Altering the DNA-binding Specificity of the Yeast Matα2 Homeodomain Protein*

Received for publication, April 9, 2001, and in revised form, June 21, 2001
Published, JBC Papers in Press, July 3, 2001, DOI 10.1074/jbc.M103097200

Jonathan R. Mathias‡, Hualin Zhong‡‡, Yisheng Jin†, and Andrew K. Vershon¶
From the Waksman Institute and the Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854-8020

Homeodomain proteins are a highly conserved class of DNA-binding proteins that are found in virtually every eukaryotic organism. The conserved mechanism that these proteins use to bind DNA suggests that there may be at least a partial DNA recognition code for this class of proteins. To test this idea, we have investigated the sequence-specific requirements for DNA binding and repression by the yeast α2 homeodomain protein in association with its cofactors, Mcm1 and Mata1. We have determined the contribution for each residue in the α2 homeodomain that contacts the DNA in the co-crystal structures of the protein. We have also engineered mutants in the α2 homeodomain to alter the DNA-binding specificity of the protein. Although we were unable to change the specificity of α2 by making substitutions at residues 47, 54, and 55, we were able to alter the DNA-binding specificity by making substitutions at residue 50 in the homeodomain. Since other homeodomain proteins show similar changes in specificity with substitutions at residue 50, this suggests that there is at least a partial DNA recognition code at this position.

Homeodomain (HD) proteins are a large family of transcriptional regulatory proteins that control many diverse cellular and developmental processes (1). HD proteins have been found in organisms ranging from fungi to plants and humans, and the DNA-binding domains of these proteins are strongly conserved. The structures of a large number of HDs have been determined alone and in complex with DNA, and each shows remarkable structural similarity to one another (2–9). The HD is usually 60 residues long and consists of three α-helices, which form a tight bundle (see Fig. 1). The third helix in the bundle, frequently termed the “recognition helix,” lies in the major groove of the DNA and makes the majority of the base-specific and sugar phosphate backbone contacts. In most of the HD structures, there is also a flexible region, called the N-terminal arm, that extends from the N terminus of the first helix and wraps around the DNA to make base-specific and phosphate backbone contacts in the minor groove. The similarity in the mechanism of DNA binding among HD proteins is due in part to highly conserved residues, such as Trp48, Arg53, Lys55, and Lys57, that are conserved in virtually all HD proteins and together make a set of phosphate and backbone contacts with both strands of the DNA (see Fig. 1). These contacts are likely to be important to help position the recognition helix within the major groove. Asn51 is absolutely conserved in all HD proteins and makes virtually identical base-specific contacts with a conserved adenine in the binding site of every HD structure that has been determined.

Given the conserved nature of the HD structure and DNA contacts among the more than 15 HDs that have been solved, along with the high sequence conservation in this family of proteins, it seems likely that most other HDs will fold and bind DNA in a similar manner. However, although HD proteins use a conserved mechanism to bind DNA, different HD proteins have very different binding specificities in vivo and in vitro. These differences in specificity are due to residues in helix 3 and the N-terminal arm that are not as well conserved among the different HD proteins. For example, residue 50 in the recognition helix makes a base-specific contact in the major groove, but, in contrast to Asn51, is relatively diversified among the different HD proteins (10). Biochemical and genetic studies suggest that residue 50 is important for dictating the preference of the dinucleotide immediately 5′ to the ATTA core sequence (5′-NNATTA-3′) that is present in many HD binding sites (11–17). These results suggest that a partial DNA recognition code may exist for residue 50 in the HD.

Although many HD proteins bind specifically to DNA in vitro, they often interact with cofactors to bind with higher affinity and specificity to their sites in vivo. Interactions with different cofactors may affect the specificity of binding by the HD protein such that its in vivo target specificity is different from its apparent binding specificity in vitro. It is important to examine the specificity of these proteins in vivo as well as in vitro. We have therefore chosen to examine the binding specificity of the Matα2 HD protein from the yeast Saccharomyces cerevisiae because the interactions with its cofactors, the Mata1 and Mcm1 proteins, and binding to its target sites have been well studied in vitro and in vivo. In the α cell type, α2 interacts with Mcm1, a member of the strongly conserved MADS box family of DNA-binding proteins (18). The α2 and Mcm1 proteins bind cooperatively as a heterotetramer complex to conserved sites upstream of α-specific genes to repress transcription (19, 20). Mcm1 binds as a dimer to the center of a partially symmetric site and is flanked on either side by monomer binding sites for the α2 HD protein. Although α2 and Mcm1 can bind to these sites on their own in vitro, both proteins are required for repression in vivo. In diploid a/a cells, α2...
bonds with α1, another HD protein, to repress transcription of haploid-specific genes. The α2 and α1 proteins form a heterodimer complex, with each HD binding to one half-site (21). The structures of the α2 HD binding DNA on its own and in combination with α1 or Mcm1 have been determined (3, 4, 22, 23). The DNA sequence requirements for recognition by these complexes have also been determined in vivo and in vitro (24–28). These structural, genetic, and biochemical data provide excellent models for how the α2 protein recognizes its site alone and in combination with its cofactors (see Fig. 1). In this study, we investigated whether the sequence recognition code that has been determined for other HD proteins (11–17) is similar for α2 in complex with its cofactors in vitro and in vivo.

MATERIALS AND METHODS

Plasmids—Transcription reporter plasmids with wild-type or mutant α2-Mcm1 or α1-α2 binding sites were previously constructed by inserting double-stranded oligonucleotides containing these sites with TCGA overhangs into the XhoI site between the UAS and TATA elements of the CYC1-lacZ promoter of pTBA23 (27–29). The wild-type α2-Mcm1 site used in these experiments is a symmetric site derived from a consensus of the wild-type sites found upstream of α-specific genes (29). Mutant derivatives of this site contain symmetric base pair substitutions in each of the α2 binding sites. The mutant sites were named by describing the original nucleotide, the positions mutated, and the substituted nucleotide. For instance, T4A is a symmetric mutant in which T at position 4 is mutated to A, and A at the symmetric position 14 is mutated to T. The α1-α2 site used in these studies is derived from a consensus of the wild-type binding sites found upstream of haploid-specific genes (28). Mutant derivatives of the α1-α2 site contain base pair substitutions only in the α2 half-site and are in the same relative position as the mutations in the α2-Mcm1 site.

Derivatives of plasmid pAV115, a yeast CEN LEU2 plasmid containing a 4.3-kilobase MATα locus with wild-type α2 or the SS01, SS05Q, or N471 mutant, have been described (30). For comparison of the α2 HD with other HD structures, we have utilized the numbering system most commonly used for the 60-residue HD. Ser1 in the sugar phosphate backbone of the DNA. We were interested in whether the strong conservation of these residues in different HD proteins as the residues that are involved in contacts with the sugar phosphate backbone of the DNA. We were interested in whether the strong conservation of these residues in different HD proteins is because they have an essential role in DNA binding. To test the role of these side chains, we constructed alanine substitutions of each residue in α2 that contacts DNA in the co-crystal structures (Fig. 1) (3, 22, 23). To monitor the ability of the mutants to repress transcription in complex with Mcm1, the α2 mutants were transformed into a malα strain bearing an integrated CYC1-lacZ reporter containing an α2-Mcm1 binding site in the promoter. The ability of the α2 mutants to repress lacZ expression was measured by liquid β-galactosidase assays (Table I). The majority of the α2 alanine substitution mutants displayed significantly lower levels of repression in complex with Mcm1. Even substitutions of residues that directly contact DNA through water-mediated contacts, such as L26A and N474, had a large effect on repression in complex with Mcm1, suggesting that they are making important contributions to DNA binding. Substitutions of residues in the N-terminal arm, such as Y3A, R4A, and G5A, appeared to have slightly less effect on repression than substitutions in the recognition helix. This suggests that the minor groove contacts made by these residues may not be as important for DNA binding as major groove contacts made by residues in the recognition helix.

The level of repression of the α2 mutants in complex with α1 was monitored in a MATα strain with an integrated CYC1-lacZ reporter containing an α1-α2 binding site in the promoter (Table I). As we have previously observed, many of the substitutions of residues that contact the DNA have little or no effect on repression in complex with α1 (28, 30). The only substitutions that showed significant reductions in α1-α2-mediated repression were F8A and W48A. In addition to contacting the DNA, these side chains make numerous contacts with other residues in the hydrophobic core of the HD; and therefore, alanine substitutions of these residues could affect the folding or stability of the protein (3, 22, 23). However, Western blot analysis

was expressed and purified from strain BL21(DE3) with plasmid pYJ173 (28). The α2 and α1 proteins were purified on nickel resin columns to >90% homogeneity according to the manufacturer's protocol (Novagen). The concentration of each protein was determined by Bradford assays (Bio-Rad) and then normalized and verified by Coomassie Blue staining of SDS-polyacrylamide gels and electrophoresed for 1.5 h at 200 V. Gels were dried and exposed to phosphor screens, and images were scanned on a Molecular Dynamics PhosphorImager. The relative α2 DNA-binding affinity for each site was calculated by comparing the percentage of fragment bound at different α2 protein concentrations.

RESULTS

Residues in the α2 HD That Contact the DNA Are Required for DNA Binding and Repression—It was previously shown that residues in the α2 HD that make base-specific contacts with the DNA have important roles in DNA binding and transcriptional repression (30). However, many of these residues are not as nearly as well conserved among the different HD proteins as the residues that are involved in contacts with the sugar phosphate backbone of the DNA. We were interested in whether the strong conservation of these residues in different HD proteins is because they have an essential role in DNA binding. To test the role of these side chains, we constructed alanine substitutions of each residue in α2 that contacts DNA in the co-crystal structures (Fig. 1) (3, 22, 23). To monitor the ability of the mutants to repress transcription in complex with Mcm1, the α2 mutants were transformed into a malα strain bearing an integrated CYC1-lacZ reporter containing an α2-Mcm1 binding site in the promoter. The ability of the α2 mutants to repress lacZ expression was measured by liquid β-galactosidase assays (Table I). The majority of the α2 alanine substitution mutants displayed significantly lower levels of repression in complex with Mcm1. Even substitutions of residues that directly contact DNA through water-mediated contacts, such as L26A and N474, had a large effect on repression in complex with Mcm1, suggesting that they are making important contributions to DNA binding. Substitutions of residues in the N-terminal arm, such as Y3A, R4A, and G5A, appeared to have slightly less effect on repression than substitutions in the recognition helix. This suggests that the minor groove contacts made by these residues may not be as important for DNA binding as major groove contacts made by residues in the recognition helix.

The level of repression of the α2 mutants in complex with α1 was monitored in a MATα strain with an integrated CYC1-lacZ reporter containing an α1-α2 binding site in the promoter (Table I). As we have previously observed, many of the substitutions of residues that contact the DNA have little or no effect on repression in complex with α1 (28, 30). The only substitutions that showed significant reductions in α1-α2-mediated repression were F8A and W48A. In addition to contacting the DNA, these side chains make numerous contacts with other residues in the hydrophobic core of the HD; and therefore, alanine substitutions of these residues could affect the folding or stability of the protein (3, 22, 23). However, Western blot analysis...
FIG. 1. Structure of the $\alpha_2$ HD and DNA contacts observed in the crystal structures. A, summary of the contacts by the $\alpha_2$ HD observed in the $\alpha_2$-Mcm1-DNA and $\alpha_1$-$\alpha_2$-DNA ternary complexes (4, 23). The DNA sequence shown is one $\alpha_2$ half-site used in the structure of the $\alpha_1$-$\alpha_2$-DNA ternary complex and is identical to the wild-type consensus $\alpha_2$ binding sites used in this study. Arrows represent base-specific contacts, and lines with circles represent sugar phosphate backbone contacts that are observed in the co-crystal structures. Water-mediated contacts are indicated by a circled W. For comparison between the different HD proteins, the numbering system we have used is the position of each residue in relation to the HD and not to the full-length protein. B and C, models of the $\alpha_2$ HD bound to DNA derived from the crystal structure of the $\alpha_1$-$\alpha_2$-DNA ternary complex (4). B, base-specific contacts in the major groove (Ser50, Asn51, and Arg54) and minor groove (Arg4, Gly5, and Arg6) of the DNA are highlighted in black. C, a model of residues involved in phosphate backbone contacts with the DNA are shown.

TABLE I

| Position | $\alpha_2$-Mcm1 repression | $\alpha_1$-$\alpha_2$ repression |
|----------|-----------------------------|-------------------------------|
|          | %                           | %                             |
| Wild-type| 100                         | 100                           |
| Y3A      | 112                         | 111                           |
| R4A      | 18                          | 123                           |
| G5A      | 32                          | 119                           |
| H6A      | 39                          | 206                           |
| R7A      | 5                           | 97                            |
| F8A      | 5                           | 43                            |
| Y25A     | 7                           | 131                           |
| L26A     | 10                          | 138                           |
| Q44A     | 5                           | 61                            |
| N47A     | 7                           | 81                            |
| W46A     | 4                           | 33                            |
| S50A     | 69                          | 113                           |
| N51A     | 7                           | 100                           |
| R53A     | 4                           | 68                            |
| R54A     | 7                           | 87                            |
| K55A     | 13                          | 109                           |
| K57A     | 15                          | 114                           |

* Values shown are percent activity of repression by each mutant relative to the wild-type protein of a CYC1-lacZ reporter promoter containing a wild-type $\alpha_2$-Mcm1 or $\alpha_1$-$\alpha_2$ binding site. For each sample, three independent transformants were assayed. For $\alpha_2$-Mcm1, wild-type $\alpha_2$ yielded 6.7 ± 0.5 $\beta$-galactosidase units, whereas a blank plasmid (no $\alpha_2$) yielded 154 ± 9.7 units. This gave a repression ratio of 23-fold, which was set at 100% repression. Fold repression values were determined for each sample in the same manner, with $\beta$-galactosidase units within 10% error. The wild-type repression ratio was then divided by the ratios determined for the mutant proteins to give percent repression values. Similar calculations were done for $\alpha_1$-$\alpha_2$, with wild-type $\alpha_2$ yielding 4.2 ± 0.4 units and a blank plasmid yielding 60 ± 7.1 units for a repression ratio of 14.5-fold.

of these mutant strains showed that proteins were present at the same level as the wild-type protein, indicating that they must only affect folding (data not shown). The observation that other alanine substitutions in $\alpha_2$ do not affect repression in complex with $\alpha_1$ suggests that these mutants do not affect expression or the ability to repress transcription. Therefore, these substitutions most likely affect DNA binding in complex with Mcm1.

We have tested a number of the mutants for their ability to bind DNA on their own or in complex with Mcm1 or $\alpha_1$ in vitro by EMSAs. As expected from the in vivo results, mutants that show large decreases in the ability to repress transcription also show large decreases in their ability to bind DNA on their own or in complex with Mcm1 (28, 30) (data not shown). Mutants that had small decreases in the level of repression, such as S50A and R132A, showed only moderate reductions in DNA-binding affinity. The in vitro DNA binding data therefore agree well with the in vivo repression results and support the model that the mutants fail to repress transcription in complex with Mcm1 because they no longer bind DNA with wild-type affinity. These results show that most of the individual protein side chain contacts with the DNA observed in the $\alpha_2$ co-crystal structures are important for the DNA-binding activity of the protein.

Residue 50 in the $\alpha_2$ HD Has Relaxed Sequence Requirements—Among the alanine substitutions in $\alpha_2$, the S50A mutant showed one of the smallest decreases in repression with Mcm1 (Table I). However, in many other HD proteins, residue 50 is an important determinant of DNA-binding specificity (11–17). We were therefore interested if other amino acid substitutions of this residue would affect the DNA-binding affinity or specificity of the protein. Ser$^{50}$ was substituted with amino acids that are often found at this position in other HD proteins (10), and the mutants were assayed for the ability to repress transcription of a promoter containing a wild-type $\alpha_2$-Mcm1 site (Table II, second column). Substitutions with relatively small side chains, such as S50A, S50G, S50C, and S50T, caused only moderate reductions in the level of repression in comparison with the wild-type protein. These results suggest that as long as the side chain is relatively small, there are no strong
Substitutions of Ser50 Have Different Sequence Preferences

We next tested whether the decreases in the levels of repression by the Ser50 substitutions are a result of decreases in the DNA-binding affinity of the mutant proteins. Each of the α2 Ser50 mutants was expressed and purified from *Escherichia coli*, and the DNA-binding affinity was assayed by EMSAs. In general, the effects of the mutations on the DNA-binding affinity in vitro correlated well with the repression activity in vivo. For example, the S50N and S50A mutants repressed the *lacZ* reporter almost as well as the wild-type α2 protein, and these proteins bound to the wild-type site with similar binding affinity as the wild-type protein (Fig. 2A). Mutants that showed lower levels of repression, such as S50C, showed further reductions in binding affinity. Finally, mutants that completely failed to repress transcription, such as S50K, S50E, and S50R, showed a >50-fold decrease in DNA-binding affinity for the wild-type site. These results show that for a small side chain, there are no stringent requirements at residue 50 in the α2 HD for binding to or repression of the wild-type site. However, there is specificity against having larger side chains at this position, presumably because of steric interference with the DNA.

**Substitutions of Ser50 Have Different Sequence Preferences for Positions 2 and 3 in the α2-Mcm1 Binding Site**—In the α1-α2 structure, the Ser50 side chain makes two water-mediated hydrogen bonds with the A5 and T6 base pairs in the α2 recognition sequence (5′-CATGTA A-3′) (4). These positions are strongly conserved among the natural α2-Mcm1 sites, and some of the base pair substitutions at these positions cause a >30-fold decrease in repression (27). To examine whether amino acid substitutions at residue 50 affect the sequence-specific preferences at positions 2 and 3 in complex with Mcm1, the α2 mutants were assayed for repression of mutant α2-Mcm1 sites containing base pair substitutions at these positions (Table II).

The S50N mutant exhibited almost the same sequence preference at positions 2 and 3 as the wild-type α2 protein. The S50N mutant showed a preference among the mutant sites for G at the second position and showed a slight preference for C over either G or A at the third position. Mutants with small amino acid side chain substitutions, such as S50A, S50G, and S50T, showed further decreases in repression with the mutant sites, indicating that despite somewhat relaxed requirements for a specific amino acid at this position for binding to the wild-type site, there is still sequence specificity for base pairs contacted by this residue. Interestingly, there were slight differences in the sequence preferences at position 2 among the different amino acid substitutions. For example, the S50A mutant had a strong preference against having T at the second position in the site, whereas the S50G mutant appeared to slightly prefer T over G or C at this position. The S50C mutant and wild-type proteins showed a slight difference in sequence specificity. These results indicate that each of these amino acid side chains has its own sequence-specific requirements and is able to at least partially discriminate among the mutant sites.

The α2 S50K and S50R Mutants Show Altered DNA-binding Specificity—Most of the residue 50 mutants with large or charged amino acid substitutions, such as S50I, S50H, S50Q, and S50E, showed very little repression activity with the wild-type or mutant binding sites (Table II). In contrast, the S50K mutant repressed the A5T and T6G sites slightly better (~2.5-fold for each) than it repressed the wild-type site and repressed significantly better (10-fold) through the A5G site. We further tested this mutant protein with the CCGGTAAA site, but found that the double GG mutant did not result in a higher level of repression compared with a single mutation at position 2 in the β-galactosidase assay (data not shown). This result suggests that the effects of S50K binding to sites with G substitutions at positions 2 and 3 may not be additive. The S50R mutant also repressed promoters containing sites with G and T substitutions at position 2 better (~3-fold) than a promoter containing the wild-type α2-Mcm1 site. Although the repression by mutant proteins through the mutant sites was not restored to wild-type levels, these results clearly indicate that both S50K and S50R have altered specificity for the mutant sites in complex with Mcm1 in vivo.

The S50K and S50R mutants were further tested for their DNA-binding affinity and specificity by EMSAs using different α2-Mcm1 binding sites. In general, the DNA-binding affinity of the S50K mutant protein for the different sites correlates well with the in vivo repression results. The S50K mutant bound to

### Table II

| α2 mutant       | Position 2 mutants | Position 3 mutants |
|-----------------|--------------------|--------------------|
| Ser50 (wild-type) |                    |                    |
| S50N            | 100                |                    |
| S50A            | 52                 |                    |
| S50G            | 24                 |                    |
| S50C            | 22                 |                    |
| S50T            | 18                 |                    |
| S50I            | 3                  |                    |
| S50H            | 2                  |                    |
| S50Q            | 2                  |                    |
| S50R            | 3                  |                    |
| S50K            | 2                  |                    |
| S50E            | 1                  |                    |

Values are the percent activity of repression by the wild-type (WT) protein relative to the wild-type protein. The repression ratio for wild-type α2 at a wild-type site (20-fold) was then divided by the ratios determined for the mutants, giving percent repression values.
the A2G site with at least 5-fold higher affinity compared with the wild-type site (Fig. 2B). This result correlates well with the observation that this mutant represses promoters with the A2G site ~10-fold better than a promoter with the wild-type site. S50K showed a 2.5-fold increase in repression of a promoter with the T5G site, and we observed a similar modest increase in the DNA-binding affinity of the mutant protein for this site (data not shown). The S50R mutant also showed a significant increase in binding affinity for the A2G site over the wild-type site, which correlates well with the increase in repression by this mutant of a promoter containing this site (Fig. 2B). Taken together, both in vivo and in vitro experiments have demonstrated that the S50K and S50R substitutions alter the DNA-binding specificity of the a2 HD.

Substitutions at Residue 50 in the a2 HD Do Not Affect a1-a2-mediated Repression—The results shown above indicate that amino acid substitutions of residue 50 alter the sequence specificity of a2 in binding alone in vitro and in complex with Mcm1 in vivo. We have previously shown that substitutions of DNA-binding residues in the a2 HD have little or no effect on a1-a2-mediated repression or DNA-binding affinity, suggesting that these side chains do not make essential contributions to the complex (30). However, substitutions of base pairs contacted by these residues have a large effect on binding and repression, showing that there are sequence-specific requirements at these positions in the DNA site (28). Therefore, even though these residues are not essential for a2 binding in complex with a1, substitutions of the amino acids at these positions may alter the binding specificity of the complex. To investigate whether residue 50 has a role in a2 DNA recognition in complex with a1, we assayed several of the a2 mutants for their effects on binding to and repression of the wild-type and mutant a1-a2 sites. Plasmids containing a2 mutants with Ser50 replaced by Ala, Gln, Ile, Lys, or Arg were co-transformed into a MATa strain and assayed with derivatives of a CYC1-lacZ reporter plasmid containing substitutions at positions 2 and 3 in the a2 half-site of the a1-a2 site in the promoter (Fig. 3A). The S50A, S50Q, and S50I mutants had essentially wild-type levels of repression of a promoter with the consensus site, indicating that these substitutions do not greatly affect the binding affinity of the complex. The wild-type protein showed slightly reduced repression of promoters containing sites with base pair substitutions at positions 2 and 3 in the a2 half-site. Interestingly, both the S50A and S50I mutants discriminated among different a1-a2 sites in roughly the same way as wild-type a2. These mutants were more tolerant to substitutions at position 2 than at position 3, with a base preference of A > T > C > G at position 2. However, the a2 S50Q mutant differed from the wild-type protein and preferred G instead of C at position 2.

In contrast to the S50A, S50I, and S50Q substitutions, the S50K mutant had significantly reduced repression with the wild-type consensus a1-a2 site. However, repression by S50K was restored to nearly wild-type levels at the A2G or T5G site (Fig. 3A). We next examined the DNA-binding affinity of the S50K mutant for the consensus and mutant a1-a2 sites (Fig. 3B). The S50K mutant bound the wild-type site with a 20-fold decrease in binding affinity compared with wild-type a2. However, the S50K protein bound the A2G mutant site with approximately the same affinity as that of the wild-type protein binding to the consensus site. We conclude that the in vitro DNA-binding affinity of the S50K mutant for different a1-a2 sites corresponds with the in vivo repression activity and that this mutant has higher affinity for a1-a2 sites with G at position 2 or 3 than for the consensus site. Although the S50R mutant did not show as large a decrease in repression of the wild-type a1-a2 site compared with the a2-Mcm1 site, it also showed changes in specificity, preferring G or T at position 2 over the presence of C. This shows that the S50K and S50R substitutions alter the DNA-binding specificity of the a2 HD in complex with both Mcm1 and a1.

Substitutions of Other Residues in the HD Are Unable to Alter a2 DNA-binding Specificity—The results shown above indicate that the DNA-binding specificity of a2 can be altered by making changes at residue 50 in the HD. We were therefore interested if it is possible to alter the binding specificity of a2 by making amino acid substitutions of other residues in the HD that contact DNA. We focused our effort on residues that make base-specific contacts with the DNA and that are variable among different HD proteins (10). Arg84 in a2 makes a base-specific contact with the N-7 group of the G4 base in the co-crystal structures (Fig. 1) (3, 4, 23). However, in many other
HD proteins, this residue is Ala and is not involved in a direct contact with the DNA. Instead, many of these HD proteins use Ile at position 47 in the HD to make a van der Waals contact with T on the other strand of DNA at position 4 in the site. In α2, there is an Asn side chain at position 47 that makes only an indirect contact with the DNA through a water molecule (4). We therefore tested whether the N47I and R54A amino acid substitutions in α2 would change the specificity of the protein from G:C at this position in the site to an A:T base pair. We constructed each of the single amino acid substitutions and the double mutant and assayed for the ability of the mutants to repress promoters containing wild-type and G4A mutant double mutant to repress promoters with the G4A site is that affinity but relaxed specificity at position 4 in site. (4)). This result suggests that the R54A mutant has decreased affinity but relaxed specificity at position 4 in site.

One possible explanation for the failure of the N47I/R54A double mutant to repress promoters with the G4A site is that residue 50 may be required to play a greater role in DNA binding in HD proteins with Ile at position 47 and Ala at position 54. In the Drosophila Engrailed HD, residue 50 is a Gln. We therefore made the N47I/S50Q/R54A triple amino acid substitution in α2 and assayed its ability to repress promoters containing the wild-type site and a mutant site that resembles the Engrailed binding site of TAATTA. However, the N47I/S50Q/R54A mutant was unable to repress reporters containing either the consensus α2-Mcm1 site or any of the chimeric Engrailed-Mcm1 sites (data not shown). This mutant also failed to show any further decrease in binding to the mutant site (Fig. 4A).

We also examined the role of Lys55 in DNA-binding specificity. In many HD proteins, residue 55 is a Lys that makes a contact with the phosphate backbone. However, in the α1, Pbx1, and Extradenticle HD proteins, this residue is an Arg that uses a similar mechanism as other HD proteins to bind DNA. Despite direct contacts with the DNA in both the α2-Mcm1 and α1-α2 crystal structures, many of the residues in the N-terminal arm have only a weak role in repression with Mcm1.
Biochemical, genetic, and structural studies have shown that residues in a flexible linker region adjacent to the N-terminal arm are involved in direct interactions with Mcm1 (23, 29, 34). In addition, two residues in the N-terminal arm, Tyr and His, make a direct contact with Mcm1. It is possible that the protein interactions by these residues replace the need for strong DNA contacts by the N-terminal arm. Another explanation for the relatively weak effect of mutations in the N-terminal arm is based on the observation that there is considerable variation among the natural a2-Mcm1 sites in the bases between the a2 and Mcm1 recognition sites (27). To bind cooperatively with Mcm1, the a2 N-terminal arm must be able to accommodate differences in the spacing between the sites as well as differences in the base pair sequences. DNA contacts by the N-terminal arm may therefore need to be relatively weak so that a loss of a contact to accommodate sites with alternate spacing will not cause a significant reduction in DNA-binding affinity. Although the N-terminal residues do not make large contributions to the binding affinity of the complex, they may still perform an important role in determining the specificity of DNA binding by the complex. In the context of the consensus a2-Mcm1 site used in these studies, mutations at the base pairs contacted by these residues cause large decreases in DNA-binding affinity and repression, indicating that there are important base-specific contacts at these positions (27).

As observed previously, although many of the side chains in the a2 HD are important for repression in complex with Mcm1, substitutions of these residues have little effect on repression in combination with a1 (28, 30). However, several of the alanine substitutions, such as W48A and F8A, show a significant decrease in a1-a2-mediated repression of the reporter promoter. In addition to directly contacting the DNA, a large portion of both of these side chains is buried in the hydrophobic core of the HD. Since Western blot analysis showed that these mutations are expressed at roughly the same level as the wild-type protein, it is likely that the alanine substitutions of these residues alter the HD structure. These changes likely disrupt multiple DNA contacts, resulting in a decrease in repression with a1.

As observed with the Engrailed HD, Ala substitution of Ser50 in a2 causes only a slight decrease in DNA-binding affinity and repression (15). We have also found that substitutions with other small side chains, such as Cys, Thr, and Asn, show relatively minor effects on repression or DNA binding. The amino acid requirements at residue 50 therefore appear to be somewhat degenerate, allowing for substitutions with small amino acids. This degeneracy may be due in part to the fact that these contacts are at the edge of the binding site, which allows the conformation of the side chain to be altered slightly to accommodate more optimal contacts with the DNA. In support of this model, solution studies of the Antennapedia HD protein and crystallographic studies of the Even-skipped HD have shown that the Glu50 side chain exists in multiple conformations when bound to DNA (8, 35). These studies, along with other HD crystal structures, have also shown that there are often water molecules that contact both residue 50 and the DNA, forming indirect protein-DNA contacts. In mutants with substitutions at residue 50 with small side chains, these water molecules may reposition to make optimal contacts between the protein and DNA, as has been observed for the Q50A co-crystal structure of the Engrailed protein (36). The smaller side chain in this mutant permits three extra water molecules to form a cage-like structure around the residue that mediates contacts between the protein and DNA. It is likely that many of the Ser50 substitutions in a2 with small side chains contain additional and/or repositioned water molecules near the side chain. The specific positioning of these water molecules by each side chain may explain why we observed subtle differences in the base pair specificity of some of the mutant proteins with small side chains. In contrast to what we have observed with the smaller side chains, substitutions with many of the larger side chains do not function in place of Ser50 in a2. Either these substitutions may sterically interfere with the protein-DNA interface, or the size of the side chain may exclude water molecules from mediating a contact between the protein and DNA. Since many HD proteins contain these larger side chains, this suggests that to accommodate the larger side chains, these proteins may dock with the DNA in a slightly different manner from a2.

Although our results suggest that the amino acid requirements for residue 50 in a2 are partially degenerate, genetic and biochemical studies have shown that residue 50 plays an important role in determining the sequence specificity of many HD proteins, enabling them to distinguish between different NNATTA sites (11–13, 15–17, 37). The most striking altered-specificity mutations usually involve substituting Lys at residue 50, such as in the Drosophila Engrailed Q50K, Paired S50K, and Fushi tarazu Q50K mutants. In each case, the Lys substitution prefers to bind a sequence of GGATTA. We observed a similar change in the DNA-binding specificity with the Lys substitution at residue 50 in the a2 HD. The a2 S50K mutant clearly prefers a G base at positions adjacent to the recognition core, such as GTGTAA, AGGTAA, and GGGTAA, in both a2-Mcm1- and a1-a2-mediated repression, whereas the
wild-type α2 protein prefers the ATGTAA site. In the crystal structure of the Engrailed Q50K mutant bound to DNA, the Lys^50 side chain makes two hydrogen bonds with the O-6 and N-7 groups of the G base at the first position (GGATTA) and a hydrogen bond with O-6 of the G base at the second position (GGATTA) in the binding site (38). These hydrogen bonds are likely to contribute much more energy to binding than the van der Waals interactions between wild-type Glu^50 and the methyl group of the T base at the first position in the TAATTA site (2). In the case of α2, although we do not know whether the Lys side chain at residue 50 would make similar hydrogen bonds with the G base, it is clear that it makes a more favorable contact with G than with other bases at this position in the site. The altered specificity preference of the α2 S50R mutant for sites with G at position 2 may also be the result of that side chain making a single hydrogen bond with the G base. However, unlike the S50K mutant, the S50R mutant is unable to recognize a site with G at position 3. Both S50K and S50R show some preference for T at the second position in the site over a G at position 2, although we do not know whether the Lys side chain at residue 50 alter the DNA-binding specificity of α2 HD proteins. These results suggest that there is at least a partial DNA recognition code for residue 50 in HD proteins.

Acknowledgments—We thank Janet Mead and Ronald Mc Cord for providing plasmid constructs and members of the laboratory for discussions on the manuscript.

References

1. Gehring, W. J., Affolter, M., and Burglin, T. (1994) Annu. Rev. Biochem. 63, 487–526
2. Krämer, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Cell 63, 579–590
3. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) Cell 67, 517–528
4. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Science 270, 262–269
5. Otting, G., Qian, Y. Q., Billette, M., Mueller, M., Affolter, M., Gehring, W. J., and Wuthrich, K. (1990) EMBO J. 9, 3085–3092
6. Qian, Y. Q., Billette, M., Otting, G., Muller, M., Gehring, W. J., and Wuthrich, K. (1989) Cell 59, 573–580
7. Wang, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) Cell 82, 709–719
8. Hirsch, J. A., and Aggarwal, A. K. (1995) EMBO J. 14, 6280–6281
9. Schott, O., Billette, M., Leitinger, B., Wider, G., and Wuthrich, K. (1997) J. Mol. Biol. 267, 673–683
10. Burglin, T. R. (1995) in Biodiversity and Evolution (Arac, R., Kato, M., and Doi, Y., eds), pp. 291–336, National Science Museum Foundation, Tokyo
11. Hanes, S. D., and Brent, R. (1989) Cell 57, 1275–1283
12. Treisman, J., Goncey, P., Vashisthath, M., Harris, E., and Desplan, C. (1989) Cell 59, 553–562
13. Percival-Smith, A., Muller, M., Affolter, M., and Gehring, W. J. (1990) EMBO J. 9, 3967–3974
14. Hanes, S. D., and Brent, R. (1991) Science 251, 426–430
15. Ades, S. E., and Sauer, R. T. (1994) Biochemistry 33, 9187–9194
16. Wilson, D. S., Sheng, G., Jin, S., and Desplan, C. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 6886–6891
17. Damante, G., Pelizzari, L., Esposto, G., Fogolari, F., Viglino, P., Fabbro, D., Tedj, G., Formisano, S., and Di Laura, R. (1996) EMBO J. 15, 4992–5000
18. Keleher, C. A., Passmore, S., and Johnson, A. D. (1989) Mol. Cell. Biol. 9, 5228–5230
19. Keleher, C. A., Gouette, C., and Johnson, A. D. (1988) Cell 53, 927–936
20. Passmore, S., Elble, R., and Tey, B. K. (1989) genes Dev. 3, 921–935
21. Gouette, C., and Johnson, A. D. (1994) EMBO J. 13, 1434–1442
22. Li, T., Jin, Y., Vershon, A. K., and Wolberger, C. (1998) Nucleic Acids Res. 26, 5707–5714
23. Tan, S., and Richmond, T. J. (1998) Nature 391, 660–666
24. Smith, D. L., and Johnson, A. D. (1992) Cell 68, 133–142
25. Smith, D. L., and Johnson, A. D. (1994) EMBO J. 13, 2373–2387
26. Acton, T. B., Zhong, H., and Vershon, A. K. (1997) Mol. Cell. Biol. 17, 1881–1889
27. Zhong, H., and Vershon, A. K. (1997) J. Biol. Chem. 272, 8402–8409
28. Jin, Y., Zhong, H., and Vershon, A. K. (1999) Mol. Cell. Biol. 19, 585–593
29. Mead, J., Zhong, H., Acton, T. B., and Vershon, A. K. (1996) Mol. Cell. Biol. 16, 2135–2143
30. Vershon, A. K., Jin, Y., and Johnson, A. D. (1995) Genes Dev. 9, 182–192
31. Jin, Y., Mead, J., Li, T., Wolberger, C., and Vershon, A. K. (1995) Science 270, 290–293
32. Passmore, J. M., Ryo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999) Nature 397, 714–719
33. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999) Cell 96, 587–597
34. Vershon, A. K., and Johnson, A. D. (1993) Cell 72, 105–112
35. Billette, M., Guntert, P., Lupinski, P., and Wuthrich, K. (1996) Cell 85, 1057–1065
36. Grant, B. A., Roudl, M. A., Klemm, J. D., and Pabo, C. O. (2000) Biochemistry 39, 8187–8192
37. Damante, G., Fabbro, D., Pelizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M., Caucci, S., Quasigiorgio, F., Formisano, S., and Di Laura, R. (1994) Nucleic Acids Res. 22, 3073–3083
38. Tucker-Kellogg, L., Roudl, M. A., Chambers, K. A., Ades, S. E., Sauer, B. T., and Pabo, C. O. (1997) Structure 5, 1047–1054
Altering the DNA-binding Specificity of the Yeast Matα2 Homeodomain Protein
Jonathan R. Mathias, Hualin Zhong, Yisheng Jin and Andrew K. Vershon

J. Biol. Chem. 2001, 276:32696-32703.
doi: 10.1074/jbc.M103097200 originally published online July 3, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M103097200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 37 references, 11 of which can be accessed free at
http://www.jbc.org/content/276/35/32696.full.html#ref-list-1