA transcript (8,34,35). The protamine Goldberg-Hogness region is conserved in all our genes and is identical to promoter regions of adenovirus major late and conalbumin genes (22).

A second conserved 5' flanking region of eukaryotic genes is located approximately 80 base pairs upstream of mRNA initiation and often contains the sequence CAAT (25). Protamine genes do not have the consensus CAAT sequence, but exhibit a large block of homology approximately 75 nucleotides 5' to the proposed mRNA initiation points. Interestingly, two of the Drosophila cuticle gene sequences, genes III and IV (36), also lack a CAAT sequence. Instead, both genes have the conserved sequence TGCATCA, a sequence also common to all our protamine clones in this region (positions -97 to -88).

There is evidence that sequences even further upstream affect the transcription level (8,10). In the protamine genes, there are two other conserved regions, an A-T rich one 115 bases upstream of the coding region and a 25 base pair region 160 nucleotides 5' to the genes.

Between the TATA box and the second block of homology is a highly variable region (positions -57 to -70) characterized by numerous deletions. Though there is a lack of homology, the length of the segments exhibit some similarities. TP17, TP21 and TP101 have 13 or 14 bases in this region while TP16 and TP15 have 3 and 5 nucleotides present, respectively. The difference between the two groups is 8-11 bases or approximately one turn of the B-form DNA helix in length. Thus, though the region seems exceedingly variable, in every clone DNA sequences upstream of this

variable region will have a similar spatial orientation to downstream sequences.

Downstream Homology

Sequences downstream of the protamine coding region show an extremely high level of sequence conservation in the 3' untranslated region of the genome. This homology extends 17 nucleotides past the end of the message, an indication that these post-message sequences may have some functional importance possibly in transcription termination.

Further downstream, the homology diminishes with little sequence conservation apparent from positions 232 to 457. The homology increases dramatically once more, approximately 250 base pairs downstream from the gene (position 458). This could represent another region important to the functioning of the protamine gene or may be the first indication of the presence of an adjacent 3' gene.

We have seen no evidence of 3' sequences analogous to the 23 base pair inverted repeat which is involved in the termination of transcription of histone mRNAs (37,38) and indeed *in vitro* transcription studies (39) of a cloned protamine gene indicate that termination does not occur at position 226 but that the RNA transcripts read through to the end of the DNA restriction fragment containing the protamine gene.

The presence of four adenines immediately adjacent to the end of the protamine message (postions 216-219) hints at an interesting possibility --- that protamine genes may have evolved from a processed gene. Inclusion of a portion of the poly A tail as well as loss of introns is often associated with these pseudogenes (40,41,42). Thus some similarities between processed genes and protamine genes are evident, however the absence of abnormalities in the coding and regulatory sequences as well as *in vitro* transcription evidence (39) indicates all these protamine genes are functional.

Coding Region Homology

Two-thirds of the amino acid residues in protamines are arginine, an amino acid that has the potential of being encoded by six different codons, CGN or AG(A/G). Thus numerous changes in the nucleotide sequence of the arginine codon would still produce the same amino acid. Considering the extremely high content of arginine residues in protamine, without some form of selection pressure, considerable variability in the nucleotide sequence of the arginine codons could be expected. As noted above, the degree of homology between the different clones is exceedingly high in the coding

region. Differences in three different amino acid positions (Figure 3) result in four different protamine polypeptides. A total of four nucleotide changes are found in the coding region --- very close to the minimum number of three nucleotide changes necessary to produce the three different residues. Only with a single residue (in TP16) is there a nucleotide shift (C to T) that could be considered "neutral" as the coding sense of the codon is not altered. The remaining 113 nucleotides of the protamine message (3' untranslated) contains only eight positions that vary. This exceptional degree of homology in the message region argues strongly for the existence of some mechanism by which this high level of homology is maintained. We feel this mechanism is likely to be maintenance of protamine mRNA secondary structure. As previously noted, protamine production is under translational control. The protamine message exists in the cytoplasm, during the early stages of spermatogenesis, as an inactive mRNP particle (33) which is accumulated and stored for a period of 3 - 4 weeks. Only during the spermatid stage of development does the population of mRNP's bind to the polysomal fraction. The structure of the inactive mRNP probably depends critically on the secondary structure of the protamine mRNA. Several other lines of evidence indicate that protamine mRNA has a well-defined secondary structure; first in the melting curve of protamine mRNA there is a high melting component (T_m=76°C) (43) and secondly a region of three guanine residues in the 3' non-coding region of native protamine mRNA is resistant to T1 ribonuclease digestion (44). Recent preliminary studies in collaboration with Dr. M. Zuker, Division of Biology, National Research Council, Ottawa, using computer analysis of the secondary structure of protamine mRNA also reveals extensive base pairing.

A POTENTIAL Z DNA REGION

The alternating purine-pyrimidine rich sequence found adjacent to one of the protamine genes (TP 15) raises the possibility of a B to Z DNA conformational change being involved in protamine gene expression. Crystallographic and circular dichroism data has shown alternating purine-pyrimidine co-polymers such as poly d(GC)-poly d(GC) (45,46,47) and poly d(AC)-poly d(GT) (48,49,50) to have the ability to assume the Z DNA or left-handed conformation both in the crystalline state and in solution. Evidence of the existence of Z DNA in vivo has been obtained through the use of Z DNA specific antibodies (51,52). Hamada et al. (53) have presented evidence indicating that A-C alternating sequences are highly conserved throughout eukaryotic evolution while G-C repeats are not

ubiquitous. Small blocks of A-C repeats have been reported for a human globin gene (54), a mouse immunoglobin gene (55), a human actin gene (53), and in our laboratory, in a region 5' to the rainbow trout histone H3 gene (56).

We have found a large (129 base pair) region exceedingly rich in alternating purine-pyrimidines, especially A-C repeats. This sequence is about one kilobase 3' to protamine gene, TP 15, --- a member of a gene family whose transcription is under strict developmental control during the maturation of sperm cells. There is evidence that a B to Z transition could have profound effects on chromatin structure (52,57). Thus this A-C rich region could potentially play a significant role in protamine gene regulation.

ACKNOWLEDGEMENTS

We thank Robert Moir and Dr. Brian Pentecost for critically reading the manuscript. We are grateful to Denise Rettie for assistance in the preparation of the manuscript. This research was supported by a Medical Research Council of Canada operating grant to G.H.D., a W.H. Davies Foundation predoctoral fellowship (J.M.A.) and the Alberta Heritage Foundation for Medical Research (pre-doctoral fellowship to D.M., visiting scientist award to H-Z.Z. and a post-doctoral fellowship to J.C.S.).

REFERENCES

1. Lewin, B. (1980) Gene Expression, Eucaryotic Chromosomes 2, John Wiley and Sons, New York.
2. Grosveld, G.C., Shewmaker, P.J. and R.A. Flavell (1981). Cell 25, 215-226.
3. Hu, S-L. and Manley, J.L. (1981) Proc. Natl. Acad. Sci. USA 78, 820-824.
4. Dierks, P., van Ooyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. USA 78, 1411-1415.
5. McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R. (1981) Cell 25, 385-398.
6. Mellon, P. Parker, V., Gluzman, Y. and Maniatis, T. (1981) Cell 27, 279-288.
7. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432-1436.
8. Benoist, C. and Chambon, P. (1981) Nature 290, 304-310.
9. Gruss, P. Dhar, R. and Khoury, G. (1981) Proc. Natl. Acad. Sci. USA 78, 943-947.
10. Grossohedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA 77, 7102-7106.
11. Iatrou, K. and Dixon, G.H. (1978) Fed. Proc. 37, 2526-2533.
12. Levy, B. and Dixon, G.H. (1978) J. Biol. Chem. 252, 2526-2533.
13. States, J.C., Connor, W., Wosnick, M.A., Aiken, J.M., Gedamu, L. and Dixon, G.H. (1982) Nuc. Acids Res. 10, 4551-4563.
14. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
15. Gedamu, L., Wosnick, M.A. Connor, W. Watson, D.C. Dixon, G.H. and

Iatrou, K. (1981) Nuc. Acids Res. 9, 1462-1482.
16. Norgard, M.V. Keem, K. and Monahan, J.J. (1978) Gene 3, 279-292.
17. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
18. Clark, L., and Carbon, J. (1978) Cell 9, 91-99.
19. Ando, T. and Watanabe, S. (1969) Int. J. Protein Res. 1, 221-224.
20. Ling, V., Jergil, B., and Dixon, G.H. (1971) J. Biol. Chem. 246, 1168-1176.
21. Louie, A.J. and Dixon, G.H. (1974) Can. J. Bioch. 52, 536-546.
22. Gregory, S.P., Dillon, N.D. and Butterworth, P.H.W. (1982) Nuc. Acids Res. 10, 7581-7592.
23. Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475.
24. Cordon, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
25. Paul, J. (1982) Bioscience Reports 2, 63-76.
26. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nuc. Acids Res. 8, 127-142.
27. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.D. and Birnstiel. M.L. (1978) Cell 14, 655-671.
28. Nagata, S., Mantei, N. and Weismann, C. (1980) Nature 287, 401-408.
29. Holland, J.P. and Holland, M.J. (1979) J. Biol. Chem. 254, 9839-9845.
30. Lipman, P.J. and Maizel, J. (1982) Nuc. Acids. Res. 10, 2723-2739.
31. Gruss, P., Lai, C-J., Dhar, R. and Khoury, G. (1979) Proc. Natl. Acad. Sci. USA 76, 4317-4321.
32. Louie, A.J. and Dixon, G.H. (1972) J. Biol. Chem. 247, 5498-5505.
33. Sinclair, G.D. and Dixon, G.H. (1982) Biochemistry 21, 1869-1877.
34. Baker, C.C., Herisse, J., Courtois, G., Galibert, F. and Ziff, E. (1979) Cell 18, 569-580.
35. Grosveld, G.C., deBoer, E., Shewmaker, C.K. and Flavell, R.A. (1982) Nature 295, 304-310.
36. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristron, J. and Davidson, N. (1982) Cell 29, 1027-1040.
37. Birchmeir, C., Grosschedl, R. and Birnstiel, M.L. (1982) Cell 28, 739-745.
38. Stunnenberg, H.G. and Birnstiel, M.L. (1982) Proc. Natl. Acad. Sci. USA 79, 6201-6204.
39. Jankowski, J. and Dixon, G.H. (1983) Manuscript in preparation.
40. Nishioka, Y., Leder, A. and Leder, P. (1980) Proc. Natl. Acad. Sci. USA 77, 2806-2809.
41. Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Nature 296, 321-325.
42. Karin, M. and Richards, R.I. (1982) Nature 299, 797-802.
43. Davies, P.L., Dixon, G.H., Ferrier, L.N., Gedamu, L. and Iatrou, K. (1976) Prog. Nuc. Acids. Res. Mol. Biol. 19, 135-155.
44. Davies, P.L., Dixon, G.H., Simoncsits, A. and Brownlee, G.G. (1979) Nuc. Acids Res. 7, 2323-2345.
45. Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686.
46. Crawford, J.L., Kolpak, F.J., Wang, A-H.J., Quigley, G.J., van Boom, J.H., van der Marel, G. and Rich, A. (1980) Proc. Natl. Acad. Sci. USA 77, 4016-4020.
47. Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R.E. 91980) Nature 286, 567-573.
48. Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L. (1980) Nature 283, 743-745.
49. Vorlickova, M., Kypr, J., Stokrova, S. and Sponar, J. (1982) Nuc. Acids Res. 10, 1071-1081.

50. Zimmer, C., Tymen, S, Marck, C. and Guschlbauer, W. (1982) Nuc. Acids Res. 10, 1081-1091.
51. Nordheim, A., Pardue, M.L., Lafer, E.M., Moller, A., Stollar, B.D. and Rich, A. (1981) Nature 294, 417-422.
52. Jovin, T.M., van de Sande, J.H., Zarling, D.A., Arndt-Jovin, D.J., Eckstein, F., Fuldner, H.H., Greider, C., Grieger, I., Hamori, E., Kalisch, B., McIntosh, L.P. and Robert-Nicoud, M. (1982) XLVII Cold Spring Harbor Symposium on Quantitative Biology.
53. Hamada, H., Petrino, M.G., and Kakunaga, T. (1982) Proc. Nat. Acad. Sci. USA 79, 6465-6469.
54. Miesfeld, R., Krystal, M. and Arnheim, N. (1980) Nuc. Acids Res. 9, 5931-5937.
55. Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255, 3691-3694.
56. Connor, W., Mezquita, J., Winkfein, R., States, J.C. and Dixon, G.H. (1983) Manuscript in preparation.
57. Nickol, J., Behe, M. and Felsenfeld, G. (1982) Proc. Nat. Acad. Sci. USA 79, 1771-1775.