6-2001

Structural and functional analysis of mutant MATal homeodomains by multidimensional nmr spectroscopy

David F. Ott
Union College - Schenectady, NY

STRUCTURAL AND FUNCTIONAL ANALYSIS OF MUTANT MATα1 HOMEODOMAINS BY MULTIDIMENSIONAL NMR SPECTROSCOPY

By

David F. Ott

****

Submitted in partial fulfillment
of the requirements for
Honors in the Department of Chemistry

UNION COLLEGE

June, 2001

ABSTRACT

OTT, DAVID Structural and Functional Analysis of Mutant MAT**a1** Homeodomains by Multidimensional NMR Spectroscopy. Department of Chemistry, June 2001.

Homeodomain proteins are transcription factors that contain a conserved 60-residue sequence, beginning with an N-terminal unstructured arm, followed by an alpha helix, a loop, and a helix-turn-helix. The yeast protein MAT**a1** is unusual among homeodomains in that, as a monomer, it binds very poorly to its DNA operator. However, the **a1**-α2 heterodimer binds to the *hsg* operator with 3000 times the affinity it has for nonspecific DNA. Studies have shown that most of the heterodimer's binding specificity is due to **a1** rather than α2 (1,2).

To identify the structural changes that transform **a1** into a strong, sequence-specific DNA binding protein, a single-point mutant (s25y) and a double-point mutant (q24r/s25y) were studied. EMSA studies showed that both mutants bind to DNA with greater affinity than wild type **a1** does. Analysis of 2-D ^{15}N-HSQC and 3-D ^{15}N-NOESY spectra showed that significant changes in the chemical shifts of the backbone amide groups of loop 1 and helix 3 occur upon mutation. The NOESY spectrum was also used to identify NOEs between amide protons of sequential residues, indicating where alpha helical conformations occurred. The NOEs showed that the third helix is extended in the **a1** mutants. Finally, titration experiments were performed by adding aliquots of the 19-residue α2 tail peptide to **a1** and to each **a1** mutant, and then recording HSQC spectra. These showed that chemical shift changes which occur in wild type **a1** upon α2 tail binding are diminished in the **a1** mutants.

ABSTRACT

OTT, DAVID Structural and Functional Analysis of Mutant MAT**a1** Homeodomains by Multidimensional NMR Spectroscopy. Department of Chemistry, June 2001.

Homeodomain proteins are transcription factors that contain a conserved 60-residue sequence, beginning with an N-terminal unstructured arm, followed by an alpha helix, a loop, and a helix-turn-helix. The yeast protein MAT**a1** is unusual among homeodomains in that, as a monomer, it binds very poorly to its DNA operator. However, the **a1**-α2 heterodimer binds to the *hsg* operator with 3000 times the affinity it has for nonspecific DNA. Studies have shown that most of the heterodimer's binding specificity is due to **a1** rather than α2 (1,2).

To identify the structural changes that transform **a1** into a strong, sequence-specific DNA binding protein, a single-point mutant (s25y) and a double-point mutant (q24r/s25y) were studied. EMSA studies showed that both mutants bind to DNA with greater affinity than wild type **a1** does. Analysis of 2-D ^{15}N-HSQC and 3-D ^{15}N-NOESY spectra showed that significant changes in the chemical shifts of the backbone amide groups of loop 1 and helix 3 occur upon mutation. The NOESY spectrum was also used to identify NOEs between amide protons of sequential residues, indicating where alpha helical conformations occurred. The NOEs showed that the third helix is extended in the **a1** mutants. Finally, titration experiments were performed by adding aliquots of the 19-residue α2 tail peptide to **a1** and to each **a1** mutant, and then recording HSQC spectra. These showed that chemical shift changes which occur in wild type **a1** upon α2 tail binding are diminished in the **a1** mutants.

ACKNOWLEDGEMENTS

I would like to recognize the following people for their contributions to this research: Susan Baxter for initiating and coordinating the project; Nicole Nall and Beverly Hart for helping to make the proteins studied; and Jon Mathias and A. K. Vershon for performing the EMSA studies. I would especially like to thank Professor Janet Anderson for her support, direction, and patience throughout the past year.

TABLE OF CONTENTS

Section Page

TABLE OF FIGURES

INTRODUCTION

Gene expression determines the size, shape, function, and behavior of cells. A cell can exercise control over the proteins it assembles anytime before, during, or after their construction. The favored method of regulating gene expression is to control which genes are transcribed into mRNA. The vast majority of genes in both prokaryotic and eukaryotic cells have gene regulatory sequences that are necessary to activate or deactivate transcription. Gene regulatory proteins recognize these sequences, bind to them, and switch the gene on or off (3).

The homeodomain is a structural motif found in many gene regulatory proteins that are present in eukaryotic cells from yeast to human. The homeodomain originates from the homeotic, or master control, genes that specify body plan and control growth of higher organisms. The homeotic genes contain a common sequence of 180 base pairs, the homeobox, which encodes the 60-residue homeodomain. This conserved 60-residue sequence is the DNA-binding domain of significantly larger gene regulatory proteins (4). The homeodomain begins with an N-terminal unstructured arm, followed by an alpha helix, a loop, and a helix-turn-helix. Helices 1 and 2 are anti-parallel to one another, while helix 3 is aligned essentially perpendicular to the first two. The third helix, the DNA recognition helix, makes the majority of the DNA contacts in the major groove via hydrogen bonds, salt bridges, and/or van der Waals interactions. The N-terminal arm of most homeodomains wraps around and makes additional favorable contacts with bases in the adjacent minor groove. Tight and specific binding of DNA by the homeodomain, as a result of these interactions, leads to regulation of transcription.

The similarity in the three-dimensional structure of homeodomains stems from a tightly packed, highly conserved core of 11 amino acids. All except one of these preserved amino acids, which are spread among residues 13 through 48, are hydrophobic (4). Given that a protein's hydrophobic core plays an important role in the determination of its three-dimensional structure, and given the fact that this core is nearly invariant among homeodomains, one can expect that various homeodomains would fold in a similar manner. It should be noted that, outside of the structurally conserved 60-residue homeodomain, DNA-binding proteins, even from the same family, have vastly different amino acid sequences and structures. These flanking structures give the different polypeptides in the homeodomain superfamily the ability to regulate a myriad of different genes in numerous types of cells.

Regulation of transcription in eukaryotic cells often requires the concerted effort of several proteins. In this process, called combinatorial control, many different polypeptides bind DNA in a sequence-specific manner and regulate transcription of a downstream gene. The cluster of polypeptides involved in combinatorial control includes sequence-specific DNA-binding proteins and additional proteins involved in protein-protein interactions with the DNA-bound transcription factors. It is thought that regulation of a single gene by more than one protein has evolved as a way to incorporate responses to a wide variety of stimuli using a limited number of proteins (5).

Two homeodomain proteins that act in tandem to regulate transcription and to determine cell type in the yeast *Saccharomyces cerevisiae* are MAT**a1** and MATα2. MAT**a1** is a 126-residue DNA binding protein with a homeodomain motif at its C-terminal end, residues 66 to 126. (In this study of **a1**, the 61 residues that compose the

homeodomain motif have been renumbered –3 through 57 to coincide with the numbering systems of other homeodomains.) Thus far, no adjacent, flanking sequences that engage in protein-protein interactions have been identified in **a1**. MATα2 is a 210-residue DNA binding protein with a homeodomain motif at residues 128 to 189 and a 21-residue C-terminal tail, residues 190 to 210. Flanking sequences adjacent to the homeodomain motif in α2 have been identified and their functions have been explored. In particular, the C-terminal tail is the flanking sequence that interacts with **a1** and facilitates combinatorial control in diploid *Saccharomyces cerevisiae* (6).

There are three cell types of *S. cerevisiae*: haploid cell types **a** and α, and diploid cell type **a**-α. Control of *S. cerevisiae* cell type begins with the MAT allele. In haploid α cells the MAT allele encodes for the α2 homeodomain. The α2 homeodomain, in haploid α cells and in diploid **a**-α cells, recruits the MCM1 protein through interactions with an N-terminal flanking sequence, binds DNA, and represses transcription of **a**-specific genes. In haploid **a** cells the MAT allele encodes for the **a1** homeodomain. At this point it is not known what function the **a1** protein has in **a** cells.

The MAT allele in diploid *S. cerevisiae* encodes different mRNA for both the **a1** and α2 homeodomains. The **a1** and α2 homeodomains combine to form a heterodimer in diploid yeast cells. The **a1**-α2 heterodimer binds tightly and specifically to the haploid-specific gene (*hsg*) operator, recruits the SSN6-TUP1 complex, and together this protein assembly represses transcription of the haploid-specific genes (7). The **a1**-α2-DNA ternary complex is depicted in Figure 1 (6). As illustrated, both **a1** (shown in shades of blue) and α2 (shown in shades of green) have the helix-loop-helix-turn-helix structure characteristic of homeodomains. The yeast **a1** protein is unique among well-studied

homeodomains in that, as a monomer, it does not bind DNA with any sequence specificity (8,9). However, upon combination with α2 the resulting **a1**-α2 heterodimer binds the *hsg* operator with a specificity of 3000 times that of nonspecific DNA (8).

Recently, Anderson et al. (10) have shown that ternary complex formation induces conformational changes in **a1** in loop 1, where the C-terminal tail of α2 binds to **a1**, and at the C-terminal end of helix 3, where **a1** contacts DNA. Figure 2 presents the

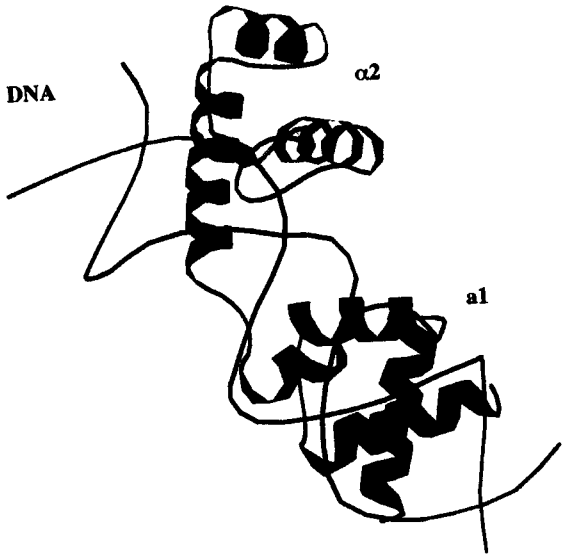

Figure 1 X-ray structure of the a1-α2-DNA ternary complex (6). Residues 128 to 205 of α2 are shown in shades of green, with the C-terminal tail shown in red contacting **a1**. Residues 77 to 125 of **a1** are shown in shades of blue.

helical axes of free **a1** (light blue) overlaid with the helical axes of **a1** (dark blue) in the ternary complex. The figure shows that the angle between helix 1 and helix 2 is significantly larger for free **a1** than it is for **a1** in the **a1**-α2-DNA ternary complex. (The angles between helix 1 and helix 2 in other homeodomains are closer to the corresponding angle of **a1** in the ternary complex than they are to the helix 1–helix 2 angle in free **a1**.) Furthermore, the packing of hydrophobic residues in loop 1 is not as favorable in free **a1** as it is in **a1** bound in the **a1**-α2-DNA ternary complex. These differences, along with the fact that the α2 tail contacts **a1** in the loop 1 region, suggest that the loop 1 region, at least partially, controls DNA binding affinity in homeodomains.

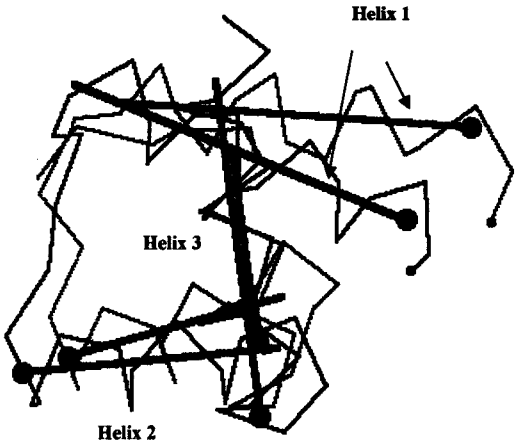

Figure 2 Free a1 (light blue with green axes) and a1 in the a1-α2-DNA ternary complex (dark blue with red axes) (10).

Comparison of **a1** to the engrailed family of homeodomains shows that **a1** differs in loop 1. Figure 3 presents the primary structures of the **a1**, engrailed (en), and antennapedia (Antp) homeodomains. In loop 1, **a1** has a glutamine at residue 24 instead of arginine, and a serine at residue 25 instead of tyrosine. Because glutamine and serine have less nonpolar surface area than arginine and tyrosine, respectively, the conserved phenylalanine at residue 20 does not have the same ability to pack against nonpolar groups in **a1** that it does with other homeodomains. This packing deficiency may lead to

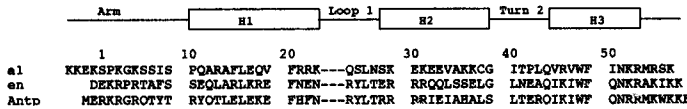

Figure 3 Primary sequences of a1, engrailed (en), and antennapedia (Antp) homeodomains. (engrailed and antennapedia are taken from drosophila melanogaster.)

the difference in angle between helix 1 and helix 2 in **a1**, and to **a1**'s unusually low affinity for the *hsg* operator. This study examines the effects on structure and function (DNA binding, heterodimerization, and ternary complex formation) of mutating residue 24 from glutamine to arginine and residue 25 from serine to tyrosine.

To analyze the structural changes elicited by the point mutations, Nuclear Magnetic Resonance (NMR) spectroscopy was employed. Structural analysis using NMR stems from the fact that some nuclei behave as if they were spinning. These spinning charges create small magnetic fields, termed nuclear spins. In isolation these spin states are degenerate; however, quantum mechanics requires that, in the presence of

an external magnetic field, those nuclei with spin quantum number I = 1/2 must be in one of two, unequal, spin states. The resonance energy is the difference between the energy of a nucleus whose spin state opposes the external magnetic field and the energy of a nucleus whose spin state is parallel to the field. The resonance energy depends on the strength of the magnetic field, B_0, to which it is subjected:

$$\Delta E = \gamma h B_0 / 2\pi$$

where h is Planck's constant and γ is the gyromagnetic ratio, a constant that depends on the identity of the nucleus. The dependence of the resonance energy on the magnetic field strength illustrates why NMR lends itself to structure determination: the magnetic field "felt" by a nucleus depends not only on the applied magnetic field, but also on the magnetic fields created by all of the nuclear spins in its immediate vicinity. Detection of the interactions among the magnetizations of nuclei in a protein, via measurement of the differences in resonance energies, leads to structure elucidation.

Once a sample of nuclei absorbs a pulse of radiofrequency radiation corresponding to their resonance energy, relaxation back to thermal equilibrium occurs. This relaxation takes place by two distinct mechanisms: spin-lattice, or longitudinal, relaxation, with relaxation time T_1; and spin-spin, or transverse, relaxation, with relaxation time T_2. The lattice is the surroundings of the nucleus, including other molecules in solution and the rest of the molecule containing the nucleus in question. The spin-lattice time, T_1, is the time it takes the irradiated system to get back to its unperturbed magnetic state via energy loss to its surroundings. The transverse relaxation time, T_2, is a measure of the adiabatic relaxation by which the perturbed nuclei distribute their absorbed energy to the other nuclei in the immediate surrounding area.

Both types of relaxation provide structural information. The NMR linewidths, the measure of an NMR signal at half-height, are inversely proportional to the transverse relaxation time, T_2, which decreases with increasing molecular size. This allows NMR peaks to be identified by shape as nuclei belonging to protein residues in certain structural motifs: broad peaks correspond to residues in turns and loops; thin, sharp peaks correspond to unstructured, flexible residues; and medium width peaks stem from structured residues, such as those in alpha helices. More important than peak shape are the results of dipole-dipole interactions that lead to spin-lattice relaxation. These dipole-dipole interactions only occur between nuclei at close distances to each other, and identifying them permits distance constraints to be assigned to protein residues.

Distance constraints, established through the measurement of nuclear Overhauser effects (NOEs), are the primary tool for protein structure determination. Through space, as opposed to through bond, nuclear dipole-dipole coupling leads to the NOE and T_1 relaxation. The strength of the NOE can be related to the distance between nuclei:

$$NOE \propto 1/r^6$$

Only nuclei in close proximity to each other will show an NOE, because the strength of the NOE is inversely proportional to the sixth power of the distance between nuclei. Indeed, an NOE will only be observed if two nuclei are less than 5 Å apart. The strengths of the NOE crosspeaks in multidimensional NMR are classified as strong (1.8-2.7 Å), medium (1.8-3.5 Å), and weak (1.8-5.0 Å). This aids protein structure determination because certain protons in residues in structural motifs exist at a characteristic distance from each other. Since the MAT homeodomains contain three alpha helices, the distances between protons in alpha helices are of particular interest. Specifically, the

distance between amide protons on sequential residues, denoted d_{NN}, in a regular α helix is 2.8 Å, corresponding to a medium NOE (11). The distances between sequential amide and alpha protons, $d_{\alpha N}$, and sequential amide and beta protons, $d_{\beta N}$, in a regular alpha helix are 3.5 Å and 2.5 to 4.1 Å, respectively (11). The sizes of the NOE crosspeaks in multidimensional, through space NMR experiments permit distance constraints between protons participating in dipolar-dipolar relaxation to be established; and, thus, makes structure determination via NMR a reality.

NMR spectroscopy was used to elucidate the structural changes that transform free **a1** into a strong, sequence-specific DNA binding protein. A single-point mutant (s25y) and a double-point mutant (q24r/s25y) were studied with heteronuclear single-quantum coherence (^{15}N-HSQC) and nuclear Overhauser effect spectroscopy (^{15}N-NOESY). The NOEs between amide protons confirmed that the characteristic homeodomain secondary structural elements are conserved in the **a1** mutants. Furthermore, these NOEs illustrated that the third helix is extended by 1 and 4 residues in s25y and q24r/s25y, respectively. Titration experiments (adding aliquots of α2 tail to **a1** and the **a1** mutants) showed that the conformational changes in **a1** caused by the binding of α2 are diminished in the **a1** mutants. The implication is that the mutants are already in the optimal conformation, normally brought about by heterodimerization, for DNA binding. EMSA studies (12) showed that 1) both s25y and q24r/s25y have significantly greater affinity for the **a1** consensus binding sites than does wild type **a1** and 2) both the mutants form the **a1**-α2 heterodimer more readily than does wild type **a1**. These results emphasize the importance of the loop 1 region in homeodomain binding to DNA.

MATERIALS AND METHODS

Analysis of ^{15}N-HSQC and ^{15}N-NOESY spectra was performed using NMRView®. The HSQC spectra of the mutant homeodomains were first compared with that of wild type **a1**. Some of the "peaks" were identified – identification means assigning the peak to the amide group of a specific residue – simply through this comparison, because the chemical shifts of some of the amide groups were virtually the same in both wild type **a1** and the **a1** mutants. The HSQC spectrum of unbound s25y is seen in Figure 4.

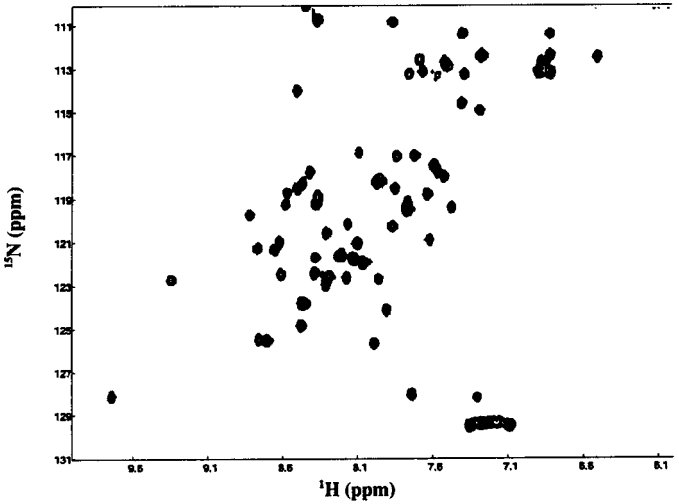

Figure 4 HSQC of s25y. The x-axis is the proton dimension and the y-axis is the nitrogen dimension. Every spot, or "peak", represents an amide group in the protein.

When peak identification could not be made by visual comparison between wild type and mutant **a1** HSQC spectra, the 3-D NOESY spectrum was used. The 3-D NOESY spectrum contains a second proton dimension and allows through-space connectivities to be ascertained. A peak, corresponding to a peak that could not be identified in the HSQC spectrum, was located in the NOESY spectrum. (Since each peak represents the same amide group they will have the same proton and nitrogen chemical shifts in both HSQC and NOESY spectra). These on-diagonal NOESY peaks had two equivalent proton dimensions. Relevant off-diagonal peaks, or crosspeaks, were then identified. The nitrogen and one of the proton dimensions of these crosspeaks were the same as those of the diagonal peak. The second proton dimension in the relevant crosspeaks corresponded to the chemical shift of an amide proton of an adjacent residue or to an α proton in the same residue. (A crosspeak is referred to as a NOE, because it results from the nuclear Overhauser effect.) Because the chemical shifts of the α protons did not change significantly upon mutation, these NOEs were used to help identify the diagonal peak. If the residues were in an alpha helix, NOEs between amide protons of adjacent residues were seen. Establishing which residues were adjacent to the unknown diagonal peak allowed the peak to be identified. Furthermore, the presence of these medium range NOEs indicated that the peak corresponded to a residue that was indeed part of an alpha helix.

HSQC spectra were taken as aliquots of α2 tail were added to s25y and q24r/s25y. Since the α2 tail had no ^{15}N labeling, no peaks for the tail are seen in the HSQCs. The peaks in the titration HSQC spectra are the weighted averages of the chemical shifts of the amide groups in the free mutants and in the mutant-α2 tail

heterodimer. Thus, by overlaying the spectra, we can see that the peaks shift along a straight path from the initial state, the free homeodomain, to the final state, the homeodomain-α2 heterodimer. Tracking the movement of the peaks in the HSQC spectra allowed the chemical shifts of the amide groups in the α2 tail-bound mutants to be determined. Figure 5 is an enlarged, overlaid section of the titration HSQCs for q24r/s25y. Each peak represents the amide group of residue K29. The red peak at the left represents K29 in free q24r/s25y, and the blue peak at the right represents K29 bound to α2 tail.

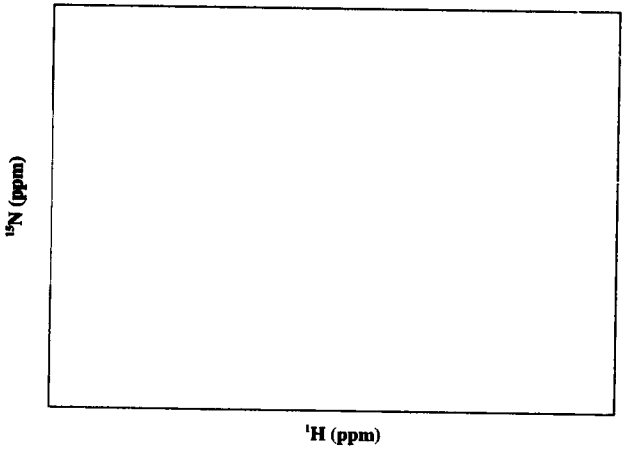

Figure 5 Overlaid HSQC spectra of q24r/s25y. All 5 peaks represent the amide group of residue K29 in q24r/s25y + increasing amounts of α2 tail. From left to right, Peak 1 (red) is completely free q24r/s25y; Peak 2 is 2:1 q24r/s25y:α2tail; Peak 3 is 1:1 q24r/s25y:α2tail; Peak 4 is 1:2 q24r/s25y:α2tail; and Peak 5 (blue) is 1:3 q24r/s25y:α2tail (representing completely bound q24r/s25y).

The next step was finding the dissociation constants for the s25y-α2 tail and q24r/s25y-α2 tail heterodimers. This was accomplished by using the chemical shift data obtained from the titration HSQCs and the following mathematical relationships. The dissociation of the heterodimer is the reverse reaction in the following equilibrium

$$\mathbf{a1} + \alpha 2 \text{ tail} \longleftrightarrow \mathbf{a1}\text{-}\alpha 2 \text{ heterodimer (HD)} \qquad (1)$$

Therefore, the dissociation constant for the heterodimer is

$$K_d = [\mathbf{a1}][\alpha 2 \text{ tail}]/[\text{HD}] \qquad (2)$$

In order to calculate the dissociation constants for the two mutant heterodimers from the chemical shifts observed in the titration data, the following were defined:

(1) ppm_o = chemical shift of proton X in free (unbound) **a1**
(2) ppm_∞ = chemical shift of proton X in **a1** with α2 tail bound
(3) ppm_{eq} = chemical shift of proton X in **a1** at a particular concentration of α2 tail

With this notation, the relationship between the bound and unbound **a1** can be represented as

$$F = (ppm_{eq} - ppm_o)/(ppm_\infty - ppm_o) \qquad (3)$$

where F is the fractional saturation of **a1**. In other words, F is the amount of **a1** in heterodimer form divided by the total amount of **a1** in solution:

$$F = [HD]/([\mathbf{a1}] + [HD]) \quad (4)$$

where [HD] is the molar concentration of heterodimer, and **[a1]** is the molar concentration of free **a1**. From equation 2 the relationship $[HD] = [\mathbf{a1}][\ \alpha 2\ \text{tail}]/K_d$ is obtained. Thus,

$$F = ([\mathbf{a1}][\ \alpha 2\ \text{tail}]/K_d)/([\mathbf{a1}] + [\mathbf{a1}][\ \alpha 2\ \text{tail}]/K_d)$$

and

$$F = [\alpha 2\ \text{tail}]/(K_d + [\alpha 2\ \text{tail}]) \quad (5)$$

The concentrations of α2 tail and the corresponding proton chemical shifts of selected residues – residues with the greatest absolute difference in backbone amide proton chemical shift between the free and α2 tail-bound states were used – were entered into a Mathematica® notebook. The chemical shift data was transformed into fractional saturation of sites. Using nonlinear regression, these values were then fitted to a nonlinear curve based on equation 5 and the K_d was obtained with a single parameter fit.

RESULTS

The HSQC spectra show that s25y and q24r/s25y are stable, folded proteins, very similar to wild type **a1**. Because the structures of s25y and q24r/s25y have not been solved, the solution structure of free **a1** solved by Anderson et al. (2000) was used to represent the mutant homeodomains in Figures 13b and c (10).

Mutation of Serine 25 to Tyrosine (s25y)

The chemical shifts of both the backbone nitrogens and amide protons in the s25y mutant homeodomain were determined (See Table 1). From these data the differences in backbone amide proton chemical shift between wild type **a1** and s25y were calculated. Figure 6 shows the solution structure of **a1** (10). The atoms depicted in space-filling format represent the residues whose backbone amide protons underwent a significant change in chemical shift (significant changes in chemical shift are >0.1 ppm) upon mutation of residue 25 from serine to tyrosine. These residues are N27, which is two residues away from the point mutation, and M54, which is at the C-terminal end of the homeodomain. The magnitudes of the differences in backbone amide proton chemical shift between s25y and wild type **a1**, for all of the homeodomain residues, are presented graphically in Figure 7.

The NOEs between amide protons of sequential residues were identified in the 3D-NOESY spectrum of s25y. The presence of a "medium" sized NOE (dependent upon the d_{NN}) will indicate that the residues are in an α-helical conformation. Figure 8 illustrates sequential NOEs for residues 44 to 57 in s25y. The arrows point from the

Table 1 s25y Backbone Nitrogen and Amide Proton Chemical Shifts

Residue	HN (ppm)	15N (ppm)
K-3	8.70	125.6
K-2	8.47	124.8
E-1	8.47	123.7
K0	8.46	123.7
S1	8.42	117.7
P2	-	-
K3	8.38	122.4
G4	8.37	110.7
K5	8.21	121.6
*S6/S7	8.38	119.2
*S7/S6	8.38	119.2
I8	8.13	121.8
S9	8.76	125.5
P10	-	-
Q11	8.47	118.3
A12	7.99	125.7
R13	8.49	118.5
*A14/K37	7.76	119.6
F15	7.61	120.9
L16	8.37	121.7
E17	8.08	116.9
Q18	7.47	119.4
V19	8.06	122.0
F20	8.76	121.3
R21	7.40	114.6
R22	7.59	117.5
K23	8.61	122.5
*Q24/N51/R53	7.95	118.2
Y25	7.28	114.9
L26	8.31	123.0
N27	9.34	122.7
S28	8.50	114.0
K29	7.90	124.1
E30	8.61	121.1
K31	8.57	118.8
E32	7.51	118.0
E33	8.10	121.1
V34	8.65	121.4
A35	8.17	122.7
K36	7.70	117.0
*K37/A14	7.76	119.6
C38	7.68	112.6
G39	7.86	110.9
I40	7.75	113.3
T41	8.46	110.2
P42	-	-
L43	7.75	119.2
Q44	7.85	118.5
V45	7.86	120.3
R46	8.31	120.6
V47	8.57	119.3
W48	8.29	122.6
F49	8.82	119.8
I50	8.16	120.2
*N51/Q24/R53	7.95	118.2
K52	7.96	122.7
*R53/Q24/N51	7.95	118.2
M54	7.56	117.7
R55	7.63	118.8
S56	7.83	117.0
K57	7.74	128.1

*Both the proton and nitrogen chemical shifts for three sets of residues are indistinguishable: S6 and S7 are represented by a single peak in the HSQC spectrum (A14 and K37 are represented by a separate single peak as are Q24, N51, and R53)

Figure 6 The residues in a1 whose backbone amide proton chemical shift changes significantly upon mutation of the 25th residue. The residues that have a significant change in backbone amide proton chemical shift upon mutation of residue 25 from serine to tyrosine (N27 and M54) are depicted as space filling. (Note that helix 1 is dark blue, helix two is light blue, and helix three is light green.)

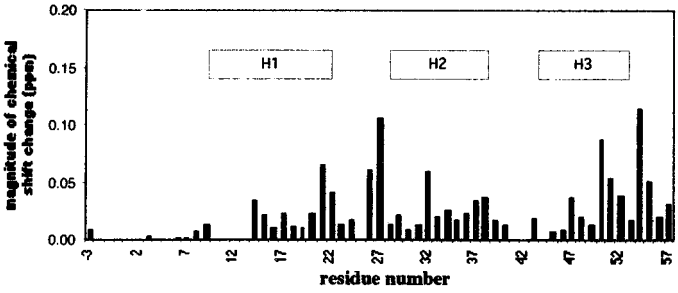

Figure 7 Magnitude of the Difference in Backbone Amide Proton Chemical Shift Between Wild Type a1 and s25y

residue indicated by the diagonal peak in the NOESY spectrum to the residue that was indicated by a cross peak to that diagonal peak. The black residues are those that are helical in wild type **a1** and the blue residues are those that are unstructured in wild type **a1**. Residue 53 is highlighted in yellow because it is unstructured in wild type **a1**, but extends the third helix in s25y. This is indicated by the NOE (arrow) between K52 and R53. Note that some NOEs between amide protons of sequential residues were not distinguishable because of overlap with other peaks in the spectrum.

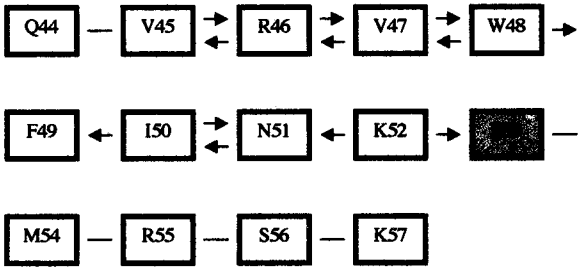

Figure 8 NOEs between amide protons of sequential residues in s25y. Note that the third helix is extended to Residue R53. The black residues are those that are helical in wild type **a1** and the blue residues are those that are unstructured in wild type **a1**. The arrows indicate an NOE between sequential residues was seen, and the dashes indicate that no NOEs were identified on the NOESY spectrum between the two residues. The residue highlighted in yellow, R53, indicates an extension of the third helix in s25y.

Mutation of Glutamine 24 to Arginine and Serine 25 to Tyrosine (q24r/s25y)

The chemical shifts of both the backbone nitrogens and amide protons in the q24r/s25y double mutant homeodomain were determined (See Table 2). From these data the differences in backbone amide proton chemical shift between wild type **a1** and q24r/s25y were calculated. Figure 9 represents **a1** (10) with the residues whose backbone amide protons underwent a significant change in chemical shift upon mutation of the 24th and 25th residues depicted in space filling format. These residues, R21, L26, R53, M54, R55, and K57, are all close in three-dimensional space to the mutated residues. The magnitudes of the differences in backbone amide proton chemical shift between q24r/s25y and wild type **a1**, for all of the homeodomain residues are presented graphically in Figure 10.

Table 2 q24r/e25y Backbone Nitrogen and Amide Proton Chemical Shifts

Residue	HN (ppm)	15N (ppm)
K-3	8.70	125.5
K-2	8.47	124.8
*E-1/K0	8.46	123.8
*K0/E-1	8.46	123.8
S1	8.41	117.7
P2	-	-
K3	8.38	122.4
G4	8.36	110.7
K5	8.20	121.5
*S6/S7	8.36	119.0
*S7/S6	8.36	119.0
I8	8.12	121.8
S9	8.75	125.5
P10	-	-
Q11	8.46	118.2
A12	7.98	125.6
R13	8.49	118.5
*A14/K37	7.76	119.5
F15	7.61	120.8
L16	8.37	121.7
E17	8.07	116.7
Q18	7.46	119.3
V19	8.03	121.9
F20	8.75	121.6
R21	7.35	114.0
R22	7.56	117.1
K23	8.55	122.6
Q24	7.98	119.4
S25	7.33	113.4
L26	8.35	122.5
N27	9.33	122.7
S28	8.49	114.0
K29	7.90	124.1
E30	8.62	120.9
K31	8.57	118.7
*E32/R55	7.49	117.8
E33	8.10	121.0
V34	8.68	121.4
A35	8.15	122.6
K36	7.70	117.0
*K37/A14	7.76	119.5
C38	7.69	112.5
G39	7.86	110.8
I40	7.75	113.1
T41	8.44	110.0
P42	-	-
L43	7.75	119.1
Q44	7.85	118.5
V45	7.85	120.3
R46	8.32	120.5
V47	8.59	119.3
W48	8.29	122.6
F49	8.83	119.8
I50	8.16	120.3
N51	7.94	118.3
K52	8.01	122.7
R53	8.04	118.0
M54	7.51	117.2
*R55/E32	7.49	117.8
S56	7.77	117.1
K57	7.87	128.7

*Residues whose backbone nitrogen and amide proton chemical shifts overlap with those of another residue

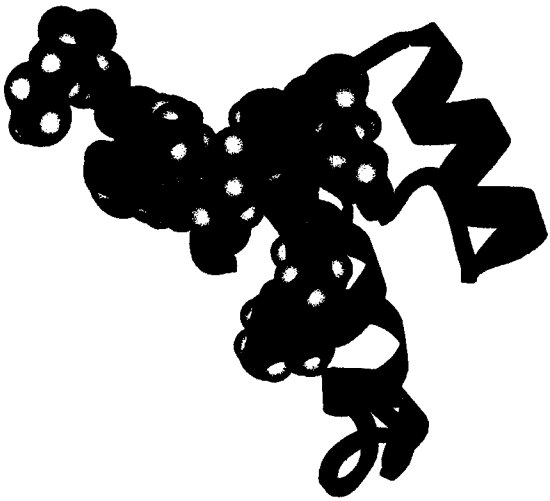

Figure 9 The residues in a1 whose backbone amide proton chemical shift changes significantly upon mutation of the 24th and 25th residues. The residues that have a significant change in backbone amide proton chemical shift upon mutation of residues 24 to arginine and 25 to tyrosine (R21, L26, R53, M54, R55, K57) are depicted as space filling. (Note that helix 1 is dark blue, helix two is light blue, and helix three is light green.)

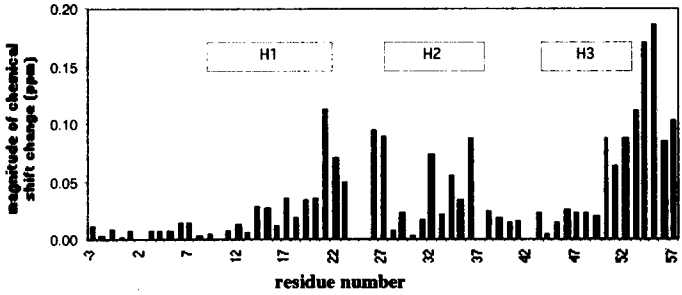

Figure 10 Magnitude of the Difference in Backbone Amide Proton Chemical Shift Between Wild Type a1 and q24r/s25y

The NOEs between amide protons of sequential residues were identified on the 3-D NOESY spectrum of q24r/s25y. Figure 11 illustrates which residues form the third helix in q24r/s25y. As in Figure 8, the arrows point from the residue indicated by the diagonal peak in the NOESY spectrum to the residue that was indicated by a cross peak to that diagonal peak. The black residues are those that are helical in wild type **a1** and the blue residues are those that are unstructured in wild type **a1**. Residues 53 through 56 are highlighted in yellow because they are unstructured in wild type **a1**, but extend the third helix in q24r/s25y.

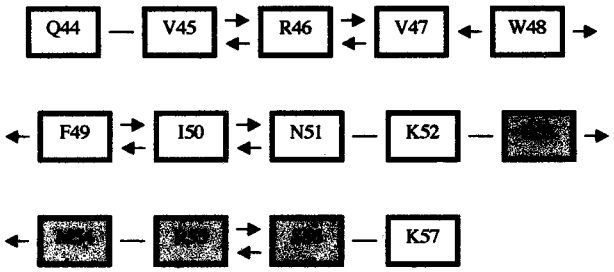

Figure 11 NOEs between amide protons of sequential residues in q24r/s25y. Note that the third helix is extended to residue S56. The black residues are those that are helical in wild type **a1** and the blue residues are those that are unstructured in wild type **a1**. The arrows indicate an NOE between sequential residues was seen, and the dashes indicate that no NOEs were identified on the NOESY spectrum between the two residues. The residues highlighted in yellow, R53-S56, indicate an extension of the third helix in q24r/s25y.

Titration with α2 tail

The changes in chemical shift of the backbone amide protons in wild type **a1** upon titration with α2 tail have been determined recently (13) (See Figure 12a). Figure 13a shows the solution structure of **a1**, (10) in which residues whose backbone amide protons underwent a significant change in chemical shift upon titration with α2 tail are shown as red spheres.

The changes in chemical shift of the backbone amide protons in both s25y and q24r/s25y upon titration with α2 tail were determined (Figures 12b and c). These data were used to create Figures 13b and c. Figure 13b is a representation of s25y in which the residues in s25y whose backbone amide protons underwent a significant change in chemical shift upon binding to α2 tail are indicated by red spheres. Likewise, Figure 13c is a representation of q24r/s25y in which the residues in q24r/s25y whose backbone amide protons underwent a significant change in chemical shift upon binding to α2 tail are also indicated by red spheres. Figure 13 shows that wild type **a1** has more residues whose amide protons underwent a significant change in chemical shift upon α2 tail binding than the mutant homeodomains have.

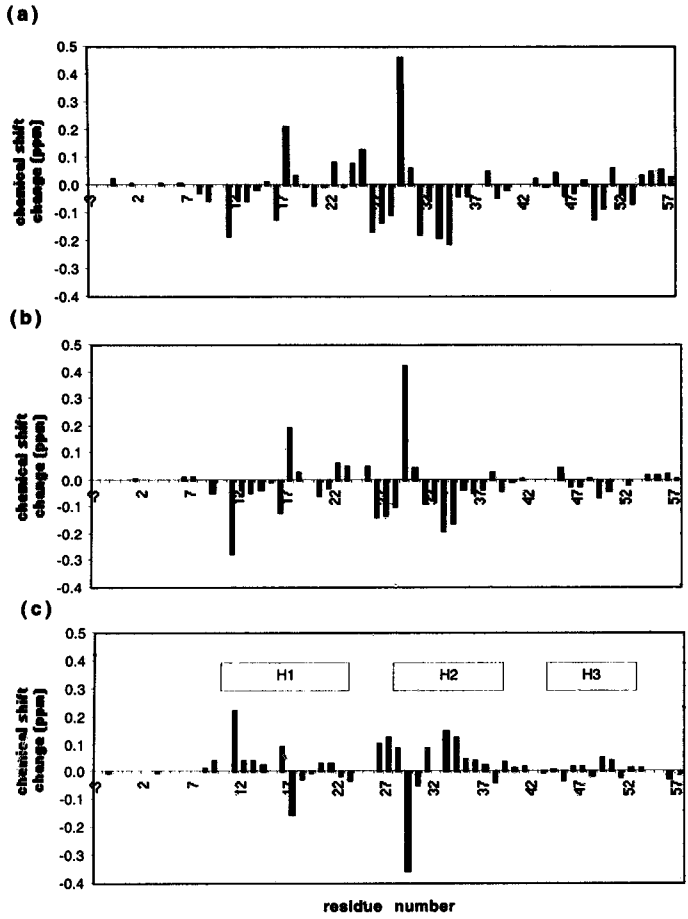

Figure 12 Changes in backbone amide proton chemical shift in (a)Wild Type a1, (b)s25y, and (c)q24r/s25y upon binding to α2 tail.

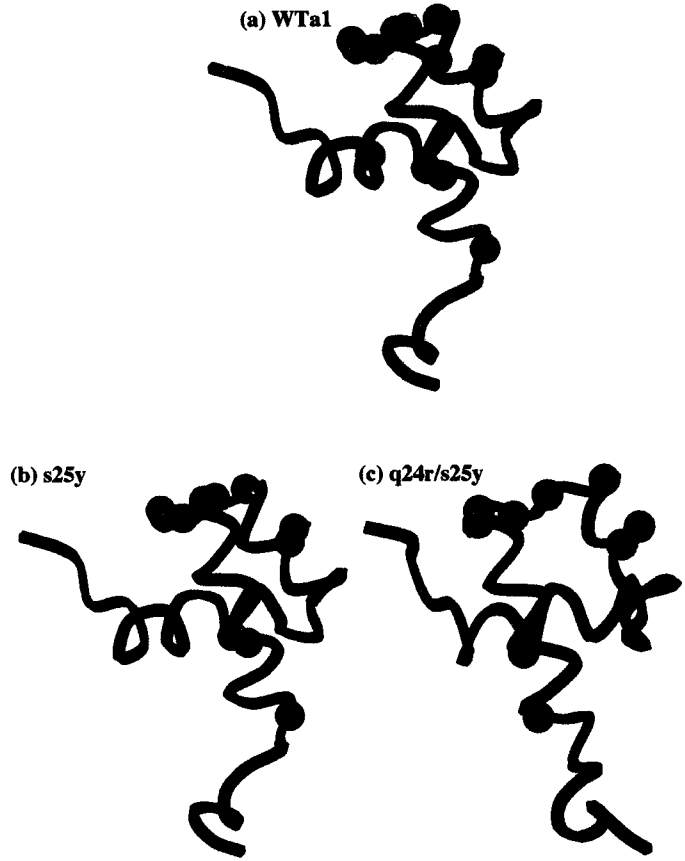

Figure 13 (a)Wild Type a1, (b)s25y, and (c)q24r/s25y. Red spheres indicate residues whose amide protons underwent a significant change in chemical shift. The orange spheres in (b) and (c) identify the mutated residues.

The residues that had significant changes in chemical shift upon titration were used to determine the K_d of the mutant **a1**-α2 tail complexes. Using Mathematica®, the K_d of the s25y-α2 tail complex was found to be 0.11 and the K_d of the q24r/s25y-α2 tail complex was found to be 0.40. (Note: the K_d of the wild type **a1**-α2 complex is 0.3 (14).) The K_d of the s25y-α2 tail complex is significantly less than the K_d of the corresponding wild type **a1** complex, while the difference between the K_ds of the wild type **a1** and q24r/s25y complexes is not significant. The decrease in K_d seen for the s25y-α2 tail complex indicates that the binding between s25y and the α2 tail is tighter than the binding between wild type **a1** and α2.

Electrophoretic Mobility Shift Assays of a1

To test if the **a1** mutants could bind DNA, Electrophoretic Mobility Shift Assays were performed by the Vershon group at Rutgers University (12). Solutions of each homeodomain along with ^{32}P-labeled DNA fragments that contained the consensus binding sites for **a1** and α2 were placed at the top of a gel. The gel, shown in Figure 14,

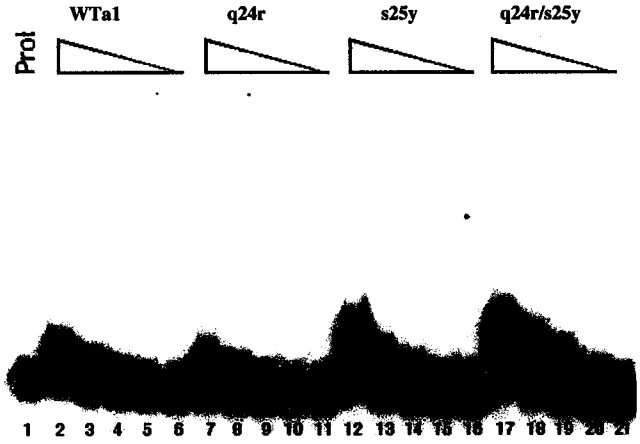

Figure 14 EMSA of Wild Type a1, q24r, s25y, and q24r/s25y with DNA containing the consensus a1 and α2 binding sites. Lane 1 is DNA alone. Lane 2 is 1:1 Wta1:DNA; Lane 3 is 1:5 Wta1:DNA; Lane 4 is 1:25 Wta1:DNA; etc. Lanes 7-11, 12-16, and 17-21 have the same ratios of homeodomain to DNA as Lanes 2-6, for q24r, s25y, and q24r/s25y, respectively. (The lower band represents free DNA and the upper band represents a DNA-homeodomain complex.)

shows that s25y and q24r/s25y bind DNA with a greater affinity than wild type **a1** does. The third lanes in each group (lanes 4,9,14 and 19) illustrate this point. The lower band in the gel represents free DNA, and the upper band represents DNA bound with

homeodomain. In Lane 4, for Wta1, almost all of the DNA is unbound, illustrated by the dark lower band and light upper band. In Lane 9, for q24r, even less DNA is bound, shown by the lighter upper band. However, in Lane 14, the densities of the upper and lower bands are practically identical, illustrating that half of the DNA is bound by s25y. The double mutant binds DNA stronger than s25y: the upper band in Lane 19 is significantly denser than the lower band, indicating that the majority of the DNA fragments are bound with q24r/s25y. The gel shows that q24r/s25y binds DNA with approximately 15 times the affinity that wild type **a1** does.

To test if the mutations affect the ability of **a1** to bind to α2 and form the **a1**-α2-DNA ternary complex, other gels were run with a constant amount of α2 present. The gel, shown in Figure 15, indicates that both s25y and q24r/s25y form the ternary complex much more readily than wild type **a1** does. The first lanes for each homeodomain group (lanes 23, 26, 29, and 32) illustrate this. The dark lower bands in Lanes 23 and 26, for wild type **a1** and q24r respectively, show that the majority of the DNA is unbound with wild type **a1** and q24r. But the dark upper band in Lane 29, for s25y, suggests that s25y has a greater affinity for ternary complex formation. The dark upper band in Lane 32, for q24r/s25y, illustrates that the double mutant has the greatest affinity for ternary complex formation of the homeodomain variants studied.

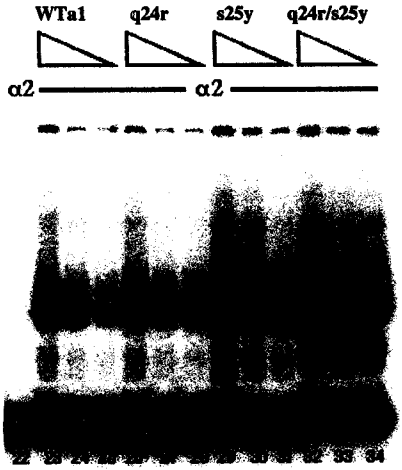

Figure 15 EMSA of Wild Type a1, q24r, s25y, and q24r/s25y with a constant amount of α2 and DNA containing the consensus a1 and α2 binding sites. Lane 22 is DNA alone. Lane 23 has the same concentration of **a1** as Lane 4, that is 1:25 Wta1:DNA, and Lanes 24 and 25 are five fold dilutions. The concentration of α2 is constant in these experiments. (The lower band represents free DNA and the upper band represents a homeodomain-α2 tail-DNA ternary complex)

DISCUSSION

The **a1** homeodomain differs from other homeodomains because it does not bind tightly to DNA on its own. This study focused on the conformational changes caused by point mutations in loop 1 and on the effects these mutations had on DNA binding. EMSAs showed that the s25y and q24r/s25y mutants both bind to the consensus binding site of **a1** with a greater affinity than does wild type **a1** (Figure 14). Furthermore, the mutations enhance the ability of the homeodomain to form the ternary complex with α2 and DNA (Figure 15). This observation was supported by the dissociation constant value of 0.11 for the s25y-α2 heterodimer, obtained from HSQC titration experiments. This value is significantly less than the K_d of the wild type **a1**–α2 heterodimer, which is 0.3, and indicates tighter binding between s25y and the α2 tail than between wild type **a1** and α2. The calculated value of the K_d of the q24r/s25y-α2 heterodimer, 0.4, is inconsistent with the gel shifts, which showed tighter binding of q24r/s25y to α2 than of wild type **a1** to α2.

Mutation of residue 25 from serine to tyrosine caused noticeable changes in the chemical environment of residues both near the loop 1 region in primary sequence and at the C-terminal end of the homeodomain (Figure 7). The changes in chemical shift of residues near the loop 1 region suggest that the mutation may indeed alter the packing of hydrophobic residues in the loop. Despite being over 20 residues away from the mutation, the changes in chemical shift of residues near the C-terminal end is not surprising: the third helix packs against the loop 1 region, and is thus close in space to the site of the mutations. In particular, the third helix in the mutant homeodomain, s25y, appears to be extended to the 53rd residue (Figure 8).

Similar to s25y, q24r/s25y had changes in the chemical environment of residues near loop 1 and near the C-terminal end of the homeodomain. Figure 10 shows that the changes in chemical shift at the C-terminal end are even more pronounced in q24r/s25y than they are in the s25y mutant. In fact, the NOEs between amide protons of sequential residues, represented in Figure 11, illustrate that the third helix in q24r/s25y extends to residue 56. This is an extension of 4 residues, indicating that the third helix is more than a complete turn longer in the double mutant than it is in wild type **a1**. This change could alter the nature of the contacts between **a1** and the major groove of DNA and may contribute to the mutant's increased affinity for DNA.

A recent study has shown that formation of the **a1**-α2-DNA ternary complex causes conformational changes in these same two general locations, the loop 1 region and the third helix (10). By titrating **a1** and the mutant homeodomains with the 19-residue α2 tail peptide, we were able to determine where conformational changes were induced in the homeodomains during heterodimerization. First, the binding of α2 tail to wild type **a1** also causes conformational changes, indicated by changes in chemical shift, in both loop 1 and in the third helix. In Figure 12a we see that the changes seen in the third helix are small and subtle, but significant because of the role of the third helix in the binding of DNA. As illustrated in Figure 13a, several other residues, concentrated in loop 1, of wild type **a1** experience a significant change in chemical environment upon heterodimerization. This implies that the changes in conformation between free **a1** and **a1** bound to DNA in the ternary complex actually occur *during heterodimerization, before binding to DNA.*

In the mutant homeodomains, there are fewer residues whose amide protons underwent a significant change in chemical shift upon binding to α2 tail than did in wild type **a1** (Figure 13). In particular, there are fewer changes seen in loop 1 and in the third helix. This important point is illustrated in Figure 12. Focusing on the third helix, we see that the changes in chemical shift that occur during heterodimerization are diminished in both s25y and q24r/s25y. The chemical shift changes caused by the mutations offer an explanation for this outcome: there are significant differences – between the mutants and wild type **a1** – in the chemical shifts of the residues in the third helix (Figures 7 and 10). This suggests that the mutations change the conformation of **a1**. The fact that the binding of α2 tail does not cause the changes in the third helix of the mutants that is does in wild type **a1** suggests that the mutations might cause the third helix to assume the optimal conformation for DNA binding. Likewise, in both mutant homeodomains, residues 24 and 25 do not experience a large change in chemical shift upon α2 tail binding. This implies that the loop 1 region is more favorably packed in the mutants than in wild type **a1**. If the third helix were in the optimal conformation for DNA binding, and if loop 1 were favorably packed, then this would explain the diminished changes in chemical shift upon α2 tail binding.

REFERENCES

1. Jin, Y., Zhong, H., and Vershon, A. K. (1999) *Mol. Cell. Biol. 19*, 585-593.

2. Stark, M. R., Escher, D., and Johnson, A. D. (1999) *EMBO J. 18*, 1621-1629.

3. Alberts, Bruce et al. Essential Cell Biology, Garland Publishing, Inc, 1998.

4. Gehring, W. J. et al. (1994) *Annu. Rev. Biochem. 63*, 487-526.

5. Wolberger, C. (1999) *Annu. Rev. Biophys. Biomol. Struct. 28*, 29-56.

6. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science 270*, 262-269.

7. Komachi, K., and Johnson, A. D. (1997) *Mol. Cell. Biol. 17*, 6023-6028.

8. Goutte, C., and Johnson, A. D. (1993) *J. Mol. Biol. 233*, 359-371.

9. Phillips, C. L., Stark, M. R., Johnson, A. D., and Dahlquist, F. W. (1994) *Biochemistry 33*, 9294-9302.

10. Anderson, J. S., Forman, M. D., Modleski, S., Dahlquist, F. W., and Baxter, S. (2000) *Biochemistry 39*, 10045-10054.

11. Wuthrich, K. NMR of Proteins and Nucleic Acids, John Wiley and Sons, Inc, 1986.

12. Vershon, A. K. Unpublished data, Rutgers University.

13. Kirkwood, J. Unpublished data, Wadsworth Center.

14. Baxter, S. M., Gontrum, D. M., Phillips, C. L., Roth, A. F., and Dahlquist, F. W. (1994) *Biochemistry 33*, 15309-15320.