Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1 Induces an Interfacial Helix in α2†

Cynthia L. Phillips,‡,§ Martha R. Stark,‖ Alexander D. Johnson,‖ and F. W. Dahlquist*,‡

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, and Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Received March 16, 1994; Revised Manuscript Received May 18, 1994®

ABSTRACT: The homeodomain proteins a1 and α2 act cooperatively to regulate cell type specific genes in yeast. The basis of the cooperativity is a weak interaction between the two proteins which forms heterodimers that bind DNA tightly and specifically. In this paper, we examine the mechanism of heterodimerization. We show that two relatively small fragments of a1 and α2 are capable of heterodimerization and tight DNA binding. The α2 fragment contains the homeodomain followed by the natural 22 C-terminal amino acids of the protein; these 22 amino acids are unstructured in the α2 fragment. The a1 fragment contains only the homeodomain, indicating that the a1 homeodomain mediates both DNA binding and protein–protein interactions with α2. We used isotope-edited NMR spectroscopy to study the interaction in solution of these two fragments. Samples in which only the α2 fragment was uniformly labeled with ^{15}N allowed us to visualize changes in the NMR spectra of the α2 fragment produced by heterodimerization. We found that the a1 homeodomain perturbs the resonances of only the C-terminal tail of α2; moreover, contact with a1 converts a portion of this tail (residues 193–203) from its unstructured state to an α-helix, as determined by J coupling and NOE measurements. Thus the heterodimerization of two homeodomain proteins involves the specific interaction between a tail of one protein and the homeodomain of the other. This interaction is accompanied by the acquisition of secondary structure in the tail.

Transcriptional regulators often act in combination to increase their range of regulatory activity. Combinatorial control of mating-type gene expression in the yeast *Saccharomyces cerevisiae* involves the cell type specific transcriptional regulators a1 and α2 [for reviews, see Herskowitz (1989), Dolan and Fields (1991), Sprague (1990), and Johnson (1992)]. The DNA-binding specificity of α2 depends on which other transcriptional regulators are present in the cell. In both α and a/α diploid cells, an α2 homodimer acts in combination with the cell type nonspecific protein MCM1, to bind DNA target sequences upstream of a-specific genes, resulting in repression of these genes. The α2 protein has a second regulatory activity in the a/α cell, in which both α2 and a1 are present. In this case, α2 acts in combination with a1, forming a heterodimer that binds DNA target sequences upstream of haploid-specific genes, resulting in repression of these genes (see above reviews).

Structurally, a1 and α2 are related, as they both contain the homeodomain DNA-binding motif. This is a 61 amino acid segment found in many eukaryotic transcriptional regulators [for reviews, see Scott et al. (1989) and Qian et al. (1989)]. On the basis of crystallographic and NMR spectroscopic studies, the homeodomains from several transcription factors have been shown to adopt a common structure, despite great variation in their amino acid sequences (Qian et al., 1989; Kissinger et al., 1990; Phillips et al., 1991; Wolberger et al., 1991). A hydrophobic core is surrounded by three helices, one of which binds in the major groove of the target DNA. An N-terminal arm, unstructured in the free homeodomain, binds in the minor groove of target DNA.

Primarily on the basis of protease sensitivity (Sauer et al., 1988), deletion mapping (Hall & Johnson, 1987; Mak & Johnson, 1993), and an analysis of point mutations (Porter & Smith, 1986; Harashima et al., 1989; Strathern et al., 1988), α2 appears to have at least four regions, each associated with different functions. The homeodomain, located near the C-terminus, has the DNA-binding function. The region N-terminal to the homeodomain is believed to contain the homodimerization contact region, responsible for the α2 dimers that bind with MCM1 to a-specific gene target sites, as well as part of the a1 contact region (Goutte & Johnson, 1988, 1992; Harashima et al., 1989). The hinge that links the N-terminal domain to the homeodomain interacts with MCM1 (Vershon & Johnson, 1992). It is the tail of α2, C-terminal to the homeodomain, that is thought to contain the major a1 contact region required for stabilization of the a1/α2 complex with the haploid-specific gene target site (A. Mak and A. D. Johnson, submitted).

The a1 and α2 proteins heterodimerize in the absence of DNA (Mak & Johnson, 1993). The a1/α2 complex subsequently binds target DNA to coregulate haploid-specific genes (Goutte & Johnson, 1993; Dranginis, 1990). The work described in this paper provides a structural explanation for the a1/α2 interaction. To study the a1/α2 complex by NMR spectroscopy, we used the smallest available fragments of a1 and α2 that were capable of both heterodimerization and specific binding to haploid-specific gene operators. Intact a1 is 126 residues, and intact α2 is 210 residues. The α2 fragment ($\alpha2_{128-210}$) contains the C-terminal 83 amino acid residues and has been previously well characterized both structurally and functionally (Sauer et al., 1988; Phillips et al., 1991; Wolberger et al., 1991). This C-terminal region includes the homeodomain (residues 128–189) and the adjacent 22 residues that comprise the C-terminal tail of the protein. NMR experiments on the α2 fragment free in solution have shown

† C.L.P. is a recipient of an NIH Training Grant in Molecular Biology. This work was supported by grants from NIH (GM37049) to A.D.J. and from the American Cancer Society (BE-74) to F.W.D.

‡ University of Oregon.

§ Present address: Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, 3181 SW Sam Jackson Park Rd., Portland, OR 97201.

‖ University of California, San Francisco.

® Abstract published in *Advance ACS Abstracts,* July 1, 1994.

0006-2960/94/0433-9294$04.50/0

that the tail is unstructured (Phillips et al., 1991). In the cocrystal with DNA, the tail of the $\alpha 2$ fragment cannot be seen, presumably because it is disordered (Wolberger et al., 1991). Furthermore, the tail of $\alpha 2$ is protease sensitive, while the adjacent homeodomain is relatively resistant to protease attack (Mak & Johnson, 1993). The a1 fragment ($a1_{66-126}$) used in these NMR experiments contains only the homeodomain. These two fragments, when mixed, are adequate for subsequent specific binding to haploid-specific gene operators (see Results). To date, no structural studies of the a1 protein have been reported.

In this paper we focus on the effects of heterodimerization on the $\alpha 2_{128-210}$ fragment. We have taken advantage of the previously established resonance assignments for the same $\alpha 2$ fragment (Phillips et al., 1991). We have also used ^{15}N-edited experiments to simplify otherwise exceedingly complex spectra. In this case, the $\alpha 2_{128-210}$ was uniformly ^{15}N-labeled, whereas the $a1_{66-126}$ remained unlabeled. The ^{15}N-labeled protein was studied with ^{15}N-edited NMR experiments designed to filter out most signals from the unlabeled a1. By comparing the spectra of $\alpha 2$ in the absence and presence of a1, we determined how the structure of $\alpha 2$ changed due to heterodimer formation. The binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ does not occur with pseudo-2-fold symmetry via complementary homeodomain contacts between a1 and $\alpha 2$. Rather, the homeodomain proper of $\alpha 2_{128-210}$ is unaffected by the binding of the a1 homeodomain, but dramatic structural changes are seen in the C-terminal tail of the $\alpha 2$ fragment. These changes in the tail are most probably a result of direct contact with the a1 homeodomain. These data also show that heterodimer formation induces an interfacial helix in the previously unstructured C-terminal tail of $\alpha 2_{128-210}$.

EXPERIMENTAL PROCEDURES

Plasmids. The $a1_{66-126}$ protein was purified from *Escherichia coli* cells containing the plasmid pCW/K66. This plasmid encodes residues 66–126 of a1, under the control of tandem P_{tac} promoters. The original $a1_{66-126}$ plasmid was under the control of the T7 promoter in the T7 expression system of Studier and Moffat (1986) and is described in the Electrophoretic Mobility Shift Assay section below. To achieve higher levels of protein expression, the a1 fragment was subsequently cloned into the *Nde*I and *Xba*I sites of pCWori+ (Muchmore et al., 1989; Gegner & Dahlquist, 1991) and subsequently transformed into *E. coli* TB1 cells. The construction of the resulting plasmid was confirmed by restriction digestion and DNA sequencing. The $\alpha 2_{128-210}$ plasmid was constructed as previously described (Phillips et al., 1991).

Protein Purification. The $a1_{66-126}$ protein was purified to >95% homogeneity from *E. coli* TB1 cells as follows. Each liter of LB media with 100 mg/L ampicillin was inoculated to an initial density of 1×10^7 cells/mL. These were grown at 37 °C to a density of 5×10^8 cells/mL, at which time the cells were induced to a concentration of 0.4 mM IPTG. The cells were harvested 6 h after induction and resuspended in 500 mM NaCl, 25 mM sodium phosphate, pH 6.5, 1 mM EDTA, 10 mM β-mercaptoethanol, 0.01% NaN_3, and 1 mM phenylmethanesulfonyl fluoride (PMSF) using 25 mL of lysis buffer per liter of harvested media. The cells were sonicated and spun at 20000*g* for 20 min. The supernatant was dialyzed overnight against 25 mM sodium phosphate, pH 6.5, 10 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol, 0.2 mM PMSF, and 0.01% NaN_3 and loaded on a column of CM Sepharose CL 6B (Sigma). The $a1_{66-126}$ protein was eluted with a gradient from 10 to 600 mM NaCl and found to be >95% pure as judged by SDS–PAGE. The $a1_{66-126}$ protein was concentrated to approximately 0.1 mM using an Amicon concentrator with a 3000 molecular weight cutoff filter and further concentrated to 1–5 mM using 3000 molecular weight cutoff concentrators from Filtron. The molecular mass of $a1_{66-126}$ is 7229 Da, and we used the calculated ϵ_{280} of 5700 cm^{-1} M^{-1} to determine the protein concentration. The yield of $a1_{66-126}$ was approximately 7 mg/L.

The $\alpha 2_{128-210}$ protein was purified to >95% homogeneity and concentrated as described previously (Phillips et al., 1991). Uniform ^{15}N labeling of the $\alpha 2_{128-210}$ was carried out as previously described (Phillips et al., 1991). Prior to use, both the $a1_{66-126}$ and $\alpha 2_{128-210}$ proteins were dialyzed against 25 mM potassium phosphate buffer, pH 4.5, with 100 mM KCl, 0.01% NaN_3, and D_2O added to 5% for the spectrometer lock.

Electrophoretic Mobility Shift Assay. DNA binding assays were performed using purified $a1_{66-126}$ and $\alpha 2_{128-210}$ proteins as previously described for full-length a1 and $\alpha 2$ (Goutte & Johnson, 1993). A DNA fragment coding for residues 66–126 of a1 was generated by polymerase chain reaction using oligonucleotides that introduced an *Nde*I site at the 5′ end and an *Xho*I site at the 3′ end of the fragment. pMSK66 was constructed by cloning the *Nde*I–*Xho*I fragment into pHB40P, a derivative of pET-3a, under the control of the T7 promoter (Studier & Moffat, 1986). The $a1_{66-126}$ protein was overexpressed and purified from BL21(DE3)-pLysS *E. coli* cells containing the plasmid pMSK66. Cells were grown to an OD_{600} of 0.5 and induced with a final concentration of 0.4 mM IPTG. The cells were harvested after 5 h, frozen in liquid nitrogen, and then resuspended in cold 100 mM Tris-HCl, pH 8.0, 1 mM EDTA, 10 mM β-mercaptoethanol, 500 mM NaCl, and 1 mM PMSF. The cells were sonicated on ice and spun at 30000*g* for 40 min. The supernatant was dialyzed overnight at 4 °C against 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.28 mM β-mercaptoethanol, and 100 mM NaCl, and then dialyzed for another 3 h against the same buffer but with 50 mM NaCl instead of 100 mM NaCl. The supernatant was cleared by spinning at 30000*g* for 20 min and then loaded onto an SP-C50 Sephadex column (Pharmacia) at 4 °C. A salt gradient from 0.5 to 1.0 M NaCl was used to elute the protein from the column. $a1_{66-126}$ came off the column between 325 and 425 mM NaCl and was >90% pure.

$\alpha 2_{128-210}$, a gift of Andrew Vershon, was overexpressed and purified from *E. coli* cells containing the plasmid pAV105 (Vershon & Johnson, 1992).

NMR Spectroscopy. All spectra were acquired at 25 °C on a General Electric Omega 500-MHz spectrometer operating at 11.9 T. The chemical shifts were set relative to an external proton reference of sodium 2,2-dimethyl-2-silapentane-5-sulfonate at 0.0 ppm and an external nitrogen standard of $^{15}NH_4Cl$ at 24.93 ppm relative to NH_3 (Levy & Lichter, 1979). The spectra were recorded with a spectral width of 6410 Hz in the ^1H dimensions and 3333 Hz in the ^{15}N dimensions (except for the 3D-HSMQC-NOESY, where it was 1300 Hz). The recycle times, including acquisition, were usually 1 s. In the ^{15}N experiments a delay of 4.8 instead of 5.4 ms was used as the nominal $(2J_{NH})^{-1}$ time period to reduce the loss of signal due to relaxation. Spectra were analyzed using FELIX software from Hare Research.

To observe line width and chemical shift changes during a titration of ^{15}N-labeled $\alpha 2_{128-210}$ with $a1_{66-126}$, single-bond ^1H–^{15}N correlation spectra were measured by heteronuclear single–multiple quantum coherence (HSMQC) (Zuiderweg

1990). Spectra were collected as described by McIntosh et al. (1990) with minor differences. The experiments were taken with 1024 complex data points in the t_2 domain and 128–256 complex increments in t_1, and they typically required 2–6 h to acquire. Spectra were processed with a 60° shifted sine-bell apodization and zero-filled to 2048 and 1024 data points in the ^1H and ^{15}N dimensions, respectively. Spectra of dilute protein where the H_2O signal was not well suppressed were processed with a baseline correction (Burg smoothing of the FID or a polynomial baseline correction) as well. ^{15}N-ω_2-edited two-dimensional COSY and NOESY spectra of the $a1_{66-126}/\alpha2_{128-210}$ complex were recorded and processed as described for T4 lysozyme by McIntosh et al. (1990). NOESY spectra were collected with 100- and 150-ms mixing times.

The 3D-HSMQC-NOESY experiment is a variation of the 3D-HMQC-NOESY experiment (Kay et al., 1989b; Zuiderweg & Fesik, 1989), in which an HSMQC ^1H–^{15}N pulse sequence is used to generate ^1H–^{15}N coherence. The mixing time was 150 ms. The spectral widths used were 6410 (^1H, ω_1), 1300 (^{15}N, ω_2), and 6410 Hz (^1H, ω_3). The 3D data matrix contained 256 (real) × 64 (real) × 512 (complex) points. Quadrature in the ω_1 and ω_2 dimensions was obtained using time-proportional phase incrementation (TPPI). Sixteen scans were collected for each increment. The recycle time was 750 ms not including the acquisition time. The spectrum was processed with a 60° shifted sine-bell apodization in ω_3 and ω_2 and a 70° shifted sine-bell apodization in ω_1. A polynomial baseline correction was applied in the direct dimension. The matrix was zero-filled to 1024 × 512 × 128 real points.

The $^3J_{HN-H\alpha}$ coupling constants of the $a1_{66-126}/\alpha2_{128-210}$ complex were measured as described previously (Phillips et al., 1991) using an HMQC-J experiment (Kay et al., 1989a; Forman-Kay et al., 1990; Kay & Bax, 1990). The measured $^3J_{HN-H\alpha}$ coupling constants have errors of up to approximately 0.5 Hz, based on the ratio $J_{measured}/J_{actual}$ calculated by Kay and Bax (1990) for a typical resonance with an 11-Hz line width in both the free and bound forms of $\alpha2_{128-210}$.

The semiselective and nonselective T_1's for the amide resonances of the complexed $\alpha2_{128-210}$ were measured using variations on the HSMQC experiment, using an inversion recovery sequence in which the final read pulse was replaced by the 2D-HSMQC sequence (Fraenkel et al., 1990; Valensin et al., 1982).

RESULTS

Cooperative DNA Binding by a1 and α2 Homeodomain Fragments. Full-length a1 and $\alpha2$ bind the haploid-specific gene (hsg) operator in a strongly cooperative fashion. The two proteins form a heterodimer in solution with a K_d of approximately 10^{-6} M (A. Mak and A. D. Johnson, submitted). This dimer then binds the hsg operator with a K_d of approximately 10^{-10} M with both homeodomains making contact with the DNA (Goutte & Johnson, 1993). The protein $\alpha2$, in the absence of a1, binds weakly to the hsg operator ($K_d \sim 10^{-6}$ M), while a1 alone shows no detectable specific binding under the same conditions (Goutte & Johnson, 1993).

We found that the cooperative binding of a1 and $\alpha2$ to the hsg operator could be reconstituted *in vitro* using short fragments of both proteins. For a1, the homeodomain alone (residues 66–126 of the full-length protein) was sufficient for this cooperative binding. For $\alpha2$, both the homeodomain and the 20 amino acid C-terminal tail (residues 128–210) were required. For example, deletion of the C-terminal tail destroyed the cooperative binding of the protein fragments to DNA [not shown; see also Mak and Johnson (1993)].

The experiment in Figure 1 illustrates the cooperative DNA binding by these minimal fragments. Lanes 13–18 and lanes

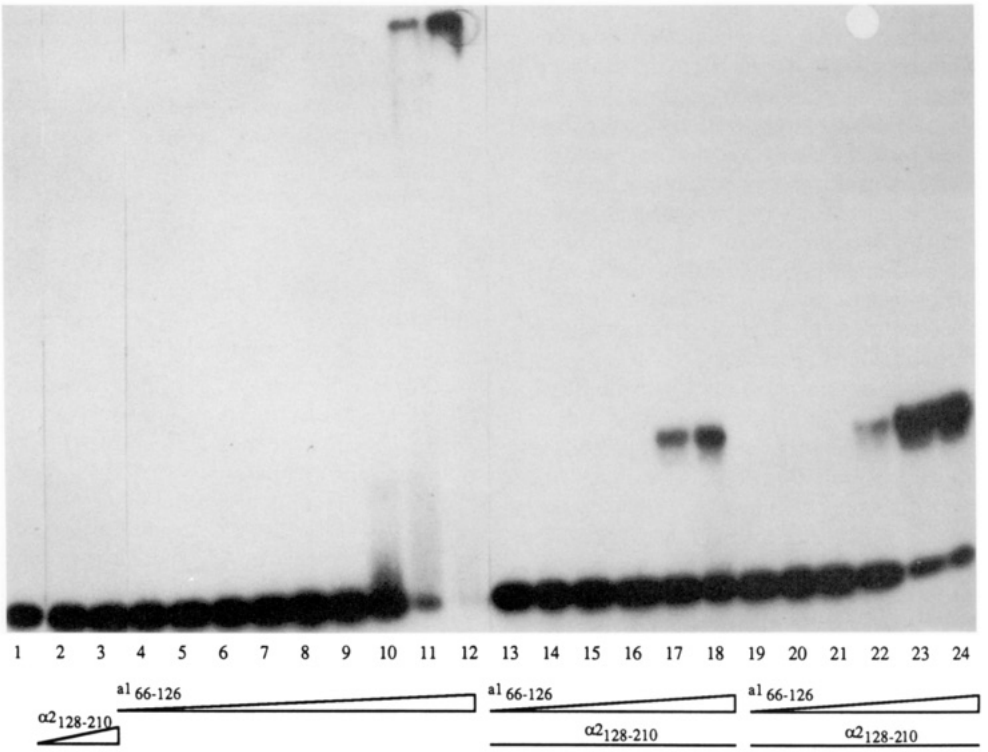

FIGURE 1: Cooperative binding of $a1_{66-126}$ and $\alpha2_{128-210}$ to the hsg operator. A 37-bp oligonucleotide duplex containing an hsg operator was used in the electrophoretic mobility shift experiment. Lane 1: labeled operator alone. Lanes 2 and 3: 3×10^{-9} M and 3×10^{-8} M $\alpha2_{128-210}$ protein, respectively. Lanes 4–12: successive 3-fold increases of $a1_{66-126}$ protein beginning with a concentration of 10^{-8} M in lane 4. Lanes 13–24 show titrations of $a1_{66-126}$ protein with a constant concentration of $\alpha2_{128-210}$ protein. Lanes 13–18: 3×10^{-9} M $\alpha2_{128-210}$ with successive 3-fold increases of $a1_{66-126}$ beginning with a concentration of 10^{-8} M. Lanes 19–24: 3×10^{-8} M $\alpha2_{128-210}$ with the same $a1_{66-126}$ titration as in previous lanes 13–18.

19–24 show the effect on DNA binding of increasing concentrations of the a1 homeodomain in a constant concentration of the α2 fragment. Both titrations show efficient operator binding at concentrations at which neither fragment alone binds to DNA (lanes 2–12). As can be seen in lanes 14–18, there is significant formation of a slower migrating form of the DNA as a result of increasing concentrations of $a1_{66-126}$ at a constant concentration of $\alpha2_{128-210}$ (3×10^{-9} M). At a 3-fold higher concentration of $\alpha2_{128-210}$, formation of the complex is observed at a corespondingly lower concentration of $a1_{66-126}$ (lanes 19–24). From these data we estimate the overall K_d for the reaction

$$a1_{66-126} + \alpha2_{128-210} + \text{hsg operator} \rightarrow a1/\alpha2/\text{operator}$$

to be approximately 10^{-14}–10^{-15} M^2. This value is approximately 10–100-fold weaker than that observed for the cooperative DNA binding of full-length a1 and α2 [see Goutte and Johnson (1993)]. We attribute this difference to protein–protein interactions made between a1 and the amino terminus of α2, which has been deleted in these experiments (Goutte & Johnson, 1988).

We conclude from these results that the a1 homeodomain is sufficient to bind the hsg operator cooperatively with α2. We know that the a1 homeodomain contacts the operator in the $a1/\alpha2$/operator complex (Goutte & Johnson, 1993); these results suggest that the a1 homeodomain is also contacted by α2. For α2, both the homeodomain and the C-terminal 20 residue tail are required for cooperative DNA binding. Since this tail does not appear to contact DNA (Mak & Johnson, 1993), the simplest model is one where the tail makes direct contact with the a1 homeodomain. In order to test these ideas experimentally, we examined the behavior of the two homeodomain fragments in solution by NMR methods.

The a1 Homeodomain, $a1_{66-126}$, Forms a Heterodimer with $\alpha2_{128-210}$ in Solution. Figure 2 shows a region of the ^{15}N–^{1}H correlated spectrum (HSMQC) of 0.5 mM uniformly ^{15}N-labeled $\alpha2_{128-210}$ as it is titrated with unlabeled $a1_{66-126}$. Since only the $\alpha2_{128-210}$ fragment is ^{15}N labeled, we observe only the $\alpha2_{128-210}$ resonances and how they are affected by addition of $a1_{66-126}$. Over the course of the titration, some $\alpha2_{128-210}$ resonances shift to new positions and experience line broadening, while others do not. For example, resonances arising from Asn 178, Thr 159, and Ser 181 in the homeodomain remain in the same position. Other resonances, such as those arising from Asp 198 and Ser 201 (shaded in the figure), shift steadily across the spectrum from a "free" position to a "bound" position. On the basis of the chemical shift and line shape differences between the spectrum of the free protein and those observed in the presence of various amounts of $a1_{66-126}$, we can estimate the exchange lifetimes for the protein–protein association. Line shapes and resonance positions were simulated using well-known relationships for a two-site exchange process (Sandstrom, 1982), suggesting that $\alpha2_{128-210}$ interconverts between the free and the bound forms approximately 300 times per second.

From the resonances that do shift during the $a1_{66-126}$ titration, we can calculate a binding constant for $a1_{66-126}$ complexing with $\alpha2_{128-210}$. Figure 3 shows the chemical shift change of the representative resonance Asp 198 as a function of the concentration of $a1_{66-126}$ added to a constant concentration of $\alpha2_{128-210}$. The smooth line drawn through the data points is a theoretical curve based on a least-squares fit of the data to a single binding site with a dissociation constant (K_d) of 2×10^{-4} M and a total shift of 196 Hz (0.39 ppm) in a 1:1 complex. Similar least-squares analysis of other resonances

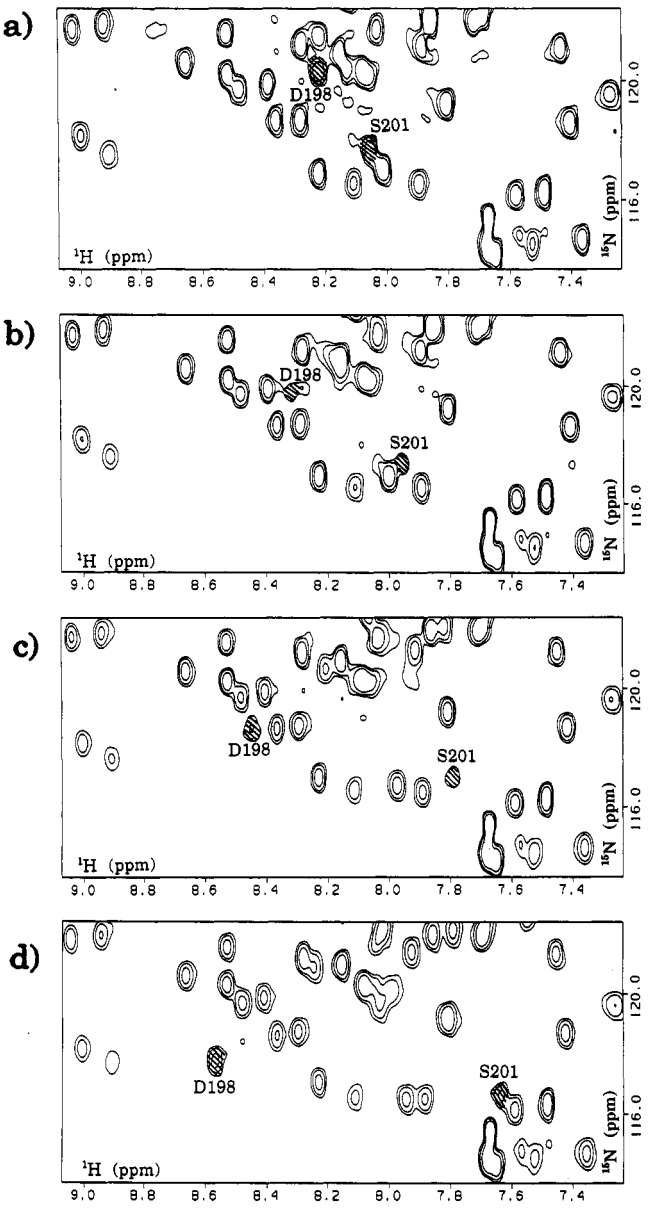

FIGURE 2: A small region of the HSMQC spectra during titration of $\alpha2_{128-210}$ by $a1_{66-126}$. The $\alpha2_{128-210}$ concentration was held constant at 0.5 mM. The resonance positions of Asp 198 and Ser 201 (shown shaded) are labeled as they shift during $a1_{66-126}$ addition. The spectra were taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. The spectra were apodized identically and are drawn at the same contour level. (a) $a1_{66-126}/\alpha2_{128-210} = 0$ (0 mM $a1_{66-126}$). (b) $a1_{66-126}/\alpha2_{128-210} = 0.25$ (0.125 mM $a1_{66-126}$). (c) $a1_{66-126}/\alpha2_{128-210} = 1.0$ (0.5 mM $a1_{66-126}$). (d) $a1_{66-126}/\alpha2_{128-210} = 4.0$ (2 mM $a1_{66-126}$).

observed to shift through the titration gave the same binding constant and 1:1 stoichiometry. Subsequent experiments on the structure of the bound form of $\alpha2_{128-210}$ were performed at an $a1_{66-126}$ concentration of 5 mM and an $\alpha2_{128-210}$ concentration of 3 mM, conditions where approximately 95% of the $\alpha2_{128-210}$ is bound to $a1_{66-126}$.

The spectral changes that occur when $a1_{66-126}$ is added to $\alpha2_{128-210}$ appear to be due to interaction between $a1_{66-126}$ and a particular region of $\alpha2_{128-210}$. Some of the resonances are perturbed, while most are not, indicating that only a small region of the protein is affected by $a1_{66-126}$ binding. Also, the total number of $\alpha2_{128-210}$ amide resonances is conserved through the titration, reflecting a fast exchange average of the free and bound forms of $\alpha2_{128-210}$. The moderate 0.2 mM

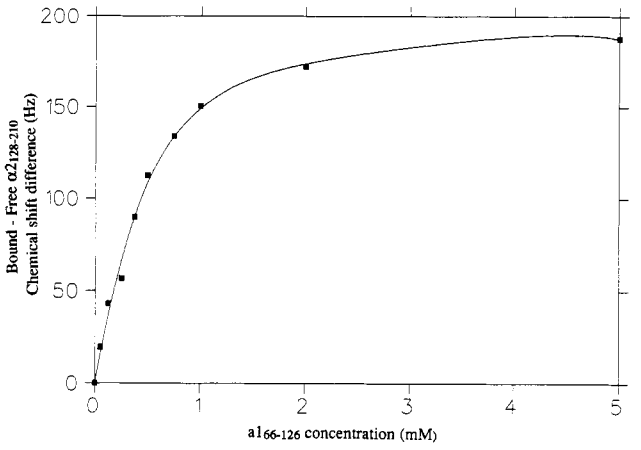

FIGURE 3: Chemical shift difference (Hz) between the bound and free forms of $\alpha 2_{128-210}$ for the representative resonance Asp 198, as a function of the concentration of added $a1_{66-126}$. The concentration of $\alpha 2_{128-210}$ was held constant at 0.5 mM. The smooth line drawn through the data points is a theoretical curve based on a least-squares fit of the data to a single binding site of $K_d = 2 \times 10^{-4}$ M and a total shift of 196 Hz (0.39 ppm).

K_d for heterodimer dissociation suggests that interactions distinct from simple electrostatics are important for heterodimer formation, especially since the two proteins are both highly positively charged. The $a1_{66-126}$ protein has a net charge of +13, and $\alpha 2_{128-210}$ has a net charge of +9, at neutral pH.

The spectra of $\alpha 2_{128-210}$ alone do not change over a concentration range of 0.2–6 mM (Phillips et al., 1991). Thus the changes we observe in the $\alpha 2_{128-210}$ spectra upon addition of $a1_{66-126}$ can only be attributed to the presence of $a1_{66-126}$ and are not a nonspecific effect of increasing total protein concentration. Finally, the HSMQC spectrum of 0.5 mM $\alpha 2_{128-210}$ with 5 mM $a1_{66-126}$ and the HSMQC spectrum of 3 mM $\alpha 2_{128-210}$ and 5 mM $a1_{66-126}$ are virtually indistinguishable in spite of the differing protein ratios (compare Figures 2d and 4). If a complex between $\alpha 2_{128-210}$ and $a1_{66-126}$ were formed with more than a single $a1_{66-126}$ per $\alpha 2_{128-210}$, this behavior would not be observed. This observation reinforces our conclusion above, based on least-squares fitting of the chemical shift data (see Figure 3), that $a1_{66-126}$ and $\alpha 2_{128-210}$ bind in a specific 1:1 complex.

Binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ Affects Only the C-Terminal Region of $\alpha 2_{128-210}$. With specific binding of the $a1_{66-126}$ and $\alpha 2_{128-210}$ fragments established, we turned to analyzing which $\alpha 2_{128-210}$ backbone resonances were affected by heterodimer formation. The 17-kDa $a1_{66-126}/\alpha 2_{128-210}$ complex is large for NMR characterization. Also, both $a1_{66-126}$ and $\alpha 2_{128-210}$ have limited chemical shift dispersion in their H^N and H^α protons, worsening the problem of spectral overlap (data for $a1_{66-126}$ not shown). Use of the added ^{15}N dimension in 2D-HSMQC spectra affords an advantage in dispersion over H^1-only spectra. In addition, the use of ^{15}N-edited experiments to observe only the uniformly ^{15}N-labeled $\alpha 2_{128-210}$ in the presence of unlabeled $a1_{66-126}$ further decreases the complexity of the $a1_{66-126}/\alpha 2_{128-210}$ heterodimer spectra. This ability to filter out signal from $a1_{66-126}$ in the $a1_{66-126}/\alpha 2_{128-210}$ complex was crucial to the assignment of the bound $\alpha 2_{128-210}$ backbone resonances.

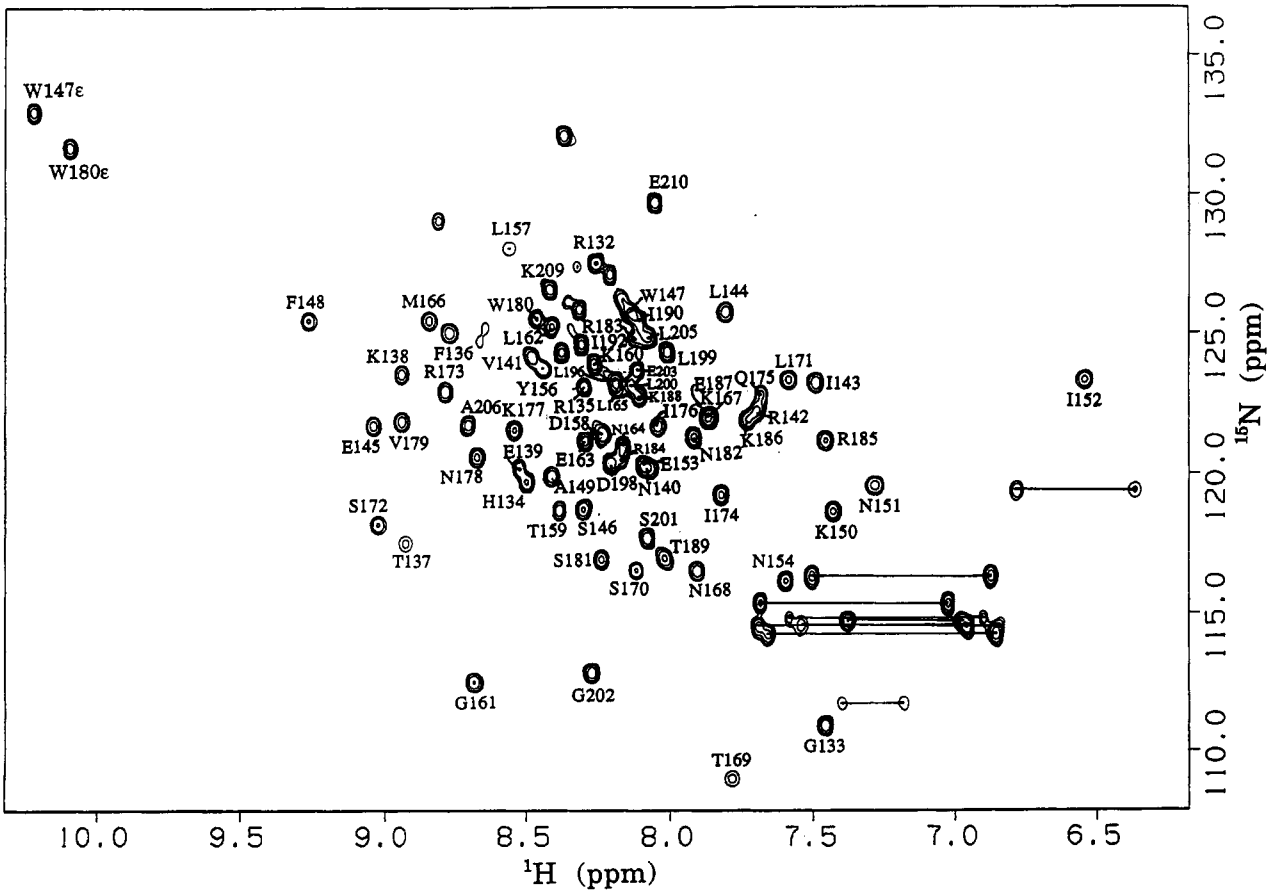

FIGURE 4: Amide region of the HSMQC of uniformly ^{15}N-labeled free $\alpha 2_{128-210}$ at 4 mM. Each cross peak arises from a proton directly bonded to a ^{15}N. In the lower right corner of the spectrum, each pair of side-chain $^{15}NH_2$'s is connected by a line. The assigned cross peaks are labeled. The spectrum was taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock.

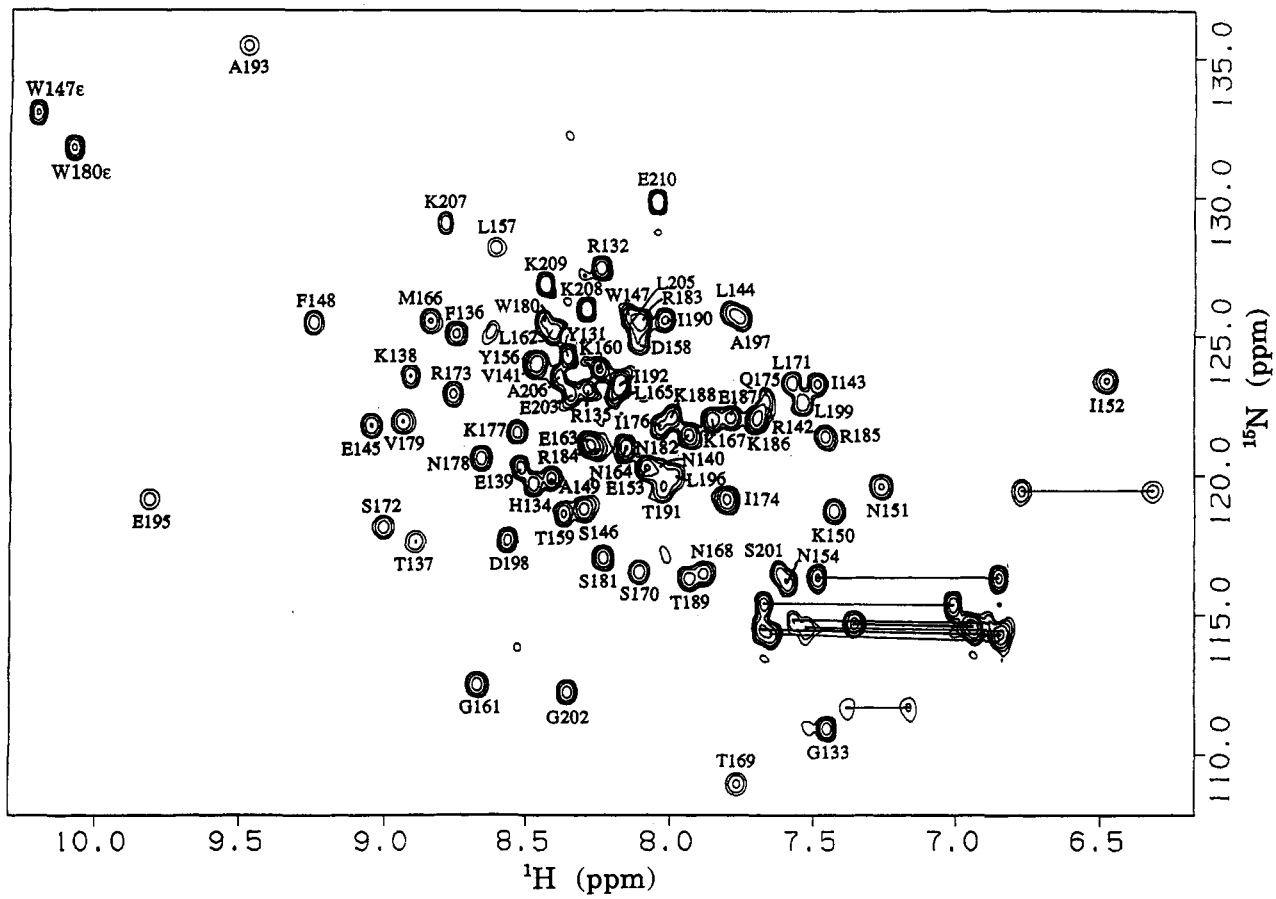

FIGURE 5: Amide region of the HSMQC of uniformly ^{15}N-labeled $\alpha 2_{128-210}$ bound to $a1_{66-126}$. The $\alpha 2_{128-210}$ is 3 mM, and the $a1_{66-126}$ is 5 mM. Greater than 95% of the $\alpha 2_{128-210}$ is in the bound form. In the lower right corner of the spectrum, each pair of side-chain $^{15}NH_2$'s is connected by a line. The assigned cross peaks are labeled. The assignments for K208 and K209 are ambiguous and may be reversed. The spectrum was taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. The resonances are broader in this figure than those in Figure 4 primarily because the contour levels shown are lower.

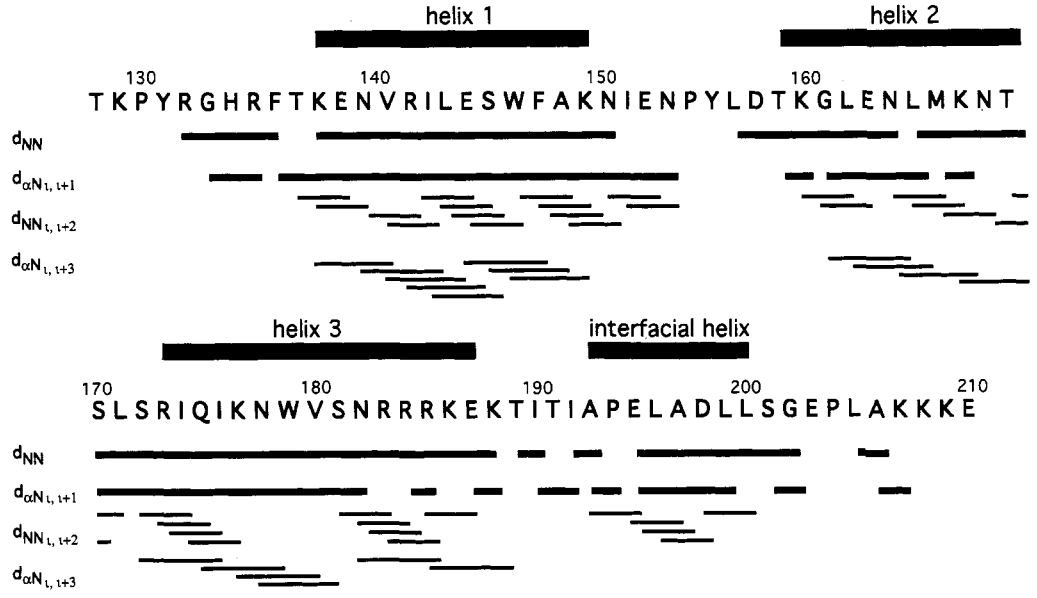

FIGURE 6: $\alpha 2_{128-210}$ amino acid sequence with sequential and intermediate-range NOE connectivities. Below the sequence, lines of varying widths indicate NOE connectivities. The uppermost two sets of broader lines indicate the d_{NN} and $d_{\alpha N}$ sequential connectivities. Below, the narrower sets of lines indicate the $d_{NN}(i, i+2)$ and the $d_{\alpha N}(i, i+3)$ connectivities. All NOE connectivities were observed in 3D-HSMQC-NOESY.

Figures 4 and 5show the full amide regions of the HSMQC spectra for the free and bound forms, respectively, of $\alpha 2_{128-210}$. The labeled resonances in Figure 5 were assigned by taking advantage of the previously determined assignments of free $\alpha 2_{128-210}$ (Figure 4). Due to spectral overlap, we were unable to simply track and reassign all of the shifting resonances in the bound form of $\alpha 2_{128-210}$ over the course of the titration. 3D-HSMQC-NOESY (not shown) of the $a1_{66-126}/\alpha 2_{128-210}$ complex, in which $\alpha 2_{128-210}$ (but not $a1_{66-126}$) was uniformly ^{15}N labeled, helped resolve resonances and allowed assignment

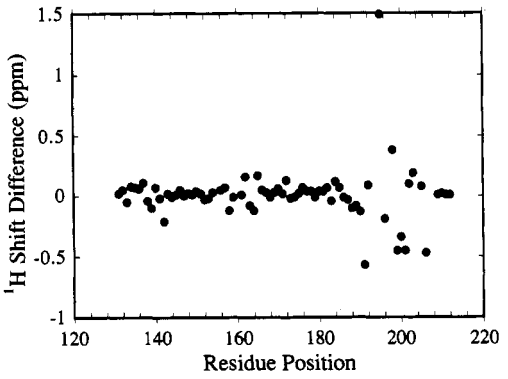

FIGURE 7: Amide ^{1}H chemical shift difference between bound and the free $\alpha 2_{128-210}$ as a function of residue position. Large chemical shift perturbations are seen in the C-terminal tail. Data is from the HSMQC spectra.

of the spectrum. Sequential amide resonances were traced using H^N_i–H^N_{i+1} (d_{NN}) connectivities in the 3D-HSMQC-NOESY spectrum. Most of these d_{NN} sequential connectivities were corroborated with sequential H^N–H^α ($d_{\alpha N}$) connectivities of the (i, i+1) and (i, i+3) types. A ^{15}N-edited 2D-DQF-COSY experiment of the $a1_{66-126}/\alpha 2_{128-210}$ complex was used to identify the shifted H^N–H^α cross peaks in the 3D-HSMQC-NOESY. The backbone NOE connectivities are summarized in Figure 6. All of the amide and most of the H^α resonances (not shown) in the bound form of $\alpha 2_{128-210}$ were assigned by this procedure (Figure 5).

Remarkably, the amide resonances from residues 132–189 that comprise the homeodomain of $\alpha 2_{128-210}$ remain virtually unperturbed by the binding of $a1_{66-126}$. In contrast, most of the resonances in the C-terminal tail region, following the homeodomain of $\alpha 2_{128-210}$, are affected by $a1_{66-126}$ binding. Figure 7is a plot of the change in $\alpha 2_{128-210}$ $^1H^N$ chemical shift as a function of residue position. The greatest chemical shift changes are seen for amide resonances belonging to residues 193–206. Residues 189–192, immediately following the C-terminal end of the homeodomain, show less dramatic perturbations. Amide chemical shift is extremely sensitive to environment and secondary structure. We conclude that the secondary structure of the $\alpha 2_{128-210}$ homeodomain proper is not affected by $a1_{66-126}$ binding. Absence of chemical shift change in the homeodomain argues strongly that the $a1_{66-126}$ homeodomain does not contact the $\alpha 2$ homeodomain. Conversely, $a1_{66-126}$ probably contacts the C-terminal tail of $\alpha 2_{128-210}$, causing the large chemical shift changes.

In the HSMQC spectrum of the free $\alpha 2_{128-210}$ (Figure 4), the C-terminal tail amide resonances are clustered in the center, where they are sharp and have chemical shifts consistent with a random coil structure. NOE connectivities and hydrogen exchange and $^3J_{HN-H\alpha}$ coupling constant data on the free $\alpha 2_{128-210}$ also indicate that this region is essentially unstructured (Phillips et al., 1991). In contrast, the HSMQC spectrum of the bound $\alpha 2_{128-210}$ (Figure 5) shows that the C-terminal tail amide resonances are considerably more disperse, and the resonances from amides 193 and 195 become the most extreme downfield shifted amide resonances in the spectrum. Such downfield chemical shifts in the proton dimension are very likely due to amide hydrogen bond formation.

In spite of a near doubling of molecular weight in the $a1_{66-126}/\alpha 2_{128-210}$ complex, the apparent line widths of the unperturbed resonances remain basically unchanged (see Figure 2, in which each spectrum in the titration is apodized identically and is drawn at the same contour level). The amide

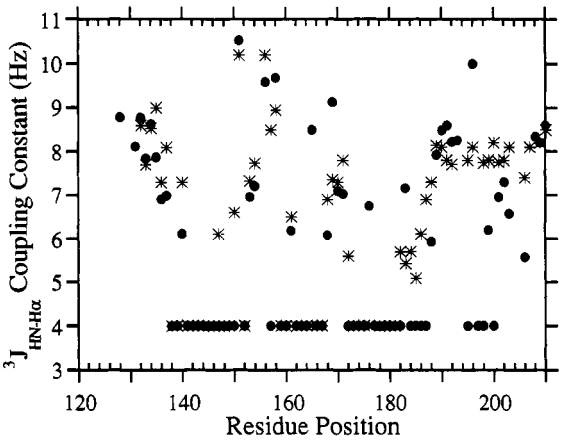

FIGURE 8: $^3J_{HN-H\alpha}$ coupling constants (Hz) as a function of residue position for free $\alpha 2_{128-210}$ (asterisks) and $a1_{66-126}$-bound $\alpha 2_{128-210}$ (filled circles). No splittings could be measured for the $^3J_{HN-H\alpha}$ coupling constants shown as 4 Hz. These $^3J_{HN-H\alpha}$ couplings are smaller than the 5.5–6.0-Hz minimum for detection and were given the value of 4 Hz for presentation purposes. $^3J_{HN-H\alpha}$ coupling constants for overlapping or unassigned resonances are not shown.

resonance line widths do not appear to be greatly influenced by the increase in correlation time upon complex formation. This may reflect relatively independent motion of the two monomers in the heterodimer. However, resonances arising from the contact region are somewhat broader than those arising from free $\alpha 2_{128-210}$. This broadening of the resonances in the contact region is probably due to chemical exchange with $a1_{66-126}$.

Structure of the $\alpha 2_{128-210}$ C-Terminal Tail in the $a1_{66-126}/\alpha 2_{128-210}$ Complex. (A) $^3J_{HN-H\alpha}$ Coupling Constants. The scalar coupling constant, $^3J_{HN-H\alpha}$, is dependent on the Ramachandran dihedral angle ϕ of the protein backbone (Pardi et al., 1983; Wüthrich, 1986), and thus is sensitive to secondary structure. Figure 8 shows the $^3J_{HN-H\alpha}$ coupling constants for the free and bound forms of $\alpha 2_{128-210}$ as a function of residue number. These couplings were obtained using an HMQC-J experiment (Kay & Bax, 1990). Due to nonlinear effects of line width (10–11 Hz for most of the amide resonances) in this experiment, observed $^3J_{HN-H\alpha}$ coupling constants may be larger or smaller by about 0.5 Hz than the actual values (Kay & Bax, 1990).

A contiguous series of residues with $^3J_{HN-H\alpha}$ coupling constants of less than 6.5 Hz is strongly indicative of helical structure (Pardi et al., 1983; Wüthrich, 1986). No other regular secondary structure features such a series of low $^3J_{HN-H\alpha}$ coupling constants. The amides of the three helices of the homeodomain in both the free and bound $\alpha 2_{128-210}$ feature these low $^3J_{HN-H\alpha}$ coupling constants. This further confirms that the helical structure of the $\alpha 2_{128-210}$ homeodomain is unperturbed by $a1_{66-126}$ binding. Differences in the $^3J_{HN-H\alpha}$ coupling constants between the free and bound forms of $\alpha 2_{128-210}$ only become apparent in the tail region of $\alpha 2_{128-210}$. The $^3J_{HN-H\alpha}$ couplings were greater than 6.5 Hz for all of the C-terminal tail in the free $\alpha 2_{128-210}$. The $^3J_{HN-H\alpha}$ coupling constants of residues 193–203 of the bound form of $\alpha 2_{128-210}$ decrease to values consistent with a helical structure.

(B) Short-Range NOE Connectivities. In helices, the sequential d_{NN} connectivites are very strong, while in β-sheets and other extended structures, where sequential amide protons are not as close in space, the connectivities are weaker (Wüthrich, 1986). We wanted to compare the strength of the sequential d_{NN} NOEs in the tail region of bound $\alpha 2_{128-210}$ with the NOEs of the amide protons in helices in the

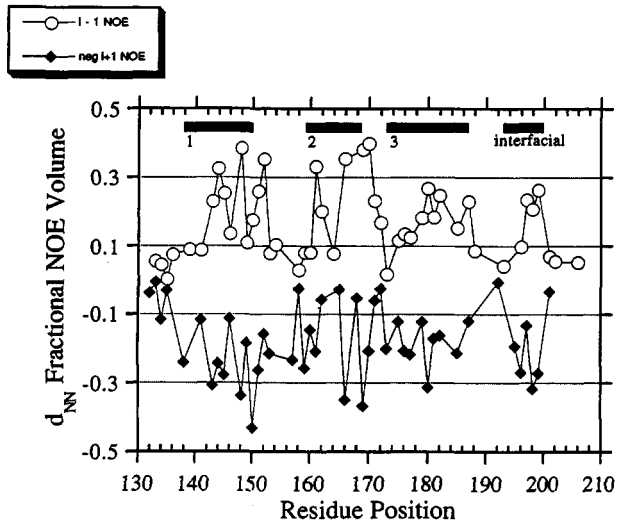

FIGURE 9: Ratio of the $d_{NN}(i, i+1)$ and $-(i, i-1)$ NOE volume to the diagonal peak volume as a function of residue position in $\alpha 2_{128-210}$ bound to $a1_{66-126}$. The $(i, i+1)$ fractional NOE volumes are shown as negative. The $(i, i+1)$ NOE values are shown as filled diamonds; the $(i, i-1)$ NOE values are shown as open circles. All data was from the 3D-HSMQC-NOESY. The positions of the helices are marked by filled bars at the top of the graph.

homeodomain that we observed to be unperturbed by $a1_{66-126}$ binding. Figure 9 shows the relative volumes of the sequential d_{NN} NOE cross peaks for the 3D-HSMQC-NOESY spectrum of bound $\alpha 2_{128-210}$. Each d_{NN} volume was divided by the volume of its own resolved "diagonal" peak for normalization. To ensure that we were measuring comparable NOEs at the NOESY mixing time of 150 ms, we measured the nonselective T_1's for the amides to be uniformly 900 ± 100 msec, and the semi-selective T_1's for the amides (also fairly uniform throughout the bound $\alpha 2_{128-210}$ form) at 250 ± 50 ms (see Experimental Procedures for a description of these experiments). Resonances 194–200 in the tail region of bound $\alpha 2_{128-210}$ have strong sequential d_{NN} NOEs, just as in the three homeodomain helices. The sequential d_{NN} NOEs of the N-terminal arm region and of the turns between the helices remain weak, as they are in free $\alpha 2_{128-210}$ (Phillips et al., 1991). Therefore, spin diffusion is not a complicating factor in this experiment on the larger complex, even at the 150-ms mixing time. These NOE cross peak volume data suggest that residues 194–200 are helical.

(C) Medium-Range NOE Connectivities. Very weak d_{NN} NOEs between an amide and its neighbor two amides away in the sequence are also commonly observed in helical structures [$(i, i+2)$ connectivities]. Figure 6 shows that these $d_{NN}(i, i+2)$ connectivities are present in part of the tail of $\alpha 2_{128-210}$ when it is bound to $a1_{66-126}$, as well as in the three helical regions of the homeodomain. The positions of these $d_{NN}(i, i+2)$ connectivities in the tail of bound $\alpha 2_{128-210}$ coincide with the position of the other helical types of NOE and the helical $^3J_{HN-H\alpha}$ coupling constants.

An NOE connectivity between an amide proton and the H^α three or four N-terminal to it in the sequence [$d^{\alpha N}(i, i+4)$ and $-(i, i+3)$] is also a marker for helical structure. These are observed in the three homeodomain helices [Figure 6; $(i, i+4)$ connectivities not shown]. Unfortunately, even with the increased dispersion of the 3D-HSMQC-NOESY, these connectivities are degenerate for most of the tail of bound $\alpha 2_{128-210}$. There are NOE cross peaks consistent with $d_{\alpha N}(i, i+3)$ and $-(i, i+4)$ connectivities, but they are degenerate with $(i, i+1)$ or other types of connectivities in this region of this experiment. We cannot be certain that the $d_{NN}(i, i+3)$ and $-(i, i+4)$ connectivities are present. However, the $^3J_{HN-H\alpha}$ coupling and the sequential d_{NN} and $d_{NN}(i, i+2)$ connectivities strongly suggest that $\alpha 2_{128-210}$ residues 194–206 of the C-terminal tail are helical in the presence of $a1_{66-126}$.

DISCUSSION

The $\alpha 2$ protein binds to two different classes of DNA target sites. In the case of a-specific gene target sites, $\alpha 2$ binds cooperatively with the cell type nonspecific MCM1 (Keleher et al., 1988; Passmore et al., 1989; Ammerer, 1990). To bind with high specificity to haploid-specific gene target sites, $\alpha 2$ protein heterodimerizes with the a1 protein. We have shown that, upon heterodimer formation with a1, a helix forms in the C-terminal tail of $\alpha 2_{128-210}$. At the $a1/\alpha 2$ interface, this helix may directly or indirectly create a new surface that is complementary to the DNA target site recognized by the heterodimer.

DNA-binding experiments and residue-specific chemical shift changes in $\alpha 2_{128-210}$ during titration with $a1_{66-126}$ show that the specific heterodimerization function is preserved in the a1 and $\alpha 2$ fragments studied in this work. Since the a1 fragment contains only the homeodomain, the homeodomain itself must serve as a protein interaction surface as well as a DNA-binding domain. We determined a moderate K_d of 2×10^{-4} M and a ratio of 1:1 for heterodimer dissociation. Since the a1 and $\alpha 2$ fragments retain this ability to specifically dimerize at a ratio of 1:1, and since the $a1_{66-126}/\alpha 2_{128-210}$ complex can bind target DNA with an affinity similar to that of intact $a1/\alpha 2$ (see Results), the $a1_{66-126}/\alpha 2_{128-210}$ complex retains the essential features of the intact $a1/\alpha 2$ complex. Given the supporting genetic evidence, the interactions we observe between the a1 and $\alpha 2$ fragments are also likely to exist between the intact a1 and $\alpha 2$.

Dividing our estimate of the overall dissociation constant of the ternary $a1_{66-126}/\alpha 2_{128-210}$ /operator complex of 10^{-14}–10^{-15} M^2 by the dissociation constant of the heterodimer of the a1 and $\alpha 2$ fragments of 2×10^{-4} M gives an estimate of the dissociation constant of the heterodimer and the hsg operator. This value of $\sim 10^{-11}$ M agrees well with estimates of the affinity of the heterodimer formed from intact a1 and $\alpha 2$ (Goutte & Johnson, 1993) and suggests that most if not all of the energetically important interactions for specific DNA binding are retained in the heterodimer formed by the fragments.

Specific Binding of $a1_{66-126}$ to $\alpha 2_{128-210}$ Induces an Interfacial Helix in $\alpha 2_{128-210}$. We assigned nearly all of the backbone ^{15}N, H^N, and H^α resonances of $\alpha 2_{128-210}$ bound to $a1_{66-126}$ and compared them with the resonances of uncomplexed $\alpha 2_{128-210}$. Only a few resonances are perturbed upon complex formation; these belong exclusively to the C-terminal tail of $\alpha 2_{128-210}$. This result suggests both that the C-terminal tail is the only region affected by $a1_{66-126}$ binding and that the homeodomain region of $\alpha 2_{128-210}$ is neither environmentally nor structurally perturbed by $a1_{66-126}$ binding.

Secondary structure determination of the bound $\alpha 2_{128-210}$ confirms this conclusion. We identified the sequential and intermediate-range NOE connectivities involving the H^N and H^α protons, roughly defining the secondary structure of the a1-bound form of $\alpha 2_{128-210}$ (Phillips et al., 1991). In addition, $^3J_{HN-H\alpha}$ coupling constants of less than 6.5 Hz helped to define the boundaries of the helices. The three helices of the homeodomain remain of the same length and position as in the free $\alpha 2_{128-210}$. Helix 1 spans residues 138–150, helix 2 spans residues 159–169, and helix 3 spans residues 173–187. The region C-terminal to the homeodomain was largely

unstructured in the free $\alpha 2_{128-210}$, whereas when bound to a1, this tail adopts a helical structure that extends from residue 193 to at least 200. This region shows comparable helical-type sequential d_{NN} NOE volumes and $d_{NN}(i, i+2)$ connectivities also seen in the homeodomain helices. Helical-type NOE connectivities are not observed beyond residue 200, but the $^3J_{HN-H\alpha}$ coupling constants are less than 6.5 Hz through residue 203, suggesting that the helix may extend this far. Absence of chemical shift change elsewhere in $\alpha 2_{128-210}$ during the titration with $a1_{66-126}$ argues strongly that direct contact with the a1 homeodomain induces formation of this fourth helix.

Recently, Spolar and Record (1994) have pointed out that there is often ordering of protein residues as a result of the specific binding of the protein to its DNA target. Our observations suggest that protein–protein interactions may also serve to initiate this ordering process. Thus heterodimer formation may "prepay" some of the entropy cost associated with the ordering of the proteins in their DNA complex.

ACKNOWLEDGMENT

We thank Susan Baxter, David M. Gontrum, Andrew K. Vershon, and Arkady Mak for communication of unpublished results and constructive discussions. We also thank Dennis Hare and Hare Research for supplying the FELIX software.

REFERENCES

Ammerer, G. (1990) *Genes Dev. 4*, 299–312.
Dolan, J. W., & Fields, S. (1991) *Biochim. Biophys. Acta 1088* (2), 155–169.
Dranginis, A. M. (1990) *Nature 347*, 682–685.
Forman-Kay, J. D., Gronenborn, A. M., Kay, L. E., Wingfield, P. T., & Clore, M. G. (1990) *Biochemistry 29*, 1566–1572.
Fraenkel, Y., Navon, G., Aronheim, A., & Gershoni, J. M. (1990) *Biochemistry 29*, 2617–2622.
Gegner, J. A., & Dahlquist, F. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 750–754.
Goutte, C., & Johnson, A. D. (1988) *Cell 52*, 875–882.
Goutte, C., & Johnson, A. D. (1993) *J. Mol. Biol. 233*, 359–371.
Hall, M. N., & Johnson, A. D. (1987) *Science 237*, 1007–1012.
Harashima, S., Miller, A. M., Tanaka, K., Kusumoto, K., Tanaka, K., Mukai, Y., Nasmyth, K., & Oshima, Y. (1989) *Mol. Cell. Biol. 9*, 4523–4530.
Herskowitz, I. (1989) *Nature 342*, 749–757.
Johnson, A. D. (1992) *Transcription Regulation* (McKnight, S. L., & Yamamoto, K. R., Eds.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.
Kay, L. E., Brooks, B., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 5488–5490.
Kay, L. E., Marion, D., & Bax, A. (1989b) *J. Magn. Reson. 84*, 72–84.
Keleher, C. A., Goutte, C., & Johnson, A. D. (1988) *Cell 53*, 927–936.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579–590.
Levy, G. C., & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance in Spectroscopy*, J. Wiley & Sons, New York.
Mak, A., & Johnson, A. D. (1993) *Genes Dev. 7*, 1862–1870.
McIntosh, L. P., Wand, A. J., Lowry, D. A., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341–6362.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44–73.
O'Neil, K. T., Shuman, J. D., Ampe, C., & Delgado, W. F. (1991) *Biochemistry 30*, 9030–9034.
Pardi, A., Wagner, G., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 445–454.
Passmore, S., Elble, R., & Tye, B. K. (1989) *Genes Dev. 3*, 921–935.
Phillips, C. L., Vershon, A. K., Johnson, A. D., & Dahlquist, F. W. (1991) *Genes Dev. 5*, 764–772.
Porter, S. D., & Smith, M. (1986) *Nature 320*, 766–768.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell 59*, 573–580.
Ransone, L. J., & Verma, I. M. (1990) *Annu. Rev. Cell Biol. 6*, 539–557.
Sandstrom, J. (1982) *Dynamic NMR Spectroscopy*, Academic Press, New York.
Sauer, R. T., Smith, D. L., & Johnson, A. D (1988) *Genes Dev. 2*, 807–816.
Scott, M. P., Tamkun, J. W., & Hartzell, G. W. (1989) *Biochim. Biophys. Acta 89*, 25–48.
Spolar, R. S., & Record, T. M., Jr. (1994) *Science 263*, 777–784.
Sprague, G. F., Jr. (1990) *Adv. Genet. 27*, 33–62.
Strathern, J., Shafer, B., Hicks, J., & McGill, C. (1988) *Genetics 120*, 75–81.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Valensin, G., Kushnir, T., & Navon, G. (1982) *J. Magn. Reson. 46*, 23–29.
Vershon, A. K., & Johnson, A. D. (1992) *Cell 72*, 105–112.
Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., & Pabo, C. O. (1991) *Cell 67*, 517–528.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Zuiderweg, E. R. P. (1990) *J. Magn. Reson. 86*, 346–357.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387–2391.