The Yeast Homeodomain Protein MATα2 Shows Extended DNA binding Specificity in Complex with Mcm1*

Hualin Zhong‡ and Andrew K. Vershon§

From the Waksman Institute and the ‡Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08855-0759

The MATα2 (α2) repressor interacts with the Mcm1 protein to turn off α-cell type-specific genes in the yeast Saccharomyces cerevisiae. We compared five natural α2-Mcm1 sites with an α2-Mcm1 symmetric consensus site (AMSC) for their relative strength of repression and found that the AMSC functions slightly better than any of the natural sites. To further investigate the DNA binding specificity of α2 in complex with Mcm1, symmetric substitutions at each position in the α2 half-sites of AMSC were constructed and assayed for their effect on repression in vivo and DNA binding affinity in vitro. As expected, substitutions at positions in which there are base-specific contacts decrease the level of repression. Interestingly, substitutions at other positions, in which there are no apparent base-specific contacts made by the protein in the α2-DNA co-crystal structure, also significantly decrease repression. As an alternative method to examining the DNA binding specificity of α2, we performed in vitro α2 binding site selection experiments in the presence and absence of Mcm1. In the presence of Mcm1, the consensus sequences obtained were extended and more closely related to the natural α2 sites than the consensus sequence obtained in the absence of Mcm1. These results demonstrate that in the presence of Mcm1 the sequence specificity of α2 is extended to these positions.

Homeodomain proteins are a family of transcription factors involved in many developmental and cellular processes and have been found in almost every eukaryotic organism (1–4). The natural target sites for many homeodomain proteins are unknown; therefore, their DNA-binding sites have been determined by NMR and x-ray crystallography studies (42, 45, 46). Although there is only 27% sequence identity between the α2 homeodomain and the homeodomain of Drosophila engrailed, the overall structures are very similar (46, 47). Moreover, these proteins bind DNA in a similar manner, and most of the conserved residues in the third helices of the homeodomains make identical contacts with DNA (46, 47). The α2 protein therefore provides a good model for studies of homeodomain protein-DNA interactions.

In this paper, we have examined in detail the sequence requirements for the α2 homeodomain protein in repression of asg in vivo and in vitro. Our results indicate that, in complex with Mcm1, the sequence specificity of DNA binding by α2 is apparently more extended than on its own. These results suggest one explanation for DNA recognition sites determined in vitro.

*This work was supported by National Institutes of Health Grant GM49265 (to A. K. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Recipient of the New Jersey Cell and Molecular Biology Fellowship.

§ To whom reprint requests should be addressed: Waksman Institute, P.O. Box 759, Piscataway, NJ 08855-0759. Tel.: 908-445-2905; Fax: 908-445-5735; E-mail: Vershon@mbcl.rutgers.edu.

1 The abbreviations used are: asg, α-cell type-specific gene(s); EMSA, electrophoretic mobility shift assay; AMSC, α2-Mcm1 symmetric consensus site; PCR, polymerase chain reaction; bp, base pair(s).
Results

The \( a2\)-Mcm1 DNA-binding Sites from Five Natural \textit{ags} Function as Repressor Sites in Either Orientation—The \( a2\)-Mcm1 binding sites have been identified in the promoter regions of five \textit{ags} (\textit{STE6}, \textit{BAR1}, \textit{STE2}, \textit{MFA1}, and \textit{MFA2}) (21–24). Although the natural sites are highly conserved, there are variations at some positions that may result in different levels of \( a2\)-Mcm1-mediated repression. To verify that these \( a2\)-Mcm1 sites are all functional repressor sites and to measure their relative strength for repression in the same promoter context, oligonucleotides containing the sites were inserted into the promoter region of a \textit{CYC1-lacZ} reporter plasmid, and the level of expression from the promoter was assayed by measuring \( \beta\)-galactosidase activity (Fig. 1). The results indicate that oligonucleotides in the shifted band that is detectable at the lowest protein concentration was extracted from the dried gel slice (49). The isolated DNA were amplified by PCR using primer W341 and end-labeled primer W340 and then purified for the next round of selection. After six rounds, the purified DNA fragments were cloned into a T-overhang vector (50). The inserted selected sites were sequenced, and their DNA binding affinity was determined by EMSA and quantitated on a PhosphorImager. The \( a2\) binding site selection in the presence of Mcm1 was performed by utilizing the same initial randomized oligonucleotide pool. In each round of selection, a titration of \( a2\) initiated at 1 \( \mu\)M and 0.5 \( \mu\)M of Mcm1 were present in the EMSA reactions. After six rounds of selection, the selected sites were cloned and sequenced, and DNA binding affinity was measured as described above.

Results

The \( a2\)-Mcm1 DNA-binding Sites from Five Natural \textit{ags} Function as Repressor Sites in Either Orientation—The \( a2\)-Mcm1 binding sites have been identified in the promoter regions of five \textit{ags} (\textit{STE6}, \textit{BAR1}, \textit{STE2}, \textit{MFA1}, and \textit{MFA2}) (21–24). Although the natural sites are highly conserved, there are variations at some positions that may result in different levels of \( a2\)-Mcm1-mediated repression. To verify that these \( a2\)-Mcm1 sites are all functional repressor sites and to measure their relative strength for repression in the same promoter context, oligonucleotides containing the sites were inserted into the promoter region of a \textit{CYC1-lacZ} reporter plasmid, and the level of expression from the promoter was assayed by measuring \( \beta\)-galactosidase activity (Fig. 1). The results indicate that although there are small variations in the level of repression between the different sites, all of the sites confer greater than 30-fold repression of \textit{lacZ} expression.

The \textit{STE6} operator was shown to function as a repressor site in either orientation. There was, however, a small difference in the level of repression between the two directions (23). Since all five natural \( a2\)-Mcm1 sites are only partially symmetric dyads, we were interested in whether the orientation of the sites with respect to the start site of transcription affects the level of repression. To address this question, we compared the level of repression conferred by these sites in both orientations (Fig. 1). Although all five sites function in either orientation, there were slight differences in the levels of repression. These small differences may be due to the asymmetry of the sites or to the flanking sequences.

An \( a2\)-Mcm1 Symmetric Consensus Site Confers Higher Repression Activity than the Natural \( a2\)-Mcm1 Sites—Although there is some asymmetry in the natural \( a2\)-Mcm1 sites, a
The co-crystal structure of the AMSC site, which is identical to the CYC1-lacZ site, is consistent with our repression data and shows that the protein makes base-specific contacts, and this suggests that the AMSC site functions as a better site for repression than the known natural operator sites by EMSA. Our results indicated that substitutions at positions T3, G4, and T5 result in a large decrease in repression of 10-, 26-, and 40-fold, respectively (Fig. 3B). 

To correlate the repression data with DNA binding activity, we assayed a2-Mcm1 DNA binding affinity for the AMSC and STE6 promoters. Substitutions at positions in which there are no base-specific contacts in the co-crystal structure, also dramatically reduce repression (Fig. 3B). For example, some substitutions at positions C1, A2, and A3, and most notably A6, reduce the level of repression by 10-50-fold. These results show that there is additional sequence specificity at these positions for a2-Mcm1-mediated repression in vivo.

To investigate if there are sequence requirements at these positions for a2 DNA binding and repression, we assayed a series of AMSC operators with symmetric base pair substitutions in both a2 half-sites were cloned into the CYC1-lacZ reporter promoter, and their effects on repression in vivo were measured using $\beta$-galactosidase assays (Fig. 2B). As expected, substitutions at positions T3, G4, and T5 result in a large decrease in repression of lacZ expression (approximately 100-fold). Substitutions at positions in which there are base-specific contacts in the minor groove, T3, T4, and T6, also significantly reduce the level of repression. Surprisingly, substitutions at positions in which there are no base-specific contacts in the co-crystal structure, also dramatically reduce repression (Fig. 2B). For example, some substitutions at positions C1, A2, and A3, and most notably A6, reduce the level of repression by 10-50-fold. These results show that there is additional sequence specificity at these positions for a2-Mcm1-mediated repression in vivo.

To determine whether the increase in repression conferred by the AMSC site is due to stronger DNA binding affinity of the a2-Mcm1 complex to the site, we compared the DNA binding affinity of a2-Mcm1 for AMSC and STE6 operators in EMSA. Our results indicated that a2-Mcm1 binds to the AMSC site slightly better (1.3-fold) than to the STE6 operator (Fig. 3). This result is consistent with our in vivo repression data and shows that the AMSC site functions as a better site for a2-Mcm1 binding and repression than the known natural aag operators.

**Saturation Mutagenesis of the a2 Recognition Sequence in AMSC**—The co-crystal structure of the a2 homeodomain bound to DNA shows that the protein makes base-specific contacts with positions T3, G4, and T5 in the major groove and with positions T3 (or A3) and T4 in the minor groove (Fig. 2A). These positions are conserved in the a2 binding sites in each of the known natural aag. However, a comparison of the natural a2-Mcm1 sites also indicates that other positions such as positions 1, 2, 6, and 7, in which there are no apparent base-specific contacts in the co-crystal structure, are also highly conserved. This strong conservation among the natural sites suggests that there is a sequence specificity at these positions. To investigate if there are sequence requirements at these positions for a2 DNA binding and repression, a series of AMSC operators with symmetric base pair substitutions in both a2 half-sites were cloned into the CYC1-lacZ reporter promoter, and their effects on repression in vivo were measured using $\beta$-galactosidase assays (Fig. 2B). As expected, substitutions at positions T3, G4, and T5 result in a large decrease in repression of lacZ expression.
Mutations | Operator Sequences | Units |
--- | --- | --- |
AMSC | CATGTAAT | 1.8 ± 0.8 | 100X |
G4A/C27T | A | 587 ± 45 | 0.3X |
G4A | A | 138 ± 10 | 1X |
A7/T24A | T | 7.3 ± 1.1 | 25X |
T24A | A | 3.6 ± 0.1 | 50X |
A7C/T24G | C | 35 ± 5.0 | 5X |
A7C | C | 4.1 ± 0.6 | 44X |
T8A/A23T | A | 9.3 ± 1.0 | 19X |
T8A | A | 2.7 ± 0.3 | 67X |
T6C/A23G | C | 36 ± 3.0 | 5X |
T6C | C | 7.6 ± 1.0 | 24X |
T9A/A23T | A | 14 ± 4.0 | 13X |
A23T | T | 4.5 ± 0.3 | 39X |
T9G/A23C | G | 95 ± 7.0 | 2X |
A22C | C | 6.1 ± 0.6 | 30X |

Fig. 4. Comparison of repression conferred by operators with single substitution in one α2 half-site of AMSC with operators with a pair of symmetric substitutions in both α2 half-sites at the corresponding positions. β-Galactosidase activity and the repression ratios were calculated as described in the legend to Fig. 2B.

Changes at positions 1, 2, 3, 6, 7, and 8 in the α2 homeodomain recognition sequence of the natural STE6 operator have only small effects on repression (36). One explanation for this discrepancy is that we have constructed operators with symmetric substitutions in both α2 half-sites of the AMSC instead of a single point mutation in one α2 half-site of the natural STE6 operator. To investigate this difference, we compared the effect on repression of single point substitutions in one α2 half-site in AMSC with symmetric mutations in both α2 half-sites (Fig. 4).

An asymmetric substitution at a position in which there is a base-specific contact, such as G4A, leads to slightly higher repression than the symmetric substitution G4A/C27T. Asymmetric substitutions at other positions, such as A7, T24, and A7C, have less effect on repression than symmetric substitutions in both half-sites. These results agree with the previous study and show that although the single mutants only have a small effect on repression, there is a sequence preference at these positions in the context of symmetric substitutions.

The α2 Homeodomain DNA-Binding Site Selection in Vitro—The natural DNA-binding sites of many homeodomain proteins are unknown. One commonly used method to determine their target sites is through in vitro DNA-binding site selection experiments utilizing randomized oligonucleotides. One important question is how well sites selected through in vitro selections correlate with the natural in vivo target sites. The mutagenesis experiments described above have precisely defined the sequence requirements for α2 binding. We therefore decided to use the site selection technique to determine whether it would identify a similar site. An oligonucleotide pool that contains an Mcm1 binding site adjacent to a randomized region was used in the site selection assay. After six rounds of selection, the sites were cloned and sequenced (Table I). An alignment of the sequences obtained in the selection arrived at a consensus site of TGT, which corresponds perfectly with the positions contacted by α2 in the major groove in the co-crystal structure (46). We have assayed the α2 DNA binding affinity to each site and have found that sites with high DNA binding affinity are closely related to the natural α2 recognition sequences (Fig. 1). If only those sites with moderate affinity (+) or better were considered, we obtained a consensus sequence of TGTA, which closely matches the natural α2 sites. These results indicate that this in vitro technique can identify sites that correspond well with the natural ones.

Although α2 binds DNA on its own in vitro, it must interact with Mcm1 to repress asg in vivo. Previous studies have shown that the cooperative DNA binding by α2 and Mcm1 requires a specific spacing and orientation between their respective DNA-binding sites (34, 35, 44). We therefore analyzed whether the sites that were obtained from the in vitro selection experiments were able to be bound cooperatively by α2 and Mcm1 (Table I). Our results show that only sites that have the proper spacing, orientation, and sequence between the α2 and Mcm1 recognition sites, such as sequences 1, 2, and 9, are bound cooperatively by α2 and Mcm1. On the other hand, those sites that do not have these sequence requirements are not bound cooperatively by α2 and Mcm1, although on their own both proteins bind to these sites with relatively high affinity. For example, sequences 3, 5, and 7 are not bound cooperatively by α2 and Mcm1 because they do not have the same orientation between the α2 and Mcm1 binding sites. Furthermore, sequences 6 and 8, which have the proper orientation and spacing between the α2 and Mcm1 binding sites, are not bound cooperatively by α2 and Mcm1 because a G or C is present at positions that are important for the α2-Mcm1 complex binding in vivo (Fig. 2B).

These results indicate that the site selection assay is also able to screen for additional sequence requirements such as spacing and orientation between sites when a protein binds DNA in a complex with other cofactors.

The in vitro selection experiment, using α2 alone, defined a consensus sequence TGT that corresponds well to the α2 recognition core sequence in natural sites. This consensus site, however, does not extend to some positions that are conserved in the natural α2-Mcm1 sites and that we have shown are important for repression in vivo. In the presence of Mcm1, it appears that there are base-specific preferences at these positions. We therefore performed the α2 site selection experiment a second time with the same pool of random oligonucleotides in the presence of Mcm1. After six rounds of selection, we obtained 50 sequences, which we aligned in two groups according to the different spacing (5 or 6 bp) between the α2 recognition core site, TGT, and the core Mcm1 binding site, CCTAAT (Table I). In each group, the sequences are listed based on the observed binding affinity of the α2-Mcm1 complex. Most of the selected α2 sites have the appropriate orientation and spacing between the α2 and Mcm1 sites. In this selection, one-third of the sites selected were sequence 1, which is not only the highest affinity site selected from the pool, but also exactly matches the α2 half-site that was used in the AMSC site (Figs. 1 and 2). Sequences that were selected from the pool (sequences 1, 4, 17, and 19) are identical to some of the natural α2 half-sites shown in Fig. 1. We obtained a consensus sequence of ATGTAAAT for sites with 5-bp spacing between the α2 TGT core site and the Mcm1 site. This sequence perfectly matches the α2 half-site sequence in the AMSC site that we derived from an alignment of the natural asg operators. A slightly different consensus site, GTGTAAAT (D represents A, G, or T) was obtained from selected sites with 6-bp spacing between the α2 TGT core site and the Mcm1 site. We have assayed one derivative of this consensus (CGTGTAAAT) for its ability to repress transcription of the CYC1-lacZ promoter in vivo and have found that the operator containing this sequence in each α2 half-site strongly represses the lacZ expression (45-fold repression).

One notable difference between the selected sites with 5-bp spacing or 6-bp spacing between the α2 TGT sequence and the Mcm1 site is the base preferences at position 2. Selected sites with 5-bp spacing predominantly have an A at this position (23 of 34), while only 6 of 34 have a G at this position. On the other hand, selected sites with 6-bp spacing predominantly contain a G (13 of 16) at this position. These results suggest that there may be a difference in the base pair specificity at this position that depends on the spacing between the α2 and Mcm1 sites. The data in Fig. 2 show that in operators with 5-bp spacing an A at position 2 represses lacZ expression 2-fold better than a

Extended DNA Binding Specificity of MATα2 in Complex with Mcm1

8405
Extended DNA Binding Specificity of MATα2 in Complex with Mcm1

Table I
The α2 binding site selection without Mcm1

| Selected sites | α2 binding affinity | α2-Mcm1 cooperativity | Spacing | Orientation |
|---------------|---------------------|------------------------|---------|-------------|
| Wild type     | CATGIAATTa-M         | ++                     | 5       | wt          |
| 1             | TATGIAATTa-M         | ++                     | 5       | wt          |
| 2             | cGTGIAATTta-M        | ++                     | 5       | wt          |
| 3             | M-taCTGTATCag        | ++                     | 3       | rev         |
| 4             | CCATGIAATTa-M        | +                      | 4       | wt          |
| 5             | M-taGTGAAATCA       | +                      | 2       | rev         |
| 6             | cGTGAAATGta-M        | +                      | 6       | wt          |
| 7             | M-taTCCGTGAAgaa    | +                      | 5       | rev         |
| 8             | TTTGCTCTLa-M         | +                      | 5       | wt          |
| 9             | cATGCTTtTta-M        | +                      | 6       | wt          |
| 10            | GGAGTAASta-M         | +                      |         |             |
| 11            | M-taAAAATATGgaat     | +                      |         |             |
| 12            | GGGGGGTAATa-M        | +                      |         |             |
| 13            | TCTCCGATA-M          | +                      |         |             |
| 14            | CAGGCTGATA-M         | +                      |         |             |
| 15            | ACCGCTGTa-M          | +                      |         |             |
| 16            | CATAGAATa-M          | +                      |         |             |
| 17            | cGTGCTATGa-M         | +                      |         |             |
| 18            | M-taCTAAAgTgaat      | +                      |         |             |

* Capitalized letters correspond to the positions of the eight randomized nucleotides (except in the case of the wild type sequence), and M represents the core Mcm1 binding site.

The α2 binding affinity is relative to that of the wild type sequence. ++ + +, >75% of wild type site DNA binding affinity; ++ ±, 40–75%; +, 20–40%; +, <20%.

The α2-Mcm1 cooperativity was determined by comparing the α2 DNA binding affinity in the presence of Mcm1 with that in the absence of Mcm1. ++, greater than 6-fold; +, between 1-fold and 6-fold; −, no cooperativity.

The spacing refers to the distance between TGT and the core Mcm1 site.

wt refers to the orientation of the TGT and Mcm1 sites found in the natural α2-Mcm1 site; rev refers to the opposite orientation.

The five natural α2-Mcm1 binding sites that have been identified in the promoter regions of asg are highly conserved. We have examined the relative strength of these sites by comparing the level of repression mediated by these sites in the same promoter context. All of these sites confer strong repression of lacZ expression from a heterologous CYC1-lacZ promoter, although there are some differences in the relative strength of repression, with MFA1 > BAR1 > STE6 > MFA2 > STE2 (Fig. 1). The strength of repression mediated by these sites correlates with the degree of similarity to a consensus α2-Mcm1 binding site; i.e., the higher the sequence similarity to the consensus site, the stronger the repression. To further test this correlation, a symmetric consensus site (AMSC) was assayed in the same context and was found to confer better repression than any of the natural sites. The AMSC site was also bound cooperatively by the α2 and Mcm1 proteins with slightly higher affinity than the STE6 operator. The level of repression is therefore, at least in part, a function of the strength of α2-Mcm1 binding, and the higher the binding affinity, the greater the repression. Although the natural α2-Mcm1 operators are not optimal binding and repressor sites, it may not be biologically necessary for these sites to function as well as the AMSC site. For example, the transcriptional activator elements in the asg promoters may be significantly weaker than the CYC1 UAS elements of the reporter promoter used in this study. These weaker promoters would not require a repressor site as strong as the AMSC site to completely turn off expression of the genes. Alternatively, the weaker natural repressor sites may enable the cells to respond faster to switches in mating type and hence the cells would quickly derepress asg and be able to mate with MATα cells.

The α2 half-sites in AMSC are identical to one of the α2 half-sites used in determining the co-crystal structure (46). In the co-crystal complex, residues Ser-50, Asn-51, and Arg-54 in the α2 homeodomain make base-specific contacts in the major groove with T5, G4, and T6, and Arg-7 in the N-terminal arm of the homeodomain makes base-specific contacts in the minor groove with T8 and T9. As expected, our mutagenesis results show that mutations in T5, G4, T6, T8, and T9 dramatically reduce the level of α2-Mcm1-mediated repression in vivo. However, we also observed that substitutions at other positions, such as C1, A2, A6 and A7, in which there are no base-specific contacts in the α2 co-crystal structure, also significantly affect repression (Fig. 2B). These results suggest that specific base pairs are also required at these positions.

Recently, a ternary crystal structure of the α1 and α2 proteins bound to DNA has been solved (42). This structure was determined at a higher resolution than the previous α2 co-crystal structure, and portions of the α2 protein, most notably the N-terminal arm and the C-terminal tail extending from the homeodomain, are more ordered in the ternary complex. The α2 half-site in the ternary complex is identical to the α2 half-sites in the AMSC consensus sequence. In the ternary structure, besides base-specific contacts at positions 3, 4, 5, 8, and 9 that are present in the co-crystal structure, there are also additional base-specific contacts at positions 2, 4, 5, and 6. It is possible that in complex with Mcm1, α2 may make similar contacts to these positions, which would explain why substitutions of these bases pairs have an effect on α2-Mcm1-mediated repression. For example, although there is no apparent base-specific contact to position 2 in the co-crystal structure, it has been shown in the structure of the α1-α2-DNA ternary complex that N-7 of A2 is contacted via a water-mediated hydrogen bond by Ser-50 of the α2 homeodomain (42). This position is strongly conserved among the α2-Mcm1 binding sites found upstream of asg, and of the 10 natural α2 half-sites, 8 contain an A and 2
contain a G at this position (Fig. 1). The observation that G, unlike C and T, functions almost as well as an A at this position is consistent with a model that in complex with Mcm1, α2 makes a similar base-specific contact to the N-7 group as is observed in the a1-α2-DNA ternary complex.

We have found that substitutions to T or A at positions A7, T8, and T9 have less effect on repression than substitutions to G or C (Figs. 2B and 4). It has been observed that A:T and T:A base pairs have a similar distribution of hydrogen bond donors and acceptors in the minor groove (51). Since in both crystal structures positions 8 and 9 are contacted in the minor groove by Arg-7, it is possible that this extended side chain is able to adjust to accommodate the slight alteration of the positions of the hydrogen bond acceptors when an A:T base pair is substituted for T:A at these positions. This model is supported by the observation that α2 binds on its own with almost equal affinity to sites with either T:A or A:T at these positions. However, substitutions from T to A at position 8 or 9 cause more than a 5-fold reduction in the level of α2-Mcm1-mediated repression (Fig. 2B). A portion of the effects of these substitutions may be due to the slight decrease in α2 DNA binding affinity. However, substitutions at these positions also affect Mcm1 binding to the site (52), and this decrease in affinity may account for most of the decrease that we observed in α2-Mcm1-mediated repression. Although no contacts were observed at position 7 in either structure, there is also an A or T preference at this base pair. It is possible that there may be base-specific contacts at this position in the α2-Mcm1-DNA complex. Alternatively, G or C substitutions at this position may interfere with the minor groove contacts at adjacent positions.

In the a1-α2-DNA ternary complex, there are only base-specific contacts in the minor groove at position 6; therefore, we might expect the A to T substitution at this position would not greatly affect repression and DNA binding affinity. However, we observed that the T substitution at this position reduces the level of repression over 30-fold. If the Arg-4 side chain makes similar contacts in the α2-Mcm1-DNA complex as observed in the a1-α2-DNA ternary complex, the position of the side chain may be fixed by its contacts with base pairs 4 and 5 (42). Therefore, unlike Arg-7, the Arg-4 side chain may not be able to alter its position to accommodate the small changes for making a hydrogen bond with the T substitution at position 6. In addition, the Gly-5 peptide backbone amide makes a hydrogen bond contact to the O-2 of thymine on the bottom strand at position 6. To maintain the hydrogen bond, the position of the peptide backbone would have to be slightly altered in the A6 to T substitution. The repositioning of the backbone may in turn weaken or destroy multiple base-specific or sugar-phosphate backbone contacts that are made by other side chains in the N-terminal arm and therefore significantly reduce the level of repression. Alternatively, the substitution may sterically interfere with the precise position of the arm for making contacts with DNA. It has been shown that a small hydrophobic region proceeding the N-terminal arm of the α2 homeodomain is required for cooperative DNA binding and protein-protein interactions with Mcm1 (44). It is possible that the interactions between the proteins fix the position of residues in the N-terminal arm so that additional contacts could be made in the minor groove that are not observed in either crystal structure. If these additional contacts are made, then that may partially contribute to the increase in α2 DNA binding specificity that is observed in the presence of Mcm1. In summary, the high degree of sequence conservation at positions 1, 2, and 6 among the natural sites along with our mutational analysis at these positions shows that they play an important role in α2 DNA recognition. Our results are consistent with a model that, in combination with Mcm1, α2 is making contacts with the DNA that

---

**TABLE II**

| Site | Selected sites<sup>b</sup> | $K_d$ | Frequency |
|------|---------------------------|-------|-----------|
| 1    | CATGTAAttta-M             | $1.9 \times 10^{-9}$ | 17        |
| 2    | GATGIAAttta-M             | $2.2 \times 10^{-9}$ | 1         |
| 3    | GTTGIAAttta-M             | $7.6 \times 10^{-9}$ | 2         |
| 4    | TGGTIAAttta-M             | $8.5 \times 10^{-9}$ | 2         |
| 5    | TATTGIAAttta-M            | $1.3 \times 10^{-8}$ | 1         |
| 6    | GTGTGIAAttta-M            | $3.8 \times 10^{-8}$ | 1         |
| 7    | TTATGIAAttta-M            | $6.9 \times 10^{-8}$ | 1         |
| 8    | CAGGIAAttta-M             | $1.5 \times 10^{-7}$ | 1         |
| 9    | AGGAAATTta-M              | $2.0 \times 10^{-7}$ | 1         |
| 10   | ACAGGIAAttta-M            | $3.5 \times 10^{-7}$ | 1         |
| 11   | ATAGGIAAttta-M            | $3.6 \times 10^{-7}$ | 1         |
| 12   | GGAATGTTta-M              | $4.4 \times 10^{-7}$ | 1         |
| 13   | TAAATCTta-M               | $1.2 \times 10^{-6}$ | 1         |
| 14   | AGCATAAttta-M             | $4.2 \times 10^{-6}$ | 1         |
| 15   | CATTTCTca-M               | $8.3 \times 10^{-6}$ | 1         |
| 16   | GATGGGTga-M               | $3.5 \times 10^{-6}$ | 1         |

<sup>a</sup> M represents the core Mcm1 binding site.

<sup>b</sup> D represents A, T, or G.
are similar to contacts observed in the a1-o2-DNA ternary complex.

We have analyzed the DNA binding specificity of o2 in complex with Mcm1 by determining the effects of mutations within the AMSC site on repression. As an alternative approach to investigate the o2 DNA binding specificity, we have performed in vitro site selection experiments for the o2 homeodomain in the presence and absence of Mcm1. In the absence of Mcm1, we obtained an o2 binding consensus site of TGT, corresponding to positions in which there are base-specific contacts by the o2 homeodomain in the major groove (46). In the presence of Mcm1, we obtained two consensus sequences, ATGTAAT and GTGTAADT (D represents A, G, or T) according to the spacing between the o2 and Mcm1 sites. These consensus sequences show extended sequence specificity compared with the consensus sequence obtained from the site selection of o2 on its own. Furthermore, most sequences obtained from the second selection have the same orientation and spacing for the o2 and Mcm1 binding sites as is found in the natural o2-Mcm1 operators (Fig. 1, Table II). Among these selected sequences, four different sequences are identical to the natural o2 half-site shown in Fig. 1. Our results demonstrate that in vitro DNA site selection technique can be utilized not only to identify binding sites of individual proteins but also to further screen for optimal binding sites for a protein complex.

Previous studies have shown that the relative positions between the o2 and Mcm1 binding sites is somewhat flexible, and while large changes in spacing are not functional, operators with 5 or 6 base pairs between the sites are bound cooperatively by the proteins and function as repressor elements in vivo (35, 44). The flexibility of the spacing between the o2 and Mcm1 sites is evident among the natural operators, since the STE2, STE6, MFA1, and MFA2 sequences have a 5-bp space between the o2 and Mcm1 sites in one half-site and 6-bp spacing in the other half-site (Fig. 1). The fact that in the presence of Mcm1 sites were selected from the random pool which have both 5- and 6-bp spacing further shows that binding by the o2-Mcm1 complex can accommodate either spacing. In contrast, the spacing requirements between the o2 and a1 binding sites of haploid-specific operators, as well as the positions of the binding sites in other homeodomain complexes such as the Drosophila Paired homodimer and the Hox-Phx heterodimer are rigidly fixed (43, 53, 54). These results suggest that either the protein-protein or protein-DNA interactions in the o2-Mcm1-DNA complex can adjust, to some extent, to accommodate the alterations in spacing between the binding sites.

Interestingly, in comparing the consensus sequences with 5- or 6-bp spacing between the o2 and Mcm1 sites, we noticed that there is a different preference for the base pair corresponding to position 2 in the AMSC site. In sites with 5-bp spacing an A is preferred (23 of 34), while sites with 6-bp spacing predominantly have a G at this position (13 of 16). We have determined that in sites with 5-bp spacing an A at position 2 results in 2-fold higher repression than a G, while in sites with 6-bp spacing, G functions as well as A. These results suggest that sites with 6-bp spacing have relaxed sequence specificity at this position in comparison with sites with 5-bp spacing. It is possible that to make the proper contacts with Mcm1 on operators with 6-bp spacing, o2 may have to alter the contacts with position 2 of the operator. In the a1-o2-DNA ternary complex structure, this base pair is contacted by Ser-50 of the o2 homeodomain via a water-mediated hydrogen bond to N-7 (42). The preference for purines at this position in operators with either 5- or 6-base pair spacing suggests that the contact to N-7 is made in both sets of operators. However, the fact that A is preferred to G in selected sites with 5-bp spacing indicates that there may be another base-specific contact to the A:T base pair at this position. In contrast, in operators with 6-bp spacing G functions as well as A, which suggests that this contact is not made in this set of operators. In other homeodomains, residue 50 makes either a direct or water-mediated hydrogen bond with the base pair corresponding to position 2, and this residue has been shown to have an important role in determining homeodomain DNA binding specificity (9, 47, 53, 56).

In summary, the in vitro site selection results support the conclusions drawn from our mutagenesis data that, in complex with Mcm1, the DNA binding specificity of the o2 protein extends to positions in which there are no apparent base-specific contacts in the co-crystal structure. Similar changes in the binding specificity of homeodomain proteins in the presence of their cofactors have also been observed in other homeodomain proteins. For example, the optimal binding site for a Hox protein in complex with Pbx1 appears to be slightly different from the site selected for Hox binding on its own (54). Likewise, the DNA binding specificity of Oct-1 appears to change upon interaction with Bob1 (13). The fact that the o2 binding sites selected in the presence of Mcm1 are extended and better defined than in the absence of Mcm1 could explain why consensus sites for some DNA-binding proteins identified in the absence of their cofactors may not function well in vivo. Site selection in the absence of the cofactor would therefore not be able to define the sequence requirements for binding by a protein complex, such as the orientation and spacing between the binding sites of each protein, as well as the sequence specificity from additional contacts made by the proteins.

Acknowledgments—We thank T. Li and C. Wolberger for providing the coordinates of the a1-o2-DNA ternary structure. We are grateful to S. Parent, C. Abate-Shen, S. Zhang, H. Tang, and members of our lab for helpful discussion and comments on the manuscript. We thank J. Mead for pJM120, which contains the AMSC site. We also thank T. B. Acton for pTBA23 and the purified Mcm1(1–96) fragment.

REFERENCES
1. Scott, M. P., Tamkun, J. W., and Hartzell, G. W. (1989) Biochim. Biophys. Acta 989, 25–48
2. Wuthrich, K., and Gehring, W. (1992) Transcriptional Regulation (Mcknight, S. L., and Yamamoto, K. R., eds) pp. 535–577, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
3. Kornberg, T. B. (1993) J. Biol. Chem. 268, 28613–28616
4. Gehring, W., Affolter, M., and Burglin, T. (1994) Annu. Rev. Biochem. 63, 487–526
5. Ekker, D. C., Young, K. E., von Kessler, D. P., and Beachy, P. A. (1991) EMBO J. 10, 1179–1186
6. Ekker, S. C., Kessler, D. P., and Beachy, P. A. (1992) EMBO J. 11, 4059–4072
7. Florence, B., Handrow, R., and Laughon, A. (1991) Mol. Cell. Biol. 11, 3613–3623
8. Catron, K. M., Iler, N., and Abate, C. (1993) Mol. Cell. Biol. 13, 2354–2365
9. Hanes, S. D., Riddihough, G., Ish-Horowicz, D., and Brent, R. (1994) Mol. Cell. Biol. 14, 3364–3375
10. Hoey, T., and Levine, M. (1988) Nature 332, 858–861
11. Nasim, S., Gross, C. T., Kanzoria, M. A., and McGinnis, W. (1992) EMBO J. 11, 991–1002
12. Czerny, T., and Busslinger, M. (1995) Mol. Cell. Biol. 15, 2858–2871
13. Gustaifer, M., Georgiev, O., van Leeuwen, H., van der Vliet, P., and Schaffner, W. (1996) EMBO J. 15, 2781–2789
14. Chang, C. P., Brocchieri, L., Shen, W. F., Largman, C., and Cleary, M. L. (1996) Mol. Cell. Biol. 16, 1734–1743
15. Tresser, M. N., He, X., and Rosenfeld, M. G. (1991) Nature 350, 577–584
16. Johnson, A. (1992) Transcriptional Regulation (Mcknight, S. L., and Yamamoto, K. R., eds) pp. 975–1006, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Xue, D., Tu, Y., and Chalifie, M. (1993) Science 261, 1324–1328
18. Chan, S., Jaffe, L., Capovilla, M., Botas, J., and Mann, R. S. (1994) Cell 78, 603–615
19. Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. L., and Beachy, P. A. (1994) EMBO J. 13, 3551–3560
20. MacKay, V. L., and Manney, T. R. (1974) Genetics 76, 272–288
21. Strathern, J., Hicks, J., and Herskowitz, I. (1981) J. Mol. Biol. 147, 357–372
22. Wilson, K. L., and Herskowitz, I. (1984.) Mol. Cell. Biol. 4, 2420–2427
23. Johnson, A. D., and Herskowitz, I. (1985) Cell 42, 237–247
24. Miller, A. M., MacKay, V. L., and Nasmyth, K. A. (1986) Nature 321, 598–603
25. Goutte, C., and Johnson, A. D. (1988) Cell 52, 875–882
26. Goutte, C., and Johnson, A. D. (1993) J. Mol. Biol. 233, 359–371
27. Keleher, C. A., Goutte, C., and Johnson, A. D. (1988) Cell 53, 927–936
28. Keleher, C. A., Passmore, S., and Johnson, A. D. (1989) Mol. Cell. Biol. 9,
5228–5230
29. Passmore, S., Elble, R., and Tye, B.-K. (1989) *Genes & Dev.* 3, 921–935
30. Wilson, K. L., and Herskowitz, I. (1986) *Proc. Natl. Acad. Sci. U. S. A.* 83, 2536–2540
31. Sprague, G. P., and Herskowitz, I. (1988) *Annu. Rev. Microbiol.* 37, 623–660
32. Jenness, D. D., Burkholder, A. C., and Hartwell, L. H. (1988) *Cell* 55, 521–529
33. Brake, A. J., Brenner, C., Najarian, R., Laybourn, P., and Merryweather, J. (1985) *Transport and Secretion of Proteins*, pp. 103–108, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
34. Sauer, R. T., Smith, D. L., and Johnson, A. D. (1988) *Genes & Dev.* 2, 807–816
35. Smith, D. L., and Johnson, A. D. (1992) *Cell* 68, 133–142
36. Smith, D. L., and Johnson, A. D. (1994) *EMBO J.* 13, 2378–2387
37. Hall, M. N., and Johnson, A. D. (1987) *Science* 237, 1007–1012
38. Kelch, C. A., Redd, M. J., Schultz, J., Carlson, M., and Johnson, A. D. (1992) *Cell* 68, 709–719
39. Komachi, K., Redd, M. J., and Johnson, A. D. (1994) *Genes & Dev.* 8, 2857–2867
40. Vershon, A. K., and Johnson, A. D. (1993) *Cell* 72, 105–112
41. Mak, A., and Johnson, A. D. (1993) *Genes & Dev.* 7, 1862–1870
42. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science* 270, 262–269
43. Jin, Y., Mead, J., Li, T., Wolberger, C., and Vershon, A. K. (1995) *Science* 270, 290–293
44. Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996) *Mol. Cell. Biol.* 16, 2135–2143
45. Phillips, C., Vershon, A. K., Johnson, A. D., and Dahlquist, F. (1991) *Genes & Dev.* 5, 764–772
46. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell* 67, 517–528
47. Kissinger, C. R., Liu, B., Martin-Blance, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell* 63, 579–590
48. Siliciano, P. G., and Tatchell, K. (1984) *Cell* 37, 969–978
49. Blackwell, T. K., and Weintraub, H. (1990) *Science* 250, 1104–1110
50. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (eds.) (1995) *Current Protocols in Molecular Biology*, pp. 15.7.1–15.7.3., Vol. 2, John Wiley & Sons, Inc., New York
51. Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. U. S. A.* 73, 804–808
52. Acton, T. B., Zhong, H., Vershon, A. K. (1997) *Mol. Cell. Biol.* 17, in press
53. Treisman, J., Genez, P., Vashishtha, M., Harris, E., and Desplan, C. (1989) *Cell* 59, 553–562
54. Knoepfler, P. S., Lu, Q., and Kamps, M. P. (1996) *Nucleic Acids Res.* 24, 2288–2284
55. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell* 82, 709–719
56. Klemm, J. D., Roule, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell* 77, 21–32

*Extended DNA Binding Specificity of MATα2 in Complex with Mcm1*