A High-Mobility-Group Protein and Its cDNAs from
*Drosophila melanogaster*

CYNTHIA R. WAGNER,¹,² KOEI HAMANA,³ AND SARAH C. R. ELGIN¹*

Department of Biology, Box 1137, Washington University, St. Louis, Missouri 63130;¹ Division of Biology and Biomedical Sciences, Washington University, St. Louis, Missouri 63110;² and College of Medical Care and Technology, Gunma University, Maebashi, Gunma 371, Japan³

Received 23 September 1991/Accepted 4 February 1992

We have identified, purified, and characterized a high-mobility-group (HMG) protein and its cDNAs from *Drosophila melanogaster*. This protein, HMG D, shares most of the characteristics of vertebrate HMG proteins; it is extractable from nuclei with 0.35 M NaCl, is soluble in 5% perchloric acid, is relatively small (molecular weight of 12,000), has both a high basic (24%) and high acidic (24%) amino acid content, and is a DNA-binding protein. HMG D exhibits characteristics of both the vertebrate HMG 1 and 2 class and the HMG 14 and 17 class of proteins. Its amino acid sequence is similar (36% amino acid identity) to that of HMG 1, while its size and selective extraction with ethidium bromide are similar to properties of the HMG 14 and 17 class of proteins. HMG D is encoded by a single-copy gene that maps to 57F8-11 on the right arm of chromosome 2. Two transcripts are observed during embryogenesis; the protein is relatively stable throughout development. By the biochemical criteria of size, solubility, and amino acid content, HMG D appears to be the major HMG protein of *D. melanogaster*.

High-mobility-group (HMG) proteins are a class of abundant nonhistone chromosomal proteins found in the eukaryotic nucleus. HMG proteins have been studied extensively in vertebrates and because of their relative abundance, high degree of conservation, and DNA-binding properties, are thought to play an important role in chromatin structure. Originally identified in the early 1960s as acidic impurities in histone HI preparations (31), HMG proteins and their genes from a variety of vertebrates have been isolated and characterized (9). Nuclei from these organisms contain at least four major types of HMGs, which are divided into two classes on the basis of their electrophoretic and biochemical properties (20, 21, 50). HMG 1 and HMG 2 have a molecular weight of approximately 28,000, have similar structures, and share greater than 80% amino acid sequence identity. HMG 14 and HMG 17 are similar to each other in size (molecular weight, 9,000 to 11,000) and structure and share greater than 90% amino acid sequence identity. Recently, a fifth class of HMG proteins has been isolated from mammalian tissue culture cells (57). The HMG I (Y) proteins have properties similar to those of the HMG 14 and 17 group. All HMG proteins bind DNA, are extractable from nuclei with 0.35 M NaCl, and are soluble in 5% perchloric acid (PCA). They also have an unusual amino acid composition, with a high content of charged amino acids (20 to 30% acidic and 25% basic) and a proline content of 7 to 10% (22). It has been estimated that there is one molecule of HMG 1 and 2 for every 3,000 bp of DNA and one molecule of HMG 14 and 17 for every 2,000 bp of DNA in vertebrate nuclei (19). HMG-like proteins and their genes from several lower eukaryotes, including *Saccharomyces cerevisiae* (23, 35, 36, 48), *Tetrahymena pyriformis* (24), and *Tetrahymena thermophila* (38, 53, 54), have also been identified. Despite the fact that much is known about the HMG proteins, their function in the nucleus remains undefined and controversial.

Several reports regarding possible functions for the HMG 1 and 2 proteins point to a role in DNA replication and repair. These proteins preferentially bind single-stranded DNA but will also bind double-stranded DNA (29). Bianchi and coworkers (7) have demonstrated that HMG 1 interacts with cruciform DNA. Other reports have suggested that the HMG 1 and 2 group might act in both nucleosome disassembly and reassembly during transcription (47). There appear to be at least two subpopulations of the HMG 1 and 2 family of proteins; one that is tightly bound to nucleosomes and one that is rapidly released from the chromatin upon digestion with micrococcal nuclease.

Investigation into the function of the HMG 14 and 17 class has been stimulated by the hypothesis that these proteins may play a role in establishing an active chromatin structure (61). Studies have shown that HMG 14 and 17 bind DNA (1, 11) but have a higher affinity for DNA within nucleosomes than for naked DNA (43, 49). Partial nucleosome reconstitution studies and DNA-protein cross-linking have demonstrated two specific binding sites for these proteins on the nucleosome, placing them on the internal side of the DNA helix where the DNA enters and exists the nucleosome (42, 49, 55). The HMG 1 (Y) class of proteins is expressed in rapidly dividing, undifferentiated cells (33). HMG I from the mouse (previously referred to as α-protein) has been shown to bind preferentially to A+T-rich, double-stranded DNA sequences (57).

While structural information and data on characteristics of DNA and nucleosome binding have been obtained for the vertebrate HMG proteins, it has not been possible to determine their function in vivo. Because of the ease with which biochemical, genetic, cytological, and molecular approaches can be combined, *Drosophila melanogaster* provides an excellent system in which to study the HMG proteins. Previous studies in *D. melanogaster* have identified two nuclear proteins with HMG-like properties; one, A13, is similar in amino acid composition to vertebrate HMG proteins (6), and the other, D1, has sequence similarity to vertebrate HMG I, although it is a much larger protein (molecular weight of 37,005 versus 10,000) (2, 5). We have

* Corresponding author.
undertaken a thorough biochemical analysis to identify the HMG protein(s) from D. melanogaster and report here the identification and characterization of HMG D (D for Drosophila) and its cDNAs. HMG D appears to be the major HMG protein of D. melanogaster. HMG D is a non-specific DNA-binding protein, is extracted from nuclei with ethidium bromide (EtBr), and has amino acid similarity to bovine HMG 1. (Ner and Travers [45a], using a protein-DNA binding assay, have independently isolated the gene encoding HMG D). Biochemical characterization and cDNA analysis indicate that HMG D shares properties with both of the two major groups of vertebrate HMG proteins.

MATERIALS AND METHODS

Protein isolation and amino acid composition. One hundred grams of 6- to 18-h-old Drosophila embryos was collected and frozen at ~80°C as previously described (45). Nuclei were isolated as described by Gilmour et al. (17). The pelleted nuclei were divided into five equal portions, and each portion was resuspended in 20 ml of one of the following reagents: 0.35 M NaCl, 5% PCA, 10 mM spermine, or 10 mM EtBr. Each nuclear suspension was homogenized, and the nuclear debris was pelleted at 20,000 × g for 10 min. The supernatant from each (containing the extracted nuclear proteins) was dialyzed separately for at least 6 hours against 0.1% acetic acid at 4°C. The extracts were lyophilized and resuspended in 500 μl of phosphate-buffered saline prior to analysis by gel electrophoresis.

The extracted nuclear proteins were analyzed by using several different polyacrylamide gel systems, including sodium dodecyl sulfate (SDS), Triton X-100-acetic acid-urea, acetic acid-urea (56), and SDS-urea (39).

The HMG D band from each extract was cut out from an SDS-12.5% polyacrylamide gel, and the protein was electrophoretically eluted by the method of Hankapiller et al. (28). The protein samples were hydrolyzed with 6 N HCl vapor for 18 to 24 hours at 100°C in vacuo. Hydrolysates were dissolved in Na-S sample dilution buffer (Beckman Instruments, Palo Alto, Calif.) and analyzed on a Beckman model 6300 amino acid analyzer.

Protein purification and amino acid sequencing. One hundred grams of 6- to 12-h-old Drosophila embryos was used to isolate nuclei as described above. The embryos were resuspended in 75 ml of 5% PCA and homogenized, and the nuclear debris was pelleted. The supernatant was concentrated (in the presence of 0.1 mM phenylmethylsulfonyl fluoride and 1 μg of leupeptin per ml) against dry polyethylene glycol 6000 at 4°C. The concentrated extract, approximately 35 ml, was equilibrated (as determined by conductivity readings) against running buffer (50 mM N-2-hydroxyethylpiperazine-N'-ethanesulfonic acid [HEPES], 50 mM NaCl, pH 7.8) at 4°C. The protein solution was then loaded onto an FPLC Mono S column (HR 5/5; Pharmacia) which had been previously equilibrated with running buffer. Protein fractions were collected over a linear gradient of 50 mM to 1 M NaCl. Most of the HMG D protein eluted from the column at approximately 300 mM NaCl, in three 1-ml fractions, as detected by SDS-12.5% polyacrylamide gel electrophoresis. These three fractions were pooled and dialyzed (in the presence of 0.1 mM phenylmethylsulfonyl fluoride and 1 μg of leupeptin per ml) against running buffer. The pooled fractions were loaded onto an FPLC Mono Q column (HR 5/5) previously equilibrated with running buffer. Fractions were collected over a linear gradient of 50 mM to 1 M NaCl. Fractions were analyzed by SDS-12.5% polyacrylamide gel electrophoresis; HMG D eluted from the column at approximately 250 mM NaCl and was free of contaminating proteins, as determined by silver staining of the gel.

Purified HMG D (as described above) was trichloroacetic acid precipitated, resuspended in 5% acetic acid containing pepsin at a concentration of 1/50 (wt/wt), and digested for 30 min at 37°C. Approximately 10 μg of the digested HMG D was trichloroacetic acid precipitated and resuspended in sample buffer; the peptides were separated by urea–SDS-12.5% polyacrylamide gel electrophoresis (25) and then transferred to Immobilon (Millipore), using 0.1 M H3PO4–15% methanol as the transfer buffer. The membrane was treated essentially as described by Matsudaira (44); the largest of the peptides was cut out for amino acid sequencing. The peptide was sequenced directly from Immobilon, using an Applied Biosystems 470A protein sequencer.

Oligonucleotide synthesis and PCR. Two degenerate oligonucleotides (5'-NCCpUtTNGpCyTCPuAAPPYTT-3' and 5'- GARPCCNAApUGCGCNAApUC3') were synthesized.

A polymerase chain reaction (PCR) was performed in a 100-μl volume containing 50 mM KCl, 10 mM Tris-Cl (pH 8.3), 1.5 mM MgCl2, 0.01% gelatin, 200 μM (each) deoxyoligonucleotide triphosphate (dNTP), and 2.5 U of Taq polymerase (Perkin Elmer/Cetus) with 100 ng of cDNA library (0- to 4-h Drosophila embryo library) and 500 ng (0.1 nmol) of each of the primers. The following program was used: 94°C, 1 min; 41°C, 2 min; increased slowly (1 min) to 72°C, 3 min; repeated for 30 cycles. The reaction products were analyzed by 1× TBE (Tris-borate-EDTA)-10% polyacrylamide gel electrophoresis.

cDNA clone isolation and sequencing. A 0- to 4-h cDNA library was screened essentially as described by Brown and Kafatos (8), using the 65-bp PCR-amplified product obtained with the degenerate oligomers. This 65-bp fragment was labelled with α-32P-labelled dNTPs, using PCR under the conditions described above. Hybridization conditions were as follows: 0.5 M Na2HPO4, 1% bovine serum albumin, 7% SDS, 1 mM EDTA (13), and 50 μg of salmon sperm DNA with the 65-bp PCR α-32P-labelled fragment. Hybridization was performed at 65°C for 12 to 24 h. The filters were washed three times for 15 min each time at 65°C with 40 mM Na2HPO4-1% SDS and exposed to film. Positive clones were picked and re-screened as described above until single colonies were obtained.

DNA was prepared from the single colonies and digested with EcoRI and HindIII to determine the size of each insert. Each insert was subcloned into either mp18, mp19, pUC118, or pUC119. The nucleotide sequences of both strands of the 0.95-kb cDNA and one strand of the 1.3-kb cDNA were determined by the ddNTP chain termination reaction (51), using Sequenase (U.S. Biochemical).

DNA sequence analysis. Sequences were compiled and analyzed by using the Genetics Computer Group computer software programs. The EMBL/GenBank data bases were searched for homologies to the DNA and protein sequences of HMG D.

Preparation of genomic DNA and Southern blot analysis. Adult flies (0.1 g) were homogenized in a ground glass homogenizer in 2 ml of grinding buffer (0.1 M Tris [pH 9.1], 0.2 M sucrose, 50 mM EDTA, 0.5% SDS). The suspension was heated for 30 min at 65°C, 0.3 ml of 8 M potassium acetate was added, and the suspension was placed on ice for 30 min. The insoluble material was pelleted by spinning at 20,000 × g for 15 min at 4°C. The supernatant was extracted with phenol-chloroform; 1 volume of ethanol was added to

Downloaded from http://mcb.asm.org on December 18, 2017 by guest
the aqueous layer, the sample was left at room temperature for 5 min, and the DNA was pelleted by spinning at 20,000 ×


g for 10 min at 4°C. The pellet was rinsed with 70% ethanol


and allowed to drain dry.

Approximately 7-μg samples of genomic DNA were di-
gested with various restriction enzymes and size separated


by electrophoresis on a 1% agarose gel; the DNA was then


transferred to Nytran (Schleicher & Schuell). cDNA inserts


were labelled with [γ-32P]ATP (15). Hybridization and


washing conditions were as described above.

Chromosomal in situ hybridization. In situ hybridization to


salivary gland polytene chromosomes was carried out as
described by Cai et al. (10), using the 1.3-kb HMG D cDNA


labelled with biotinylated dCTP.

Preparation of RNA and Northern (RNA) blot analysis.


Total RNA from different developmental stages of D. melano-
gaster was prepared by hot phenol-chloroform extraction


(34). Approximately 5 μg of the staged RNA was size


separated by electrophoresis on a 1.2% agarose–2% formal-
dehyde gel and transferred to Nytran. The 65-bp HMG D


fragment obtained from PCR as described above was la-


belled with α-32P-labelled dNTPs, using PCR under the same


conditions. Hybridization and washing conditions were as


previously described. To control for equal loadings, the filter


was simultaneously hybridized with the [α-32P]dCTP-la-


belled RP49 clone (58).

Developmental protein profile. Nuclei were isolated from
different embryonic stages of D. melanogaster, and nuclear


proteins were extracted with 5% PCA as described above.
Separate samples of larvae, pupae, adult females, and adult


males were frozen and ground in liquid nitrogen, and total


proteins were extracted with 5% PCA. The insoluble mate-


rial was removed by centrifugation, and the soluble proteins


were precipitated by trichloroacetic acid (final concentra-


tion, 20%). The samples could not be standardized for equal


loading. The proteins were resuspended in 10 μl of sample


buffer (0.1 M HCl, 9 M urea, 0.715 M β-mercaptoethanol,


0.02% pyronin Y) and size separated by acetic acid-urea-


polyacrylamide gel electrophoresis; the proteins were visu-


alized by staining the gel with Coomassie blue.

Gel retardation assays. The histone H3/H4 S/X promoter


fragment (17) was labelled with [γ-32P]ATP and T4 poly-


nucleotide kinase. Column-purified HMG D was added to a


12.5-μl reaction mixture containing approximately 2 ng of


labelled DNA, 5 mM HEPES (pH 7.9), 0.2 mM dithiothre-


itol, 0.02 mM EDTA, 5% glycerol, and between 70 and 120


mM NaCl and was incubated for 15 min at room tem-


terature. The reaction products were loaded directly onto


a preelectrophoresed, 4% nondenaturing, 25 mM TBE–poly-


acrylamide gel. After electrophoresis for 2 h at 200 V at 4°C,


the gel was dried and exposed for autoradiography.

RESULTS

Drosophila melanogaster contains one abundant HMG pro-


tein, HMG D. Our initial search for Drosophila HMG pro-


teins was based on the biochemical characteristics that define


HMG proteins in higher eukaryotes: molecular


weight, solubility properties, and amino acid composition.
Nuclei from 6- to 18-h D. melanogaster embryos were iso-


lated (17), and nuclear proteins were extracted by homog-


enizing the nuclei in one of four reagents known to extract


HMG proteins from vertebrate organisms: 0.35 M NaCl (32),


5% PCA, (32), 10 mM spermine (24), and 10 mM EtBr (52).


Each of these extracts contained one low-molecular-weight


protein which migrated identically, as analyzed by several
different gel systems, including acetic acid-urea-polyacryl-


amide gel electrophoresis (Fig. 1A), SDS-polyacrylamide gel


electrophoresis (Fig. 1B), and Triton X-100-acetic acid-


urea-polyacrylamide gel electrophoresis (data not shown).


This abundant protein did not comigrate with any of the


known histones (Fig. 1A), and it is the only protein recov-


ered by all four extraction procedures. HMG D has an


estimated molecular weight of 14,500, as determined by


SDS-urea-polyacrylamide gel electrophoresis (data not


shown) (39).

To establish that the putative HMG D protein identified


from the different extracts was the same protein, we deter-


mined the amino acid compositions of the protein recovered


from each of the extraction procedures. Nuclear protein


extracts were prepared by using either 10 mM spermine, 10


mM EtBr, or 5% PCA and were subjected to SDS-polyacryl-


amide gel electrophoresis. The gel was stained with


Coomassie blue, and the putative HMG D bands from the
different extracts were cut out of the gel. The proteins were


electroeluted from the gel (28) and subjected to acid hyd-


rosis. The amino acid compositions from these three bands


were the same and confirmed the identification of HMG D.


HMG D has an amino acid composition similar to that of


previously identified HMG proteins, with 24% total acidic


amino acids and 24% total basic amino acids. HMG D


appears to be the same protein as the previously reported


A13 (6). We have classified this protein as a member of the


HMG protein family on the basis of its solubility properties,
molecular weight, and amino acid composition.


Using column chromatography, we purified HMG D. Figu-


re 2A displays a summary of the purification protocol


(see Materials and Methods for details), and Fig. 2B is a


Coomassie-stained SDS-polyacrylamide gel showing the ex-


tent of purification of HMG D through each step in the


procedure. HMG D elutes from the Mono Q column in two


peaks, suggesting a variation in the charge of the protein, as


would occur with some types of posttranslational modifica-
tion (such as phosphorylation). When two-dimensional gel electrophoresis (isoelectrofocusing followed by SDS-polyacrylamide gel electrophoresis) was performed on the two peak fractions containing HMG D, several (at least three) different-charge isomers of the protein were seen (data not shown).

HMG D is encoded by a single-copy gene and has amino acid sequence similarity to bovine HMG 1. To determine the amino acid sequence of HMG D, a cDNA clone was obtained. Purified HMG D was partially digested with pepsin to generate proteolytic peptides. The peptides were separated by SDS-urea-polyacrylamide gel electrophoresis and were transferred to nylon membrane for direct sequencing. Edman degradation of the largest of these proteolytic fragments yielded a peptide sequence of 33 amino acids (Fig. 3).

On the basis of the peptide sequence obtained, two degenerate oligonucleotides (Fig. 3) were synthesized for use in a PCR. The template for the reaction was a 0- to 4-h Drosophila embryo cDNA library (8). The expected product of the correctly amplified DNA fragment would be 65 bp. When the products of the PCR were separated by polyacrylamide gel electrophoresis, the expected 65-bp fragment was seen. This fragment was eluted from the gel, subcloned, and sequenced. The deduced peptide sequence from the obtained DNA sequence corresponded exactly to the previously determined amino acid sequence.

We then screened the same cDNA library to obtain a full-length cDNA encoding the HMG D protein, using the 65-bp fragment as a probe. Of three colonies showing strong hybridization, two had an insert of 0.95 kb and the other had an insert of 1.3 kb.

The 0.95-kb insert was sequenced completely (Fig. 3). We identified an open reading frame which had a translation initiation codon in a good context based on the consensus sequences put forth by Cavener (12) for Drosophila nuclear genes. This open reading frame encodes a protein with a molecular weight of 12,300, which corresponds well to the estimated molecular weight of 14,500 (given the highly charged nature of the protein) for the isolated HMG D protein. When the deduced protein sequence was compared

---

**FIG. 2.** Purification of HMG D. (A) Strategy used to purify HMG D. PAGE, polyacrylamide gel electrophoresis. (B) Coomassie blue-stained SDS-polyacrylamide gel showing the extent of purification of HMG D at each step in the purification procedure. Lanes: 1, low-molecular-weight protein standards (Bethesda Research Laboratories); 2, total PCA nuclear extract; 3, pooled fractions from the Mono S column containing HMG D; 4, fraction from the Mono Q column containing only HMG D.

**FIG. 3.** Nucleotide sequence and deduced amino acid sequence of the 0.95-kb HMG D cDNA. The underlined amino acids represent the peptide sequence obtained from one of the pepsin digest products of purified HMG D. The two arrows under the italicized amino acids correspond to the peptide sequences used in making the two degenerate oligonucleotides. The arrows are 5' to 3'.
with sequences of known proteins, two proteins, bovine HMG 1 (22) and yeast NHPP6 (35), were identified as having significant similarity to HMG D, 36 and 33% amino acid identity (if one considers conservative changes, 54 and 50%, respectively (Fig. 4). In addition, a previously identified DNA-binding motif, referred to as the HMG box (30), is present in the HMG D protein sequence (indicated in Fig. 4).

Partial DNA sequencing of the 1.3-kb insert demonstrated the same 5' end and protein-coding region as in the 0.95-kb insert but showed an extended 3' untranslated region, suggesting alternative polyadenylation (data not shown). There is a good polyadenylation signal (16) in both clones 17 bp upstream of the poly(A) tail.

Southern analysis of D. melanogaster genomic DNA, using either the 0.95-kb HMG D cDNA (data not shown) or the 1.3-kb HMG D cDNA as a probe, indicates that HMG D is encoded by a single-copy gene (Fig. 5). The two bands in the BglII digest are anticipated, since there is a BglII site within the cDNA. Two or three substantially weaker hybridization signals in each digest are consistently observed (using stringent conditions), suggesting that HMG D may be a member of a small group of related genes.

The chromosomal location of the HMG D gene was determined by in situ hybridization to the polytene chromosomes of larval salivary glands (10), using the 1.3-kb HMG D cDNA as a probe; the signal was mapped to region 57F8-11 on the right arm of chromosome 2 (Fig. 6).

**HMG D mRNA is expressed predominantly in the embryo, while the HMG D protein is present in all developmental stages of D. melanogaster.** Total RNA from sequential developmental stages of D. melanogaster was isolated, electrophoresed, transferred to Nytran, and hybridized with either the 65-bp PCR product (Fig. 7A), the 1.3-kb cDNA (data not shown), or the 0.95-kb cDNA (data not shown). In each case, the same two transcripts can be seen throughout embryonic development, with the peak of mRNA levels occurring between 8 and 10 h. The sizes of the two transcripts correspond well to the sizes of the isolated cDNAs encoding HMG D. The relative abundances of the two transcripts are similar in all samples except those from adult females, 0- to 2-h and 2- to 4-h samples, which show more of the smaller transcript. HMG D mRNA levels begin to decrease in late embryonic development and drop dramatically in the first-instar larval stage (although faint signal is detected in the larval and pupal stages), concomitant with the decrease in cell division (40). Neither HMG D mRNA is seen in the adult male, but both transcripts are present in the adult female. The observation that the transcripts are abundant in the adult female and in the 0- to 2-h embryo suggests maternal loading of the HMG D transcripts into the oocyte (41).

To examine the developmental profile of the HMG D protein, we obtained proteins soluble in 5% PCA from isolated nuclei (embryonic stages) or from whole organisms (larvae, pupae, and female and male adults). Proteins were electrophoresed and visualized by Coomassie blue staining. It was difficult to standardize for equal loading of the PCA protein extracts; however, the amount of histone H1 per haploid genome can be assumed to be approximately constant, which provides each lane with an internal control. The results (Fig. 7B) indicate that the HMG D protein is rela-
HMG D binds DNA nonspecifically but has a fivefold preference for A+T-rich DNA in vitro. All previously identified HMG proteins have been shown to be nonspecific DNA-binding proteins. HMG D preparations starting either from the 0.35 M NaCl extract or from the 5% PCA extract were used in these gel retardation experiments; similar results were obtained with both preparations. The labelled DNA used was a 442-bp fragment which includes the divergent Drosophila histone H3 and H4 gene promoters (17, 18).

When equal amounts of labelled DNA were incubated with decreasing amounts of HMG D, two kinds of DNA binding were evident. At high concentrations of HMG D, high-molecular-weight complexes were formed; at lower concentrations of HMG D, ladders were formed (Fig. 8A).

Several other DNA fragments used as probes, differing in length and sequence, gave similar results (data not shown); these included Xenopus ribosomal DNA promoter and enhancer fragments (46) and a 438-bp Sphi-EcoRI fragment (26) from the promoter of the Drosophila ftz gene.

FIG. 6. In situ localization of the HMG D gene. The biotinylated 1.3-kb HMG D cDNA hybridized to a single location at position 57F8-11 on the right arm of chromosome 2.

FIG. 7. Expression of HMG D throughout development. (A) Northern analysis. Total RNA from organisms at different stages as indicated above the lanes was isolated and size separated by electrophoresis on a denaturing agarose gel; the Nytran blot was probed with a 65-bp HMG D fragment labelled by PCR. To control for equal sample loading, the filter was simultaneously probed with a labelled RP49 clone. (B) Developmental protein profile. Nuclear PCA-extracted proteins (embryonic samples) and total PCA-extracted proteins (larvae, pupae, and adults) were electrophoresed on an acetic acid-urea gel, and the were proteins visualized by staining the gel with Coomassie blue. Histone H1 and HMG D are indicated with arrows. Different exposure times were necessary during printing to bring out detail in the samples from embryonic stages.

FIG. 8. DNA-binding studies with HMG D. (A) Decreasing amounts of NaCl-extracted, column-purified HMG D incubated with a constant amount (2 ng) of labelled histone H3/H4 promoter. (B) Increasing amounts of competitor DNA were incubated with a constant amount of both labelled histone H3/H4 promoter DNA (2 ng) and HMG D protein (4 ng). Lanes: 1 and 8, free DNA; 2 and 9, no competitor added; 3 to 7, 2, 10, 20, 100, and 200 ng, respectively, of poly(dA-dT) added as competitor; 10 to 14, 2, 10, 20, 100, and 200 ng, respectively, of poly(dI-dC) added as competitor. Different exposure times were necessary during printing to bring out the detail in lanes 2 (A and B) and 9 (B).
Since HMG D displayed this pattern of DNA binding irrespective of DNA sequence, the results suggested that HMG D binds with very little, if any, sequence specificity. Competition experiments were performed to test the question more directly. Increasing amounts of competitor DNA, either poly(dA-dT) or poly(dI-dC) (Fig. 8B), were incubated with a constant amount of labelled DNA. The binding of HMG D to the histone H3/H4 promoter was competed for by a 10-fold excess of poly(dA-dT), whereas a 50-fold excess of poly(dI-dC) was necessary to abolish the binding of HMG D to the promoter fragment. This result indicates that HMG D has a slight preference for, but is not limited to, A+T-rich DNA in vitro.

**DISCUSSION**

We have isolated a *Drosophila* HMG protein and its cDNAs. This protein, HMG D, possesses properties similar to those of the two major groups of vertebrate HMG proteins: its small size and selective extraction with EtBr relate it to the HMG 14 and 17 group, while its amino acid sequence is similar to that of HMG 1. HMG D is the only abundant protein consistently extracted from *Drosophila* nuclei when the classical reagents known to extract vertebrate HMG proteins (0.35 M NaCl, 10 mM EtBr, 10 mM spermine, and 5% PCA) are used and is therefore most likely the major HMG protein in *D. melanogaster*. HMG D appears to be the same protein as the previously reported A13 (6), as judged by size, solubility, and amino acid composition. The previously identified D1 protein from *D. melanogaster* (2, 5) is extracted with NaCl and PCA (clearly seen as the largest protein in the PCA extract of Fig. 1B) but is much larger than the vertebrate HMG proteins. The D1 protein has sequence similarity with HMG I but no sequence similarity to HMG D. It is worth noting that HMG D can be extracted from nuclei with spermine, a component of the Hewish and Burgoyne (27) buffer which has been often used to isolate *Drosophila* nuclei. Even at the low concentrations of spermine and spermidine included in that buffer (0.15 mM spermine and 0.15 mM spermidine), we have found that HMG D is almost completely extracted (data not shown). This may explain the failure of previous attempts to obtain HMG-like proteins from *D. melanogaster*.

Two-dimensional gel electrophoresis shows several (at least three) charge isoforms of HMG D; we do not know the type of modification(s) present or how they might relate to the function of HMG D in vivo. HMG proteins from vertebrates are known to be highly modified (3), but the role of these modifications in the function of the proteins is not understood. This is clearly an area that warrants further study.

The conclusion that HMG D is the only abundant HMG protein in *D. melanogaster* is supported by two additional observations. We have screened a *Drosophila* cDNA library at low stringency, using a chicken HMG 17 open reading frame sequence (37) as the probe. No homologous cDNA clones were recovered (59), suggesting that *D. melanogaster* does not have a protein that is similar to the vertebrate HMG 14 and 17 group by the criteria of DNA sequence. There is also no evidence for the existence of an HMG protein from *D. melanogaster* with a molecular weight similar to that of the vertebrate HMG 1 and 2 class: the major band migrating on an SDS-polyacrylamide gel between HMG D and histone H1 in the PCA extract was electroeluted from the gel, sequenced, and found to be a breakdown product of histone H1 (59). However, the presence of weakly hybridizing signals in the Southern blot leaves open the possibility that HMG D is a member of a small group of related genes.

The presence of two transcripts is puzzling, since the protein coding region is unaffected. It may be that the transcripts have different translation efficiencies in different cell types. It remains to be determined why the 0.95-kb HMG D transcript is more abundantly expressed than is the 1.3-kb transcript in adult females and in 0- to 4-h-old embryos. The presence of abundant HMG D mRNA in adult females and 0- to 2-h-old embryos suggests maternal pools that the developing embryo draws upon during early embryogenesis, when nuclear division is very rapid and little or no RNA synthesis is occurring.

A family of transcription factors that have in common a DNA-binding motif with similarity to HMG I, the so-called HMG box, has recently been identified (30). Some of these transcription factors are very specific in their pattern of tissue expression and in their biological role; for example, T-cell-specific transcription factor TCF-1a, present only in the thymus, has an HMG box that directs binding to 5'PyCTTTG-3' (60). In contrast, UBF, a nucleolar transcription factor with an HMG box, binds preferentially to G+C-rich regions of DNA (30). It appears that the HMG box indicates a common ancestor from which UBF and other specific DNA-binding structures have evolved (30). The major vertebrate HMG proteins show relatively little sequence specificity in binding to DNA, but a strong preference for a sequence of six or more consecutive A+T base pairs has been reported for the mammalian HMG I (57).

In this case, we have observed two types of DNA binding which are dependent on the concentrations of HMG D in the binding reaction. When a low concentration of HMG D is used, ladders of protein-DNA complex are seen. A high-molecular-weight complex is seen when higher concentrations of HMG D are used. The results reported here for purified *Drosophila* HMG D indicate no DNA sequence specificity for binding but show that HMG D has a fivefold preference for A+T-rich sequences in vitro. The gene encoding HMG D has also been identified by screening an expression library for protein binding with an A+T-rich DNA fragment (45a). The DNA fragment used in that screen was a 50-bp A+T-rich sequence in the upstream element of the *Drosophila* *ftz* gene.

Two-dimensional gel electrophoresis shows several (at least three) charge isoforms of HMG D; we do not know the type of modification(s) present or how they might relate to the function of HMG D in vivo. HMG proteins from vertebrates are known to be highly modified (3), but the role of these modifications in the function of the proteins is not understood. This is clearly an area that warrants further study.

The conclusion that HMG D is the only abundant HMG protein in *D. melanogaster* is supported by two additional observations. We have screened a *Drosophila* cDNA library at low stringency, using a chicken HMG 17 open reading frame sequence (37) as the probe. No homologous cDNA clones were recovered (59), suggesting that *D. melanogaster* does not have a protein that is similar to the vertebrate HMG 14 and 17 group by the criteria of DNA sequence. There is also no evidence for the existence of an HMG protein from *D. melanogaster* with a molecular weight similar to that of the vertebrate HMG 1 and 2 class: the major band migrating on an SDS-polyacrylamide gel between HMG D and histone H1 in the PCA extract was electroeluted from the gel, sequenced, and found to be a breakdown product of histone H1 (59). However, the presence of weakly hybridizing signals in the Southern blot leaves open the possibility that HMG D is a member of a small group of related genes.

The presence of two transcripts is puzzling, since the protein coding region is unaffected. It may be that the transcripts have different translation efficiencies in different cell types. It remains to be determined why the 0.95-kb HMG D transcript is more abundantly expressed than is the 1.3-kb transcript in adult females and in 0- to 4-h-old embryos. The presence of abundant HMG D mRNA in adult females and 0- to 2-h-old embryos suggests maternal pools that the developing embryo draws upon during early embryogenesis, when nuclear division is very rapid and little or no RNA synthesis is occurring.

A family of transcription factors that have in common a DNA-binding motif with similarity to HMG I, the so-called HMG box, has recently been identified (30). Some of these transcription factors are very specific in their pattern of tissue expression and in their biological role; for example, T-cell-specific transcription factor TCF-1a, present only in the thymus, has an HMG box that directs binding to 5'PyCTTTG-3' (60). In contrast, UBF, a nucleolar transcription factor with an HMG box, binds preferentially to G+C-rich regions of DNA (30). It appears that the HMG box indicates a common ancestor from which UBF and other specific DNA-binding structures have evolved (30). The major vertebrate HMG proteins show relatively little sequence specificity in binding to DNA, but a strong preference for a sequence of six or more consecutive A+T base pairs has been reported for the mammalian HMG I (57).

In this case, we have observed two types of DNA binding which are dependent on the concentrations of HMG D in the binding reaction. When a low concentration of HMG D is used, ladders of protein-DNA complex are seen. A high-molecular-weight complex is seen when higher concentrations of HMG D are used. The results reported here for purified *Drosophila* HMG D indicate no DNA sequence specificity for binding but show that HMG D has a fivefold preference for A+T-rich sequences in vitro. The gene encoding HMG D has also been identified by screening an expression library for protein binding with an A+T-rich DNA fragment (45a). The DNA fragment used in that screen was a 50-bp A+T-rich sequence in the upstream element of the *Drosophila* *ftz* gene.
used in establishing the correct chromosomal context, allowing
the transcriptional machinery to function properly (36).

What might be the role of HMG D, an abundant DNA-
binding protein? Wustemann et al. (62) have shown that when
the region 57D-58D, containing the HMG D gene, is present
in three copies, one sees a suppression of position effect
variegation in D. melanogaster. This finding indicates that
the chromatin of loci susceptible to position effect in a strain
carrying that duplication is being packaged in a form more
accessible to transcription. The idea that HMG D may be
involved in packaging the DNA at a level that makes a
distinction between euchromatin and heterochromatin is an
intriguing one. The abundance of the HMG D transcripts
only in early development reflects a common pattern seen
for transcripts of other proteins involved in DNA packaging,
most notably the histones and their variants (4). The obser-
vation that the HMG D protein is relatively stable is consist-
tent with a structural role. P-element transformation with the
genomic clone of HMG D (work in progress) should deter-
mine whether the suppression of position effect is caused by
the HMG D gene product.

Continued genetic and biochemical analysis of the HMG D
protein should allow us to increase our understanding of the
role of HMG D in the Drosophila nucleus and the role of
HMG proteins in general in eukaryotic chromatin.

ACKNOWLEDGMENTS
We are grateful to Sally Ameo for help in the early stages of
the project and to Peter Burgers, Howard Granok, John Majors,
and Charlie Rice for technical advice as well as helpful discussions. We
thank Linda Pike and the WUMS Howard Hughes Protein Facility for
amino acid sequencing, and we thank Haini Cai, Joyce Yee, and
Ian Duncan for help with the in situ hybridization. We are also
grateful to Robert Landick, John Majors, and the members of the
Elgin laboratory for critical reading of the manuscript, and we thank
Carolyn Craig for photographic assistance. We thank Sarb Ner and
Andrew Travers (MRC Laboratory, Cambridge, United Kingdom) for
communication and discussion of their results prior to publica-
tion.

This research was supported by NIH grant GM31532 (S.C.R.E.)
and by an NIH predoctoral training grant 5-T32-GM07067 (C.R.W.).

REFERENCES
1. Abercrombie, B. D., G. G. Kneale, C. Crane-Robinson, E. M.
1978. Studies on the confirmational properties of the
high-mobility group chromosomal protein HMG17 and its
Isolation, properties, and cellular distribution of DI, a chromo-
Johns (ed.), The HMG chromosomal proteins. Academic Press,
New York.
sion in Drosophila development: multiple levels of gene regula-
tion, p. 135–161. In G. S. Stein and J. S. Stein (ed.), Histone
genes: structure, organization and regulation. John Wiley &
Sons, New York.
5. Ashley, C. T., C. G. Pendleton, W. W. Jennings, A. Saxena,
and C. V. C. Glover. 1989. Isolation and sequencing of cDNA clones
encoding Drosophila chromosomal protein DI. J. Biol. Chem.
264:8393–8401.
group proteins of Drosophila melanogaster embryo nuclei.
recognition of cruciform DNA by nuclear protein HMG I.
features of the HMG chromosomal proteins and their genes.
lished data.
11. Cary, P. D., D. S. King, C. Crane-Robinson, E. M. Bradbury,
studies on two high mobility group proteins from calf thymus,
HMG14 and ubiquitin, and their interaction with DNA. Eur.
J. Biochem. 112:577–588.
flanking translational start sites in Drosophila and vertebrates.
and J. T. Kadonaga. 1991. Sequence-specific antirepression of his-
tone H1-mediated inhibition of basal RNA polymerase II tran-
radiolabeling DNA restriction endonuclease fragments to high
AAUAAA-3' forms part of the recognition site for polyaden-
box-dependent protein-DNA interactions are detected on heat
shock and histone gene promoters in nuclear extracts derived
structure and function, p. 193–218. In E. W. Johns (ed.), The
An improved large scale fractionation of high mobility group non-
291.
group of chromatin-associated proteins with a high content of
mobility group (HMG) non-histone chromosomal proteins, p.
Press, New York.
cerevisiae ACP2 gene encodes an essential HMG-like protein.
chromosomal proteins also exist in Tetrahymena. J. Biochem.
86:789–794.
electrophoresis, p. 1–86. In D. B. Hanes (ed.), Gel electropho-
binding sites for potential regulatory proteins in the upstream
enhancer element of the Drosophila fushi tarazu gene. Nucleic
Acids Res. 16:11403–11416.
ture: the digestion of chromatin DNA at regularly spaced sites
by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Com-
mun. 52:504–510.
1983. Isolation of microgram quantities of proteins from poly-
crylamide gels for amino acid sequence analysis. Methods
Enzymol. 91:227–236.
1979. Preferential affinitiy of high molecular weight high mobility
group non-histone chromatin proteins for single-stranded DNA.
46. Ner, S., and A. Travers. Personal communication.