Nucleic Acids Research, 1994, Vol. 22, No. 3 285–292

Mutational analysis of the DNA binding domain A of chromosomal protein HMG1

Luca Falciola[1], Alastair I.H.Murchie[2], David M.J.Lilley[2] and Marco E.Bianchi[1,3,*]
[1]DIBIT, San Raffaele Scientific Institute, via Olgettina 58, 20132 Milan, Italy, [2]Department of Biochemistry, The University, Dundee DD1 4HN, UK and [3]Dipartimento di Genetica e di Biologia dei Microrganismi, Universitá di Milano, via Celoria 26, 20133 Milan, Italy

Received November 9, 1993; Revised and Accepted January 10, 1994

ABSTRACT

We have mutated several residues of the first of the two HMG-boxes of mammalian HMG1. Some mutants cannot be produced in *Escherichia coli*, suggesting that the peptide fold is grossly disrupted. A few others can be produced efficiently and have normal DNA binding affinity and specificity; however, they are more sensitive towards heating and chaotropic agents than the wild type polypeptide. Significantly, the mutation of the single most conserved residue in the rather diverged HMG-box family falls in this 'in vitro temperature-sensitive' category, rather than in the non-folded category. Finally, two other mutants have reduced DNA binding affinity but unchanged binding specificity. Overall, it appears that whenever the HMG-box can fold, it will interact specifically with kinked DNA.

INTRODUCTION

HMG1 and its close relative HMG2 belong to the 'high mobility group' class of mammalian chromatin proteins. They are present in lower amounts than histones, but are nonetheless very abundant: in some cells there is an HMG1 molecule every 3 kb of DNA on average (1). Similar proteins exist in fungi, protozoa, plants and arthropods. Their sequence conservation, ubiquity and abundance all point to an important function for HMG1-like proteins: roles have been suggested in replication, transcription and chromatin assembly (reviewed in (2)), but none has been demonstrated beyond doubt. Proof that HMG1-like proteins are essential for cell life is also missing.

HMG1 binds to single-stranded and double-stranded DNA with moderate affinity and apparent lack of sequence specificity (3, 4); however, it binds preferentially to a limited set of sites on supercoiled plasmids (5). We and others have shown that in fact HMG1 has a considerable selectivity for DNA structures containing sharp bend or kinks: four-way junctions (6, 7), DNA covalently modified by the antitumour drug cisplatin (8), or DNA duplexes containing unpaired bases ('bulges') (Pontiggia and Bianchi, unpublished).

The structure of HMG1 is tripartite: the N-terminal A-domain and the central B-domain are positively charged and bind DNA, while the C-terminal domain is acidic and has been shown to interact with histones (9, 10). Domains A and B can be expressed independently in *E. coli*: each of them retains the full DNA binding affinity and the same structure selectivity as the entire protein (11). The sequences of the two domains are similar to each other and to segments of about 70 amino acids called 'HMG-box motifs' (12) present in UBF, a general transcription factor for RNA polymerase I (13), in the testis-determining factor SRY (14), and in a growing number of other proteins (15).

The HMG-box motif is the hallmark of a new class of proteins with unusual DNA binding properties: the HMG-box proteins. They comprise chromatin proteins, general transcription factors and gene-specific transcriptional regulators, all characterised by the ability to recognise kinked DNA or to generate kinks upon binding to specific sequences (16). In particular, both the lymphoid-specific protein LEF-1 and SRY recognise sites related to the consensus sequence AACAAAG (17, 18), and deflect the axis of the double strand by about 130° and 85° respectively (19, 20). The production of such distortions in DNA is quite extraordinary and is possibly the actual biological function of these proteins, which may control transcription by creating the correct spatial arrangement of specific DNA segments and transcriptional factors bound to them.

The structure of one HMG-box domain, box B of HMG1, has been recently solved using NMR techniques by two independent groups (21, 22). Its fold is dominated by three helical segments linked by short turns, and arranged in a remarkably flat and L-shaped form. There is no discernible similarity to helix-turn-helix proteins or other previously described folds. This, together with the sequence conservation and the common DNA binding properties, confirms that the HMG-box domain is indeed a new structural element. There is at present no structural information on how HMG-boxes lock onto DNA, beyond the obvious complementarity between the arrowhead form of the B-domain of HMG1 and the angle formed by the targets of SRY and LEF-1. However, methylation interference, base substitution and hydroxyl radical protection experiments suggest that the

*To whom correspondence should be addressed at DIBIT, San Raffaele Scientific Institute, via Olgettina 58, 20132 Milan, Italy

interaction between HMG-boxes and DNA occurs mainly through the minor groove (19, 23, Pontiggia *et al.*, unpublished).

Given the inherent interest of HMG-box domains, and in order to investigate the mode of their interaction with DNA, we undertook a mutational analysis of HMG-box A of HMG1. Our choice of mutations predated the release of the structural data on the B domain of HMG1, and was primarily based on the HMG-box consensus and on a computer prediction of secondary structure (12). Although our mutations cannot be directly related to the published structure of box B, which is only 28% identical in sequence to box A, they test several critical features which appear to be common to HMG-boxes in general. In particular, our results show that some conserved residues are important for domain folding, while others are important for the structure-specific DNA binding. Surprisingly, the best-conserved residue in HMG-boxes, the tryptophan that sits in the vertex of the arrowhead-shaped fold, is not critical for either folding or DNA binding.

MATERIALS AND METHODS

Construction of plasmids coding for mutants of the HMG1bA polypeptide

Plasmid pT7-HMG1bA, directing the synthesis of polypeptide HMG1bA, has been described (11). Polypeptide HMG1bA-K82Z is an alternative name for HMG1/M1-P81, whose synthesis is directed by plasmid pT7-HMG1/M1-P81 (11). Mutants of both plasmids were constructed using a directed mutagenesis technique based on the Polymerase Chain Reaction (24). For each mutant, we synthesised two oligonucleotides bearing the same name and distinguished by the suffix 'dir' (for the coding strand sequence) or 'rev' (for the non-coding strand sequence), as indicated below:

oligonucleotide name	oligonucleotide sequence
R10Gdir	AAGAAGCCGGGAGGCAAAAT
R10Grev	TTTGCCTCCCGGCTTCTTA
Y16Sdir	GTCCTCATCTGCATTCTTTGTG
Y16Srev	GCACAAAGAATGCAGATGAGGAC
F19Adir	CATATGCATTCGCGGTGCAA
F19Arev	TTTGCACAAAGAATGCATATG
K28Edir	GCCGGGAGGAGCACGAGAAGAAGCAC
K28Erev	GCTTCTTCTCGTGCTCCTCCCGGC
P32Adir	AAGAAGAAGCACGCGGATGCTTCT
P32Arev	GACAGAAGCATCCGCGTGCTTCTTCTT
F38Sdir	TCTGTCAACTCCTCAGAGTTCT
F38Srev	GGAGAACTCTGAGGAGTTGACAGA
S42Fdir	CAGAGTTCTTCAAGAAGTGCTC
S42Frev	GAGCACTTCTTGAAGAACTCT
S46Pdir	CCAAGAAGTGCCCCGAGAGGTGG
S46Prev	CCACCTCTCGGGGCACTTCTTGG
W49Rdir	CTCAGTGTGGCGTAAGACC
W49Rrev	CATGGTCTTACGCCTCTCTGAGC
F60Sdir	AAGGGGAAATCTGAAGATATG
F60Srev	GGCCATTATCTTCAGATTTCCCCTT
Y71Gdir	GACAAGGCTCGTGGTGAAAGAGAA
Y71Grev	TTCTCTTTCACCACGAGCCTTGTC

In the first round of PCR, two fragments were amplified: the 'rev' fragment using the T7 primer (5'-TAATACGACTCAC-TATAGGGAGA-3') and the 'rev' primer on pT7-HMG1bA template, and the 'dir' fragment using the M13(−20) primer (5'-TTGTAAAACGACGGCCAGTG-3') and the 'dir' primer on pRNHMG1bA or pRNHMG1/M1-P81 templates (11). The dir and rev PCR products were gel purified, mixed and used as template in a second round of PCR using the T7 and M13(−20) primers. The recombined PCR product was purified by agarose gel electrophoresis, cut with *Xba*I and *Hind*III and cloned between the *Hind*III and *Xba*I sites of plasmid pT7-7 (25). Ligation products were cloned into the *E.coli* DH5α strain, recovered and checked by completely sequencing the insert. Final constructs were then introduced into the *E.coli* BL21(DE3) strain (26).

Preparation of peptide HMG1bA and derived mutants

Peptides were prepared according to a published protocol (11). Briefly, cultures of *E.coli* strain BL21(DE3) bearing the appropriate plasmids were grown in LB broth plus 100 μg/ml ampicillin at 37°C to an optical density of about 0.7−1.0 at 590 nm, and induced with 1−2 mM IPTG for 100 minutes. The cells were then harvested, resuspended on ice in lysis buffer and sonicated. The lysates were batch-adsorbed to DEAE-cellulose equilibrated in lysis buffer; the supernatants were mixed with solid ammonium sulphate (70% of maximum solubility at 0°C) and centrifuged. The supernatants were fractionated by FPLC on Phenyl-superose (Pharmacia). The fractions containing the pure peptides were concentrated with Centricon-10 cartridges and diluted with buffer F (20 mM HEPES pH 7.8, 100 mM NaCl, 0.5 mM DTT, 0.1 mM EDTA, 5% glycerol). No particular difference was found in the purification of the various mutants, apart for the lower expression of some of them and minor variations in the elution profile on Phenyl-superose. The concentration of peptides was estimated in two ways: using the biuret assay for ones produced in large quantities (HMG1bA , HMG1bA-K82Z, HMG1bA-R10G), and absorbance at 280 nM. The relative concentration of polypeptides was also checked by visually comparing their amounts on PAGE-SDS gels.

Construction of junction and linear DNAs

All oligonucleotides were synthesised by the phosphotriester method and purified either by HPLC or gel electrophoresis. 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 (7). Junction z is composed of four strands of 30 nucleotides each; as controls for structure-specific binding we used linear duplex DNAs az and bz, with the same sequences of the junction (20).

Electrophoretic mobility shift assay

DNA binding buffer (10 μl of final volume) contained 8% Ficoll, 200 mM NaCl, 10 mM HEPES, pH 7.9, 0.05% NP-40, 0.5 mM DTT. To these components, we added in various combinations synthetic four-way junction DNAs, linear duplex DNAs and purified polypeptides (final concentrations are indicated in the legends to the figures). After mixing on ice, samples were applied to vertical 6.5% polyacrylamide gels in 0.5×TBE and electrophoresed at 11 V/cm at room temperature. The gel was then dried, and autoradiographed at −80°C for 12−24 hours with intensifying screens.

The DNA binding assays in the presence of urea were performed at room temperature by mixing together the polypeptide and the urea solution buffered with HEPES to pH 7.8 (5 μl), and adding 5 μl of mix containing all the other components. The indicated urea concentrations refer to the final 10 μl mix.

Heat resistance experiments were performed by incubating 10−100 ng of polypeptides in 3−5 μl of storage buffer F (20 mM HEPES pH 7.8, 100 mM NaCl, 0.5 mM DTT, 0.1 mM EDTA, 5% glycerol) at the indicated temperature for 30 minutes, cooling on ice, and immediately adding the other components.

RESULTS

Design of the mutations in the HMG-box domain A of HMG1

We previously showed that the segment spanning from Met1 to Phe89 of HMG1 retains the full DNA binding capabilities of the entire protein, and so by definition comprises the HMG-box domain A. We used this peptide, which we will call HMG1bA for brevity, as the wild type against which to compare our set of mutations. Chromosomal proteins most similar to HMG1 (class I) are assumed to perform similar if yet undefined functions: their comparison allows us to define the most conserved residues (a selection of these proteins is shown in Fig 1A). In addition to these, a database search identifies close to 100 HMG-boxes in proteins of the ribosomal and mitochondrial transcription factor class (class II), and in gene-specific transcription regulator proteins (class III). This sequence information was also considered in the selection of residues to be mutated.

Arg10, Tyr16, Phe19, Pro32, Trp49, Phe60 were chosen as targets for mutagenesis because of their strong conservation across class boundaries. Tyr71 and Lys28 are strongly conserved in class

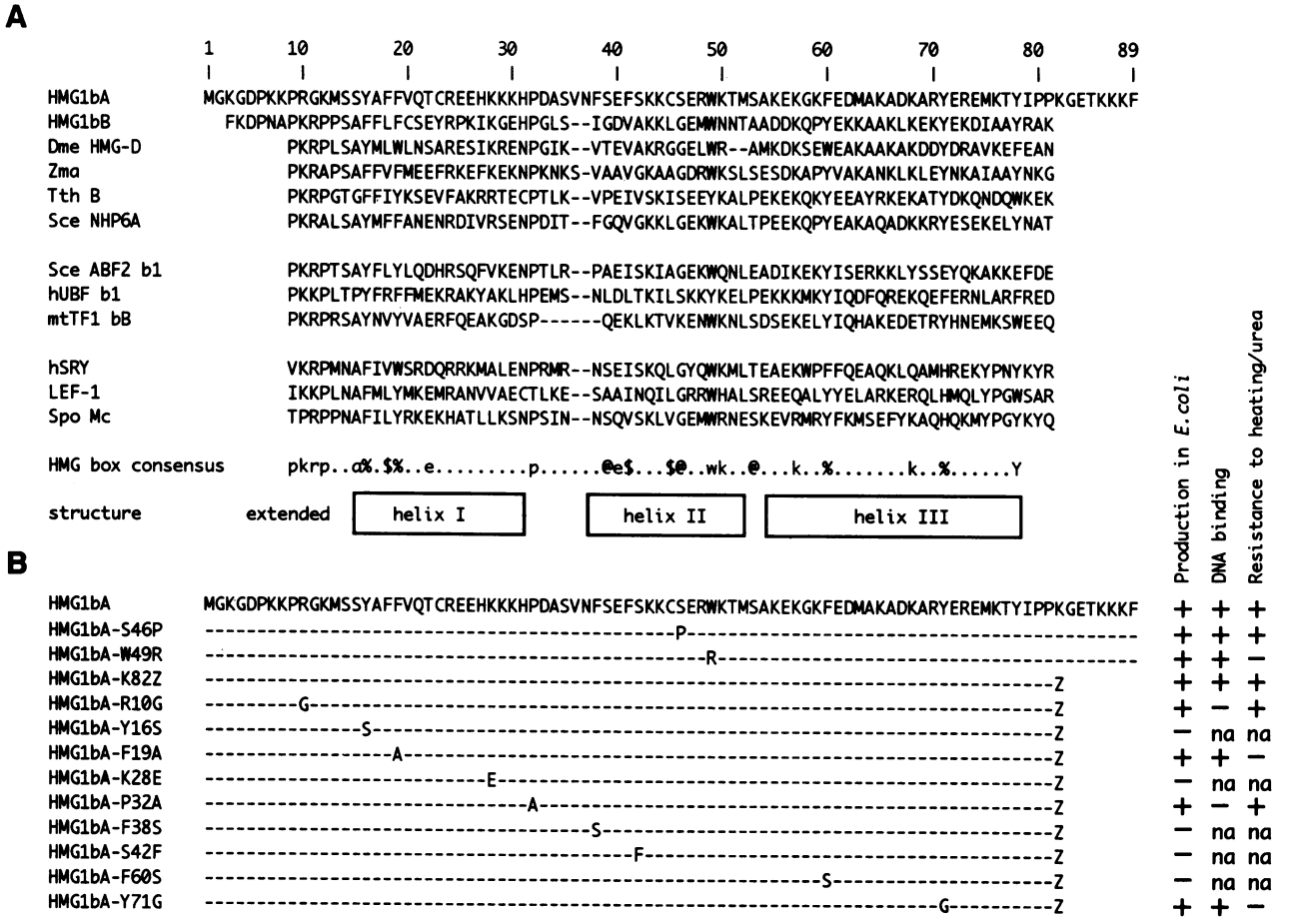

Figure 1. **A)** Sequence alignments of representative HMG-boxes. The first group of sequences are HMG-boxes are from chromatin proteins of various organisms. HMG1bA and HMG1bB are the two boxes of mammalian HMG1s. HMG1bA is the fragment of HMG1 (residues 1 to 89) used as wild type in this work, while HMG1bB is the fragment of rat HMG1 (residues 88 to 164) whose structure has been solved by Weir *et al.* (21); their sequence is written out in full: the last amino acid in HMG1bA is the same phenylalanine which is the first amino acid in HMG1bB. HMG-D is from *Drosophila melanogaster* (34), Zma from maize (35), Tth B from the protozoan *Tetrahymena thermophila* (36), Sce NHP6A from the budding yeast *Saccharomyces cerevisiae* (37). The second group of sequences are HMG-boxes from general transcription factors. Sce ABF2-b1 is first of the two boxes of a yeast mitochondrial protein (38), hUBF-b1 is the first of the four boxes of the human RNA polymerase I transcription factor UBF (13), mtTF1-bB is the second of the two boxes of the human mitochondrial transcription factor mtTF1 (39). The third group of sequences are HMG-boxes from sequence-specific transcription factors: hSRY is the single box of the human sex-determining factor (14), LEF-1 is the single box of the mouse T-cell specific LEF-1 protein (17), Spo Mc is the single box of the Mc protein of the fission yeast *Schizosaccharomyces pombe* (40). Gaps in the alignment are indicated by dashes. The consensus sequence for the HMG box motif was obtained from a much wider set of sequences: one letter symbols indicate individual amino acids present in more than half of the sequences; conservative substitutions (at least 75% of the occurrences at a particular position) are indicated as follows: @ for proline, alanine, glycine, serine and threonine; % for phenylalanine, tyrosine and tryptophan; $ for methionine, valine, leucine and isoleucine. The secondary structure refers to box B of mammalian HMG1 (21). **B)** Mutagenesis of HMG1bA. The sequences of the mutants are aligned to wild type HMG1bA (top line); dashes indicate unchanged amino acids, Z indicates a stop codon. On the right the effects of the mutagenesis are summarised (see text). A minus sign indicates that the peptide was produced in extremely low amounts (column 'Production in *E. coli*'), that its four-way junction binding affinity was reduced compared to HMG1bA (column 'DNA binding'), or that its DNA binding activity was severely reduced after heating or in the presence of 4 M urea (column 'Resistance to heating/urea'); 'na' stands for 'not assayed'.

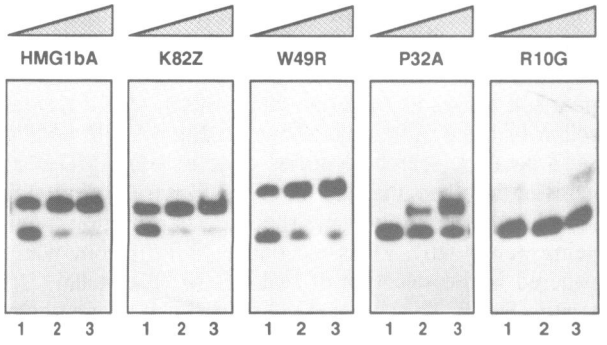

Figure 2. DNA binding properties of some HMG1bA mutants. DNA binding mixtures were set up to contain 1.5 nM labelled four-way junction DNA z. Each triplet of lanes contains the following amounts of the indicated peptide: lanes 1, 2 nM; lanes 2, 10 nM; lanes 3, 20 nM. After incubation on ice, the samples were assayed by electrophoresis as described in Materials and Methods. The faster migrating band is free junction DNA, the slower is the HMG box-junction complex. The uneven relative position of the protein−DNA complexes in this figure is due to uneven migration conditions in different gels; however, the various mutations do alter the mobility of the complexes in a very small but reproducible way.

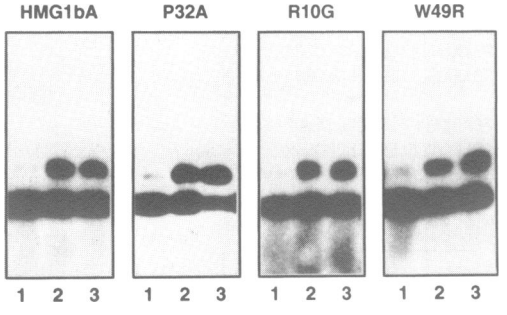

Figure 3. Structure-specificity of selected HMG1bA mutants. DNA binding mixtures were set up to contain 1.5 nM labelled four-way junction DNA z (lanes 2), or the same concentration of labelled junction plus 200 nM unlabelled junction (lanes 1), or the same concentration of labelled junction plus unlabelled linear duplexes az and bz (each 200 nM) (lanes 3). Final concentrations of polypeptides were 2 nM for HMG1bA and HMG1bA-W49R, 10 nM for HMG1bA-P32A, and 25 nM for HMG1bA-R10G. After incubation on ice, the samples were assayed by electrophoresis as described in Materials and Methods. The faster migrating band is free junction DNA, the slower is the HMG box-junction complex.

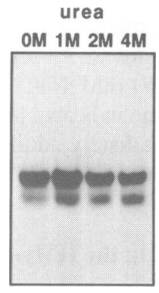

Figure 4. DNA binding activity of HMG1bA in the presence of urea. Urea was added at the indicated final concentrations to the DNA binding mixtures containing 1.5 nM labelled four-way junction DNA z and 2 nM HMG1bA; the samples were incubated at room temperature for 15 minutes and assayed as described.

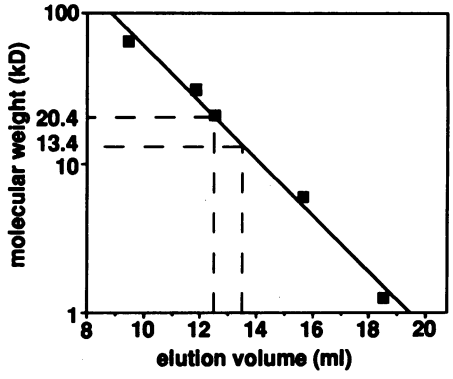

Figure 5. Sensitivity of the DNA binding activity of mutants towards heating and urea. Sensitivity to urea: DNA binding mixtures were set up to contain 1.5 nM labelled four-way junction DNA z and 2 nM of protein, except for HMG1bA-R10G (20 nM) and HMG1bA-P32A (10 nM), and the indicated final concentrations of urea. The assays were performed as described in Material and Methods. Sensitivity to heating: peptides were incubated at the indicated temperatures for 30 minutes in their storage buffer (see Materials and Methods). Samples were then added to DNA binding mixtures containing 1.5 nM labelled four-way junction DNA z and immediately assayed. The final concentration of polypeptides was: HMG1bA-R10G: 20 nM; HMG1bA-P32A, 10 nM; all others, 2 nM. The remaining activity was normalised to the activity of samples kept on ice. The points in the figure are the average of several independent experiments. Peptides are indicated as follows: wt for HMG1bA and HMG1bA-K82Z; F for HMG1bA-F19A; P for HMG1bA-P32A; R for HMG1bA-R10G; S for HMG1bA-S46P; W for HMG1bA-W49R; Y for HMG1bA-Y71G.

I but less so in classes II and III, while Phe38 and Ser42 are altogether not strongly conserved. All these residues were changed to amino acids of opposite hydrophobicity, size or charge, in order to maximise the differences (Figure 1B). The substitution matrix of Bordo and Argos (27) suggests that the changes we introduced should not be readily accommodated within the three-dimensional fold of the protein if these residues are indeed structurally important. Another residue, Ser46, which is mildly conserved, was mutagenized to proline with the intention to break or kink alpha helix 2, which was strongly predicted by several computer programs in most HMG-boxes (12) and in fact turned out to be one of the salient features of the B-domain of HMG1 (21, 22). We also trimmed the last 8 amino acids from HMG1bA, which are not strongly conserved, as an example of a modification which should be readily tolerated.

For technical reasons, these mutations were generated in two rounds, and the second round used as a starting material the shortened version of box A, which by that time had been already shown to be competent in DNA binding. Therefore, some mutant boxes are 89 amino acids long, while others are 81 amino acids long.

Several mutants cannot be produced in *E.coli*

Plasmid pT7-HMG1bA can direct the efficient synthesis of HMG1bA in *E.* coli BL21(DE3), and this peptide can be readily purified by ammonium sulphate precipitation and a single chromatographic step on Phenyl-Superose (11). The sequence of pT7-HMG1bA was therefore modified to direct the synthesis of the mutated forms of box A. Lysates of induced cells were analysed by SDS-gel electrophoresis: peptides HMG1bA-K82Z, -R10G, -F19A, -P32A, -S evel comparable to that of wild type boxA, while peptides HMG1bA-Y16S, -K28E, -F38S, -S42F and -F60S were present in extremely low quantities or not at all.

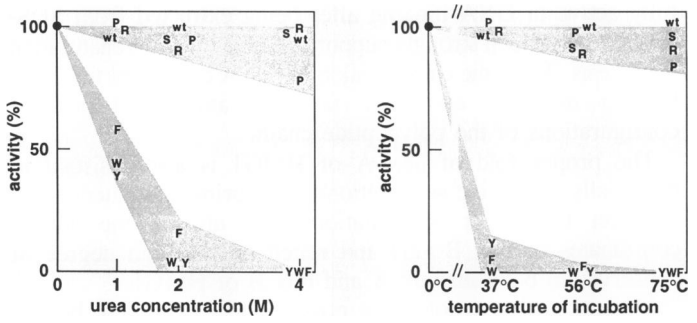

Figure 6. The last eight amino acids in HMG1bA are responsible for its anomalous hydrodynamic properties. About 5 μg of purified HMG1bA or HMG1bA-K82Z were applied to an FPLC Superdex-75 column (Pharmacia) equilibrated with buffer D-500 (20 mM HEPES, pH 7.9, 500 mM NaCl, 0.2 mM EDTA, 0.05% NP40, 10% glycerol and 0.5 mM DTT). The elution profiles of marker molecules of known native molecular weight (bovine serum albumin, soybean trypsin inhibitor, myoglobin, cytochrome c, aprotinin and vitamin B12) were determined by recording the absorbance at 280 nm in separate runs. Fractions of 0.4 ml were collected and analysed for total protein content (by SDS-PAGE) and DNA binding activity towards four-way junction DNA z.

Messenger RNA levels for the various constructs were similar, and short pulses of labelled methionine immediately followed by cell lysis showed that some protein was produced (data not shown). Therefore, we assume that all peptides were actually synthesised in *E.coli*, but some were misfolded and immediately degraded by proteases.

All the peptides that were produced efficiently could be purified to homogeneity following the same protocol used for wild type HMG1bA. Special care was used to ensure that all preparations were devoid of *E. coli* protein HU, which has biochemical properties extremely close to those of HMG boxes (28).

DNA binding properties of mutants

The purified mutant forms of HMG1bA were tested for binding to four-way junction DNA in a bandshift assay. Dissociation constants were calculated using a limiting amount of four-way junction DNA and titrating the polypeptides. The amounts of radioactivity in the free DNA and protein-DNA complex bands were measured, and the ratios fitted to the equation for bimolecular complexes. Mutants HMG1bA-K82Z, -F19A, -S46P, -W49R and -Y71G showed an affinity for four-way junction DNA z very close to the affinity displayed by HMG1bA (K_D close to 3×10^{-9} M), while the affinities of HMG1bA-P32A and -R10G were reduced about 5-fold and 10-fold, respectively (Figure 2 shows some examples of the results). The lack of effect of the substitution of arginine for triptophan at position 49 was unexpected, since tryptophans are present at this position in most HMG-boxes reported to date, with tyrosine as the only allowed alternative. In addition, the equivalent tryptophan in the B domain of HMG1 is located in a hydrophobic region at the vertex and is reputed to be critical for the fold of the HMG-box (21).

The structure selectivity of the mutants was also checked in a competition experiment. The binding to labelled four-way junction DNA was challenged with a 130-fold excess of unlabelled junction DNA, or a 130-fold excess of unlabelled control linear DNAs. As is the case for wild type HMG1bA, only unlabelled junction DNA is able to compete, while the addition of linear DNA is ineffective (Figure 3 shows the results for the most representative mutants). Therefore, it appears that structure specificity is retained also in the mutants with reduced affinity.

Resistance of the DNA binding activity of mutants to heating cycles and urea

The DNA binding activity of box A of HMG1 is remarkably resistant to chemical and physical stresses. Lyophilization, cycles of freezing and thawing, heating at 65°C for 3 hours do not affect it measurably (data not shown). HMG1bA retains its junction DNA binding activity even in high urea concentrations (Figure 4). In addition, box A renatures rapidly and completely when eluted from SDS-PAGE gels (data not shown).

The junction DNA binding activity of mutants HMG1bA-W49R, -F19A and -Y71G, however, is strongly reduced after incubation at 37°C for 30 minutes (Figure 5). In contrast, the 2 mutants with lower affinity for four-way junction DNA (HMG1bA-R10G and -P32A) and mutant HMG1bA-S46P retain their activity after heating up to 75°C, like HMG1bA.

The same pattern of results was obtained when urea was added in the binding buffer (Figure 5): HMG1bA, HMG1bA-K82Z, -R10G, -P32A and -S46P bound to the junction with high affinity in 4M urea, whereas the binding activity of HMG1bA-W49R, -F19A and -Y71G was strongly reduced by 1M urea and was almost completely abolished in 2M urea.

Thus, there appear to be two distinct classes of mutation effects: in one class the DNA binding capabilities of the polypeptides are normal in terms of binding affinity and specificity, but not of resistance to denaturing conditions; in the other class, mutant polypeptides bind to DNA less avidly, but the residual activity is not affected by heating cycles or urea. The mutant HMG1bA-S46P belongs to neither class and does not differ from wild type HMG1bA: the proline introduced in helix II does not disrupt the folding or the activity of the polypeptide in any appreciable way.

HMG1bA-K82Z is a monomer

We have previously shown that segments 1–89 and 91–176 of HMG1 have the apparent molecular weight of dimers in gel filtration chromatography, and that the subunits of the putative homodimers can be crosslinked with glutaraldehyde (11). However, a shorter form of the B-domain (residues 88–164) showed no sign of dimerization in its NMR spectrum, sedimented with the apparent molecular weight of the monomer (22), and could not be crosslinked with a lysine-specific bifunctional reagent (21). Likewise, the HMG-box domain of SRY behaves as a monomer in solution (20). These conflicting results might be attributed to the presence or absence of the amino acids C-terminal to the 'minimal' HMG-boxes of HMG1. We tested this possibility by comparing the hydrodynamic properties of HMG1bA and HMG1bA-K82Z. In fact, HMG1bA elutes from a Superdex-75 column with an apparent molecular weight of 20 kDa, against an expectation of 10.3 kDa for a globular monomer (Figure 6). In contrast, HMG1bA-K82Z elutes with an apparent weight of 13 kDa, against an expectation of 9.5 kDa. Although this elution is still slightly anomalous (perhaps because of the distinctly non-globular shape of the HMG-box domain), it is clearly not compatible with a dimeric structure for the HMG-box per se.

Full-length HMG1, HMG1bA and HMG1bA-K82Z form a single complex with four-way junctions at low protein concentration, and an additional complex at higher concentration (11, and data not shown). This behaviour is consistent with the

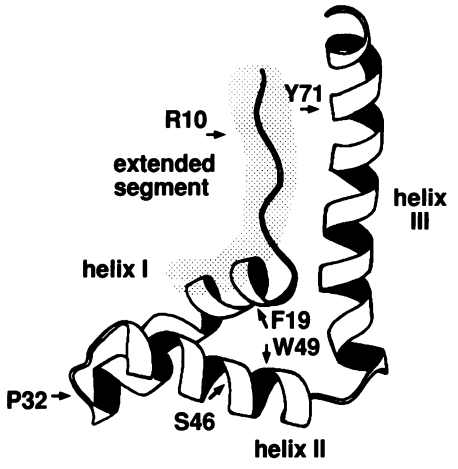

Figure 7. Three dimensional structure of the HMG-box B of rat HMG1 (21). The positions of the amino acids of box A we mutated in this study are indicated, on the assumption that the structure of box A can be superimposed on that of box B. We propose that the extended segment fits in the cleft of the minor groove, and that the first helix is posited to follow the minor groove without penetrating into it. The shaded area represents the proposed DNA binding surface.

successive, independent binding of two protein molecules to structurally quasi-equivalent sites on the junction. Single HMG-boxes are thus entirely competent for DNA binding.

DISCUSSION

We have analysed the effect of a range of replacement mutations on the biochemical properties of one HMG-box, boxA of the archetypal HMG-box protein HMG1. The mutated residues fell in three categories: those conserved in all classes of HMG boxes, those conserved within the HMG-box chromosomal proteins, and those not strictly conserved. The criterion used for substitution was to maximise the difference with respect to the conserved residue, so that the mutant should either not be able to fold or to bind DNA, if the chosen residue were involved in one of these functions.

The first result is that amino acids 82–89, which are mildly conserved in chromosomal HMG-box proteins, are not a part of the core HMG-box domain. Our 'minimal' HMG1 domain A behaves as a monomer in solution and in DNA binding, like the single HMG-box domain of SRY and LEF-1 proteins (19, 20) and like the HMG1 B domains whose structure has been solved by NMR (21, 22). The two HMG-boxes of HMG1 therefore appear to be structurally and functionally independent of each other in DNA binding. However, in vitro we have never observed complexes containing one HMG1 full-length molecule and two DNA molecules, as if the two HMG-boxes of HMG1 could not be occupied at the same time.

The second observation is that the DNA binding activity of the HMG-box is extremely resistant to regimes that usually cause protein denaturation. We could show that HMG1 boxA binds to four-way junction DNA in very high concentrations of urea, and displays DNA-binding activity immediately after being exposed to high temperatures. These data indicate that the HMG-box does not denature easily, or that it renatures rapidly and efficiently after being denatured, or both. We did not follow the folding status of the polypeptide directly, and so we cannot directly prove any of these possibilities. However, HMG1bA is fully active in DNA binding after being extracted from SDS-PAGE gels, which strongly supports at least the fast-renaturation hypothesis. Thus, the conformation of box A competent for DNA binding must be extremely favoured among all possible configurations of the polypeptide chain.

The proper fold of box A of HMG1 is also resistant to potentially disruptive substitutions in its primary sequence. To interpret the effect of mutations, we utilised the atomic coordinates of box B (21) and relied on the high degree of conservation between box A and box B of HMG1.

Mutants unable to fold have possibly been observed, but we cannot prove this point as the only observable phenotype of mutations of Tyr16, Lys28, Phe38, Ser42 and Phe60 is the inability to accumulate in *E. coli*. Amongst them, the case of Tyr16 is particularly interesting. The equivalent residue in box B, a phenylalanine, is involved in hydrophobic interactions that connect helix I to helix II; the absence of the aromatic ring in the mutant HMG1bA-Y16S may cause the two helices to pack incorrectly or not at all. In a similar manner, the role of Phe60 in the packing of the vertex of the domain was pointed out both by Weir *et al.* (21) and Read *et al.* (22), and the elimination of its aromatic ring in the mutation to serine may have caused a dramatic loss of folding ability. It is more difficult to account for the lack of expression of mutant HMG1bA-F38S, since residue Phe38 is not widely conserved and is located in a region between helices I and III where the number of amino acids may vary from box to box, and one might expect some degree of conformational tolerance.

A second group of mutants showed no alteration of the DNA binding capabilities in standard conditions, but enhanced sensitivity to heating and urea. These may be regarded as in vitro temperature-sensitive mutants; like traditional ts mutants, they identify residues or interactions which are not absolutely essential, but that nonetheless are important. Residue Tyr71 is engaged in hydrophobic contacts between the extended N-terminal region and helix III; it is not conserved in group 2 and 3 boxes, but its substitution with a very small residue as glycine likely reduces the opportunity for favourable van der Waals packing contacts. Mutations F19A and W49R are more surprising. Residues Phe17 and Trp45 (the corresponding residues in boxB, (21)) are part of the cluster of aromatic residues at the vertex of the box which maintains the angle between the 'wings' of the fold. Together with the third aromatic residue involved in the same set of interactions, Phe60, they are extremely well conserved; indeed, the tryptophan residue is the single most conserved residue among HMG-boxes, and the only alternative present in the boxes identified so far is tyrosine. However, not only are mutants F19A and W49R produced efficiently in *E. coli*, but they have a normal affinity and selectivity for four-way junction DNA if tested in standard conditions (0°C to room temperature). The broadly correct folding of both mutants may be rationalised in one of two ways: either the packing interactions within the hydrophobic interior of the vertex of the HMG-box are extensive and redundant, so that the substitution of one residue does not alter the structure considerably, or the side-chain interactions in this region are dynamic and able to accommodate large perturbations without transmitting gross variations to the rigid wings of the domain. We have examined models of the B box where an arginine has been substituted for the tryptophan: it appears that the charged amino group might project outside of the hydrophobic pocket into the solvent without gross steric clashes. However, the number of favourable interactions in the vertex interior

decreases, as does the number of atoms per unit volume. In our opinion, the correct packing density within the vertex interior might be preserved if the angle between the wings varies. This angle is somewhat different in the models of Weir *et al.* (21) and Read *et al.* (22) ($\approx 80°$ versus $\approx 70°$), and in reality might be somewhat variable or 'elastic', causing the vertex to work more like a hinge than like a rigid frame. The correspondence between a possible flexibility of the vertex of the HMG-box and the flexibility of the linear DNA which is the target of SRY and other group 3 HMG-box proteins is attractive, but its demonstration needs both the resolution of additional HMG-box structures, and more detailed investigations on HMG1 boxes A and B.

A third group of mutants can be produced efficiently in *E. coli* and have a low sensitivity to heating and urea. The properties of HMG1bA-S46P did not differ from those of the wild type, suggesting that the probable distortion introduced by proline into helix II is not critical. The same applies to helices I and III, since in a few HMG-boxes they contain prolines. Two other mutations have a modified affinity for DNA. Both of them affect amino acids on the surface of the protein, which can potentially interact with the ligand.

Mutation P32A modifies the strongly conserved proline which terminates helix I, but reduces the affinity for four-way junction DNA mildly. Proline therefore is not essential for the formation of the turn between helix I and II, or else its substitution should grossly affect the folding of the polypeptide. The reduction of the binding affinity for the ligand can be a direct effect if proline contacts the DNA, but more probably is an indirect effect due a possible alteration of the relative positioning of helices I and II.

Mutation R10G affects one of the several basic residues that are on or close to the concave surface formed by the wings of the HMG-box. This single point mutation reduces the affinity of boxA for junctions at least tenfold — a strong confirmation of the specificity of the interaction between the box and the DNA. Significantly, the selectivity of the HMG1bA-R10G mutant for junction over linear DNA does not appear to be impaired. We interpret this as an indication that the overall structure of the polypeptide still governs the interaction with structural features of the four-way junction DNA that differ from normal B-form DNA. The reduction of the affinity for the junction, on the other hand, may indicate that residue Arg10 is in close proximity to the DNA in the protein-DNA complex. Mutations of SRY and LEF-1 that reduce the affinity for DNA without obviously being involved in structural folding interactions also fall in the extended segment of the HMG-box (18, 19).

We have tried out on a computer graphics station several possible modes of interaction between box B of HMG1 and an example of strongly kinked DNA: the double helix containing a cisplatin 1,2-intrastrand crosslink. Cisplatin-modified DNA is a good binding target for the HMG-box proteins HMG1, SSRP-1 and Ixr1 (8, 29, 30), and the X-ray structure of one such molecule is available (31, 32). In the covalent adduct there is a bend towards the major groove of 32° and correspondingly the minor groove is expanded on the outside of the bend. The entire extended segment of the HMG box fits quite well into the enlarged minor groove of the platinated DNA, and the concave surface of the box follows the turn of the DNA helix for at least 120°. The quality of the fit depends to a very considerable degree from the strong kink and wide narrow groove of the cisplatin-modified DNA; straight B-form DNA segments retrieved from the Brookhaven structure library do not dock satisfactorily to the HMG-box. A wide narrow groove in the helices at the crossover point is also prominent in a computer-generated model of four-way junction DNA (33), and may underlie the affinity of HMG-boxes for four-way junctions. Our prediction of the mode of interaction of the HMG-box with DNA is based on limited information; however, if one uses the coordinates of box B as a guide to the predicted spatial positions of the residues in box A (Figure 7), the results of mutagenesis are compatible with this working model. The residue most critical for DNA binding, Arg10, is in the extended segment which is predicted to fit in the minor groove.

In conclusion, our results suggest that the peculiar conformation of the HMG-box is remarkably favoured and can accommodate several nonconservative substitutions; because of its peculiar shape, it cannot but discriminate strongly for the very features of kinked DNA which set it apart from normal B-form DNA.

ACKNOWLEDGEMENTS

We thank Andrea Pontiggia for considerable help during the course of this work, Jean Thomas for providing the coordinates of domain B of HMG1, Ermanna Rovida and Paolo Arosio for leading us through the joys and sorrows of computer graphics. We are grateful to Domenico Bordo, Annalisa Pastore, Martino Bolognesi and Simona Ferrari for useful discussions. The financial contributions of Telethon-Italy (grant A.07) to M.E.B., of the Cancer Research Campaign to D.M.J.L., and of the British Council/MURST programme for British-Italian collaborations are gratefully acknowledged.

REFERENCES

1. Johns, E. W. (1982) The HMG chromosomal proteins. Academic Press, London.
2. Bustin, M., Lehn, D. A. and Landsman, D. (1990) *Biochim. Biophys. Acta*, **1049**, 231–243.
3. Isackson, P. J., Fishback, J. L., Bidney, D. L. and Reeck, G. R. (1979) *J. Biol. Chem.*, **254**, 5569–5572.
4. Bonne, C., Sautiere, P., Duguet, M. and De Recondo, A.-M. (1982) *J. Biol. Chem.*, **257**, 2722–2725.
5. Hamada, H. and Bustin, M. (1985) *Biochemistry*, **24**, 1428–1433.
6. Bianchi, M. E., Beltrame, M. and Paonessa, G. (1989) *Science*, **243**, 1056–1059.
7. Bianchi, M. E. (1988) *EMBO J.*, **7**, 843–849.
8. Pil, P. M. and Lippard, S. J. (1992) *Science*, **256**, 234–236.
9. Reeck, G. R., Isackson, P. J. and Teller, D. C. (1982) *Nature*, **300**, 76–78.
10. Carballo, M., Puigdomenech, P. and Palau, J. (1983) *EMBO J.*, **2**, 1759–1764.
11. Bianchi, M. E., Falciola, L., Ferrari, S. and Lilley, D. M. J. (1992) *EMBO J.*, **11**, 1055–1063.
12. Bianchi, M. E., Beltrame, M. and Falciola, L. (1992) In Eckstein, F. and Lilley, D. M. J. (ed.), Nucleic Acids and Molecular Biology. Springer Verlag, Berlin, Vol. 6, pp. 112–128.
13. Jantzen, H.-M., Admon, A., Bell, S. P. and Tjian, R. (1990) *Nature*, **344**, 830–836.
14. 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.
15. Ner, S. S. (1992) *Curr. Biol.*, **2**, 208–210.
16. Lilley, D. M. J. (1992) *Nature*, **357**, 282–283.
17. Travis, A., Amsterdam, A., Belanger, C. and Grosschedl, R. (1991) *Genes Dev*, **5**, 880–894.
18. Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R. and Goodfellow, P. N. (1992) *Science*, **255**, 453–456.
19. Giese, K., Cox, J. and Grosschedl, R. (1992) *Cell*, **69**, 1–20.
20. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1992) *EMBO J.*, **11**, 4497–4506.

21. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D. and Thomas, J. O. (1993) *EMBO J.*, **12**, 1311–1319.
22. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C. and Norman, D. G. (1993) *Nucleic Acids Res.*, **21**, 3427–3436.
23. van de Wetering, M. and Clevers, H. (1992) *EMBO J.*, **11**, 3039–3044.
24. Horton, R. and Pease, L. R. (1991) In Mc Pherson, M. J. (ed.), Directed mutagenesis A practical approach. Oxford University Press.
25. Tabor, S. and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1074–1078.
26. Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1991) *Methods Enzymol.*, **185**, 60–89.
27. Bordo, D. and Argos, P. (1992) *J. Mol. Biol.*, **217**, 721–729.
28. Pontiggia, A., Negri, A., Beltrame, M. and Bianchi, M. E. (1993) *Mol. Microbiol.*, **7**, 343–350.
29. Bruhn, S. L., Pil, P. M., Essigman, J. M., Housmann, D. and Lippard, S. J. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 2307–2311.
30. Brown, S. J., Kellett, P. J. and Lippard, S. L. (1993) *Science*, 603–605.
31. Sherman, S. E., Gibson, D., Wang, A. H. J. and Lippard, S. J. (1985) *Science*, **230**, 412–417.
32. Sherman, S. E., Gibson, D., Wang, A. H. J. and Lippard, S. J. (1988) *J. Am. Chem. Soc.*, **110**, 7368–7381.
33. von Kitzing, E., Lilley, D. M. J. and Diekmann, S. (1990) *Nucleic Acids Res.*, **18**, 2671–2683.
34. Wagner, C. R., Hamana, K. and Elgin, S. C. R. (1992) *Mol. Cell. Biol.*, **12**, 1915–1923.
35. Grasser, K. D. and Feix, G. (1991) *Nucleic Acids Res.*, **19**, 2573–2577.
36. 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.
37. Kolodrubetz, D. and Burgum, A. (1989) *J. Biol. Chem.*, **265**, 3234–3239.
38. Diffley, J. F. X. and Stillman, B. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7864–7868.
39. Parisi, M. A. and Clayton, D. A. (1991) *Science*, **252**, 965–969.
40. Kelly, M., Burke, J., Smith, M., Klar, A. and Beach, D. (1988) *EMBO J.*, **7**, 1537–1547.