MOLECULAR AND CELLULAR BIOLOGY, Jan. 1991, p. 166–174
0270-7306/91/010166-09$02.00/0

Vol. 11, No. 1

Macronuclei and Micronuclei in *Tetrahymena thermophila* Contain High-Mobility-Group-Like Chromosomal Proteins Containing a Highly Conserved Eleven-Amino-Acid Putative DNA-Binding Sequence

IRA G. SCHULMAN,[1]† TONGTONG WANG,[1] MIN WU,[2] JOSEPHINE BOWEN,[2] RICHARD G. COOK,[3] MARTIN A. GOROVSKY,[2] AND C. DAVID ALLIS[1]*

Department of Biology, Syracuse University, Syracuse, New York 13244-1220[1]; Department of Immunology, Baylor College of Medicine, Houston, Texas 77030[3]; and Department of Biology, University of Rochester, Rochester, New York 14627[2]

Received 19 September 1990/Accepted 2 October 1990

HMG (high-mobility-group protein) B and HMG C are abundant nonhistone chromosomal proteins isolated from *Tetrahymena thermophila* macronuclei with solubilities, molecular weights, and amino acid compositions like those of vertebrate HMG proteins. Genomic clones encoding each of these proteins have been sequenced. Both are single-copy genes that encode single polyadenylated messages whose amounts are 10 to 15 times greater in growing cells than in starved, nongrowing cells. The derived amino acid sequences of HMG B and HMG C contain a highly conserved sequence, the HMG 1 box, found in vertebrate HMGs 1 and 2, and we speculate that this sequence may represent a novel, previously unrecognized DNA-binding motif in this class of chromosomal proteins. Like HMGs 1 and 2, HMGs B and C contain a high percentage of aromatic amino acids. However, the *Tetrahymena* HMGs are small, are associated with nucleosome core particles, and can be specifically extracted from macronuclei by elutive intercalation, properties associated with vertebrate HMGs 14 and 17, not HMGs 1 and 2. Thus, it appears that these *Tetrahymena* proteins have features in common with both of the major subgroups of higher eucaryotic HMG proteins. Surprisingly, a linker histone found exclusively in transcriptionally inactive micronuclei also has several HMG-like characteristics, including the ability to be specifically extracted from nuclei by elutive intercalation and the presence of the HMG 1 box. This finding suggests that at least in *T. thermophila*, proteins with HMG-like properties are not restricted to regions of transcriptionally active chromatin.

High-mobility-group (HMG) proteins are abundant nonhistone chromosomal proteins that characteristically contain a high percentage of basic and acidic amino acids and are easily extracted from nuclei with low concentrations of salt or dilute acids (21). In higher eucaryotes, HMG proteins have been subdivided into three main groups based primarily on differences in molecular weight, amino acid composition, and primary sequence. The first subgroup is composed of two relatively large HMG proteins, HMGs 1 and 2 (25 to 29 kDa). Potential roles for HMGs 1 and 2 in transcription, DNA replication, chromatin assembly, and recombination have been inferred from in vitro studies (for reviews, see references 21 and 43). Recent studies (39a) have indicated that HMGs 1 and 2 may be required for transcription of class II genes in vitro. However, the functions of HMGs 1 and 2 in vivo remain in question.

Two smaller proteins, HMGs 14 and 17 (7 to 14 kDa), represent a well-studied second subgroup. Both proteins can interact directly with the nucleosome core particle, and several groups have reported that HMGs 14 and 17 are specifically associated with nucleosome particles in transcriptionally active or transcriptionally competent chromatin (for reviews, see references 13, 31, and 43). These results, however, are still controversial.

More recently, HMG I (or α protein), a fifth HMG protein that appears distinct from the other HMG proteins, has been identified in a number of vertebrate cells (22, 40). D1, a *Drosophila* heterochromatin-associated protein (1), has recently been shown to be structurally related to this class of chromosomal proteins (6). HMG I and D1 bind preferentially to A+T-rich satellite DNA in vitro (10, 24, 32, 34), but the significance of their affinity for A+T-rich DNA and their functions in vivo are not clear.

The ciliated protozoan *Tetrahymena thermophila* provides a model eucaryotic system in which to study the structure and function of HMG proteins. Like most ciliates, *T. thermophila* contains two nuclei: a transcriptionally active macronucleus which governs the phenotype of the cell and a transcriptionally silent micronucleus which acts as a germ line nucleus. If HMG proteins are specifically associated with transcriptionally active or competent chromatin as has been suggested for several of the vertebrate HMG proteins, one would expect to find these proteins in transcriptionally active macronuclei but not in transcriptionally inactive micronuclei.

We and others have previously reported the existence of two abundant nonhistone chromosomal proteins, HMGs B and C (Fig. 1), isolated from *Tetrahymena* macronuclei which have solubility properties and amino acid compositions characteristic of higher eucaryotic HMG proteins (15, 17, 25, 33, 39). HMGs B and C resemble vertebrate HMGs 14 and 17 in size and in the ability to interact with nucleosome core particles (15, 25, 38, 39). They also share with HMGs 14 and 17 the ability to be specifically extracted from

* Corresponding author.
† Present address: Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280.

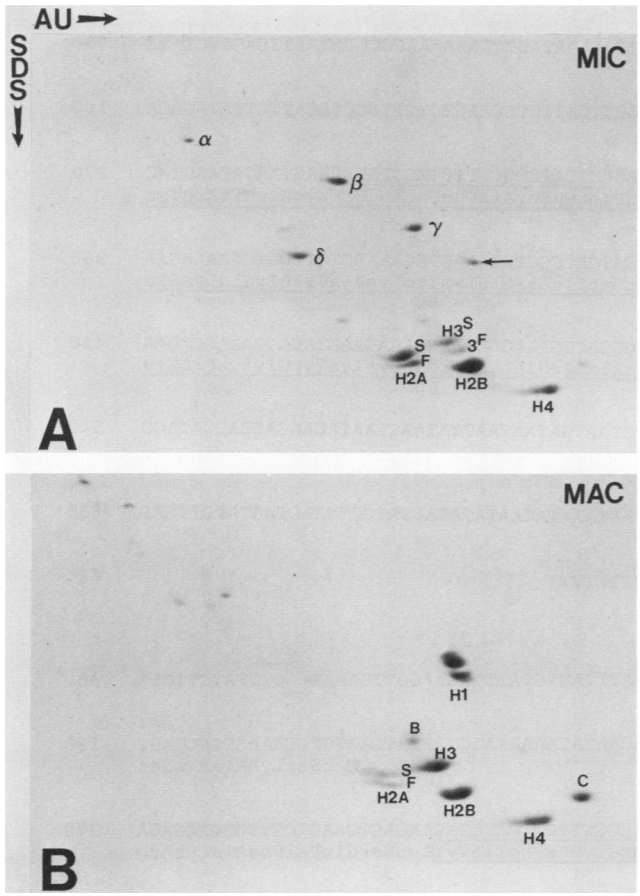

FIG. 1. Two-dimensional electrophoretic analysis of macro- and micronuclear total acid-soluble protein. Two-dimensional electrophoretic analysis (acid-urea by SDS) of micronuclear (MIC) total acid-soluble protein (A) and macronuclear (MAC) total acid-soluble protein (B). HMG B and HMG C were originally described as LG2 and LG1, respectively (15), but were later renamed (25). The arrow in panel A points to contaminating histone H1 in the micronuclear preparation. Note that appreciable amounts of delta (roughly equal to those of beta and gamma) are recovered when micronuclei are isolated in buffers lacking spermidine (compare this figure with Fig. 1 of reference 5). Histone variants hv1 and hv2 (4) are not observed because the first-dimension gel lacks the detergent Triton.

nuclei with intercalating drugs such as ethidium bromide (EtBr) (39), suggesting that they bind near the ends of core particle DNA (28).

To further study the structures and functions of HMG-like proteins, we have cloned and sequenced the genes that encode HMGs B and C. These genes are single copy, and Northern (RNA) analyses show that the abundance of both HMG B and HMG C messages is 10- to 15-fold greater in growing cells than in starved, nondividing cells. As for other HMGs, the derived sequences indicate that these proteins contain a high percentage of basic and acidic amino acids. HMGs B and C also share an 11-amino-acid sequence homology with HMGs 1 and 2. Quite unexpectedly, this sequence is also present in delta, one of the stable, linker-associated proteins found in micronuclei (Fig. 1; 49). Interestingly, delta, like HMGs B and C, is soluble in 5% perchloric acid (PCA) and can be specifically extracted from micronuclei with intercalating agents or with spermidine-containing buffers. These observations suggest the possibility that an HMG-like polypeptide is associated with linker regions of transcriptionally inactive chromatin. This surprising result causes us to question the relationship between HMG proteins and transcriptionally competent chromatin.

MATERIALS AND METHODS

Cell culture and labeling. Genetically marked strains of *T. thermophila*, CU 427 (Mpr/Mpr[6-mp-s]VI) and CU 428 (Chx/Chx[cy-s]VII), were used in all experiments. These strains were kindly provided by P. Bruns (Cornell University). Cells were grown axenically in 1% (wt/vol) enriched proteose peptone as described by Gorovsky et al. (12). All cultures were maintained at 30°C.

Extraction of nuclear proteins and gel electrophoresis. Macronuclei and micronuclei were isolated from growing cells by the method of Gorovsky et al. (12), with the following changes: spermidine was omitted from all buffers, and 10 mM Tris (pH 6.8), 1 mM phenylmethylsulfonyl fluoride, 10 mM sodium butyrate, and 10 mM iodoacetamide were added to the nucleus isolation buffer (medium A). Acid extraction and precipitation of soluble proteins were performed as described by Schulman et al. (39).

For elutive intercalation, macronuclei or micronuclei were washed in buffer E (10 mM Tris [pH 7.0], 5 mM $MgCl_2$, 3 mM $CaCl_2$, 1 mM phenylmethylsulfonyl fluoride, 10 mM iodoacetamide, 10 mM sodium butyrate) before being resuspended in 20 mM EtBr in buffer E as described by Schröter et al. (37). After a 30-min extraction at 4°C, nuclei were pelleted and the supernatant was converted to 20% trichloroacetic acid. The remaining pellet was then extracted to produce PCA-soluble and -insoluble fractions as described above. Precipitated proteins were collected, washed, and dried. The effects of other intercalating drugs or buffer conditions were not examined.

The first-dimension (acid-urea or sodium dodecyl sulfate [SDS]) and second-dimension (SDS) gels used have been described previously (4, 5). All gels were stained with Coomassie blue.

Recombinant DNA methods. Standard methods were used for DNA isolation and Southern and Northern blotting (38). DNA sequencing was carried out by using Sequenase. All DNA fragments were sequenced twice in both directions.

Oligonucleotide hybridization. To clone the HMG B gene, two synthetic oligonucleotides were built from existing protein sequence information (39). One oligonucleotide was a single-sequence 36-mer (GATCCATGGAAGGAAAAGT ATGGTGATATCGAAAAG) constructed from the codon usage bias of previously sequenced *Tetrahymena* genes (27). This oligonucleotide covers amino acids 83 to 94 of the intact protein. Upon sequencing of the cloned gene, this oligonucleotide was found to be 92% homologous (33 of 36 nucleotides), with the last 21 nucleotides correct. The second oligonucleotide was a mixed 17-mer ($TA^T_CAGAAA^A_GGA^A_G AA^A_GGC$) covering a region beginning with amino acid 71 through the first two bases of the codon for amino acid 76. Optimum hybridization conditions for the oligonucleotides were determined by genomic Southern analyses (38). Under the conditions used, the single-sequence oligonucleotide hybridized to a single band in multiple restriction digests. The mixed oligonucleotide hybridized to between two and four bands per restriction digest; however, this oligonucleotide always hybridized to the single band detected by the unique oligonucleotide.

To clone the gene for HMG C, one unique-sequence

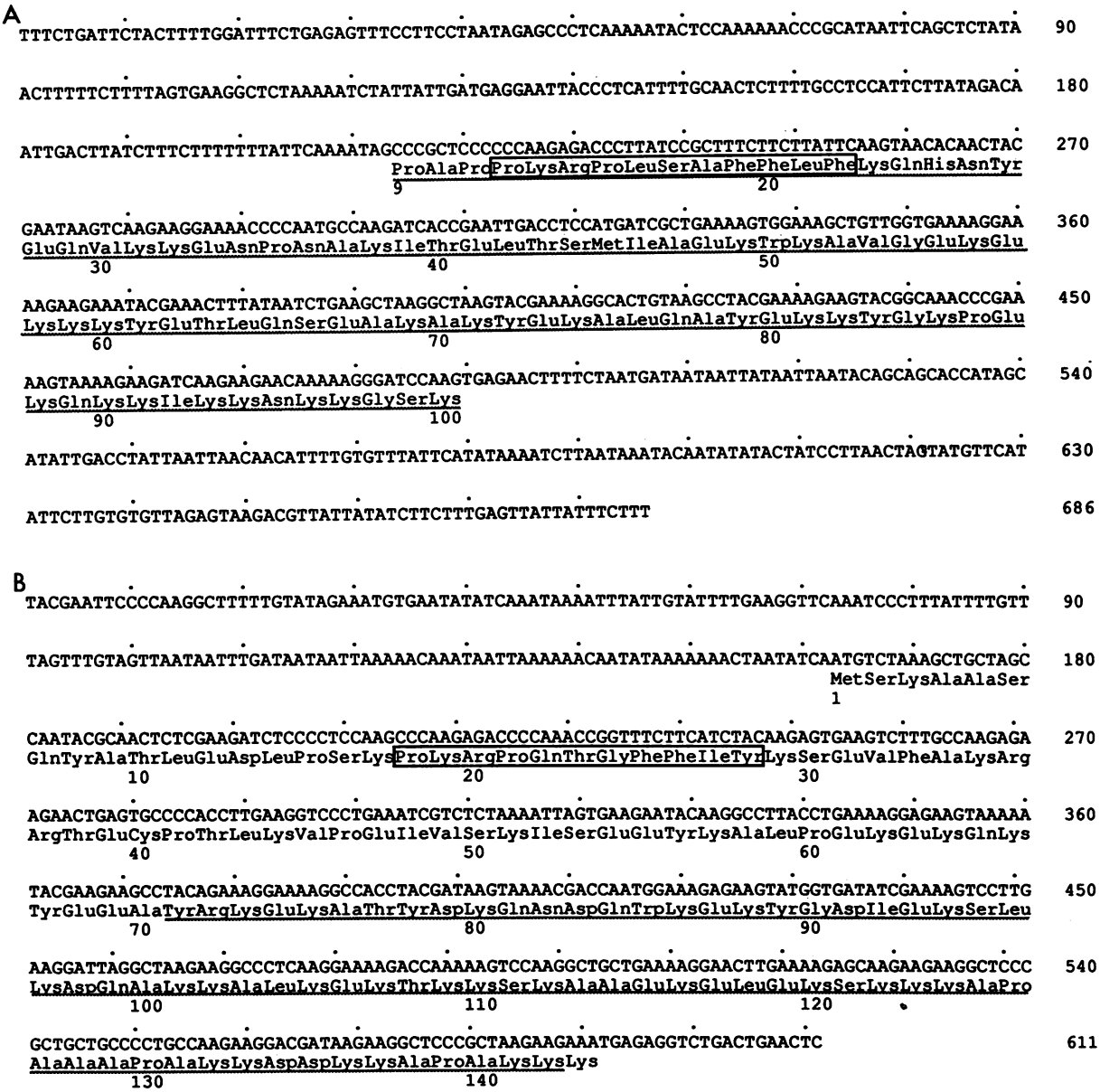

FIG. 2. DNA and derived protein sequences of the *T. thermophila* HMG B and C genes. Shown are the sequences of flanking and intron (HMG C only) DNA and the derived protein sequences of the HMG B and HMG C genes. Underlined residues in the HMG B sequence are those determined by direct protein sequencing; the complete protein sequence of HMG C has been published (33). Shown for both HMG B and C is the position of a highly conserved HMG 1 box (see text and Fig. 4 for details). Intron sequence (in HMG C) is found between nucleotides 1 and 213. Numbers at the right refer to nucleotides, and numbers below the sequences refer to amino acids.

39-mer (TTCCTTTTCACCGACAGCCTTCCACTTTTCAG CGATCAT) was constructed from the codon usage bias of previously sequenced *Tetrahymena* genes (27). This oligonucleotide covers amino acids 45 to 57 of the HMG C protein sequence (33). Upon sequencing of the cloned gene, this oligonucleotide was found to be 95% correct (37 of 39 nucleotides), with the first 12 and the last 20 nucleotides correct. Under the conditions used, this oligonucleotide hybridized to a single band in multiple restriction digests.

Size-selected library construction and screening. A size-selected genomic library was constructed by single or double-digestion of macronuclear DNA with the appropriate restriction enzyme(s). The digested DNA was then resolved in a 0.8% low-melting-point agarose gel, and the region of the gel containing the hybridizing fragment of interest was excised, extracted by using GeneClean, and ligated at an insert-to-vector ratio of 1:1 into the Bluescript vector (Stratagene, La Jolla, Calif.). Following transformation into competent XL1-blue cells (Stratagene), approximately 10,000 colonies (70 to 90% recombinant) were replica plated to nitrocellulose filters (approximately 2,000 colonies per filter) and hybridized with the appropriate oligonucleotides, using the conditions optimized for oligonucleotide hybridization.

RNA isolation, nuclease protection, and primer extension analyses. RNA was isolated from growing or starved cells by standard methods. RNase protection mapping was per-

formed as described by Melton et al. (29) and the Riboprobe system protocols provided by Promega Biotec. For mapping the 5′ ends of the HMG B transcripts, antisense ^{32}P-labeled RNA transcripts were prepared by using the Amersham SP6 system with plasmids containing the appropriate flanking region of HMG B (a *Hin*dIII/*Bgl*II restriction fragment which covers all of the 5′-flanking region shown in Fig. 2 as well as +37 nucleotides of coding information). Cold UTP was included in the reaction mixtures; transcripts were run on acrylamide gels, dried, and autoradiographed to assay production of full-length transcripts. Optimum RNase digestion conditions were empirically determined to be 4 μg of RNase A per ml and 90 U of RNase T_1 per ml at 0°C for 10 min.

Primer extension using reverse transcriptase of avian myeloblastosis virus was done as described by Ghosh et al. (11). The synthetic primer used (complementary to +10 to +34 of the HMG B transcript) was ^{32}P 5′ end labeled with T4 polynucleotide kinase and [^{32}P-γ]ATP and annealed to total RNA from growing cells in 0.1 M NaCl–20 mM Tris hydrochloride (pH 7.9)–0.1 mM EDTA at 37°C for 4 h. After hybridization, the reverse transcription reaction was conducted in a final volume of 50 μl of buffer, consisting of 50 mM Tris (pH 8.0), 10 mM dithiothreitol, 80 mM KCl, 6 mM $MgCl_2$, 0.5 mM each deoxynucleoside triphosphate, 50 μg of actinomycin D per ml, and 20 U of avian myeloblastosis virus reverse transcriptase (Life Sciences Inc., St. Petersburg, Fla.) at 45°C for 1.5 h. The extended products were ethanol precipitated and electrophoresed on 10% polyacrylamide–8 M urea gels.

RESULTS AND DISCUSSION

Isolation and identification of the HMG B and C genes. The genes encoding HMGs B and C were cloned by using oligonucleotide probes derived from the amino acid sequence of each. Figure 2 shows the DNA sequences and deduced protein sequences of the genes encoding *Tetrahymena* HMGs B and C. The available derived protein sequence of HMG C is in complete agreement with the published amino acid sequence of HMG C (33). However, because of the presence of an intron in the 5′ portion of the HMG C gene, the cloned genomic fragment is missing the information encoding 9 N-terminal amino acid residues of the protein (NH_2-AKSKDDSK; 33, 39). The derived protein sequence of HMG B contains 70 amino acids, also determined by microsequencing several internal fragments of purified HMG B (underlined in Fig. 2) The complete correspondence of these two sequences demonstrates that the cloned DNA fragments contain the genes for HMGs B and C. Southern blots using macronuclear DNA digested with a variety of restriction enzymes (Fig. 3) indicate that both HMG B and HMG C are encoded by single-copy genes, which is somewhat unusual for genes encoding abundant histone and nonhistone chromosomal proteins. In addition, the Southern hybridization pattern for micronuclear DNA was the same as that for macronuclear DNA (Fig. 3), indicating that both HMG B and HMG C genes are not associated with regions that are grossly rearranged in the development of the macronucleus from the micronucleus during the sexual cycle (see reference 7 for references).

It is likely that the amino terminus of HMG B begins with the ATG codon at bp 162. This ATG is the first start codon in frame with the known amino acid sequence of HMG B, and its placement as the amino-terminal residue is consistent with our inability to detect methionine in HMG B by CNBr cleavage or labeling in vivo with [^{35}S]methionine. The calculated molecular size of the protein initiated at this ATG (approximately 16.3 kDa) is in close agreement with the molecular size determined for HMG B by gel filtration chromatography (15 ± 0.5 kDa [15]), and the predicted amino acid composition agrees with the published amino acid composition of HMG B (15, 25). Removal of the initiator methionine is common in eucaryotes and in this case would produce a serine as the amino-terminal residue of the mature protein. Amino-terminal serines are often blocked by acetylation (42), and blocking of this serine would be consistent with our inability to sequence the amino terminus of HMG B.

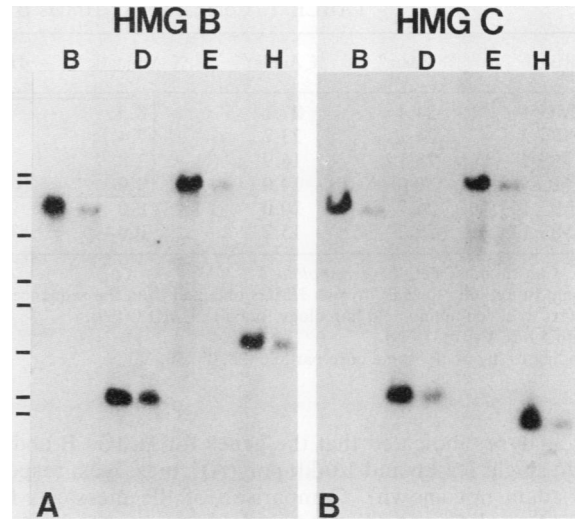

FIG. 3. Analyses showing that HMG B and HMG C are encoded by single-copy genes that are not developmentally rearranged. Macronuclear (left lane of each pair) and micronuclear (right lane of each pair) DNAs were digested with *Bgl*II (B), *Dra*I (D), *Eco*RI (E), and *Hin*dIII (H). Blots were hybridized with appropriate random-primer-labeled probes to HMG B (A) or HMG C (B), washed, and exposed to film with an intensifying screen for 12 h at −80°C. Lines on the left show positions of DNA marker fragments; sizes from top to bottom are 11.4, 10.5, 6.3, 4.4, 3.7, 2.6, 1.7, and 1.1 kb. Although the mobilities of *Bgl*II, *Dra*I, and *Eco*RI fragments from HMG B and HMG C are remarkably similar, appropriate double digests and digestion with other restriction enzymes (for example, *Bam*HI) indicate that the HMG B and HMG C genes are not closely linked.

There is only one other potential in-frame start codon (31) in 160 bp of upstream sequence. This ATG, however, is followed directly by an in-frame stop codon. Furthermore, nuclease protection and primer extension analyses both show a heterogeneous collection of major transcription start sites about 50 bp upstream of the proposed initiator ATG in HMG B (both techniques detect sites at −55, −50, −47, and −44; data not shown). Finally, the DNA sequence 5′ to the proposed initiator ATG is a very A+T rich, consistent with the high A+T character of *Tetrahymena* noncoding DNA (18). The proposed initiator ATG in HMG B, however, is not preceded by the putative translation start signal found in a number of *Tetrahymena* histone genes (18).

Northern blot analysis. To examine the messages for HMGs B and C, total RNA was isolated from growing cells, fractionated into poly(A)$^+$ and poly(A)$^-$ RNA by oligo(dT)-cellulose chromatography, and resolved on a denaturing agarose gel. The RNA was then transferred to nitrocellulose and hybridized with antisense RNA probes. Results from

TABLE 1. Comparison of HMGs B and C with calf thymus HMGs 1, 2, 14, and 17

HMG[a]	% Basic[b]	% Acidic[b]	% Aromatic[b]	HMG 1 box	HMG 14 and 17 homology[c]	Molecular size (kDa)	Elutive intercalation[d]	Nucleosome bound[e]
CT HMG 1	24.4	27.1	8.5	+	−	29.6	−	−
CT HMG 2	24.1	23.7	7.4	+	−	29.2	−	−
Tet HMG B	28.7	16.9	7.8	+	+	16.5	+	+
Tet HMG C	30.0	14.0	9.0	+	+	11.7	+	+
CT HMG 14	26.7	20.0	0.0	−	+	10.7	+	+
CT HMG 17	29.2	15.7	0.0	−	+	9.2	+	+

[a] CT, Calf thymus; Tet, *T. thermophila*.
[b] Composition data for calf thymus HMGs obtained from the sequences in reference 45.
[c] Refers to seven-amino-acid homology found in HMG C (39).
[d] With 5 mM EtBr (37, 39).
[e] Binding to the nucleosome core particle (26, 35, 36, 38).

these analyses indicated that the genes for HMGs B and C encode single 1.2-kb and 1.0-kb poly(A)$^+$ messages, respectively (data not shown). Comparison of the messages for HMGs B and C in total RNA isolated from growing and starved cells indicated that there was 10 to 15 times more B and C messages present in growing cells (data not shown), a result which correlates well with the difference in HMG B and C protein synthesis observed between growing and starved cells (38, 39). Since HMG B and HMG C message levels are partly correlated with cell growth (as in the case with histone message in *T. thermophila* [47, 50]), it seems reasonable to suspect that new synthesized HMGs B and C associate with newly replicated macronuclear DNA.

Compositional and sequence analyses. Analyses of the amino acid sequences of HMGs B and C (Fig. 2) indicate that both are relatively small (12 to 16 kDa) and highly charged (30% basic, ~15% acidic) proteins (Table 1). The ability to extract HMGs B and C from nuclei by elutive intercalation (39), their interactions with nucleosome core particles (25, 38), and their smaller sizes suggest that HMGs B and C more closely resemble the smaller pair of higher eucaryotic HMG proteins, HMGs 14 and 17, than the larger pair, HMGs 1 and 2 (Table 1). The high percentage of aromatic amino acid residues in HMGs B and C (HMGs 14 and 17 are devoid of aromatic amino acids; Table 1) and the presence of the HMG 1 box in both proteins (see below) indicate that *Tetrahymena* HMGs B and C also have features that closely resemble HMGs 1 and 2 (Table 1). HMGs B and C do not appear to be strongly related to HMG I or D1. Like HMGs 14 and 17, HMG I and D1 are devoid (or essentially devoid) of aromatic acids, and they contain characteristic short repeats of the triplet (Gly-Arg-Pro, which are not present in either HMG B or C (6, 22). Thus, it seems reasonable to view HMGs B and C in *T. thermophila* as proteins sharing properties of both HMGs 14 and 17 and HMGs 1 and 2 (see below).

Even if this view is correct, we stress that HMGs B and C also differ considerably from all known vertebrate HMG counterparts. For example, in vertebrate HMG proteins, acidic amino acids are characteristically concentrated in the carboxy-terminal domain (45). The carboxy-terminal regions of both HMGs B and C are, however, strikingly basic. The carboxy-terminal third of HMG B is also rich in alanine (23%) and proline (6%). Clustering of lysine, alanine, and proline is found in the amino- and carboxy-terminal domains of a number of histone H1 proteins, including *Tetrahymena* H1 (2, 16, 46, 48).

Chromatin-associated proteins with properties similar to those of vertebrate HMG proteins have now been described from several lower eucaryotes other than *T. thermophila*. However, the precise relationships of these proteins to each other and to the well-characterized vertebrate HMG proteins are not clear (see reference 43 for an excellent review and discussion of this issue), and this problem is further confounded by the absence of functional assays for this class of proteins. For instance, *Tetrahymena* macronuclear histone H1 is unusual in that it lacks the central hydrophobic domain conserved in H1 proteins of multicellular eucaryotes (17, 48). Since H1 and HMG proteins overlap considerably in size, solubility, and aspects of amino acid composition (Table 2), it is difficult to classify *Tetrahymena* H1 being more H1- or HMG-like. Presently, the only distinctive H1-like property of macronuclear H1 is its association with linker DNA (38). The finding that delta, a linker-associated micronuclear protein, shares many properties in common

TABLE 2. Amino acid compositions of micronuclear linker histones and macronuclear HMG B, HMG C, and histone H1

Protein	Amino acid composition (%)																			
	Basic			Acidic		Aromatic			Nonpolar hydrophobic							Polar				
	Lys	Arg	His	Asp	Glu	Tyr	Phe	Trp	Leu	Ile	Val	Ala	Pro	Gly	Met	Thr	Ser	Asn	Gln	Cys
Micronuclear[a]																				
Beta	24.2	7.2	0.9	3.4	5.5	0.4	0.0	0.0	0.0	0.4	0.9	7.2	2.1	4.7	1.3	6.4	23.3	8.5	3.8	0.0
Gamma	20.0	13.5	0.5	3.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	4.5	0.0	4.5	1.0	2.5	39.0	5.5	3.5	0.0
Delta	20.0	3.6	1.5	3.1	8.2	5.6	4.6	0.5	6.2	2.6	5.1	5.6	4.1	3.6	0.5	2.6	6.7	8.7	7.2	0.0
Macronuclear																				
HMG B	25.9	2.8	0.0	5.0	11.9	5.0	2.1	0.7	5.0	2.8	2.1	12.6	6.3	1.4	0.7	4.2	6.3	0.7	4.2	0.7
HMG C	28.0	1.0	1.0	3.0	11.0	5.0	3.0	1.0	4.0	3.0	2.0	9.0	6.0	3.0	2.0	3.0	6.0	4.0	5.0	0.0
H1[b]	32.5	1.8	2.5	5.5	4.3	0.0	0.0	0.0	1.8	2.5	3.7	16.0	7.4	1.8	0.6	10.4	6.8	1.8	0.6	0.0

[a] Sequences from Wu et al. (49).
[b] Sequence from Wu et al. (48).

with macronuclear HMGs B and C (see below) only serves to blur the distinction between H1- and HMG-type proteins even further.

The absence of clear distinctions between H1-like and HMG-like proteins in *T. thermophila* and the observation that HMGs B and C share properties with two different subgroups of higher eucaryotic HMG proteins (Table 1) suggest that lower eucaryotes like *T. thermophila* contain a family of related proteins (HMG B, HMG C, H1, and delta) whose structural and functional properties have been imparted to multiple, more specialized proteins during evolution. This observation may account for the inability to identify a "typical H1" protein in some lower eucaryotes like yeasts. Interestingly, deletion of *ACP2* from *Saccharomyces cerevisiae* is lethal and suggests a vital, yet undetermined function for HMG-like proteins in yeast cells (14). Although NHP6A and NHP6B from *S. cerevisiae* appear to be related to *Tetrahymena* HMG C (23), genetic and biochemical experiments addressing the role of this protein in yeast cells have not yet been reported. Unfortunately, it is not known whether ACP2 or NHP6 is associated with linker regions of yeast chromatin.

Homology to vertebrate HMGs 1 and 2. Comparison of the amino acid sequences of HMGs B and C to each other and to other sequenced eucaryotic HMG proteins identified an 11-amino-acid sequence near the amino terminus, the HMG 1 box, that is highly conserved among HMGs B, C, 1, 2, and T (Fig. 4). Also shown for comparison is the sequence of two yeast proteins, ACP2 (14) and NHP6 (23), which both show partial homology to HMG 1. NHP6 proteins A and B show significant homology to *Tetrahymena* HMG C (~30% amino acid identity), and clearly both NHP6 proteins contain a sequence with strong homology to the HMG 1 box. Interestingly, strong homology to the HMG 1 box is not found in the predicted translation product of *ACP2*, an essential yeast gene suggested to encode a protein analogous to calf thymus HMG 1 (14).

The HMG 1 box is contained within a broader region shown to be involved in the DNA binding of HMGs 1 and 2 (8, 9, 19), and its striking conservation raises the possibility that the HMG 1 box represents an important, novel DNA-binding motif. Evidence suggesting that this is likely to be correct has recently been provided by Jantzen and associates in their study of the nucleolar transcription factor hUBF (20). Three domains were identified in hUBF (hUBF boxes 1 to 3, each roughly 80 to 90 amino acids long) which display significant sequence similarity to HMGs 1 and 2. Interestingly, these domains each contain the shorter (11-amino-acid), highly conserved HMG 1 box that we have identified in *Tetrahymena* HMGs B and C (Fig. 4). Importantly, DNA binding experiments (20) have shown that a single copy of the hUBF HMG box is necessary and sufficient for DNA binding.

To resolve the apparent paradox between the sequence-independent DNA binding of HMG proteins and sequence-specific DNA binding of transcription factors like hUBF, Jantzen et al. (20) point out that the HMG domains of hUBF may have evolved from a primordial, nonspecific DNA-binding structure. Because of the abundance of *Tetrahymena* HMG B, HMG C, and delta (see below), it is unlikely that their HMG 1 boxes function in a sequence-specific fashion. The high degree of conservation illustrated by the 11-amino-acid region shown in Fig. 4 suggests that this narrower region could be the primitive core structure from which more specific DNA-binding functions evolved by addition of sequences conferring specificity. Despite the likely importance of the HMG 1 box, it cannot account for all of the characteristic HMG-like properties shared between the *Tetrahymena* and vertebrate HMG proteins. HMGs 14 and 17 share with *Tetrahymena* HMGs B and C the ability to interact with nucleosome core particles and the property of being extracted from nuclei by elutive intercalation (Table 1). However, neither HMG 14 nor HMG 17 contains the HMG 1 box.

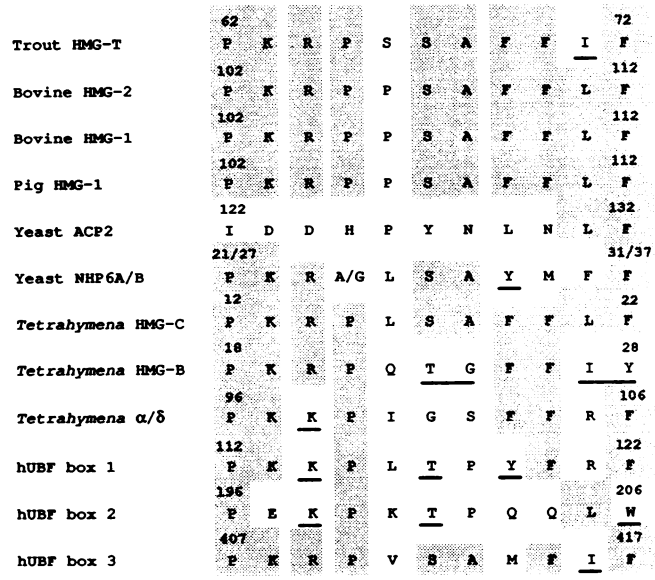

FIG. 4. The HMG 1 box. Shown is an 11-amino-acid residue sequence conserved among bovine HMGs 1 and 2, pig HMG 1, trout testis HMG T, yeast NHP6A and -B, *Tetrahymena* HMG B, *Tetrahymena* HMG C (the complete amino acid sequence of HMG C has also been obtained from *Tetrahymena pyriformis* [17]; the HMG 1 box in it is identical to that presented here for *T. thermophila*). For comparison, the same region of a yeast HMG 1-like protein ACP2, is presented. Also shown is a smaller region of three domains recently identified in the human nucleolar transcription factor hUBF as showing sequence similarity to HMGs 1 and 2 (hUBF boxes 1 to 3). Shaded amino acids represent residues conserved in most (at least 6 of 12) of the sequences shown above. Underlined amino acids indicate conservative replacements. Numbers above the first and last amino acids of each sequence indicate the position of this sequence in the intact protein. Proteolytic processing of the micronuclear linker histone polyprotein is such that alpha is further processed into delta and gamma (3, 49). Because of this, the HMG 1 box is present in both alpha and delta, and therefore we refer to it as alpha/delta. Two NHP6 genes encoding closely related yeast proteins, NHP6A and NHP6B, have been described (23); differences between these proteins are indicated as A/B. The sequences used were obtained from the following sources: bovine HMGs 1 and 2, Walker (45); pig HMG 1 Tsuda et al. (41); trout HMG T, Pentecost et al. (30); yeast ACP2, Haggren and Kolodrubetz (14); yeast NHP6A/B, Kolodrubetz and Burgum (23); and hUBF boxes 1 to 3, Jantzen et al. (20).

The hypothesis that the amino-terminal regions of HMGs B and C are involved in DNA binding is further supported in the case of HMG C by the existence of a seven-residue stretch (residues 7 to 13) that partially overlaps the HMG 1 box and contains residues completely homologous to residues 28 to 35 of HMG 17 (six of seven homologous to the same region of HMG 14). This small region has been determined by physical studies to be a component of a broader region of these two smaller HMG proteins that interacts with DNA in vitro (8), suggesting that HMG C may

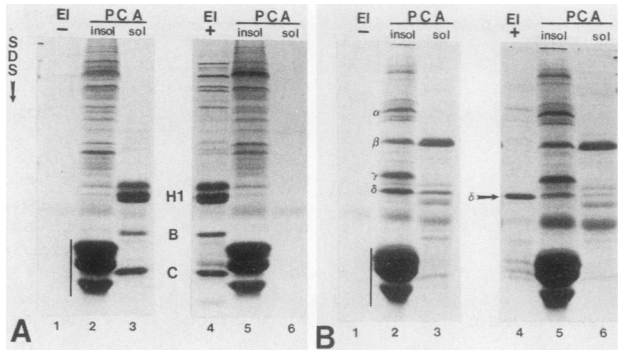

FIG. 5. Release of delta from micronuclei by elutive intercalation with 20 mM EtBr. Isolated macronuclei (A) or micronuclei (B) were washed into buffer E and subjected to elutive intercalation (EI) for 30 min in the presence (+; lanes 4) or absence (−; lanes 1) of 20 mM EtBr. The nuclear pellets remaining after elutive intercalation were then acid extracted to produce a PCA-soluble fraction (lanes 3 and 6) and a PCA-insoluble fraction (lanes 2 and 5) (see Materials and Methods for details). Each fraction was then electrophoresed in a 15% SDS-gel and visualized by Coomassie staining. The arrow points to delta, which is the only micronuclear linker protein extracted by elutive intercalation with 20 mM EtBr. Vertical lines to the left of lanes 2 identify macro- and micronuclear core histones.

have a DNA-binding domain that combines features of the DNA-binding domains of two classes of higher eucaryotic HMG proteins. Other than the HMG 1 box and the short homology to HMGs 14 and 17 mentioned above, no other significant homologies to *Tetrahymena* HMGs B and C have been detected in computer searches of existing DNA and protein sequence data banks. However, it is likely that the HMG 1 box itself will be found in other unrelated DNA-binding proteins if this motif provides a general DNA-binding function(s).

An HMG-like protein exists in transcriptionally inactive micronuclei. Wu et al. (49) recently reported the complete protein sequence derived from the micronuclear linker histone gene mic LH. This gene encodes a large polyprotein, initially referred to as X, that is proteolytically processed to yield the micronuclear linker polypeptides alpha, beta, gamma, and delta (Fig. 1; 3, 49). Quite surprisingly, this sequence encodes an HMG 1 box between amino acids 96 and 106 in the region of the precursor protein which encodes delta (Fig. 4; 49).

By itself, the presence of 11 amino acids of sequence homology may not be a strong criterion for classifying a protein as being HMG-like. Delta, however, also has a number of other HMG-like properties. It is small (approximately 23 kDa) and contains a high concentration of both basic (25%) and acidic (11%) residues. The overall amino acid composition of delta is similar to that of HMGs B and C and strikingly different from those of beta, gamma, and macronuclear histone H1 (Table 2). This point is particularly evident when one focuses on the percentage of aromatic acids in these polypeptides. The percentage of aromatic amino acids in delta is strikingly high (11%) compared with the percentages in beta (0.4%), gamma (0%), and macronuclear H1 (0%) and is very similar to those of HMG B (8%) and HMG C (9%). Delta also is the only product of the mic LH gene that is partially soluble in 5% PCA and that can be extracted from micronuclei by elutive intercalation (Fig. 5). This procedure specifically extracts HMGs 14 and 17 from vertebrate nuclei (37), an HMG-like protein from *Drosophila* nuclei (44), and HMGs B and C from *Tetrahymena* macronuclei (39). Elutive intercalation of delta from micronuclei requires a higher concentration of EtBr than that used to extract HMGs B and C from macronuclei (20 versus 5 mM; compare Fig. 5 with Fig. 3 of reference 39), although the requirement for higher concentrations of EtBr may be due simply to differences in accessibility of the drug to the DNA in the condensed chromatin of micronuclei. When macronuclei are extracted with 20 mM EtBr, histone H1 is removed along with HMG B and HMG C (Fig. 5, lanes 4). Thus, at higher concentrations of EtBr, elutive intercalation is not specific for small HMG-like proteins, a result also observed for chicken nuclei (37). These results suggest, however, that delta, HMG B, HMG C, and H1 may all interact with chromatin in a similar fashion (see below). Finally, the presence of spermidine in nucleus isolation buffers specifically depletes delta (but not alpha, beta, and gamma) isolated from purified micronuclei (compare Fig. 1 with Fig. 1 of reference 5). Sensitivity to spermidine has been observed for HMGs B and C from purified *Tetrahymena* macronuclei (15, 39) and for the *Drosophila* HMG-like protein (44). Taken together, these observations strongly suggest that a protein with many properties characteristic of *Tetrahymena* HMGs B and C and of higher eucaryotic HMG proteins is present as an abundant acid-soluble protein in transcriptionally inactive micronuclear chromatin.

Conclusion. Several lines of evidence suggest that HMGs B and C may play a role(s) in packaging macronuclear chromatin into a transcriptionally competent conformation (43). Consistent with this notion is the recent finding that a eucaryotic transcription factor, hUBF, contains several short DNA-binding motifs with sequence similarity to HMGs B and C and vertebrate HMGs 1 and 2 (20; Fig. 4). However, the fact that delta in micronuclei also contains this putative DNA-binding domain and shares other properties with HMGs B and C serves to complicate this issue. Superficially, the mere presence of delta in micronuclei suggests that HMG- or linker-like polypeptides also exist in chromatin that is packaged into a transcriptionally inactive state. The function(s) of delta in micronuclei is not known, nor is that of any of the other micronuclear linker-associated polypeptides (alpha, beta, and gamma). The distinct amino acid composition of delta suggests it may have a unique function(s) similar to that of macronuclear HMG-like molecules. Although it is likely that this function(s) is not related to transcription, we point out that the region of homology between all of these polypeptides, the HMG 1 box (Fig. 4), is very short and may be related to only the DNA-binding potential of each polypeptide. Other domains outside of the HMG 1 box could impart other, more specific functions (protein-protein and protein-DNA) to each polypeptide which could be very specific. Detailed genetic and biochemical analyses will be necessary to clarify these points.

ACKNOWLEDGMENTS

I.G.S. thanks Scott Rose for his excellent assistance in cloning of the HMG B and C genes and Sue Berget for providing some of the reagents used in this study. We thank Susan Goike for her help in typing the manuscript.

This work was supported by NIH grants HD16259 and GM40922 to C.D.A. and by grants from the American Cancer Society and NIH to M.A.G.

REFERENCES

1. **Alfagaeme, C. R., G. T. Gudkin, and L. H. Cohen.** 1980. Isolation, properties, and cellular distribution of D1, a chromosomal protein of *Drosophila*. Chromosoma **78**:1–31.

2. **Allan, J., T. Mitchell, N. Harborne, L. Bohm, and C. Crane-Robinson.** 1986. Roles of H1 domains in determining higher order chromatin structure and H1 location. J. Mol. Biol. **187:** 591–601.
3. **Allis, C. D., R. L. Allen, J. C. Wiggins, L. G. Chicoine, and R. Richman.** 1984. Proteolytic processing of H1-like histones in chromatin: a physiological and developmentally regulated event in *Tetrahymena* micronuclei. J. Cell Biol. **99:**1669–1677.
4. **Allis, C. D., C. V. C. Glover, J. K. Bowen, and M. A. Gorovsky.** 1980. Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eukaryote, *Tetrahymena thermophila*. Cell **20:**609–617.
5. **Allis, C. D., C. V. C. Glover, and M. A. Gorovsky.** 1979. Micronuclei of *Tetrahymena* contain two types of histone H3. Proc. Natl. Acad. Sci. USA **76:**4857–4861.
6. **Ashley, C. T., C. G. Pendelton, W. W. Jennings, A. Saxena, and C. V. C. Glover.** 1989. Isolation and sequencing of cDNA clones encoding *Drosophila* chromosomal protein D1. J. Biol. Chem. **264:**8394–8401.
7. **Blackburn, E. H., and K. M. Karrer.** 1986. Genomic reorganization in ciliated protozoans. Annu. Rev. Genet. **20:**501–521.
8. **Bradbury, E. M.** 1982. Conformations and interactions, p. 89–110. *In* E. W. John (ed.), The HMG chromosomal proteins. Academic Press, London.
9. **Butler, A. P., J. K. W. Mardian, and D. E. Olins.** 1985. Nonhistone chromosomal protein HMB 1 interactions with DNA. Fluorescence and thermal denaturation studies. J. Biol. Chem. **260:**10613–10620.
10. **Elton, T. S., M. S. Nissen, and R. Reeves.** 1987. Specific A-T DNA sequence binding of RP-HPLC purified HMG-1. Biochem. Biophys. Res. Commun. **143:**260–265.
11. **Ghosh, P. K., V. B. Reddy, M. Platak, P. Lebowitz, and S. M. Weissman.** 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. **65:**580–595.
12. **Gorovsky, M. A., M.-C. Yao, J. B. Keevert, and G. L. Pleger.** 1975. Isolation of micro- and macronuclei of *Tetrahymena pyriformis*. Methods Cell Biol. **9:**311–327.
13. **Gross, D. S., and W. T. Garrard.** 1987. Poising chromatin for transcription. Trends Biochem. Sci. **12:**293–297.
14. **Haggren, W., and D. Kolodrubetz.** 1988. The *Saccharomyces cerevisiae ACP2* gene encodes an essential HMG-like protein. Mol. Cell. Biol. **8:**1281–1289.
15. **Hamana, K., and K. Iwai.** 1979. High mobility group nonhistone chromosomal proteins also exist in *Tetrahymena*. J. Biochem. (Tokyo) **86:**789–794.
16. **Hayashi, T., H. Hayashi, and K. Iwai.** 1987. *Tetrahymena* histone H1. Isolation and amino acid sequence lacking the central hydrophobic domain conserved in other H1 histones. J. Biochem. (Tokyo) **102:**369–376.
17. **Hayashi, T., H. Hayashi, and K. Iwai.** 1989. *Tetrahymena* HMG nonhistone chromosomal protein. Isolation and amino acid sequence lacking the N- and C-terminal domains of vertebrate HMG 1. J. Biochem. (Tokyo) **105:**577–581.
18. **Horowitz, S., J. K. Bowen, G. A. Bannon, and M. A. Gorovsky.** 1987. Unusual features of transcribed and translated regions of the histone H4 gene family of *Tetrahymena thermophila*. Nucleic Acids Res. **15:**141–160.
19. **Isackson, P. J., J. L. Fishback, D. L. Bidney, and G. R. Reek.** 1979. Preferential affinity of high molecular weight mobility group non-histone chromatin proteins for single-stranded DNA. J. Biol. Chem. **254:**5569–5572.
20. **Jantzen, H.-M., A. Admon, S. P. Bell, and R. Tjian.** 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature (London) **344:** 830–836.
21. **Johns, E. W.** 1982. History, definitions and problems, p. 1–9. *In* E. W. Johns (ed.), The HMG chromosomal proteins. Academic Press, London.
22. **Johnson, K. R., D. A. Len, T. S. Elton, P. J. Barr, and R. Reeves.** 1988. The chromosomal high mobility group protein HMG-1(Y): complete murine cDNA sequence, genomic structure, and tissue expression. J. Biol. Chem. **263:**18338–18342.
23. **Kolodrubetz, D., and A. Burgum.** 1990. Duplicated NHP6 genes of *Saccharomyces cerevisiae* encode proteins homologous to bovine high mobility group protein I. J. Biol. Chem. **265:**3234–3239.
24. **Levinger, L., and A. Varshavsky.** 1982. Protein D1 preferentially binds A + T-rich DNA *in vitro* and is a component of *Drosophila melanogaster* nucleosomes containing A + T-rich satellite DNA. Proc. Natl. Acad. Sci. USA **79:**7152–7156.
25. **Levy-Wilson, B., M. S. Denker, and E. Ito.** 1983. Isolation, characterization, and postsynthetic modifications of *Tetrahymena* high mobility group proteins. Biochemistry **22:**1715–1721.
26. **Mardian, J. K. W., A. E. Baton, G. J. Bunick, and D. E. Olins.** 1980. Nucleosomes have two specific binding sites for nonhistone chromosomal proteins HMG 14 and HMG 17. Science **209:**1534–1536.
27. **Martindale, D. W.** 1989. Codon usage in *Tetrahymena* and other ciliates. J. Protozool. **36:**29–34.
28. **McMurray, C., and K. E. van Holde.** 1986. Binding of ethidium bromide causes dissociation of the nucleosome. Proc. Natl. Acad. Sci. USA **83:**8472–8476.
29. **Melton, D. A., P. A. Drieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12:**7035–7056.
30. **Pentecost, B. T., J. M. Wright, and G. H. Dixon.** 1985. Isolation and sequence of cDNA clones coding for a member of the family of high mobility group proteins (HMG-T) in trout and analysis of HMG-T-mRNA's in trout tissues. Nucleic Acids Res. **13:**4871–4888.
31. **Reeves, R.** 1984. Transcriptionally active chromatin. Biochim. Biophys. Acta **782:**343–393.
32. **Reeves, R., T. S. Elton, M. S. Nissen, D. Lehn, and K. R. Johnson.** 1987. Posttranscriptional gene regulation and specific binding of the nonhistone protein HMG-1 by the 3′ untranslated region of bovine interleukin 2 cDNA. Proc. Natl. Acad. Sci. USA **84:**6531–6535.
33. **Roth, S. Y., I. G. Schulman, R. G. Cook, and C. D. Allis.** 1987. The complete amino acid sequence of an HMG-like protein isolated from the macronucleus of *Tetrahymena*. Nucleic Acids Res. **15:**8112.
34. **Russnak, R. H., E. P. M. Candido, and C. R. Astell.** 1988. Interaction of the mouse chromosomal protein HMG-1 with the 3′ ends of genes *in vitro*. J. Biol. Chem. **263:**6392–6399.
35. **Sandeen, G., W. I. Wood, and G. Felsenfeld.** 1980. The interaction of high mobility group proteins HMG 14 and 17 with nucleosomes. Nucleic Acids Res. **8:**3757–3778.
36. **Schröter, H., and J. Bode.** 1982. The binding sites for large and small high mobility group (HMG) proteins. Studies on HMG-nucleosome interactions *in vitro*. Eur. J. Biochem. **127:**429–436.
37. **Schröter, H., G. Maier, H. Ponsting, and A. Nordheim.** 1985. DNA intercalators induce specific release of HMG 14 and HMG 17 and other DNA-binding proteins from chicken erythrocyte chromatin. EMBO J. **4:**3867–3872.
38. **Schulman, I. G.** 1989. Isolation and characterization of two high mobility group-like proteins from the ciliated protozoan, *Tetrahymena thermophila*. Ph.D. thesis, Baylor College of Medicine, Houston, Tex.
39. **Schulman, I. G., R. G. Cook, R. Richman, and C. D. Allis.** 1987. *Tetrahymena* contain two distinct and unusual high mobility group (HMG)-like proteins. J. Cell Biol. **104:**1485–1494.
39a. **Singh, J., and G. H. Dixon.** 1990. High mobility group proteins 1 and 2 function as general class II transcription factors. Biochemistry **29:**6295–6302.
40. **Strauss, F., and A. Varshavsky.** 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought to mutual proximity by DNA folding in the nucleosomes. Cell **37:**889–901.
41. **Tsuda, K.-I., M. Kikuchi, K. Mori, S. Waga, and M. Yoshida.** 1988. Primary structure of non-histone protein HMG 1 revealed by the nucleotide sequence. Biochemistry **27:**6159–6163.
42. **Tsunasawa, S., and F. Saklyama.** 1984. Amino-terminal acetylation of proteins: an overview. Methods Enzymol. **106:**165–

170.
43. **van Holde, K. E.** 1988. Chromatin. Springer-Verlag-New York, Inc., New York.
44. **Wagner, C. R., S. A. Amero, K. Hamana, and S. C. R. Elgin.** 1987. Isolation of a putative HMG protein from *Drosophila*. J. Cell. Biol. **105**:152a.
45. **Walker, J. M.** 1982. Primary structures, p. 69–88. *In* E. W. Johns (ed.), The HMG chromosomal proteins. Academic Press, London.
46. **Wells, D. E.** 1986. Compilation analysis of histones and histone genes. Nucleic Acids Res. **14**:r119–r123.
47. **White, E. M., D. L. Shaprio, C. D. Allis, and M. A. Gorovsky.** 1988. Sequence and properties of the message encoding *Tetrahymena* hv1, a highly evolutionarily conserved histone H2A variant that is associated with active genes. Nucleic Acids Res. **16**:179–198.
48. **Wu, M., C. D. Allis, R. Richman, R. G. Cook, and M. A. Gorovsky.** 1986. A canonical intervening sequence in an unusual histone H1 gene of *Tetrahymena thermophila*. Proc. Natl. Acad. Sci. USA **83**:8674–8678.
49. **Wu, M., C. D. Allis, R. Richman, R. G. Cook, and M. A. Gorovsky.** Unpublished data.
50. **Yu, S.-M., and M. A. Gorovsky.** 1986. *In situ* dot blots: quantitation of mRNA in intact cells. Nucleic Acids. Res. **14**:7497–7615.