Volume 17 Number 15 1989 Nucleic Acids Research

Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are capable of binding to the octamer sequence motif

Richard Eckner and Max L.Birnstiel*

Institute of Molecular Pathology, Dr Bohrgasse 7, A-1030 Vienna, Austria

Received May 26, 1989; Revised and Accepted July 6, 1989 EMBL accession nos X14957, X14958

ABSTRACT

In human B lymphocytes and placenta HMG I and its smaller isoform HMG Y are encoded by two distinct but structurally highly similar mRNAs which arise most likely by alternative splicing of a single primary transcript. Both have been cloned as cDNAs. On Northern blots an abundant mRNA species 2000 nucleotides in length was detected in all cell lines examined. Exclusively in erythroid cells an additional rare 3800 nucleotides long mRNA species was noted. In quiescent cells the mRNA levels of HMG I/Y were not significantly down-regulated. Southern blot analysis indicated that at least four genes are present per haploid human genome. Both proteins when expressed in bacteria bind specifically to A-T rich stretches of DNA suggesting that no posttranslational modifications are necessary for specific DNA binding. Interestingly, HMG I as well as HMG Y are capable of binding to the octamer transcriptional regulatory sequence motif.

INTRODUCTION

High mobility group proteins are abundant and small non-histone chromosomal proteins of higher eukaryotes. They are classified into high molecular mass HMG proteins (HMG 1 and HMG 2, molecular weight around 25kD) and low molecular mass HMG proteins (HMG 14, HMG 17, HMG I and HMG Y; molecular weight around 10kD) (1, 2). Characteristically both classes of HMGs are highly charged, the NH_2 terminus and the central part being basic and the COOH terminus being highly acidic. It is thought that the basic domain is responsible for DNA binding whereas for the acidic part no function has been established that is valid for all HMG proteins. In the case of HMG 14 and 17, however, it was shown that the acidic tail contacts the basic histone proteins in the nucleosome core particle primarily through electrostatic interactions (3, 4, 5).

Not much is known about the precise biological role of HMG proteins, but it is thought that they modify the chromatin structure of active genes in a way that facilitates access by RNA polymerases (6). This idea is consistent with the fact that the amount of HMGs present in a cell (for example per 10 nucleosome particles only one molecule of HMG 14 or 17 exists) is sufficient to bind only to a fraction of the chromatin. From results obtained by the functional dissection of transcriptional activator proteins it is also conceivable that the acidic tail of HMG proteins may be directly involved in transcriptional activation (7). Ptashne and colleagues have demonstrated that in principle a protein fulfilling two requirements can work as a transcriptional activator. First it must be attached to DNA (by a DNA binding domain) and second it has to expose a negatively charged surface (8 and references therein). As both HMG 1, HMG 2 and HMG I/Y bind to DNA (9, 10) and possess a negatively charged COOH terminus, they have the structure of a Ptashne type activator protein.

HMG I and its isoform HMG Y are the most recently discovered low molecular weight HMG proteins (2). HMG I (in monkey also called α-protein) binds specifically to A-T rich runs of double stranded DNA (9) and it was suggested that it is responsible for the observed nucleosome phasing (11). Furthermore HMG I/Y proteins and mRNAs are present at high levels in undifferentiated, rapidly dividing cells and are virtually absent in terminally differentiated non-dividing cells (12, 13, 14). This observation was interpreted to mean that HMG I/Y proteins might be required either for chromosome segregations during rapid cell divisions, or alternatively for maintaining the undifferentiated state of chromatin, or both.

The human HMG I protein has been partially sequenced (15) and recently a mouse cDNA most probably coding for HMG Y protein has been isolated (14).

To allow the examination of the properties of HMG I/Y in more detail, we have isolated human cDNA clones encoding these proteins. In this paper we describe the structure of these clones together with their expression in E. coli, and we also investigate whether HMG I/Y mRNA levels respond to the proliferation state of undifferentiated cells.

MATERIALS AND METHODS

Expression screening

As a screening probe a double stranded 22mer oligonucleotide comprising a human immunoglobulin light chain promoter octamer was used. It had the following sequence:

5′ CTAGATGAATATGCAAATAACT 3′
3′ TACTTATACGTTTATTGAGATC 5′,

the fos SRE (serum response element) oligonucleotide used as a control had the sequence:

5′ TCGAGGATGTCCATATTAGGACATCT 3′
3′ CCTACAGGTATAATCCTGTAGAAGCT 5′.

Both oligonucleotides were multimerized and subsequently labelled by nick translation to a specific activity of $1-2\times10^8$ cpm per μg of DNA. Approximately 10^6 plaques from a B lymphocyte (RPMI 4265) λgt11 cDNA library (bought from Clontech, Palo Alto, CA) were screened. Screening was performed as described (16) except that $MgCl_2$ was replaced by 5mM EDTA in the binding buffer. The tetrameric probes for the octamer sites of the murine IgH enhancer (described in reference 17, a kind gift of E. Schreiber, University of Zürich) and of the histone H2B promoter (described in reference 18) were end-labelled with T4-polynucleotide kinase.

Screening by hybridisation, Southern and Northern blotting

For phage screening by hybridisation plaques were transfered to Gene Screen Plus membranes; for capillary blotting of Southern and Northern gels Gene Screen membranes (New England Nuclear) were used. The conditions for running and transfering the gels were as described (19). Twice gel purified fragments were labelled with random priming oligonucleotides (20). Hybridisations were carried out at 65°C for 18h using Church and Gilbert SDS hybridisation buffer (21). Washes were done for 1h at 65°C in two changes of washing buffer (21).

Sequencing

For sequencing, the cDNA inserts of clones λB41 and λB48 were resected from each side with Bal 31 to produce a nested set of deletions. Sequencing reactions were performed with T7 Sequenase according to the recommendations of the manufacturer (USB). Both strands were entirely sequenced, and regions of ambiguity were resequenced with adjacent priming oligonucleotides. For the remaining cDNA clones the sequence of the first 300

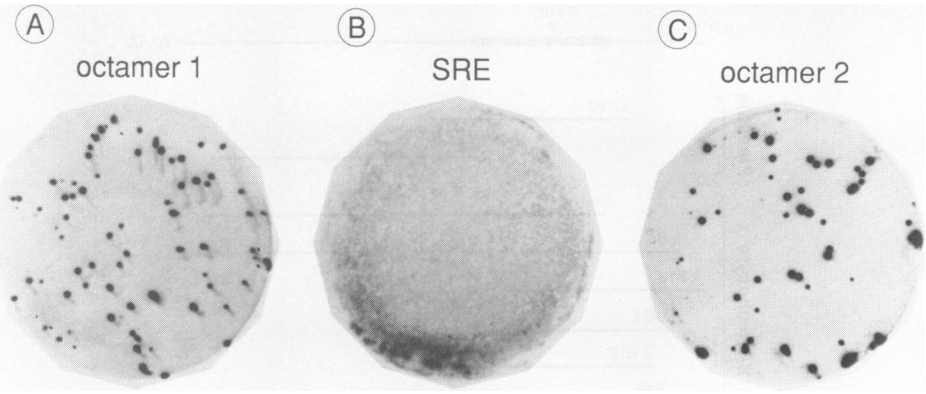

Figure 1. *Sequence specific DNA binding of bacterially expressed HMG I and HMG Y protein.* Phages encoding HMG Y (panels A and B, clone λB48) or HMG I (panel C, clone λB41) were plated out and transfered to nitrocellulose filters.
A) plaque lift probed with the multimerized octamer oligonucleotide derived from the Ig light chain promoter.
B) plaque lift probed with a multimerized SRE oligonucleotide (an overexposure is shown).
C) filter exposed to end-labelled tetrameric octamer site from a histone H2B promoter.

nucleotides at the 5′ and 3′ end was determined. For sequence comparisons the programs of the Genetics Computer Group (GCG, reference 22) were used.

Cell culture and RNA extraction

Exponentially growing NIH 3T3 cells were cultured in DMEM plus 10% fetal calf serum (FCS), Hela cells in DMEM plus 5% FCS. For serum starvation experiments half confluent cultures were incubated for 3 days in DMEM supplemented with only 1% serum. Exponentially growing NIH 3T3 cells were arrested in G1 by exposing them for three days to 0.4mM indomethacin (23) in complete medium. Cytoplasmic RNA was extracted according to Favoloro et al. (24).

RESULTS

Isolation and characterisation of HMG I/Y cDNAs

Complementary DNAs encoding human HMG I/Y were isolated due to the ability of these proteins to bind A-T rich stretches of duplex DNA. A human B lymphocyte λgt11 cDNA library was screened with a multimerized octamer oligonucleotide probe (see Methods for sequence) having two A-T runs per monomer. This probe was labelled to high specific activity by nick translation and used for an expression screening according to Vinson et al. (16). From the initial screening 12 clones were repeatedly positive and were reprobed with a multimerized oligonucleotide comprising an SRE site (see Methods for sequence) in order to distinguish between general DNA binding proteins and specific ones. Out of 12 clones, 10 reacted also with the SRE probe and were not further investigated. The remaining two (λB41 and λB48) did not bind to the SRE probe and therefore encoded sequence specific DNA binding proteins (fig. 1). A subsequent analysis of the binding specificities by methylation interference revealed that the β-galactosidase fusion proteins of λB41 and λB48 make contacts to DNA which are centered over the octamer motif. They recognized mainly the minor groove of DNA but made more extended contacts in

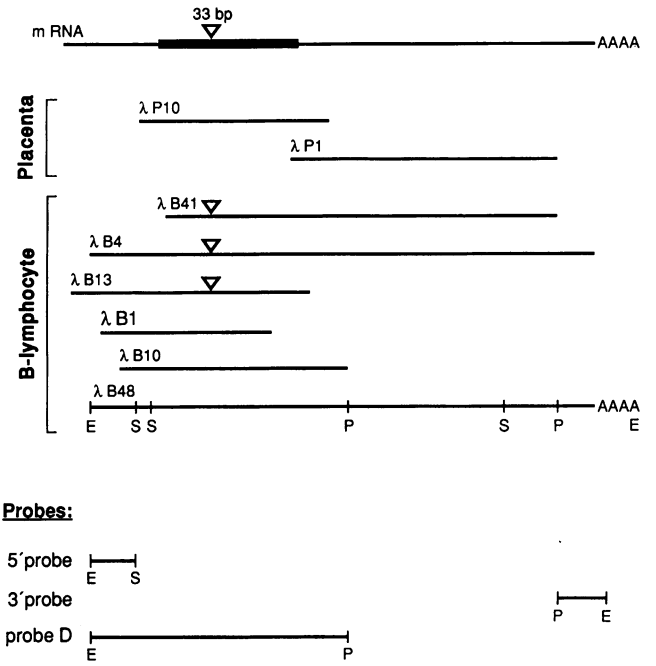

Figure 2. *Restriction map and schematic representation of some of the isolated cDNA clones.* The top line shows the structure of the full length HMG I/Y mRNA as deduced from the isolated cDNA clones. The thick bar represents the coding body, the triangle indicates the location of the additional 33 bp fragment of HMG I mRNAs. The placental clones λP1 and λP10 were recovered by screening with the insert of clone λB48. The positions of EcoRI (E), SacI (S) and PstI (P) sites are indicated for the clone λB48 from which all hybridisation probes were derived. The EcoRI sites were added during the cloning process to the cDNA insert. DNA fragments used as hybridisation probes are listed below. The figure is not drawn to scale.

it than the described octamer binding proteins OTF 1 and OTF 2 (25, 26; results not shown). We concluded therefore that λB41 and λB48 do not encode OTF 1 or OTF 2.

Phage λB41 had a 1.5 kb insert, while λB48 harboured a 1.85 kb insert. Restriction mapping and sequencing of the two inserts showed that they were identical in an 1.5 kb overlapping region, except for a 33bp fragment in clone λB41 that was absent in λB48 (fig. 2). A sequence search with GenBank found a sequence having 91.5% identity in the coding region on the nucleic acid level and 96% identity on the protein level with λB48: the recently cloned cDNA for murine HMG I(Y) (14). Furthermore the deduced amino acid sequence from the open reading frame of λB41 showed 99% identity with a published partial amino acid sequence of human HMG I protein (15). We therefore concluded that λB41 codes for human HMG I. Since both the published murine sequence and clone λB48 are lacking the same 33bp in their coding regions, we strongly suspect that they represent cDNAs for HMG Y, the smaller isoform of HMG I (see Discussion).

The finding that HMG I and HMG Y both can bind to the octamer motif of the immunoglobulin light chain promotor prompted us to investigate whether the two proteins can also bind to octamer motifs from other genes having different flanking sequences. Both proteins are capable of binding to the tetrameric octamer motifs of the histone H2B

```
-231  GGCCGGTGCTGCGCTCCTCTAATTGGGACTCCGAGCCGGGGCTATTTCTGGCGCTGGCCGGGCTCCAAGAAGGCA

-156  TCCGCATTTGCTACCAGCGGCGGCCGCGGCGGAGCCAGGCCGGTCCTCAGCGCCCAGCACCGCCGCTCCCGGCAA

 -81  CCCGGAGCGCGCACCGCAGCCGGCGGCCGAGCTCGCGCATCCCAGCCATCACTCTTCCACCTGCTCCTTAGAGAA

  -6  GGGAAGATGAGTGAGTCGAGCTCGAAGTCCAGCCAGCCCTTGGCCTCCAAGCAGGAAAAGGACGGCACTGAGAAG
          M  S  E  S  S  S  K  S  S  Q  P  L  A  S  K  Q  E  K  D  G  T  E  K

additional 33bp       GTGAGTCCCGGGACAGCGCTGGTAGGGAGTCAG
of HMG I              V  S  P  G  T  A  L  V  G  S  Q

 69  CGGGGCCGGGGCAGGCCGCGCAAGCAGCCTCCGAAGGAGCCCAGCGAAGTGCCAACACCTAAGAGACCTCGGGGC
     R  G  R  G  R  P  R  K  Q  P  P  K  E  P  S  E  V  P  T  P  K  R  P  R  G

144  CGACCAAAGGGAAGCAAAAACAAGGGTGCTGCCAAGACCCGGAAAACCACCACAACTCCAGGAAGGAAACCAAGG
     R  P  K  G  S  K  N  K  G  A  A  K  T  R  K  T  T  T  T  P  G  R  K  P  R
           A
219  GGCAGACCCAAAAAACTGGAGAAGGAGGAAGAGGAGGGCATCTCGCAGGAGTCCTCGGAGGAGGAGCAGTGACCC
     G  R  P  K  K  L  E  K  E  E  E  E  G  I  S  Q  E  S  S  E  E  E  Q  *

294  ATGCGTGCCGCCTGCTCCTCACTGGAGGAGCAGCTTCCTTCTGGGACTGGACAGCTTTGCTCCGCTCCCACCGCC

369  CCCGCCCCTTCCCCAGGCCCACCATCACCACCGCCTCTGGCCGCCACCCCCATCTTCCACCTGTGCCCTCACCAC

444  CACACTACACAGCACACCAGCCGCTGCAGGGCTCCCATGGGCTGAGTGGGGAGCAGTTTTCCCCTGGCCTCAGTT

519  CCCAGCTCCCCCCGCCCACCCACGCATACACACATGCCCTCCTGGACAAGGCTAACATCCCACTTAGCCGCACCC

594  TGCACCTGCTGCGTCCCCACTCCCTTGGTGGTGGGGACATTGCTCTCTGGGCTTTTGGTTTGGGGGCGCCCTCTC

669  TGCCTCCTTCACTGTTCCCTCTGGCTTCCCATAGTGGGGCCTGGGAGGGTTCCCCCTGGCCTTAAAAGGGGCCCA

744  AGCCCATCTCATCCTGGCACGCCCTACTCCACTGCCCTGGCAGCAGCAGGTGTGGCCAATGGAGGGGGGTGCTGG

819  CCCCCAGGATTCCCCCAGCCAAACTGTCTTTGTCACCACGTGGGGCTCACTTTTCATCCTTCCCCAACTTCCCTA

894  GTCCCCGTACTAGGTTGGACAGCCCCCTTCGGCTACAGGAAGGCAGGAGGGGTGAGTCCCCTACTCCCTCTTCAC

969  TGTGGCCACAGCCCCCTTGCCCTCCGCCTGGGATCTGAGTACATATTGTGGTGATGGAGATGCAGTCACTTATTG

1044 TCCAGGTGAGGCCCAAGAGCCCTGTGGCCGCCACCTGAGGTGGGCTGGGGCTGCTCCCCTAACCCTACTTTGCTT

1119 CCGCCACTCAGCCATTTCCCCCTCCTCAGATGGGGCACCAATAACAAGGAGCTCACCCTGCCCGCTCCCAACCCC

1194 CCTCCTGCTCCTCCCTGCCCCCCAAGGTTCTGGTTCCATTTTTCCTCTGTTCACAAACTACCTCTGGACAGTTGT
                    G  G
1269 GTTGTTTTTTGTTCAATGTTCCATTCTTCGACATCCGTCATTGCTGCTGCTACCAGCGCCAAATGTTCATCCTCA

1344 TTGCCTCCTGTTCTGCCCACGATCCCCTCCCCCAAGATACTCTTTGTGGGGAAGAGGGGCTGGGGCATGGCAGGC

1419 TGGGTGACCGACTACCCCAGTCCCAGGGAAGGTGCCCTGCCCCTAGGATGCTGCAGCAGAGTGAGCAAGGGGGCC

1494 CGAATCGACCATAAAGGGTGTAGGGGCCACCTCCTCCCCCTGTTCTGTTGGGGAGGGGTAGCCATGATTTGTCCC

1569 AGCCTGGGGCTCCCTCTCTGGTTTCCTATTTACAGTTACTTGAATAAAAAAAATATCCTTTTCTGGAApoly(A)
```

Figure 3. *Nucleotide and deduced amino acid sequence of HMG I/Y cDNAs.* The amino acid sequence is written in the single letter code. The numbering refers to the first nucleotide of the ATG initiation codon as number 1. The 5′ non-translated region is displayed with negative numbers. The position of the additional 33bp fragment of HMG I cDNA is indicated, but is not included in the numbering. At locations with polymorphisms between cDNAs derived from B cell and placental tissue the alternative nucleotide of the placental cDNAs is written above the sequence. The polyadenylation signal is underlined.

promoter (fig. 1c) and of the immunoglobulin heavy chain enhancer (data not shown). Furthermore binding was also observed with monomeric octamer probes in gel mobility shift and phage-plaque binding assays albeit at lower efficiency than with multimeric probes

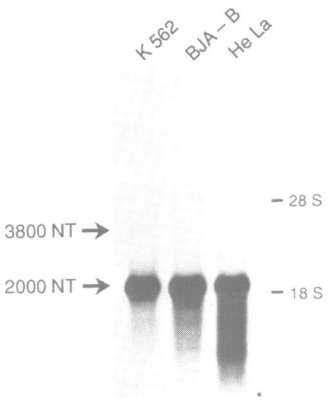

Figure 4. *Northern blot analysis.* Equal amounts of cytoplasmic RNA from K562 (human erythroblast cell line), BJA-B (human B lymphoblast cell line) and HeLa (human cervical carcinoma cell line) cells were fractionated on a 1% formaldehyde gel. The full length insert of clone λB48 was used as probe. Positions of the ribosomal RNAs are indicated on the right, sizes of hybridizing bands on the left. The abundant 2kb species of HMG I/Y mRNA migrates slightly above 18S ribosomal RNA.

(results not shown). Taken together these results suggest that HMG I/Y can in principle bind to all octamer elements independent of the flanking sequences.

Northern blot analysis indicated that a 2000 nucleotides RNA represents the smallest

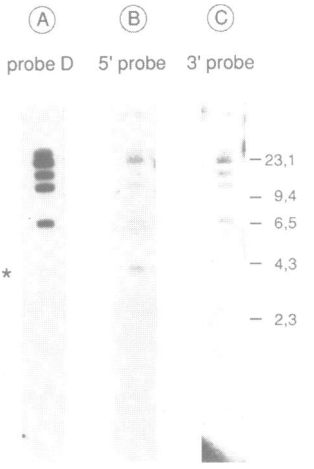

Figure 5. *Southern blot analysis.* Genomic DNA from HeLa cells was digested with EcoRI, separated and subsequently blotted by alkaline transfer. Size markers are displayed on the right in kb. A) Probe D spanning the coding body of HMG Y was used for hybridisation. Washing conditions were more stringent (68°C) than in panel B and C (65°C); under these conditions the band around 4.3 kb (marked with a star) is hardly visible. B) The 5′ 130bp EcoRI-SacI fragment was used as hybridisation probe. C) Hybridisation with the 3′ specific probe (170bp PstI-EcoRI fragment).

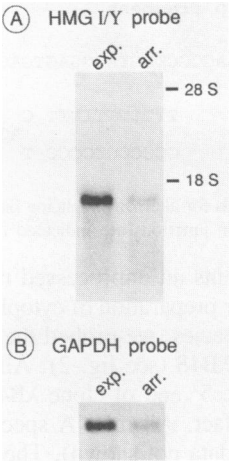

Figure 6. *Northern blot analysis of RNA derived from exponentially growing and quiscient NIH 3T3 cells.* Equal amounts of RNA were resolved on a 1% formaldehyde gel and blotted. The filter was first hybridized with the total insert of λB48 and, after stripping off the signal, reprobed with a GAPDH probe. EXP. and ARR. indicate the lanes of RNA from exponentially growing and serum arrested cells respectively. The positions of the ribosomal RNAs are marked on the right side, the murine HMG Y mRNA has a length of about 1.7kb (14).

full length HMG I/Y RNAs (see below and fig. 4). Since this is longer than either of our original clones, we rescreened by hybridisation the B lymphocyte λgt11 and a placental library with a probe spanning the first 130bp of the 5′ end of clone λB48 (EcoRI−SacI fragment; see fig. 2). Four additional clones were recovered from the B cell library, and one from the placental library. Two of the clones isolated from the B cell library had the additional 33bp fragment, the other two were without (fig. 2). Clone λB13 had a 5′ endpoint 47bp longer than clone λB48. The overall structure and sequence of HMG I and HMG Y appears to be identical, except the additional 33bp fragment only present in HMG I cDNAs. We noted, however, several sequence polymorphisms between cDNAs recovered from the B cell library and those isolated from the placenta library, though these could not be attributed specifically to HMG I or HMG Y cDNAs. The assembled complete sequence of HMG I/Y cDNA and the deduced amino acid sequence is shown in figure 3. The human HMG I protein is encoded by 107 amino acids (11,674 Dalton molecular weight), and human HMG Y protein consists of 96 amino acid residues (10,676 Dalton).

In order to get an idea of the abundance of HMG I/Y mRNAs in B lymphocyte and placental tissue we probed 10^5 plaques of each library with the full length insert of λB48. In both libraries HMG I/Y message is present at a frequency of 1 in 8000 plaques. Assuming a proportional representation of cellular mRNAs in our cDNA libraries, HMG I/Y mRNA represents about 0.0125% of the total mRNA population of a cell.

Expression of HMG I/Y RNA

To investigate the expression pattern of HMG I/Y RNAs in various tissues, we fractionated cytoplasmic RNA on a denaturing gel and probed transfers using three different probes derived from different parts of λB48 insert. The result obtained with the whole insert as hybridisation probe is shown in figure 4. All tissues examined contain an abundant mRNA species 2000 nucleotides in size. Erythroid cells (K562) show in addition a rare species migrating with an estimated length of 3800 nucleotides. To determine whether the minor

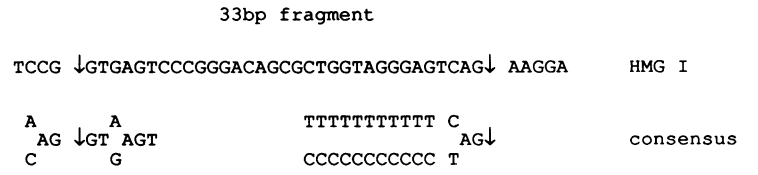

Figure 7. Comparison of the putative signals for alternative splicing flanking the 33bp fragment with the consensus boundary sequences (32). The exon/intron junctions are indicated by arrows.

erythroid specific species represents an unprocessed mRNA precursor (which might be leaked out of the nuclei during our preparation of cytoplasmic RNA) or a mature transcript distinct from the abundant 2kb species, we probed the same blot with fragments derived from the 5′ and 3′ end of clone λB48 (see fig. 2). All mRNA species hybridised to the 5′ end probe, confirming that the 5′ end of clone λB48 is really represented in cellular mRNA and is not a cloning artefact. All mRNA species, except the erythroid specific, gave a signal with the 3′ probe (data not shown). Therefore the erythroid 3.8kb species must contain another 3′ end and cannot simply be a precursor of the abundant 2kb species in erythroid cells. The weak band around 4.8 kb that is present in each lane exactly comigrated on all gels with the 28S ribosomal RNA. It also hybridised with the 5′ end probe but not with the one derived from the 3′ end. We think that this band is an artefact signal generated by some cross-hybridisation of the G/C rich 5′ end of HMG I/Y mRNA with the abundant 28S ribosomal RNA. Such a phenomenon was already observed with other G/C rich probes (27).

Genomic analysis

To examine the copy number of HMG I/Y genes in the human genome, we probed a Southern blot of EcoRI-digested genomic DNA with probe D (see figure 2) at high stringency. On a short exposure 4 EcoRI fragments are visible, on a longer exposure an additional weakly hybridising fragment appears (fig. 5a). In order to distinguish between the possibilities that either one huge gene with many exons/introns or a gene family codes for human HMG I/Y, we reprobed the blot with the same short 5′ and 3′ probes used already for Northern analysis. The 5′ fragment detects three bands (5b) whereas the 3′ fragment recognizes four bands (5c), two EcoRI fragments in the human genome hybridise to both probes and must therefore contain the complete information of HMG I/Y cDNA. Because it is highly unlikely that the 170bp probe derived from the 3′ end is encoded by 4 large genomic EcoRI fragments, we conclude that there are most likely four genes (perhaps including pseudogenes) comprising the human HMG I/Y gene family.

Terminally differentiation or growth arrest as cause for HMG I/Y mRNA down regulation?

The observation of Johnson et al. (14) that in terminally differentiated cells no HMG I/Y mRNA is detectable on Northern blots, raised the question of whether this down-regulation is caused by the differentiation process itself or by the fact that these cells had stopped growing. If the latter hypothesis were true, one should be able to obtain a strong down-regulation of HMG I/Y message by arresting the growth of, for example, undifferentiated fibroblasts. With such an experimental approach one is able to separate the contribution of proliferation arrest from that of differentiation on HMG I/Y mRNA levels.

To test the above hypothesis, we arrested NIH 3T3 fibroblasts in G_0 phase by serum starvation for three days. Approximately equal amounts of cytoplasmic RNA from exponentially growing cells and from serum-starved cells were separated, blotted and probed

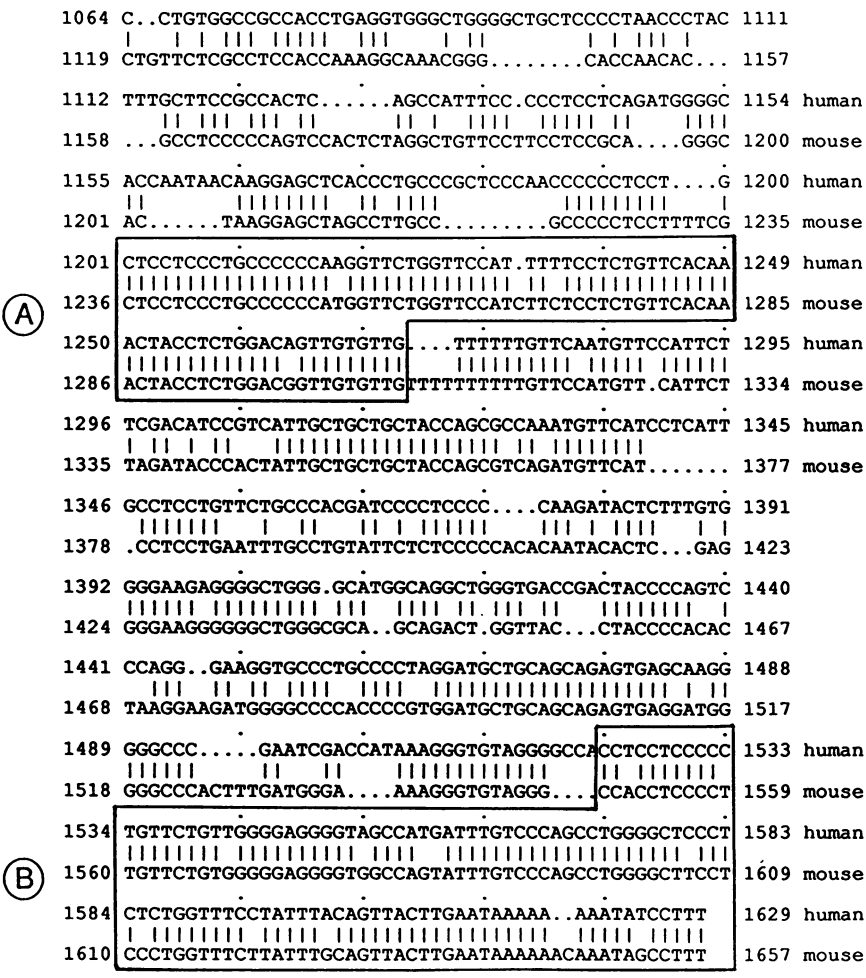

Figure 8. *Two highly conserved sequences at the 3' end of HMG I/Y mRNA.* Comparison of the nucleotide sequences of human and mouse HMG Y cDNAs at the 3' end. The two conserved regions are boxed; the conservation of the sequence in box A is 94%, in box B 88%. The numbering of the human sequence is the same as in figure 3, the numbering of murine HMG Y refers to Johnson et al. (14).

sequentially with a HMG I/Y specific probe and with a GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) gene fragment. The GAPDH probe served as internal standard to correct for different mRNA levels, as its messenger is not down regulated in fibroblast cells arrested for two days in G_0 phase by serum starvation (28). From figure 6 it is evident that both the signals obtained with the HMG I/Y probe and with the GAPDH probe are about 10× weaker under the condition of serum starvation as compared with exponential growth. Parallel reduction of both mRNAs is compatible with a general down-regulation of the whole cellular mRNA pool under conditions of prolongated G_0 arrest (29), rather than a specific down-regulation of HMG I/Y mRNA. Most important, even

after three days of proliferation arrest, HMG I/Y mRNA is still readily detectable. Essentially the same results were obtained by arresting cells for 3 days with the G1 phase specific inhibitor indomathacin (data not shown). Thus, we conclude that HMG I/Y mRNA is not specifically down regulated by a stop of proliferation and that the mere being quiescent of terminally differentiated cells can only partially (if at all) account for the absence of HMG I/Y message in these cells.

DISCUSSION

By screening an expression library using an octamer binding site as a probe, we have cloned cDNAs corresponding to human HMG I and (most likely) HMG Y proteins. Our analysis demonstrates that two distinct but related mRNAs are encoding these proteins. Since at present the human HMG Y protein has not been sequenced, we are not absolutely certain that the mRNA without the 33bp fragment codes for it. However the following arguments speak in favour of this assumption: i) HMG Y migrates faster than HMG I on both SDS- and acid urea gels (30); this is in agreement with the lower molecular weight of HMG Y as predicted from the cDNA sequence. ii) Amino acid analysis and peptide mapping showed that the two proteins are highly similar (e. g. four out of five peptides are identical) (30, 31). A final test of this assumption has to come from amino acid sequence analysis.

The occurence of the two highly similar mRNAs raises the question of whether there is one gene encoding both proteins by alternative splicing or two distinct genes. We favour the first possibility because the 33bp fragment is flanked by sequences which match closely the consensus sites (32) for splice donors and acceptors (figure 7). According to the model of alternative splicing, the 33bp fragment is recognized from the splicing machinery once as an exon to generate HMG I mRNA and once as an intron to give rise to HMG Y mRNA. The reason for this leaky splicing event is most probably to be found in the missing poly-pyrimidine stretch that is usually located upstream of the 3′ end of an constitutively spliced out intron.

What might be the function of the additional 11 amino acids in HMG I ? Clearly this peptide is distinct from the rest of the protein in that it does not contain any charged amino acid and in that it has a slightly hydrophobic character. It therefore could be used as a dimerisation domain for the formation of homo- or heterodimers. The 11 amino acid difference between HMG I and HMG Y has no influence on the principal ability of the two proteins to bind to A-T rich stretches of DNA, as both proteins gave strong signals during the expression screening. But the presence or absence of this region might modulate the affinity or specificity with which the two isoforms bind to various sites in the genome. The posttranslational modifications shown to occur at HMG I protein (33) do not appear to be necessary for specific DNA binding, since HMG I/Y proteins synthesized in E. coli are unlikely to be appropriately modified, yet they bound tightly to DNA.

It was interesting to note that HMG I and HMG Y proteins seem to be capable of binding in general to the octamer sequence motif but not to an SRE motif. Both the SRE and the octamer site of the immunoglobulin light chain promoter used in this study have a run of six A-T basepairs. Binding can therefore not only be directed by any run of six contiguous A-T basepairs as suggested by Solomon et al. (9), but may also be dependent on other criterias such as flanking sequences and the ability of a DNA fragment to adopt a sterical conformation allowing binding of HMG I/Y. The octamer sequence motif seems to provide

such a conformation, which might include an enlarged minor groove because OTF 1, OTF 2 and HMG I/Y (α-protein) exhibit several minor groove contacts which are absolute essential for their binding (25, 9, our unpublished data). Our results are supported by a finding of Solomon et al. (9) who showed by footprinting experiments that one of the binding sites of monkey HMG I has the sequence gcctATaCAAATctac. This sequence corresponds exactly to the octamer motif (ATGCAAAT) except one mismatch towards the middle and it has only four contiguous A-T basepairs as does the octamer element from the histone H2B promoter. It therefore seems that binding sites for HMG I/Y include also sequences that are only moderately A-T rich. At the moment one can only speculate about the biological significance of the observed binding of HMG I/Y to the octamer motif. Clearly OTF 1 and OTF 2 bind with higher affinity to the octamer site than HMG I/Y (unpublished data); but binding of HMG I/Y proteins to the octamer motif might still contribute to the regulation of genes containing this sequence element.

In respect to their coding region HMG I/Y mRNAs possess unusually long 5′ and especially 3′ untranslated regions. Only 16% of the HMG I messenger encodes amino acids. The 5′ leader sequence has a length of at least 231bp, which is twice as long as a typical leader sequence (average length of about 100bp; reference 34). Messenger RNAs with long 5′ non coding regions are often subjected to translational control mechanisms, as documented for the transcription factors SRF (serum response factor) and GCN4 (35, 36). The 3′ untranslated region has a length of 1.3kb and is devoid of long A-T rich stretches that were shown to destabilize other mRNAs such as fos and GM-CSF (37, 38). However it is intriguing to note the presence of two regions at the 3′ trailer of the human and mouse HMG I/Y cDNAs which are as highly conserved in nucleotide sequence as the sequence of the coding body (see figure 8). These conserved sequence elements might play some regulatory function either at the transcriptional or posttranscriptional level. Further experiments are needed to elucidate this issue.

We also noted the presence of 9 HpaII restriction enzyme cutting sites within the first 420bp, compared with none in the remaining 1.45kb of the cDNA. The clustered presence of HpaII sites is diagnostic for undermethylated CpG-rich DNA, the so called CpG islands (39). However, it remains to be established whether these CpG clusters are really unmethylated in the genome and whether the methylation state of them or of other regions of the gene changes during differentiation. The last point is especially interesting because methylation is often associated with the inactivation of the particular gene. Since there is apparently no HMG I/Y mRNA present in terminally differentiated cells, this might be one possibility of inactivating HMG I/Y genes.

During the course of our studies we investigated whether a proliferation arrest per se is sufficient for down regulating HMG I/Y transcripts. Our results demonstrate that no specific down-regulation occurs upon stopping cell proliferation of undifferentiated fibroblasts. It therefore seems that the switch off mechanism, responsible for the absence of HMG I/Y mRNAs and proteins in terminally differentiated cells, is tightly coupled to the differentiation process of a cell and does not specifically respond to changes in the proliferation state alone. This mode of control is clearly different from that used for the regulation of histone mRNA levels. In contrast to HMG I/Y mRNAs, histone transcripts are specifically and strongly down regulated by an arrest in G_0-phase (40). Both messengers code for chromosomal proteins, yet they are differently regulated. It will be interesting to investigate the molecular basis for this different regulation and to determine how the switch off mechanism for HMG I/Y expression exactly works.

ACKNOWLEDGMENTS

We thank Drs Denise Barlow, Meinrad Busslinger, Fu Xiao-Cang, Gulio Superti-Furga and Edgar Schreiber for their kind gifts of materials used in this study. We are indebted to Dr Gotthold Schaffner for oligonucleotide synthesis and to Sissy Ender, Ivan Botto and Robert Kurzbauer for performing sequencing reactions. We are grateful to Dr Adrian Bird for his comments on the manuscript. The expert graphic work of Ingeburg Hausmann is greatly acknowledged. Last but not least we wish to thank Dr Anton Beyer for his patient help with computing.

*To whom correspondence should be addressed

REFERENCES

1. John, E.W. (ed) (1982) The HMG Chromosomal Proteins, Academic Press, London.
2. Lund, T., Holtlund, J., Fredriksen, M., and Laland, S.G. (1983) FEBS Lett. 152, 163–167.
3. Weisbrod, S. and Weintraub, H. (1981) Cell 23, 391–400.
4. Shick, V.V., Belyavski, A.V., and Mirzabekov, A.D. (1985) J. Mol. Biol. 185, 329–339.
5. Cook, G.R., Minch, M., Schroth, G.P., and Bradbury, E.M. (1989) J. Biol. Chem. 264, 1799–1803.
6. Weisbrod, S. (1982) Nature 297, 289–295.
7. Tremethick, D.J., and Molloy, P.L. (1988) Nucleic Acids Res. 16, 11107–11123.
8. Ptashne, M. (1988) Nature 335, 683–689.
9. Solomon, M., Strauss, F., and Varshavsky, A. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 1276–1280.
10. Bianchi, M.E., Beltrame, M., and Paonessa, G. (1989) Science 243, 1056–1059.
11. Strauss, F., and Varshavsky, A. (1984) Cell 37, 889–901.
12. Goodwin, G.H., Cockerill, P.N., Kellam, S. and Wright, C.A. (1985) Eur. J. Biochem. 149, 47–51.
13. Vartianen, E., Palvimo, J., Mahonen, A., Linnala-Kankkunen, A., and Mäenpää, P.H. (1988) FEBS Lett. 228, 45–48.
14. Johnson, K.R., Lehn, D.A., Elton, T.S., Barr, P.J., and Reeves, R. (1988) J. Biol. Chem. 263, 18338–18342.
15. Lund, T., Dahl, K.H., Mork, E., Holtlund, J., and Laland, S. G. (1987) Biochem. Biophys. Res. Commun. 146, 725–730.
16. Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H., and McKnight, S.L. (1988) Genes Dev. 2, 801–806.
17. Gerster, T., Matthias, P., Thali, M., Jiricny, J., and Schaffner, W. (1987) EMBO J., 6, 1323–1330.
18. Barberis, A., Superti-Furga, G. and Busslinger, M. (1987) Cell, 50, 347–359.
19. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning–A Laboratory Manual (Cold Spring Harbour, New York).
20. Feinberg, A. and Vogelstein, B. (1984) Anal. Biochem. 137, 266–267.
21. Church, G. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 1991–1995.
22. Devereux, J., Haeberli, P., and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
23. Bayer, B.M., Kruth, H.S., Vaughan, M., and Beaven, M.A. (1979) J. Pharmacol. Exp. Therap. 210, 106–111.
24. Favoloro, J., Treisman, R. and Kamen, R. (1980) Meth. Enzymol. 65, 718–749.
25. Staudt, L.M., Singh, H., Sen, R., Wirth, T., Sharp, P.A. and Baltimore, D. (1986) Nature 323, 640–643.
26. Müller, M.M., Ruppert, S., Schaffner, W. and Matthias, P. (1988) Nature 336, 544–551.
27. Lindsay, S. and Bird, A.P. (1987) Nature 327, 336–338.
28. Almendral, J.M., Sommer, D., MacDonald-Bravo, H., Burckhardt, J., Perera, J., and Bravo, R. (1988) Mol. Cell. Biol. 8, 2140–2148.
29. Marcus, M., Fainsod, A., and Diamond, G. (1985) Ann. Rev. Genetics 19, 389–421.
30. Lund, T., Holtlund, J., and Laland, S.G. (1985) FEBS Lett. 180, 275–279.
31. Lehn, D.A., Elton, T.S., Johnson,K.R., and Reeves, R. (1988) Biochem. Int. 16, 963–971.
32. Mount, S.M. (1982) Nucleic Acids Res. 10, 459–475.
33. Elton, T.S., and Reeves, R. (1986) Anal. Biochem. 157, 53–62.
34. Kozak, M.(1987) Nucleic Acids Res. 15, 8125–8148.
35. Norman, C., Runswick, M., Pollock, R., and Treisman, R. (1988) Cell 55, 989–1003.
36. Mueller, P.P., and Hinnebusch, A.G. (1986) Cell 45, 201–207.

37. Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989) Genes Dev. 3, 60–72.
38. Shaw, G. and Kamen, R. (1986) Cell 46, 659–667.
39. Bird, A.P. (1986) Nature 321, 209–213.
40. Schümperli, D. (1988) Trends Genet. 4, 187–191.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.