Secondary structure of the homeo domain of yeast $\alpha 2$ repressor determined by NMR spectroscopy

Cynthia L. Phillips,^{1,3} Andrew K. Vershon,² Alexander D. Johnson,² and F.W. Dahlquist^{1,3}

¹Institute of Molecular Biology, University of Oregon, Eugene, Oregon 98403 USA; ²Department of Microbiology and Immunology, University of California, San Francisco, California 94143 USA

The yeast $\alpha 2$ protein is a regulator of cell type in *Saccharomyces cerevisiae*. It represses transcription of a set of target genes by binding to an operator located upstream of each of these genes. The $\alpha 2$ protein shares weak sequence similarity with members of the homeo domain family; the homeo domain is a 60-amino-acid segment found in many eukaryotic transcriptional regulators. In this paper we address the question of whether $\alpha 2$ is structurally related to prototypical members of the homeo domain family. We used solution ¹H and ¹⁵N nuclear magnetic resonance (NMR) spectroscopy to determine the secondary structure of an 83-amino-acid residue fragment of $\alpha 2$ that contains the homeo domain homology. We have obtained resonance assignments for the backbone protons and nitrogens of the entire 60-residue region of the putative homeo domain and for most of the remainder of the $\alpha 2$ fragment. The secondary structure was determined by using NOE connectivities between backbone protons, ³J_{HN-H α} coupling constants, and dynamical information from the hydrogen exchange kinetics of the backbone amides. Three helical segments exist in the $\alpha 2$ fragment consisting of residues 11–23, 32–42, and 46–60 (corresponding to residues 138–150, 159–169, and 173–187 of the intact protein). The positions of these three helices correspond extremely well to those of the *Drosophila Antennapedia* (*Antp*) and *engrailed* (*en*) homeo domains, whose three-dimensional structures have recently been determined by NMR spectroscopy and X-ray crystallography, respectively. This study reveals that in spite of low sequence similarity between the homeo domain of $\alpha 2$ and that of higher eukaryotes, the secondary structures of these proteins are well conserved.

[Key Words: $\alpha 2$; homeo domain; helix–turn–helix; yeast; NMR; secondary structure]

Received December 26, 1990; revised version accepted February 19, 1991.

The $\alpha 2$ protein (the product of the *MAT $\alpha 2$* gene) is a transcriptional repressor of cell type-specific genes in the budding yeast *Saccharomyces cerevisiae* (for review, see Herskowitz 1989). Although α itself is a sequence-specific DNA-binding protein, it must interact with other regulatory proteins to bind its DNA operators with high affinity. The specificity of DNA binding by $\alpha 2$ depends on which of these other proteins are present in the cell. In the α cell, $\alpha 2$ acts in combination with the MCM1 protein to bind to DNA operator sequences upstream from α -specific genes, resulting in repression of these genes (Keleher et al. 1988, 1989). In the α/α cell, $\alpha 2$ works with the protein $\alpha 1$. Together, $\alpha 1$ and $\alpha 2$ recognize a second set of operators to turn off haploid-specific genes (Goutte and Johnson 1988).

The $\alpha 2$ protein is folded into two structural domains, separable by proteolysis. The carboxy-terminal domain (residues 132–210) is sufficient for DNA binding (Sauer et al. 1988). On the basis of amino acid sequence homology comparisons, a 61-residue portion of the carboxy-

terminal domain of $\alpha 2$ was proposed to be related to the homeo domain superfamily of proteins (Laughon and Scott 1984; Shepherd et al. 1984). It is now recognized that the homeo domain is a DNA-binding motif of ~61 residues found in many transcriptional regulators (for review, see Hayashi and Scott 1990). There are >90 known members of the homeo domain superfamily, many with extremely similar amino acid sequences (Scott et al. 1989). This conservation can extend between species; for example, the *Drosophila Antennapedia* (*Antp*) homeo domain is identical to the mouse *Hox-1.1* homeo domain at 56 of 61 positions. Other members show less sequence similarity; the *Drosophila Antp* and *engrailed* (*en*) homeo domains are identical at 30/61 positions. The putative homeo domain of $\alpha 2$ has much weaker sequence similarity to the prototypical homeo domains (14/61 identical residues with *Antp*), with most identities in a small region at the carboxy-terminal end of the homeo domain. This observation places $\alpha 2$ on the fringes of the homeo domain superfamily and raises the question of the relatedness of the $\alpha 2$ homeo domain to those of other members of the superfamily.

³Corresponding author.

Recently, the three-dimensional structures of the homeo domains from the *Antp* and *en* proteins have been determined by using nuclear magnetic resonance (NMR) spectroscopy (Otting et al. 1989; Qian et al. 1989; Billeter et al. 1990; Gehring et al. 1990) and X-ray crystallography (Kissinger et al. 1990), respectively. The structures are very similar, with both homeo domains having three α -helices at the same positions in the aligned sequences. Helices 2 and 3 are connected by a tight 3-residue turn and resemble the helix–turn–helix motif first identified in prokaryotic DNA-binding proteins (Laughon and Scott 1984; Pabo and Sauer 1984). However, helix 3 is somewhat longer than that found in prokaryotic helix–turn–helix proteins. In both homeo domain proteins the three helices are packed around a hydrophobic core of conserved residues.

In this paper we examine the structure of the DNA-binding domain of $\alpha 2$. We constructed an 83-residue fragment ($\alpha 2_{128-210}$) from the carboxy-terminal domain of $\alpha 2$ (residues 128–210), overexpressed it in bacteria, and purified it to apparent homogeneity. We used ^1H and ^{15}N NMR spectroscopy to determine the secondary structure of this fragment and some features of its three-dimensional organization. Specifically, we examined the NOE connectivities between backbone protons, $^3J_{\text{HN-H}\alpha}$ coupling constants, which depend on the Ramachandran dihedral angle, ϕ , and hydrogen exchange rates of the backbone amides. We found three α -helical segments, comprised of residues 11–23, 32–42, and 46–60 of $\alpha 2_{128-210}$. (These helices correspond to residues 138–150, 159–169, and 173–187 of the intact protein.) The relative positions of these helices correspond well to those of the *Antp* and *en* homeo domains, with the exception that $\alpha 2$ contains a 3-residue insert relative to the other two. In addition, we find that the helices of $\alpha 2$ are packed about a hydrophobic core, suggesting a very similar tertiary structure to the other two proteins. These results indicate that $\alpha 2$, despite limited sequence similarity, shares a common structural motif with members of the homeo domain superfamily.

Results

Assignments

We began our study of $\alpha 2_{128-210}$ with the traditional ^1H two-dimensional NMR NOESY- and COSY-type experiments. Using COSY, RELAY, and TOCSY experiments, we identified the H^{N} , H^{α} , and H^{β} resonances of many amino acid spin systems. However, poor dispersion of the H^{N} – H^{α} cross peaks in the “fingerprint” region of these spectra, possibly reflecting disordered regions of the protein, prohibited the complete sequential assignment of the spectra. To resolve the amide proton resonances and to allow the selective detection of only resonances from amide protons, we turned to heteronuclear ^1H – ^{15}N NMR experiments by using ^{15}N -enriched $\alpha 2_{128-210}$ (Griffey and Redfield 1987).

Figure 1 shows a two-dimensional ^1H – ^{15}N HSMQC spectrum, in which every directly bonded ^1H – ^{15}N pair is

detected as a single cross peak at its ^1H and ^{15}N chemical shifts. In the HSMQC spectrum, the two tryptophan indole $^{15}\text{N}^{\epsilon 1}\text{H}$ cross peaks are evident in the far downfield region of the spectrum. The cross peak pairs, arising from $^{15}\text{NH}_2$ groups, cluster in the more upfield region of the spectrum. These are joined in a line in the figure, each with two distinct proton shifts and one common nitrogen shift. These pairs correspond to the seven asparagine and single glutamine side chains. There is a ninth NH_2 cross peak pair, which may be due to conformational heterogeneity at one of these asparagine or glutamine residues. The remaining cross peaks in the HSMQC spectrum arise from backbone amide ^{15}NH pairs. We expect a maximum of 78 ^1H – ^{15}N amide cross peaks to be visible in the HSMQC spectrum of $\alpha 2_{128-210}$, as the amino-terminal ammonium group is exchange-broadened and the four proline residues have no amide protons. We observe a total of 81 amide cross peaks. Of these, 72 are assigned specifically. The nine unassigned cross peaks are all sharp and arise from the ends of the fragment, where there may be some conformational heterogeneity or disorder.

We used two parallel approaches to assign the backbone resonances of $\alpha 2_{128-210}$. In the first approach, we used a ^1H DQF–COSY spectrum to define the positions of the H^{N} – H^{α} cross peaks in the ^1H NOESY spectrum. We then used ^{15}N -edited COSY and NOESY spectra of the $\alpha 2_{128-210}$ samples selectively labeled with [^{15}N]leucine, [^{15}N]isoleucine, or [α - ^{15}N]lysine to identify cross peaks from those three residue types. Using the resonances from labeled residues as “anchor points,” we worked through the sequential connectivities in the H^{N} – H^{N} and H^{N} – H^{α} regions of the ^1H NOESY spectrum by following the “main chain-directed approach” (Englander and Wand 1987). For example, we identified the amide resonances of Lys11–Leu17 as a result of NOE connectivities joining a lysine to an isoleucine five H^{N} – H^{N} “steps” away and found a leucine adjacent to that isoleucine.

The second approach enabled us to check the assignments from the first approach and to make additional backbone proton assignments. We used a HSMQC spectrum of the fully ^{15}N -labeled $\alpha 2_{128-210}$ to define the positions of the ^1H – ^{15}N cross peaks in the HSMQC–NOESY spectrum of the fully ^{15}N -labeled $\alpha 2_{128-210}$. In this experiment proton NOE interactions are observed as weak cross peaks interspersed among the strong ^1H – ^{15}N cross peaks also seen in an HSMQC experiment alone. We then used HSMQC spectra of the [^{15}N]leucine, [^{15}N]isoleucine, or [α - ^{15}N]lysine selectively labeled $\alpha 2_{128-210}$ to identify ^1H – ^{15}N resonances from those residues in the HSMQC–NOESY spectrum. With the ^1H – ^{15}N cross peaks from selectively labeled residues as anchor points, we could work through the H^{N} – H^{N} sequential connectivities in the HSMQC–NOESY spectrum. Figure 2 shows these H^{N} – H^{N} connectivities in the first helix sketched on an HSMQC–NOESY spectrum. The Lys11 ^1H – ^{15}N cross peak is connected to the Glu12 ^1H – ^{15}N cross peak via the less intense NOE cross peak at the ^{15}N chemical shift of the Lys11 amide and the ^1H chem-

Figure 1. HSMQC spectrum of fully ^{15}N -labeled $\alpha 2_{128-210}$. Each cross peak arises from a proton directly bonded to a nitrogen-15. In the lower right corner of the spectrum, each pair of $^{15}\text{NH}_2$ s is connected by a line. Assigned cross peaks are labeled. There are nine cross peaks in the spectrum as yet unassigned to amides, but only 6 residues remain unassigned. We attribute the additional cross peaks to conformational heterogeneity (not to impurities) in a short section of the protein backbone, probably near the amino or carboxyl terminus.

ical shift of the Glu12 amide (also the ^1H chemical shift of the Lys11 amide and the ^{15}N chemical shift of the Glu12 amide). Glu12 is similarly connected to Asn13, and so on, through the helix. Because of favorable multiple quantum relaxation in the HSMQC-NOESY experiment relative to the ^1H -NOESY experiment, we were able to observe weak $\text{H}^{\text{N}}-\text{H}^{\text{N}}$ cross peaks in the HSMQC-NOESY spectrum that we were unable to detect in the ^1H NOESY spectrum. Thus, sequential assignments could be made in the end regions of $\alpha 2_{128-210}$, where there are no strong $\text{H}^{\text{N}}-\text{H}^{\text{N}}$ and $\text{H}^{\text{N}}-\text{H}^{\alpha}$ connectivities as are typical in α -helices. Having specifically identified the ^{15}N and H^{N} resonances of $\alpha 2_{128-210}$ from the HSMQC-NOESY spectra, we then assigned the corresponding H^{α} and, if possible, H^{β} , resonances of each residue by using COSY experiments.

At this time, almost all of the backbone ^{15}N , H^{N} , and H^{α} resonances are assigned. In addition, $\sim 50\%$ of the H^{β} protons and most of the aromatic protons are assigned from the COSY and NOESY spectra of unlabeled $\alpha 2_{128-210}$.

NOE connectivities

Figure 3 shows the backbone NOE connectivities iden-

tified from the ^1H NOESY and HSMQC-NOESY spectra of $\alpha 2_{128-210}$. There are three clusters of contiguous $d_{\alpha\text{N}}(i, i+3)$ and strong d_{NN} connectivities. Such connectivities are diagnostic of helices (Wüthrich 1986). Regions of α -helix tend to have $d_{\alpha\text{N}}(i, i+4)$ connectivities in addition to the $d_{\alpha\text{N}}(i, i+3)$ connectivities, whereas 3_{10} -helices and turns tend to have $d_{\alpha\text{N}}(i, i+2)$ connectivities. The three helices appear to be α -helical, with several $d_{\alpha\text{N}}(i, i+4)$ connectivities observed.

Residues 43–45 have relatively strong $d_{\alpha\text{N}}(i, i+2)$ connectivities and no detectable $d_{\alpha\text{N}}(i, i+3)$ connectivities, consistent with a turn conformation.

Measurement of $^3J_{\text{HN-H}\alpha}$ coupling constants

To better define the secondary structure of $\alpha 2_{128-210}$ we measured the $^3J_{\text{HN-H}\alpha}$ coupling constants using the HMQC-J experiment. This scalar coupling constant is dependent on the Ramachandran dihedral angle ϕ of the protein backbone. The results of these measurements are also shown in Figure 3. We were able to observe $^3J_{\text{HN-H}\alpha}$ coupling constants of greater than 5 Hz. Cross-peak splittings of less than this value were not resolved in the spectra of $\alpha 2_{128-210}$, which has $^1\text{H}-^{15}\text{N}$ multiple quantum line widths of ~ 11 Hz. In Figure 3, solid circles

Figure 2. HSMQC-NOESY spectrum of fully ^{15}N -labeled $\alpha 2_{128-210}$. The higher intensity cross peaks correspond to protons directly bonded to a nitrogen-15. The weaker cross peaks arise from NOEs between these protons. Thus, the weaker cross peaks connect to amides that are close in space. Traced are the sequential connectivities from residues 11 through 25.

indicate a $^3J_{\text{HN-H}\alpha}$ coupling of less than 6.5 Hz. A contiguous series of residues with $^3J_{\text{HN-H}\alpha}$ couplings of less than 6.5 Hz is strongly indicative of a helical structure (Wüthrich 1986). There are three such sequences in $\alpha 2_{128-210}$, coincident with the three helices identified by NOE connectivities.

Residues 43–45 have larger $^3J_{\text{HN-H}\alpha}$ couplings of 7–8 Hz, indicating a break between helices 2 and 3. These couplings are consistent with a tight turn between the helices.

Amide hydrogen exchange

The results of the hydrogen exchange experiment at pH 4.5, 25°C, are also shown in Figure 3. Solvent-inaccessible and hydrogen-bonded amide protons are generally protected from exchange with the solvent (Englander and Kallenbach 1984). Regions of α -helical structure are typically protected from hydrogen exchange due to the intrahelical hydrogen bonds that begin after the third residue. The three regions of contiguous residues in $\alpha 2_{128-210}$ that are protected from hydrogen exchange are consistent with there being three α -helices in the structure. Both the amino- and carboxy-terminal regions of

the protein exchange rapidly, consistent with these regions having little defined structure.

Tertiary connectivities

In addition to these short-range NOE connectivities, we have identified several long-range NOE connectivities as a consequence of the aromatic ring assignments. There are NOE cross peaks between the aromatic ring protons of Phe9 and the $\text{H}^{\delta 1}$ and $\text{H}^{\epsilon 1}$ of Trp52, between the aromatic ring protons of Phe21 and those of Trp52, and between the $\text{H}^{\epsilon 3}$ of Trp20 and the $\text{H}^{\zeta 2}$ or $\text{H}^{\zeta 3}$ of Trp52. Both Leu30 and Leu17 side-chain protons show NOEs to the aromatic ring protons of Trp20 as well. These are probably primary NOEs (at 100 msec mixing time), restricting the distance between the protons to <5.0 Å. Thus, the side chain of Leu30 and Leu17, along with the aromatic rings of Phe9, Trp20, Phe21, and Trp52, pack to form a hydrophobic core of $\alpha 2_{128-210}$.

Discussion

Secondary structure of $\alpha 2_{128-210}$

We assigned nearly all of the backbone amide nitrogen-15 and proton resonances of $\alpha 2_{128-210}$ and have identified

Figure 3. The $\alpha_{2_{128-210}}$ amino acid sequence with sequential and medium-range NOE connectivities, $^3J_{\text{HN-H}\alpha}$ coupling constants, and hydrogen exchange data. Below the sequence, lines of varying widths indicate NOE connectivities. The narrower lines indicate $d_{\alpha\text{N}}(i,i+2)$, $d_{\alpha\text{N}}(i,i+3)$, and $d_{\alpha\text{N}}(i,i+4)$ connectivities. The broader lines indicate d_{NN} and $d_{\alpha\text{N}}$ sequential connectivities. Both the d_{NN} connectivities obtained from the HSMQC–NOESY (d_{NN}^a) and those from the ^1H –NOESY (d_{NN}^b) are shown to demonstrate how many more d_{NN} connectivities were observable in the HSMQC–NOESY than in the ^1H –NOESY. Directly below the sequence, large solid circles indicate $^3J_{\text{HN-H}\alpha}$ coupling constants of <6.5 Hz; open circles indicate $^3J_{\text{HN-H}\alpha}$ coupling constants of >6.5 Hz. Below the coupling constant data are the hydrogen exchange data, taken at pH* (observed pH meter reading, uncorrected for the deuterium isotope effect) 4.50, 25°C, in 100 mM KCl, 0.05% NaN_3 , and 25 mM potassium phosphate. One small dot indicates a rate of hydrogen exchange faster than 0.14 min^{-1} . Two small dots indicate a rate between 0.14 and 0.05 min^{-1} , and three small dots indicate a rate of hydrogen exchange slower than 0.05 min^{-1} . The thick bars above the $\alpha_{2_{128-210}}$ sequence show the positions of the three helices as determined by these data.

the sequential and intermediate-range NOE connectivities involving the H^{N} and H^{α} protons. These NOEs roughly define the secondary structure of $\alpha_{2_{128-210}}$ as three α -helices separated by two turns. Helix 1 spans residues 11–23, helix 2 spans residues 32–42, and helix 3 spans residues 46–60. The observation of $^3J_{\text{HN-H}\alpha}$ coupling constants of <6.5 Hz provides additional information concerning the boundaries of α -helices. We used this information to confirm and define more rigorously the end points of the three helices. This was particularly useful for defining the carboxy-terminal end of helix 3, because fewer helical-type NOEs were detected in this region between residues 54 and 60. The coupling constants in this region are <6.5 Hz, helping to define it as an extension of helix 3.

The hydrogen exchange properties of $\alpha_{2_{128-210}}$ are also consistent with there being three helices. Protection from exchange typically begins at the fourth residue of an α -helix. At this position, the amide protons begin to form intrahelical $(i,i+4)$ hydrogen bonds, thereby providing protection from hydrogen exchange (Wüthrich 1986). Examination of the exchange properties shown in Figure 3 reveals helix 2 to be weakly protected from hydrogen exchange, relative to helix 1 and helix 3. This suggests

that helix 2 may have more flexibility in the backbone relative to helices 1 and 3. The other striking hydrogen exchange property of $\alpha_{2_{128-210}}$ is seen at the end of helix 3 (residues 54–59), where the exchange is considerably faster than that at the beginning of the helix. This region appears to exchange at a rate intermediate to that of the slowly exchanging first part of helix 3 and the very quickly exchanging regions at both ends of $\alpha_{2_{128-210}}$. The pattern of helical NOE connectivities also breaks down in residues 54–60, suggesting that this carboxy-terminal end of helix 3 is probably less regular and perhaps more flexible than the amino-terminal end of helix 3.

Interhelix connections

We now consider the structure of the residues that connect the helices. The backbone NOE connectivities for residues 43–45 (linking helices 2 and 3) show that the $\text{H}\alpha$ protons of these residues are near the amide protons 2 residues toward the carboxyl terminus. These short $d_{\alpha\text{N}}(i,i+2)$ distances, combined with the apparent lack of $d_{\alpha\text{N}}(i,i+3)$ NOEs and the relatively large $^3J_{\text{HN-H}\alpha}$ couplings observed for residues 43–45, suggest that these

residues form a turn. The few tertiary NOEs that we have unambiguously assigned support this conclusion. We observe NOEs between the ring protons of Trp52 and protons from residues Phe9, Trp20, and Phe21. The ring protons of residue 20 show NOEs to the Leu17 and Leu30 side chains. To bring Trp52 to the required distance for NOE energy transfer, residues 43–45 must form a tight turn. Similarly, to bring Trp52 to within 5 Å of residues 20 and 21, there must be a loop or turn between helices 1 and 2. Thus, we conclude that helices 1 and 2 are joined by a loop and helices 2 and 3 are joined by a tight turn.

The data do not allow us to unambiguously define the secondary structure of the carboxy-terminal end of helix 1. The NOE connectivities and coupling constants for residues 24–27 are not those of an ideal α -helix but could be consistent with a distorted helical conformation. Asp24 has a very large $^3J_{\text{HN-H}\alpha}$ coupling of 10 Hz and is most likely in an extended conformation, but the next residue, Ile25, has a small $^3J_{\text{HN-H}\alpha}$ coupling with strong sequential amide NOEs and is most likely in a helical conformation. To be conservative, we define the end of helix 1 at residue 23. The chain could continue, in principle, in a distorted helix for 2 or 3 residues, although it surely ends by Pro28.

Comparison with *Antp* and *en*

In spite of the low sequence homology, the secondary structure of the $\alpha 2$ homeo domain is remarkably similar to those of *Antp* and *en*, determined by NMR spectroscopy (Otting et al. 1989; Qian et al. 1989; Billeter et al. 1990; Gehring et al. 1990) and X-ray crystallography (Kissinger et al. 1990), respectively. Figure 4 shows the aligned homeo domain sequences of $\alpha 2$, *Antp* and *en*, with the positions of the three helices in each. The homology of $\alpha 2$ to *Antp* and *en* is slightly higher if residues 24–26 (NIE) are deleted from the $\alpha 2$ sequence (Shepherd et al. 1984). The sequences in Figure 4 are aligned accordingly.

This alignment places the beginning of helix 1 at the same position in all three proteins and aligns the sequences N-YL- in the loop region of each homeo domain. The addition of the triplet NIE in the $\alpha 2$ homeo domain sequence could either extend helix 1 into a longer distorted helix or could contribute to a larger loop between helices 1 and 2 than is seen in the *Antp* and *en* homeo domains. With the NIE addition in $\alpha 2$, helix 2 in the $\alpha 2$ homeo domain is at the same position and is of the same length in both the *Antp* and *en* homeo domains. The following 3-residue turn between helices 2 and 3 is conserved in all three homeo domains. This turn may be critical to the angle between helices 2 and 3, reminiscent of the helix–turn–helix motif of prokaryotes (Laughon and Scott 1984; Pabo and Sauer 1984). Helix 3 again begins at the same position in all three homeo domains but is marginally shorter in $\alpha 2_{128-210}$ than in the *Antp* (considering helices 3 and 4 of *Antp* as a single unit) and *en* homeo domains. We see increased hydrogen exchange rates of the amides from residues 55–60 of helix 3. This same region in *Antp* (helix 4) also shows increased amide exchange kinetics. Although the $^3J_{\text{HN-H}\alpha}$ coupling constants in this region of $\alpha 2_{128-210}$ are all <6.5 Hz (indicating an α -helix), they are all measurably higher than those in helices 1, 2, and the amino-terminal portion of helix 3 (data not shown). These hydrogen exchange and coupling constant data also suggest that the carboxy-terminal region of helix 3 is exposed to solvent and extended, as is the same region (helix 4) of the *Antp* homeo domain. The residues in this region are mostly positively charged and are thought to contact DNA (Billeter et al. 1990; Kissinger et al. 1990; Otting et al. 1990). The flexibility of this region in $\alpha 2$ may therefore be important for DNA recognition.

A region of $\alpha 2_{128-210}$ lying outside of the homeo domain proper has been proposed to contact the *a1* protein (Strathern et al. 1988; A.K. Vershon and A. Mak, unpubl.). As can be seen in Figure 3, residues 70–73 show some protection from hydrogen exchange, as well as d_{NN} and d_{aN} connectivities. This is not seen anywhere else

Figure 4. Comparison of the amino acid sequence and helix positions in $\alpha 2_{128-210}$ and the homeo domains of *Antp* and *en*. The *Antp* and *en* homeo domain sequences are aligned to skip the triplet NIE sequence of $\alpha 2$ to maximize the sequence and structural homology. The boxed sequences show the helical regions determined for each protein. The third helical section of the *Antp* homeo domain has two boxes marking where helix 3 turns into the extended helix 4. A kink at Arg57 (as numbered above) of *Antp* has been proposed (Qian et al. 1989) as the basis for that distinction between helices 3 and 4. We lack sufficient evidence to prove or disprove the existence of a kink in $\alpha 2_{128-210}$. No evidence for such a kink was observed in the X-ray crystal structure of the *en* homeo domain complex.

outside of the homeo domain in $\alpha 2_{128-210}$. We attribute these observations to a protected, somewhat structured region. This region may interact directly with **a1**. Excluding the homeo domain, this is the only region of $\alpha 2_{128-210}$ where we detect any structural connectivities. The remaining residues, 1–9, 61–69, and 75–83, give sharp resonances in all spectra and show very weak, sparse connectivities, suggesting that these regions may be disordered.

The helices of the $\alpha 2$ homeo domain pack about a hydrophobic core. The highly conserved tryptophan at position 52 in $\alpha 2_{128-210}$ contributes to this core in all three homeo domain structures. The NOE connectivities observed between Trp20, Phe21, and Trp52 in $\alpha 2_{128-210}$ suggest that helix 1 packs against helix 3. This is consistent with the tertiary structures of the *Antp* and *en* homeo domains, in which helices 1 and 3 cross each other in the vicinity of these residues.

The work described here demonstrates that the secondary structure of the $\alpha 2$ homeo domain is very similar to those of the *Antp* and *en* homeo domains. The limited tertiary distance constraints obtained to date suggest that the overall fold of the $\alpha 2$ homeo domain is similar to the fold of the *Antp* and *en* homeo domains. We therefore conclude that $\alpha 2$ is a genuine member of the homeo domain superfamily.

Materials and methods

Vector construction

The protein $\alpha 2_{128-210}$ was purified from *Escherichia coli* cells containing the plasmid pAV105. This plasmid encodes for residues 128–210 of $\alpha 2$, under the control of tandem P_{tac} promoters. A gene cassette encoding for these residues was synthesized via polymerase chain reaction and digested with *NdeI* and *SalI* (MacFerrin et al. 1990). This fragment was cloned into the *NdeI* and *SalI* sites of pCWori+ (Muchmore et al. 1989; Gegner and Dahlquist 1991) and transformed into *E. coli* HB101. The construction of the resulting plasmid was confirmed by restriction digestion and DNA sequencing. Expression of the cloned $\alpha 2_{128-210}$ was verified by Western blot analysis with antibody raised against intact $\alpha 2$ protein.

Protein purification and ^{15}N labeling

The $\alpha 2_{128-210}$ protein was purified to >95% homogeneity as described elsewhere (Wolberger et al. 1991), with the following minor modifications. The cells were lysed by French press. The supernatant was first loaded on a CM Sepharose CL 6B (Sigma) column and eluted with a gradient from 0 to 0.4 M NaCl in S buffer (50 mM Tris-HCl, 1 mM EDTA at pH 8.0). The protein was then loaded on a Whatman cellulose phosphate P11 column and eluted with a gradient from 0 to 0.75 M NaCl in S buffer. After elution from a subsequent phenyl Sepharose (Sigma) column at 2 mM NaCl, the purified $\alpha 2_{128-210}$ was dialyzed against 25 mM potassium phosphate buffer (pH 6.5) with 50 mM KCl and 0.05% NaN_3 . Finally, the protein was concentrated, using a 1-ml column of SP Sephadex C50 (Sigma). The protein was eluted with a small volume of 25 mM potassium phosphate buffer (pH 6.5) with 500 mM KCl and 0.05% NaN_3 and stored frozen at -70°C until needed. The molecular mass of $\alpha 2_{128-210}$ is 9907 daltons,

and we used the calculated A_{280} of 1.0 for 0.74 mg/ml to determine the protein concentration.

The protein was concentrated further to 2–5 mM by using Amicon Centricon 3 microconcentrators and subsequently dialyzed against 25 mM potassium phosphate buffer (pH 6.5) with 100 mM KCl, 0.05% NaN_3 . The pH was adjusted as necessary by using HCl, and D_2O was added to 5% for the spectrometer lock. $\alpha 2_{128-210}$ is soluble as a monomer to >6 mM. We determined that the fragment is largely folded under the conditions used: 4–30°C, pH* (observed pH meter reading, uncorrected for the deuterium isotope effect) 4.5–7.0.

Selective and uniform ^{15}N labeling (>95% enrichment) of $\alpha 2_{128-210}$ was carried out using *E. coli* strains, media, and methods described previously (Muchmore et al. 1989; McIntosh and Dahlquist 1990; McIntosh et al. 1990). We selectively enriched $\alpha 2_{128-210}$ with [^{15}N]leucine, [^{15}N]isoleucine, or [α - ^{15}N]lysine by using the auxotrophic *E. coli* strain DL39 (LeMaster and Richards 1988) grown in synthetic rich media containing one of these isotopically labeled amino acids (Muchmore et al. 1989). The $\alpha 2_{128-210}$ was uniformly ^{15}N -labeled by using a prototrophic *E. coli* strain grown in M9T media containing 1 g/liter of $(^{15}\text{NH}_4)_2\text{SO}_4$ as the sole nitrogen source (McIntosh et al. 1990). The $(^{15}\text{NH}_4)_2\text{SO}_4$ was obtained from Monsanto Research Corporation, and the ^{15}N -enriched amino acids were purchased from Tracer Technology.

NMR spectroscopy

All spectra were acquired with 2–5 mM $\alpha 2_{128-210}$ at 25°C on a General Electric Omega 500 spectrometer operating at 11.9 tesla. The recycle times, not including acquisition, were 850 msec. The chemical shifts were set relative to an external proton reference of sodium 2,2-dimethyl-2-silapentane-5-sulfonate and an external nitrogen standard of $^{15}\text{NH}_4\text{Cl}$ (Levy and Lichter 1979). Spectra were analyzed by using FTNMR software from Hare Research.

Conventional ^1H DQF-COSY, NOESY, and TOCSY spectra of $\alpha 2_{128-210}$ were recorded and processed as described in detail elsewhere (Wüthrich 1986; Oppenheimer and James 1989). ^{15}N - ω_2 -edited COSY and NOESY spectra of uniformly and selectively labeled $\alpha 2_{128-210}$ were recorded and processed as described for T4 lysozyme by McIntosh et al. (1990). The single-bond ^1H - ^{15}N correlation spectra were measured by the heteronuclear single-multiple quantum coherence experiment (HSMQC) (Zuiderweg 1990). Most of the sequential amide connectivities were obtained from an HSMQC-NOESY spectrum. The HSMQC-NOESY experiment is a variation of the HMQC-NOESY experiment (Gronenborn et al. 1989; Muchmore et al. 1989), in which an HSMQC ^1H - ^{15}N pulse sequence is used to generate ^1H - ^{15}N multiple quantum coherence. The HSMQC-NOESY has improved resolution significantly in the ^{15}N dimension over the HMQC-NOESY.

The spectra were recorded with a spectral width of 6410 Hz in the ^1H dimensions and 3333 or 4000 Hz in the ^{15}N dimensions. The TOCSY spectra were recorded with a mixing time of 70 msec by using an MLEV-17 pulse train, and the NOESY spectra were recorded with a 100-msec mixing time.

The $^3J_{\text{HN-H}\alpha}$ coupling constants were measured through the use of an HMQC-J experiment (Kay et al. 1989; Forman-Kay et al. 1990; Kay and Bax 1990). The data were collected with 1024 complex points in T_1 and T_2 by using TPPI (Marion and Wüthrich 1983). The spectra were zero-filled to 8K in the ^{15}N dimension to give a digital resolution of 0.5 Hz/point. To determine which $^3J_{\text{HN-H}\alpha}$ coupling constants were greater than or less than 6.5 Hz, the measured $^3J_{\text{HN-H}\alpha}$ coupling constants were adjusted approximately by using the ratio of $J_{\text{measured}}/J_{\text{actual}}$ calcu-

lated by Kay and Bax (1990) for a typical resonance with 11 Hz line width.

The hydrogen exchange experiments required fast transfer of $\alpha 2_{128-210}$ from H_2O to D_2O . This was accomplished by passing the $\alpha 2_{128-210}$ through a Quik-Sep (Isolab, Inc.) spin column containing G-50 Sephadex (Sigma) pre-equilibrated in a D_2O buffer of the same pH and salt content. Immediately following the transfer to D_2O , 10 HSMQC spectra, each requiring 37 min, were recorded at 25°C over a period of 24 hr. A graph of the volume of each amide cross peak as a function of the time after solvent transfer was fit to an exponential curve, $V(t) = V_\infty + V_0 e^{-kt}$. V_0 and V_∞ , the initial and final cross peak volumes, respectively, were allowed to vary to give the best fit of the data to obtain the rate of exchange, k , of each amide proton.

Acknowledgments

We thank Lawrence McIntosh for advice and help with the NMR spectroscopy, A. Mak for communication of unpublished results, and Dennis Hare and Hare Research for supplying the FTNMR software. A.K.V. is a Smith, Kline, Beecham Fellow of the Life Sciences Research Foundation. This work was supported by a grant from the American Cancer Society (BE-74) to F.W.D., and grants from the National Institutes of Health (GM37049) and Pew Charitable Trust to A.D.J.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Billeter, M., Y.Q. Qian, G. Otting, M. Müller, and K. Wüthrich. 1990. Determination of the three-dimensional structure of the *Antennapedia* homeo domain from *Drosophila* in solution by 1H nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **214**: 183–197.
- Englander, S.W. and N.R. Kallenbach. 1984. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**: 521–655.
- Englander, S.W. and J.A. Wand. 1987. Main-chain-directed strategy for the assignment of 1H NMR spectra of proteins. *Biochemistry* **26**: 5953–5958.
- Forman-Kay, J.D., A.M. Gronenborn, L.E. Kay, P.T. Wingfield, and M.G. Clore. 1990. Studies on the solution conformation of human thioredoxin using heteronuclear ^{15}N - 1H nuclear magnetic resonance spectroscopy. *Biochemistry* **29**: 1566–1572.
- Gegner, J.A. and F.W. Dahlquist. 1991. Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. *Proc. Natl. Acad. Sci.* **88**: 750–754.
- Gehring, W.J., M. Müller, M. Affölter, A. Percival-Smith, M. Billeter, Y.Q. Qian, G. Otting, and K. Wüthrich. 1990. The structure of the homeo domain and its functional implications. *Trends Genet.* **6**: 323–329.
- Goutte, C. and A.D. Johnson. 1988. $\alpha 1$ alters the DNA binding specificity of $\alpha 2$ repressor. *Cell* **52**: 875–882.
- Griffey, R.H. and A.G. Redfield. 1987. Proton-detected heteronuclear edited and correlated nuclear magnetic resonance and nuclear Overhauser effect in solution. *Q. Rev. Biophysics* **19**: 51–82.
- Gronenborn, A.M., P.T. Wingfield, and G.M. Clore. 1989. A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N - 1H multiple quantum coherence spectroscopy. *FEBS Lett.* **243**: 93–98.
- Hayashi, S. and M.P. Scott. 1990. What determines specificity of action of *Drosophila* homeodomain proteins? *Cell* **63**: 883–894.
- Herskowitz, I. 1989. A regulatory hierarchy for cell specialization in yeast. *Nature* **342**: 749–757.
- Kay, L.E. and A. Bax. 1990. New methods for the measurement of NH-C α H coupling constants in ^{15}N -labeled proteins. *J. Magn. Reson.* **86**: 110–126.
- Kay, L.E., B. Brooks, S.W. Sparks, D.A. Torchia, and A. Bax. 1989. Measurement of NH-C α H coupling constants in staphylococcal nuclease by two-dimensional NMR and comparison with x-ray crystallographic results. *J. Am. Chem. Soc.* **111**: 5488–5490.
- Keleher, C.A., C. Goutte, and A.D. Johnson. 1988. The yeast cell-type-specific repressor $\alpha 2$ acts cooperatively with a non-cell-type-specific protein. *Cell* **53**: 927–936.
- Keleher, C.A., S. Passmore, and A.D. Johnson. 1989. Yeast repressor $\alpha 2$ binds to its operator cooperatively with yeast protein MCM1. *Mol. Cell. Biol.* **9**: 5228–5230.
- Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo. 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.
- Laughon, A. and M.P. Scott. 1984. Sequence of a *Drosophila* segmentation gene: Protein structure homology with DNA binding proteins. *Nature* **310**: 25–31.
- LeMaster, D.M. and F.M. Richards. 1988. NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteration. *Biochemistry* **27**: 142–150.
- Levy, G.C. and R.L. Lichter. 1979. *Nitrogen-15 nuclear magnetic resonance spectroscopy*. Wiley & Sons, New York.
- MacFerrin, K.D., M.P. Terranova, S.L. Schreiber, and G.L. Verdine. 1990. Overproduction and dissection of proteins by the expression-cassette polymerase chain reaction. *Proc. Natl. Acad. Sci.* **87**: 1937–1941.
- Marion, D. and K. Wüthrich. 1983. Application of phase sensitive correlated spectroscopy for measurement of proton-proton spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.* **113**: 967–974.
- McIntosh, L.P. and F.W. Dahlquist. 1990. Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q. Rev. Biophys.* **23**: 1–38.
- McIntosh, L.P., A.J. Wand, D.A. Lowry, A.G. Redfield, and F.W. Dahlquist. 1990. Assignment of the backbone 1H and ^{15}N NMR resonance of bacteriophage T4 lysozyme. *Biochemistry* **29**: 6341–6362.
- Muchmore, D.C., L.P. McIntosh, C.B. Russell, D.E. Anderson, and F.W. Dahlquist. 1989. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **177**: 44–73.
- Oppenheimer, N.J. and T.L. James, eds. 1989. Nuclear magnetic resonance parts A and B. *Methods Enzymol.* **176/177**.
- Otting, G., Y.Q. Qian, M. Müller, W.J. Gehring, and K. Wüthrich. 1989. Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif. *EMBO J.* **7**: 4305–4309.
- Otting, G., Y.Q. Qian, M. Billeter, M. Müller, M. Affölter, W.J. Gehring, and K. Wüthrich. 1990. Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**: 3085–3092.
- Pabo, C.O. and R.T. Sauer. 1984. Protein-DNA recognition.

Phillips et al.

- Annu. Rev. Biochem.* **53**: 293–321.
- Qian, Y.Q., M. Billeter, G. Otting, M. Müller, W.J. Gehring, and K. Wüthrich. 1989. The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: Comparison with prokaryotic repressors. *Cell* **59**: 573–580.
- Sauer, R.T., D.L. Smith, and A.D. Johnson. 1988. Flexibility of the yeast $\alpha 2$ repressor enables it to occupy the ends of its operator, leaving the center free. *Genes & Dev.* **2**: 807–816.
- Scott, M.P., J.W. Tamkun, and G.W. Hartzell. 1989. The structure and function of the homeodomain. *Biochim. Biophys. Acta* **89**: 25–48.
- Shepherd, J.C., W. McGinnis, A.E. Carrasco, R.R.M. De, and W.J. Gehring. 1984. Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. *Nature* **310**: 70–71.
- Strathern, J., B. Shafer, J. Hicks, and C. McGill. 1988. α/α -Specific repression by *MAT $\alpha 2$* . *Genetics* **120**: 75–81.
- Wolberger, C., C.O. Pabo, A.K. Vershon, and A.D. Johnson. 1991. Crystallization and preliminary X-ray diffraction studies of a *MAT $\alpha 2$* -DNA complex. *J. Mol. Biol.* **217**: 11–13.
- Wüthrich, K. 1986. *NMR of proteins and nucleic acids*. John Wiley & Sons, New York.
- Zuiderweg, E.R.P. 1990. A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity. *J. Magn. Reson.* **86**: 346–357.

Secondary structure of the homeo domain of yeast alpha 2 repressor determined by NMR spectroscopy.

C L Phillips, A K Vershon, A D Johnson, et al.

Genes Dev. 1991, **5**:

Access the most recent version at doi:[10.1101/gad.5.5.764](https://doi.org/10.1101/gad.5.5.764)

References

This article cites 33 articles, 5 of which can be accessed free at:
<http://genesdev.cshlp.org/content/5/5/764.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).