Volume 13 Number 13 1985 Nucleic Acids Research

Isolation and sequence of cDNA clones coding for a member of the family of high mobility group proteins (HMG-T) in trout and analysis of HMG-T-mRNA's in trout tissues

B.T.Pentecost[1], J.M.Wright[2] and G.H.Dixon*

Department of Medical Biochemistry, Health Sciences Centre, University of Calgary, 3330 Hospital Drive N.W., Calgary, Alberta T2N 4N1, Canada

Received 25 February 1985; Revised and Accepted 27 May 1985

ABSTRACT

A specific oligonucleotide has been used to isolate a cDNA prepared from the mRNA for a trout High Mobility Group (HMG) protein closely related to trout HMG-T and bovine HMG 1 and 2 proteins. The sequence isolated more closely resembles bovine HMG-1 than the previously sequenced HMG-T protein in regions corresponding to the N terminal half of the protein. Northern blot analysis at low stringency indicated that 2 related sequences are expressed in total trout testis mRNA. Southern blots of total trout DNA indicate that several different forms of the homologous sequence are present in the trout genome and an estimate of copy number by dot-blot shows 4 HMG-T genes per trout sperm DNA equivalent. Analysis of mRNA from several trout tissues, including testis, liver and kidney indicates that expression of genes for histones and the larger HMG proteins in trout is not closely coupled.

INTRODUCTION

The high mobility group (HMG-) proteins are abundant (1) ($\simeq 10^6$ molecules per nucleus), widely distributed, chromosomal proteins. Two sub-classes have been described and characterized in detail from several different vertebrate chromatins. These two sub-classes differ in molecular weight, amino acid sequence and location in chromatin (2-4). The sub-class exemplified by bovine thymus HMG-1 and -2 (3) and trout testis HMG-T (2, 5) has a M.W. close to 29,000 daltons and an unusual three-domain structure (6). From the N-terminus, there are two tandemly arranged globular domains whose amino acid sequences show evidence of having been derived by partial duplication (6) followed at the C-terminus by a third domain possessing a characteristic, highly acidic sequence composed of long tracts of aspartyl- and/or glutamyl- residues (2, 3, 7, 8). The second sub-class of HMG- proteins has a smaller M.W. of 8-10,000 and members of this sub-class that have been characterized in detail are HMG-14 and -17 from bovine thymus (3) and H6 from trout testis (2).

In order to complement and extend our on-going studies of the

structure and function of trout testis HMG-T at the protein level (2, 5, 7 and D.C. Watson and G.H. Dixon, in preparation), we have isolated and sequenced a trout testis cDNA clone containing a reading frame which predicts an amino acid sequence very similar but not identical to that previously partially deduced for trout testis HMG-T (2, 5, 7). Our conclusion is that this cDNA codes for a second member of the HMG-T family of proteins.

We had previously isolated and characterized (8) a cDNA to the 3' coding and 3' untranslated regions of the message for bovine HMG-1. Isolation of that cDNA was aided by our ability to design an oligonucleotide probe with very low redundancy. We had thought that this bovine HMG-1-cDNA would facilitate isolation of a similar clone for the trout analogue of HMG-1, HMG-T, our eventual aim being to examine the expression of genes for HMG proteins during the terminal differentiation of trout testis, and in relation to histone genes, as well as to study the organization of such genes. Northern blot analysis unfortunately indicated that the bovine HMG-1 cDNA did not cross-hybridize specifically to total trout testis RNA, which we considered to be a good source of HMG-T message as testis nuclei are an excellent source of the protein (2, 5).

The oligonucleotide approach was, therefore, employed again to isolate a trout HMG cDNA using data from this laboratory on the primary structure of HMG-T (2, 7), and on codon usage in a trout histone H3 gene (9) which suggested preferential use of the GAC codon for aspartic acid. The selected portion of the protein sequence and the derived message and cDNA options are shown below:

Known protein sequence	Met-Asp-Asp-Asp-Asp-Asp
	U U U U U
Potential mRNA sequences	5' AUGGACGACGACGACGAC 3'
	A A A A A
DNA complement of mRNA	5' GTCGTCGTCGTCGTCCAT 3'

Only one, 16 mer, oligonucleotide sequence, TCGTCGTCGTCGTCCA, was prepared (a custom synthesis by Bio-logicals Ltd.). We considered that this oligonucleotide might bind, in addition to the specific Met-$(Asp)_5$ sequence in HMG-T (2, 7), at additional points, with varying degrees of mismatch, within cDNA sequences for the characteristic acidic domain of HMG-T which had the empirical formula $(GAX)_{28}$.

MATERIALS AND METHODS

Isolation of RNA

RNA was prepared from tissues other than testis by the method of Chirgwin et al. (10), and from testis by adjusting post-nuclear supernatants to 7.5 M guanidinium hydrochloride followed by further extractions with reducing volumes of guanidinium hydrochloride, as in the standard procedure, after precipitation with 0.5 volumes of ethanol. Polyadenylated RNA was selected from total RNA by oligo-dT cellulose affinity chromatography using the procedure of Bantle et al. (11).

Cloning Procedures

Double stranded cDNAs were prepared from trout testis polyadenylated RNA by the method of Wickens et al. (12). cDNAs were ligated into the Hind III site of linearized, dephosphorylated pBR322 following S1 nuclease treatment, attachment of synthetic Hind III linkers (13, 14), and selection of material of >500 bp on Sepharose 4B (14). The constructs were introduced into the E. coli strain DH1 by the method of Hanahan (15).

Screening

Bacterial colonies were screened by hybridization of the 5'-^{32}P radiolabelled oligonucleotide probe (8) to nitrocellulose replica platings of randomly arrayed colonies. Colonies were also screened, after isolation and analysis of the first positive clone, with nick-translated (16) cDNA using the hybridization medium described below under RNA blot analysis, but including 10% dextran sulphate, hybridising at 45°C and washing at 0.5x SSC, 0.5% SDS (SSC = 0.15 M NaCl, 0.015 M Na Citrate, pH 7.0) at 45°C. Positive colonies were rescreened until considered pure.

RNA Analysis

Electrophoresis of RNA was performed under denaturing conditions in agarose gels (17) and electroblotted to nylon membranes (Pall Biodyne A) without staining. Pre-hybridization and hybridization was carried out in 0.1% ficoll 400, 0.1% PVP360, 0.1% BSA, 0.75 M sodium chloride, 50 mM sodium phosphate, pH 6.5, 0.1% SDS, and 500 µg/ml yeast RNA with formamide at 50%. Typically nick-translated cDNA (2 x 10^8 dpm/µg) was used at 2 x 10^6 dpm/ml and hybridization carried out at 42°C for 12 hours, followed by washing to 0.2x SSC, 0.5% SDS at 50°C. Hybridization at low stringency was performed at 37°C and filters washed to 0.5x SSC at 50°C. When required, filters were prepared for reprobing by 'stripping' in 0.1x SSC, 0.1% SDS at 100°C, followed by pre-hybridization. Hybridization of ^{32}P-labelled oligonucleotide to RNA blots, and subsequent washing, was

performed using the conditions previously described for colony hybridization with oligonucleotide probes (8).

Analysis of Genomic DNA

Genomic DNA samples from individual trout testes were digested with restriction enzymes (the digestion was complete as judged by co-digestion of lambda DNA), subjected to electrophoresis in 0.8% agarose gels and electroblotted to nylon membrane (Pall Biodyne A). Hybridization to the nick-translated insert of pBP2 was carried out at 50°C in the medium described under RNA analysis and washing was to 0.1x SSC, 0.5% SDS at 60°C.

RESULTS AND DISCUSSION

Isolation of cDNAs

Six cDNAs, from 3000 colonies screened, hybridized to the oligonucleotide probe, suggesting a message level of ≈1% of the total poly A(+) fraction. These colonies were purified, DNA isolated and restricted to determine the insert size. In Fig. 1 the restriction map of the largest clone, pBP2 is shown together with the DNA sequencing strategy and in Fig. 2 the sequence is presented together with the translation product of the longest open reading frame (LORF). The 713 bp sequence was determined by the chemical procedures of Maxam and Gilbert (18) using the strategy shown in Fig. 1. Approximately 80% of the sequence was derived from both strands and analysis of other, shorter, cDNAs has confirmed the sequence

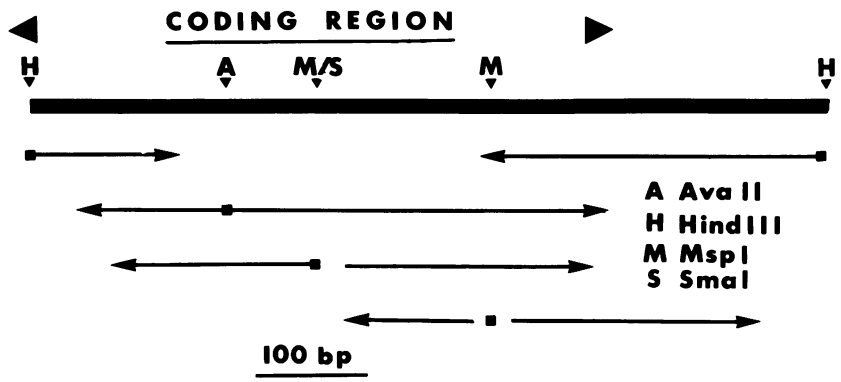

Figure 1
Restriction map and sequencing strategy for the trout specific HMG protein cDNA insert of pBP2.

```
Coding region
                  15                  30                  45                  60
GCT TCC GTC AAC TTC TCA GAG TTC TCC AAG AAG TGC TCT GAG AGA TGG AAG ACC ATG TCT
Ala Ser Val Asn Phe Ser Glu Phe Ser Lys Lys Cys Ser Glu Arg Trp Lys Thr Met Ser

                  75                  90                 105                 120
GCC AAG GAG AAG GGG AAG TTT GAG GAT CTG GCC AAA CTG GAC AAG GTG CGA TAT GAG AGG
Ala Lys Glu Lys Gly Lys Phe Glu Asp Leu Ala Lys Leu Asp Lys Val Arg Tyr Glu Arg

                 135                 150                 165                 180
GAG ATG AGG AGC TAC ATT CCT CCC AAG GGA GAG AAG AAG AAG AGG TTC AAG GAC CCC AAC
Glu Met Arg Ser Tyr Ile Pro Pro Lys Gly Glu Lys Lys Lys Arg Phe Lys Asp Pro Asn

                 195                 210                 225                 240
GCC CCC AAG AGA CCA TCG TCC GCG TTC TTC ATC TTC TGC GCT GAC TTC AGA CCC CAG GTG
Ala Pro Lys Arg Pro Ser Ser Ala Phe Phe Ile Phe Cys Ala Asp Phe Arg Pro Gln Val

                 255                 270                 285                 300
AAG GGG GAG ACC CCG GGC CTG TCT ATC GGT GAT GTG GCC AAG AAG CTG GGA GAG AAG TGG
Lys Gly Glu Thr Pro Gly Leu Ser Ile Gly Asp Val Ala Lys Lys Leu Gly Glu Lys Trp

                 315                 330                 345                 360
AAC AAC CTA ACA GCG GAG GAC AAG GTA CCC TAT GAG AAG AAG GCT TCC AAG CTG AAG GAG
Asn Asn Leu Thr Ala Glu Asp Lys Val Pro Tyr Glu Lys Lys Ala Ser Lys Leu Lys Glu

                 375                 390                 405                 420
AAG TAC GAG AAG GAC ATC ACT GCG TAC CGC AAC AAG GGT AAG GTG CCG GTC AGC ATG CCA
Lys Tyr Glu Lys Asp Ile Thr Ala Tyr Arg Asn Lys Gly Lys Val Pro Val Ser Met Pro

                 435                 450                 465                 480
GCG AAG GCC GCC GCA CCC GCC AAG GAC GAC GAC GAC GAC GAT GAT GAC GAC GAC GAT GAT
Ala Lys Ala Ala Ala Pro Ala Lys Asp Asp Asp Asp Asp Asp Asp Asp Asp Asp Asp Asp

                 495                 510
GAG GAC GAC GAT GAT GAC GAC GAT GAG GAT GAC GAG TAG
Glu Asp Asp Asp Asp Asp Asp Asp Glu Asp Asp Glu Ter

3'-noncoding region
       529         539         549         559         569         579         589
ATTACCGTTG TTACGTATAG CACTTACAAT GTCTTGTTTA TAAAGCATTT AACCCCCCCC CCACCCCCTC

       599         609         619         629         639         649         659
TTGTACACAC AACACTTACT GAAAACAAAA CAAGATAAAG GAAGCCGAGG GGCAAGATGT CCTCCTATGC

       669         679         689         699         709
CTACTTTGTC CAGACCTGTC GGGAGGAGCA CAAGAAGAAA CACCCAGAAG CT
```

Figure 2
Nucleotide sequence of the cDNA insert of pBP2, the translation product of the LORF is shown below the nucleotide sequence.

shown up to base 590. The indicated reading frame shows significant homology with the N-terminal regions of bovine HMG-1 protein (3) and the trout equivalent, HMG-T (2, 7 and Fig. 3), and encodes a C-terminal polyacidic region characteristic of the group of 29-30,000 M.W. HMG proteins. Interestingly, the oligonucleotide and cDNA are mismatched at the 5' base of the oligonucleotide since the residue preceding the polyacidic region is Lys (AAG) rather than Met (ATG), as previously determined in the protein sequence of HMG-T (2, 7). However, the terminal location presumably allowed the mismatch to have only a minimally destabilising

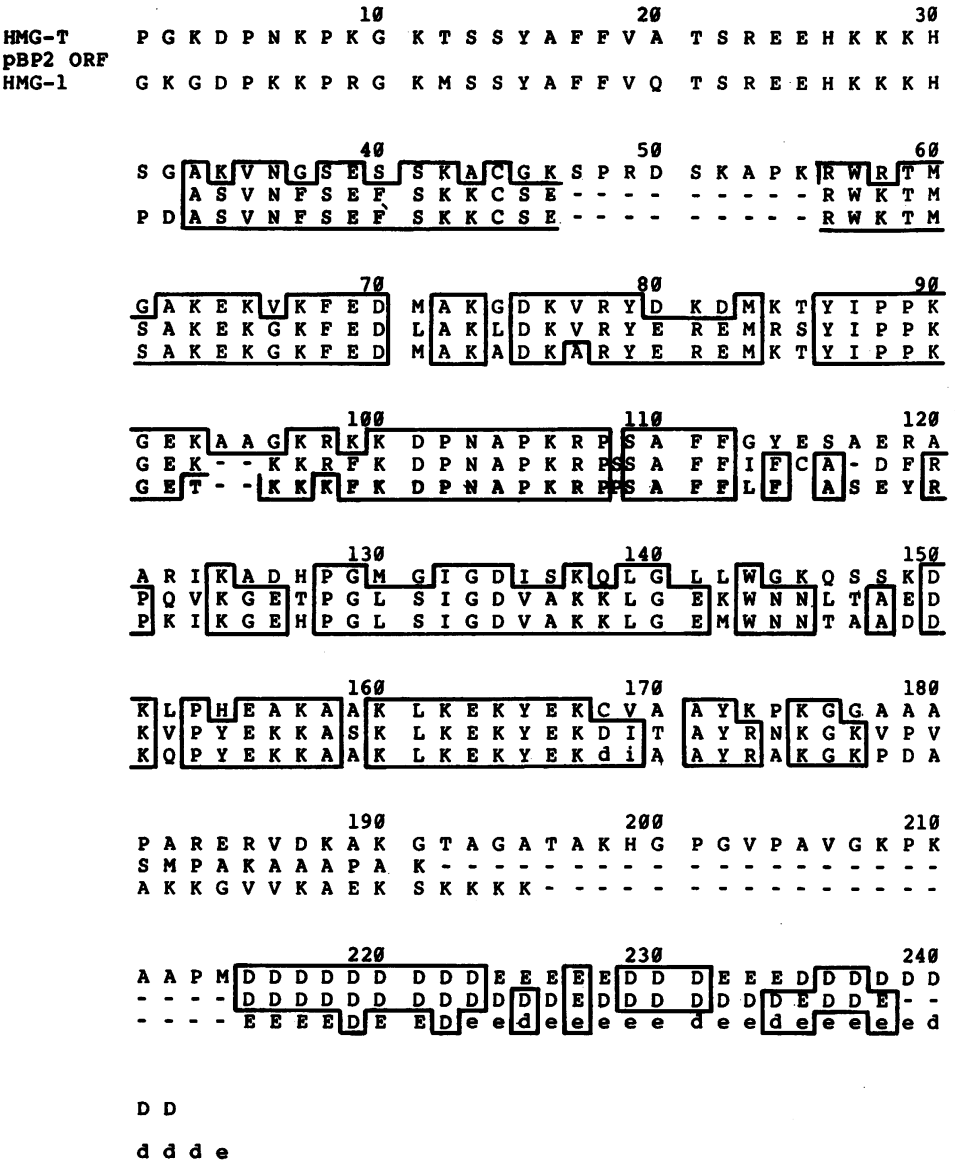

Figure 3
A comparison of the LORF of pBP2 insert with Bovine HMG-1 and trout HMG-T protein sequences.
(The HMG-1 sequence is based on published (3) protein data [upper case] supplemented and modified using information from the published (8) bovine cDNA sequence [lower case letters]. The trout protein sequence is from reordered published data (2, 7) and unpublished data (Watson and Dixon) that will be presented elsewhere.)

effect on hybridization (19). There is distinct preferential use of the GAG codon for Glu (100%) but usage of the codon GAC for Asp (67%) is not as pronounced as in the trout Histone H3 clone (9) considered when deciding upon codon usage in the oligonucleotide. Interestingly, the bovine HMG-1 cDNA clone previously sequenced (8) also showed exclusive usage of a single codon for the less frequent acidic residue, but in that case the GAT codon for Asp was used exclusively.

An ORF in pBP2 Encodes an HMG Protein

A comparison of the longest open reading frame of the pBP2 insert with bovine HMG-1 and trout HMG-T protein sequences (Fig. 3) suggests that the insert of pBP2 begins at residue 33 of the protein. Since three of the cDNA's isolated begin at this site we consider that the Hind III site at this position in the nucleic acid sequence to be an intrinsic internal site rather than being derived from the linkers used to insert the cDNA in the cloning procedure (Hind III linkers were used in the cDNA cloning procedure). We find that up to amino acid residue 177, numbering from the HMG-T protein data, homology of the insert LORF to HMG-1 (80%) is greater than to the HMG-T protein. This estimate of homology excludes amino acid analysis data which indicates that there may be additional residues in HMG-1 near position 47 (3). Many of the differences between the three sequences up to position 177 are conservative rather than radical in terms of charge, maintenance of hydrophobic areas and location of proline residues known to be important in the tertiary structure, and presumably function, of this group of proteins. We particularly note that a strongly basic region between residues 96-100 is maintained in all members of this group of proteins. This region closely resembles a peptide in the SV40 large T antigen that has recently been shown, by *in vitro* mutagenesis, to be important for the nuclear localization of that protein (20).

Beyond residue 177 homology is marginal up to the acidic C-terminal region, and no definite alignments can be made. The number of residues in the acidic domains is variable. The poorly conserved acidic domain, which has a variable number of residues, lies outside the second globular domain thought to exist in the native form of these proteins (6, 21) and, since it may exist in a random coil form, is probably under reduced conservative selection pressure compared with the folded regions. The acidic domain of the HMG protein predicted from pBP2 LORF is marginally shorter (24 residues) than that predicted for HMG-1 (1) (30 residues) or the 28 residues sequenced in trout HMG-T protein (2, 7 and D.C. Watson and G.H.

Dixon, in preparation), but does consist of an unusually long run of solely acidic residues. The composition, being mainly aspartate residues (21 Asp/3 Glu) is more similar to the of HMG-T (20 Asp/8 Glu) than that of HMG-1 (9 Asp/21 Glu). This is in contrast to the substantially greater homology in the more N-terminal regions between the LORF and bovine HMG-1 as compared with trout HMG-T (Fig. 3). However, the lower glutamate content of the LORF tail makes it clearly different, both as regards sequence and empirical amino acid composition (2, 7) data, from the sequenced trout protein, HMG-T. The polyacidic region of pBP2 is followed by a termination codon, as in the HMG-1 cDNA pBP1 (8), indicating that it lies at the C-terminus of the equivalent protein as with the protein sequence data for HMG-T (2, 7 and in preparation). There is, however, strong conservation of the exclusively polyacidic nature of this region despite variation in the ratio of Glu and Asp residues and it is likely that the selective forces involved in the evolutionary development of the polyacidic domain have not discriminated between the two residues. The significance of differences in length of the polyacidic tracts is unclear. We suggest that there is a lower limit to the length of this region which can give a stable interaction with, probably, basic regions of histones in chromatin, but that increases beyond that minimum have a decreasingly important effect. In addition, the C-terminal location of the polyacidic tract must undoubtedly affect the range of conformations that can be adopted by the HMG-proteins in their interactions with other chromatin components.

The differences between the sequenced protein, HMG-T, and the LORF of the pBP2 insert cannot be accounted for by assuming simple sequencing misreads at either the protein or nucleic acid levels. The majority of the cDNA has been determined from both strands and we have total confirmation of the sequence of the presented coding region from additional clones. Both our and other groups (22) have previously noted the existence of multiple forms of the larger HMG proteins in trout, although there has been some dispute over their relationship with HMG-1 and -2. The protein encoded in the pBP2 insert specifies 172 amino acid residues and probably begins at residue 33 (by comparison with HMG-1 and HMG-T proteins), suggesting that the entire protein contains a total of 204 amino acids. This is 38 residues shorter than the 242 residues previously determined for HMG-T (2, 7 and in preparation) and 10 residues shorter than bovine HMG-1 (3 and Fig. 3).

In Fig. 3 it is evident that two additional sequences appear to be inserted in HMG-T for which equivalent sequences do not occur either in bovine HMG-1 (3) or pBP2 ORF. The first, 9 residues in length, occurs at pos. 47-55 (HMG-T numbering) and the second, much longer, insert of 23 residues from pos. 192 to 214 is adjacent to the C-terminal polyacidic tail. While it is, of course, possible that these additional sequences in HMG-T represent protein sequencing errors, it is clear that the pBP2 ORF would, if expressed, code for a different trout HMG-T than the one previously sequenced (2, 7 and in preparation). The main evidence for this comes from a comparison of the derivation and sequence of the polyacidic tails. In the case of HMG-T, the polyacidic tail peptide was derived either by CNBr cleavage of intact HMG-T at methionine 214 (2, 7) or by tryptic cleavage at the Lys-Ala bond at pos. 210-211 (in preparation). Unequivocal automated Edman sequencing of the two easily purified acidic peptides showed that a methionine residue was indeed present at the 4th position of the tryptic peptide and the acidic region comprised 9 successive aspartate residues followed by 5 glutamate residues. This is clearly different from the pBP2 ORF cDNA sequence which predicts a Lys residue (at pos. 191 in Fig. 3) replacing the Met at pos. 214 (the difference in numbering is due to the presence of the putative additional sequence in the protein sequence determination of HMG-T). Furthermore, instead of 9 Asp residues followed by 5 Glu residues in the protein sequence, the cDNA sequence predicts 12 successive Asp residues followed by a single Glu. In fact, the empirical amino acid composition of the two peptides is totally different, HMG-T (protein determination) shows 20 Asp and 8 Glu residues while the cDNA predicts 21 Asp and only 3 Glu residues. A search is currently underway to isolate additional trout HMG cDNA clones which might correspond more closely to the HMG-T protein previously sequenced. A point worthy of note, however, is that the yield of the HMG-T acidic peptide derived by CNBr cleavage of HMG-T was very low (7), a finding that could indicate heterogeneity at pos. 214 with only a small proportion of HMG-T molecules having Met at this position. It is equally possible that some proportion of the total HMG-T pool could possess the additional sequences at pos. 47-55 and/or pos. 192-214 while another component lacked these sequences and might be more similar to the HMG-T predicted by the pBP2 insert.

3' Untranslated Region

The insert of pBP2 contains a 3' untranslated region of 192 bp

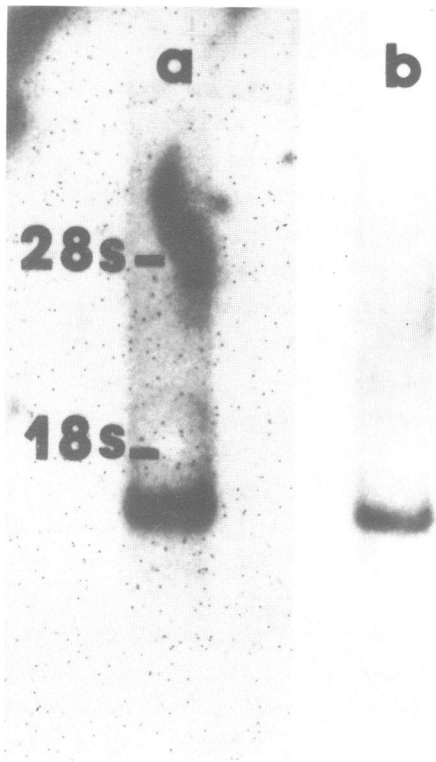

Figure 4
Hybridisation of trout HMG specific oligonucleotide (track B) and nick-translated pBP2 insert (track A) to trout testis RNA (15 µg Poly A+).

following the termination signal (TAG) of the LORF (Fig. 2). This sequence contains no poly-A tract or recognizable polyadenylation signal, despite construction of the cDNA library with poly dT primer; it is likely, therefore, that the complete untranslated region is not represented in the clones analysed. The mRNA sizes, estimated on RNA gels probed with HMG-T cDNA (see Figs. 4-6 below) indicate that the mRNA is probably ~2000 nucleotides long and that there should be untranslated regions of considerable (>1000 bases in total) size. An unusually long 3' untranslated region (~450 bp) was also seen in the bovine HMG-1 cDNA clone (8) and it is possible that the corresponding region is even longer in the HMG-T mRNA. In support of this, an HMG-1 cDNA has recently been isolated from a Chinese hamster ovary cell cDNA library which also shows an extremely long 1.5 kb 3' untranslated region (K-L.D. Lee and G.H. Dixon,

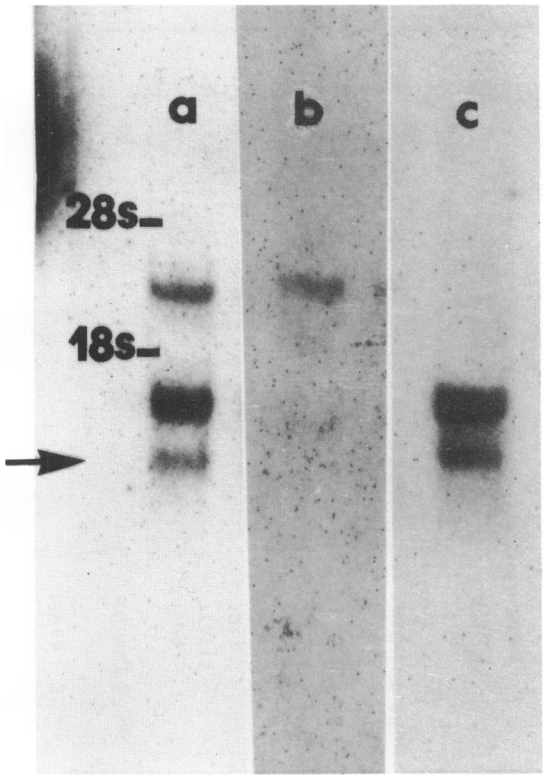

Figure 5
Analysis of trout and mammalian testis RNA's with bovine and trout probes.
Track A: bovine testis RNA (10 µg) probed with the bovine homologous pBP1 insert at high stringency.
Track B: bovine testis RNA (10 µg) probed with trout homologous pBP2 insert at low stringency.
Track C: trout testis RNA (10 µg) probed with trout homologous pBP2 insert at low stringency.

in preparation). The sequence is significantly more GC rich than the untranslated region of the HMG-1 (46% compared to 33%) (7), and no homologies between the two 3' regions can be recognized, except that both contain poly-pyrimidine tracts close to the termination codon.

Trout HMG mRNA's

The nick-translated insert of pBP2 hybridized to an RNA band of ≈2 Kb when used to probe Northern blots of trout testis RNA under standard stringency conditions (Fig. 4, track a) and a similar band was detected using the ^{32}P-labelled oligonucleotide probe (Fig. 4, track b). If the

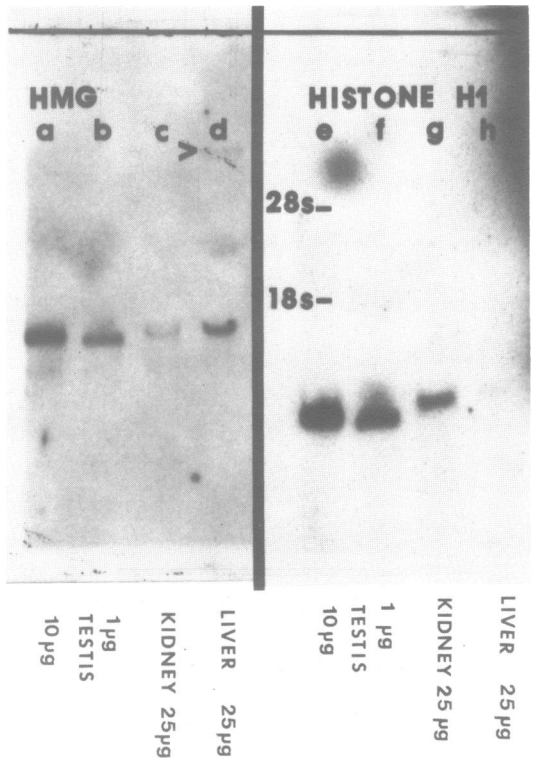

Figure 6
A comparison of the relative levels of trout HMG protein message and histone H1 message in a number of trout tissues.
A & E - 10 µg total testis RNA, B & F - 1 µg testis RNA, C & G - 25 µg total kidney RNA, D & H - 25 µg total liver RNA.
The filter (A, B, C, D) was probed with pBP2 insert, stripped and reprobed using a genomic trout histone H1 fragment (E, F, G, H).

stringencies of hybridization and washing were reduced when using the nick-translated insert a second, smaller, band of ≈1600 nts was also detected (Fig. 5, track c [arrowed]), supporting the hypothesis that there is more than one related form of mRNA, a finding consistent with the observation (22) of more than one of the larger HMG proteins in trout. We predict that the weakly hybridizing RNA species will prove to be a message coding for a protein more closely related to the HMG-T protein sequenced (2, 7 and unpublished) and that its acidic domain is likely to contain one or more GAT Asp codons to account for the failure of this second message to hybridize significantly to the oligonucleotide probe.

Under standard hybridization conditions we failed to detect cross-

hybridization of the trout probe to mammalian RNA, but at low stringency there was poor but distinct binding to an mRNA in total RNA from bovine testis (Fig. 5, track b) of similar size (~2000 nts) to the largest band visualized by the bovine clone pBP1 (Fig. 5, track a). Further work on more fully characterizing the various mammalian messages with homology to nucleic acid sequences for HMG proteins will be presented elsewhere.

Tissue Levels of Trout HMG-mRNA's

Levels of the RNA's hybridizing to the pBP2 insert have been compared for a number of tissues (Fig. 6, tracks a-d). We found, as in the case of histone mRNA's (23, and Fig. 6, tracks e-h), that HMG-mRNA levels were higher in testis than in liver or kidney (which also contains the fish haemopoetic tissue (24) by at least a factor of ten. The estimate of relative levels is rather crude as it takes into account neither variation in RNA content between tissues nor the ratio of messenger to ribosomal RNA, and we find the poly A+ mRNA content of total testis RNA to be very low.

The high HMG-mRNA level in testis suggests that the high HMG-T protein content of testis (25) must be due to active transcription of HMG-T-mRNA's, rather than simply persistence of the HMG-T protein during spermatogenesis. We consider this to be significant in light of observations of high levels of HMG-2 in rat testis (26, 27) during spermatogenesis. It is also of particular interest that Bucci et al. (27) showed that HMG-2 levels increased in meiotic cells suggesting that rapid new synthesis of one subcomponent, the "fast" form of HMG-2 is the major cause of the increased level. Our analysis also indicates that the distribution of the homologous HMG message does not precisely parallel that of histone H1 (Fig. 6) or the other core histone mRNA's (data not shown), since levels of the trout HMG-mRNA are greater in liver than kidney (Fig. 6, lanes c and d), while the reverse is observed for histone mRNA's (Fig. 6, lanes g and h). An additional high molecular weight band can be detected in liver RNA in the original autoradiogram corresponding to Fig. 6, track d (arrowed). This does not appear to be due to DNA contamination as it is not visualised by a probe to histone H1 (Fig. 6, track h), particularly since histone genes have a high copy number in the trout genome (9).

Analysis of Genomic DNA

Analysis of genomic DNA was carried under conditions more stringent than those employed for RNA analysis. It is very unlikely that material homologous to the smaller (1.5 kb) mRNA species, detectable only under

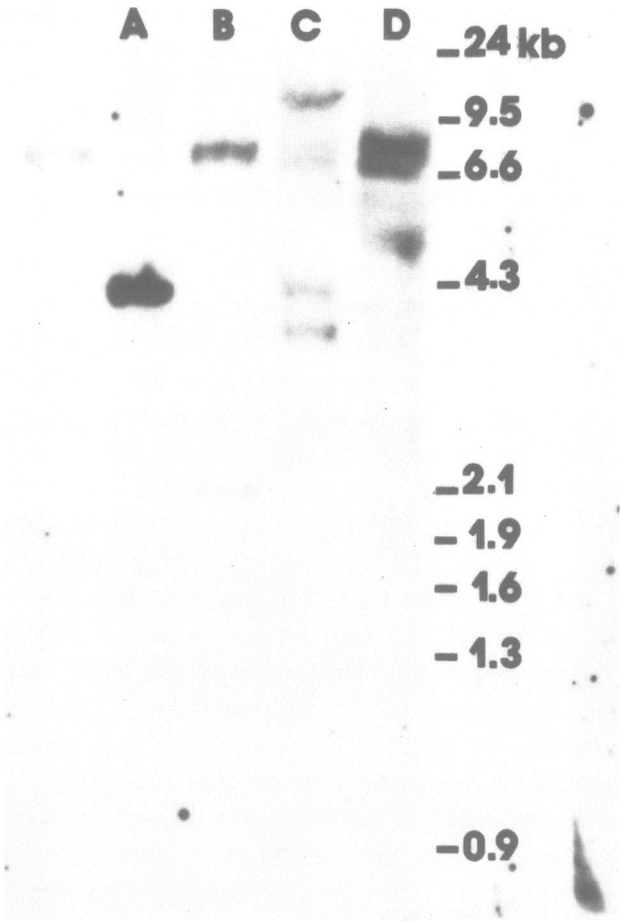

Figure 7
Analysis of trout genomic DNA with pBP2 insert.
A - Bam H1, B - EcoR I, C - Bgl II, D - Hind III restriction enzyme digests.

very relaxed conditions, would hybridize significantly under these conditions. Hybridizing fragments can therefore be considered to be homologous to, or closely related to, the pBP2 insert and the 2 kb mRNA. Results shown in Figure 7 indicate that several different hybridizing fragments are produced by EcoR1, Bgl II and Hind III. This observation is consistent with there being several copies of the gene corresponding to the pBP2 insert. There are also likely to be additional sequences which would encode the weakly homologous 1.5 kb message and are probably not

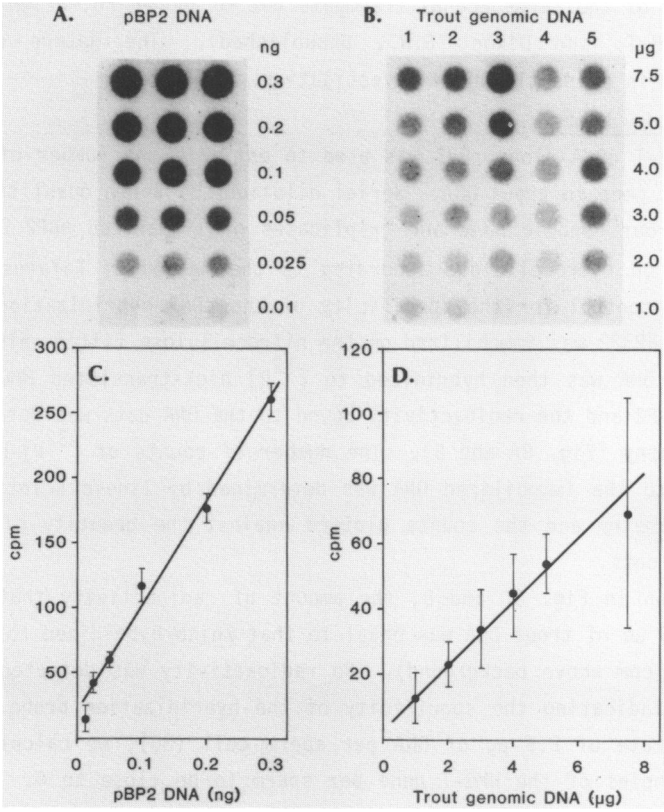

Figure 8. A and B
Autoradiogram of 0.01-0.3 ng of pBP2 DNA (A) and 1.0-7.5 µg of five individual trout DNA's (B) dotted on nitrocellulose and hybridized to pBP2 insert.

C and D
Graph showing the amount of radioactivity that hybridized to pBP2 DNA dots (C) or trout DNA dots (D) as a function of the quantity of DNA immobilized.

detected here. If Southern blots are performed at lower stringency, a number of additional bands hybridize (data not shown) consistent with this idea. Thus, the larger HMG's in trout may represent a part of a multi-gene family, a situation similar to that situation found for the genes of other trout nuclear proteins (9, 28). There is variation in the size of fragments produced by Eco R1, Bgl II, and Hind III enzymes with genomic DNA from different Rainbow trout (data not shown), and we have recently isolated, from a trout genomic library, clones that contain one

or the other of the large Eco R1 fragments (~2 kb and ~6 kb; Wright, J.M., Pentecost, B.T. and Dixon, G.H., unpublished). The nature of this polymorphism is presently under investigation.

HMG-T Gene Copy Number

The HMG-T cDNA clone pBP2 was used to estimate the number of copies of the HMG-T gene in trout DNA. Serial dilutions of known quantities of 5 different trout genomic DNAs and triplicates of linearized pBP2 DNA were immobilized on nitrocellulose according to the method of Kafatos et al. (29). As a control for the specificity of the cDNA hybridization probe, 0.25 ng of pBR322 was immobilized on the nitrocellulose filter, also. The immobilized DNA was then hybridized to [^{32}P] nick-translated HMG-T cDNA insert of pBP2 and the radioactivity bound to the DNA dots was detected by autoradiography (Fig. 8A and B). The number of counts of [^{32}P]-DNA that hybridized to the immobilized DNA was determined by liquid scintillation spectrophotometry and the counts plotted against the quantity of DNA in each of the dots.

As shown in Fig. 8C and D, the amount of radioactivity that hybridized to 4.9 μg of trout DNA was equal to that which hybridized to 0.04 ng of pBP2 (50 cpm above background). No radioactivity was detected in the pBR322 dot indicating the specificity of the hybridization probe. Based on the estimate of 2.5 pg of DNA per sperm cell (30), we calculate the number of copies of the HMG-T gene per sperm to be close to 4. The low copy number for the HMG-T gene is in agreement with the limited number of fragments seen to hybridize to the HMG-T probe in Southern blots of trout DNA (Fig. 7).

It may be concluded, therefore, that at least two mRNA's coding for related trout high mobility group proteins of the 29-30 K-class are transcribed in trout tissues (Fig. 5c). This finding is consistent with the clear differences between the coding region predicted by the sequenced cDNA clone, pBP2, and the amino acid sequence determined previously for the major isolated 29 K trout HMG-T particularly in the C-terminal polyacidic region (2, 7). Additional evidence suggesting that the class of 29 K HMG- proteins is coded by a small family of genes comes from the determination of the copy number as being close to 4 per sperm DNA equivalent. The observation of multiple bands on Southern blots of total trout DNA with three different restriction enzymes (EcoR1, Bgl II and Hind III) is also consistent with this estimate. The situation in trout appears, therefore, to resemble that for the only other well-characterized group of

29 K HMG-proteins, those of bovine thymus, where two proteins, HMG-1 and HMG-2 have been well-characterized at the protein level (1, 3 and 4) and are clearly closely related family members.

ACKNOWLEDGEMENTS

We thank J. Mezquita for histone H1 probes, D. McKenzie for trout genomic DNA and M. Cudmore for final preparation of the manuscript. Sequence analysis programmes were constructed by D. McKay. This work was generously supported by grants from the MRC of Canada to G.H. Dixon.

[1]Present address: NIEHS, M.D. 1301, P.O. Box 12233, Research Triangle Park, NC 27709, USA
[2]Alberta Heritage Foundation for Medical Research Post-Doctoral Fellow
*To whom correspondence should be addressed

REFERENCES

(1) Goodwin, G.H. and Johns, E.W. (1978). Biochim. Biophys. Acta 519:279-284.
(2) Dixon, G.H. (1982). In: The HMG Chromosomal Proteins, pp. 149-182. London and New York: Academic Press.
(3) Walker, J.M. (1982). In: The HMG Chromosomal Proteins, pp. 69-88. London and New York: Academic Press.
(4) Mayes, E.L.V. (1982). In: The HMG Chromosomal Proteins, pp. 9-40. London and New York: Academic Press.
(5) Watson, D.C., Peters, E.H. and Dixon, G.H. (1977). Eur. J. Biochem. 74:53-60.
(6) Reeck, E.R., Isakson, P.J. and Teller, D.C. (1982). Nature 300:76-78.
(7) Watson, D.C. and Dixon, G.H. (1981). Bioscience Reports 1:167-175.
(8) Pentecost, B. and Dixon, G.H. Bioscience Reports 4:49-57.
(9) Connor, W., Mezquita, J., Winkfein, R.J., States, J.C. and Dixon, G.H. (1984). J. Mol. Evol. 20, 227-235.
(10) Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979). Biochemistry 18:5294-5304.
(11) Bantle, J.A., Maxwell, I.H. and Hahn, W.E. (1976). Anal. Biochem. 72:413-427.
(12) Wickens, M.P., Buell, G.N. and Schimke, R.T. (1978). J. Biol. Chem. 253:2483-2495.
(13) Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D. and Itakusk, K. (1977). Science 196:177-180.
(14) Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). In: Molecular Cloning, A Laboratory. Cold Spring Harbour Lab. Manual, pp. 230-246.
(15) Hanahan, D. (1983). J. Mol. Biol. 166:557-580.
(16) Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg. P. (1977). J. Mol. Biol. 113:237-251.
(17) Bailey, J.M. and Davidson, N. (1976). Anal. Biochem. 70:75-80.
(18) Maxam, A.M. and Gilbert, W. (1980). In: Methods in Enzymology, Vol. 65, pp. 499-559. New York: Academic Press.
(19) Szostak, J.W., Stiles, J.I., Tye, B-K., Chia, P., Sherman, F. and Wu, R.(1979). In: Methods in Enzymology, Vol. 68, pp. 419-428.

(20) Kalderon, D., Richardson, W.D., Markham, A.F. and Smith, A.E. (1984). Nature 311:33-38.
(21) Cary, P.D., Turner, C.H., Leung, I., Mayes, E. and Crane-Robinson, C. (1984). Eur. J. Biochem., in press.
(22) Brown, E., Goodwin, G.H., Mayes, E.L.V., Hastings, J.R.B. and Johns, E.W. (1980). Biochem. J. 191:661-664.
(23) Connor, W., States, J.C., Mezquita, J. and Dixon, G.H. (1984). J. Mol. Evol. 20, 236-250.
(24) Catton, W.T. (1951). Blood 6:39-60.
(25) Christensen, M.E. and Dixon, G.H. (1981). J. Biol. Chem. 256:7549-7556.
(26) Seyedin, S.M. and Kistler, W.S. (1979). J. Biol. Chem. 254:11265-11271.
(27) Bucci, L.R., Brock, W.A., Goldknopf, I.L. and Meistrich, M.L. (1984). J. Biol. Chem. 259:8840-8846.
(28) Aiken, J.M., McKenzie, D., Zhao, H-Z., States, J.C. and Dixon, G.H. (1983). Nucl. Acids Res. 11:4907-4922.
(29) Kafatos, F., Jones, C. and Efstratiadis, A. (1979). Nuc. Acids Res. 7:1541-1552.
(30) Louie, A.J. and Dixon, G.H. (1972). J. Biol. Chem. 247:5490-5497.