The EMBO Journal vol.11 no.3 pp.1055-1063, 1992

The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins

Marco E.Bianchi, Luca Falciola, Simona Ferrari and David M.J.Lilley[1]

Dipartimento di Genetica e Microbiologia, Universita' di Pavia, via Abbiategrasso 207, I-27100 Pavia, Italy, and [1]Department of Biochemistry, The University, Dundee DD1 4HN, UK

Communicated by D.M.J.Lilley

The mammalian nuclear protein HMG1 contains two segments that show a high sequence similarity to each other. Each of the segments, produced separately from the rest of the protein in *Escherichia coli*, binds to DNA with high specificity: four-way junction DNA of various sequences is bound efficiently, but linear duplex DNA is not. Both isolated segments exist as dimers in solution, as shown by gel filtration and chemical crosslinking experiments. HMG1-like proteins are present in yeast and in protozoa: they consist of a single repetition of a motif extremely similar to the DNA binding segments of HMG1, suggesting that they too might form dimers with structural specificity in DNA binding. Sequences with recognizable similarity to either of the two DNA binding segments of HMG1, called HMG boxes, also occur in a few eukaryotic regulatory proteins. However, these proteins are reported to bind to specific sequences, suggesting that the HMG box of proteins distantly related to HMG1 might differ significantly from the HMG box of HMG1-like proteins.

Key words: cruciform DNA/high mobility group proteins/HMG1/HMG box/protein-DNA interactions

Introduction

High mobility group 1 protein (HMG1) is an abundant protein which is present in all vertebrate nuclei, and whose sequence is extremely conserved. Calf, pig, rat and human HMG1s have only two divergent amino acids (Kaplan and Duncan, 1988; Tsuda *et al.*, 1988; Bianchi *et al.*, 1989; Wen *et al.*, 1989) and the trout protein is extremely similar to the mammalian ones (Pentecost *et al.*, 1985). A related protein, HMG2, is also present in vertebrate nuclei, although with a lower abundance in most tissues (Einck and Bustin, 1985). In addition, HMG1-like proteins exist in *Saccharomyces cerevisiae* (Kolodrubetz and Burgum, 1989) and *Tetrahymena thermophila* (Roth *et al.*, 1987); similar proteins may well be present in all eukaryotic cells.

Its ubiquity and sequence conservation point to a fundamental role for HMG1, but despite intensive study its function has not been identified unequivocally. Roles in DNA replication and chromatin assembly (Bonne-Andrea *et al.*, 1984; Kohlstaedt *et al.*, 1986) or RNA transcription (Kleinschmidt *et al.*, 1983) have been proposed. Reconstitution studies have shown that HMG1 and 2 may act *in vitro* as general transcription factors for RNA polymerase II and III and overcome the inhibition caused by histones (Tremethick and Molloy, 1986, 1988; Singh and Dixon, 1990). In addition, both HMG1 and 2 stimulate *in vitro* transcription of the adenovirus major late promoter by facilitating the binding of the specific transcription factor MLTF (Watt and Molloy, 1988).

Recently, we discovered that HMG1 binds with high selectivity to four-way junction DNA, a specific structure of DNA that can be generated as an intermediate in recombination as well as from inverted repeat sequences under the effect of supercoiling (Bianchi *et al.*, 1989). The binding to four-way junction DNA is structure specific and sequence independent, since HMG1 can bind to several four-way junctions of unrelated sequence, but does not bind to linear duplex or single-stranded DNA containing the same sequences that are recognized in cruciform configuration.

In this paper, we analyse the mode of interaction of rat HMG1 with DNA. We demonstrate that HMG1 actually contains two peptides that are involved in DNA recognition; sequences clearly related to them exist in all HMG1-like proteins. While this work was in progress, a number of investigators reported that several eukaryotic transcription factors share a region of considerable sequence similarity to amino acids 92-172 of HMG1, which correspond to one of the DNA binding peptides. Such conserved sequences, called HMG boxes, are present in SRY, the product of the Y-chromosome gene that determines the formation of the testis in mammalian males (Sinclair *et al.*, 1990) and in hUBF, a protein involved in rDNA transcription by human RNA polymerase I (Jantzen *et al.*, 1990). These proteins, however, show a defined sequence-specific DNA binding: this suggests the existence of two different subclasses of HMG boxes, with different modes of protein-DNA interaction.

Results

Interaction of partially deleted HMG1 proteins with four-way junction DNA

In order to identify the DNA binding domain of rat HMG1, we constructed several sets of derivatives of the original plasmid bearing the cDNA sequence. Each plasmid with a variant sequence was then transcribed and translated *in vitro*, producing radioactively labelled polypeptides which were tested for binding to four-way junction DNA by means of an electrophoretic assay.

A set of 13 plasmids coding for proteins with progressive deletions from the C-terminus were generated by terminating the sequence of HMG1 cDNA with a TAG stop codon (Figure 1). The corresponding deletions in HMG1 protein spanned from a minimum of 24 amino acids (pRNHMG1/M1-E191) to a maximum of 188 amino acids (pRNHMG1/M1-N37). *In vitro* translation of polypeptides shorter than 90 amino acids turned out to be very inefficient, either using rabbit reticulocyte lysates or wheat germ

```
1   MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKTMSAKEKGKFEDMAKADKARYEREMKTYIP  80
81  PKGETKKKFKDPNAPKRPPSAFFLFCSEYRPKIKGEHPGLSIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIA 160
161 AYRAKGKPDAAKKGVVKAEKSKKKKEEEDDEEDEEDEEEEEEEEDEDEEEDDDDE                          215
```

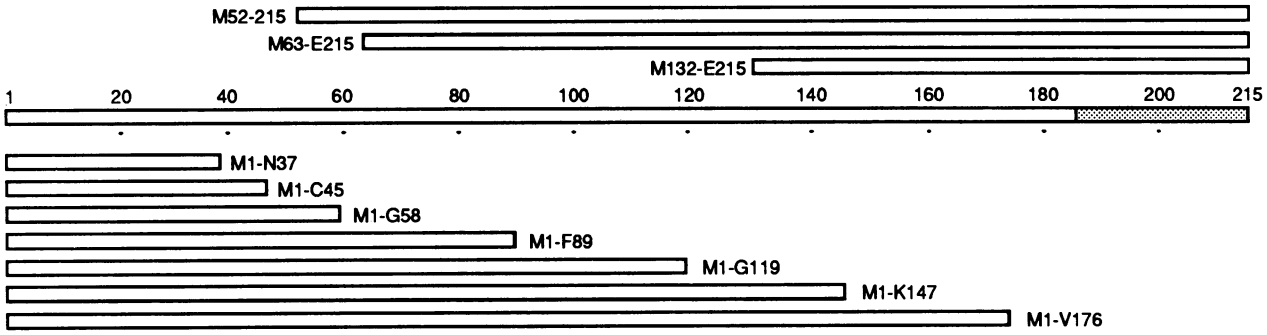

Fig. 1. Protein HMG1 and its partially deleted derivatives. **Upper part**: sequence of rat HMG1 (Paonessa *et al.*, 1988; corrected by Bianchi *et al.*, 1989). **Lower part**: schematic representation of wild-type and partially deleted HMG1 proteins. Plasmids coding for the deleted proteins were constructed as described in Materials and methods. All polypeptides are identified by their first and last amino acids. Complete HMG1 is represented in the middle of the diagram; the shaded box represents the stretch of 30 acidic residues.

extracts; all other polypeptides were produced efficiently. The labelled polypeptides were then incubated on ice with carrier DNA, in the presence or the absence of unlabelled four-way junction DNA, and were electrophoresed in an agarose gel. In the absence of four-way junction DNA all translation products did not migrate far from the well, but in the presence of the specific probe a variable proportion formed an additional, faster migrating band, which is indicative of the formation of a specific protein–DNA complex (Figure 2). Although this test is essentially qualitative, the intensity of the band corresponding to the protein–DNA complex can be taken as a rough indication of the binding abilities of the various deletion proteins. Protein HMG1/M1–V176 binds four-way junction DNA even better than full-length HMG1, indicating that, as expected, the stretch of 30 consecutive acidic residues at the C-terminus are not required for DNA binding nor for protein stability. HMG1/M1–F89 appears to bind four-way junction DNA at least as efficiently as full-length HMG1, suggesting that the DNA binding domain might be located in the N-proximal third of the wild-type protein.

To confirm the localization of the binding domain, we constructed deletions in pRNHMG1 that removed the 5′ untranslated portion of the cDNA together with part of the N-terminal coding sequence, allowing translation to start from an internal ATG triplet. We synthesized polypeptides starting from the third, the fourth and the sixth methionine residue of HMG1 (HMG1/M52–E215, HMG1/M63–E215 and HMG1/M132–E215, respectively). The smallest of these polypeptides is unable to bind four-way junction DNA, but both HMG1 proteins with deletions of 51 and 62 amino acids still retain strong binding activity (Figure 2).

The identification of the HMG box

The results reported in the preceding section might be taken to locate the DNA binding site of HMG1 to the stretch of amino acids between M63 and K90, but the limited size and lack of apparent structure of the peptide involved made this conclusion rather unappealing. Moreover, comparative sequence analysis of HMG1 and related proteins pointed to

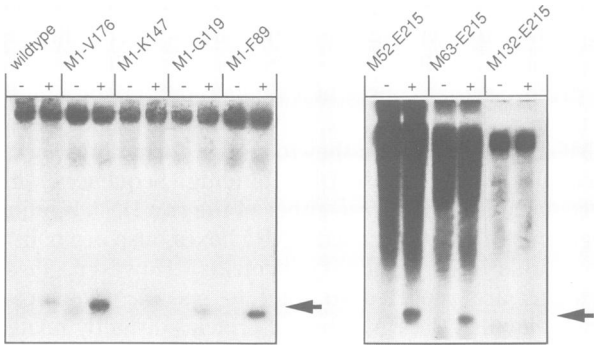

Fig. 2. Binding assays for HMG1 N-terminal and C-terminal deletions. Plasmids of the pRNHMG1 series were cut with *Hin*dIII restriction endonuclease and transcribed with T7 RNA polymerase; RNAs were translated *in vitro* by a rabbit reticulocyte lysate with [^{35}S]methionine (see Materials and methods for details). An aliquot from the unfractionated translation mixtures was added to DNA binding buffer containing 250 μg/ml of sonicated salmon sperm DNA (lanes indicated with −) or 250 μg/ml of sonicated salmon sperm DNA plus 5 μg/ml four-way junction DNA (lanes indicated with +), and assayed by electrophoresis in agarose gels. The arrow indicates protein–DNA complexes.

a different interpretation of the binding data. It had been noted earlier that two segments within protein HMG1 are remarkably similar to each other (Reeck *et al.*, 1982). We confirmed this observation and found that both segments are similar to one or two segments of a number of HMG1-like protein sequences, including proteins HMG-T from trout (Pentecost *et al.*, 1985), HMG2 from pig (Shirakawa *et al.*, 1990), a protein from the macronucleus of *Tetrahymena* (Roth *et al.*, 1987), and two nuclear proteins from the yeast *S.cerevisiae* (Kolodrubetz and Burgum, 1989). The consensus in this set of sequences, which we will call the canonical HMG box, extends for almost 80 amino acids and is very strict: 12 residues are absolutely conserved and 20 more are present in the majority of the sequences (Figure 3A). In addition, while this work was in progress, several authors reported that sequences related to the canonical HMG box, but less stringently conserved, were

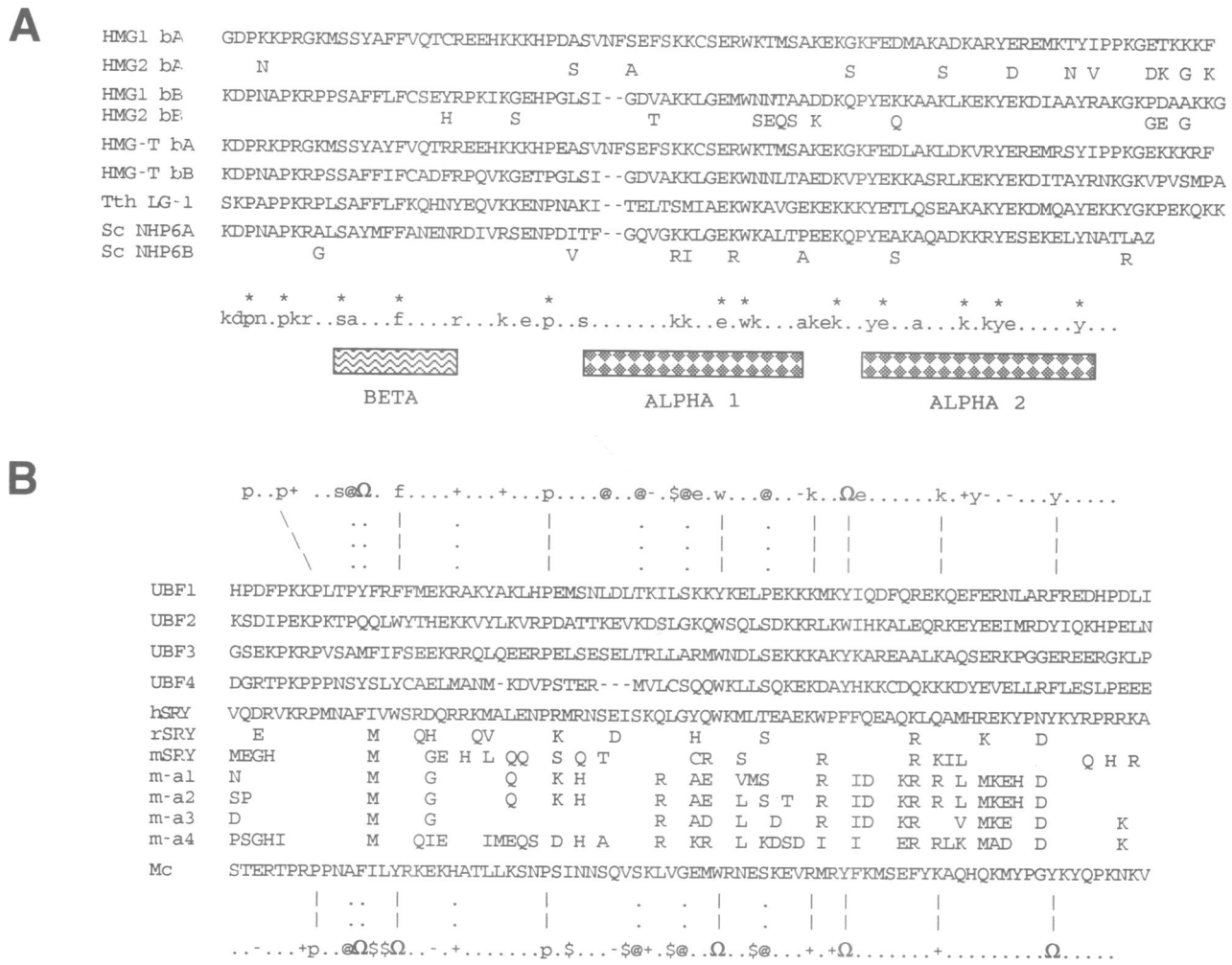

Fig. 3. HMG1 contains two amino acid sequences which are similar to each other and to a number of other proteins. (**A**) Canonical HMG boxes. Alignment of the regions of similarity in amino acids 4–89 (HMG1 bA) and 90–174 (HMG1 bB) of rat HMG1 (Paonessa *et al.*, 1988; Bianchi *et al.*, 1989) with corresponding regions in pig HMG2 (Shirakawa *et al.*, 1990) (HMG2 bA and bB), trout HMG-T (Pentecost *et al.*, 1985) (HMG-T bA and bB), macronuclear protein LG-1 of *T.thermophila* (Roth *et al.*, 1987) (Tth LG-1), proteins NHP6A and -B of *S.cerevisiae* (Kolodrubetz and Burgum, 1990) (Sc NHP6A and Sc NHP6B). Since HMG2 is almost identical to HMG1, and NHP6B to NHP6A, only the amino acids that are different are indicated; an empty space means that the corresponding residue is identical to that of the complete line above it. Z indicates the end of the protein (a stop codon in the sequence of the gene). Gaps introduced in the sequences to align them are indicated with a dash. The consensus sequence for canonical HMG boxes contains only amino acids which are identical in most of the proteins; conservative substitutions have not been considered. The 12 amino acids which are identical in all HMG1-like proteins have been highlighted with an asterisk. Several different computer programs predicting secondary structures were run for all of the proteins: the consensus prediction is indicated below the consensus sequence. (**B**) Generalized HMG boxes. The first line is the consensus sequence for the canonical HMG box: only residues which are always present or which are only conservatively substituted are indicated. Below it are aligned the four HMG boxes (UBF1–4) in the sequence of UBF, the human transcription factor for RNA polymerase I (Jantzen *et al.*, 1990); the single HMG boxes of human (hSRY), rabbit (rSRY) and mouse (mSRY) SRY proteins (Sinclair *et al.*, 1990; Gubbay *et al.*, 1990); the four HMG boxes encoded by four different genes which cross-hybridize to the cDNA for mouse SRY (m-a1 to m-a4) (Gubbay *et al.*, 1990). Since the SRY-related sequences are very similar to each other, for this group only the complete sequence of hSRY and the amino acids of the other proteins which differ from hSRY are indicated. Mc indicates the single HMG1 box in the product of the mating-type gene *Mc* of *S.pombe* (Kelly *et al.*, 1988). Gaps in the alignment are indicated by dashes. Conservative substitutions are indicated as follows: @ for proline, alanine, glycine, serine and threonine; − for glutamic acid and aspartic acid; + for arginine and lysine; Ω for tryptophan, phenylalanine and tyrosine; $ for methionine, valine, leucine and isoleucine. The last line represents the consensus sequence of the generalized HMG box. The dashed lines across the set of sequences indicate those residues which are very strongly conserved. Other conserved residues are indicated by a dotted line.

present in a number of other proteins (Figure 3B). Four such sequences have been identified within the human RNA polymerase I transcription factor UBF (Jantzen *et al.*, 1990), and one each in the Mc protein from *Schizosaccharomyces pombe* (Kelly *et al.*, 1988), in human SRY protein, the product of the gene for testis determination (Sinclair *et al.*, 1990), and in mouse SRY-related proteins (Gubbay *et al.*, 1990).

The evolutionary conservation of the HMG box implies that it constitutes a well-defined functional and/or structural module. The entire amino acid sequence of HMG1 is accounted for by two HMG boxes plus the polyacidic domain, which is not required for protein binding (Figure 2). This suggests that HMG1 contains two DNA binding modules, each corresponding to one of the HMG boxes. The deletion analysis reported in the preceding section is in accordance with this interpretation, but does not prove it. We therefore proceeded to examine whether each HMG box, in isolation from the rest of the protein, would bind to four-way junction DNA.

Both HMG boxes of HMG1 protein bind efficiently to four-way junction DNA

Plasmids capable of directing the synthesis in *Escherichia coli* of boxes A and B of HMG1 protein were constructed by cloning the relevant sequences of the rat HMG1 cDNA (coding for amino acids M1−F89 and D91−V176, respectively) in plasmid pT7-7 (see Materials and methods). Cells of the BL21(DE3) strain (Studier and Moffat, 1986) were transformed with plasmids pT7-HMG1bA and pT7-HMG1bB; after induction with IPTG, they produced the polypeptides corresponding to boxes A and B very efficiently (Figure 4A).

The crude lysates from the overproducing strains could be tested directly for binding to four-way junction DNA (Figure 4B): the patterns of retardation they formed in polyacrylamide gels, as well as their specific activity, were almost indistinguishable. Thus, the high level of sequence similarity between the two HMG boxes of HMG1 protein is matched by a similar binding activity towards four-way junction DNA.

Peptide HMG1bA is 89 amino acids long, and corresponds to the shortest polypeptide that could be translated efficiently *in vitro*. However, it extends a little beyond the region of strict sequence conservation defined by the canonical HMG box, which ends with tyrosine 78. To test whether a shortened version of box A would retain its DNA binding properties, we mutagenized *in vitro* lysine 82 to a stop codon and produced in *E.coli* the corresponding peptide, HMG1/M1−P81. This 'minimal HMG box' also efficiently binds four-way junction DNA (Figure 4C), supporting the conclusion drawn from the analysis of primary sequences. Still shorter versions of box A were difficult to produce in *E.coli* as well as in cell-free systems.

Both HMG boxes of HMG1 protein exhibit structure discrimination

We showed previously that HMG1 protein can distinguish four-way junction DNA from duplex or single-stranded DNA of the same sequence (Bianchi, 1988; Bianchi *et al.*, 1989); we therefore tested whether this property was also intrinsic to the isolated boxes. Labelled HMG1bA was synthesized *in vitro* with rabbit reticulocyte extract; it was then mixed with 1.0 μg of *Hin*fI-cleaved plasmid pUC19, or with the same amount of DNA fragments plus 0.05 μg of four-way junction DNA (Figure 5). In the absence of four-way junction DNA, most of the labelled box A formed complexes with the various fragments of plasmid pUC19 (lane 5), and only a minor amount migrated as the diffuse smear of unbound protein. The labelled polypeptide did not appear to bind completely aspecifically, since the DNA fragments of 65 and 75 bp did not form complexes, while the fragments of 214 and 1419 bp were bound more avidly than the fragments of 396 and 517 bp. When a 20-fold lower concentration of four-way junction DNA was added to the linear pUC DNA fragments, most of the labelled polypeptide appeared to form a complex with this species, rather than with the much more abundant pUC fragments (lane 6). It therefore appears that box A, even isolated from the rest of the protein, retains a considerable selectivity towards four-way junction DNA. On the other hand, the control experiment performed with wild-type HMG1 (lanes 1−3) shows that the full-length protein is essentially ineffective in binding linear fragments, even in the absence of four-way junction DNA, and is therefore even more selective.

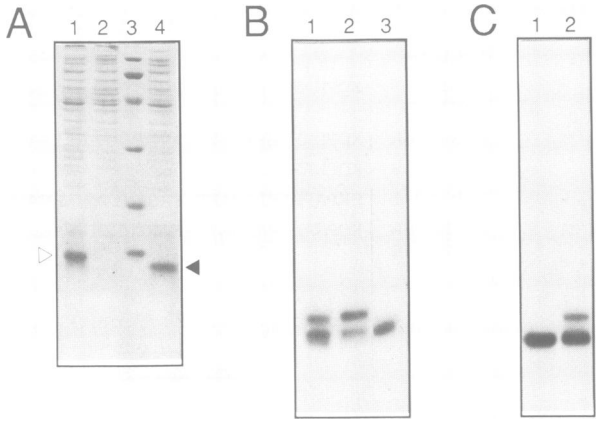

Fig. 4. Isolated A and B boxes of HMG1 produced by *E.coli* bind to four-way junction DNA. (**A**) Production of HMG boxes in *E.coli*. BL21(DE3) cells, and the same cells transformed with plasmids pT7-HMG1bA, pT7-HMG1bB and pT7-HMG1/M1−P81, were grown and induced with IPTG as described in Materials and methods. Induced cells were lysed by sonication and centrifuged; the clear lysates were applied to 16% polyacrylamide−SDS gels (~100 μg per lane), electrophoresed and stained with Coomassie Blue. Lane 1: lysate from BL21(DE3)[pT7-HMG1bA] cells; lane 2: lysate from control BL21(DE3) cells; lane 3: mol wt markers (rabbit phosphorylase B, bovine serum albumin, hen ovalbumin, bovine carbonic anhydrase, soybean trypsin inhibitor and hen egg lysozyme); lane 4: lysate from BL21(DE3)[pT7-HMG1bB] cells. (**B**) Binding of boxes A and B to four-way junction DNA. About 0.2 μg of total protein from the clear lysates of BL21(DE3)[pT7-HMG1bA] cells (lane 1), BL21(DE3)[pT7-HMG1bB] cells (lane 2), or control BL21(DE3) cells (lane 3) were added to DNA binding mixtures containing 1.5 nM labelled four-way junction DNA **c** and 50 μg/ml sonicated salmon sperm DNA as aspecific competitor. The samples were then electrophoresed and autoradiographed as described in Materials and methods. (**C**) Binding of the 'minimal HMG box' to four-way junction DNA. About 0.2 μg of total protein from the clear lysates of BL21(DE3) cells (lane 1), or BL21(DE3)[pT7-HMG1/M1−P81] cells (lane 2) were added to DNA binding mixtures containing 1.5 nM labelled four-way junction DNA **c** and 50 μg/ml sonicated salmon sperm DNA as aspecific competitor. The samples were then electrophoresed and autoradiographed as described in Materials and methods.

The proteins which conform to the canonical HMG consensus are not more divergent from either the A or B box of HMG1 than the latter are between each other: we then tested the structure discrimination of box B as a proxy for all other canonical HMG boxes. Roughly equivalent concentrations of purified peptides HMG1bA and HMG1bB were mixed with labelled four-way junction DNA, or with the same amount of labelled four-way junction DNA plus varying concentrations of unlabelled four-way junction DNA or unlabelled linear duplex DNAs of identical sequence (Figure 6). The binding selectivity of both HMG boxes for four-way junction DNA was essentially identical: binding to the probe was not competed by a 1000-fold higher concentration of linear DNA, but was competed as expected by cold four-way junction DNA.

Both HMG boxes of HMG1 protein form dimers in solution

Each HMG1 molecule contains two HMG boxes; it was therefore important to determine whether the individual boxes (produced as separate polypeptides and in different cells) would associate spontaneously to form dimers.

An aliquot of a purified preparation of peptide HMG1bB was applied to a gel filtration FPLC column; the fractions

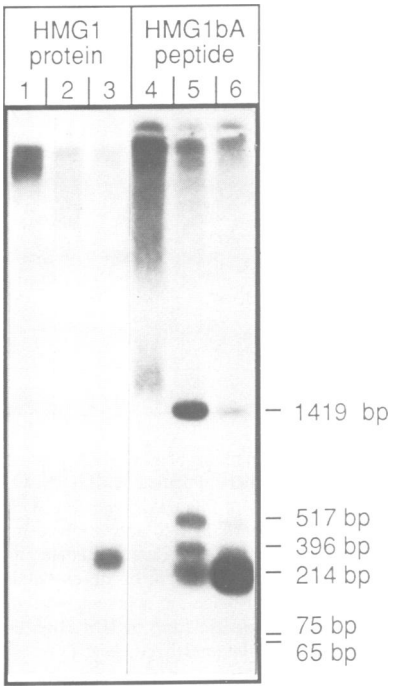

Fig. 5. DNA binding selectivity of isolated box A. Aliquots from unfractionated translation mixtures containing labelled HMG1 protein (lanes 1–3) or labelled HMG1bA peptide (lanes 4–6) were added to DNA binding buffer containing either no DNA (lanes 1 and 4), 100 μg/ml *Hin*fI-cleaved pUC19 plasmid DNA (lanes 2 and 5), or 100 μg/ml *Hin*fI-cleaved pUC19 plasmid DNA plus 5 μg/ml synthetic four-way junction DNA **f** (lanes 3 and 6), and assayed by electrophoresis in an agarose gel. The mobilities of the pUC19/*Hin*fI fragments, as visualized by ethidium bromide staining before drying the gel, are indicated to the side.

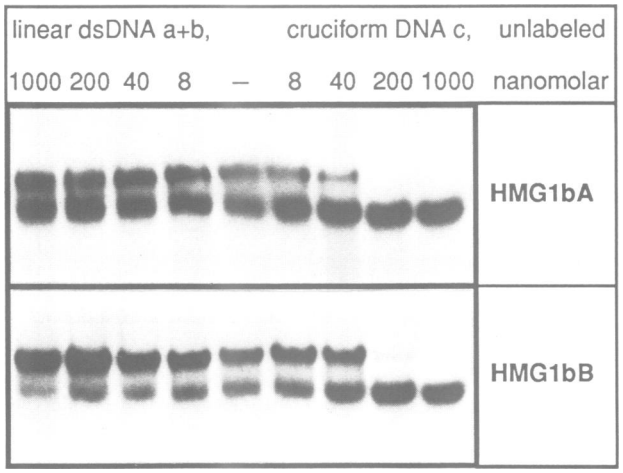

Fig. 6. DNA binding selectivity of isolated box B. To DNA binding mixtures containing 1.5 nM labelled four-way junction DNA **c**, we added the indicated concentrations of unlabelled four-way junction DNA, or the indicated concentration of unlabelled linear duplex **a** plus the same concentration of unlabelled linear duplex **b**. To each, we then added peptides HMG1bA or HMG1bB to a final concentration of 6 ng/ml (~0.6 nM polypeptide chains for each, from the preparations purified to homogeneity). After incubation on ice, the samples were assayed by electrophoresis as described in Materials and methods.

of the eluate were assayed by gel retardation and Western blots (Figure 7). In comparison with molecular weight standards, the activity emerged as a single peak centred around a M_r of 20 kDa. The amount of material detected in Western blots followed the same distribution; in particular, no immunoreactive material was detected in the fractions corresponding to the M_r of 9.9 kDa calculated from the known amino acid sequence of HMG1bB.

An equivalent experiment was performed for peptide HMG1bA, and produced an estimate of M_r = 19 kDa against an expectation of M_r = 10.3 kDa for a globular monomer (results not shown).

As a control, we determined the apparent M_r of a sample of full-length HMG1 prepared from rat liver as described by Bianchi *et al.* (1989). The activity emerged as a peak centred at ~30 kDa (Figure 7B and C). This molecular weight is roughly compatible with the hypothesis that the DNA binding peptides of HMG1 dimerize (20 kDa for the A+B 'internal dimer' plus 5 kDa for the acidic tail), and less compatible with the hypothesis that each of them possesses an extremely extended, almost rod-like monomeric conformation (20 kDa for peptide A + 20 kDa for peptide B + 5 kDa for the acidic tail).

Both boxes A and B contain cysteines, and these could be involved in dimerization. We therefore performed SDS–polyacrylamide electrophoresis of the purified peptides HMG1bA and HMG1bB under reducing and non-reducing conditions. Their electrophoretic mobilities were not affected by reduction, implying that the putative dimers are not stabilized through S–S bonds (data not shown).

To substantiate further the finding that HMG boxes form dimers in solution, we performed crosslinking experiments in the absence of DNA. Dilute samples of purified HMG1bA were exposed to low concentrations of glutaraldehyde at 28°C for 1 h, and the reaction products were fractionated by gel electrophoresis. The apparent molecular weight of monomeric HMG1bA in SDS–polyacrylamide gels is 12.5 kDa. Exposure to glutaraldehyde in the range of 20–50 μM results in the appearance of a new diffuse band at an apparent molecular weight of 27.6 kDa, which we interpret as a crosslinked dimeric species of HMG1bA (Figure 8A). Exposure of streptavidin (a homotetrameric protein) to this range of glutaraldehyde concentrations results in the formation of bands corresponding to crosslinked dimers and trimers, while the same treatment of soybean trypsin inhibitor (a monomeric protein) does not lead to the appearance of crosslinked species (results not shown). In addition, both the crosslinked and the non-crosslinked species of HMG1bA emerged in the same fractions from gel filtration (not shown).

Although the results described in the preceding paragraph indicate that HMG boxes form dimers in solutions, they do not rule out the possibility that monomers might bind to DNA. To obtain monomers, we fractionated glutaraldehyde-treated HMG1bA samples by SDS–polyacrylamide electrophoresis and blotted them onto a membrane. The membrane was washed with a strongly denaturing solution to remove all protein not bound directly to the plastic texture. The protein still bound to the filter was then incubated in a solution promoting renaturation (peptide HMG1bA can be reversibly denatured and renatured in solution with minimal loss of activity, unpublished results). Finally, the membrane was incubated with radioactive four-way junction DNA and briefly washed several times: upon exposure, a strong band appeared in the position corresponding to the crosslinked HMG1bA species, whereas no signal was detected in the position corresponding to monomeric HMG1bA not treated with glutaraldehyde (Figure 8B). The faint band

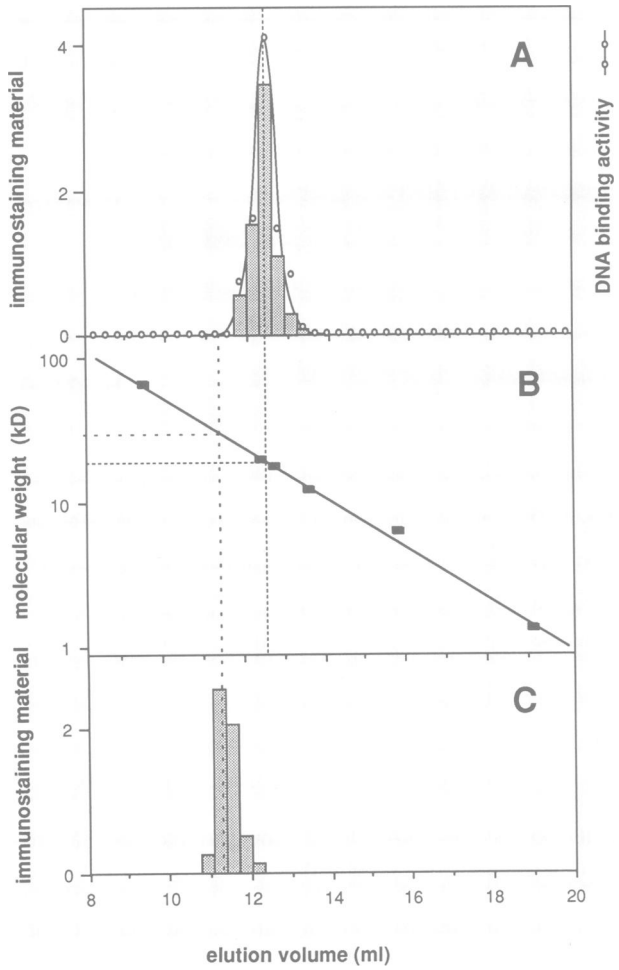

Fig. 7. Determination of the native molecular weight of HMG1bB and full-length HMG1 by gel filtration. **(A)** Peptide HMG1bB was produced by inducing with IPTG *E.coli* cells of the BL21(DE3) strain, transformed with plasmid pT7-HMG1bB. The peptide was partially purified by DEAE-Sepharose chromatography, the preparation was concentrated and 100 μl were applied to an FPLC Superdex 75 column (Pharmacia) as described in Materials and methods. Fractions of 300 μl were collected and analysed for total protein content by SDS−PAGE. The concentration of HMG1bB in each fraction (shaded boxes) was determined by Western blotting with affinity-purified anti-HMG1 rabbit antibody, biotinylated goat anti-rabbit secondary antibody and streptavidin-conjugated alkaline phosphatase. The four-way junction DNA binding activity of each fraction (-○-○-) was assayed by electrophoresis on polyacrylamide gels as described in Materials and methods, followed by densitometry of the autoradiographic signals. The amount of immunostaining material and DNA binding activity are expressed in arbitrary units. **(B)** The elution volume of marker molecules of known native molecular weight (bovine serum albumin, soybean trypsin inhibitor, myoglobin, cytochrome *c*, aprotinin and vitamin B12) were determined by measuring OD_{280nm} in separate runs under identical conditions. **(C)** HMG1 protein was prepared from rat liver as described by Bianchi *et al.* (1989); the chromatographic run and the assays were identical to those reported in (A).

corresponding to the position of monomeric HMG1bA treated with glutaraldehyde is probably an artefact.

Discussion

Two kinds of HMG boxes

Recent work has suggested that several proteins interact with DNA by means of domains (called HMG boxes) that bear a detectable sequence similarity to protein HMG1 of mammals. In this paper we show that rat HMG1 protein

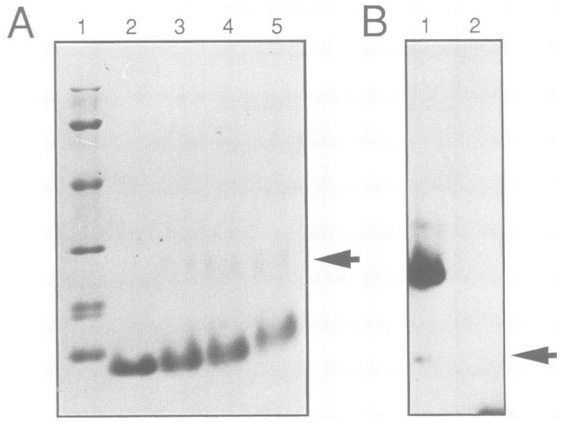

Fig. 8. Glutaraldehyde crosslinking of HMG1bA peptide. **(A)** Formation of crosslinked dimers. Purified HMG1bA was exposed for 1 h to various concentrations of glutaraldehyde (0 μM, lane 2; 20 μM, lane 3; 50 μM, lane 4; 100 μM, lane 5) and resolved in a 16% denaturing SDS−polyacrylamide gel. Lane 1 contains molecular weight markers. The apparent molecular weight of the crosslinked dimer (arrow) is a little higher than twice the apparent molecular weight of the monomer. **(B)** The monomeric form of HMG1bA does not bind DNA. Crosslinked (50 μM glutaraldehyde, lane 1) or non-crosslinked (lane 2) samples of HMG1bA peptide were resolved by SDS−PAGE and blotted onto a membrane as described in Materials and methods. The position of the proteins immobilized on the filter was marked, and the filter was subjected to a denaturing treatment. The filter was then incubated in a solution allowing renaturation of HMG1bA, and was exposed to radioactive four-way junction DNA (c). Radioactivity was retained in the position corresponding to the crosslinked dimer species, with only a trace amount being localized in the position of the glutaraldehyde-treated monomer species. No radioactivity was found in the position corresponding to the monomer species not treated with glutaraldehyde (arrow).

contains two such HMG boxes, and that each box is a free-standing unit capable of binding to DNA. Therefore, mammalian HMG1 is composed almost entirely of two DNA binding peptide modules plus a 30 residue polyacidic tail. As expected, the polyacidic tail is not required for DNA binding, and might be involved in protein−protein interactions.

Our sequence analysis indicates the existence of two types of HMG boxes, which we have called canonical and generalized. Canonical HMG boxes occur in mammalian and fish HMG1-like proteins, protein LG-1 of *T.thermophila* and proteins NHP6A and -B of *S.cerevisiae*. They conform very strictly to the consensus (~50% of residues are conserved) and appear to make up most of the protein to which they belong. The protozoan and yeast proteins indeed contain little else than a single HMG box (and so lack the polyacidic tail). In addition to their compact architecture, proteins displaying the canonical HMG box also share the properties of being abundant and located in the nucleus. Their evolutionary spread suggests that one or more proteins of this type might be present in every eukaryote, and that they might serve an essential function.

Generalized HMG boxes, on the other hand, occur in a rather diversified set of proteins. They too might be present in a single or multiple copies per polypeptide chain, but they represent a limited proportion of the total mass of the protein to which they belong. They diverge considerably among themselves and from the canonical HMG box; however, they are characterized by eight conserved residues (two invariant prolines plus two positively charged and four aromatic amino acids) (Figure 3).

The structure of canonical HMG boxes

We have applied to the nine members of the canonical HMG box family a number of algorithms for the prediction of protein secondary structure. The PepPlot (Gribskov *et al.*, 1986) and SEGMENT83 (C.Kabsch and C.Sander, unpublished) programs are based on statistical considerations (Chou and Fasman, 1978; Garnier *et al.*, 1978), while the ALB program (Finkelstein, 1975; Ptitsyn and Finkelstein, 1983) is based on the physical and stereochemical properties of the amino acids. They produced essentially the same prediction for all sequences: a possible β-strand near the N-terminal part of the sequence and two strongly predicted α-helices comprising the C-terminal half of the sequence (see Figure 3). In a few predictions the two α-helices were fused, but the presence of a proline in one of the yeast proteins appears to confirm the interruption predicted by the large majority of simulations.

The hydrodynamic properties of both HMG boxes of HMG1 suggest that they form dimers in solution. This hypothesis is further supported by the observation that HMG1bA or HMG1bB peptides produce band retardation patterns that are almost indistinguishable from the pattern produced by full-length HMG1, or HMG1 minus the acidic tail. Additional proof of the dimeric nature of HMG boxes was obtained through crosslinking experiments with glutaraldehyde, a rather non-specific bifunctional crosslinking agent (Figure 8). We were unable to detect monomeric subunits in solution, suggesting that the dimeric conformation is strongly favoured. In addition, the monomeric form is inactive in binding DNA, at least when immobilized onto a filter. Dimeric subunit structures are common among DNA binding proteins and are often essential for their DNA binding properties; for example, this is the case for the leucine zipper proteins, such as Fos/Jun (Halazonetis *et al.*, 1988) and the helix-loop-helix families of proteins (Jones, 1990).

The avalaibility of plasmids that direct the efficient expression of HMG boxes and the limited size of the HMG1bA and HMG1bB peptides will make it possible to execute accurate biochemical and biophysical studies, as well as to produce targeted mutations of the HMG box.

The binding specificities of canonical and generalized HMG boxes

Our data show unequivocally that both HMG boxes of HMG1 display a very marked structure discrimination. This property is also exhibited by full-length HMG1: in fact, it was only after we had purified and sequenced a *bona fide* cruciform DNA binding protein (Bianchi, 1988; Bianchi *et al.*, 1989) that we discovered its identity with HMG1. The binding of HMG1 and its individual HMG boxes is structure specific and sequence independent: both the entire protein and its subdomains bind to four-way junction DNAs of unrelated sequence, but not to linear duplex or single-stranded DNAs containing the same sequences as the cruciform molecules (Bianchi, 1988; this paper).

HMG-T, the trout equivalent of mammalian HMG1s, also contains two canonical HMG boxes (Figure 3). The best interpretation of the published data is that HMG-T also binds to cruciform DNA, rather than A-T rich sequences as originally proposed (Wright and Dixon, 1988). HMG-T binds to a stretch of alternating A-T residues upstream of its own gene only when the DNA is supercoiled; under these conditions this DNA site is sensitive to nuclease S1 and is cleaved by endonuclease VII of phage T4, indicating that it adopts a cruciform conformation.

We feel confident that our results may be extrapolated to the other proteins containing canonical HMG boxes. Their sequence conservation is striking, allowing us to predict similar biochemical properties. In addition, preliminary results on yeast NHP6A and B (M.E.Bianchi, unpublished) and one HMG1-like protein cloned from rice (P.Quail, personal communication) appear to validate our assumption.

What about generalized HMG boxes? All existing information indicates that proteins that contain them recognize specific sequences. Human SRY protein appears to bind the consensus sequence AACAAAG (Harley *et al.*, 1992). Human, rat and frog UBF proteins produce distinctive footprints upstream of the transcription start site of rRNA (Jantzen *et al.*, 1990; Pikaard *et al.*, 1990). To date, no structure-specific binding of transcriptional regulators has been demonstrated, although several of these proteins do cause significant three-dimensional distortion of the DNA tract to which they bind (Liu-Johnson *et al.*, 1986; Schroth *et al.*, 1989; Kerppola and Curran, 1991). In our opinion, affinity of some classes of binding domains, including HMG boxes, for unusual DNA conformations is a likely possibility.

The significance of the binding specificity of the canonical HMG box is at present not clear, and neither is the function of HMG-T and HMG1 proteins, although some studies have associated them with transcription as accessory factors of RNA polymerases (Kleinschmidt *et al.*, 1983; Singh and Dixon, 1990). The four-way junction DNA is a complex structure, with several planes of pseudo-symmetry and several charged surfaces and crevices where protein–DNA contacts might be established (Murchie *et al.*, 1989). Locally equivalent three-dimensional surfaces are formed by other DNA conformations: for example, the specific angles formed by the axes of the DNA helices projecting from the central junction of four-way DNA are also present in the kinked helices formed by DNA double strands with bulged bases (Bhattacharyya and Lilley, 1989). In fact, we have preliminary evidence that kinked DNA molecules can be bound by HMG boxes (M.E.Bianchi and D.M.J.Lilley, unpublished results). Therefore, four-way junction DNA might not be the only or the main physiological target for canonical HMG boxes: the high concentration of HMG1-like proteins in nuclei suggests that they might recognize structures commonly present in DNA, perhaps as a consequence of perturbations introduced by the movement of DNA and RNA polymerases along the double helix of DNA. We also wish to stress that although the structure specificity of HMG1 and its boxes is not in doubt, it cannot be taken for granted that the entire protein or its isolated component parts do not possess some sort of binding preference for one or more specific sequences, as yet undiscovered.

Conclusion

We have demonstrated experimentally that two peptides which belong to a proposed new class of DNA binding modules, the HMG box, do bind specifically to DNA and form dimers in solution, even in the absence of the DNA ligand. The specificity of the peptides (towards cruciform structures) is the same as that of the protein they belong to, but presumably different from the binding specificities (towards particular sequences) of the proteins that contain other, more degenerate members of the same class of DNA

binding module. The availability of milligram quantities of the two HMG boxes of HMG1 protein will allow the biochemical and structural analysis of this novel type of DNA binding module to proceed.

Materials and methods

Construction of synthetic junction and linear DNAs

Oligonucletides were synthesized by the phosphotriester method and purified by HPLC. Duplex and four-way junction DNA molecules were obtained by annealing the appropriate oligonucleotides, and were purified by gel electrophoresis. When appropriate, one of the strands was labelled with T4 polynucleotide kinase before annealing. Details of these preparations have been described by Bianchi (1988).

Two different four-way junction DNA molecules have been used, identified by the letters **c** and **f**. Junction **c** is composed of four strands of 30, 35, 40 and 46 nucleotides. Junction **f** is composed of four strands of 46, 47, 50 and 50 nucleotides, without any sequence similarity to those composing junction **c**. As controls for structure-specific binding, two linear duplex DNAs were used, called **a** and **b**. Molecule **a** is composed of the 35 nucleotide strand of junction **c** annealed to its anti-parallel complement; molecule **b** is composed of the 40 nucleotide strand of junction **c** annealed to its antiparallel complement. Therefore molecules **a** and **b** together have the same sequences present on junction **c**, the same number of double-stranded ends, the same approximate mass, but a different three-dimensional structure. Details on the construction of these molecules are given in Bianchi *et al.* (1989).

Construction of C-terminal deletion derivatives from plasmid pRNHMG1

Plasmid pRNHMG1 (~ 10 μg) was digested with restriction endonuclease *Xba*I, and its ends were filled in with DNA polymerase I large fragment in the presence of dCTP, dGTP, TTP and dATP-αS (from Pharmacia). The DNA was then purified by phenol extraction, recut with restriction endonuclease *Bam*HI, re-extracted with phenol, and purified by spinning through a Sepharose G-100 column equilibrated with 10 mM Tris−HCl (pH 7.5), 1 mM EDTA. The DNA was then incubated at 30°C with *E.coli* exonuclease III (60 U) in 400 μl of a buffer containing 50 mM Tris−HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT. At 30 s intervals, 20 μl aliquots were withdrawn from the exonuclease reaction mixture and mixed with 124 μl H_2O and 16 μl of 10 × mung bean nuclease buffer (300 mM Na acetate pH 4.6, 500 mM NaCl, 10 mM $ZnCl_2$). Nineteen such samples were then heated at 68°C for 15 min to inactivate exonuclease III. To each sample we then added 1 μg of yeast tRNA and 2 U of mung bean nuclease, and we incubated them at 30°C for 30 min. The samples were purified by phenol extraction and ethanol precipitation, ligated with T4 ligase and introduced by transformation in *E.coli* cells. DNA was prepared by a fast method from ~ 150 individual transformants and the size of the deletion was gauged by electrophoresis on agarose gels. DNAs of the appropriate size were sequenced with T7 DNA polymerase (Pharmacia): we found 34 different deletions, ranging from 64 to 802 bp. Thirteen of the deletions had brought the stop codon TAG in frame to the amino acid sequence of HMG1 protein.

Construction of other plasmids coding for derivatives of HMG1 protein

Plasmid pRNHMG1/M52−E215 was constructed by cleaving plasmid pRNHMG1 with restriction endonucleases *Eco*RI and *Tth*111I, filling in the ends with DNA polymerase I large fragment and ligating.

Plasmid pRNHMG1/M63−E215 was constructed by cleaving plasmid pRNHMG1 with restriction endonucleases *Eco*RI and *Tth*111I, blunting in the ends with mung bean nuclease and ligating.

Plasmid pRNHMG1/M132−E215 was constructed by ligating the *Hph*I−*Hin*dIII fragment of plasmid pRNHMG1 (after blunting the *Hph*I end with mung bean nuclease) between the *Eco*RI and *Hin*dIII sites of plasmid pTZ18R (after blunting the *Eco*RI end with mung bean nuclease).

Plasmid pT7-HMG1bA was constructed by swapping the *Nsi*I−*Hin*dIII fragments of plasmids pT7-RNHMG1 (Bianchi, 1991) and pRNHMG1/M1−F89. It directs in *E.coli* the synthesis of a polypeptide corresponding to the first 89 amino acids of HMG1 (HMG1/M1−F89).

Plasmid pT7-HMG1bB was constructed by ligating the *Ava*II−*Hin*dIII fragment of plasmid pRNHMG1/M1−V176 (after filling in the *Ava*II end with DNA polymerase I large fragment) with plasmid pT7-7, which had been cut with restriction endonucleases *Eco*RI and *Hin*dIII, and whose *Eco*RI end had been filled in with DNA polymerase I large fragment. It directs in *E.coli* the synthesis of a polypeptide corresponding to the first four amino acids of the *gene10* product of phage T4 (MARI) followed by amino acids D91−V176 of HMG1.

Plasmid pT7-HMG1/M1−P81 was constructed by *in vitro* mutagenesis of the K82 codon (AAA) of plasmid pT7-HMG1bA to a stop codon (TAA). It directs in *E.coli* the synthesis of a polypeptide corresponding to the first 81 amino acids of HMG1 (HMG1/M1−P81).

All constructions were checked by sequencing with T7 DNA polymerase.

Transcription and translation in vitro

Plasmids were linearized with restriction endonuclease *Hin*dIII and transcribed with T7 RNA polymerase (Stratagene). RNAs were translated with rabbit reticulocyte lysate or wheat germ extracts (Promega) following the instructions of the producer. L-[^{35}S]methionine (specific activity 70−85 Ci/mmol, from Amersham) was used as radioactive precursor. Translation products were checked by electrophoresis on 18% SDS−polyacrylamide gels and autoradiography.

The concentration of HMG1-related protein present in each translation mixture was estimated by determining the amount of ^{35}S label incorporated in TCA-precipitable material, and dividing this figure by the number of methionine residues in the polypeptide chain and the specific activity of the methionine precursor used. Since the amount of translation products present in different translation mixtures was variable, we equalized the concentration of HMG1-related protein by diluting the RNA-programmed translation mixtures with a control mixture where no RNA had been added; typical concentrations were 0.1−0.5 nM.

Assay for DNA binding of in vitro synthesized HMG1 and HMG1-derived polypeptides

DNA binding mixtures (10 μl total volume) contained 8% Ficoll, 200 mM NaCl, 10 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 5 mM KCl, 1 mM EDTA, 1 mM spermidine, 0.5 mM DTT and, as indicated in the legends to the figures, sonicated salmon DNA, *Hin*fI-cleaved pUC19 DNA and synthetic four-way junction DNA (molecule **f**). To these, we added 1 μl of unfractionated *in vitro* translation products. After incubation for 10 min on ice, samples (7 μl) were applied to a horizontal 2% agarose gel in 0.5 × TBE and electrophoresed at 5.5 V/cm for 3 h at 0°C. The gel was then washed extensively in 25% ethanol, 10% acetic acid to remove free methionine, soaked in Amplify fluorographic reagent (Amersham), dried and autoradiographed with Kodak XAR-5 films at −80°C for 24−72 h with intensifying screens.

Preparation of peptides HMG1bA and HMG1bB

Peptides HMG1bA, HMG1bB and HMG1/M1−P81 were synthesized in *E.coli* strain BL21(DE3) (Studier and Moffat, 1986). Plasmid-bearing colonies were selected by growth on LB plates containing 0.4% glucose and 100 μg/ml ampicillin; one colony of each type was diluted in LB broth plus 15% glycerol and kept frozen at −80°C, serving as reference stock to avoid plasmid instability problems. Glucose was always added to LB broth when these cells were grown, in order to avoid expression of the sequences controlled by the T7 promoter. To obtain expression, cells were resuspended to an optical density of ~ 0.05 at 590 nm in 150 ml of LB broth plus 100 μg/ml ampicillin, grown with strong aeration at 37°C to an optical density of ~ 0.7, and induced with 0.5 mM IPTG. After 100 min at 37°C, the cells were harvested, resuspended on ice in 7.5 ml of a buffer containing 50 mM Tris pH 8.0, 20 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF and sonicated. We added 0.75 ml of 5 M NaCl to the lysate and centrifuged it for 10 min at 4°C in Eppendorf tubes. We recovered the supernatant, mixed it with 0.5 ml of glycerol, and kept it at −20°C until further use (fraction I). An aliquot was analysed by SDS−PAGE to measure the amount of HMG1-related polypeptides produced. At this level, they always represented the major protein species in the extract.

For purification to homogeneity, the extracts were batch-adsorbed to DEAE-cellulose equilibrated in the same buffer; the supernatants were mixed with solid ammonium sulfate (70% of maximum solubility at 0°C) and centrifuged. The supernatant was subjected to FPLC on phenyl-Superose (Pharmacia). The first peak eluted (~ 1.5 M ammonium sulfate) contained the HMG boxes, which appeared completely homogeneous by SDS−polyacrylamide electrophoresis. The preparations of purified peptides were concentrated with Centricon-10 cartridges and diluted with buffer G, containing 10 mM potassium phosphate, pH 7.5, 100 mM NaCl, 0.5 mM DTT and 0.1 mM EDTA. Several cycles of dilution−concentration ensured the complete exchange of buffers. The preparations were stable in buffer G for several days at room temperature. For longer term storage, they were frozen and kept at −20°C. Several cycles of freezing and thawing produced no detectable loss of DNA binding activity.

Assay for DNA binding of HMG box peptides

DNA binding mixtures (10 μl of final volume) contained 8% Ficoll, 200 mM NaCl, 10 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 5 mM KCl, 1 mM EDTA, 1 mM spermidine, 0.5 mM DTT. To these components, we added in various combinations 1.5 nM synthetic four-way junction DNA (molecule **c**), labelled with T4 polynucleotide kinase, unlabelled four-way junction DNA (molecule **c**), unlabelled linear duplex DNA (molecules **a** plus **b**) or sonicated salmon sperm DNA (final concentrations are indicated in the legends to the figures). HMG1bA or HMG1bB polypeptides were incorporated at the indicated final concentrations. The order of addition was (i) fixed components, (ii) polypeptides, (iii) DNAs (labelled probe plus cold competitors). In separate experiments (not shown) the order of addition was varied in order to verify that the equilibrium between free and bound ligands had been reached. After incubation for 10 min on ice, samples (5 μl) were applied to vertical 6.5% polyacrylamide gels in 0.5 × TBE and electrophoresed at 11 V/cm for ~3 h at room temperature. The gel was then fixed in 10% acetic acid, dried, and autoradiographed with Kodak XAR-5 films at −80°C for 24−72 h with intensifying screens.

Glutaraldehyde crosslinking

Purified HMG1bA peptide (500 μg/ml in buffer G) was diluted to a final concentration of ~10 μg/ml in a buffer containing 10 mM potassium phosphate, pH 7.5 and 250 mM NaCl. Aliquots of the protein solution (500 μl each) were mixed with glutaraldehyde at final concentrations of 100, 50, 20 or 0 μM, and incubated at 28°C for 1 h. Similar reactions were set up for samples containing control proteins (streptavidin and soybean trypsin inhibitor). The reactions were then quenched with 5 mM ethanolamine pH 8.3. The contents of each tube were dried down in a Savant concentrator, resuspended in 40 μl of distilled water, precipitated with 10 μl of trichloroacetic acid (100% w/v), washed with acetone, dried and eventually loaded onto a 16% SDS−polyacrylamide gel.

To test the DNA binding ability of the monomeric and crosslinked dimeric forms of HMG1bA, samples were treated as indicated in the preceding paragraph and separated by SDS−PAGE. The separated proteins were blotted onto an Immobilon membrane, which was sequentially washed at room temperature with distilled water, Ponceau Red solution (obtained from Sigma) to locate the transferred proteins, 100 mM Tris pH 8.0 to destain the membrane, and with a solution containing 6 M guanidinium chloride, 50 mM Tris pH 8.0, 10 mM DTT, 1 mM EDTA and 0.1% (v/v) Nonidet P-40. The membrane was then extensively washed with buffer D-200 (containing 20 mM HEPES pH 7.9, 200 mM NaCl, 0.5 mM DTT, 0.2 mM EDTA, 0.05% Nonidet P-40 and 10% glycerol) and incubated overnight in the same buffer. The membrane was then incubated for 15 min in DNA binding solution (8% Ficoll, 200 mM NaCl, 10 mM $MgCl_2$, 10 mM HEPES, pH 7.9, 5 mM KCl, 1 mM EDTA, 1 mM spermidine, 0.5 mM DTT) containing 10 nM synthetic four-way junction DNA (molecule **c**), labelled with T4 polynucleotide kinase to a specific activity of nearly 5 μCi/pmol. The membrane was then washed three times for 2 min in buffer D-200, dried and autoradiographed with Kodak XAR-5 film at −80°C for 10 days with intensifying screens.

Acknowledgements

We thank Alastair Murchie for synthesizing the oligonucleotides for *in vitro* mutagenesis, Enzo Notarianni for help in the construction of some of the plasmids, Dr Raffaele De Francesco for helpful discussions, Dr Monica Beltrame for her encouragement and suggestions, and Drs Fabio Cobianchi and Martino Bolognesi for carefully reading the manuscript. This work was partially supported by a grant from Ministero della Ricerca Scientifica e Tecnologica of Italy (MURST). M.E.B. gratefully acknowledges the receipt of an EMBO Short Term Fellowship that enabled him to work temporarily in Dundee. The collaboration was then continued with the financial help of a British Council/MURST programme.

References

Bhattacharyya,A. and Lilley,D.M.J. (1989) *Nucleic Acids Res.*, **17**, 6821−6840.
Bianchi,M.E. (1988) *EMBO J.*, **7**, 843−849.
Bianchi,M.E. (1991) *Gene*, **104**, 271−275.
Bianchi,M.E., Beltrame,M. and Paonessa,G. (1989) *Science*, **243**, 1056−1059.
Bonne-Andrea,C., Harper,F., Sobczak,J. and De Recondo,A.-M. (1984) *EMBO J.*, **5**, 1193−1199.
Chou,P.J. and Fasman G.D. (1978) *Adv. Enzymol.*, **47**, 45−147.
Einck,L. and Bustin,M. (1985) *Exp. Cell Res.*, **156**, 295−310.
Finkelstein,A.V. (1975) *Dokl. Akad. Nauk SSSR*, **223**, 744−747.
Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97−120.
Gribskov,M., Burgess,R.R. and Devereux,J. (1986) *Nucleic Acids Res.*, **14**, 327−334.
Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Muensterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245−250.
Halazonetis,T.D., Georgopoulos,K., Greenberg,M.E. and Leder,P. (1988) *Cell*, **55**, 917−924.
Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R., Berkovitz,J.D., Sockanathan,S., Lovell-Badge,R. and Goodfellow,P.N. (1992) *Science*, in press.
Jantzen,H.-M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830−836.
Jones,N. (1990) *Cell*, **61**, 9−11.
Kaplan,D.J. and Duncan,C.H. (1988) *Nucleic Acids Res.*, **16**, 10375.
Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537−1547.
Kerppola,T.K. and Curran,T. (1991) *Cell*, **66**, 317−326.
Kleinschmidt,J.A., Scheer,U., Dabauvalle,M.-C., Bustin,M. and Franke,W.W. (1983) *J. Cell Biol.*, **97**, 838−848.
Kohlstaedt,L.A., Sung,E.C., Fujishige,A. and Cole,R.D. (1986) *J. Biol. Chem.*, **262**, 524−526.
Kolodrubetz,D. and Burgum,A. (1989) *J. Biol. Chem.*, **265**, 3234−3239.
Liu-Johnson,H.-N., Gartenberg,M.R. and Crothers,D.M. (1986) *Cell*, **47**, 995−1005.
Murchie,A.I.H., Clegg,R.M., von Kitzing,E., Duckett,D.R., Diekmann,S. and Lilley,D.M.J. (1989) *Nature*, **341**, 763−766.
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.
Pikaard,C.S., Smith,S.D., Reeder,R.H. and Rothblum,L. (1990) *Cell. Mol. Biol.*, **10**, 3810−3812.
Ptitsyn,O.B. and Finkelstein, A.V. (1983) *Biopolymers*, **22**, 15−25.
Reeck,G.R., Isackson,P.J. and Teller,D.C. (1982) *Nature*, **300**, 76−78.
Roth,S.Y., Schulman,I.G., Cook,R.G. and Allis,C.D. (1987) *Nucleic Acids Res.*, **19**, 8112.
Schroth,G.P., Cook,G.R., Bradbury,E.M. and Gottesfeld,J.M. (1989) *Nature*, **340**, 487−488.
Shirakawa,H., Tsuda,K. and Yoshida,M. (1990) *Biochemistry*, **18**, 4419−4423.
Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischauf,A.-M., Lovell-Badge,R. and Goodfellow,P.N. (1990) *Nature*, **346**, 240−244.
Singh,J. and Dixon,G.H. (1990) *Biochemistry*, **29**, 6295−6302.
Studier,F.W. and Moffatt,B.A. (1986) *J. Mol. Biol.*, **189**, 113−130.
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.*, **4**, 1471−1486.
Tsuda,K., Kikuchi,M., Mori,K., Waga,S. and Yoshida,M. (1988) *Biochemistry*, **27**, 6159−6163.
Watt, F. and Molloy,P.L. (1988) *Nucleic Acids Res.*, **16**, 1471−1486.
Wen,L., Huang,J.-K., Johnson,B.H. and Reeck,G.R. (1989) *Nucleic Acids Res.*, **17**, 1197−1214.
Wright,J.M. and Dixon,G.H. (1988) *Biochemistry*, **27**, 576−581.

Received on August 14, 1991; revised on November 26, 1991