Lin, S. H., Konishi, Y., Denton, M. E., & Scheraga, H. A. (1984) *Biochemistry 23*, 5504–5512.
Marks, C. B., Vasser, M., Ng, P., Henzel, W., & Anderson, S. (1986) *J. Biol. Chem. 261*, 7115–7118.
Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., & Anderson, S. (1987a) *Science 235*, 1370–1373.
Marks, C. B., Naderi, H., Kosen, P. A., Kuntz, I. D., & Anderson, S. (1987b) in *Protein Structure, Folding and Design* (Oxender, D. L., Ed.) Vol. 2, pp 335–340, Liss, New York.
Matouschek, A., Kellis, J. T., Jr., Serrano, L., & Fersht, A. R. (1989) *Nature 340*, 122–126.
Matsumura, M., Becktel, W. J., & Matthews, B. W. (1988) *Nature 334*, 406–410.
Moses, E., & Hinz, H.-J. (1983) *J. Mol. Biol. 170*, 765–776.
Motulsky, H. J., & Ransnas, L. A. (1987) *FASEB J. 1*, 365–374.
Nall, B. T. (1985) *Comments Mol. Cell. Biophys. 3*, 123–143.
Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 3*, 1–43.
Pace, C. N., Grimsley, G. R., Thomson, J. A., & Barnett, B. J. (1988) *J. Biol. Chem. 263*, 11820–11825.
Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J.-T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., & Kraut, J. (1987) *Biochemistry 26*, 2674–2682.
Ramdas, L., & Nall, B. T. (1986) *Biochemistry 25*, 6959–6964.
Richardson, J. A. (1981) *Adv. Protein Chem. 34*, 167–339.
Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700–704.
Schellman, J. A. (1978) *Biopolymers 17*, 1305–1322.
Schwarz, H., Hinz, H.-J., Mehlich, A., Tschesche, H., & Wenzel, H. R. (1987) *Biochemistry 26*, 3544–3551.
Stackhouse, T. M., Onuffer, J. J., Matthews, C. R., Ahmed, S. A., & Miles, E. W. (1988) *Biochemistry 27*, 824–832.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121–282.
Tanford, C. (1970) *Adv. Protein Chem. 24*, 1–95.
Tüchsen, E., & Woodward, C. (1985) *J. Mol. Biol. 185*, 405–419.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694–699.
Utiyama, H., & Baldwin, R. L. (1986) *Methods Enzymol. 131*, 51–71.
Vincent, J.-P., Chicheportiche, R., & Lazdunski, M. (1971) *Eur. J. Biochem. 23*, 401–411.
Wells, J. A., & Estell, D. A. (1988) *Trends Biochem. Sci. 13*, 291–297.
White, T. B., Berget, P. B., & Nall, B. T. (1987) *Biochemistry 26*, 4358–4366.
Wlodawer, A., Deisenhofer, J., & Huber, R. (1987) *J. Mol. Biol. 193*, 145–156.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol.* Biol. *155*, 311–319.
Yutani, K., Ogasahara, K., Tsujita, T., & Sugino, Y. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4441–4444.

Primary Structure of Non-Histone Chromosomal Protein HMG2 Revealed by the Nucleotide Sequence†

Hitoshi Shirakawa,‡ Koh-ichiroh Tsuda,§ and Michiteru Yoshida*,‡

Department of Biological Science and Technology, Science University of Tokyo, Yamazaki, Noda-shi, Chiba 278, Japan, and Department of Agricultural Chemistry, Tohoku University, 1-1 Tsutsumidori-Amamiyamachi, Sendai 980, Japan

Received June 22, 1989; Revised Manuscript Received September 13, 1989

ABSTRACT: The isolation and sequencing of a cDNA clone for the entire sequence of pig thymus non-histone protein HMG2 are described. cDNA the size of 1153 nucleotides contains an open reading frame of 627 nucleotides. The 5′-untranslated region of 146 nucleotides is extremely rich in GC residues whereas the 3′-untranslated region of 380 nucleotides is rich in AT residues. The open reading frame encodes 209 amino acids, which contain a unique continuous run of 23 acidic amino acids at the C-terminal. The deduced amino acid sequence is 79% homologous to that of HMG1 protein from the same source which we reported [Tsuda, K., Kikuchi, M., Mori, K., Waga, S., & Yoshida, M. (1988) *Biochemistry 27*, 6159–6163]. In addition, the hydropathy index profiles of both proteins are very similar, supporting that they have similar structural features. Northern analysis of poly(A+) RNA reveals that a single-sized mRNA codes for HMG2 protein. Southern analysis suggests that the HMG2 coding gene is homogeneous within the pig thymus genome.

The relatively low molecular weight proteins in eukaryotic non-histone nuclear proteins are called high mobility group (HMG)[1] proteins. HMG1 and HMG2 in four major HMG proteins have remarkable structural and functional similarities to each other. These proteins show a preferential binding to single-stranded DNA (Bidney & Reeck, 1978; Isackson et al., 1979; Yoshida & Shimura, 1984; Hamada & Bustin, 1985), unwind double-stranded DNA structure (Yoshida & Shimura, 1984; Makiguchi et al., 1984; Javaherian et al., 1978, 1979), and remove the transcriptional block caused by left-handed

† The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02895.

* Correspondence should be addressed to this author.
‡ Science University of Tokyo.
§ Tohoku University.

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

0006-2960/90/0429-4419$02.50/0

Z-form DNA (Waga et al., 1988). In addition, these proteins may effect in vitro nucleosome assembly (Bonne-Andrea et al., 1984; Waga et al., 1989). However, the functions and roles of HMG1 and HMG2 in vivo are still controversial. Walker et al. showed 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 incomplete in having the undetermined sequences presumably because of the difficulty in determining the runs of similar amino acids by the chemical procedure, in addition to microheterogeneity of the proteins. Recently, we reported the molecular cloning of a cDNA coding for HMG1 from pig thymus and the entire amino acid sequence of the protein deduced from the nucleotide sequence of the cDNA. So far as we know, the complete primary structure of HMG2 protein from any eukaryotic source has not been elucidated.

The present paper describes the cloning of a cDNA coding for non-histone protein HMG2 from pig thymus, the nucleotide sequence of the cDNA, and the deduced amino acid sequence. The primary structure of HMG2 was compared with that of HMG1 from the same material (pig thymus). We use the HMG2 cDNA to analyze the size and multiplicity of the transcript, and to determine the homogeneity of this gene in the pig thymus genome.

Materials and Methods

Cloning and Sequencing of the cDNA for HMG2 Protein. The cloning of cDNA for HMG2 was performed at the same time with that for HMG1 as described previously (Tsuda et al., 1988). Shortly, the complementary DNA library for the whole poly(A+) RNA was obtained by using a plasmid developed by Okayama and Berg (1982). Two synthetic 17-mer oligonucleotides were used as probes for the cloning of cDNAs for HMG2 and HMG1. Two positive clones to both probes were obtained in the screening of approximately 3×10^3 colonies. One positive clone (pcD-MK02) contained the entire nucleotide sequence coding for HMG1 (Tsuda et al., 1988), and another clone (pcD-MK10) was submitted to the present experiments. The inert DNA fragment excised with *Xho*I or *Bam*HI from the positive clone, pcD-MK10, was digested with several restriction endonucleases to identify the cloning sites for recombination into pUC strain cloning vectors. Several overlapping segments of the insert were ligated into multicloning sites of plasmids pUC118 and pUC119 DNA prior to transformation into *Escherichia coli* MV1184. Single-stranded DNA was prepared according to Vieira and Messing (1987) and sequenced by the dideoxy method (Sanger et al., 1977).

Southern and Northern Hybridization. For DNA blot analysis, high molecular weight genomic 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). For RNA blot analysis, the poly(A+) RNA fraction was electrophoresed on a 1.1% agarose gel, followed by transfer to a nitrocellulose filter (Thomas, 1983). The full-length pcD-MK10 insert or the restriction fragments of the insert, ^{32}P-labeled with the random primer DNA labeling system of Takara (Kyoto), were used as the probes for hybridization.

Results and Discussion

Nucleotide Sequence Analysis of the cDNA Clone for HMG2 Protein and the Deduced Amino Acid Sequence. As described previously (Tsuda et al., 1988), a cDNA library for the poly(A+) RNA fraction from pig thymus was obtained, and a positive clone to two sets of probe of 17-mer oligonucleotide, pcD-MK10, was analyzed for the nucleotide sequence. A restriction endonuclease map and the sequencing strategy for the pcD-MK10 insert are shown in Figure 1. The complete nucleotide sequence of 1153 nucleotides except for the poly(A) sequence is shown in Figure 2. After a 5′-noncoding sequence of 143 nucleotides, the first ATG codon is located at nucleotides 144–146. This ATG codon may not be translated into the final protein moiety as in the case of HMG1 (Tsuda et al., 1988), because the N-terminal amino acid of HMG2 is glycine from the result of direct amino acid se-

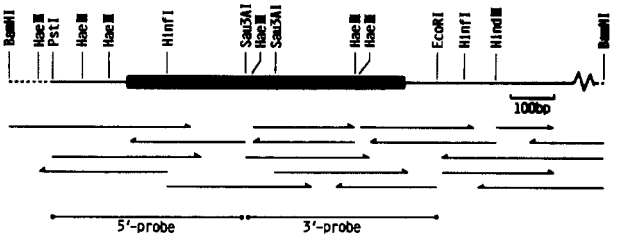

FIGURE 1: Restriction endonuclease map and sequencing strategy for the pig thymus HMG2 cDNA insert. The dotted lines at both ends of the insert (thin solid line) are the adjacent vector-primer DNA fragments excised by *Bam*HI digestion. A zig-zag line represents the poly(dA)/poly(dT) tail. The thick region indicates the position of the nucleotide region coding for HMG2. Only relevent 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 the Southern blot hybridization experiments in Figure 7.

```
                                  1
                CTGCAGGGGGGGGGGGGGGGACAAACCAGTTCACGCCGGAGCCGCA   26
GGTAGGCAGCGTCGCGGTCGGACCCGGCCGCTGCGCGGGAGCCTGAGGAGAAGCTACCAC   86
CAGGCAAGAGACCCCTCCGGGCCCCGGTGGACGCGTCGTCGCCGGCGCCGCAGCACCATG  146
GGGAAGGGCGACCCCAACAAGCCGCGGGGCAAGATGTCCTCGTACGCCTTCTTCGTGCAG  206
GlyLysGlyAspProAsnLysProArgGlyLysMetSerSerTyrAlaPhePheValGln   (20)
ACCTGCCGGGAGGAGCACAAGAAGAAACACCCCGATTCCTCGGTCAACTTCGCCGAGTTC  266
ThrCysArgGluGluHisLysLysLysHisProAspSerSerValAsnPheAlaGluPhe   (40)
TCCAAGAAGTGCTCCGAGCGATGGAAGACTATGTCTGCCAAGGAAAAGTCCAAGTTTGAA  326
SerLysLysCysSerGluArgTrpLysThrMetSerAlaLysGluLysSerLysPheGlu   (60)
GATATGGCAAAAAGTGACAAAGCTCGCTATGACCGGGAGATGAAAAATTACGTCCCTCCC  386
AspMetAlaLysSerAspLysAlaArgTyrAspArgGluMetLysAsnTyrValProPro   (80)
AAGGGTGACAAGAAGGGCAAGAAAAAAGATCCCAATGCTCCCAAAAGGCCTCCATCTGCC  446
LysGlyAspLysLysGlyLysLysLysAspProAsnAlaProLysArgProProSerAla  (100)
TTCTTCCTGTTTTGCTCTGAACATCGCCCAAAGATCAAAAGTGAACACCCTGGCTTATCC  506
PhePheLeuPheCysSerGluHisArgProLysIleLysSerGluHisProGlyLeuSer  (120)
ATTGGGGATACTGCAAAGAAATTGGGTGAAATGTGGTCTGAGCAGTCAGCCAAAGATAAA  566
IleGlyAspThrAlaLysLysLeuGlyGluMetTrpSerGluGlnSerAlaLysAspLys  (140)
CAACCGTATGAACAGAAAGCAGCTAAGCTAAAGGAAAAATATGAAAAGGATATTGCTGCA  626
GlnProTyrGluGlnLysAlaAlaLysLeuLysGluLysTyrGluLysAspIleAlaAla  (160)
TACCGTGCCAAGGGTAAGGGTGAAGCAGGAAAGAAGGGCCCTGGCAGGCCAACAGGCTCT  686
TyrArgAlaLysGlyLysGlyGluAlaGlyLysLysGlyProGlyArgProThrGlySer  (180)
AAGAAGAAGAATGAACCAGAAGATGAGGAGGAAGAGGAAGAGGAGGAAGAAGATGAAGAC  746
LysLysLysAsnGluProGluAspGluGluGluGluGluGluGluGluAspGluAsp     (200)
GAGGAGGAAGAAGATGAAGATGAGGAATAAATGGCTATCCTGTAATGTTGTGTGTGGAGT  806
GluGluGluGluAspGluAspGluGlu                                   (209)
GTGTGTGTGTGCTCAGGCAATTATTTTGCTAAGAATGTGAATTCAAGTGCAGCTCAATAT  866
TAGCTTCAGTATAAAAACTGTACAGATTTTTGTATAGCTAATAAGATTCTTTGTAGAGAA  926
AATACTTTTTTTAAAAGTGCAGGTTGCAGCTTTTTGAGGGGCTACTACATACAGTTAGAT  986
TTTAAAGCTTCTGATGTTGAATGTTTCTAAATATTTAATGGTTTCTTTAATTTCTTGTGT 1046
ATGGTAACAGCAAACTCGTAGAAATAAGTATCAATAGCAAATTTTGGGTTTTCTAGAATG 1106
TTGCATTTTGTTTTTTAAAAAAATTTTGTAATAAAATTATGTATATTA(n)           1154
```

FIGURE 2: Complete nucleotide sequence of the cDNA coding for pig thymus HMG2 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)/oligo(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. Perfectly conserved putative polyadenylation signals, AATAAA, are boxed.

quencing of the protein. The nucleotide sequence of an open reading frame of 627 nucleotides (nucleotides 147–773) terminated at a TAA codon (nucleotides 774–776). The deduced amino acid sequence from the open reading frame is also shown in Figure 2. We previously analyzed the amino acid sequence of the N-terminal region of the isolated HMG2 from pig thymus by the chemical procedure (data not shown). The deduced amino acid sequence from the cDNA was consistent with that of pig thymus HMG2 determined by the chemical procedure for the first 32 amino acid residues obtained, but not with that of pig thymus HMG1. In addition to this, the deduced sequence is not homologous with that of pig thymus HMG1 in 43 amino acid residues, as will be mentioned later (see Figure 4), but has higher homology to the calf thymus HMG2 sequence reported by Walker (1982): only 10 amino acid residues are not homologous in 194 residues within comparable sequences. These results support that the cDNA codes for the pig thymus HMG2 protein. The unique carboxyl terminal of HMG2 protein is shown to be a continuous run of 23 acidic amino acids, consisting of 18 glutamic acid and 5 aspartic acid residues. The protein has the following values for moles of amino acid residues per mole: Asp, 15; Thr, 4; Ser, 16; Glu, 34; Asn, 5; Gln, 4; Pro, 15; Gly, 14; Ala, 15; Cys, 3; Val, 3; Met, 5; Ile, 3; Leu, 4; Tyr, 6; Phe, 8; His, 4; Lys, 40; Trp, 2; Arg, 9. The molecular weight of the protein of 24 000, calculated from the deduced amino acid composition, is similar to that obtained by SDS–polyacrylamide gel electrophoresis (data not shown). The 3′-untranslated region, containing 380 nucleotides (nucleotides 774–1153) except for the poly(A) sequence, is not as long as that of HMG1. A consensus polyadenylation signal, AATAAA, is present in the 3′-untranslated region at 18 nucleotide residues upstream (nucleotides 1136–1141) of the poly(A) tail. Forty-eight percent of the nucleotides in the open reading frame coding for HMG2 are either G or C. The 5′-untranslated region is extremely rich in GC residues; 73% of 146 nucleotides are either G or C. In contrast, the 3′-untranslated region is highly enriched in A and T residues, comprising 71% of A and T in total. A similar GC-rich composition in the 5′-untranslated region and AT-rich composition in the 3′-untranslated region was also observed in cDNAs for HMG1 (Tsuda et al., 1988), HMG14, and HMG17 (Landsman & Bustin, 1986).

Comparison of the Nucleotide Sequences of cDNA Coding for HMG2 and HMG1 Proteins. The preliminary structural analysis has indicated the primary sequences of HMG1 and HMG2 protein are relatively homologous (Walker, 1982). This suggested to us that both nucleotide sequences of the cDNA may also show a homology. The nucleotide sequence of 627 nucleotides within the open reading frame in cDNA coding for HMG2 is presented in Figure 3 along with that of the 642-nucleotide sequence within HMG1 cDNA from pig thymus which we determined previously (Tsuda et al., 1988). A computer analysis of the two sequences for best alignment reveals that there are 465 matched bases (i.e., 74% of the sequence is identical). However, the homology within the 3′-untranslated regions of both the proteins is not high (45% of the nucleotide are identical).

Comparison of Structural Characteristics of HMG2 with HMG1 Protein. The deduced amino acid sequence of HMG2 is presented in Figure 4 in order to compare with that of HMG1. Between both proteins, the primary sequence is relatively conserved; the HMG2 sequence shows 79% homology with the HMG1 sequence (166/209 residues). Twenty-one amino acid residues in 43 nonhomologous residues may have resulted from a single point mutation of the code position.

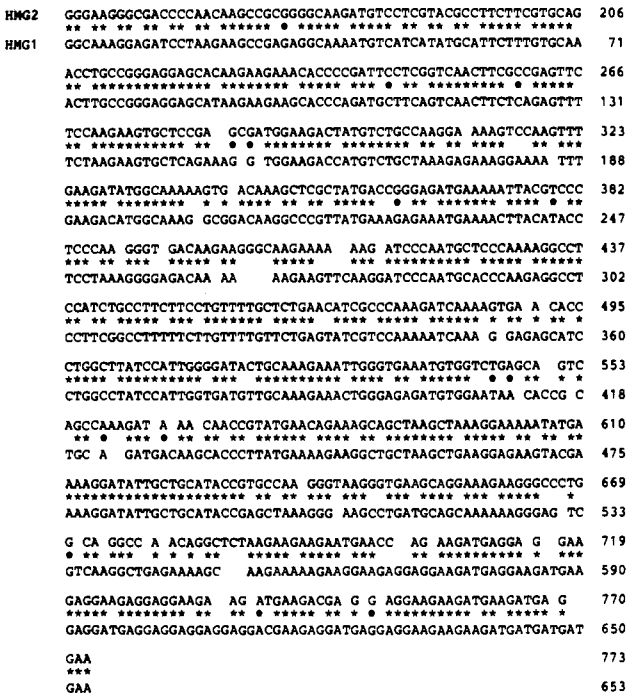

FIGURE 3: Comparison of the nucleotide sequence within the open reading frame in HMG2 cDNA with that in HMG1 cDNA. With the use of DNASIS software (Hitachi Software Engineering), the sequences were aligned to show greatest nucleotide homology with the first nucleotides within the open reading frames at the heads. The numbers at the right column were those of nucleotides in each cDNA. Identical bases are denoted by asterisks.

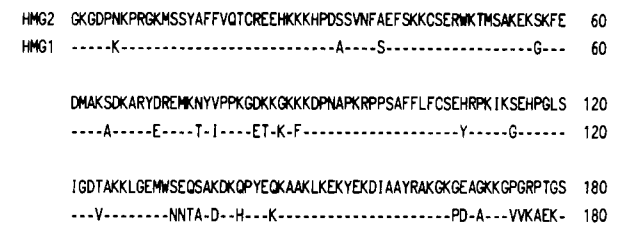

FIGURE 4: Comparison of the deduced amino acid sequence of pig thymus HMG2 with that of HMG1. Sequences are aligned directly from the N-terminal amino acid, glycine, determined by protein sequencing. The dashed lines represent identical residues.

The sequences of the N-terminus, corresponding to one-third of the regions in both the proteins, exhibit a high homology. All the changes in acidic amino acids in the C-terminus are the result of one base substitution at the third codon position. When the structural feature of HMG2 was compared with that of HMG1 by analyzing their hydropathy index profiles which were determined by the method of Kyte and Doolittle (1982), as shown in Figure 5, the profiles were significantly homologous. Reeck et al. (1982) suggested that calf thymus HMG1 consists of a three-domain structure; domains A and B of compact and globular conformations and domain C of highly acidic property. The hydropathy index profiles in Figure 5 and the predicted secondary structure data of pig thymus HMG2 (not shown) suggest that HMG2 protein has a similar domain structure to HMG1 and that domains A and B in HMG1 and HMG2 have very similar structural features to one another. The primary sequences of HMG1 protein are highly conservative, suggesting that there are evolutionary constraints on the conformation of the proteins (Tsuda et al., 1988). This may be equally true of HMG2 protein, because

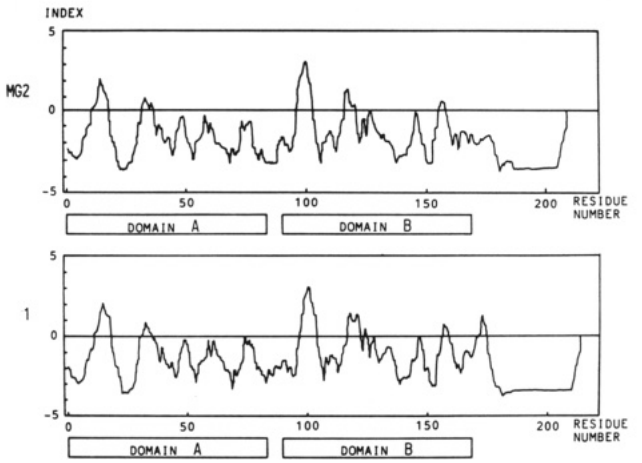

FIGURE 5: Hydropathy index profiles of pig thymus HMG2 and HMG1 proteins deduced from their cDNA sequences. Hydropathy indexes were determined by the method of Kyte and Doolittle (1982). Domains A and B were boxed by reference to Reeck et al. (1982).

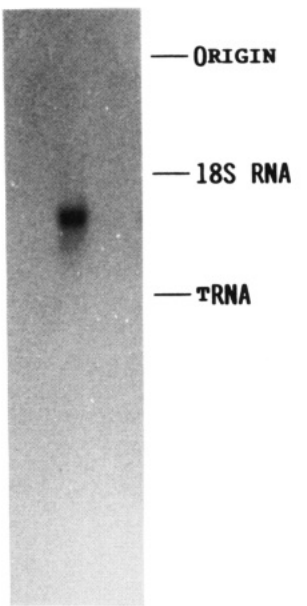

FIGURE 6: Northern blot analysis of pig thymus RNA. Fifteen micrograms of poly(A+) RNA from pig thymus was fractionated on 1.1% agarose gel, blotted on a nitrocellulose membrane, and probed with ^{32}P-pcD-MK10 insert. The migrations of size marker RNAs are indicated in the right column.

the HMG2 sequence is homologous to that of HMG1.

Northern Analysis of mRNA for HMG2 Protein. The size of mRNA coding for HMG2 was estimated by Northern blot hybridization of whole poly(A+) RNA with the ^{32}P-labeled pcD-MK10 insert as a probe. A single band of RNA species of approximate size 1.0K nucleotides was observed as shown in Figure 6. The size of the mRNA is similar to that of the pcD-MK10 insert, suggesting that this insert represents the full-length transcript. In our previous experiment, two bands of hybridization of sizes 1.2K and 2.3K nucleotides were observed when the cDNA for HMG1 was used as a probe, suggesting that the molecular sizes of mRNA for both proteins are different and that the nucleotide sequences coding for HMG1 and HMG2 are not cross-hybridized to one another.

Analysis of Genomic DNA. Southern blot analyses of genomic DNA from pig thymus were examined to identify possible multiplicity in the HMG2 gene. High molecular weight DNA was digested with various restriction enzymes that have a single or no recognition site in the HMG2 cDNA insert, separated electrophoretically, and blotted to the membrane. A DNA fragment that contains the 5′ half of the coding region (5′ probe; *Pst*I–*Sau*3AI fragment in Figure 1) for the initial hybridization and another fragment that contains the 3′ half of the coding region (3′ probe; *Sau*3AI–*Eco*RI fragment) for the rehybridization were used as probes. The results presented in Figure 7 indicate that each restriction enzyme produced one or two fragments hybridizable with the 5′ probe as well as the 3′ probe. In addition, no additional band except for the bands observed by the 5′ probe and the 3′ probe was detected when the whole insert was used as a probe. These results suggest that the HMG2 coding gene is homogeneous within genomic DNA. Moreover, the Southern hybridization may suggest that the genomic DNA contains at least one intervening sequence no longer than 0.7K nucleotide's near the *Sau*3AI site at the 478 nucleotide position. In similar experiments using HMG1 cDNA as a probe, heterogeneity in the HMG1 gene was indicated (Tsuda et al., 1988). The human genes for HMG14 and HMG17 were regarded as multigenes (Landsman et al., 1986a,b).

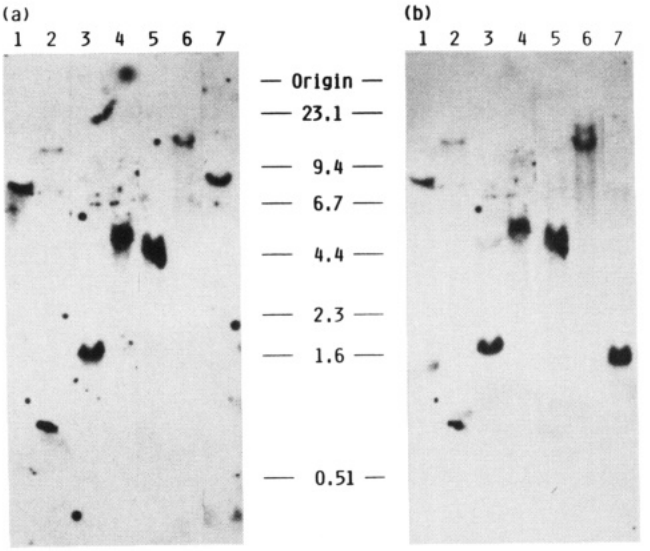

FIGURE 7: Southern blot analysis of pig genomic DNA. High molecular weight DNA from pig thymus was digested with *Pst*I (lane 1), *Bam*HI (lane 2), *Pvu*II (lane 3), *Hin*dIII (lane 4), *Eco*RI (lane 5), *Eco*RV (lane 6), and *Bgl*II (lane 7), analyzed on an agarose gel, and transferred to a GeneScreen Plus hybridization transfer membrane. The resulting hybridization was probed with (a) a 5′ probe (a *Pst*I–*Sau*3AI fragment of 427 bp in Figure 1) and (b) a 3′ probe (a *Sau*3AI–*Eco*RI fragment of 431 bp). DNA markers are indicated in the middle of the figures in kilobase pairs.

Most of the studies on HMG1 and HMG2 underline the similarities between the two proteins. They are both extracted from chromatin at the same ionic strength, their cellular distribution is similar [for a review, see Einck and Bustin (1985)], they exist in three domains of comparable lengths (Reeck et al., 1982), and their binding to DNA and to nucleosome is indistinguishable (Yoshida & Shimura, 1984; Waga et al., 1988, 1989; Bonne-Andrea et al., 1984). They show a transcription stimulation on a similar level (Waga et al., 1988; Tremethick & Molly, 1988) and so on. However, differences in their amino acid sequences by about 20%, and in their molecular weights, have been noted (Walker et al., 1980; Walker, 1982). The differences between the two proteins are also demonstrated by their immunological specificity (Bustin et al., 1978, 1982). Thus, the structures of HMG1 and HMG2 have been considered to be significantly different so as to display significant immunological specificity.

Comparison of the present data for HMG2 with our previous data for HMG1 from the same source (Tsuda et al.,

1988) clearly confirmed the differences in the primary structure between the two proteins. These structural differences may be reflected in the significant immunological specificity. Interestingly, their hydropathy index profiles (Figure 5) are rather identical, suggesting that these proteins are similar in their overall structures. The indistinguishable functional characteristics of both proteins may be reflected in the similarity in their overall structure. However, the difference between the two proteins in primary structure, especially in the sequences preceding the acidic carboxyl terminal (residues 167–179), raises a possibility that they are involved in distinguishable cellular functions.

Analysis of the available data concerning the pig thymus HMG2 and HMG1 genes and cDNAs points out an interesting situation where two distinct gene families have different features. The pig genome contains the genes which encode a single-sized transcript for HMG2 and various-sized transcripts for HMG1. The transcripts from HMG2 and HMG1 genes are different in size. The cDNAs have 74% homology for the nucleotide within their coding regions, while the 3′-untranslated regions are low in homology.

It is conceivable that the two proteins evolved from a common ancestral gene. Recently, Hayashi et al. (1989) reported that *Tetrahymena* HMG is rather similar to the central part of vertebrate HMG1 in terms of the amino acid sequence and the hydropathy profile. HMG1 and HMG2 may have arisen from such a primitive HMG protein. The high conservation of HMG1 protein sequence suggests that they are in a slow rate of evolution (Tsuda et al., 1988). This may be true of HMG2 protein. These evolutionary constraints on the conformation of the protein suggest that these proteins share a fundamental role in the nuclear functions.

As stated in the introduction, HMG2, as well as HMG1, is one of the most abundant and ubiquitous nuclear proteins associated with chromatin. So far as we know, the complete primary structure of the protein has not been reported. The availability of the cDNA and amino acid sequence for HMG2 will allow studies on the gene structure, on the transcriptional regulation of HMG2, and on elucidation of the cellular function of the protein.

Acknowledgments

We thank Dr. H. Nagayama, Hirosaki University, for his critical reading of the manuscript.

References

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

Bonne-Andrea, C., Harper, F., Sobczak, J., & De Recondo, A. M. (1984) *EMBO J. 3*, 1193–1199.

Bustin, M., Hopkins, R. B., & Isenberg, I. (1977) *J. Biol. Chem. 252*, 1694–1699.

Bustin, M., Dunn, B., Gillette, R., Mendelsohn, E., & Soares, N. (1982) *Biochemistry 21*, 6773–6777.

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

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

Hayashi, T., Hayashi, H., & Iwai, K. (1989) *J. Biochem. (Tokyo) 105*, 577–581.

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.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Landsman, D., & Bustin, M. (1986) *J. Biol. Chem. 261*, 16087–16091.

Landsman, D., Soares, N., Gonzalez, F. J., & Bustin, M. (1986a) *J. Biol. Chem. 261*, 7479–7484.

Landsman, D., Srikantha, T., Westermann, R., & Bustin, M. (1986b) *J. Biol. Chem. 261*, 16082–16086.

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

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

Reeck, G. R., Isackson, P. J., & Teller, D. C. (1982) *Nature 300*, 76–78.

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

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

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

Tremethick, D. J., & Molly, P. L. (1988) *Nucleic Acids Res. 16*, 11107–11123.

Tsuda, K., Kikuchi, M., Mori, K., Waga, S., & Yoshida, M. (1988) *Biochemistry 27*, 6159–6163.

Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.

Waga, S., Mizuno, S., & Yoshida, M. (1988) *Biochem. Biophys. Res. Commun. 153*, 334–339.

Waga, S., Mizuno, S., & Yoshida, M. (1989) *Biochim. Biophys. Acta 1007*, 209–214.

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

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

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