

# Isolation and characterization of maize cDNAs encoding a high mobility group protein displaying a HMG-box

Klaus D.Grasser and Günter Feix\*

Institut für Biologie III, Schänzlestr. 1, D-7800 Freiburg, FRG

Received March 22, 1991; Revised and Accepted April 26, 1991

EMBL accession no. X58282

## ABSTRACT

**cDNAs encoding a nonhistone chromosomal high mobility group (HMG) protein corresponding to the animal HMG1 family were isolated from a maize cDNA library using an immunoscreening approach. The cDNAs revealed an open reading frame of 471 base pairs together with 413 base pairs of flanking region, in agreement with the size of mRNA detected by Northern analysis of maize endosperm RNA. Like its animal counterparts the 17146 Da maize HMG protein contains a basic aminotermus and an acidic carboxyterminus. The HMG-box region of this plant HMG protein shows striking sequence similarity to members of the vertebrate HMG1 family. Based on Southern blot hybridization analysis of genomic DNA, the isolated cDNA appears to be derived from a single or low copy gene.**

## INTRODUCTION

High mobility group (HMG) proteins represent a family of nonhistone chromosomal proteins with relative low molecular weights common to eukaryotic organisms. In animals they consist of at least two proteins with approximate Mr 25000 (HMG1 and HMG2), and three proteins with approximate Mr 11000 (HMG14, HMG17 and HMGI) (1,2). HMG1 and HMG2 bind to single-stranded and double-stranded DNA, while HMG14 and HMG17 are associated with nucleosomes (3,4). HMGI and its alternatively processed isoform HMGY bind to A/T-rich regions of DNA (2,5). HMG proteins appear to be preferentially associated with transcriptionally active chromatin (1,6) and are more abundant in undifferentiated, rapidly proliferating cells (7,8,9,10). These properties, together with their different posttranslational modifications, indicate an essential regulatory role for HMG proteins (1). This was recently confirmed by the finding that the yeast ACP2 gene product, which is of structural similarity to HMG1 and HMG2, proved to be required for cell viability (11). HMG1, HMG2 and the HMGI-like protein p16 were, furthermore, reported to activate *in vitro* transcription of polymerase I, II and III genes (12,13,14,15). The formation of active initiation complexes is therefore possibly assisted by HMG proteins by stimulating the binding of specific transcription factors to their recognition sites (16) or by altering DNA topology (2,17).

Alternatively, the activation may occur by changing nucleosome positioning (18,19) or by removing the transcriptional block caused by cruciform DNA (20). However none of these proposed functions has so far been experimentally confirmed *in vivo*.

Plant HMG proteins have been isolated and biochemically characterized from different sources (21,22,23,24,25). Since no immunological crossreactions between these plant HMG proteins and animal HMG proteins could be observed, no major structural similarity was anticipated (26,27). Wheat HMG proteins have been shown to be preferentially associated with actively transcribed chromatin like their animal counterparts (28). Furthermore, some plant HMG proteins bind to A/T-rich stretches of double-stranded DNA (25,29,30). Maize HMG proteins were recently shown to bind specifically to CCAAT- and TATA-boxes of a zein gene promoter (31).

In search for a gene coding for a plant HMG protein, we screened a maize cDNA library with an antiserum against a HMG protein from maize endosperm (32) corresponding to the animal HMG1 family as estimated from its running position in SDS-PAGE (31,32). This led to the isolation of cDNA clones encoding a maize HMG protein. Although the protein sequence derived from these clones proved to be different from animal HMG protein sequences, it displays striking similarity in the conserved putative DNA binding motif, the so-called HMG-Box.

## EXPERIMENTAL PROCEDURES

### Screening of a maize library for HMG protein encoding cDNAs

A maize Uni-ZAP cDNA library (Stratagene) was screened by the immunoscreening procedure essentially as described by Sambrook et al. (33). The Uni-ZAP phages contain the plasmid vector pBluescript (Stratagene) with cDNAs inserted at the EcoRI/XhoI site of the multiple cloning site. Approximately 150.000 IPTG induced phage plaques were screened on nitrocellulose filters (Schleicher & Schuell) using a mouse antiserum against the largest HMG protein detected in maize endosperm as described recently (32). Filters were incubated in TBST (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 0.05% Tween 20) containing 1% low fat milk powder for 30 minutes before the antiserum was added at a final dilution of 1:2500 in TBST for at least 6 h. Then filters were washed twice in TBST for 10

\* To whom correspondence should be addressed

minutes followed by addition of alkaline phosphatase conjugated antimouse antiserum (Sigma) diluted 1:2000 with TBST. After an incubation for 2 h filters were washed twice with TBST before antibody binding was established by the standard alkaline phosphatase colour reaction. Positive phage clones were purified by rescreening the plaques until complete purity of the clones was obtained.

**Nucleotide sequence analysis**

The recombinant pBluescript plasmids were isolated from purified phage clones by *in vivo* excision with helper phage R408 as outlined by the manufacturer (Stratagene). After amplification of the plasmids, the cDNA inserts were subjected to restriction enzyme analysis. Nucleotide sequence analysis was carried out on an EMBL automated fluorescent DNA sequencer (34) according to a modified dideoxy method using fluorescently labeled primers (35). The obtained nucleotide sequences were submitted to further computer analysis using the program PC/Gene (IntelliGenetics and Genofit).

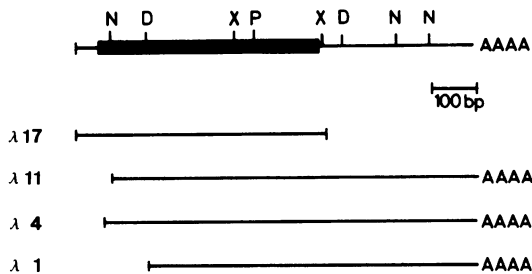
**Aminoterminal amino acid sequence determination**

The amino acid sequence of the aminoterminal of the maize HMG protein purified to homogeneity from endosperm tissue was determined as described by Eckerskorn et al. (36).

**Southern and Northern blot analysis**

**Southern blot analysis:** 3 µg of high molecular weight total DNA isolated from maize seedling leaves (37) was digested with restriction enzymes, separated in 0.8% agarose gels and transferred to a Hybond-N membrane (Amersham) by capillary blotting (33).

**Northern blot analysis:** 4 µg of poly(A) RNA from endosperm tissue was separated in 1% agarose gels and transferred to membrane as mentioned above. The cDNA insert of one of the isolated cDNA clones was labeled with digoxigenin (DIG) (Boehringer) by random priming and used as probe for Southern and Northern blot hybridizations. Hybridizations were performed at 42°C in a buffer containing 5×SSC, 50% formamide, 0.1% sarcosyl, 0.02% SDS and 2% blocking powder (Boehringer) for at least 6 h followed by a wash at 68°C in the same buffer but without formamide. Hybridization bands were detected by the addition of an alkaline phosphatase conjugated anti-DIG antiserum (Boehringer) followed by the chemiluminiscent substrate AMPPD



**Figure 1.** Restriction enzyme map and schematic alignment of several of the isolated cDNA clones. The top line represents the structure of the maize HMG cDNA deduced from the analysis of different overlapping clones. The black bar represents the coding region. The positions of restriction sites for NarI, DraII, XhoI or PstI are indicated by N, D, X, P respectively. The alignment of four of the isolated cDNAs according to their sequences is shown schematically below.

(Boehringer) and subsequent visualization by exposition to a X-ray film for 30 to 120 minutes.

Standard DNA manipulations were performed according to Sambrook et al. (33).

**RESULTS**

**Isolation of maize HMG cDNA clones**

An antiserum against the largest HMG protein isolated from maize endosperm tissue with an apparent molecular weight of about 20 kDa described recently (32) was used to screen approximately 150.000 phage plaques of a maize Uni-ZAP cDNA library and revealed 16 phage clones reacting with the antiserum. Eight of these clones were purified by rescreening at low plaque density. The specificity of the antiserum was confirmed by a parallel screening with preimmune serum. The cDNA inserts obtained by the *in vivo* excision procedure from four of the recombinant phages and the results of the restriction mapping are summarized schematically in figure 1.

**Nucleotide sequence analysis of the HMG cDNA and the deduced amino acid sequence**

The four cDNA inserts shown in figure 1 were sequenced using an automated fluorescent DNA sequencer (34,35). The sequence of 885 bp obtained from the sequences of the four different overlapping clones is shown in figure 2. The deduced 157 amino acid protein sequence of the open reading frame indicates a

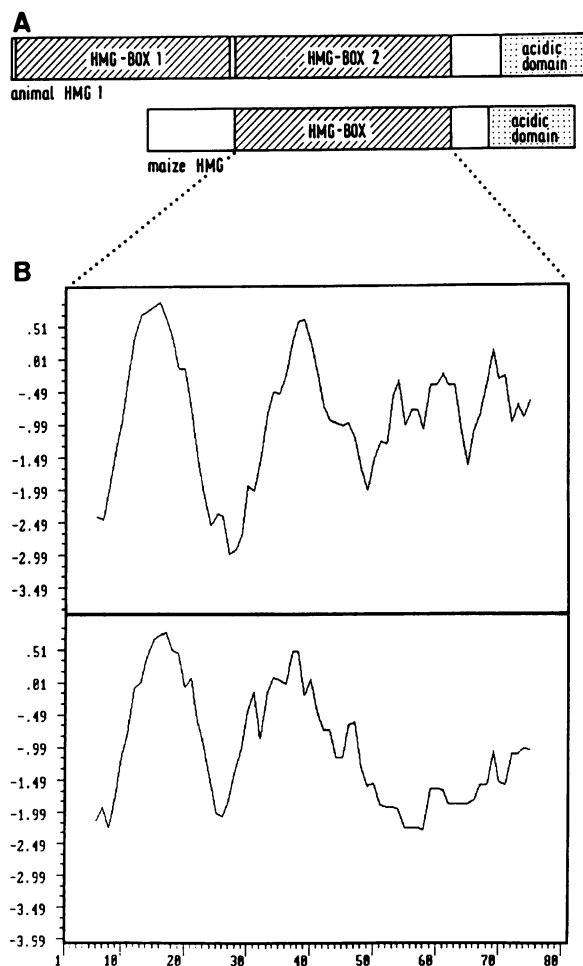
```

-39  GGCACGAGCGCTGCGCTCCGTGTGCGCGCTCGAATCCGCC
1  ATGAAGGGGGCCAAATCCAAGGGCGCCGCAAGGCCGATGCCAAGCTGGCGGTGAAGGC
   M K G A K S K G A A K A D A K L A V K E
   10                                     20
61  AAGGGGGCGGAGAAGCCCGCAAGGGCAGGAAGGGCAGGCAAGCCCAACCAAG
   K G A E K P A R G R K G K A G K D P N K
   30                                     40
121 CCTAAGAGGGCTCCACGCGCTTTCTCGTGTTCATGGAGGAATTCGCAAGGAGTTCAG
   P K R A P S A P F V F M E E F R K E F K
   50                                     60
181 GAGAAGAACCTAAGAACAAGTCTGTCGCTGCGGTGGGAAAGCTGCTGGCAGCAGGTGG
   E K N P K N K S V A A V G K A A G D R W
   70                                     80
241 AAATCCCTGAGCGAGTCCGACAGGGCTCCCTATGTAGCCAGGCTAACAGCTCAAGCTC
   K S L S E S D K A P Y V A K A N K L K L
   90                                     100
301 GAGTACAACAGGCCATCGCTGCTACACAAAGGGCAGAGCACTGCAAGCTAAGAGGCT
   E Y N K A I A A Y N K G E S T A A K K A
   110                                    120
361 CCTGCGAAGGAGGAAGAGGAGGAGATGAAGAGGAGTCTGACAAAGTCCAAGTCCGAGGTC
   P A K E E E E E D E E E S D K S K S E V
   130                                    140
421 AATGACGAGGATGATGAAGAGGGTAGTGAGGAGGATGAAGATGATGACGAGTGAATGGAGC
   N D E D D E E G S E E D E D D D E
   150
481 TCCTCGAGACAATGGACCGTGTCTTCAACCAATGGAGCGGCTACACAAAGGCCCGTGG
541 CGATCACAAAAAGGAGCCCTATATCCATGTACTAGAAATATTCAAGTTTCACTCCACATCG
601 TGATGTTTTATTTTACTTTTGTGCTATAACGGATAGCGCTCCTCGTTGGCGCCACT
661 GGCGGGTGGTTCTGCGCTGGTGTGATGTTTGTGTGGTCCACACTTCCAGCCACG
721 AGAGCGCCTGCTTCTACCTAGATTACTGTTTCCATTGCTGCTCATCGGCTAACATTG
781 TCATAATGTCAGTTGGGTAATGTTAGATTAAGTAATTTGTTGTTCAAAAAAAAAAAAA
841 AAAAAA
    
```

**Figure 2.** Nucleotide sequence of the maize HMG cDNA. The nucleotide numbering refers to the ATG initiation codon. The termination codon (TGA) is underlined. Flanking EcoRI and XhoI sites are not shown. The underlined aminoterminal part of the deduced amino acid sequence corresponds perfectly with that determined previously by amino acid sequence analysis of purified HMG protein.



termed HMG-box (47) between amino acids 36 to 109. Members of the animal HMG1 family contain the two HMG-boxes 1 and 2 of about 80 amino acid residues, while the maize HMG protein contains only one such element. The maize HMG-box shows 42% identity to HMG-box2 of pig HMG1 which is localized in the middle of the protein, while it is 26% identical to the aminoterminal HMG-box1. An alignment of the maize HMG-box to HMG-box2 of various vertebrate HMG1 and HMG2 proteins is shown in figure 4. Only a low similarity of the aminoterminal region of the maize HMG protein to HMG-box1 of the vertebrate proteins could be observed, while there is some similarity in the carboxyterminal acidic domain. This is schematically represented in the alignment of the maize HMG protein to pig HMG1 shown in figure 5A. A potential structural and functional similarity of the different proteins is also evident from a comparison of the hydropathy index profiles of their HMG-box regions determined by the method of Kyte and Doolittle (48) and depicted in figure 5B.



**Figure 5.** Schematic representation of the maize HMG protein and pig HMG1 and the hydropathy index profiles of their HMG-boxes. A) Comparative alignment of the structure of the maize HMG protein and pig HMG1 protein. HMG-boxes and the acidic domains are indicated. B) Hydropathy index profiles of the maize HMG protein (upper panel) and pig HMG1 (lower panel) determined by the method of Kyte and Doolittle (46). Relative amino acid positions are indicated below, corresponding to positions 36 to 109 for the maize HMG and positions 89 to 162 for pig HMG1 of figure 4.

## DISCUSSION

Genes encoding various HMG proteins have been cloned and sequenced from several vertebrates, from *Tetrahymena* and from yeast (2). The maize HMG gene, presented in this communication adds now the first plant HMG gene to this list. The protein encoded by this gene displayed in band shift experiments a strong binding to the A/T-rich P2 promoter region of the zein gene *PMS1* (29) with a preference for CCAAT- and TATA-boxes of this promoter region (31).

The amino acid sequence deduced from the cloned maize HMG cDNA inserts displays 22.8% basic and 23.4% acidic residues which is similar to the amino acid content of animal HMG1 and HMG2 proteins. A search of the EMBL protein sequence database confirmed the similarity to members of the vertebrate HMG1 family. An alignment of the HMG1 and HMG2 sequences with the maize HMG protein revealed high similarity especially in the so-called HMG-box (47) of the putative DNA binding region. However, in contrast, to the animal proteins which contain two HMG-boxes, the plant HMG protein reported here displays only one such element. This HMG-box is more similar to the animal HMG-box2 than to the aminoterminal HMG-box1. The similarity of this plant HMG-box to the HMG-boxes located in the middle of animal HMG1 and HMG2 is also evident from a comparison of their hydropathy index profiles. This, in addition to the smaller size of the maize HMG protein, indicates to an aminoterminal shorter protein because a carboxyterminal acidic domain containing 73% acidic residues is also present in the maize protein (compare schema of figure 5A). There are also similarities in the HMG-box regions of this plant HMG protein and the chromosomal nonhistone proteins NHP6a and NHP6b from yeast (49) and to a lower extent to the HMGb and HMGC proteins from *Tetrahymena* (50), but these proteins are smaller and display no clear acidic domain. Almost no sequence similarity exists to the yeast ACP2 protein (11) although this protein displays several structural features of the HMG1 family, but no conserved HMG-box. The search of the EMBL protein sequence database revealed no significant similarity to the animal HMG14/17 and HMG1/Y proteins.

The apparent molecular weight of about 20 kDa of the maize HMG protein as estimated by SDS-PAGE (31) compared with the 17 kDa deduced from the cDNA clone may be an indication for posttranslational modifications. This would be in line with results from animal HMG proteins, which are subject to different types of modifications (1). The serine at position 149 which lies within an almost perfect recognition sequence for casein kinase II (41) is a good candidate for the phosphorylation of the HMG protein. This maize HMG protein had been shown previously to be phosphorylated by a protein kinase of the casein type II isolated from maize endosperm nuclei (24).

Southern blot hybridization analysis suggests that the HMG gene may occur in the genome in one or only few copies. The observed abundance of this HMG protein in nuclei of maize tissue indicates therefore to a high expression of the gene(s) and/or a remarkable stability of the gene products. In the case of animal HMG proteins a rather long half-life could be demonstrated (9). Since the existence of the aminoterminal methionine in the mature maize HMG protein was shown by aminoterminal amino acid sequence determination a half-life greater than 20 h was calculated (Grasser and Feix, unpublished observation).

As shown earlier, the HMG protein described here is one of the most abundant chromosomal nonhistone proteins of maize

nuclei (24,31). The availability of this HMG cDNA sequence will allow studies on structure and expression of the gene. Furthermore, the deduced amino acid sequence may facilitate investigations on the involvement of this HMG protein in chromatin structure and transcriptional regulation.

## ACKNOWLEDGEMENTS

We would like to thank Mrs.E.Bruzer for excellent technical assistance, Dr.G.Igloi and Mr.R.Maier of this department for DNA sequence determination, Drs.C.Eckerskorn and F.Lottspeich of the Max Planck Institut für Biochemie, Martinsried for protein sequence determination and Drs.T.Quayle and U.Theopold of this department for helpful discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie.

## REFERENCES

- Einck, L. and Bustin, M. (1985) *Exp. Cell Res.* 156, 296–310.
- Bustin, M., Lehn, D.A. and Landsman, D. (1990) *Biochim. Biophys. Acta* 1049, 231–243.
- Brown, J.W. and Anderson, J.A. (1986) *J. Biol. Chem.* 261, 1349–1354.
- Bustin, M., Crippa, M.P. and Pash, J.M. (1990) *J. Biol. Chem.* 265, 20077–20080.
- Johnson, K.R., Lehn, D.A. and Reeves, R. (1989) *Mol. Cell. Biol.* 9, 2114–2123.
- Weisbrod, S. (1982) *Nature* 297, 289–295.
- Giancotti, V., Pani, B., Andrea, P.D., Berlingeri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G., Philip, R., Crane-Robinson, C., Nicolas, R.H., Wright, C.A. and Goodwin, G.H. (1987) *EMBO J.* 6, 1981–1987.
- Mosevitsky, M.I., Novitskaya, V.A., Iogannsen, M.G. and Zabezhinsky, M.A. (1989) *Eur. J. Biochem.* 185, 303–310.
- Begum, N., Pash, J.M. and Bhorjee J.M. (1990) *J. Biol. Chem.* 265, 11936–11941.
- Johnson, K.R., Diney, J.E., Wyatt, C.R. and Reeves, R. (1990) *Exp. Cell Res.* 187, 69–76.
- Haggren, W. and Kolodrubetz, D.C. (1988) *Mol. Cell. Biol.* 8, 1282–1289.
- Tremethick, D.J. and Molloy, P.L. (1986) *J. Biol. Chem.* 261, 6986–6992.
- Tremethick, D.J. and Molloy, P.L. (1988) *Nucleic Acids Res.* 16, 11107–11123.
- Yang-Yen, H.-F. and Rothblum, L.I. (1988) *Mol. Cell. Biol.* 8, 3406–3414.
- Singh, J. and Dixon, G.H. (1990) *Biochemistry* 29, 6295–6302.
- Watt, F. and Molloy, P.L. (1988) *Nucleic Acids Res.* 16, 1471–1486.
- Sheflin, L.G. and Spaulding, S.W. (1989) *Biochemistry* 28, 5658–5664.
- Strauss, F. and Varshavsky, A. (1984) *Cell* 37, 889–901.
- Waga, S., Mizuno, S. and Yoshida, M. (1989) *Biochim. Biophys. Acta* 1007, 209–214.
- Waga, S., Mizuno, S. and Yoshida, M. (1990) *J. Biol. Chem.* 265, 19424–19428.
- Spiker, S. (1984) *J. Biol. Chem.* 259, 12007–12013.
- Vincentz, M. and Gigot, C. (1985) *Plant Mol. Biol.* 4, 161–168.
- Moehs, C.P., McElwain, E.F. and Spiker, S. (1988) *Plant. Mol. Biol.* 11, 507–515.
- Grasser, K.D., Maier, U.-G. and Feix, G. (1989) *Biochem. Biophys. Res. Commun.* 162, 456–463.
- Jacobsen, K., Laursen, N.B., Jensen, E.O., Marcker, A., Poulsen, C. and Marcker, K.A. (1990) *Plant Cell* 2, 85–94.
- Vanderbilt, J.N. and Andersen, J.N. (1985) *J. Biol. Chem.* 260, 9336–9345.
- Spiker, S. and Everett, K.M. (1987) *Plant Mol. Biol.* 9, 431–442.
- Spiker, S., Murray, M.G. and Thompson, W.F. (1983) *Proc. Natl. Acad. Sci.* 80, 815–819.
- Maier, U.-G., Grasser, K.D., Haass, M.M. and Feix, G. (1990) *Mol. Gen. Genet.* 221, 164–170.
- Pedersen, T.J., Arwood, L.J., Spiker, S., Guiltinan, M.J. and Thompson, W.F. (1991) *Plant Mol. Biol.* 16, 95–104.
- Grasser, K.D., Maier, U.-G., Haass, M.M. and Feix, G. (1990) *J. Biol. Chem.* 265, 4185–4188.
- Grasser, K.D. and Feix, G. (1991) submitted.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Ansorge, W., Sproat, B.S., Stegemann, J. and Schwager, C. (1986) *J. Biochem. Biophys. Meth.* 13, 315–323.
- Voss, H., Schwager, C., Kristensen, T., Duthie, S., Olsson, A., Erfle, H., Stegemann, J., Zimmermann, J. and Ansorge, W. (1989) *Meth. Mol. Cell. Biol.* 1, 155–159.
- Eckerskorn, C., Mewes, W., Goretzki, H. and Lottspeich, F. (1988) *Eur. J. Biochem.* 176, 509–519.
- Quayle, T.J.A., Brown, J.W.S. and Feix, G. (1989) *Gene* 80, 249–257.
- Tsuda, K., Kikuchi, M., Mori, K., Waga, S. and Yoshida, M. (1988) *Biochemistry* 27, 6159–6163.
- Shirakawa, H., Tsuda, K. and Yoshida, M. (1990) *Biochemistry* 29, 4419–1123.
- Doolittle, R.F. (1986) *Of urfs and orfs: A primer on how to analyze derived amino acid sequences*, University Science Books, Mill Valley, CA.
- Kuenzel, E.A., Mulligan, J.A., Sommercorn, J. and Krebs, E.G. (1987) *J. Biol. Chem.* 262, 9136–9140.
- Kaplan, D.J. and Duncan, C.H. (1988) *Nucleic Acids Res.* 16, 10375.
- Wen, L., Huang, J.-K., Johnson, B.H. and Reeck, G.R. (1989) *Nucleic Acids Res.* 17, 1197–1214.
- Lee, K.-L.D., Pentecost, B.T., Anna, J.A.D., Tobey, R.A., Gurley, L.R. and Dixon, G.H. (1987) *Nucleic Acids Res.* 15, 5051–5068.
- Paonessa, G., Frank, R. and Cortese, R. (1987) *Nucleic Acids Res.* 15, 9077.
- Pentecost, B.T., Wright, J.M. and Dixon, G.H. (1985) *Nucleic Acids Res.* 13, 4871–4888.
- Jantzen, H.-M., Admon, A., Bell, S.P. and Tijan, R. (1990) *Nature* 344, 830–836.
- Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105–132.
- Kolodrubetz, D. and Burgum, A. (1990) *J. Biol. Chem.* 265, 3234–3239.
- Schulman, I.G., Wang, T., Wu, M., Bowen, J., Cook, R.G., Gorovsky, M.A. and Allis, C.D. (1991) *Mol. Cell. Biol.* 11, 166–174.