MOLECULAR AND CELLULAR BIOLOGY, Mar. 1994, p. 1979–1985
0270-7306/94/$04.00+0

Vol. 14, No. 3

Importance of a Flanking AT-Rich Region in Target Site Recognition by the GC Box-Binding Zinc Finger Protein MIG1

MARIA LUNDIN,† JAN OLOF NEHLIN, AND HANS RONNE*

Ludwig Institute for Cancer Research, Uppsala Branch, Uppsala Biomedical Center, S-751 24 Uppsala, Sweden

Received 11 August 1993/Returned for modification 17 September 1993/Accepted 1 December 1993

MIG1 is a zinc finger protein that mediates glucose repression in the yeast *Saccharomyces cerevisiae*. MIG1 is related to the mammalian Krox/Egr, Wilms' tumor, and Sp1 finger proteins. It has two fingers and binds to a GCGGGG motif that resembles the GC boxes recognized by these mammalian proteins. We have performed a complete saturation mutagenesis of a natural MIG1 site in order to elucidate its binding specificity. We found that only three mutations within the GC box retain the ability to bind MIG1: G_1 to C, C_2 to T, and G_5 to A. This result is consistent with current models for zinc finger-DNA binding, which assume that the sequence specificity is determined by base triplet recognition within the GC box. Surprisingly, we found that an AT-rich region 5′ to the GC box also is important for MIG1 binding. This AT box is present in all natural MIG1 sites, and it is protected by MIG1 in DNase I footprints. However, the AT box differs from the GC box in that no single base within it is essential for binding. Instead, the AT-rich nature of this sequence seems to be crucial. The fact that AT-rich sequences are known to increase DNA flexibility prompted us to test whether MIG1 bends DNA. We found that binding of MIG1 is associated with bending within the AT box. We conclude that DNA binding by a simple zinc finger protein such as MIG1 can involve both recognition of the GC box and flanking sequence preferences that may reflect local DNA bendability.

Zinc finger proteins of the C_2H_2 type are a family of DNA-binding proteins which are present in all eukaryotic cells (5). Zinc finger proteins that bind to GC boxes (also known as GSG boxes) are a subfamily of evolutionarily related proteins whose fingers are highly similar in sequence. These proteins include Sp1 (16, 17), the mammalian Krox/Egr and Wilms' tumor proteins (13, 14), the yeast MIG1 repressor (28), and CREA, the MIG1 homolog in *Aspergillus nidulans* (4). The number of fingers in these proteins varies from two in MIG1 and CREA to four in the Wilms' tumor protein.

On the basis of a comparison of Krox-20 and Sp1, which bind with high affinity to the nonamers GCGGGGGCG and GGGGCGGGG, a simple model in which each finger binds to a base triplet was proposed (24). According to this model, there are two types of fingers in these proteins that recognize either GGG or GCG, with the specificity being determined by residues 18 and 21. In particular, residue 18 is always a histidine in fingers that bind GGG and always a glutamic acid in fingers that bind GCG. The importance of residues 18 and 21 was confirmed by an altered DNA specificity of proteins with mutations in these two positions (24).

The three-dimensional structure of a complex between the three fingers of Zif268 (Krox-24) and its binding site has been determined (30). The structure confirms that each finger binds to a base triplet and shows that residues 21, 18, and 15 can form specific contacts with the first, second, and third base, respectively, in a triplet. The three contact residues are located on the surface of an α-helix that interacts with the major groove of the DNA. Interestingly, each finger in Zif268 forms only two of three possible base contacts. This suggests that the specificities of these fingers could be less stringent than first thought. Indeed, a mutagenesis of the second finger in Krox-20 revealed that this is the case (23). Thus, the two types of fingers in the Krox/Egr proteins were shown to recognize the degenerate triplets N(G/A)G and G(C/A/T)G. The mutagenesis confirmed the importance of residues 15, 17, 18, and 21 for DNA specificity.

The base triplet recognition principle has been extended to other zinc finger proteins. Thus, it was suggested that the yeast activator ADR1 (33) and the *Drosophila* transcription factor CF2 (10) recognize DNA in this way. These two proteins are less similar in sequence to the GC box-binding proteins, and they bind to motifs that are more AT rich, TTGG(G/A)G and GTATATATA. It has therefore been proposed that base triplet recognition is a universal principle for zinc finger-DNA binding and that the specificity of new proteins could be determined theoretically once this principle is fully understood (1, 15, 20).

We have previously identified several binding sites for the MIG1 repressor in the *SUC2*, *GAL1*, and *GAL4* genes (25, 28). A comparison of these sites revealed a common motif: (G/C)(C/T)GGGG. This is consistent with the fact that MIG1 has two zinc fingers and therefore is expected to recognize a 6-bp GC box. The MIG1 homolog CREA also binds to the same target sequence (21). However, the four MIG1 sites in the *SUC2* and *GAL* genes also share an AT-rich sequence 5′ to the GC box (25). This prompted us to investigate whether flanking sequences could be important for MIG1 binding.

We have now determined the specificity of MIG1 by saturation mutagenesis of a MIG1 site in the *SUC2* gene. As expected, we found that the GC box is essential: all but three mutations within this box abolish MIG1 binding. In contrast, no single mutation outside the GC box prevents binding. However, when several possible MIG1 sites in yeast promoters were tested, those that had more than one G or C within the AT box all failed to bind MIG1. A scan of the *SUC2* site with CC pairs confirmed that the presence of more than one C within the AT box interferes with MIG1 binding. We conclude

* Corresponding author. Mailing address: Ludwig Institute for Cancer Research, Uppsala Branch, Uppsala Biomedical Center, Box 595, S-751 24 Uppsala, Sweden. Phone: 46-18-174239. Fax: 46-18-506867. Electronic mail address: RONNE@BMC.UU.SE.

† Present address: Department of Morphology, Genetics and Aquatic Biology, Norwegian College of Veterinary Medicine, N-0033 Oslo, Norway.

that the specificity of MIG1 is determined in part by base triplet recognition in the GC box and in part by the presence of an adjacent AT box. We further found that MIG1 bends its target site within the AT box. This may explain the requirement for an AT-rich sequence, which would be more susceptible to protein-induced bending (34).

MATERIALS AND METHODS

Oligonucleotide gel shift assays. The oligonucleotides were synthesized, annealed, and labelled with [^{32}P]dCTP by using fill-in reaction mixtures with the Klenow enzyme, as previously described (26). Labelled DNA was purified on Bio-Spin 6 columns (Bio-Rad) prior to use. In vitro-translated MIG1 protein was prepared as previously described (28). The protein was incubated with ^{32}P-labelled oligonucleotides for 1 h at 20°C in 150 mM Tris-HCl buffer (pH 7.5) containing 30% glycerol, 450 mM KCl, 35 mM $MgCl_2$, 5 mM dithiothreitol, and 50 μM $ZnCl_2$. The reaction mixtures contained 1 μl of reticulocyte lysate, 500 to 1,000 pg (5,000 cpm) of labelled DNA, and 0.5 μg of poly(dI-dC) carrier DNA, in a total volume of 10 μl. Protein-DNA complexes were separated on nondenaturing 6% (19:1) polyacrylamide gels run at 20°C for 30 min at 40 V/cm in 10 mM Tris-borate buffer (pH 8.3) with 125 mM EDTA. The gels were dried and analyzed in a Molecular Dynamics 400S PhosphorImager.

Bent DNA assay. Bent DNA experiments were performed as described by Kim et al. (19). The following double-stranded oligonucleotide, JN82, containing the *SUC2* A MIG1 site, was synthesized:

5′-CTAGAGTAATAAAAATGCGGGGAATAATGG-3′
3′-TCATTATTTTTACGCCCCTTATTACCAGCT-5′

JN82 was cloned between the *Sal*I and *Xba*I sites of pBEND2 (19) to generate pJN127. This plasmid was cut with different restriction enzymes to generate circularly permutated 147-bp fragments containing the MIG1 site. The fragments were purified by low-melting-point agarose gel electrophoresis and labelled with [^{32}P]dCTP by filling in the ends with the Klenow enzyme. Labelled fragments were incubated with in vitro-made MIG1 protein as described above and then loaded on 12% (75:1) nondenaturing acrylamide gels. The gels were run for 2 h at 20°C and 150 V in 0.89 M Tris-borate buffer (pH 8.3) containing 4 mM ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA).

RESULTS

Deletion mapping of a MIG1-binding site. As a starting point for our analysis, we used the sequence of site A, one of two previously described MIG1-binding sites in the *SUC2* promoter (28). DNase I footprints have shown that 15 bases on the G-rich strand of this site (bases −8 to +7, counted from the first G in the GC box) are protected by MIG1. On the other strand, the protected region is slightly longer, extending another few bases in each direction (28). To determine the minimal sequence required for MIG1 binding, we tested progressively shorter double-stranded oligonucleotides in gel shift assays. We found that deletions beyond base −6 prevent MIG1 binding. Deletions from the other side retained the ability to bind MIG1 until base +6, the last G in the GC box, was removed (data not shown). The shortest oligonucleotide tested that could bind MIG1 well had the 15-bp sequence GGTAAAAATGCGGGG (the 5′ GG is not derived from *SUC2* but was added to allow fill-in labelling with [^{32}P]dCTP). This is approximately the same region that is protected by MIG1 in DNase I footprints with *SUC2* DNA. The requirement for the AT-rich region 5′ to the GC box could not be alleviated by including 3′ sequences instead. Thus, our results suggest that the GC box and the 5′ AT-rich region both are positively required for MIG1 binding, while sequences 3′ to the GC box are dispensable. However, the deletion mapping does not reveal what specific sequences within the minimal site are essential for binding.

Saturation mutagenesis of the MIG1-binding site. We proceeded with a saturation mutagenesis of the *SUC2* site, using a 17-bp oligonucleotide with the sequence GGTAAAAATG CGGGGAA. This includes the minimal binding site defined above and also the first two bases 3′ to the GC box. Double-stranded oligonucleotides containing all possible point mutations between bases −4 and +8 were synthesized and tested for MIG1 binding. The results are shown in Fig. 1. We found that some mutations outside the GC box have quantitative effects. For example, a T in position −4 enhances the binding, and a C in position −3 reduces binding. However, none of these mutations completely prevents MIG1 binding. We conclude that no single base outside the GC box motif is essential for binding. In contrast, most mutations within the GC box abolish MIG1 binding. Only three mutations retain the ability to bind MIG1: G_1 to C, C_2 to T, and G_5 to A. The first two mutations bind MIG1 at least as well as the wild type, while the third has a reduced binding (Fig. 1).

The saturation mutagenesis was subsequently extended to include also positions −5 and −6, in order to cover the entire AT box (Fig. 2). Again, we found that no single base outside the GC box is essential for MIG1 binding. However, some quantitive effects were observed. Thus, a T in position −5 or −6 enhances binding, as does a T in position −4, while a C in position −5 reduces binding (Fig. 2). We conclude that while some mutations outside the GC box have quantitative effects, none of them prevents MIG1 binding. Our finding that some flanking mutations enhance binding is consistent with the fact that naturally occurring MIG1 sites vary considerably in their binding affinities. There is therefore no reason to expect the wild-type *SUC2* A site to be strongest of all possible MIG1-binding sequences.

These results suggest that the consensus motif for MIG1 binding is a GC box with the redundant sequence (G/C)(C/T)GG(G/A)G. To test whether all variants of this motif can bind MIG1, we proceeded to assay all possible double and triple mutations for MIG1 binding (Fig. 3). Indeed, we found that all variants can bind MIG1, but they do so with different affinities. Thus, mutations that include the G_5-to-A substitution are less efficient in the gel shift assay. We conclude that MIG1 recognizes (G/C)(C/T)GG(G/A)G, with a G being preferred in position 5. This is consistent with the fact that all naturally occurring MIG1 sites found so far have a G in this position (see below).

MIG1-binding sites in yeast promoters. It is thought that GC box-binding proteins acquire their DNA specificity through base triplet recognition in the GC box. With a precise knowledge of the consensus GC box bound by MIG1, we therefore proceeded to look for MIG1 sites in yeast promoters. A number of possible sites were tested for binding, by using different methods. The sequences tested included several sites that have been proposed to bind MIG1, such as the URS-C element in *GAL1* (7), site 2 in the *GAL4* promoter (12), and sites in the *CIF1* and *FBP1* promoters (11, 22). Other sites were chosen from genes known to be repressed by glucose, such as the *GAL*, *MAL*, *MEL*, and *SUC* genes (9, 35). Also included were several possible MIG1 sites in the *MIG1* pro-

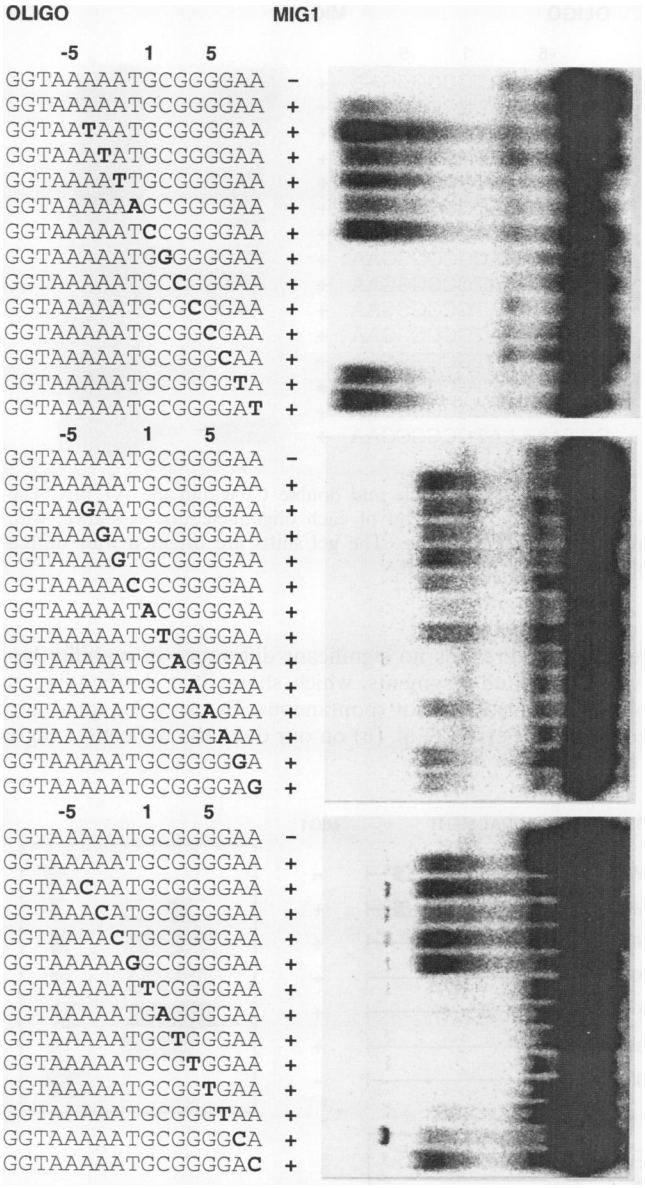

FIG. 1. Saturation mutagenesis of MIG1 site A in the *SUC2* promoter. The sequence of the top strand of each oligonucleotide (OLIGO) is shown, with mutated bases in boldface. Nucleotides are numbered starting with the first G in the GC box. The gel shifts are oriented with shifted oligonucleotides to the left.

moter, which is autoregulated (27), and finally a MIG1 site in the *PDC1* gene, which is induced by glucose (2, 18).

Surprisingly, we found that many of the tested sequences fail to bind MIG1, even though they have perfect MIG1-binding GC boxes. A sequence comparison revealed a highly consistent pattern. First, many of these sequences have an A in position 5. In contrast, all confirmed MIG1-binding sites have a G in position 5 (Fig. 4). Possibly, sites with an A in this position are less preferred in vivo because of a lower binding affinity. However, this does not explain the failure of several sites with a G in position 5 to bind MIG1. A likely explanation for this is, instead, the absence of a conserved AT box adjacent to these sites. Thus, sites that failed to bind MIG1 all have Gs or Cs in this region. In contrast, every confirmed MIG1-binding site has

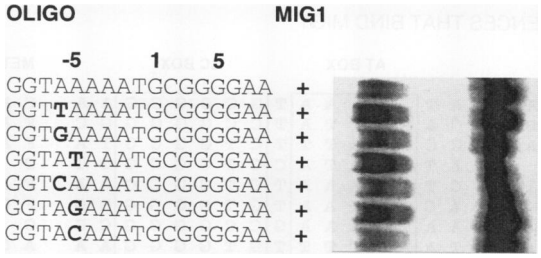

FIG. 2. Effects of mutations in positions −5 and −6. The sequence of the top strand of each oligonucleotide (OLIGO) is shown with mutated bases in boldface. The gel shifts are oriented with shifted oligonucleotides to the left.

a well-conserved AT box (Fig. 4). In particular, bases −3 to −5 are always A or T in MIG1-binding sites but are not conserved in any of the sites that failed to bind MIG1. It should be noted that while most of these sites have multiple Gs and Cs within the AT box, three of them have only a single G or C (Fig. 4). Since the corresponding mutant *SUC2* sites were able to bind MIG1 (Fig. 1), it suggests that other factors, such as the strength of binding to the GC box, may decide how many Gs or Cs can be tolerated within the AT box. In conclusion, these observations suggest that the AT box is important for MIG1 binding, even though no single base within it is strictly conserved.

The AT box is important for MIG1 binding. We proceeded to examine more thoroughly the effects of substitutions within the AT box of the *SUC2* site. Our finding that no single mutation in this region prevents MIG1 binding was confirmed by scanning the AT box with single-C substitutions (Fig. 5). However, it is clear that a C in positions −2 to −5 reduces binding, and in particular, position −3 is sensitive to this substitution. Since most of the promoter sequences that failed to bind MIG1 had several Gs or Cs within the AT box, we also tested the effect of inserting pairs of Cs (Fig. 5). We found that this has a much more pronounced effect on MIG1 binding than single Cs. Thus, all double-C substitutions between positions −1 and −6 either prevent or strongly reduce MIG1 binding. Only the last CC pair, in positions −6 and −7, was without effect. We conclude that while no single base in the AT box is essential, binding of MIG1 is prevented when the AT-rich nature of this sequence is disrupted by the presence of several Cs.

MIG1 bends its target site. AT-rich sequences are known to

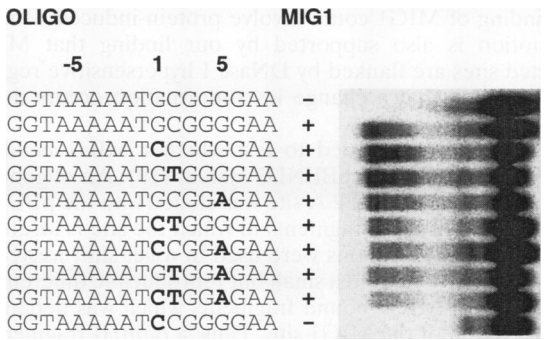

FIG. 3. Effects of double and triple mutations in the GC box. The sequence of the top strand of each oligonucleotide (OLIGO) is shown, with mutated bases in boldface. The gel shifts are oriented with shifted oligonucleotides to the left.

SEQUENCES THAT BIND MIG1

	AT BOX	GC BOX	METHOD
SUC2 A	A A T A A A A A T	G C G G G G A A	F A O
SUC2 B	G G A A A T T A T	C C G G G G G C	F A
GAL1 A	G C C T T A T T T	C T G G G G T A	F A
GAL3	T A T G A A T A C	C T G G G G T T	O
GAL4 1	G C T G A A A A T	C T G G G G A A	F A
MEL1	G A C A T T A A T	G T G G G G G G	F O
MAL2R	C A A T A A A A G	C C G G G G C G	O
MAL61	G T A A T T T T T	G T G G G G A A	A O
MAL62	T T C A A T T G T	G T G G G G T C	A O
MAL63	T T A T T A A A C	C C G G G G T A	A O
FBP1	T A A T A T A G T	G T G G G G A A	F A
PDC1	T A A A T A A A C	G C G G G G T G	F A
HAP4	G A T A T A A A A	C T G G G G T T	O
FPS1	T A G T T A A A A	G C G G G G A A	O

BASE	-5	1	5
A	1 8 6 7 7 9 9 10	2 0 0 0 0 0 5	7
T	6 2 4 5 7 5 5 2	8 0 8 0 0 0 0 6	2
C	1 2 3 0 0 0 0 0	3 7 6 0 0 0 0 1	2
G	6 1 1 2 0 0 0 2	1 7 0 14 14 14 14 2	3

SEQUENCES THAT DO NOT BIND MIG1

		METHOD
SUC2	G G C C T G G A C G T G G G G T C	O
GAL1 C	G G A A G G T T T G T G G G G C C	F O
GAL4 2	G G A C A C A A T C T G G A G A T	F O
FBP1	T G C G G A C A C C C G G A G T T	A
ADH1	A T G T A G G T G G C G G A G G G	A O
CIF1	A C C T T C C C A G C G G G G T G	O
MIG1	G A G T T T C T C G C G G G G C C	F A O
MIG1	T A C G C C A C T C T G G A G T T	F A O
MIG1	A G C A G A C A C G T G G A G A A	O
MIG1	G C G A G A A A A G T G G G G A A	O
MIG1	G A A T G A G T A C T G G G G T G	O
MIG1	G T G A G T A T A G T G G A G A C	O
MAL61	A G T C T A C A T G T G G A G A C	A O
MAL6T	T G T T T T C C C T G G A G T A	A O

FIG. 4. Sites in yeast promoters that were tested for MIG1 binding. Sites that were confirmed to bind MIG1 in vitro are shown at the top, and a table of base frequencies is shown underneath the sequences. Sites that failed to bind MIG1 are listed at the bottom. The GC boxes and AT boxes are enclosed, with Gs and Cs within the AT boxes being shown in boldface. The figure is a compilation of data from many different experiments. The methods used to assay MIG1 binding are listed to the right. Abbreviations: F, DNase I footprint; A, agarose gel shift; O, oligonucleotide gel shift.

facilitate both spontaneous bending and protein-induced bending of the DNA (34). The fact that MIG1 requires an AT-rich region adjacent to the GC box therefore raised the possibility that binding of MIG1 could involve protein-induced bending. This notion is also supported by our finding that MIG1-protected sites are flanked by DNase I hypersensitive regions, which suggests that a change in conformation occurs during MIG1 binding (25, 28).

We therefore proceeded to test whether MIG1 bends its target site, by using the pBEND2 system (19). An oligonucleotide containing the *SUC2* A site was cloned into the pBEND2 vector, and restriction fragments in which the site is located at different internal positions were used in a gel shift assay with MIG1 protein. We found a small but reproducible difference in the mobility of MIG1-bound fragments which was dependent on the position of the MIG1 site. Thus, a *Bam*HI fragment, in which the site is close to one end, migrates faster than a *Xma*I fragment, in which the site is close to the center (Fig. 6). This difference is consistent with protein-induced bending, which should cause fragments with centrally located sites to be more

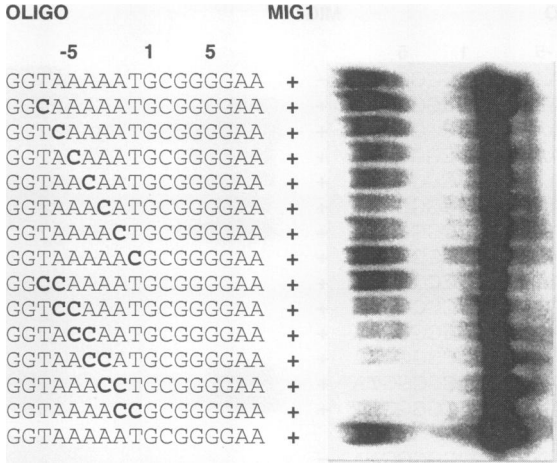

FIG. 5. Effects of single and double Cs within the AT box. The sequence of the top strand of each oligonucleotide is shown, with mutated bases in boldface. The gel shifts are oriented with shifted oligonucleotides to the left.

retarded. There was no significant difference in mobility between unshifted fragments, which shows that the bending is protein induced and not spontaneous. By using the quadratic equation of Ferrari et al. (6) on our data, the deflection angle

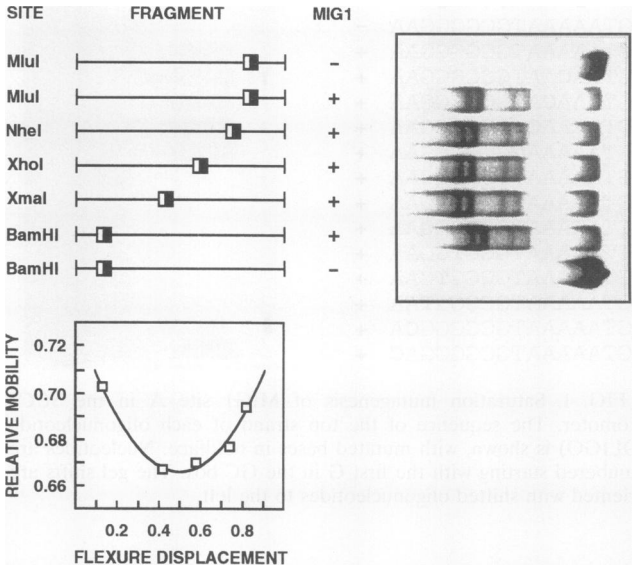

FIG. 6. Induced bending of the *SUC2* A site by MIG1. The restriction fragments tested are shown at the upper left, with GC boxes drawn as filled boxes and AT boxes drawn as open boxes. The gel shift is at the upper right, with shifted fragments oriented to the left. The gel system used (19) differs from that in previous figures, because of the much larger size of the DNA. The deflection angle and the locus of flexure were determined by plotting relative mobilities against flexure displacement, after which the least-squares method was used to fit a parabola to the datum points (6). The fitting equation was $y = 0.265x^2 - 0.266x + 0.733$ ($r^2 = 0.975$). The deflection angle α was 35° (bend angle θ = 145°) when calculated from either the linear or the quadratic coefficient, indicating a good fit of the model to the data (6). The locus of flexure was calculated by solving the first-order derivative of the fitting equation for $y' = 0$ to obtain the minimum of the parabola. This minimum is located at position −4 relative to the first G in the GC box.

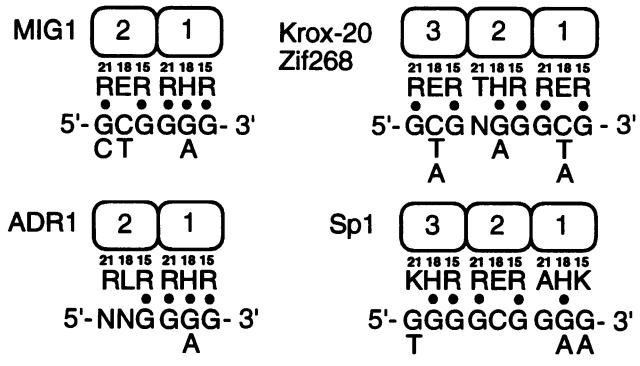

FIG. 7. Proposed interactions of zinc fingers with base triplets in GC boxes. Only the three contact residues 15, 18, and 21 in each finger are shown. The dots indicate possible base-specific contacts as inferred from the structure of the Zif268-DNA complex (29).

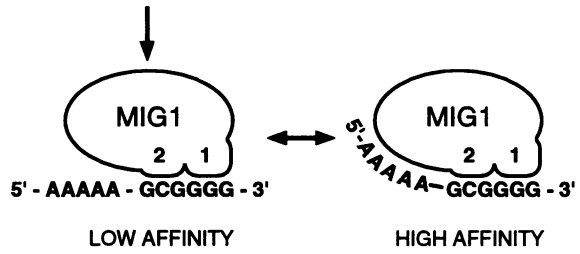

FIG. 8. Two-step model for target site recognition by MIG1.

was calculated to 35° (bend angle, 145°). Interestingly, the position of the bend, as calculated from the minimum of the fitting parabola, is located at base pair −4, the central position in the AT box. While this estimate has a possible error of a few base pairs, it suggests that the locus of bending is closely associated with the AT box.

Finally, it should be noted that the shifted DNA fragments appear as doublets (Fig. 6). The reason for this is not clear. Possibly, it reflects the binding of a second MIG1 molecule. We have previously suggested that two MIG1 molecules may bind cooperatively to adjacent sites (28). Similar interactions could also occur between a DNA-bound and a free MIG1 molecule, in which case the extra mass of the second molecule would cause a further small retardation of the DNA. However, other explanations, such as the existence of conformationally different complexes, are also conceivable.

DISCUSSION

Base triplet recognition in the GC box. Pavletich and Pabo (30) have shown that residues 21, 18, and 15 in each finger of Zif268 can form specific contacts with the first, second, and third base, respectively, in a triplet. Site-directed mutagenesis of finger 2 in Krox-20 showed that an arginine in position 21 recognizes a G as base 1, while a threonine in the same position is without specificity (23). Similarly, an arginine in position 15 recognizes a G as base 3, while a glutamine is without specificity. In the middle position 18, a histidine conferred specificity for either G or A as base 2, while a glutamic acid recognized C, A, or T, and a glutamine was without specificity (23).

Finger 1 of MIG1 has an arginine in position 21, a histidine in position 18, and an arginine in position 15. On the basis of the above principles, finger 1 should therefore recognize the triplets GGG and GAG. This is in excellent agreement with our results (Fig. 7). A mutated Krox-20 finger 2 with the same three contact residues as MIG1 recognized these two triplets, but also recognized TGG (23). MIG1, however, does not allow a T in position 4, which suggests that its finger 1 does not bind TGG (Fig. 2). Interestingly, finger 1 of ADR1 also has the same three contact residues as finger 1 of MIG1, and it, too, recognizes G(G/A)G, even though it differs considerably in sequence from MIG1 (33). This suggests that the recognition principle is correct and also that it has some general validity.

Finger 2 of MIG1 has an arginine in position 21, a glutamic acid in position 18, and an arginine in position 15. According to the Krox-20 results, it should therefore recognize the triplet G(C/A/T)G. In fact, the specificity of MIG1 for the middle base is restricted to C or T. Fingers 1 and 3 of the Krox/Egr proteins and finger 2 of Sp1 also have the same three contact residues (Fig. 7). Christy and Nathans (3) found that the mutant sequences GAGGGGGCG and GCGGGGGAG compete very poorly with a wild-type site for binding of Zif268 (Krox-24). This suggests that GAG is a much less preferred binding site also for fingers 1 and 3 in Zif268. Possibly, the reason MIG1 does not tolerate an A in position 2 is that it has only two fingers and therefore may require high-affinity binding to both triplets. The first base recognized by finger 2 of MIG1 is either G or C (some very weak binding was also seen with an A in this position [Fig. 1]). In contrast, the Krox-20 experiments predict a unique specificity for G. It is possible that the broader specificity of MIG1 in this position is due to an increased flexibility at the end of the GC box.

Role of the AT box in MIG1 binding. Our finding that the AT box is important for MIG1 binding was unexpected. A similar requirement for certain flanking sequences has not previously been observed for other zinc finger proteins. A possible explanation would be that base-specific contacts, similar to those involved in recognition of the GC box, are formed between MIG1 and the AT box. However, MIG1 has only two zinc fingers, and they are most likely used for recognition of the GC box (see above). Moreover, the fact that no single base in the AT box is essential for binding suggests that a certain DNA structure rather than a unique sequence is recognized. Our finding that MIG1 bends the DNA within the AT box (Fig. 6) further suggests that this unique structure could be bent DNA.

It is formally possible that the AT box acts indirectly to enhance binding of MIG1 to the GC box. For example, this binding might require bending or twisting of the DNA within the GC box, which could be facilitated by an adjacent AT-rich region. Alternatively, bending within the AT box could trigger a conformational change in MIG1 which enhances its affinity for the GC box. However, three facts argue against these possibilities. First, there is no major deformation of the DNA in the Zif268-DNA complex (30). Since the fingers of MIG1 are highly similar to those of Zif268, they should bind DNA in a similar way. Second, there is a positive requirement for the AT box: MIG1 does not bind to oligonucleotides that lack this region. If the AT box serves only to increase DNA flexibility, replacing it with a free end should also permit binding. Third, the AT box is always protected by MIG1 in footprints (25, 28). This suggests a direct interaction between MIG1 and the AT box.

We therefore consider it more likely that MIG1 recognizes the AT box directly, but in a way that requires bendable DNA rather than a unique sequence motif. Conceivably, target recognition by MIG1 could be a two-step process (Fig. 8). First, the zinc fingers would bind to the GC box, as predicted

by current models. Second, the interaction would be stabilized by bending of the DNA, which would allow further protein-DNA contacts. This could explain why an AT-rich region is necessary for high-affinity binding even though no single base within this region is conserved. The interaction could be similar to that of the *trp* repressor with its operator, which binds to the phosphate backbone of the DNA and recognizes the sequence only indirectly, through its effects on the overall geometry (27). There is a stretch of basic residues following the second finger in MIG1 (28). Since this finger is supposed to bind the 5′ half of the GC box, the adjacent basic residues would be well positioned to interact with the phosphate groups of the AT box.

It should be emphasized that the failure of MIG1 to bind sites without an AT box does not exclude a possible role for these sites in vivo, where cooperativity and interactions with other proteins may stabilize low-affinity binding. In fact, there is genetic evidence that two sites which lack AT boxes and therefore fail to bind MIG1, URS-C in *GAL1* and site 2 in *GAL4* (Fig. 4), are involved in glucose repression (7, 12). These two sites are both adjacent to confirmed strong MIG1 sites, which could facilitate cooperative binding (25). However, it has not yet been formally shown that repression at URS-C or *GAL4* site 2 is mediated by MIG1.

Targets of MIG1 regulation in vivo. Yeast promoters that contain confirmed MIG1 sites include the *SUC*, *GAL*, *MAL*, and *MEL* genes, which are involved in fermentation of alternative sugars (Fig. 4). These genes all respond to glucose repression, and a regulatory role for MIG1 has been confirmed for several of the genes (25, 28). It therefore seems that genes involved in sugar metabolism are major targets for MIG1 regulation in vivo. A different set of glucose-repressed genes are those that are required for growth on gluconeogenic carbon sources. They include the genes for the gluconeogenic enzymes FBP1 and PCK1, but they also include a number of genes that are required for respiration, the Krebs cycle, and other mitochondrial functions (9, 35). Many of the latter are dependent on the HAP4 activator for their expression (8). The mechanism by which these genes are repressed by glucose is less well understood. However, the sugar fermentation and gluconeogenic genes both require the SNF1 (CAT1) protein kinase for their expression, which suggests that a common mechanism is involved in regulating these two sets of genes (9, 35).

Interestingly, we found that both the *FBP1* and the *HAP4* promoters contain strong MIG1-binding sites (Fig. 4), indicating a possible role for MIG1 in repression of these genes. *HAP4* is a particularly interesting target, since it would mean that MIG1 functions at the top of a transcriptional cascade, similar to its role in control of the *GAL* genes (25). Despite these findings, a disruption of *MIG1* did not cause significant derepression of either the *FBP1* or the *HAP4* mRNAs on glucose (data not shown). Consistent with this, mutations in *MIG1* (*CAT4*) are epistatic over *snf1* (*cat1*) mutations for growth on alternative sugars but not for growth on gluconeogenic carbon sources (27, 32). A likely explanation is that repression of gluconeogenic and respiratory genes involves at least two parallel mechanisms downstream of the SNF1 kinase. One mechanism would act through MIG1, while the other mechanism remains to be identified. It should be noted that there is a MIG1-independent component also in glucose repression of the sugar fermentation genes (28). Thus, the difference between the two sets of genes could be quantitative rather than qualitative.

Implications for other zinc finger proteins. An important question is whether our findings apply also to other zinc finger proteins. Our finding that MIG1 bends its target site is not unprecedented, since other zinc finger proteins are known to bend DNA (31). However, it has not been shown whether such induced bending is a mere consequence of binding or whether it is required for high-affinity binding. Our finding that the AT box is important for MIG1 binding suggests that the latter is the case and that flanking sequences can contribute to target site recognition by a zinc finger protein. Clearly, future attempts to predict target sites on the basis of finger sequences alone will be frustrated if this is a common phenomenon.

It is possible that the role of flanking DNA is a unique development among fungal zinc finger proteins that have only two fingers and therefore may require additional DNA contacts for high-affinity binding. The mammalian GC box-binding proteins frequently bind to very GC-rich sequences (GC islands) and are therefore unlikely to require flanking AT-rich DNA (3, 17). However, other preferences for certain flanking sequences cannot be excluded, and they could play a more prominent role than so far realized. The specificities of many zinc finger proteins are only incompletely known. In most cases, they were deduced by comparing the sequences of known binding sites. We have for the first time performed a complete saturation mutagenesis of a zinc finger protein binding site, including its flanking DNA. Our results reveal that flanking sequences can play a very subtle role, with a preference for a certain base composition rather than for a unique sequence. Clearly, such preferences would escape detection in sequence comparisons and also in most site-directed mutagenesis studies. It was recently suggested that the apparent lack of specificity for some contact residues predicts a poor triplet discrimination for many known zinc fingers (23). Flanking sequence preferences could play an important role in such cases. Conceivably, the relative importance of base-specific contacts and local DNA conformation in binding could vary for different zinc finger proteins, as seems to be the case with the HMG box proteins (6).

ACKNOWLEDGMENTS

We thank Christer Wernstedt and Aino Ruusala for making the large number of oligonucleotides required for this study and Odd Gabrielsen for providing us with the pBEND2 plasmid.

The first two authors made equal contributions to this work.

REFERENCES

1. **Berg, J. M.** 1992. Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites. Proc. Natl. Acad. Sci. USA **89:**11109–11110.
2. **Butler, G., and D. McConnell.** 1989. Identification of an upstream activation site in the pyruvate decarboxylase structural gene. Curr. Genet. **14:**405–412.
3. **Christy, B., and D. Nathans.** 1989. DNA binding site of the growth-factor-inducible protein Zif268. Proc. Natl. Acad. Sci. USA **86:**8737–8741.
4. **Dowzer, C. E. A., and J. M. Kelly.** 1991. Analysis of the *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. Mol. Cell. Biol. **11:**5701–5709.
5. **El-Baradi, T., and T. Pieler.** 1991. Zinc finger proteins: what we know and what we would like to know. Mech. Dev. **35:**155–169.
6. **Ferrari, S., V. R. Harley, A. Pontiggia, P. N. Goodfellow, R. Lovell-Badge, and M. E. Bianchi.** 1992. SRY, like HMG1, recognizes sharp angles in DNA. EMBO J. **11:**4497–4506.
7. **Flick, J. S., and M. Johnston.** 1992. Analysis of URS_G-mediated glucose repression of the *GAL1* promoter of *Saccharomyces cerevisiae*. Genetics **130:**295–304.
8. **Forsburg, S. L., and L. Guarente.** 1989. Identification and characterization of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. Genes Dev. **3:**1166–1178.

9. **Gancedo, J. M.** 1992. Carbon catabolite repression in yeast. Eur. J. Biochem. **206**:297–313.
10. **Gogos, J. A., T. Hsu, J. Bolton, and F. C. Kafatos.** 1992. Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger protein. Science **257**:1951–1955.
11. **Gonzalez, M. I., R. Stucka, M. A. Blazquez, H. Feldmann, and C. Gancedo.** 1992. Molecular cloning of *CIF1*, a yeast gene necessary for growth on glucose. Yeast **8**:183–192.
12. **Griggs, D. W., and M. Johnston.** 1991. Regulated expression of the *GAL4* activator gene in yeast provides a sensitive genetic switch for glucose repression. Proc. Natl. Acad. Sci. USA **88**:8597–8601.
13. **Haber, D. A., and A. J. Buckler.** 1992. *WT1*: a novel tumor suppressor gene inactivated in Wilms' tumor. New Biol. **4**:97–106.
14. **Herschman, H. R.** 1991. Primary response genes induced by growth factors and tumor promoters. Annu. Rev. Biochem. **60**: 281–319.
15. **Jacobs, G. H.** 1992. Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J. **11**:4507–4517.
16. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51**:1079–1090.
17. **Kadonaga, J. T., K. A. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends Biochem. Sci. **11**:20–23.
18. **Kellermann, E., and C. P. Hollenberg.** 1988. The glucose- and ethanol-dependent regulation of *PDC1* from *Saccharomyces cerevisiae* are controlled by two distinct promoter regions. Curr. Genet. **14**:337–344.
19. **Kim, J., C. Zwieb, C. Wu, and S. Adhya.** 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene **85**:15–23.
20. **Klevit, R. E.** 1991. Recognition of DNA by Cys_2, His_2 zinc fingers. Science **253**:1367.
21. **Kulmberg, P., M. Mathieu, C. Dowzer, J. Kelly, and B. Felenbok.** 1993. Specific binding sites in the *alcR* and *alcA* promoters of the ethanol regulon for the CREA repressor mediating carbon catabolite repression in *Aspergillus nidulans*. Mol. Microbiol. **7**:847–857.
22. **Mercado, J. J., O. Vincent, and J. M. Gancedo.** 1991. Regions in the promoter of the yeast *FBP1* gene implicated in catabolite repression may bind the product of the regulatory gene *MIG1*. FEBS Lett. **291**:97–100.
23. **Nardelli, J., T. Gibson, and P. Charnay.** 1992. Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic Acids Res. **20**:4137–4144.
24. **Nardelli, J., T. J. Gibson, C. Vesque, and P. Charnay.** 1991. Base sequence discrimination by zinc-finger DNA-binding domains. Nature (London) **349**:175–178.
25. **Nehlin, J. O., M. Carlberg, and H. Ronne.** 1991. Control of yeast *GAL* genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J. **10**:3373–3377.
26. **Nehlin, J. O., M. Carlberg, and H. Ronne.** 1992. Yeast *SKO1* gene encodes a bZIP protein that binds to the CRE motif and acts as a repressor of transcription. Nucleic Acids Res. **20**:5271–5278.
27. **Nehlin, J. O., M. Carlberg, and H. Ronne.** Unpublished data.
28. **Nehlin, J. O., and H. Ronne.** 1990. Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. EMBO J. **9**:2891–2898.
29. **Otwinowski, Z., R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler.** 1988. Crystal structure of *trp* repressor/operator complex at atomic resolution. Nature (London) **335**:321–329.
30. **Pavletich, N., and C. O. Pabo.** 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science **252**:809–817.
31. **Schroth, G. P., G. R. Cook, E. M. Bradbury, and J. M. Gottesfeld.** 1989. Transcription factor IIIA induced bending of the *Xenopus* somatic 5S gene promoter. Nature (London) **340**:487–488.
32. **Schüller, H.-J., and K. D. Entian.** 1991. Extragenic suppressors of yeast glucose derepression mutants leading to constitutive synthesis of several glucose-repressible enzymes. J. Bacteriol. **173**:2045–2052.
33. **Thukral, S. H., M. L. Morrison, and E. T. Young.** 1992. Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transcription. Mol. Cell. Biol. **12**:2784–2792.
34. **Travers, A. A.** 1989. DNA conformation and protein binding. Annu. Rev. Biochem. **58**:427–452.
35. **Trumbly, R. J.** 1992. Glucose repression in the yeast *Saccharomyces cerevisiae*. Mol. Microbiol. **6**:15–21.