her excellent help in preparing the manuscript.

Registry No. DC, 3571-34-4; DCOR, 91449-01-3; DHI, 3131-52-0; DHICA, 4790-08-3; Cu, 7440-50-8; Ni, 7440-02-0; Co, 7440-48-4.

REFERENCES

Barber, J., Townsend, D., Olds, D., & King, R. (1984) *J. Invest. Dermatol. 83*, 145–149.

Ito, S. (1986) *Biochim. Biophys. Acta. 883*, 155–161.

Kiss, T., & Gergely, A. (1985) *J. Inorg. Biochem. 25*, 247–259.

Korner, A., & Pawelek, J. (1980) *J. Invest. Dermatol. 75*, 192–195.

Korner, A., & Gettins, P. (1985) *J. Invest. Dermatol. 85*, 229–231.

Mason, H. S. (1955) *Adv. Enzymol. Relat. Subj. Biochem. 16*, 105–184.

Napolitano, A., Chioccara, F., & Prota, G. (1985) *Gazz. Chim. Ital. 115*, 357–359.

Palumbo, A., d'Ischia, M., Misuraca, G., & Prota, G. (1987) *Biochim. Biophys. Acta 925*, 203–209.

Pierce Chemical Co. (1986) BCA and BCA Protein Assay Reagent—Instruction 23230, 23225.

Raper, H. S. (1927) *Biochem. J. 21*, 89–96.

Raper, H. S. (1928) *Physiol. Rev. 8*, 245–282.

Solomon, E. (1981) in *Copper Proteins* (Spiro, T. G., Ed.) pp 41–108, Wiley-Interscience, New York.

Tietz, N. W. (1982) in *Cecil Textbook of Medicine* (Wyngaarden J. B., & Smith, L. H., Jr., Eds.) 16th ed. pp 2320–2354, W. B. Saunders, Philadelphia.

Townsend, D., Guillery, P., & King, R. A. (1984) *Anal. Biochem. 139*, 345–352.

Primary Structure of Non-Histone Protein HMG1 Revealed by the Nucleotide Sequence†

Koh-ichiroh Tsuda, Masahiro Kikuchi, Kazuyuki Mori, Shou Waga, and Michiteru Yoshida*

Laboratory of Biochemistry, Department of Agricultural Chemistry, Tohoku University, 1-1 Tsutsumidori-Amamiyamachi, Sendai 980, Japan

Received November 19, 1987; Revised Manuscript Received March 16, 1988

ABSTRACT: The isolation and sequencing of a cDNA clone coding for the entire sequence of pig thymus non-histone protein HMG1 are described. The sequence analysis reveals a complete 2192-nucleotide sequence with a 5′-terminal untranslated region of 11 nucleotides, 642 nucleotides of an open reading frame that encoded 214 amino acids, and a 3′-terminal untranslated region of 1539 nucleotides. The HMG1 protein, deduced from the nucleotide sequence, has a molecular weight of 24 785 and a C-terminal of a continuous run of 30 acidic amino acids, encoded by a simple repeating sequence of $(GAN)_{30}$. The predicted amino acid sequence is homologous to HMG1, HMG2, and HMG-T sequences from several sources, suggesting that the protein conformation is under evolutionary constraints. Northern blot analysis reveals that another hybridizable RNA species of smaller size is present. Southern blot analyses suggest that pig genome contains several HMG1 gene equivalents.

High mobility group (HMG)[1] proteins are a family of non-histone components in chromatin of relatively low molecular weights. Proteins complying with some or all of the criteria used for identification of HMGs are ubiquitously distributed in relative abundance among various organisms of eukaryotic kingdoms. HMG1 and HMG2 are a pair of proteins with a remarkable structural similarity and presumably share a common ancestral gene (Johns et al., 1975). In vitro, HMG1 and HMG2 proteins show a preferential binding to single-stranded DNA (Bidney & Reeck, 1978; Isackson et al., 1979; Yoshida & Shimura, 1984; Hamada & Bustin, 1985) and unwind double-stranded DNA structure (Yoshida & Shimura, 1984; Makiguchi et al., 1984; Javaherian et al., 1978, 1979). The high glutamic and aspartic acid regions in the C-terminal of the proteins are the active site in the DNA-unwinding reaction (Yoshida, 1987). The functions of HMG1 and HMG2 in vivo, however, have not yet been clarified, but there are some indications of their involvement in transcription or replication of chromatin (Einck & Bustin, 1985). It is very important to know the primary sequence and the structural characteristics of these proteins to understand their actual cellular roles. Walker et al. reported the primary amino acid sequences of HMG1 and HMG2 from calf thymus analyzed by the procedure of Edman degradation (Walker et al., 1980; Walker, 1982). The sequences, however, were ambiguous in having the undetermined sequences presumably because of the microheterogeneity of the proteins, in addition to the difficulty in determining the runs of similar amino acid by the chemical procedure. Recently, Pentecost and Dixon (1984) have reported the partial amino acid sequence of the C-terminal half of bovine testis HMG1 deduced from the nucleotide sequence of the cDNA missing the 5′-half of coding region. The trout testis HMG proteins, HMG-T, a second member of the protein, have homologous sequences to HMG1 and HMG2 (Dixon, 1982; Pentecost et al., 1985).

† This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

* To whom correspondence should be addressed.

[1] Abbreviations: HMG, high mobility group; bp, base pair(s); cDNA, complementary DNA; SDS, sodium dodecyl sulfate.

0006-2960/88/0427-6159$01.50/0

The present paper describes the molecular cloning of a cDNA coding for the non-histone protein HMG1 from pig thymus and the determination of the complete nucleotide sequence of the cDNA. We use this cDNA to probe the size of the mRNA and to examine the multiplicity of this gene in the pig genome.

Materials and Methods

RNA Preparation, in Vitro Protein Synthesis, and Analysis of Translation Products. Total RNA was extracted with guanidine hydrochloride as described by Harding et al. (1977). The RNA was fractionated into the poly (A+) and poly(A−) RNA by oligo(dT)–cellulose column chromatography as described by Aviv and Leder (1972). The RNAs were translated in vitro by using nuclease-treated wheat germ extract from Amersham. The translation product was immunoprecipitated with anti-HMG1 and anti-HMG2 antisera, followed by SDS–polyacrylamide gel electrophoresis (Laemmli, 1970), and by fluorography with sodium salicylate for the detection of [^{3}H]HMG proteins (Chamberlain, 1979).

Amino Acid Sequencing of N-Terminal Region of HMG1 and HMG2. HMG(1+2) was prepared from pig thymus and fractionated into HMG1 and HMG2 by column chromatography using Polybuffer exchanger PBE94 (Pharmacia; by unpublished results). The direct amino acid sequencing of the N-terminal region of both proteins was carried out by Beckman protein/peptide sequenator system 890ME.

Cloning and Sequencing of cDNA for HMG1 Protein. The complementary DNA library for the whole poly(A+) RNA was obtained by using a plasmid DNA vector which itself serves as the primer for the first- and ultimately second-strand cDNA synthesis developed by Okayama and Berg (1982). For the cloning of cDNAs for HMG1 and HMG2, the following two synthetic 17-mer oligonucleotide probes were prepared by the phosphodiester method on a solid support by Beckman Japan (Tokyo):

probe 1: 5′TCT(C)TCT(C)TCT(C)TCT(C)TCT(C)TC3′

probe 2: 5′GCCATA(G)TCT(C)TCA(G)AAT(C)TT3′

Initial screening of approximately 3×10^3 colonies by the colony hybridization with [5′-^{32}P]probe 1 gave 12 positives. Secondary screening was performed by the Southern blot hybridization with [5′-^{32}P]probe 2 on nitrocellulose filter to the inserts excised by *Xho*I digestion from the clones followed by agarose gel electrophoresis. The insert DNA of a positive clone containing the longer insert (pcD-MK02) was fragmented by several restriction enzymes to identify the convenient cloning sites for inserting various segments into bacteriophage M13 cloning vectors. Several overlapping segments of the insert were ligated by site-directed cloning into M13-mp10 and M13-mp11 replicative form DNA. After transformation into *Escherichia coli* JM101 or JM105 cells, single-stranded DNA was prepared (Messing, 1983) and sequenced by the dideoxy method (Sanger et al., 1977).

Northern and Southern Blot Hybridization. For RNA blot analysis, the poly(A+) RNA fraction was subjected to agarose gel electrophoresis, followed by transfer to a nitrocellulose filter (Thomas, 1983). For DNA blot analysis, high molecular weight DNA from pig thymus was digested with restriction enzymes, electrophoresed on a 0.7% agarose gel, and transferred to a GeneScreen Plus hybridization-transfer membrane (NEN Research Products) according to Southern (1975). 5′-^{32}P-Labeled pcD-MK02 insert or the restriction fragments of the insert, labeled by the multiprime labeling system of Amersham, were used as the probes for hybridization.

Results and Discussion

Isolation of the cDNA Clone for HMG1 Protein. For the initial investigation into whether HMG1 mRNA has a polyadenylate tail, the whole RNA from pig thymus was fractionated into the poly(A+) RNA and poly(A−) RNA fractions. Both the RNAs were expressed in an in vitro translation system for protein synthesis. The protein, translated from the poly(A+) RNA fraction followed by immunoprecipitation with anti-HMG1 antiserum, revealed the same mobility on SDS–polyacrylamide gel electrophoresis (data not shown). This result confirmed the previous one that the HMG1 mRNA has the polyadenylate tail (Bustin et al., 1981). Using the polyadenylate tail to anneal to vector–primer DNA for first-strand complementary DNA (cDNA) synthesis, a cDNA library for the whole polyadenylate RNA was obtained. The amino acid sequences of the N-terminal region of purified HMG1 and HMG2 from pig thymus were analyzed as an initial step in searching the suitable probes for screening cDNA clone for HMG1 from the library. The respective sequences of the first thirty amino acids determined were exactly identical with those of calf thymus (Walker et al., 1980; Walker, 1982), suggesting to us that the internal sequence of the pig thymus proteins has considerable homology to those of calf thymus. Two sets of synthetic probe of 17-mer oligonucleotide, therefore, were designed with reference to the amino acid sequences of calf thymus proteins. Probe 1 is a putative DNA complement of mRNA coding for the sequence Glu-Glu-Glu-Glu-Glu-Asp (or Glu) in the polyacidic amino acid region in the C-terminal (at residue positions 196–201 and 205–210 in the deduced amino acid sequence in Figure 2). Probe 2 is a putative DNA complement coding for the sequence Lys-Phe-Glu-Asp-Met-Ala in a one-third region from the N-terminal (at residue positions 58–63). Screening about 3000 clones from the cDNA library with probe 1 yielded 12 positive clones. The secondary screening with probe 2 gave two positive clones. A clone containing a larger insert, 2.2 kbp in size, pcD-MK02, was chosen for further analysis.

Nucleotide Sequence Analysis and Deduced Amino Acid Sequence. A restriction endonuclease map of pcD-MK02 together with the sequencing strategy is shown in Figure 1. The complete nucleotide sequence of 2192 nucleotides except the poly(A) sequence of the cDNA coding for pig thymus HMG1 is presented in Figure 2. The first ATG codon is located at nucleotides 9–11. The ATG codon may not be translated into the final protein moiety, because the N-terminal amino acid of HMG1 is glycine. The nucleotide sequence of an open reading frame of 642 nucleotides (nucleotides 12–653), to which two synthetic probes hybridized, terminated at a TAA codon (nucleotides 654–656). The 3′-untranslated region contains 1539 nucleotides (nucleotides 654–2192) except for poly(A) sequence. Two consensus polyadenylation signals, AATAAA, are present in the 3′-untranslated region at nucleotides 1285–1290 and 24 residues upstream (nucleotides 2164–2169) of the poly(A) tail. The deduced amino acid sequence from the open reading frame is also shown in Figure 2. The thirty amino acid sequence from the N-terminal is consistent with that of HMG1, but not with that of HMG2 determined beforehand by the chemical procedure. As was expected, the deduced sequence is highly homologous to the published tentative sequence of calf thymus HMG1 (Walker et al., 1980; Walker, 1982; also see Figure 3), but not to the total. The protein has the following values for mol of amino acid residues per mole: Asp, 19; Thr, 5; Ser, 11; Glu, 37; Asn, 4; Gln, 1; Pro, 13; Gly, 11; Ala, 19; Cys, 3; Val, 5; Met, 5; Ile, 4; Leu, 4; Tyr, 7; Phe, 9; His, 4; Lys, 43; Trp, 2; and Arg,

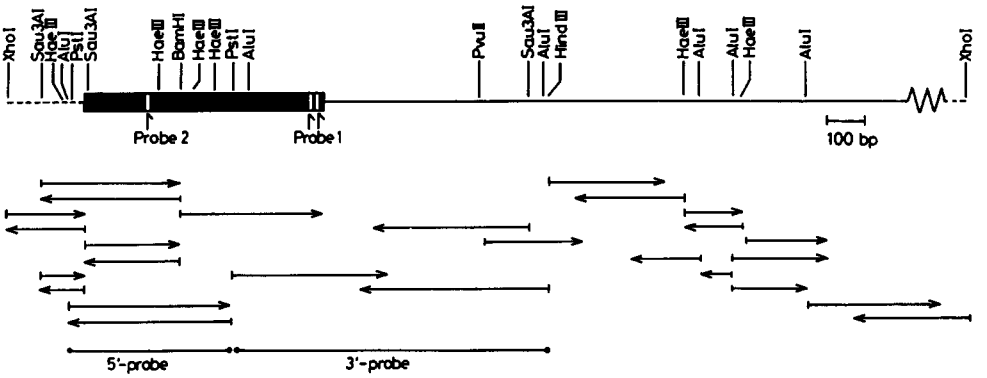

FIGURE 1: Restriction endonuclease map and sequencing strategy for the pig thymus HMG1 cDNA insert. The dotted lines at both ends of the insert (thin solid line) are the adjacent vector-primer DNA fragments excised by *Xho*I digestion. A zigzag line represents the poly(dA)/(dT) tail. The thick region indicates the position of the nucleotide region coding for HMG1. The open boxes in the thick region indicate the positions to which the synthetic probes hybridized. Only relevant restriction sites are shown. The horizontal arrows indicate the direction and extent of each sequence analysis. The bars at the bottom indicate the DNA fragments used as probes for Southern blot hybridization experiments in Figure 5.

```
                                                    1
                                CTGCAGGGGGGGGGGGGGGGGAATTAAACATG          11

GGCAAAGGAGATCCTAAGAAGCCGAGAGGCAAAATGTCATCATATGCATTCTTTGTGCAAACTTGCCGGGAGGAG   86
GlyLysGlyAspProLysLysProArgGlyLysMetSerSerTyrAlaPhePheValGlnThrCysArgGluGlu  (25)

CATAAGAAGAAGCACCCAGATGCTTCAGTCAACTTCTCAGAGTTTTCTAAGAAGTGCTCAGAAAGGTGGAAGACC  161
HisLysLysLysHisProAspAlaSerValAsnPheSerGluPheSerLysLysCysSerGluArgTrpLysThr  (50)

ATGTCTGCTAAAGAGAAAGGAAAATTTGAAGACATGGCAAAGGCGGACAAGGCCCGTTATGAAAGAGAAATGAAA  236
MetSerAlaLysGluLysGlyLysPheGluAspMetAlaLysAlaAspLysAlaArgTyrGluArgGluMetLys  (75)

ACTTACATACCTCCTAAAGGGGAGACAAAAAAGAAGTTCAAGGATCCCAATGCACCCAAGAGGCCTCCTTCGGCC  311
ThrTyrIleProProLysGlyGluThrLysLysLysPheLysAspProAsnAlaProLysArgProProSerAla  (100)

TTTTTCTTGTTTTGTTCTGAGTATCGTCCAAAAATCAAAGGAGAGCATCCTGGCCTATCCATTGGTGATGTTGCA  386
PhePheLeuPheCysSerGluTyrArgProLysIleLysGlyGluHisProGlyLeuSerIleGlyAspValAla  (125)

AAGAAACTGGGAGAGATGTGGAATAACACCGCTGCAGATGACAAGCACCCTTATGAAAAGAAGGCTGCTAAGCTG  461
LysLysLeuGlyGluMetTrpAsnAsnThrAlaAlaAspAspLysHisProTyrGluLysLysAlaAlaLysLeu  (150)

AAGGAGAAGTACGAAAAGGATATTGCTGCATACCGAGCTAAAGGGAAGCCTGATGCAGCAAAAAAGGGAGTCGTC  536
LysGluLysTyrGluLysAspIleAlaAlaTyrArgAlaLysGlyLysProAspAlaAlaLysLysGlyValVal  (175)

AAGGCTGAGAAAAGCAAGAAAAAGAAGGAAGAGGAAGATGAGGAAGATGAAGAGGATGAGGAGGAGGAGGAG  611
LysAlaGluLysSerLysLysLysGluGluGluGluAspGluGluAspGluGluAspGluGluGluGluGlu  (200)

GACGAAGAGGATGAGGAGGAAGAAGAAGATGATGATGATGAATAAGTTGGTTCTAGCGCAGTTTTTTTTTCTTGT  686
AspGluGluAspGluGluGluGluGluAspAspAspAspGlu                                 (214)

CTATAAAGCATTTAACCCCCCCTGTACACAACTCACTCCTTTTAAAGAAAAAAAATTGAAATGTAAGGCTGTGTA  761

AGATTTGTTTTTAAACTGTACAGTGTCTTTTTTTTGTATAGTTAACACACTACCGAATGTGCTTTAGATAGCCCT  836

GTCCTGGTGGTATTTTCAATAGCCACTAACCTTGCCTGGTACAGTATGGGGGTTGTAAATTGGCATGGAAATTTA  911

AAGCAGGTTCTTGTTGTGCACAGCACAAATTAGTTATATATGGGGTTGGTAGTTTTTTCATCTTCAGTTGTCTCT  986

GATGCAGCTTATACGAAATAATTGTTGTTCTGTTAACTGAATACCACTCTGTAATTGCAAAAAAAGAAAAAAAGT 1061

TGCAGCTGTTTTGTTGACATTCTGAATGCTTCTAAGTAAATACAATTTTTTTTATTAGTATTGTTGTCCTTTTCA 1136

TAGGTCTGAAATTTTTCTTCTTAAGGGGAGGCTAGTCTTTTGCTTTTGCCCATTTTGGATCACACGAATTATTAC 1211

AGTGTTTATCTTTCATATAGTTTAGCTGATAAAAAGCTTTTGTCTATACACCCTGCATACTCATGACGAGGGTTAA 1286

TAAAAGTTGGAATTGAGACAGTTTCATCCATAACTGAAACTCAAATCTTGATCGGTTGATAATCAGATTTCACCT 1361

AGCCCAGTTACATTTACAAAGTGAAGAGTAACATATCTACTCACAGCATGGGATTATTAGAATCAAACATTTTGA 1436

AAGTCTGTCCTTGAAGGACTAATAGAAAAGTATGTTCTAATCTTTACATGAGGACTCTACATTCTTTAACTCCCA 1511

TTACCATGTAATGGCAGTTATATTTGCAGTTCCCAAGTTAAAGACCTGAGAATGTATCCCCAAAAGCGTGAGATT 1586

TAAAATACAAGACGGGCCGTGTTATGTTTTTTTGTTGACATTAGTCCCAGCAAAGCTCTGGGGAAAAAAAAAGTG 1661

CTGGCTGTAAATGTCTTCTCCTGTTATCTAAATATGGATTGTTGAGGAAACTTGAGACCCACCATTAAAGCTATT 1736

TTTAATTAATTGGGCCAACTTTTAGAACTGTATACCACATTTTAAACCAGGTATATTTTCCTATATCATGGTTTG 1811

CCCCTGTATAAACCAAAATAGACATGAGGGAAGGAAGGACACTTAAACTTTGTATCTCAGTATGAATTATCTGAT 1886

TTATTTGAATTTGATTTTTCTTTACAAAAGTCAGCTCATTCATTAGGGTCATATGTTTATCTGCTTAACAGTGTA 1961

GGGAACAATTTGGCAATTTTGTGGTTTTCGAGATTATCGTTCTCTTAAAGTGCCAGTGTTTTAAAATAGCGTTCT 2036

TGTAATTTTACACGCTTTTGTGATGGAGTGCTGTTTTGTTATATAATTTTGACTTGGATTCTTTCCATTTGCATT 2111

TGTTTATGTAATTTCAGGAGGAATACTGAACATCTGAGTCCTGGATGATACTAATAAACTAATAATTTGCAGGAG 2186

GTTTTTA(n)                                                                  2193
```

FIGURE 2: Complete nucleotide sequence of the cDNA coding for pig thymus HMG1 and the deduced amino acid sequence. Nucleotides are numbered in the 5′ to 3′ direction, beginning with the first nucleotide of the cDNA insert preceded by the oligo(dG)/(dC) linker. The deduced amino acid residues are numbered in parentheses beginning with the N-terminal amino acid (Gly) next to the ATG initiation codon (underlined). The termination codon, TAA, is also underlined. Two perfectly conserved putative polyadenylation signals, AATAAA, are boxed. The doubly underlined sequence of residues 185–214 is the polyacidic amino acid region.

8. The molecular weight of the protein of 24 785, calculated from the deduced amino acid composition, is similar to that obtained by SDS–polyacrylamide gel electrophoresis. The acidic tail is a run of 30 amino acids total, consisting of 21 glutamic acid and 9 aspartic acid residues. The usage of Glu codons in the region is almost equally divided between 12 GAG and 9 GAA codons, while Asp codon usage is definitely nonrandom since GAT (8/9) is dominant. Surprisingly, this domain of the protein is coded for by a simple repeating sequence, $(GAN)_{30}$.

Structural Homology to Other HMG Proteins. The deduced amino acid sequence of pig thymus HMG1 was compared to the tentative sequences of HMG1 and HMG2 from calf thymus (Walker et al., 1980; Walker, 1982), to the partial sequence of bovine testis HMG1 (Pentecost & Dixon, 1984), and to the sequences of HMG-T (Dixon, 1982; Pentecost et al. 1985) as shown in Figure 3. The sequence of bovine testis HMG1 is different in five amino acid residues from the corresponding region of pig thymus HMG1. Except for the sequences not comparable to each other, only three amino acid residues differed by one base substitution in the respective amino acid codon between pig and calf thymus HMG1s. The degree of homology to calf thymus HMG2 is not extremely high as in calf thymus HMG1. In addition, the sequences of HMG-T and a member of the family of HMG-T from trout testis are somewhat homologous to the HMG1. The high conservation of these protein sequences may suggest that there are evolutionary constraints on the conformation of the proteins.

The partial sequence homology of the acidic tail of pig thymus HMG1 to the acidic region in the sequence of DNA-destabilizing protein from bacteriophage T7 (Dunn & Studier, 1981) is noticeable. The acidic tail of pig thymus HMG1 is an active region in the DNA-unwinding reaction (Yoshida, 1987), while the unwinding active site of the DNA-destabilizing protein from bacteriophage T7 has not been clarified. This sequence homology, however, suggests a possibility that these runs of acidic amino acid residues are fundamental to exhibit such a DNA-unwinding activity.

The deduced HMG1 sequence contains two sequences (residues 5–11 and 180–186) considerably homologous to the putative nuclear localization signals (Kalderon et al., 1984; Smith et al., 1985; Richardson et al., 1986). The sequence between residues 110 and 126 seems to be an alignment to maintain the helix–turn–helix structure contained in general DNA binding proteins (Pabo & Sauer, 1984).

Analysis of mRNA for HMG1. The size of mRNA coding for HMG1 was estimated by Northern blot hybridization of whole poly(A+) RNA. When synthetic probe 1 was used at the initial step of analysis, two bands of hybridization of sizes 1.2K and 2.3K were obtained (data not shown). Similar bands were observed when the pcD-MK02 insert was used as another probe, as shown in Figure 4. These results suggest that there are two species of HMG1 mRNA encoded by independent

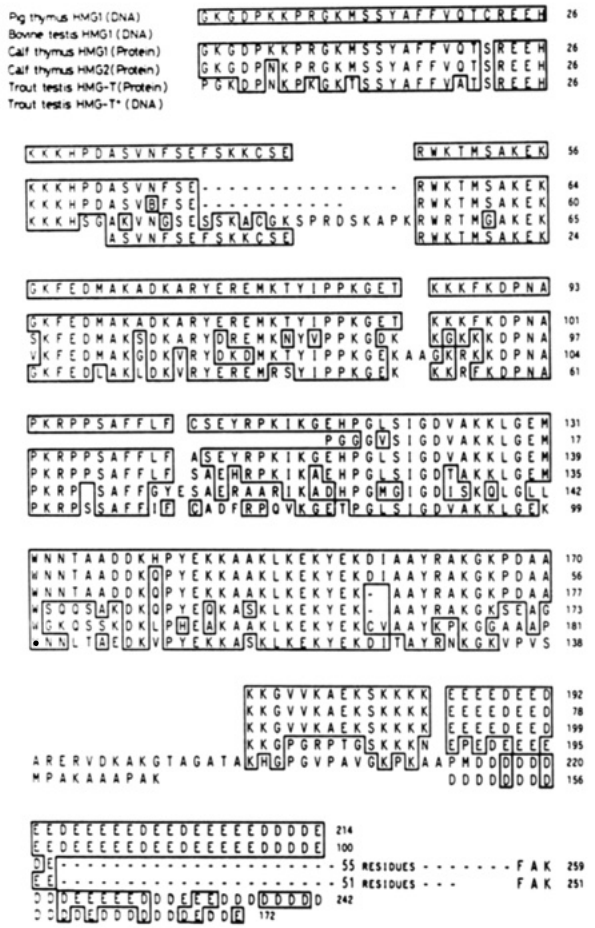

FIGURE 3: Comparison of the deduced amino acid sequence of pig thymus HMG1 with those of several HMG proteins. Sequences are aligned to show the greatest amino acid homology. Identical amino acids are boxed. "(DNA)" marked after the HMG source indicates sequence deduced from the cDNA nucleotide sequence; "(protein)" indicates sequence determined by the chemical procedures for the protein. Sequence references: bovine testis HMG1 (Pentecost & Dixon, 1984); calf thymus HMG1 and HMG2 (Walker et al., 1980; Walker, 1982); trout testis HMG-T (Dixon, 1982); and a member of the family of HMG-T* (Pentecost et al., 1985).

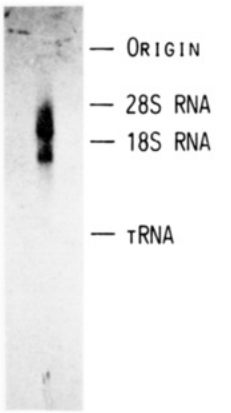

FIGURE 4: Northern blot analysis of pig thymus RNA. Fifteen micrograms of poly(A+) RNA from pig thymus was fractionated on 1.1% agarose gel, blotted on a nitrocellulose membrane, and probed with [^{32}P]pcD-MK02 insert. The migrations of size marker RNAs are indicated in the right column.

genes, and that the insert in pcD-MK02 is complementary to the longer mRNA molecule. Some other possibilities for the smaller RNA species, however, are considered. The cDNA contained two perfectly conserved putative polyadenylation signals, AATAAA, in the 3′-untranslated region (Figure 2). The primary transcript polyadenylated at the signal of nucleotides 1285–1290 may give a band of size 1.2K, while the product polyadenylated at the signal of nucleotides 2164–2169 could be the longer RNA species. Second, the smaller RNA molecule may be an mRNA coding for other members of HMG1, because HMG1 protein comprises some microheterogeneous molecules (Nicolas & Goodwin, 1982). Third, the smaller RNA may be an mRNA coding for HMG2, because HMG2 is a homologous protein to HMG1 (see Figure 3). The cloning and sequencing studies of cDNA for the smaller RNA may elucidate the character of the RNA species.

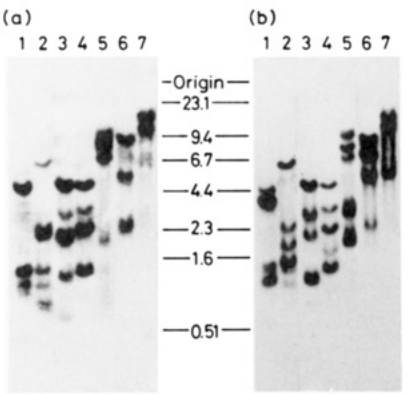

FIGURE 5: Southern blot analysis of pig genomic DNA. High molecular weight DNA from pig thymus was digested with *Pst*I (lane 1), *Bam*HI (lane 2), *Msp*I (lane 3), *Pvu*II (lane 4), *Hin*dIII (lane 5), *Eco*RI (lane 6), and *Eco*RV (lane 7), analyzed on an agarose gel, and transferred to GeneScreen Plus hybridization-transfer membrane. The resulting hybridization was probed with (a) 5′-probe (*Pst*I–*Pst*I fragment of 438 bp in Figure 1) and (b) 3′-probe (*Pst*I–*Hin*dIII fragment of 827 bp). DNA markers are indicated in the middle of the figure in kilobase pairs.

Southern Blot Analysis. Southern blot analyses of genomic DNA were examined so as to identify possible multiplicity in the HMG1 gene. High molecular weight DNA was digested with various restriction enzymes that have a single or no recognition site in the HMG1 cDNA insert. Initially, the Southern blots were probed with the whole cDNA insert. The results (not shown) indicated that each digest produced a complex mixture of restriction fragments that hybridized with the cDNA insert, suggesting that the genomic DNA contained more than one hybridizable sequence. In order to determine the DNA part to give the multiplicity of the hybridizing fragments, a DNA fragment that contains the 5′-half of the coding region for the initial hybridization (5′-probe; *Pst*I–*Pst*I fragment in Figure 1) and another fragment that contains the 3′-half of the coding region followed by a part of the 3′-untranslated sequence for the rehybridization (3′-probe; *Pst*I–*Hin*dIII fragment) were used as probes. The results presented in Figure 5 indicate that each restriction enzyme produced several fragments hybridizable with 5′-probe as well as 3′-probe. *Pst*I, which has a single recognition site in the coding sequence, produced four distinct fragments, (4.5, 1.3, 1.0, and 0.9 kbp) with 5′-probe and four distinct fragments (4.4, 3.6, 1.4, and 1.1 kbp) with 3′-probe. *Bam*HI, which has a single cleavage site in the coding sequence, produced six distinct fragments (6.2, 2.4, 2.1, 1.3, 1.0, and 0.75 kbp) with 5′-probe and seven distinct fragments (6.2, 5.1, 2.3, 1.9, 1.5, 1.3, and 1.0 kbp) with 3′-probe. The largest fragment (6.2 kbp) was common to both the probes. *Pvu*II and *Hin*dIII, which have a single recognition site in the 3′-untranslated region, produced respectively four common bands with both the probes and an extra band with 3′-probe. *Msp*I, which has no recognition site in the insert, produced four common distinct fragments with both the probes. *Eco*RI and *Eco*RV, which have no recognition site in the insert, produced three and two common

distinct fragments with both the probes and an extra restriction fragment with 3′-probe, respectively. The multiplicity of the fragments does not arise from partial digestion since the band pattern was reproducible. In addition, both 5′-probe and 3′-probe do not contain any repetitive sequence which could show multiple hybridization bands (Figure 2). These patterns, therefore, could be not most readily explained by the presence of multiple introns in the HMG1 gene, suggesting that the pig genome contains several types of sequences homologous with the HMG1 gene. It is not clear whether all the sequences can be transcribed to yield the two species of mRNA described above, or whether some fragments are generated from cross-hybridizing pseudogenes or from HMG2 gene, which may be expected to have homologous sequences to those of HMG1.

As stated in the introduction, HMG1 is the most abundant, widely distributed protein associated with chromatin. So far as we know, the complete primary structure of the protein has not been reported. The present study primarily showed the nucleotide sequence of cDNA coding for HMG1 and the entire amino acid sequence deduced from it. The availability of cDNA sequence for HMG1 will allows studies on the gene structure and transcriptional regulation of this chromosomal protein. In addition, the amino acid sequence of HMG1 will give much information in understanding the mechanisms of the in vitro DNA-unwinding reaction, as well as the cellular roles of this major non-histone protein.

ACKNOWLEDGMENTS

We thank Y. Hirota (Beckman Japan, Tokyo) for his synthesis of oligonucleotides and determination of amino acid sequences. We also thank Dr. S. Mizuno for his continuous interest and valuable comments, H. Ono for his technical advice, and Dr. H. Nagayama for his critical reading of the manuscript.

Registry No. HMG1 protein, 115203-95-7; HMG1 cDNA, 115203-94-6; HMG1 precursor protein, 115203-93-5.

REFERENCES

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Bidney, D. L., & Reeck, G. R. (1978) *Biochem. Biophys. Res. Commun. 85*, 1211–1218.

Bustin, M., Neihart, N. K., & Fagan, J. B. (1981) *Biochem. Biophys. Res. Commun. 101*, 893–897.

Chamberlain, J. P. (1979) *Anal. Biochem. 98*, 132–135.

Dixon, G. H. (1982) in *The HMG Chromosomal Proteins* (Johns, E. W., Ed.) pp 149–182, Academic, London.

Dunn, J. J., & Studier, F. W. (1981) *J. Mol. Biol. 148*, 303–330.

Einck, L., & Bustin, M. (1985) *Exp. Cell Res. 156*, 295–310.

Hamada, H., & Bustin, M. (1985) *Biochemistry 24*, 1428–1433.

Harding, J. D., MacDonald, R. J., Przybyla, A. E., Chirgwin, J. M., Pictet, R. L., & Rutter, W. J. (1977) *J. Biol. Chem. 252*, 7391–7397.

Isackson, P. J., Fishback, J. L., Bidney, D. L., & Reeck, G. R. (1979) *J. Biol. Chem. 254*, 5569–5572.

Javaherian, K., Liu, L. F., & Wang, J. C. (1978) *Science (Washington, D.C.) 199*, 1345–1346.

Javaherian, K., Sadeghi, M., & Liu, L. F. (1979) *Nucleic Acids Res. 6*, 3569–3580.

Johns, E. W., Goodwin, G. H., Walker, J. M., & Sanders, C. (1975) *Ciba Found. Symp. 29*, 95–112.

Kalderon, D., Roberts, B. L., Richardson, W. D., & Smith, A. E. (1984) *Cell (Cambridge, Mass.) 39*, 499–509.

Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.

Makiguchi, K., Chida, Y., Yoshida, M., & Shimura, K. (1984) *J. Biochem. (Tokyo) 95*, 423–429.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Nicolas, R. H., & Goodwin, G. H. (1982) in *The HMG Chromosomal Proteins* (Johns, E. W., Ed.) pp 41–68, Academic, London.

Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol. 2*, 161–170.

Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.

Pentecost, B., & Dixon, G. H. (1984) *Biosci. Rep. 4*, 49–57.

Pentecost, B. T., Wright, J. M. & Dixon, G. H. (1985) *Nucleic Acid Res. 13*, 4871–4888.

Richardson, W. D., Roberts, B. L., & Smith, A. E. (1986) *Cell (Cambridge, Mass.) 44*, 77–85.

Sanger, F., Nicklen, S., & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Smith, A. E., Kalderon, D., Roberts, B. L., Colledge, W. H., Edge, M., Gillett, P., Markham, A., Paucha, E., & Richardson, W. D. (1985) *Proc. R. Soc. London, B 226*, 43–58.

Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.

Thomas, P. S. (1983) *Methods Enzymol. 100*, 255–266.

Walker, J. M. (1982) in *The HMG Chromosomal Proteins* (Johns, E. W. Ed.) pp 69–87, Academic, London.

Walker, J. M., Gooderham, K., Hasting, J. R. B., Mayes, E., & Johns, E. W. (1980) *FEBS Lett. 122*, 264–270.

Yoshida, M. (1987) *J. Biochem. (Tokyo) 101*, 175–180.

Yoshida, M., & Shimura, K. (1984) *J. Biochem. (Tokyo) 95*, 117–124.