Volume 17 Number 3 1989 Nucleic Acids Research

A human placental cDNA clone that encodes nonhistone chromosomal protein HMG-1

Lisa Wen, Jenq-Kuen Huang, Barbara H.Johnson and Gerald R.Reeck

Department of Biochemistry, Kansas State University, Manhattan, KA 66506, USA

Received July 21, 1988; Revised and Accepted January 10, 1989 EMBL accession no. X12597

ABSTRACT

From a human placental λgt11 cDNA library, we have isolated a cDNA clone that encodes the entire 215-residue amino acid sequence of HMG-1. Analysis of an internal sequence similarity suggests that the DNA-binding domains of HMG-1 are separated by a rather long and flexible linker segment. Southern blotting of DNA digested with BamHI indicated a highly variable number of genes (or pseudogenes) for HMG-1 in different species. Characterization of HMG-1 mRNA expression by Northern blotting showed that three mRNA species of approximately 1.0, 1.4 and 2.4 kb were expressed in all mammalian organs and cell lines examined. These included several rat organs at different stages of development. Northern analysis also suggested the occurrence of HMG-1 mRNA in an invertebrate and a plant species.

INTRODUCTION

The High Mobility Group (HMG) proteins are perhaps the most extensively studied nonhistone chromosomal proteins (for reviews see Ref. 1,2). They comprise two unrelated families: the HMG-1 family (of higher molecular weight, around 25,000), and the HMG-14 family (of lower molecular weight, around 10,000). The proteins within each family are similar in their sequences and conformations (3). Although no bona fide HMG proteins have been isolated from invertebrates, HMG proteins have been identified from lower vertebrates, most prominently from trout. Trout homologs of HMG-1 and HMG-14 are HMG-T and H6, respectively (4).

Members of the HMG-1 family appear to have a three domain structure (5). The N-terminal and central domains are basic, DNA-binding regions. The highly acidic C-terminal domain -- or module (6) -- is believed to interact with histones. The functions of the HMG-1 family are not yet firmly established. Suggested functions include an involvement in transcription or DNA replication (1,2). These suggestions have been based in

part on interactions in which HMG-1 and its homologs participate. For instance, the proteins exhibit preferential affinity for single-stranded DNA (7,8). They are capable of introducing negative supercoils into nicked, circular DNA (9). Furthermore, HMG-1 has nucleosome assembly factor activity (10).

Cloned cDNA's encoding HMG-1 or HMG-T have been isolated from bovine (11), CHO cells (12), rat (13), and trout (14). Only the rat sequence appears to encode an entire protein. We report here the isolation and characterization of a cDNA clone that encodes the entirety of human HMG-1. We discuss the possible structural implications of an internal sequence similarity and use the cDNA as a probe for Northern and Southern analyses.

MATERIALS AND METHODS

Library Screening Approximately 6 x 10^5 recombinant plaques from a human placental λgt11 cDNA library (Clontech Laboratories, Palo Alto, CA) were screened by plaque hybridization with two synthetic oligonucleotide probes (99 nucleotides long). The probes were based on the sequence of a cDNA clone for bovine HMG-1 (11) and correspond to nucleotide positions 10 to 108 and 145 to 243, respectively. These encode residues 130 to 162 and 175 to 220 of bovine HMG-1 (using the numbering scheme of Figure 5).

The plaques were lifted twice onto colony/plaque screen hybridization transfer membranes (New England Nuclear, Boston, MA) and the phage were lysed by autoclaving. Blocking and hybridization were carried out at 37°C in a solution containing 50% (v/v) formamide, 1% (w/v) SDS, 1 M NaCl, 100 µg/ml denatured salmon sperm DNA and 10% (w/v) dextran sulfate. The hybridization probes were labeled with ^{32}P-dCTP (3,000 Ci/mmol; New England Nuclear, Boston, MA) by oligo-labeling (15,16) and added to 10^6 cpm/ml of hybridization solution. Membranes were washed three times at room temperature in 2X SSC, 1% SDS for 10 min each and then used for autoradiography.

Template Isolation and DNA Sequence Analysis The cDNA inserts from positive clones were excised with EcoRI, separated on a 1% agarose gel, purified and subcloned into M13mp18 (17).

These constructions were used to transform JM 109 competent cells (17). White plaques were picked and the single-stranded DNA from the phage containing the inserts in both orientations were isolated as described (18). The nucleotide sequences of both strands of cDNA were determined by the dideoxynucleotide chain termination method (19) utilizing T7 DNA polymerase (United States Biochemical Corporation, Cleveland, Ohio) and synthetic primers (20).

RNA Isolation and Northern Blot Analysis Total RNA was isolated from various sources by a guanidine thiocyanate method (21). Thirty micrograms of total cellular RNA were fractionated on 1.2% agarose gels containing 2.2 M formaldehyde and blotted onto nitrocellulose membranes (22). Blots were treated for 6 h in formamide (see figure legends for concentrations), 5X SSC, 5X Denhardt's solution, 0.1% SDS, 10% dextran sulfate and 100 ug/ml denatured herring sperm DNA (23). The probe (the cDNA insert of clone HP3A-1) was added to the hybridization solution at 10^6 cpm/ml and incubated overnight at 42°C. Blots were washed as described in figure legends. Blots were exposed to Kodak X-Omat AR film with intensifying screens.

DNA Isolation and Southern Blot Analysis High molecular weight genomic DNA was isolated from HeLa cells, bovine liver, chicken liver, and catfish (Pimelodella gracilis) muscle by a guanidinium thiocyanate ultracentrifugation method (21). The DNA was digested with BamHI, fractionated on a 0.7% agarose gel, and blotted onto Gene Screen Plus nylon membrane in 0.4 M NaOH, 1.5 N NaCl (24). The blot was incubated for 4 h at 42°C in 25% formamide (deionized), 1 M NaCl, 1% SDS, 10% dextran sulfate, and 100 ug/ml denatured herring sperm DNA. The blot was then allowed to hybridize to the bovine HMG-1 probe (the 99-mer corresponding to nucleotides 10 to 108 of bovine HMG-1 cDNA). The probe was added to the hybridization solution at 10^6 cpm/ml and incubated at 42°C for 16 h. The blot was washed twice at room temperature in 2X SSC, 1% SDS for 5 min, followed by washing under increasingly stringent conditions by increasing the washing temperature in 0.5X SSC, 1% SDS. The blot was exposed to Kodak X-Omat AR film with intensifying screens at -70°C.

Sequence Alignments The Sankoff algorithm (25) as implemented by de Haen et al. (26) was used as previously described (1,5) to compare amino acid or nucleotide sequences. Scores of McLachlan (27) were used. The Sankoff algorithm evaluates the number of statistically justified gaps without using a gap penalty. The number of allowed gaps is identified by comparing the increases in the score of the alignment of the actual sequences to increments of scores of alignments of pairs of randomized sequences of the same lengths and compositions as the actual sequences. We select that number of gaps at which the incremental score loses its statistical significance (26).

In investigating the internal sequence similarity in HMG-1, we examined all possible dispositions of residues 80-89, including dividing it between the two similar segments. These residues did not, however, make a positive contribution to the alignment, regardless of where they were placed.

We will use the following term to describe the level of similarity between two sequences: "Identity index" is the percentage of positions in an alignment at which the amino acids or nucleotides are identical.

RESULTS

Isolation and Characterization of a Human HMG-1 cDNA Clone Two synthetic oligonucleotide probes based on a bovine HMG-1 cDNA sequence (11) were used to screen a human placental cDNA library. Of $6x10^5$ plaques screened, approximately 100 hybridized to both probes. Thirty strongly hybridizing clones were plaque purified and DNA was isolated digested with EcoRI and analyzed on a 1% agarose gel. The DNA was transferred to Gene Screen Plus and positive clones were confirmed by Southern blot analysis. One of the positive clones with a cDNA insert about 1 kb was subjected to further analysis.

This cDNA insert (HP3A-1) was subcloned into M13mp18 and sequenced according to the strategy shown in Figure 1. The nucleotide sequence of clone HP3A-1 and the predicted amino acid sequence of human HMG-1 are shown in Figure 2. The cDNA insert contained 52 bp of 5'-noncoding sequence followed by an open reading frame of 645 bp encoding a polypeptide of 215 amino

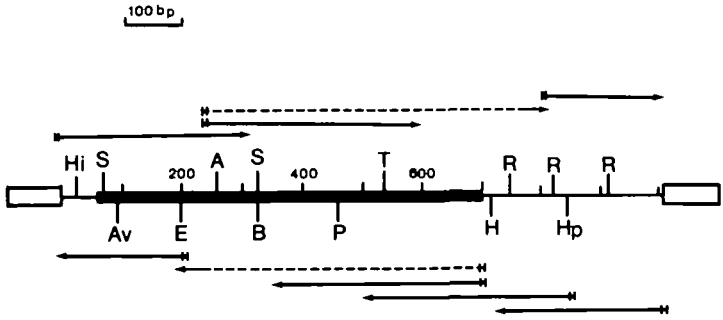


FIGURE 1. Restriction Map and Sequencing Strategy for Human HMG-1 cDNA Clone HP3A-1. The DNA insert in the original clone, HP3A-1, was subcloned into M13mp18, and the single stranded templates from both orientations were isolated and used for single stranded DNA sequencing using synthetic oligonucleotide primers. After the initial round of sequencing using universal primers, sequences were extended using synthetic 21-mer oligonucleotide primers whose sequences were derived from the previous round of sequencing. The open boxes refer to the M13 sequences; the thin line represents the noncoding region of clone HP3A-1, and the thick line represents coding sequences. The locations of several restriction sites are indicated, as are the locations of the sequencing primers used and the extent of sequencing derived from each primer. Reaction conditions were altered according to the manufacturer's recommendations to obtain long sequences (dashed lines). The primers are represented by the 2 bars at the beginning of each arrow. Abbreviations for the restriction endonuclease sites are: A, AsuI; Av, AvaIII; B, Bam H1; E, Eco PI; H, HhaI; Hi, HinfI; Hp, HpaI; P, PstI; R, RsaI; S, Sau3a; T, TaqI.

acids with a calculated molecular mass of 24,896. The 3'-noncoding region (309 bp) of the insert contains a possible polyadenylation signal, AAGAAA, at a site 76 nucleotides 3' from the stop codon.

Expression of HMG-1 mRNA in Several Species We performed Northern analyses of RNA from human placenta, and from kidney or testis from several animal species (bovine, sheep, rat, hamster, mouse, chicken, turtle), as well as of RNA from several mammalian cell lines, fish muscle, insect muscle, and developing rice seeds (Figure 3). The human HMG-1 cDNA probe reacted with mRNA from all sources examined. Three bands of hybridization, of approximate sizes 1.0, 1.4, and 2.4 kb, were detected in all mammalian sources (i.e., human, bovine, sheep, rat, hamster, and

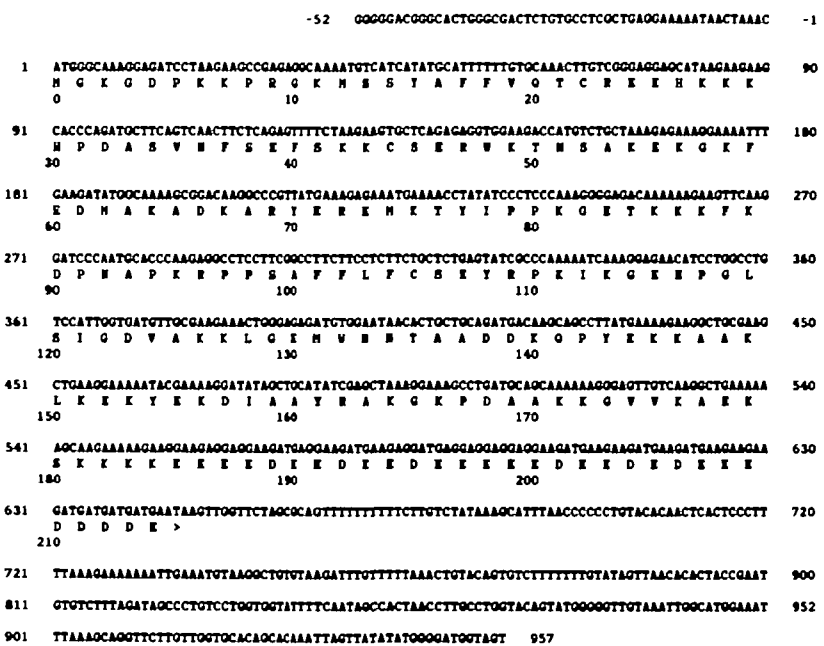


FIGURE 2. Primary Sequence of Human HMG-1 cDNA and the Predicted Amino Acid Sequence of Human HMG-1. The primary sequences of both human HMG-1 cDNA and its encoded protein are given. Position 0 represents the initiator methionine. The 5'-untranslated sequence is shown as negative numbers. The coding sequences begins at nucleotide +1 and the 3'-untranslated sequence terminates at 957.

mouse). The hybridizing bands from other species showed species-dependent heterogeneity and variation in sizes of the bands of hybridization.

HMG-1 mRNA Levels in Developing Rat Organs Northern analysis of total RNA from several rat organs at several stages of development revealed mRNA species of the 3 characteristic sizes in all tissues examined (arrows in Figure 4). The 1.4 kb mRNA species was the predominant form in all cases. The 2.4 kb mRNA species was always less abundant and the 1.0 kb mRNA was always least abundant. Kidney appeared to differ from other tissues examined in having a higher ratio of the 2.4 kb species to the 1.4 kb species.

Tissue-specific differences were evident in the develop-

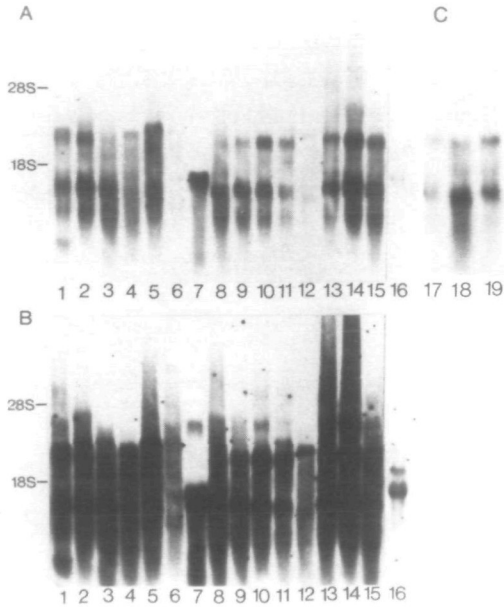
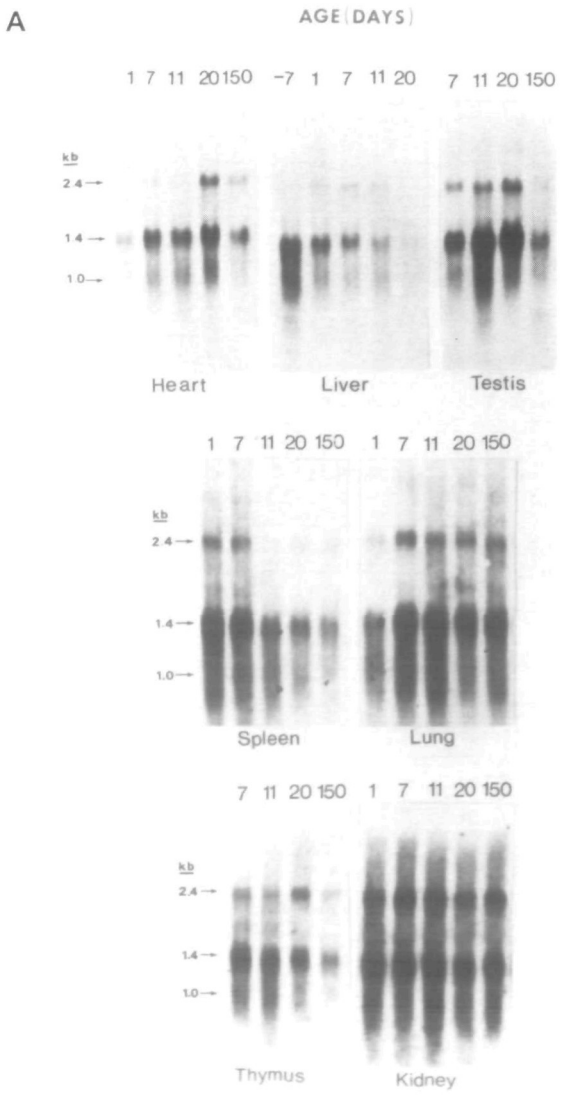


FIGURE 3. Northern Blot Analysis of RNA from Various Species. The samples were total RNA from: lane 1, rat testis; lane 2, sheep testis; lane 3, bovine testis; lane 4, turtle testis; lane 5, chicken testis; lane 6, catfish muscle; lane 7, tobacco hornworm (*Manduca sexta*) muscle; lane 8, golden hamster kidney; lane 9, mouse kidney; lane 10, rat kidney; lane 11, sheep kidney; lane 12, turtle kidney; lane 13, NIH 3T3 fibroblasts; lane 14, 3T3-L1 fibroblast; lane 15, H4IIE hepatoma cells; lane 16, immature rice seeds; lane 17, bovine kidney; lane 18, HeLa cells; lane 19, human placenta. RNA samples (30 µg/lane except lane 13, which contained 20 µg total RNA) were subjected to electrophoresis, transferred to nitrocellulose, and hybridized at 42°C in a solution containing 25% formamide, using the insert from clone HP3A-1 as a probe. In A, the blot was washed at room temperature in 2X SSC, 0.1% SDS for 20 min each , and then washed at 50°C in 2X SSC and exposed for 16 h; in B, for 3 days. In C, the blot was hybridized at 42°C in 50% formamide using the same probe. The blot was washed twice at room temperature in 2X SSC, 0.1% SDS and then washed once at 65°C in 0.1X SSC, 0.1% SDS.

mental pattern of HMG-1 mRNA abundance (Figure 4). In liver, thymus, and spleen, HMG-1 mRNA levels were highest before birth or within the first week of life, then declined. In testis, the levels of HMG-1 mRNA were low on day 7 and increased in amount from this age to day 20. The levels then decreased dramatically

after sexual maturation, i.e., between 20 and 150 days of age. In heart, levels of HMG-1 mRNA were low on day 1 or day 150, and were higher between days 7 and 20. In lung, levels of HMG-1 mRNA were lowest on day 1, significantly increased on day 7, and remained high thereafter. In kidney, the amount of HMG-1 mRNA remained essentially unchanged throughout development.



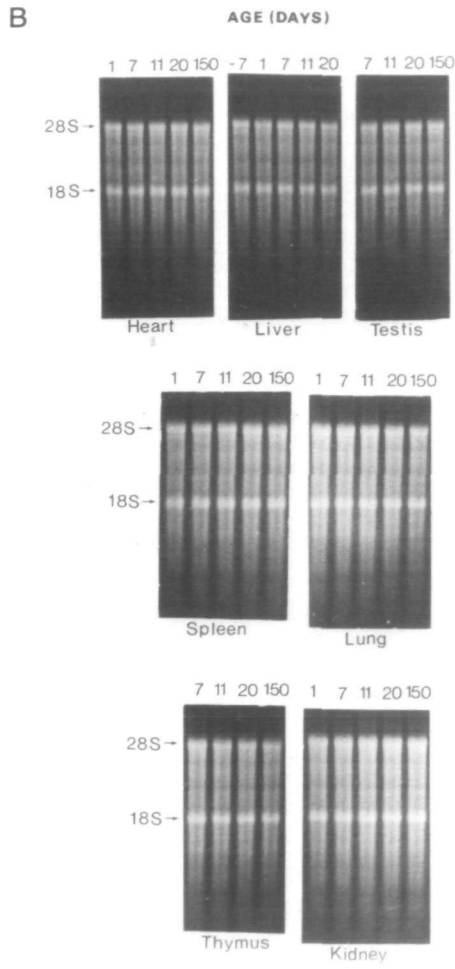


FIGURE 4. Northern Blot Analysis of RNA from Rat Organs during Development. Total RNA was obtained from rat tissues of male rats in utero and at various ages. 30 µg of total RNA was loaded into each gel lane. Ages, in days, are indicated at the top of each series of blots, and the relative molecular sizes of hybridizing species are indicated. A: Northern blot probed with insert DNA of clone HP3A-1 in a solution containing 50% formamide at 37°C. The blot was washed as described in figure 3C. B: Gel stained with ethidium bromide to demonstrate constancy of amount of total RNA loaded.

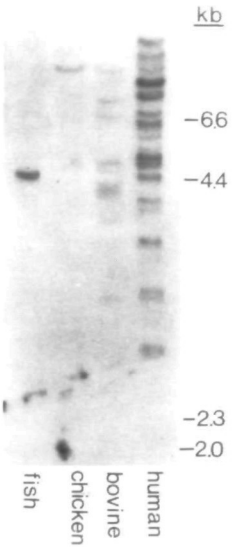


FIGURE 5. Southern Hybridization of Bovine HMG-1 Probe to Human, Bovine, Chicken, and Fish Genomic DNA's. Genomic DNA (20 µg) from HeLa cells, bovine liver, chicken liver or catfish was digested with BamHI, separated on a 0.7% gel, transferred to Gene Screen Plus membrane, and allowed to hybridize to the synthetic bovine HMG-1 probe (99-mer). The blot was washed under increasingly stringent conditions. The autoradiogram showed here represents the blot that was washed in 0.5X SSC, 1% SDS at 50°C and exposed for 18 days. The positions of molecular size markers are indicated to the right of the blot.

Southern Analysis of DNA from Several Species Human (HeLa cell), bovine, chicken and fish DNA's were subjected to Southern analysis following digestion with BamHI. The results (Fig. 5) demonstrate that, in different species, widely different numbers of bands hybridize with a probe for HMG-1.

Comparison of Amino Acid Sequences of HMG-1's Six complete or partial amino acid sequences of members of the HMG-1 family are shown in an alignment in Figure 6. The sequences include bovine HMG-1 (3,11,14), rat HMG-1 (13), HMG-1 from CHO cells (12), trout HMG-T as inferred from a cDNA clone (14), HMG-T determined by amino acid sequencing (14,28,29) and human HMG-1 (this work). The sequences were aligned using a computerized algorithm that determines the number and placement of statistically significant gaps without using a gap penalty (see Methods). The sequences were sufficiently similar, however, that only in aligning the two HMG-T sequences (one inferred from a cDNA sequence and the other determined by amino acid sequencing) was there any ambiguity from simple visual inspection in how many gaps were appropriate. With respect to the human HMG-1 sequence the identity indexes were 98, 97, 98, 67, and 53%, respectively, for bovine HMG-1, rat HMG-1, CHO cell

```
                1        10        20        30        40        50
 Human HMG-1:   GKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKC------
Bovine HMG-1:   GKGDPKKPRGKMSSYAFFVQTSREEHKKKHPDASVNFSEFSKKC------
   Rat HMG-1:   GKGDPKKPRRKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKC------
   CHO HMG-1:   --------------------------------------------------
HMG-T (cDNA):   ----------------------------------ASVNFSEFSKKC------
HMG-T (prot):   PGKDPNKPKGKTSSYAFFVATSREEHKKKHSGAKVNGSESSKACGKSPRD
                                                  * ** ** ** * *

               51        60        70        80        90       100
 Human HMG-1:   ---SERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGET--KKKFK
Bovine HMG-1:   ---SERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGET--KKKFK
   Rat HMG-1:   ---SERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPK-ET--KKKFK
   CHO HMG-1:   ...SERWKTMSAKEKGKFEDMAKADKARYEREMKTYIPPKGET--KKKFK
HMG-T (cDNA):   ---SERWKTMSAKEKGKFEDLAKLDKVRYEREMRSYIPPKGEK--KKRFK
HMG-T (prot):   SKAPKRWRTMGAKEKVKFEDMAKGDKVRYDKDMKTYIPPKGEKAAGKRKK
                     ** ** **** **** ** ** **   *  *******    *  *

              101       110       120       130       140       150
 Human HMG-1:   DPNAPKRPPSAFFLFCSEYR-PKIKGEHPGLSIGDVAKKLGEMWNNTAAD
Bovine HMG-1:   DPNAPKRPPSAFFLFASEYR-PKIKGEHPGLSIGDVAKKLGEMWNNTAAD
   Rat HMG-1:   DPNAP-RASSAFFLFCSEYR-PKIKGEHPGLSIGDVAKKLGEMWNNTAAD
   CHO HMG-1:   DPNAPKRPPSAFFLFCSEYR-PKIKGEHPGLSIGDVAKKLGEMWNNTAAD
HMG-T (cDNA):   DPNAPKRPSSAFFIFCADFR-PQVKGETPGLSIGDVAKKLGEKWNNLTAE
HMG-T (prot):   DPNAPKRP-SAFFGYESAERAARIKADHPGMGIGDISKQLGLLWGKQSSK
                ******** ****      *    *   **  ***  * **  *

              151       160       170       180       190       200
 Human HMG-1:   DKQPYEKKAAKLKEKYEKDIAAYRAKGKPDA----------------AKK
Bovine HMG-1:   DKQPYEKKAAKLKEKYEKDIAAYRAKGKPDA----------------AKK
   Rat HMG-1:   DKQPYEKKAAKLKEKYEKDIAAYRAKGKPDA----------------AKK
   CHO HMG-1:   DKQPYEKKAAKLKEKYEKDIAAYRAKGKPDA----------------AKK
HMG-T (cDNA):   DKVPYEKKASKLKEKYEKDITAYRNKGKVP--------------------
HMG-T (prot):   DKLPHEAKAAKLKEKYEKCVAAYKPKGGAAAPARERVDKAKGTAGATAKH
                ** * * ** ********   **  **   *                 **

              201       210       220       230       240
 Human HMG-1:   GVVKAEKSKKKKEEEEDEEDEEDEEEEEDEEDEDEEEDDDDE
Bovine HMG-1:   GVVKAEKSKKKKEEEEDEEDEEDEEEEEDEEDEEEEEDDDDE
   Rat HMG-1:   GVVKAEKSKKKKEEEDDEEDEEDEEEEEEEEDEDEEEDDDDE
   CHO HMG-1:   GVVKAEKSKKKKEEEDDEEDEEDEEEEEEEEDEDEEEDDDDE
HMG-T (cDNA):   ---VSMPAKAAAPAKDDDDDDDDDDDDEDDDDDDDEDDE
HMG-T (prot):   GPGVPAVGKPKAAPMDDDDDDDDDEEEEEDDDEEEDDDDDDDD
                *       *       * *  *    *   *     * **
```

FIGURE 6. Comparison of Amino Acid Sequences of HMG-1's from Human, Bovine, Rat, and CHO cells, and of HMG-T from Trout. Amino acid sequences from human HMG-1, as shown in Fig. 2, were compared with published HMG-1 sequences from bovine (11,12,14), rat (13), CHO cells (12) and HMG-T sequences from analysis of a trout cDNA clone (14) and the trout protein (14,28,29). The numbers at both ends of each line are shown for reference and do not represent the original numbers of individual amino acid sequences from the published reports. Hyphens indicate gaps and an asterisk indicates that all amino acids that are identical at that position in the alignment.

HMG-1, HMG-T inferred from the cDNA, and the HMG-T determined by amino acid sequencing.

Internal Sequence Similarity in Human HMG-1 Using the inferred amino acid sequence of human HMG-1, we re-examined an internal similarity that was first noticed in the incomplete

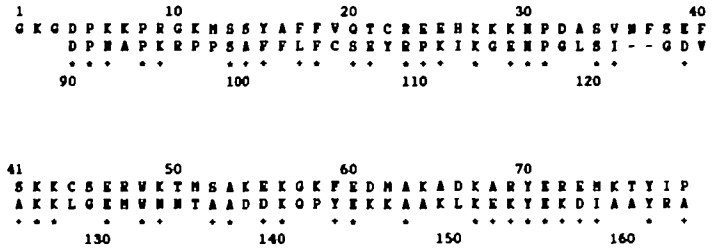


FIGURE 7. Internal Sequence Similarity in HMG-1. Two stretches of human HMG-1 that exhibit sequence similarity are shown in an alignment given by the Sankoff algorithm. Asterisks indicate identities; plus signs indicate positions at which two amino acid residues have higher than average scores.

sequences of bovine HMG-1 and HMG-2 and that formed part of the initial body of evidence for a tri-partite organization in this family of proteins (5). The Sankoff algorithm clearly established that only one gap was justified statistically in aligning the stretches encompassing residues 1-79 and 90-163 (Figure 7). The McLachlan score of the internal similarity was 5.9 standard deviations above the mean score from 200 pairs of randomized cases. The internal similarity is thus highly significant from a statistical standpoint.

Comparison of Nucleotide Sequences In the coding regions, the nucleotide sequences for the cDNA's encoding bovine HMG-1, rat HMG-1, CHO cell HMG-1, and trout HMG-T had identity indexes with respect to the human HMG-1 cDNA 91, 89, 92, and 69%, respectively (alignment not shown). A comparable level of similarity was evident in the 3'-noncoding regions for the bovine, rat, and CHO cell proteins, which had identity indexes with respect to the human sequence of 96, 99, and 86%. The trout HMG-T 3'-noncoding region was markedly more divergent, however, and had an identity index of 38% with respect to the human sequence.

DISCUSSION

The Inferred Amino Acid Sequence In this work we have isolated and characterized a cDNA clone (HP3A-1) that encodes human HMG-1. This clone, along with a recently reported rat

cDNA clone (13), are the first for any members of the HMG-1 family that encode the entire protein. It should be noted that a complete amino acid sequence has been reported for a homolog of HMG-1 (namely, trout HMG-T) (14). This sequence was from amino acid sequence analysis of the protein itself. The details of the determination of that sequence have not been described to our knowledge. Because it contains features not found in any of the HMG-1 or HMG-T sequences deduced from nucleotide sequences of cDNA clones, it is conceivable that the proposed sequence for HMG-T contains extra residues, just as did the sequences for bovine HMG-1 and HMG-2 determined at the protein level (30, and see below).

The amino acid sequence of human HMG-1 inferred from the sequence of clone HP3A-1 reinforces an important conclusion already drawn by Dixon and coworkers (11,12): the highly acidic region of members of the HMG-1 family is at the very C-terminus of the proteins. That is, the poly-Glu/Asp sequence does not have appended to it a stretch of polypeptide of mixed amino acid composition, as was proposed from the initial amino acid sequencing on bovine HMG-1 and HMG-2 (30).

As we have previously pointed out (1), proteins that interact with nucleic acids are over-represented in the group of proteins that have Glu/Asp-rich stretches. Earnshaw (31) has recently made a systematic analysis of DNA-binding proteins that have such stretches (which he called A- regions). Although numerous examples of proteins with A- regions now are clearly established, the A- regions of HMG-1 and its homologs are unique in being 100% Glu or Asp and notable in that they occur at the very C terminus of the proteins. The function of the A- region in HMG-1 is not known (nor is a function clearly assigned to any other A- region). Nonetheless, it seems clear that the A- region in HMG-1 is the site of interaction with histones (2), at least in vitro. Our proposal (5) that it functions in nucleosome assembly continues to seem reasonable (10).

Cys Residues in HMG-1 The availability of a complete sequence for a human HMG-1 molecule allows us to draw some conclusions about the Cys residues in members of the HMG-1 family. The human HMG-1 amino acid sequence inferred from clone

HP3A-1 contains 3 Cys residues (positions 22, 44, and 105 in Fig. 2). Only one (at position 44) is conserved in all members of the HMG-1 family for which we have large stretches of amino acid sequence (Figure 5). Thus, the total number of Cys residues is variable: the bovine sequence contains 1; the trout sequence, 2; and the human and rat, 3.

Cys-22 and Cys-44 are surrounded in the amino acid sequence by hydrophilic amino acids and are therefore likely to be on the surface of the proteins. They seem likely to participate in the intramolecular disulfide formation studied by Kohlstaedt et al. (32). A long and flexible connection between domains A and B, of the sort we postulate to exist (see below), could be important in allowing Cys residues on the two domains to come into juxtaposition for disulfide formation.

Internal Sequence Similarity and Domain Structure The availability of the complete amino acid sequence for human HMG-1 has led us to re-examine more critically an internal sequence similarity that we previously proposed (5) based on an analysis of the incomplete sequences of bovine HMG-1 and HMG-2 (30).

Our basic conclusion remains the same: two essentially contiguous stretches of HMG-1 exhibit a similarity in sequence that is highly significant statistically. The complete sequence has, however, allowed us to delimit the regions of similarity more reliably than was possible previously. This is directly relevant to thinking about the three-dimensional organization of the protein. In human HMG-1 the regions of internal sequence similarity comprise residues 1-79 and 90-163. Residues 80-89 (having the sequence PKGETKKKFK) are not accommodated well in the internal alignment. We suggest that these 10 residues constitute a linker between domains A and B. Domains A and B are believed to be globular regions and are known to have high helix contents (2,5). If the last portion of domain A were helical (as is predicted (5) by the Chou and Fasman rules) the PP sequence at residues 79 and 80 would thoroughly disrupt that helix and lead into the highly hydrophilic linker.

Our view at this point of the overall architecture of the first two domains is that of dumbbell with a long and flexible linker. Ten linker residues, fully extended, would cover 38 Å,

which is considerably greater than the diameter of either domain A or B, if they are approximately spherical. The notion of a long and flexible linker between domains A and B is rather different from any previous models for HMG-1, including ours (5), and could have important functional consequences.

If we divide the entire molecule into three regions based on the internal similarity, the third region (or module) would consist of residues 164-214. The last 29 residues of this module comprise the A- region. The first portion of the C-terminal module (residues 164-184) is highly basic in character (with a net charge of +8). Its highly charged nature suggests that it, like the A- region, is probably devoid of defined three-dimensional structure (at least when HMG-1 is not complexed with other macromolecules).

The architecture of HMG-1 and its homologs, as depicted in our model, is striking in its resemblance to Sigler's model of the "generic transcriptional-regulatory protein" (33). One of the striking similarities between the models is the presence of what Sigler calls an "acid blob."

Multiple mRNA Species for HMG-1 A pattern of three bands has been reported by Dixon and coworkers from Northern analyses of RNA from bovine and CHO cells and attributed to use of alternate polyadenylation sites (12). Using human HMG-1 cDNA as a probe, we have found a similar pattern in Northern blots of RNA from several mammalian organs and cell lines. Whereas this pattern may conceivably be common to RNA from most or all mammalian sources, markedly different patterns are evident in RNA from organisms other than mammals (Fig. 3).

Developmental Expression of Rat HMG-1 mRNA The 3 different sizes of mRNA were revealed in Northern blots of RNA from several rat organs examined at different times of development (Fig.4).

In liver, thymus, and spleen, the levels of HMG-1 mRNA were highest before birth or within the first week of life, then declined. This result would seem to suggest a correlation between the level HMG-1 mRNA's with the state of cell proliferation within those organs. In heart, lung, kidney, and testis, however, HMG-1 mRNA levels showed no such correlation.

Thus, no simple insight is offered into the possible functions of HMG-1.

It is interesting to note that the levels of HMG-1 mRNA in rat testis decreased dramatically after sexual maturation. It has been reported that there is a developmentally regulated decrease in HMG-1 protein levels in germ cells (34).

Multiple Genes for HMG-1 Two-dimensional gel electrophoresis has indicated that HMG-1 (and other members of the HMG-1 family) each consist of several components (35). Furthermore, Southern analysis has indicated that several genes (or pseudogenes) for HMG-1 exist in pig (36). The potentially important insight from our Southern analysis (Figure 5) is that the number of genes (or pseudogenes) in different species varies widely. In this respect, the situation for HMG-1 resembles that for HMG-14 and HMG-17, for which only single gene copies occur in chicken (37,38) but for which approximately 50 to 70 copies occur in the human genome (39).

Conservation of HMG-1 during Evolution It has been recognized for several years that HMG-1, HMG-2, HMG-E and HMG-T are members of a family of proteins that are derived from a common ancestor. Bona fide members of the HMG-1 family have, however, not been identified outside of vertebrates (for discussion of this point, see Ref. 5).

As additional sequences accumulate, we are gradually gaining a clearer picture not only of structure of the proteins, but of the evolution of the family. The picture that is emerging suggests that the proteins are rather highly conserved in sequence. The human, bovine, rat, and hamster sequences are identical in well over 90% of their aligned positions. HMG-T, from trout, is identical in 58% of the aligned positions with respect to human HMG-1. It is therefore somewhat surprising that proteins with clear-cut sequence similarity to HMG-1 have not yet been identified outside vertebrates (1). In our Northern blotting of RNA preparations from an insect and a grass, we observed distinct and strong hybridization signals. This suggests that recombinant DNA methods may reveal the existence of members of the HMG-1 family from sources other than vertebrates. In fact, this may already been realized in the

recent description of a genomic DNA clone from yeast that encodes a protein that is a possible homolog of HMG-1 (40).

ACKNOWLEDGMENTS

This work was supported by the Kansas Agricultural Experiment Station (publication number 88-573-J) and by a research grant from the National Institutes of Health (CA-17782). We thank Kirk Clark for assistance with sequence alignments and their display and Michael Denton for stimulating discussions.

REFERENCES

1. Reeck, G.R. and Teller, D.C. (1985) In Bekhor J. (ed), Progress in Nonhistone Protein Research, CRC Press, Boca Raton, Florida, Vol. II, pp. 1-22.
2. Puigdomenech, P. and Jose, M. (1985) In Reeck, G.R. Goodwin, G.H. and Puigdomenech, P. (eds), Chromosomal Proteins and Gene Expression, Plenum Press. New York, pp. 249-262.
3. Walker, J.M. (1982) In Johns, E.W. (ed), The HMG Chromosomal Proteins, Academic Press, London, pp. 69-89.
4. Dixon, G.H. (1982) In Johns, E.W. (ed) The HMG Chromosomal Proteins, Academic Press, London, pp. 149-192.
5. Reeck, G.R., Isackson, P.J. and Teller, D.C. (1982) Nature 300, 76-78.
6. Reeck, G.R. (1985) In Reeck, G.R., Goodwin, G.H. and Puigdomenech P. (ed), Chromosomal Proteins and Gene Expression, Plenum Press, New York, pp. 1-17.
7. Isackson, P.J., Fishback, J.L., Bidney, D.L. and Reeck, G.R. (1979) J. Biol. Chem. 254, 5569-5572.
8. Isackson, P.J., Cox, D.J., Manning, D. and Reeck, G.R. (1985) In Bekhor, J. (ed) Progress in Nonhistone Protein Research, CRC Press, Boca Raton, Florida, Vol. II, pp. 23-39.
9. Javaherian, K., Liu, L.F. and Wang, J.C. (1978) Science 199, 1345-1347.
10. Bonne-Andrea, C., Harper, F., Sobczak, J. and DeReconde, A-M. (1984) EMBO J. 3, 1193-1199.
11. Pentecost, B. and Dixon, G.H. (1984) Biosci. Reports 4, 49-57.
12. Lee, K-L.D., Pentecost, B.T., D'Anna, J.A., Tobey, R.A., Gurley, L.R. and Dixon, G.H. (1987) Nuc. Acids Res. 15, 5051-5068.
13. Paonessa, G., Frank, R., and Cortese, R. (1987) Nuc. Acids Res. 15, 9077.
14. Pentecost, B., Wright, J.M. and Dixon, G.H. (1985) Nuc. Acids Res. 13, 4871-4888.
15. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
16. Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.

17. Yanisch-perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
18. Williams, S.A., Slatko, B.E., Moran, L.S. and Desimone, S.M. (1986) Biotechniques 4, 138-147.
19. Sanger, F., Nicklen, S. and Coulson B.G. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
20. Sanchez-Pescador, R. and Mickey, S.U. (1984) DNA 3, 339-343.
21. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W. J. (1979) Biochemistry 18, 5294-5299.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
23. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
24. Reed, K.C., and Mann, D.A. (1985) Nuc. Acids Res. 13, 7207-7221.
25. Sankoff, D. (1972) Proc. Natl. Acad. Sci. USA 69, 4-6.
26. de Haen, C., Swanson, E. and Teller, D.C. (1976) J. Mol. Biol. 106, 639-661.
27. McLachlan, A.D. (1971) J. Mol. Biol. 61, 409-424.
28. Dixon, G.H. (1982) In Johns, E.W. (ed) The HMG Chromosomal Proteins, London and New York, Academic Press, pp 149-182.
29. Watson, D.C. and Dixon, G.H. (1981) Bioscience Reports 1, 167-175.
30. Walker, J.M., Gooderham, K., Hastings, J.R.B., Mayes, E. and Johns, E.W. (1980) FEBS Lett. 122, 264-270.
31. Earnshaw, W. (1988) Cell 105, 1479-1482.
32. Kohlstaedt, L.A., King, D.S. and Cole, R.D. (1986) J. Biol. Chem. 25, 4562-4565.
33. Sigler, P.B. (1988) Nature 333, 210-212.
34. Bucci, L.R., Brock, W.A., Goldknopf, I.L., Meistrich, M.L. (1984) J. Biol. Chem. 259, 8840-8846.
35. Wen, L., Tweeten, R.K., Isackson, P.J., Iandolo, J.J., and Reeck, G.R., (1983) Anal. Biochem. 132, 294-304.
36. Tsuda, K., Kikuchi, M., Mori, K., Waga, S., and Yoshida, M. (1988) Biochemistry 27, 6159-6163.
37. Dodgson, J.B., Browne, D.L., and Black, A.J. (1988) Gene 63, 287-295.
38. Landsman, D., Srikantha, T., and Bustin, M. (1988) J. Biol. Chem. 263, 3917-3923.
39. Landsman, D., and Bustin, M. (1986) J. Biol. Chem. 261, 16087-16091.
40. Haggren, W., and Kolodrubet, D. (1988) Mol. Cell. Biol. 8, 1282-1289.