Volume 15 Number 13 1987 Nucleic Acids Research

Characterization of cDNA sequences corresponding to three distinct HMG-1 mRNA species in line CHO Chinese hamster cells and cell cycle expression of the HMG-1 gene

K.-L.D.Lee, B.T.Pentecost*, J.A.D'Anna[1], R.A.Tobey[1], L.R.Gurley[1] and G.H.Dixon

Department of Medical Biochemistry, Faculty of Medicine, Health Sciences Centre, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada and [1]Toxicology Group, Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received April 10, 1987; Revised and Accepted June 2, 1987 Accession no. Y00365

ABSTRACT

We have isolated cDNA clones encoding the high mobility group (HMG) protein HMG-1 in line CHO Chinese hamster cells. The cDNA clones correspond to the three HMG-1 mRNA species detected on Northern blots. Three different polyadenylation sites are found to be used. The three mRNA species of sizes 1.05, 1.45 and 2.45 kb are generated by differential polyadenylation at sites 115 nucleotides, 513 nucleotides and 1515 nucleotides downstream from the stop codon. A perfectly conserved putative poly(A) signal AAUAAA is present upstream of only one of the three poly(A) sites. Two homologous but imperfect sequences exist upstream from the other two poly(A) sites. All three HMG-1 mRNA species maintain significant levels throughout the M, G1 and S phases of the cell cycle and the rate of large HMG protein (HMG-1 and HMG-2) synthesis increases approximately two-fold from G1 to S phase.

INTRODUCTION

The high mobility group proteins are the better studied group of non-histone chromosomal proteins. HMG proteins characterized so far can be subdivided into two classes. The small HMG proteins have M.W. of 8-10,000 daltons. The bovine HMG-14, HMG-17 and the trout H6 belong to this class (1, 2) and have been found enriched in transcriptionally active chromatin (3, 4). HMG-1 and HMG-2 from calf thymus and HMG-T from trout testis are well characterized HMG proteins of the group having higher M.W.'s of 28-30,000 daltons (1, 2). They have characteristic polyacidic domains at their C-termini containing typically a continuous tract of about 30 glutamic or aspartic acid residues. Although large HMG proteins have been found enriched in active chromatin (5) and it has also been suggested that they might be involved in nucleosome assembly or DNA replication (5, 6), their exact function(s) has yet to be elucidated.

In cultured CHO cells, HMG-1 and HMG-2 proteins have been identified (7) but no amino acid sequence data have been reported. In preliminary studies, we found that the bovine HMG-1 cDNA clone isolated earlier in our laboratory

(8) cross-hybridized with CHO mRNA. However, the trout HMG-T mRNA showed only a very weak cross-hybridization signal to the bovine HMG-1 cDNA (9).

We then isolated CHO HMG-1 cDNA clones by screening a cDNA library using a 500 bp BstNI-BstNI fragment from the bovine HMG-1 cDNA clone (8). Three different sets of CHO HMG-1 cDNA clones have been characterized and show three substantially different lengths of their 3' untranslated regions. The three different sets of cDNA clones correspond to the three different-sized positive bands observed on RNA blots. Three bands of similar sizes have also been observed on bovine RNA blots (9). We have now identified all three bands as CHO HMG-1 mRNA species with 3' heterogeneity as a result of differential processing and polyadenylation at three different sites.

By analyzing RNA extracted at various times from synchronized CHO cell cultures, we observed that all three mRNA species were present during the mitotic stage, the G1 phase and the S phase of the cell cycle and that the levels of the three HMG-1 mRNA's showed a maximum of two-fold increase between the G1 and S phase. The HMG-1 mRNA levels correlate well with the synthesis rates of the HMG-1 and HMG-2 proteins. From across the G1 phase to the S phase of the CHO cell cycle, relative synthesis rates of the large HMG proteins (HMG-1 and HMG-2) increase by approximately two-fold. This is in strong contrast to the levels of histone H4 mRNA whose level is closely coupled with the S phase.

We have observed homologous sequences immediately upstream from two of the three sites at which polyadenylation occurs. We believe these homologous sequences might have significant roles in the processing and polyadenylation of HMG-1 transcripts. A perfectly conserved putative polyadenylation signal, AAUAAA, is present at only one of the three sites. 3' heterogeneity generated by differential processing and polyadenylation has also been observed in mouse dihydrofolate reductase mRNAs, in chicken vimentin mRNAs and in rat brain myelin proteolipid protein mRNAs (10, 11, 12). We suggest that transcription of HMG-1 gene in CHO cells might terminate (or undergo processing) heterogeneously in the 3' non-coding region and that polyadenylation of the transcripts occurs at the nearest available poly(A) site.

MATERIALS AND METHODS

Library Screening

A CHO cDNA library synthesized using the Okayama and Berg method (13) and transformed into *E. coli* X1776 cells was kindly provided by Dr. P. Berg, Department of Biochemistry, Stanford University. A 500 bp BstNI-BstNI DNA

fragment from a bovine cDNA for HMG-1 pBP1 (8) was labelled with ^{32}P by nick-translation (14) and used as a probe to screen duplicated nitrocellulose replicas of colonies. Prehybridization and hybridization were carried out at 42°C in 50% (v/v) formamide, 5x SSC (0.75 M NaCl, 75 mM NaCitrate, pH 7.0), 50 mM Na phosphate (pH 6.5), 0.1% (w/v) SDS, 5x Denhardt's solution, 200 µg/ml yeast tRNA and without or with labelled probe fragment, respectively. Filters were washed three times at room temperature in 2x SSC for 5 min. each, followed by washing twice at 42°C in 1x SSC for 15 min. each. Positive colonies (present on both replicas) were rescreened and positive clones confirmed.

DNA Sequencing

Positive clones were restriction mapped and overlapping DNA fragments were subcloned into pUC12 plasmids. The cDNA fragments were sequenced using the chemical cleavage method of Maxam and Gilbert (15). Four of the positive clones have been fully sequenced on both strands. Sequence analysis and sequence comparisons were performed with the Microgenie (Beckman) computer programmes (16).

RNA Blot Analysis

RNA was extracted from CHO cells using the method of Chirgwin et al. (17) and electrophoresed in 1% agarose gels containing 10 mM methylmercuric hydroxide as described (18). RNA was visualized by staining the gel with 10 µg/ml ethidium bromide in 0.5 M ammonium acetate and then capillary-blotted onto Hybond-N membranes (Amersham). A 280 bp PstI-PstI fragment (fragment "a" in Figure 1) was "oligo-labelled" with ^{32}P as described (19). Prehybridization and hybridization were performed at 42°C as described above. The membranes were washed at 42°C in 5x SSPE (0.9 M NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.0) for 15 min. and repeated once, followed by washing for 30 min. at 42°C in 1x SSPE, 0.1% (w/v) SDS and in 0.1x SSPE, 0.1% (w/v) SDS at room temperature for 15 min.

Cell Cycle Analysis of HMG-1 mRNA Levels

CHO cell cultures were synchronized by shaking monolayers every 10 min. for 6 hrs. to collect mitotic cells as described (20). Collected cells were kept on ice and then released from mitosis by suspending them in warm medium. Aliquots of cells were withdrawn every one or one-and-a-half hours and centrifuged through an iced medium to arrest their cell cycle traverse. Cell pellets were frozen on dry ice and stored at -80°C. RNA samples were prepared as described above and HMG-1 mRNA levels were analyzed on RNA blots. The following amounts of total RNA were loaded onto the gel: M - 3.5 µg; 1 hr -

3.0 µg; 2 hr - 5.5 µg; 3 hr - 5.5 µg; 4 hr - 5.5 µg; 5.5 hr - 6 µg; 7 hr - 5.5 µg; 8.5 hr - 6.0 µg; 10 hr - 8.5 µg. HMG-1 mRNA levels are determined by scanning the autoradiographs of the Northern bolts using a videodensitometer (Biorad, model 620). Relative mRNA levels were calculated as (optical density of labelled RNA band)/(total RNA loaded onto gel) and normalized to a range of 0 to 1.0. In a separate experiment, the mRNA levels of the core histone H3 were assayed in the same manner. A 900 bp EcoRI-SalI mouse histone H3.2 DNA fragment was used as probe. The mouse H3.2 DNA clone was kindly provided by Dr. G. Schultz, Department of Medical Biochemistry, University of Calgary (21) and was first isolated by Sittman et al. (22). Synthesis of DNA in synchronized CHO cell cultures was assayed by measuring the incorporation of [^{3}H] thymidine.

Cell Cycle Analysis of HMG Protein Synthesis

Relative rates of protein synthesis for fractions containing HMG-1 plus HMG-2 were measured the same as previously described for histones H1 and H1° (23) following the release of cells synchronized by mitotic selection. Cells were synchronized by mitotic selection, and at appropriate times, they were pulse-labeled for 1 hr with 4.2 µCi mL^{-1} [^{3}H]lysine before harvest. The HMG proteins were extracted by blending radiolabeled and carrier cells in 0.83 M $HClO_4$. The total $HClO_4$-soluble, Cl_3CCO_2H-precipitable proteins were then separated by high performance liquid chromatography as described in D'Anna et al. (23). Fractions containing HMG-1 plus HMG-2 were collected and quantified by measuring the [^{3}H]-incorporation by liquid scintillation spectrometry. The rates of synthesis are based on the incorporation per cell relative to the maximum rate in S phase.

Specific activities of HMG-1 were also measured following the release of G1 cells synchronized by the isoleucine deprivation method (24). At various times after release from G1 arrest, cells were pulse-labeled with 1 µCi mL^{-1} [^{3}H]lysine for 1 hr before harvest. HMG-1 and other proteins were extracted with 0.83 M $HClO_4$ from the chromatin of cells blended in isotonic saline and precipitated by the addition of Cl_3COO_2H to 20% (w/v). HMG-1 was resolved by electrophoresis, and the specific activities of HMG-1 were calculated from the [^{3}H]lysine incorporated into the HMG-1 as previously described for histone H1 and H1° (25). The fraction of cells in S phase was determined by autoradiography of cells that had incorporated [^{3}H]thymidine for one hour.

RESULTS AND DISCUSSION

Isolation and Characterization of cDNA Clones

Using the 500 bp BstNI-BstNI fragment from the bovine HMG-1 cDNA clone (pBP1) to probe the CHO cell cDNA library, we have identified seven positive CHO cDNA clones from a total of approximately 30,000 colonies. The positive clone with the largest insert size, pCH3, was picked and sequenced. Sequencing data showed that pCH3 contained a long 3' untranslated region of 1515 nucleotides (Fig. 3) plus 192 nucleotides of the coding region (Fig. 2). The longest open reading frame (LORF) showed that the polyacidic domain characteristic of this class of HMG-1 proteins was encoded at the C-terminus and is composed of 20 glutamic acid residues and 10 aspartic acid residues.

Three other positive cDNA clones were then fully sequenced. One of these three clones, pCH1, was found to contain 513 nucleotides of the coding region (Fig. 2) and a 3' untranslated region of only 115 nucleotides. The other two clones sequenced both have the same 3' untranslated region of 513 nucleotides. One of these two clones contains 231 nucleotides and the other only 200 nucleotides of the coding region. The longer of these two cDNA clones was

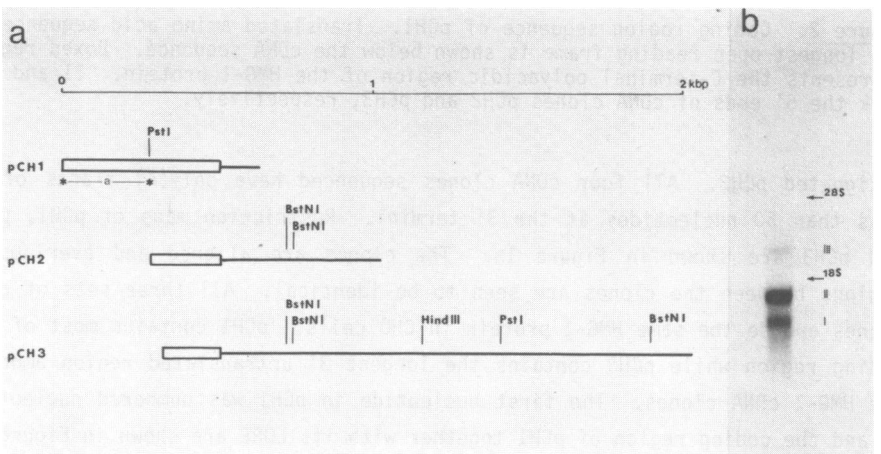

Figure 1: (a) Restriction maps of CHO HMG-1 cDNA inserts. Coding regions are shown as boxes. 3' untranslated regions are represented by solid lines. The three clones are aligned. Overlapping regions are identical. (b) Northern blot of CHO RNA probed with the 280 bp PstI-PstI fragment (marked as "a" on the restriction map). Three positive bands observed are marked as I, II, and III, respectively.

```
                                                                      60
TCA GAA AGG TGG AAG ACC ATG TCT GCT AAA GAA AAG GGA AAA TTT GAG GAC ATG GCA AAG
Ser Glu Arg Trp Lys Thr Met Ser Ala Lys Glu Lys Gly Lys Phe Glu Asp Met Ala Lys

                                                                     120
GCT GAC AAA GCT CGT TAT GAA AGA GAA ATG AAA ACC TAC ATC CCC CCC AAA GGG GAG ACC
Ala Asp Lys Ala Arg Tyr Glu Arg Glu Met Lys Thr Tyr Ile Pro Pro Lys Gly Glu Thr

                                                                     180
AAA AAG AAG TTC AAG GAC CCC AAT GCA CCC AAG AGG CCT CCT TCG GCC TTC TTC TTG TTC
Lys Lys Lys Phe Lys Asp Pro Asn Ala Pro Lys Arg Pro Pro Ser Ala Phe Phe Leu Phe

                                                                     240
TGT TCT GAG TAT CGC CCA AAA ATC AAA GGA GAA CAT CCA GGG CTG TCC ATT GGT GAT GTT
Cys Ser Glu Tyr Arg Pro Lys Ile Lys Gly Glu His Pro Gly Leu Ser Ile Gly Asp Val

                                                    II
                                                    ↓
                                                                     300
GCA AAG AAA CTG GGA GAG ATG TGG AAC AAC ACT GCT GCA GAT GAC AAG CAG CCC TAT GAA
Ala Lys Lys Leu Gly Glu Met Trp Asn Asn Thr Ala Ala Asp Asp Lys Gln Pro Tyr Glu

                          III
                          ↓
                                                                     360
AAG AAG GCT GCT AAA CTG AAG GAG AAG TAT GAA AAG GAT ATT GCT GCT TAC AGA GCT AAA
Lys Lys Ala Ala Lys Leu Lys Glu Lys Tyr Glu Lys Asp Ile Ala Ala Tyr Arg Ala Lys

                                                                     420
GGA AAA CCC GAT GCA GCG AAA AAG GGG GTG GTT AAG GCA GAA AAG AGC AAG AAA AAG AAG
Gly Lys Pro Asp Ala Ala Lys Lys Gly Val Val Lys Ala Glu Lys Ser Lys Lys Lys Lys

                                                                     480
GAA GAG GAA GAT GAT GAG GAG GAT GAA GAG GAT GAG GAA GAG GAG GAA GAA GAG GAA GAC
Glu Glu Glu Asp Asp Glu Glu Asp Glu Glu Asp Glu Glu Glu Glu Glu Glu Glu Glu Asp

GAA GAT GAA GAA GAA GAT GAT GAT GAT GAA TAA
Glu Asp Glu Glu Glu Asp Asp Asp Asp Glu End
```

Figure 2: Coding region sequence of pCH1. Translated amino acid sequence of the longest open reading frame is shown below the cDNA sequence. Boxed region represents the C-terminal polyacidic region of the HMG-1 protein. II and III mark the 5' ends of cDNA clones pCH2 and pCH3, respectively.

designated pCH2. All four cDNA clones sequenced have poly(A) tracts of no less than 50 nucleotides at the 3' termini. Restriction maps of pCH1, pCH2 and pCH3 are shown in Figure 1a. The clones are aligned and overlapping regions between the clones are seen to be identical. All three sets of cDNA clones encode the same HMG-1 protein in CHO cells. pCH1 contains most of the coding region while pCH3 contains the longest 3' untranslated region amongst the HMG-1 cDNA clones. The first nucleotide in pCH1 was numbered nucleotide #1 and the coding region of pCH1 together with its LORF are shown in Figure 2. The 5' end of pCH2 and pCH3 are marked by arrows II and III. The 3' untranslated region of pCH3 is shown in Figure 3 with the polyadenylation sites of pCH1, pCH2 and pCH3 marked by arrows, I, II, and III, respectively. A sequence termed pCH-HMG1 was formed by combining sequences from all three clones. It is composed of essentially the coding region in pCH1 and the 3' untranslated region of pCH3. This sequence was employed in sequence comparisons and analyses.

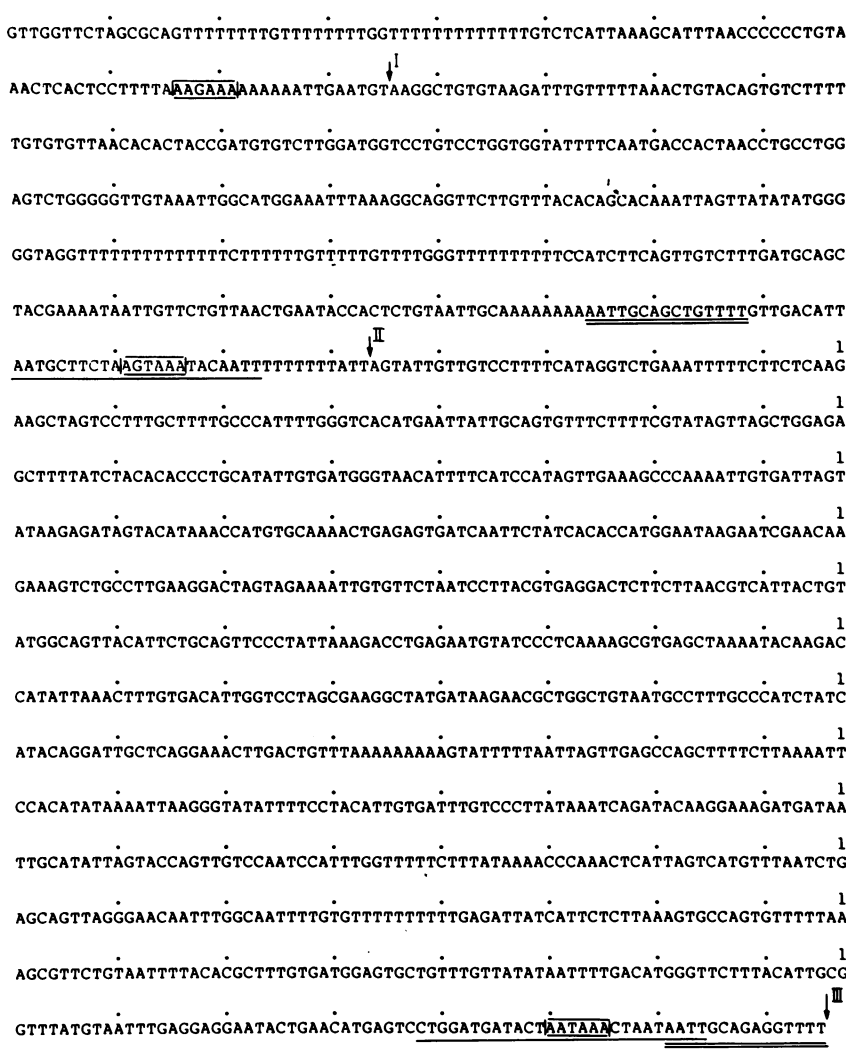

Figure 3: 3' untranslated region sequence of pCH3. I, II and III mark the polyadenylated sites in cDNA clones pCH1, pCH2 and pCH3, respectively. The putative hexanucleotide poly(A) signals are boxed. Singly- and doubly-underlined regions are homologous sequences associated with poly (A) sites II and III.

Three HMG-1 mRNA Species Generated by Differential Polyadenylation

Three positive bands were observed when CHO RNA blots were probed with a 280 bp PstI-PstI fragment from pCH1 (Figures 1b and 5). The three bands are marked as I, II and III. Their sizes as estimated on the RNA gel are 1.05 kb,

Table 1 : HMG-1 mRNA Sizes and Poly(A) Sites

(a)

mRNA Band	Size(kb)
I	1.05
II	1.45
III	2.45

(b)

Poly(A) Site	Nucleotide Number
I (pCH1)	628
II (pCH2)	1026
III (pCH3)	2028

1.45 kb and 2.45 kb, respectively. As shown in Figure 3, the poly(A) sites of pCH1, pCH2 and pCH3 are at nucleotide #628, nucleotide #1026 and nucleotide #2028, respectively (Table 1). The size difference between mRNA bands I and II is 400 nucleotides, which is the same as the number of nucleotides between the poly(A) sites I and II. Similarly, the size difference between mRNA bands II and III is the same as the number of nucleotides (1002) between the poly(A) sites II and III. poly(A) site III is 1.4 kb downstream from poly(A) site I and mRNA band III has a size 1.4 kb larger than that of mRNA band I. We believe, therefore, that the cDNA clones pCH1, pCH2 and pCH3 represent the HMG-1 mRNA species I, II and III, respectively.

A perfectly conserved putative poly(A) signal, AAUAAA, is present only at poly(A) site III. At poly(A) sites I and II, modified sequences AAGAAA and

Figure 4: Polyadenylation signal sequences and conserved adjacent regions. (a) and (b): Homologous sequences found at poly(A) sites II and III. (c): Comparison of the hexanucleotide sequences at the three poly(A) sites. Locations of sequences are indicated by nucleotide numbers. Negative numbers illustrate number of nucleotides upstream from the respective poly(A) site.

AGUAAA are present, respectively (Fig. 3 and Fig. 4). Therefore, it appears that a perfect hexanucleotide AAUAAA is not an absolutely essential element in the efficient polyadenylation of HMG-1 transcripts. On RNA blots, mRNA species II is always seen to be present in a much higher quantity than the mRNA species I and III. This might be a result of greater stability of mRNA species II or it might be a result of higher efficiency of termination and/or polyadenylation at site II.

We also observed that at poly (A) sites II and III, the hexanucleotides AGUAAA and AAUAAA are embedded in a 26-27 bp AT-rich sequence and there is a 75% homology between the two sites (Fig. 3, singly-underlined sequences and Fig. 4a for homology comparison). A second shorter sequence of 15 bp associated with sites II and III is indicated by the double underline in Fig. 3 and shows 87% identity in the two cases (Fig. 4b). However, this sequence is located upstream (5') of the hexanucleotide AGUAAA at poly (A) site II but downstream (3') of the AAUAAA sequence at site III. Since neither of these regions is associated with poly (A) site I, their possible functional significance in polyadenylation is difficult to assess although similar sequences are also found associated with the poly A site in the bovine HMG-1 cDNA sequence (8).

It is generally recognized that synthesis of mature poly(A)$^+$ mRNAs involves termination of transcription, specific cleavage of the primary transcript and polyadenylation at the poly(A) sites. In the present situation, several alternatives that might explain the 3' heterogeneity of the HMG-1 mRNA's are apparent. First, it is possible that there is a single primary transcript which encompasses all three poly (A) sites I, II and III and which is then cleaved and polyadenylated at each of the three sites with different efficiencies. This model would predict that cleavage or polyadenylation at site II was most efficient thus accounting for the predominance of the 1.45 kb mRNA II. The second possibility is that transcription of the HMG-1 gene may terminate at three distinct sites, producing three primary transcripts which are, in turn, processed into the three mRNA species at the poly A site closest 5' to the termination site. In this scenario, transcription termination would be most efficient 3' to the poly A site II. The third possibility is that transcripts of the HMG-1 gene may terminate heterogeneously throughout a wide 3' untranslated region as has been observed with other eukaryotic genes (26,27). Specific cleavage and polyadenylation would then take place at the closest poly A site on the 5' side of the point of termination. Since the distance between sites II and III

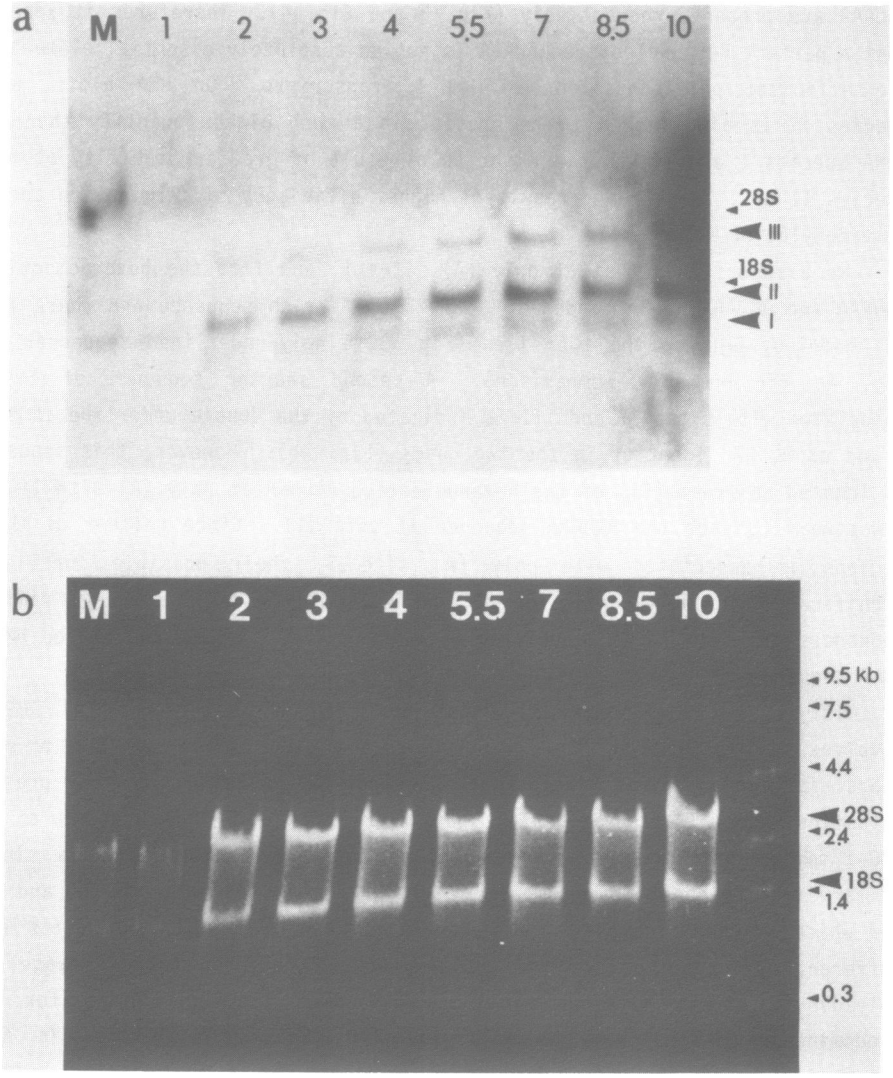

Figure 5: (a) Northern blot of RNA from synchronized CHO cells. The lanes are labelled as number of hours after mitosis (M). cDNA fragment "a" (Figure 1) was used as probe. (b) Ethidium bromide staining of RNA gel before blotting. Amounts of CHO RNA loaded are: M - 3.5 µg; 1 hr - 3.0 µg; 2 hr - 5.5 µg; 3 hr - 5.5 µg; 4 hr - 5.5 µg; 5.5 hr - 6.0 µg; 7 hr - 5.5 µg; 8.5 hr - 6.0 µg; 10 hr - 8.5 µg.

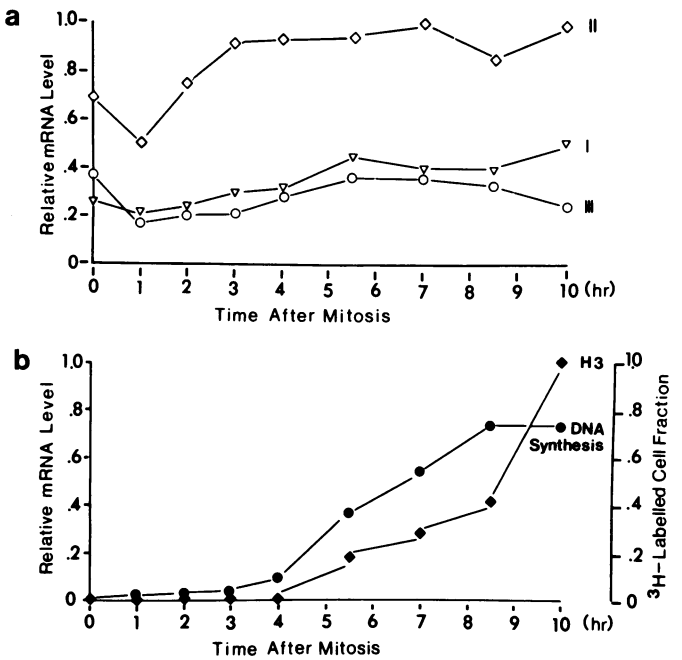

Figure 6: (a) Expression pattern of HMG-1 in synchronized CHO cells. Relative levels of HMG-1 mRNA species I, II, and III are plotted against culture time after mitosis. (b) Histone H3 mRNA levels and DNA synthesis (measured as ^{3}H-labelled cell fraction) in synchronized CHO cells.

is much greater (1002 bp) than between sites I and II (400 bp), the probability of termination, processing and polyadenylation at site II is much higher than that for the other two sites and hence more HMG-1 mRNA II would result. Another reasonable explanation would be that cleavage of the transcript may occur prior to termination of transcription and there is a high probability that the transcript is cleaved while the RNA polymerase is transcribing the DNA between site II and site III. Site I is probably a less efficient cleavage site as it lacks the homologous sequences found at site II and site III (Fig. 4a and 4b). In addition, at site I there is a base change of U to G in the putative poly(A) hexanucleotide signal, AAUAAA. Until we have more information about the mechanism of transcription termination and the nature of the primary HMG-1 mRNA transcripts, it is not possible to choose between these alternatives. It is worth noting that whatever the mechanism is

for generating the three HMG-1 mRNA's in CHO cells, a similar situation must apply to bovine cells since three HMG-1 mRNA species of similar size, with mRNA II predominant, have also been observed on bovine RNA blots (a).

It has also been observed in eukaryotic systems that alternative splicing of the primary transcript is a mechanism often proposed when heterogeneous families of mRNA's are observed (28). In the present case, since the overlapping regions up to the poly (A) sites in the three HMG-1 cDNA species have identical sequence, this would not be consistent with an alternative splicing model which would predict that regions of sequence representing "skipped" exons would be missing in the shorter mRNA's. Other mRNAs, e.g. dihydrofolate reductase mRNAs (10), vimentin mRNAs (11) and brain myelin proteolipid protein mRNAs (12) have already been reported to demonstrate 3' heterogeneity similar to the HMG-1 mRNAs and there is no indication that this phenomenon of 3' heterogeneity is generated by alternative splicing.

HMG-1 mRNA Levels in the CHO Cell Cycle

Cultured CHO cells arrested at the mitotic stage were released into culture synchronously. Levels of the three HMG-1 mRNA's were assayed as described in Materials and Methods using RNA (Northern) blots. We observed that all three HMG-1 bands were present at the mitotic, G1 and S phases of the CHO cell cycle (Figure 5a). RNA loaded onto the gel was visualized by ethidium bromide staining of the agarose gel before capillary blotting (Figure 5b). Relative mRNA levels were plotted against the time after synchronization (Figure 6a). There appear to be significant levels of all three HMG-1 mRNA species during the cell cycle and between early G1 phase and S phase of the cell cycle, there is apparently an approximately two-fold increase in the levels of the three mRNA species. This increase is most noticeable in the major species, mRNAII, between 1 and 5.5 hr after release from mitosis (Fig. 6a).

The HMG-1 mRNA levels detected correlate very well with the relative rate of synthesis of both the HMG-1 and HMG-2 proteins in CHO cells synchronized by mitotic selection or isoleucine deprivation (Fig. 7). In synchronized CHO cells released from mitotic selection, there is a two-fold increase in the relative synthesis rate of the large HMG proteins, HMG-1 and HMG-2 (Fig. 7a) while a similar synthesis rate increase of HMG-1 is also observed in CHO cells released from isoleucine deprivation arrested at S1 (Fig. 7b).

In contrast, the expression of the nucleosomal, core histone H3 gene, follows a totally different pattern. H3 mRNA is not detected until 5.5 hr after mitosis and sharply increases to a high level at 10 hr (Figure 6b,

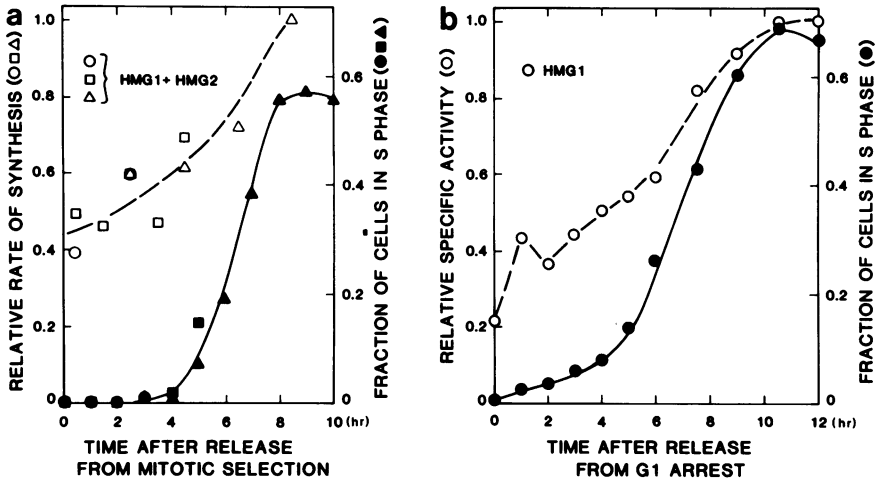

Figure 7: Relative rates of HMG protein synthesis and relative specific activities. (a) Relative rate of HMG-1 plus HMG-2 synthesis (○ , □ , △) presented as a fraction of maximum rate and the fraction of S phase cells (● , ■ , ▲) as functions of time after chilled mitotically selected cells were resuspended in warm medium to resume transverse of the cell cycle. The opened and closed form of each symbol denote measurements on cells taken from same mitotic selection populations. (b) Relative specific activity of HMG-1 (○) and the fraction of S phase cells (●) as functions of time after the release of synchronized cells from isoleucine deprivation G1 arrest.

autoradiograph not shown). Synthesis of DNA in cultured CHO cells follows a pattern similar to the H3 mRNA level. As new core histones are required for assembling newly synthesized DNA molecules into chromatin, it is not surprising that synthesis of DNA and an increase of the level of core histone mRNAs are tightly coupled and this has been consistently observed (29, 30).

Since clear biological roles for the class of larger M.W. HMG proteins are not yet established, the functional significance of the presence of significant levels of HMG-1 mRNA's and the increasing synthesis rate of the HMG-1 and HMG-2 proteins across G1 to S phase of the cell cycle, is not yet clear. However, the control of HMG-mRNA levels must be entirely different from that of the core histone mRNA's whose levels and translational activity are closely coupled to DNA synthesis in the S phase (29, 30). Two possible functional advantages might derive from the presence of appreciable levels of HMG-1 protein prior to the beginning of assembly of chromatin in early S phase. It is known that synthetic polymers of aspartic and glutamic acids greatly facilitate *in vitro* chromatin assembly (31) and by virtue of their

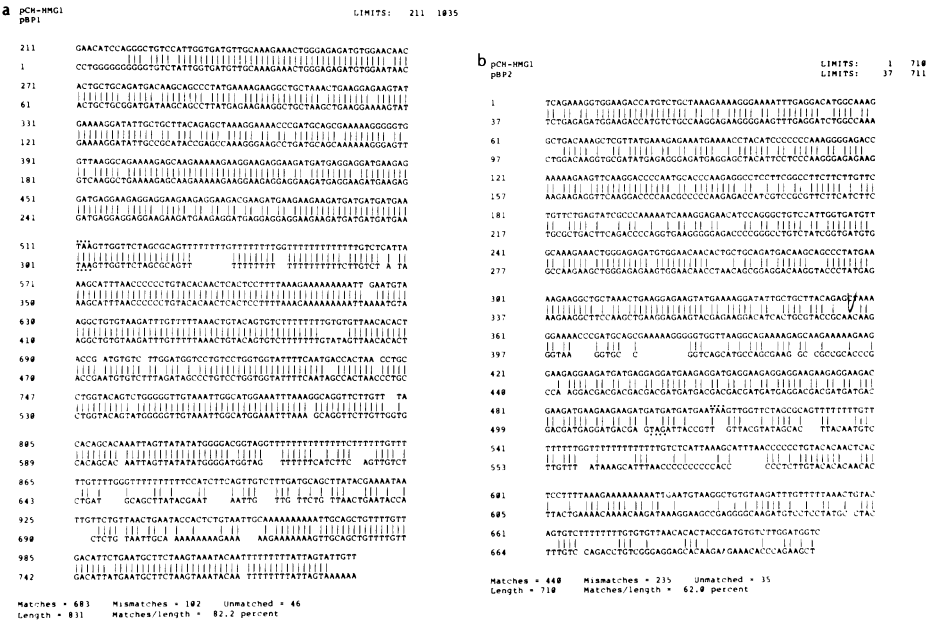

Figure 8: Alignment of comparable regions (a) between the CHO (pCH-HMG1, upper line) and the bovine (pBP1, lower line) HMG-1 cDNA sequences; (b) between the CHO HMG-1 cDNA (pCH-HMG1, upper line) and the trout HMG-T cDNA (pBP2, lower line). The stop codons in the cDNA sequences are marked with dots on top or below.

C-terminal poly-acidic tails, HMG's-1 and -2 have the same effect (6) as does HMG-T (F.D. Miller, D.C. Watson and G.H. Dixon, unpublished observations). Thus the presence of the HMG-1 protein at high levels at the moment chromatin assembly begins in S-phase could facilitate chromatin assembly. It has also been reported recently by Yoshida (32) that the tryptic polyacidic C-terminal peptide from HMG-1/2 can markedly lower the melting temperature of DNA in the presence of Mg^{2+} or Ca^{2+}. This DNA unwinding activity possessed by the HMG-1 or HMG-2 polyacidic domain might have a very significant role in DNA synthesis in G1-S phase of the cell.

It has also been shown that the large HMG proteins appear to be asymmetrically-distributed in chromatin and can be released by very mild micrococcal nuclease digestion (2). Furthermore, recent experiments in which chromosomal proteins can be cross-linked to DNA by UV irradiation of intact nuclei and the adduct recovered by a specific antibody to the protein in question (33), has shown that in trout liver nuclei, HMG-T is bound

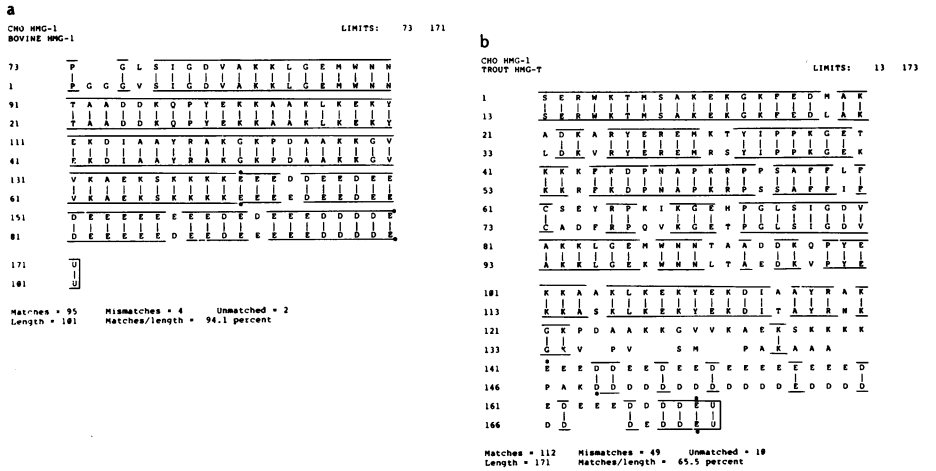

Figure 9: Alignment of amino acid sequences derived from the LORF of comparable regions in (a) the CHO HMG-1 cDNA (upper line) and the bovine HMG-1 cDNA (lower line); (b) the CHO HMG-1 cDNA (upper line) and the trout HMG-T cDNA (lower line).

specifically to certain DNA sequences. This evidence suggests that at the moment of chromatin assembly, the HMG proteins must bind to certain specific DNA sequences and therefore it would be advantageous to have HMG proteins present at sufficiently high levels to select such sites on the newly-synthesized DNA before they become fully complexed with newly-synthesized histones later in S phase.

Sequence Comparison

We have been able to isolate the bovine HMG-1 cDNA clone (pBP1) (8), the trout HMG-T cDNA clone (pBP2) (9) and the CHO HMG-1 cDNA clones (pCH-HMG1). Significant homology has been found between these three sequences. CHO cell pCH-HMG1 and bovine pBP1 are highly homologous in both the coding region and the 3' untranslated region (Fig. 8a). However, between CHO cell pCH-HMG1 and the trout HMG-T cDNA (pBP2), homology exists only in the coding region. The 3' untranslated regions are not homologous (Fig. 8b).

One feature of note in the 3' untranslated regions of both CHO cell and bovine cDNA clones is the presence of extensive and almost perfect homopolymeric regions of T and A. In pCH-HMG1, the T regions are more extensive and range in length up to 45 residues (e.g., in Fig. 3, 29 T residues from 530-561 interrupted by 3 G's and 34 T's from 840-884 interrupted by 5 G's and 1 C) while A regions are shorter ranging in length from 9-12

residues (e.g., 12 A's from 608-620, 10 A's from 959-968 and 9 A's from 1585-1593). The general features of these homopolymeric T and A regions are conserved in both CHO and bovine 3' untranslational regions (Fig. 8a) although they appear to be shorter and more interrupted in the bovine cDNA. In the trout HMG-T 3' untranslated sequence (Fig. 8b), where there is much less homology with the mammalian HMG-1 cDNA's, the long homopolymeric T regions are not apparent. However, one of the A-rich regions (608-620) is well-conserved between the CHO and trout HMG sequences. Interestingly enough, homopolymeric regions of T's are also found in the 3' untranslated regions of the human HMG-14 and HMG-17 cDNA clones (34). Whether these runs of T's and A's observed have functional significance is not clear. However, base-pairing between runs of T's and A's could be involved in the secondary structure and hence possibly the stability of the mRNA. As proposed by Shaw and Kamen (35), A-T rich sequences in the 3' untranslated region may be involved in selective degradation of mRNAs.

Comparisons have also been made between amino acid sequences deduced from the LORF of the cDNA sequences. Alignments of comparable regions are shown in Figure 9. There are three amino acid substitutions between the CHO and the bovine HMG-1 sequences (Fig. 9a) all in the C-terminal polyacidic domain. Mismatches observed at the 5' ends of the LORF's probably result from an ambiguity at the 5' end of the pBP1 cDNA sequence (8) and will be resolved when sequence data for the bovine HMG-1 gene becomes available. The CHO HMG-1 and the trout HMG-T amino acid sequences differ substantially in the C-terminal region (Fig. 9b). The polyacidic domain in CHO HMG-1 is composed of 20 glutamic acid residues and 10 aspartic acid residues while the polyacidic domain of the trout HMG-T contains 21 aspartic acid residues and only 3 glutamic acid residues. The highly conserved region (residues #1-122 in CHO HMG-1 and residues #13-134 in trout HMG-T) corresponds to the central domain of the protein and this domain forms highly-folded structures in solution (36). There is a highly divergent region between the central domain and the polyacidic domain at the C-terminus. This region of the protein might serve as a buffer or transition zone between the highly-folded central domain and the highly-destabilizing polyacidic domain such that the two domains may maintain their individual structures both of which are likely to be essential for the HMG-1 protein to function.

ACKNOWLEDGEMENTS

We thank Dr. P. Berg (Stanford) for the CHO cDNA library, Dr. C.C. Lin for a CHO cell line, Dr. G. Schultz for the H3 probe, M. Herfort and W. Connor

for technical support, and M. Cudmore for preparation of the manuscript. This work was supported by grants from the M.R.C. of Canada to G.H. Dixon and by the U.S. Department of Energy at the Los Alamos National Laboratory.

*Present address: NIEHS, MD 1301, PO Box 12233, Research Triangle Park, NC 27709, USA

REFERENCES

(1) Walker, J.M. (1982). In: The HMG Chromosomal Proteins, pp. 69-88. Academic Press, London, New York.
(2) Dixon, G.H. (1982) In: The HMG Chromosomal Proteins, pp. 149-182. Academic Press, London, New York.
(3) Levy, W.B. and Dixon, G.H. (1978). Nucl. Acids Res. 15:4155-4163.
(4) Weisbrod, S., Groudine, M. and Weintraub, H. (1980). Cell 19:289-301.
(5) Levy, W.B., Connor, W. and Dixon, G.H. (1979). J. Biol. Chem. 254:609-620.
(6) Bonne-Andrea, C., Harper, F., Sobczak, J. and DeReconde, A-M. (1984). EMBO J. 3:1193-1199.
(7) D'Anna, J.A., Becker, R.R., Tobey, R.A. and Gurley, L.R. (1983). Biochim. Biophy. Acta 739:197-206.
(8) Pentecost, B. and Dixon, G.H. (1984). BioSci. Rep. 4:49-57.
(9) Pentecost, B.T., Wright, J.M. and Dixon, G.H. (1985). Nucl. Acids Res. 13:4871-4888.
(10) Setzer, D.R., McGrogan, M., Nunberg, J.H. and Schimke, R.T. (1980). Cell 22:361-370.
(11) Zehner, Z.E. and Paterson, B.M. (1983). Proc. Natl. Acad. Sci. USA 80:911-915.
(12) Milner, R.J., Lai, C., Nave, K-A., Lenoir, D., Ogata, J. and Sutcliffe, J.G. (1985). Cell 42:931-939.
(13) Okayama, H. and Berg, P. (1982). Mol. Cell Biol. 2:161-170.
(14) Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975). Proc. Natl. Acad. Sci. USA 72:1184-1188.
(15) Maxam, A.M. and Gilbert, W. (1980). Methods Enzymol. 65:499-559.
(16) Queen, C. and Korn, L.J. (1984). Nucl. Acids Res. 12:581-599.
(17) Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979). Biochemistry 18:5294-5304.
(18) Chandler, P.M., Rimkus, D. and Davidson, N. (1979). Anal. Biochem. 99:200-206.
(19) Feinberg, A.P. and Vogelstein, B. (1983). Anal. Biochem. 132:6-13.
(20) Tobey, R.A., Anderson, E.C. and Petersen, D.F. (1967). J. Cell. Physiol. 70:63-68.
(21) Giebelhaus, D.H., Heikkila, J.J. and Schultz, G.A. (1983). Dev. Biol. 98:148-154.
(22) Sittman, D.B., Chin, I-M., Pan, C-J., Cohn, R.H., Kede, L.H. and Marzluff, W.F. (1981). Proc. Natl. Acad. Sci. USA 78:4078-4082.
(23) D'Anna, J.A., Thayer, M.M., Tobey, R.A. and Gurley, L.R. (1985). Biochemistry 24:2005-2010.
(24) Tobey, R.A. and Key, K.D. (1971). Cancer Res. 31:46-51.
(25) D'Anna, J.A., Gurley, L.R. and Tobey, R.A. (1982). Biochemistry 21:3991-4001.
(26) Citron, B., Falck-Pedersen, E., Salditt-Georgieff, M. and Darnell, J.E., Jr. (1984). Nucl. Acids Res. 12:8723-8731.
(27) Frayne, E.G., Leys, E.J., Crouse, G.F., Hook, A.G. and Kellems, R.E. (1984). Mol. Cell Biol. 4:2921-2924.
(28) Leff, S.E. and Rosenfeld, M.G. (1986). Ann. Rev. Biochem. 55:1091-1117.

(29) Heintz, N., Site, H.L. and Roeder, R.G. (1983). Mol. Cell. Biol. 3:539-550.
(30) Alterman, R-B.M., Ganguly, S., Schulze, D.H., Marzluff, W.F., Schildkraut, C.L. and Skoultchi, A.I. (1984). Mol. Cell. Biol. 4:123-132.
(31) Stein, A., Whitlock, J.P. and Bina, M. (1979). Proc. Natl. Acad. Sci. USA 76:5000-5004.
(32) Yoshida, M. (1987). J. Biochem. 101:175-180.
(33) Blanco, J., States, J.C. and Dixon, G.H. (1985). Biochemistry 24:8021-8028.
(34) Landsman, D. and Bustin, M. (1986). J. Biol. Chem. 261:16087-16091.
(35) Shaw, S. and Kamen, R. (1986). Cell 46:659-667.
(36) Cary, P.D., Turner, C.H., Leung, I., Mayes, E. and Crane-Robinson, C. (1984). Eur. J. Biochem. 143:323-330.