Engineered improvements in DNA-binding function of the MATa1 homeodomain reveal structural changes involved in combinatorial control

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · MARCH 2002

Impact Factor: 4.33 · DOI: 10.1006/jmbi.2001.5333 · Source: PubMed

7 AUTHORS, INCLUDING:

Jonathan R Mathias
Georgia Health Sciences University

Andrew K Vershon
Rutgers, The State University of New Jersey

Susan M Baxter
California State University, Chancellor's Off...

doi:10.1006/jmbi.2001.5333 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 316, 247–256

JMB

Engineered Improvements in DNA-binding Function of the MATa1 Homeodomain Reveal Structural Changes Involved in Combinatorial Control

Beverly Hart[1]†, Jonathan R. Mathias[2]†, David Ott[3], Lynn McNaughton[1] Janet S. Anderson[3], Andrew K. Vershon[2] and Susan M. Baxter[1]*

[1]*Wadsworth Center, NY State Department of Health, Empire State Plaza, Albany NY 12201-0509, USA*

[2]*Waksman Institute and the Department of Molecular Biology and Biochemistry Rutgers University Piscataway, NJ 08854-8020, USA*

[3]*Department of Chemistry Union College, Schenectady NY 12308, USA*

**Corresponding author*

We have engineered enhanced DNA-binding function into the **a**1 homeodomain by making changes in a loop distant from the DNA-binding surface. Comparison of the free and bound **a**1 structures suggested a mechanism linking van der Waals stacking changes in this loop to the ordering of a final turn in the DNA-binding helix of **a**1. Inspection of the protein sequence revealed striking differences in amino acid identity at positions 24 and 25 compared to related homeodomain proteins. These positions lie in the loop connecting helix-1 and helix-2, which is involved in heterodimerization with the α2 protein. A series of single and double amino acid substitutions (**a**1-Q24R, **a**1-S25Y, **a**1-S25F and **a**1-Q24R/S25Y) were engineered, expressed and purified for biochemical and biophysical study. Calorimetric measurements and HSQC NMR spectra confirm that the engineered variants are folded and are equally or more stable than the wild-type **a**1 homeodomain. NMR analysis of **a**1-Q24R/S25Y demonstrates that the DNA recognition helix (helix-3) is extended by at least one turn as a result of the changes in the loop connecting helix-1 and helix-2. As shown by EMSA, the engineered variants bind DNA with enhanced affinity (16-fold) in the absence of the α2 cofactor and the variant α2/**a**1 heterodimers bind cognate DNA with specificity and affinity reflective of the enhanced **a**1 binding affinity. Importantly, *in vivo* assays demonstrate that the **a**1-Q24R/S25Y protein binds with fivefold greater affinity than wild-type **a**1 and is able to partially suppress defects in repression by α2 mutants. As a result of these studies, we show how subtle differences in residues at a surface distant from the functional site code for a conformational switch that allows the **a**1 homeodomain to become active in DNA binding in association with its cofactor α2.

Keywords: homeodomain; combinatorial control; MAT**a**1; DNA-binding affinity; HSQC

Introduction

The combinatorial control of transcription achieved by the *Saccharomyces cerevisiae* mating type proteins is a textbook example of a set of protein-protein interactions that leads to tightly regulated gene expression *in vivo*.[1] For instance, the MATα2, or α2, homeodomain has two regions flanking its homeodomain that are involved in protein-protein interactions and are important for two entirely separate functions. In both haploid α and diploid **a**-α yeast cell types, a linker region N-terminal to the α2 homeodomain interacts with the Mcm1 protein, leading to repression of **a**-specific haploid genes.[2,3] In diploid **a**-α cells, a

†These authors contributed equally to this work.

Present address: S. M. Baxter, Geneformatics, Inc., 5830 Oberlin Drive, Suite 200, San Diego, CA 92121, USA.

Abbreviations used: EMSA, electrophoretic mobility shift assay; HSQC, heteronuclear single quantum coherence; DSC, differential scanning calorimetry, ChIP, chromatin immunoprecipitation; NOE, nuclear Overhauser effect.

E-mail address of the corresponding author: susanbaxter@geneformatics.com

tail C-terminal to the α2 homeodomain binds the MAT**a**1, or **a**1, homeodomain protein to form a heterodimer that recognizes sites in the promoters of haploid-specific genes (*hsg*) to repress their transcription.[4,5] Although both α2 and **a**1 have homeodomain DNA-binding motifs, neither protein can function to repress transcription of haploid-specific genes alone and therefore must work in combination. We are interested in how the interaction between these proteins confers this regulatory activity.

The yeast **a**1 protein is distinct among the well-studied homeodomain family because it binds DNA very poorly *in vitro*.[4,6] However, in combination, the α2/**a**1 heterodimer is 3000 times more specific for *hsg* operators than for non-specific DNA.[4] Structural studies[7] of the **a**1 monomer showed that the protein has a characteristic homeodomain fold consisting of an unstructured N-terminal arm and three helices linked by two loops. The structure of the α2/**a**1/DNA ternary complex[5] confirmed that the **a**1 protein binds the major groove of DNA in a manner similar to other homeodomains and with a similar number of hydrogen bond contacts between the protein and DNA substrate. Surprisingly, although **a**1 binds DNA weakly on its own, in complex with α2, **a**1 provides the majority of the DNA specificity and affinity for the α2/**a**1- heterodimer complex.[8,9] This result explains the finding that an α2 variant, with alanine substituting for residues that make base-specific DNA contacts, can still repress transcription in complex with **a**1.[10]

Biochemical and biophysical data suggest that the **a**1 homeodomain protein is structurally stable.[7] This finding led to the hypothesis that the **a**1 protein may have evolved to be a poor DNA-binding protein as a monomer and that **a**1 undergoes a conformational change upon binding α2, effecting an increase in DNA-binding affinity. The α2/**a**1/DNA ternary complex[5] revealed that only a short alpha-helical segment of α2 contacts the **a**1 homeodomain fold. The α2 tail contacts a hydrophobic surface, formed by the helix-1-loop-1-helix-2 region, distant from the **a**1 DNA-binding surface. NMR studies[7] indicated a structural link between this surface loop, involved with protein-protein recognition, and the distal DNA-binding helix, suggesting that the α2 interaction at loop-1 could cause structural changes in **a**1 that underlie increased DNA affinity.

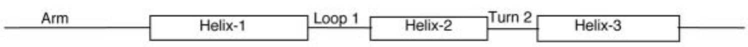
Comparison of the amino acid sequence in loop-1 with other homeodomains that do not require cofactors for DNA binding reveals some striking differences in amino acid identity; in particular two positions are notable. The highly conserved tyrosine at position 25 in the homeodomain family[11] is serine in the **a**1 sequence. At the neighboring position 24, there is glutamine in **a**1 in place of arginine found in many other homeodomain proteins. Using the link between loop-1 and the DNA-binding helix, we hypothesized that we may be able to convert **a**1 into a more competent DNA-binding homeodomain by exchanging the amino acids at positions 24 and 25 to more commonly conserved residues.

We describe here the design and characterization of a series of variant **a**1 proteins. *In vitro* and *in vivo* studies demonstrate that the variant proteins have enhanced DNA-binding affinity and retain biological specificity. Our results show one instance of how nature has evolved an on/off switch triggered by the presence of a protein cofactor. The switch is accomplished using only a few amino acid residues distant from, but structurally linked to, the DNA-binding surface of the protein. This work shows in part how combinatorial interactions can enhance the activity of a protein.

Results

Stability and structure of engineered a1 variants

Inspection of the **a**1 protein sequence revealed amino acid differences at positions 24 and 25 in loop-1, connecting helix-1 and helix-2, compared to the corresponding sequence in other homeodomain proteins (Figure 1). We hypothesized that we could design variant **a**1 proteins with enhanced DNA-binding affinity if we could mimic the loop-1 sequences of similar homeodomain sequences that bind DNA without cofactors. Both the *Drosophila melanogaster* engrailed (en) and Antennapedia (Antp) homeodomain proteins have arginine at position 24 and tyrosine at position 25. A series of **a**1 variants with single and double amino acid substitutions (**a**1-Q24R, **a**1-S25Y, **a**1-S25F, and **a**1-Q24R/S25Y) was engineered, replacing the wild-type **a**1 Q24 and S25 residues with those found in en and Antp. The variant proteins were expressed and purified for biochemical and biophysical study.




```
            1          10          20          30          40          50
a1     KKEKSPKGKSSIS  PQARAFLEQV  FRRKQSLNSK  EKEEVAKKCG  ITPLQVRVWF  INKRMRSK
en         DEKRPRTAFS  SEQLARLKRE  FNENRYLTER  RRQQLSSELG  LNEAQIKIWF  QNKRAKIKK
Antp      MERKRGRQTYT  RYQTLELEKE  FHFNRYLTRR  RRIEIAHALS  LTERQIKIWF  QNRRMKWKKEN
```

Figure 1. Sequence and approximate secondary structure alignment of the *S. cerevisiae* MAT**a**1, and *D. melanogaster* engrailed (en) and Antennapedia (Antp) homeodomains. The C-terminal **a**1 homeodomain fragment used in these studies has been numbered −3 to 57 in accordance with previous studies of homeodomains.[5,7] Positions 24 and 25 are highlighted in yellow.

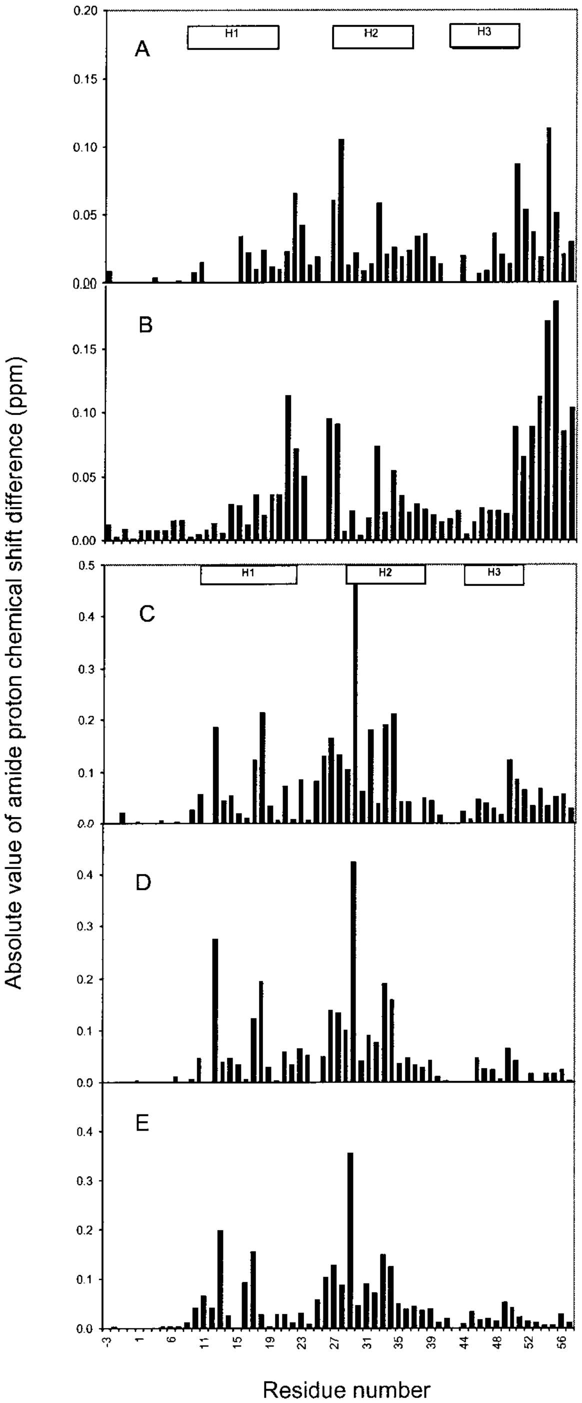


Figure 2. Absolute values of proton chemical shift changes associated with effects of amino acid substitutions in variant **a**1 proteins and α2 tail peptide binding. Amide chemical shifts are compared to the resonance shifts for wild-type **a**1 sequence. Chemical shift changes are plotted against residue position in the primary sequence of the **a**1 homeodomain. Approximate helix positions in the **a**1 homeodomain are illustrated by boxes. For (a) and (b), amide chemical shifts are compared to the resonance shifts for wild-type **a**1 sequence. (a) Amide proton chemical shift changes due to the Ser to Tyr change engineered into the **a**1-S25Y variant. (b) Amide proton chemical shift changes due to the two changes (Gln to Arg and Ser to Tyr) engineered into the **a**1-Q24R/S25Y variant. For (c), (d), and (e), amide chemical shifts for the heterodimer are compared to those for the free **a**1 homeodomain. (c) Amide proton chemical shift changes observed as a result of α2 tail peptide binding to wild-type **a**1 homeodomain. (d) Amide proton chemical shift changes observed as a result of α2 tail peptide binding to **a**1-S25Y homeodomain. (e) Amide proton chemical shift changes observed as a result of α2 tail peptide binding to **a**1-Q24R/S25Y homeodomain.

To test that the purified variant proteins are relatively stable and well folded, differential scanning calorimetric (DSC) and NMR data (not shown) were collected. Each variant was found to be equally stable as wild-type **a**1, since all have transition temperatures around 62 °C. Proton NMR spectra for all **a**1 variants (**a**1-Q24R, **a**1-S25Y, **a**1-S25F, and **a**1-Q24R/S25Y) are well resolved, have relatively sharp resonance peaks, and are very similar in appearance to wild-type **a**1 spectra. Both **a**1-S25Y and **a**1-Q24R/S25Y express well in minimal media and *U*-^{15}N-labeled samples were prepared for higher-resolution studies. The fingerprint ^{1}H-^{15}N HSQC spectra for **a**1-S25Y and **a**1-Q24R/S25Y are well resolved (data not shown) suggesting that the variant proteins are folded. In fact, there are no doubled resonances and fewer broadened resonances compared to wild-type **a**1 HSQC spectra, suggesting perhaps that the variant proteins are less flexible than wild-type **a**1 on the μs-ms timescale.

Secondary structure analysis, based on NOESY patterns and $^{3}J_{HNH\alpha}$ coupling constants, established that **a**1-S25Y and **a**1-Q24R/S25Y have the characteristic homeodomain fold, an unstructured N-terminal arm and three alpha helices connected by two loops (data not shown). Using NOESY data, the backbone amide resonances were assigned sequentially for both **a**1-S25Y and **a**1-Q24R/S25Y. Chemical shift information for the backbone amides identified the specific residues and locations in the structure that are affected by these amino acid changes (Figure 2 (a) and (b)). In both proteins, residues whose backbone amide protons underwent a significant change in chemical shift (>0.1 ppm) are located in the loop connecting helix-1 and helix-2 and at the end of helix-3, the DNA recognition helix. The magnitude of chemical shift changes in the third helix is much greater for the double mutant (**a**1-Q24R/S25Y) than for the single mutant, **a**1-S25Y.

Analysis of the NOE patterns and $^{3}J_{HNH\alpha}$ coupling constants for the residues in helix-3 for **a**1-S25Y and **a**1-Q24R/S25Y suggested a gain of helical conformation compared to wild-type **a**1 (data not shown). Wild-type **a**1 is unstructured along the backbone from R52 to the end of the amino acid sequence at K57. In **a**1-S25Y sequential backbone NOEs connect K52 and R53, suggesting that residue 53 is no longer completely unstructured. However, in the **a**1-Q24R/S25Y double mutant sequential backbone NOEs connect a stretch of residues from K52 to S56. Coupling constants ($^3J_{HNH\alpha} < 6$ Hz) indicate that residues R53 and M54 have helical backbone conformations. The amide resonance for R55 is overlapping with another resonance, preventing accurate measurement of its coupling constant. These results suggest that reordering in loop-1 due to substitutions at residues 24 and 25 leads to a conformational change in the DNA-binding helix.

Interactions of the a1 variants with the α2 C-terminal tail

Wild-type **a**1 binds the C-terminal tail of the α2 protein to form a heterodimer that binds DNA and regulates gene expression. A peptide containing only 19 amino acid residues (residues 189 to 207) from the α2 C-terminal tail is sufficient to convert wild-type **a**1 from a weak DNA-binding protein to a strong one.[9] NMR studies (J.S.A., S.M.B. & G. Hernandez, unpublished results) confirm that this 19mer peptide (α2 tail peptide) binds wild-type **a**1 in the same manner as the homeodomain construct, $\alpha 2_{128\text{-}210}$, and with the same affinity.[12] Using the [*U*-^{15}N]**a**1 proteins and unlabeled α2 tail peptide, we performed a series of HSQC experiments at a variety of peptide concentrations, keeping the **a**1 protein concentration constant. The HSQC titration series showed that the protein-peptide complex was in fast exchange on the NMR timescale, as seen for other studies of the α2/**a**1 heterodimer.[12] The two variant proteins, **a**1-S25Y and **a**1-Q24R/S25Y, also bind the α2 tail peptide with dissociation constants ($K_d = 0.3$ (± 0.1) mM) similar to wild-type **a**1.

The locations and overall pattern of chemical shift changes observed upon α2 tail peptide binding are similar for all **a**1 proteins studied, suggesting that the peptide binds to the same surface in the wild-type and variant **a**1 proteins. Significant backbone amide proton chemical shift changes are observed upon peptide binding for residues in the helix-1-loop-1-helix-2 region of all the **a**1 proteins (Figure 2(c)-(e)). In wild-type **a**1, relatively small, but obvious changes occur at the C-terminal end of the protein. These changes are not observed in either of the **a**1 variants, S25Y and Q24R/S25Y. Fewer residues shift significantly (>0.1 ppm) for the variant proteins compared to wild-type **a**1, even though the regions of the proteins affected by peptide binding are the same (Figure 3).

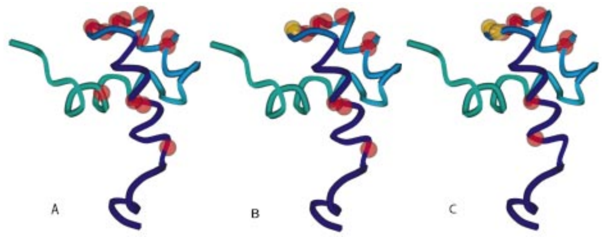


Figure 3. Regions in the **a**1 variant structures affected by α2 cofactor binding. Residues in which the amide proton resonances shift significantly (>0.1 ppm) upon α2 cofactor binding are represented by red balls. Yellow balls represent positions of the S25Y (b) and Q24R/S25Y (c) amino acid changes. The DNA-binding helix (at the back of the protein as pictured here) is colored green. Wire-frame cartoon representations, generated by MolScript[29] are based on the wild-type **a**1 homeodomain NMR structure (RSCB accession no. 1F43) and are not meant to represent the exact variant protein structures. (a) Residues in the wild-type **a**1 homeodomain affected by α2 binding. (b) Residues in the **a**1-S25Y homeodomain affected by α2 binding. (c) Residues in the **a**1-Q24R/S25Y homeodomain affected by α2 binding.

DNA-binding activity of a1 variants

To test whether the conformational changes observed between the wild-type **a**1 protein and the **a**1-Q24R/S25Y variants translate to higher DNA-binding affinity, we performed electrophoretic mobility shift assays (EMSA). The probe for these assays was a radiolabeled DNA fragment containing an α2/**a**1 consensus site derived from the sequences of 17 naturally occurring α2/**a**1 binding sites (Figure 4(a)).[8,10] In the absence of α2, the **a**1-S25Y protein has fourfold greater DNA-binding affinity than wild-type **a**1, while **a**1-Q24R/S25Y binds DNA with 16-fold greater affinity (Figure 4(b)). These results show that the mutations that alter the folding of the **a**1 homeodomain also work to increase the DNA-binding affinity of the protein. In contrast, the **a**1-Q24R protein shows slightly weaker DNA-binding affinity than the wild-type protein, indicating that this change alone is not sufficient to confer DNA binding.

Although the **a**1 variants are binding DNA with higher affinity, it is possible that they are doing so non-specifically. To address this possibility, we assayed the binding of the proteins to mutant probes that replace bases essential for DNA-binding by α2 (T_7G) or **a**1 ($T_{19}G$) (Figure 4(c)).[8] The α2 mutant half-site has no effect on the DNA-binding affinity of either wild-type **a**1 or the **a**1-Q24R/S25Y variant. However, the **a**1 mutant half-site clearly prevents binding by either protein, demonstrating that the **a**1 variants are binding specifically to the **a**1 half-site in these assays.

We reasoned that the S25Y substitution increases DNA-binding affinity of **a**1 by altering the structure of the **a**1 homeodomain. However, in many homeodomains there is a tyrosine at residue 25 that contacts DNA through the phosphate backbone.[13,14] In fact, a substitution from threonine to tyrosine at this position in the *S. cerevisiae* Pho2 homeodomain results in increased DNA-binding affinity *in vitro* and transcriptional activation *in vivo*.[15] It is therefore possible that tyrosine at position 25 of the **a**1 homeodomain could also make a direct contact with the DNA, and that this additional contact may be responsible for the

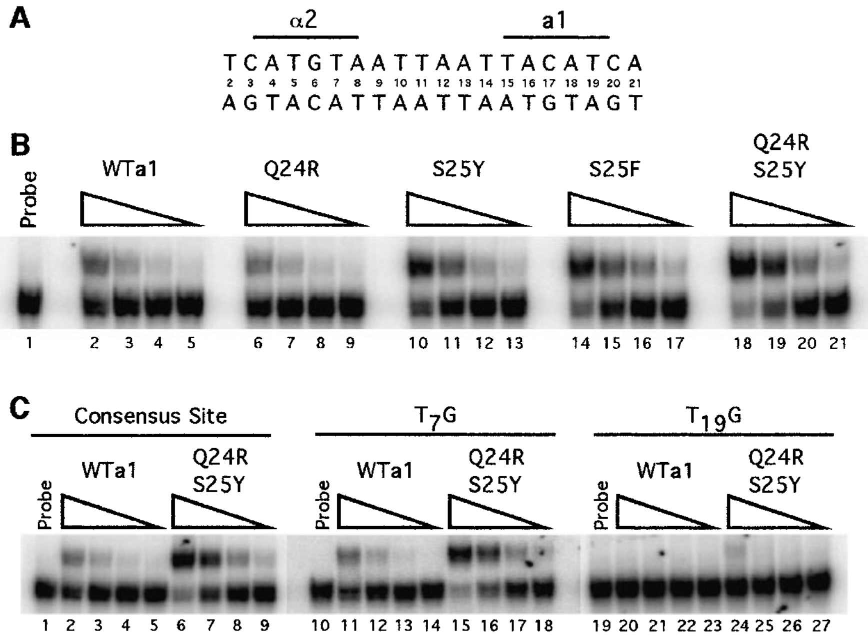


Figure 4. EMSA of wild-type and variant **a**1 proteins. (a) The wild-type α2/**a**1 binding site used as a probe for these assays with the α2 and **a**1 half-sites indicated by black bars. (b) EMSA of wild-type and variant **a**1 proteins on a wild-type site. Protein concentrations range by fourfold dilutions from 1 μM (lanes 2, 6, 10, 14, and 18) to 16 nM (lanes 5, 9, 13, 17, and 21) for each sample. (c) EMSA of wild-type **a**1 and **a**1-Q24R/S25Y on wild-type (1-9), T_7G (10-18) and $T_{19}G$ (19-27) mutant sites. Protein concentrations range by fourfold dilutions from 1 μM (lanes 2, 6, 11, 15, 20, and 24) to 16 nM (lanes 5, 9, 14, 18, 23, and 27).

increased DNA-binding affinity of the S25Y variants. To test this model we constructed an S25F amino acid substitution in **a**1. If the changes in conformation by the S25Y substitution cause an increase in DNA-binding affinity, then **a**1-S25F should bind as well as the **a**1-S25Y variant, since the phenylalanine residue would provide similar hydrophobic packing as the tyrosine at this position. However, if the increase in binding affinity is due to the DNA contact by the hydroxyl group of the tyrosine residue, then the **a**1-S25F variant should bind DNA with an affinity similar to wild-type **a**1. The **a**1-S25F variant has roughly the same DNA-binding affinity as the **a**1-S25Y protein (Figure 4(b)). This suggests that the increased DNA-binding affinity demonstrated by the **a**1-S25Y and **a**1-Q24R/S25Y proteins is due to differences in the folding of the protein, and not simply because of the gain of a phosphate contact by the tyrosine substitution.

Although the Q24R/S25Y substitution alters the folding of **a**1 so that it can bind DNA with higher affinity and the NMR titrations suggest that α2/**a**1 heterodimers can form, it is possible that these amino acid substitutions in the loop block the ability of the mutant protein to cooperatively bind DNA in combination with α2. We therefore performed EMSAs in the presence of a constant amount of α2 (Figure 5). The **a**1-Q24R/S25Y protein is able to bind DNA cooperatively with α2 with significantly better affinity (roughly 16-fold) than wild-type **a**1. This indicates that the Q24R/S25Y substitution does not block the formation of the ternary complex and may in fact form a more optimal hydrophobic interface between **a**1 and α2. The larger, bulky side-chains should allow for better packing in this loop region, as observed in other homeodomains like engrailed.[7,13] Neither wild-type nor variant heterodimer demonstrated cooperative DNA-binding to T_7G and $T_{19}G$ mutant sites (data not shown), again demonstrating that the complex binds the DNA in a sequence-specific manner.

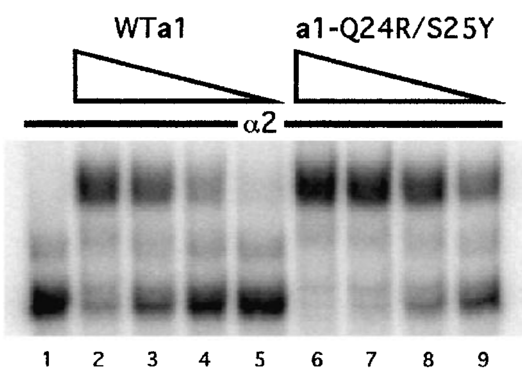


Figure 5. EMSA of wild-type **a**1 and **a**1-Q24R/S25Y binding in complex with α2. The probe used for this assay is the same as in Figure 4(a). Concentrations of **a**1 proteins range by fivefold dilutions from 0.8 μM (lanes 2, 6) to 6.4 nM (lanes 5, 9). A constant α2 concentration of 0.125 μM was used.

In vivo activity of a1 variants

Having shown that the **a**1-Q24R/S25Y protein binds DNA *in vitro* with higher affinity than wild-type **a**1, we next tested the variant protein for DNA-binding *in vivo* by chromatin immunoprecipitation (ChIP) assays using an antibody to the **a**1 protein. Plasmid-based copies of wild-type and variant *MATa1* were transformed into both *mat*Δ and *MAT*α yeast strains to assay for DNA binding by **a**1 alone and in combination with α2, respectively. Primer sets were designed to amplify: (1) an α2/**a**1 site at the distal end of the *HO* promoter, which contains a total of ten potential binding sites for α2/**a**1[16], and (2) a region of the *YDL223c* open reading frame (~10 kb from *HO*) to control for non-specific immunoprecipitation (Figure 6(a)).

As shown by the presence of a PCR product using immunoprecipitated DNA as a template, the **a**1-Q24R/S25Y mutant binds to the *HO* site in the absence of α2 (Figure 6(b)). No PCR product is

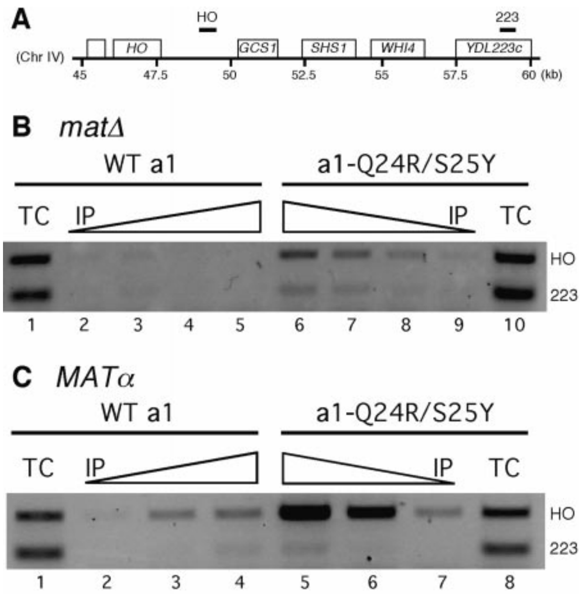


Figure 6. Chromatin immunoprecipitation of wild-type **a**1 and **a**1-Q24R/S25Y. (a) Schematic of a section of *S. cerevisiae* chromosome IV (from base-pairs 45,000 to 60,000), with the positions of PCR products indicated by small bars: HO, a 262 bp PCR product containing an α2/**a**1 site from the *HO* promoter; 223, a 121 bp PCR product from the *YDL223c* open reading frame. (b) ChIP from *mat*Δ yeast cultures transformed with plasmid-based copies of wild-type **a**1 and **a**1-Q24R/S25Y. Lanes 1 and 10 are PCR products using 1 μl of total chromatin (TC) as a template. Varying amounts of immunoprecipitated (IP) DNA were used as templates for PCR: 1 μl (lanes 2 and 9), 2 μl (lanes 3 and 8), 4 μl (lanes 4 and 7) and 8 μl (lanes 5 and 6). (c) ChIP from *MAT*α yeast cultures transformed with plasmid-based copies of wild-type **a**1 and **a**1-Q24R/S25Y. Lanes 1 and 8 are PCR products using 1 μl of TC as a template. Varying amounts of IP DNA were used as templates for PCR: 0.25 μl (lanes 2 and 7), 1 μl (lanes 3 and 6), and 4 μl (lanes 4 and 5).

observed for wild-type **a**1, suggesting that it cannot bind to the site without α2. At higher concentrations of template, a slight band can be observed for the *YDL223c* primer set for the **a**1-Q24R/S25Y samples. Since this band is not present in the wild-type samples, we do not interpret this as non-specific immunoprecipitation, but rather that the **a**1-Q24R/S25Y protein binds to potential **a**1 binding sites near the area of the *YDL223c* primer set. There is one putative **a**1 binding site (TACATC) and several degenerate sites within 500 bp (the average fragment length generated in the chromatin preparation employed) of the *YDL223c* primer set. If any of these sites are occupied by **a**1 then the fragment will be immunoprecipitated to some extent. However, the *HO* bands in Figure 6(b) (lanes 6-9) are all significantly stronger than the corresponding *YDL223c* bands, suggesting that the ChIP assay demonstrates **a**1 binding specifically to its cognate site.

In the presence of α2, the **a**1-Q24R/S25Y mutant binds to the *HO* site with at least fivefold greater affinity than wild-type **a**1 (Figure 6(c)). Very little immunoprecipitation of the non-specific fragment is observed, suggesting that most of the **a**1 protein is binding DNA in combination with α2. We therefore conclude that the **a**1-Q24R/S25Y mutant protein is binding the *HO* site *in vivo* in the absence of α2, and with significantly higher affinity than wild-type in combination with α2.

We next made the **a**1-Q24R/S25Y variant in a yeast expression vector and tested the ability of the protein to repress transcription in complex with α2 (Figure 7). In combination with α2, the **a**1-Q24R/S25Y variant represses at a level comparable to wild-type **a**1 (Figure 7(b)), demonstrating again that the mutant protein binds DNA *in vivo* in combination with α2. To test if the variant **a**1 protein confers increased *in vivo* DNA-binding affinity in complex with α2, we also assayed repression in combination with several α2 variants that show reduced α2/**a**1-mediated repression. If the **a**1-Q24R/S25Y protein can bind DNA with higher affinity than wild-type **a**1 *in vivo*, then it may be able to partially suppress the defects in repression caused by the α2 variants. Two of the α2 variants that we tested, α2-F136A and α2-W179A, have changes at positions that are buried in the hydrophobic core of the α2 homeodomain yet still contact DNA.[5] We have shown that alanine substitutions of these α2 residues decrease repression in combination with **a**1 (Figure 7).[17] In the case of both α2 variants, the **a**1-Q24R/S25Y mutant represses slightly better than wild-type **a**1, suggesting that it can partially suppress the defects of these α2 variants. Suppression of these defects in α2 by **a**1-Q24R/S25Y may indicate an increased contribution by the **a**1 variant to the overall DNA-binding affinity of the heterodimer. More importantly, the **a**1-Q24R/S25Y variant partially suppresses the defects in repression of an α2-L196S variant, a substitution that affects protein-protein interactions between α2 and **a**1.[10] The L196 residue

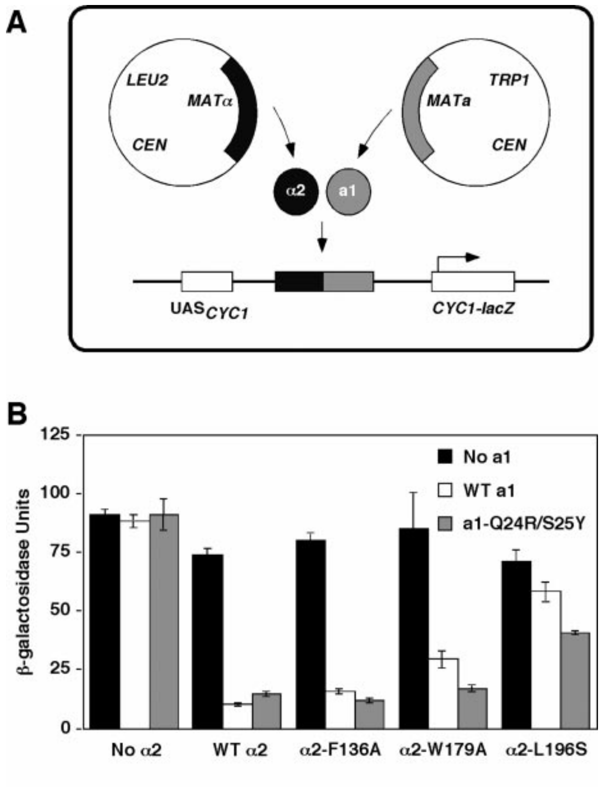


Figure 7. *In vivo* repression assays. (a) Schematic of β-galactosidase assay. Strain AJ126/79 (*mat*Δ) contains an integrated *CYC1-lacZ* reporter with an α2/**a**1 site placed in between the UAS and TATA of *CYC1*. Wild-type and mutant **a**1 and α2 proteins are expressed from *CEN* plasmids with the indicated auxotrophic markers. (b) Summary of β-galactosidase assay results. α2 mutants are listed on the *X*-axis. The average of β-galactosidase expression (in Miller units, min^{-1} ml^{-1}) for three independent transformants is shown on the *Y*-axis. Black bars indicate the level of expression in the absence of **a**1. White bars indicate the level of expression in the presence of wild-type **a**1, and gray bars indicate expression levels in the presence of the **a**1-Q24R/S25Y mutant.

is in the C-terminal tail of the α2 protein, which folds into a short α-helix that interacts with **a**1, thereby altering the conformation of the **a**1 homeodomain.[5,12] Suppression of the α2-L196S variant suggests that the **a**1-Q24R/S25Y variant has already undergone this conformational change and has increased DNA-binding affinity *in vivo*.

Discussion

To convert **a**1 from a weak binding to a strong DNA-binding protein, substitutions were made in the loop connecting helix-1 and helix-2 in the homeodomain. These substitutions are not conservative, but instead, involve relatively significant changes in side-chain bulk and charge. At position 24, glutamine is replaced by arginine, which is a charged, bulky residue instead of a neutral, relatively compact one. At position 25, a short, polar serine is replaced by an aromatic residue, either phenylalanine or tyrosine. However, the protein fold is able to accommodate these changes and the variant proteins are stable and well-folded. Furthermore, comparison of the HSQC spectra of the wild-type protein to the **a**1 variants suggests that the variants are less flexible based on the reduced number of doubled or broadened NMR resonances observed.

Although the mutational changes did not decrease protein stability, NMR chemical shift changes suggest conformational changes in the loop-1 region and the DNA-recognition helix (helix-3). Undoubtedly changes in the loop-1 region result from repacking due to the mutations. Conformational changes in helix-3, however, confirm that the DNA-binding helix and loop-1 are linked structurally. NMR data show that **a**1-Q24R/S25Y has at least one more full turn of helix-3 compared to wild-type **a**1.

To examine the effects of the loop-1 mutations on α2 cofactor binding, we titrated the **a**1 variants with α2 tail peptide. No measurable differences in dissociation constants were observed compared to wild-type α2-peptide/**a**1 heterodimer formation, suggesting that the key functional features of the hydrophobic protein-protein interaction surface are still intact. Surprisingly, fewer changes in chemical shifts upon peptide binding were observed throughout the **a**1 variant structures compared to wild-type **a**1. Importantly, no significant chemical shift changes were observed in the DNA-binding helix of **a**1-Q24R/S25Y upon α2 tail peptide binding. These observations suggest that the mutations effect the same conformational changes as those that occur when the natural cofactor, α2, binds the **a**1 homeodomain. In other words, the loop-1 mutations have triggered conformational changes that result in the extension of the DNA-binding helix, independent of α2 cofactor binding.

Although NMR studies suggest that loop-1 mutations effect conformational changes at the DNA-binding surface, it was necessary to describe the effects of the mutations on **a**1 DNA-binding function. EMSAs clearly demonstrate that the variant **a**1 proteins have enhanced DNA-binding affinity and retain their specificity for target DNA sequences in the absence of α2. In fact, the **a**1-Q24R/S25Y protein binds DNA with 16-fold greater affinity than wild-type **a**1. This enhanced **a**1 monomer affinity translates to increased DNA-binding affinity in complex with α2, showing that the mutations in loop-1 do not interfere with cooperative binding. Thus, the conformational changes noted in the **a**1 variants, i.e. the repacking of loop-1 and extension of helix-3, serve to enhance DNA-binding affinity.

Experiments were next done to test the DNA-binding activity of the **a**1-Q24R/S25Y protein *in vivo*. Chromatin immunoprecipitation assays show that **a**1-Q24R/S25Y binds the *HO* promoter in the absence of α2 and with fivefold greater affinity than wild-type **a**1 in combination with α2. Furthermore, *in vivo* repression assays demonstrate that **a**1-Q24R/S25Y binds cooperatively with α2 and is able to suppress defects in a series of α2 mutants. In particular, **a**1-Q24R/S25Y partially suppresses the defects in repression of α2-L196S, which contains a mutation in the short helix that contacts **a**1. Therefore the loop-1 mutations in the **a**1 homeodomain are able to partially compensate for a defective α2 cofactor in effecting repression *in vivo*. In summary, the engineered changes in amino acid identity at positions 24 and 25 in the **a**1 homeodomain, distant from the DNA recognition helix, have triggered conformational changes that enhance DNA-binding *in vitro* and biological function *in vivo*.

Some details of our studies of the **a**1 homeodomain illustrate the idea that a myriad of compensatory forces combine to direct macromolecular recognition events. In particular, key functional residues that drive recognition events do not necessarily form the hydrogen bonds or salt bridges observed at protein-protein interfaces or substrate surfaces. The mutations that we engineered at loop-1 in the **a**1 homeodomain likely alter the surface terrain or properties of the hydrophobic patch that binds the α2 cofactor.[5] Nonetheless, we observed tight and specific heterodimer and ternary complex formation for the **a**1 variants tested. This case illustrates that exposed non-polar, but adaptive, surfaces can be used[18,19] by a protein to enhance its function in cooperation with a particular cofactor. Both the **a**1-S25Y and **a**1-S25F proteins have enhanced DNA-binding affinity compared to wild-type **a**1, suggesting that an important functional contribution of position 25 is to drive packing and folding, rather than the ability to make polar contacts with the DNA substrate. The folding induced by an aromatic residue at position 25 enhances the DNA-binding affinity of the homeodomain. When the wild-type serine is present at position 25 the ability to make these contacts is lost by the **a**1 protein, therefore requiring cofactor to drive the conformational changes that lead to DNA recognition and binding.

Even with high-resolution protein structures in hand, mechanisms underlying protein-protein and protein-DNA interactions are difficult to unravel. Cellular life necessarily requires exquisite spatial and temporal control of protein function, which is often controlled by requisite interactions with cofactor proteins expressed at the same time and, in higher eukaryotes, the same location. The **a**1 homeodomain seems to have evolved so it folds into a fully functional form only when a cofactor is present to work cooperatively and in combination. Identification of the adaptive mechanisms used by multifunctional and multi-use proteins is key to providing the next level of understanding into how proteins network and function cooperatively and dynamically.

Materials and Methods

a1 variant engineering

a1-Q24R, **a**1-S25Y and **a**1-Q24R/S25Y variants were constructed using the polymerase chain reaction-splicing overlap extension (PCR-SOE) method.[20] A mutagenic primer was designed in both the forward and reverse direction with a unique *Eco*RI site for each mutant as described.[21] The pCW/**a**1$_{66\text{-}126}$ plasmid[12] was used as a template in the initial amplifications. Products were digested with *Xba*I and *Bam*HI, and ligated to pCW.[22] Constructs were confirmed by restriction enzyme digestion and automated DNA sequencing. The mutant **a**1-S25F was isolated using single PCR mutagenesis starting with pCW/**a**1-S25Y as a DNA template. Products were digested with *Eco*RI and *Bam*HI, and ligated into pCW/**a**1-S25Y, and constructs were confirmed by automated DNA sequencing.

Peptide

HPLC-purified α2 tail peptide (TITIAPELADLLSGEPLAK) was supplied as a lyophilized powder by Anaspec. Lyophilized peptide was dissolved in 25 mM deuterated acetate (pH 4.5), 100 mM KCl, 0.01 % (w/v) sodium azide.

NMR measurements

The wild-type and variant **a**1 homeodomains were overexpressed and purified as described.[6,21] MALDI-TOF mass spectroscopic analysis confirmed that all proteins were correctly expressed and, for the U-^{15}N-labeled samples, at >99.99 % ^{15}N isotopic enrichment. NMR data were acquired at 25 °C on Bruker Avance DRX 500 and 600 MHz spectrophotometers. Uniformly labeled [U-^{15}N]protein samples were dissolved in 25 mM deuterated acetate (pH 4.5), 100 mM KCl, 0.01 % sodium azide. DSS was added as the internal chemical shift standard. Spectra were processed using Felix 2000 (Molecular Simulations, Inc.).

For assignment of NOEs, ^{1}H-^{15}N NOESY-HSQC spectra, collected with 100 ms mixing times, were used. For all ^{1}H-^{15}N HSQC and J-modulated ^{1}H-^{15}N HSQC[23] datasets, 2048 × 256 complex matrices were collected, with eight scans and recycle times of 1.5 seconds. Watergate pulses[24] were used for water suppression. The three-dimensional NOESY-HSQC experiment was recorded as a 1024 × 256 × 64 complex matrix with eight scans and a recycle time of 1.5 seconds. For all spectra an exponential line-broadening function was applied in ω_2; shifted sine bell functions were used to apodize data in ω_1.

$^{3}J_{HNH\alpha}$ coupling constants were measured using a set of ten J-modulated ^{1}H-^{15}N HSQC experiments with delay times between 0.01 and 0.215 second. Volumes were measured and plotted as a function of delay time. Non-linear least-squares fits (*Mathematica*) to the delay-dependent resonance peak volumes were used to obtain $^{3}J_{HNH\alpha}$.[23]

Electrophoretic mobility shift assays

DNA probes used in the EMSAs were synthesized by PCR using oligonucleotides that anneal to opposite sides of the α2/**a**1 site of pYJ103 and its derivatives[10] and contain (5′) flanking *Eco*RI or *Hin*dIII sites. Following PCR amplification the 80 bp products were gel-purified, digested with *Eco*RI and *Hin*dIII, and then end-labeled with Klenow polymerase and [α-^{32}P]ATP. The labeled probes were then purified using the QIAquick Nucleotide Removal Kit (Qiagen). Each probe was diluted to around 100 cpm/μl in assay buffer (20 mM Tris (pH 7.6), 0.1 mM EDTA, 5 mM $MgCl_2$, 10 mg/ml of BSA, 5% (v/v) glycerol, 0.1% IGEPAL, 10 μg/ml of sheared salmon sperm DNA). Proteins were normalized to 0.1 mM in NMR buffer (25 mM sodium acetate (pH 4.5), 100 mM KCl, 0.1% sodium azide) and then diluted in protein dilution buffer (50 mM Tris (pH 7.6), 500 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 1 mg/ml of BSA). For each assay 40 μl of probe was mixed with 5 μl of **a**1 protein and either 5 μl of protein dilution buffer (for EMSA of **a**1 alone) or 5 μl of $\alpha 2_{128\text{-}210}$ protein (for cooperative EMSA).[25] The assay mix was incubated at room temperature for 2.5 hours, then electrophoresed on a 6% (w/v) polyacrylamide gel in 0.5 × TBE. Gels were dried and exposed to phosphorimager plates overnight, then developed using a Molecular Dynamics phosphorimager. Gels were quantified using IPLabGel H software.

In vivo repression assays

Plasmid pYJ210, a *CEN*, *LEU2* plasmid containing a 4.3 kb *MAT**a*** fragment with a cassetted version of *MAT**a**1* has been described.[8] To clone *MAT**a*** into a plasmid with a *TRP1* marker, a 6 kb *Bgl*I fragment (containing *MAT**a***) of pYJ210 was subcloned into a 3 kb *Bgl*I fragment (containing *TRP1*) of pRS414, making pJR064. Constructs were confirmed by restriction digest analysis. Oligonucleotide-directed mutagenesis of *MAT**a**1* in both pYJ210 and pJR064 was done using the QuikChange Kit (Stratagene). Mutations (and full gene sequences) were confirmed by automated sequencing. Mutations in *MAT*α*2* were made in pAV115 or pJM130 (*CEN LEU2*) and were described previously.[10,17,26] Plasmids were transformed into AJ126/79, a *mat*Δ *ura3 trp1 leu2 his4* yeast strain with a *lacZ*-reporter containing an α2/**a**1 site.[27] β-Galactosidase assays were performed as described.[2]

Chromatin immunoprecipitation (ChIP) assays

Strains AJ126/79 (*mat*Δ) and AJ82 (*MAT*Δ) were transformed with plasmids bearing wild-type and mutant copies of *MAT**a**1*. These strains were grown in selective media to an A_{600} of ~0.5. Formaldehyde (37%) was then added to a final concentration of 1%, and the cultures were fixed for 20 minutes at room temperature under gentle agitation. Sheared chromatin was then isolated as described by Kuo *et al*.[28] with the following minor modifications. Briefly, the cells were lysed by glass bead breakage, then sonicated six times for ten seconds at each interval. The sonicated lysate was then clarified by centrifugation and brought to 1 ml with lysis buffer: 100 μl of each sample was frozen at −20 °C to be used as a total chromatin (TC) fraction. The remaining 900 μl of each sample was then pretreated with 50 μl of recombinant protein G (rProtG) agarose beads (Gibco) for one hour at 4 °C. Following centrifugation, 2.5 μl of rabbit antiserum to Mat**a**1p (a gift from Sandy Johnson) was added to each supernatant and then incubated overnight at 4 °C: 50 μl of rProtG-Agarose beads was added to each sample, followed by incubation at 4 °C for two hours. The beads were washed and the immunoprecipitate (IP) fraction was eluted from the beads.[28] For both the IP and TC fractions, crosslinks were reverted by addition of 5 M NaCl to a concentration of 0.2 M, followed by incubation at 65 °C for four hours. Samples were then treated with proteinase K, phenol/chloroform-extracted and ethanol-precipitated as described by Kuo *et al*.[28] Samples were dried and then resuspended in distilled water, with 50 μl for IP samples and 500 μl for TC samples.

Primers were synthesized to amplify a 262 bp fragment (base pairs 49,639 to 49,901 of *S. cerevisiae* chromosome IV) of the *HO* promoter, containing an α2/**a**1 site.[16] A second primer set that generates a 121 bp segment (base-pairs 59,244 to 59,365 of chromosome IV, containing no putative binding sites for **a**1) of the *YDL223c* open reading frame was designed to control for non-specific immunoprecipitation with the Mat**a**1 antibody. PCR were then performed using the IP and TC samples as templates. PCR cycles were as follows: 94 °C for 1.5 minutes (1×); 94 °C for 30 seconds, 51 °C for one minute, 72 °C for 30 seconds (25×); 72 °C for ten minutes (1×). PCRs were then electrophoresed on 2% (w/v) agarose/TAE. Bands were quantified using IPLabGelH software.

Acknowledgments

We thank Jacquelyn Fetrow for helpful discussions on protein folding and packing, and Griselda Hernandez for scientific support. We thank James Kirkwood for preliminary work with the α2 tail peptide titrations. We acknowledge Nicole Nall for technical assistance and the Wadsworth Center Molecular Genetics and Mass Spectroscopy Cores for analytical assistance. We thank Gary Pielak and Ashutosh Tripathy of the University of North Carolina Macromolecular Interactions Facility for calorimetric measurements. We thank the Wadsworth Center for support and use of the NMR Structural Biology Core. This work was supported by grants from the National Institute of General Medical Sciences (GM55361) to S.M.B. and (GM49265) to A.K.V.

References

1. Johnson, A. D. (1992). A combinatorial regulatory circuit in budding yeast. In *Transcriptional Regulation* (McKnight, S. L. & Yamamoto, K. R., ed.), pp. 975-1006, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Keleher, C. A., Goutte, C. & Johnson, A. D. (1988). The yeast cell-type-specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. *Cell*, **53**, 927-936.
3. Vershon, A. K. & Johnson, A. D. (1993). A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. *Cell*, **72**, 105-112.
4. Goutte, C. & Johnson, A. D. (1993). Yeast a1 and alpha 2 homeodomain proteins form a DNA-binding activity with properties distinct from those of either protein. *J. Mol. Biol.* **233**, 359-71.
5. Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. (1995). Crystal structure of the MAT**a**1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262-269.
6. Phillips, C. L., Stark, M. R., Johnson, A. D. & Dahlquist, F. W. (1994). Heterodimerization of the

yeast homeodomain transcriptional regulators alpha 2 and **a**1 induces an interfacial helix in alpha 2. *Biochemistry*, **33**, 9294-9302.
7. Anderson, J. S., Forman, M. D., Modleski, S., Dahlquist, F. W. & Baxter, S. M. (2000). Cooperative ordering in homeodomain-DNA recognition: solution structure and dynamics of the MAT**a**1 homeodomain. *Biochemistry*, **39**, 10045-10054.
8. Jin, Y., Zhong, H. & Vershon, A. K. (1999). The yeast **a**1 and alpha2 homeodomain proteins do not contribute equally to heterodimeric DNA binding. *Mol. Cell Biol.* **19**, 585-593.
9. Stark, M. R., Escher, D. & Johnson, A. D. (1999). A trans-acting peptide activates the yeast **a**1 repressor by raising its DNA-binding affinity. *EMBO J.* **18**, 1621-1629.
10. Vershon, A. K., Jin, Y. & Johnson, A. D. (1995). A homeo domain protein lacking specific side chains of helix 3 can still bind DNA and direct transcriptional repression. *Genes Dev.* **9**, 182-192.
11. Gehring, W. J., Affolter, M. & Burglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487-526.
12. Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F. & Dahlquist, F. W. (1994). Heterodimerization of the yeast homeodomain transcriptional regulators alpha 2 and **a**1: secondary structure determination of the a1 homeodomain and changes produced by alpha 2 interactions. *Biochemistry*, **33**, 15309-15320.
13. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579-590.
14. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, **67**, 517-528.
15. Justice, M. C., Hogan, B. P. & Vershon, A. K. (1997). Homeodomain-DNA interactions of the Pho2 protein are promoter-dependent. *Nucl. Acids Res.* **25**, 4730-4739.
16. Miller, A. M., MacKay, V. L. & Nasmyth, K. A. (1985). Identification and comparison of two sequence elements that confer cell- type specific transcription in yeast. *Nature*, **314**, 598-603.
17. Mathias, J. R., Zhong, H., Jin, Y. & Vershon, A. K. (2001). Altering the DNA-binding specificity of the yeast MAT alpha 2 homeodomain protein. *J. Biol. Chem.* **276**, 32696-32703.
18. Burz, D. S. & Hanes, S. D. (2001). Isolation of mutations that disrupt cooperative DNA binding by the *Drosophila* bicoid protein. *J. Mol. Biol.* **305**, 219-230.
19. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000). Convergent solutions to binding at a protein-protein interface. *Science*, **287**, 1279-1283.
20. Horton, R. M. (1997). In vitro recombination and mutagenesis of DNA. SOEing together tailor-made genes. *Methods Mol. Biol.* **67**, 141-149.
21. Forman, M. D., Stack, R. F., Masters, P. S., Hauer, C. R. & Baxter, S. M. (1998). High level, context dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in *Escherichia coli*. *Protein Sci.* **7**, 500-503.
22. Gegner, J. A. & Dahlquist, F. W. (1991). Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. *Proc. Natl Acad. Sci. USA*, **88**, 750-754.
23. Billeter, M., Neri, D., Otting, G., Qian, Y. Q. & Wuthrich, K. (1992). Precise vicinal coupling constants 3_{JHN} alpha in proteins from nonlinear fits of J-modulated [15N,1H]-COSY experiments. *J. Biomol. NMR*, **2**, 257-274.
24. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661-665.
25. Phillips, C. L., Vershon, A. K., Johnson, A. D. & Dahlquist, F. W. (1991). Secondary structure of the homeo domain of yeast alpha 2 repressor determined by NMR spectroscopy. *Genes Dev.* **5**, 764-772.
26. Mead, J., Zhong, H., Acton, T. B. & Vershon, A. K. (1996). The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. *Mol. Cell Biol.* **16**, 2135-2143.
27. Mak, A. & Johnson, A. D. (1993). The carboxy-terminal tail of the homeo domain protein alpha 2 is required for function with a second homeo domain protein. *Genes Dev.* **7**, 1862-1870.
28. Kuo, M. H. & Allis, C. D. (1999). *In vivo* cross-linking and immunoprecipitation for studying dynamic protein:DNA associations in a chromatin environment. *Methods*, **19**, 425-433.
29. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Edited by P. E. Wright

(Received 21 September 2001; received in revised form 5 December 2001; accepted 7 December 2001)