Heterodimerization of the Yeast Homeodomain Transcriptional Regulators α2 and a1: Secondary Structure Determination of the a1 Homeodomain and Changes Produced by α2 Interactions†

S. M. Baxter, D. M. Gontrum, C. L. Phillips,‡ A. F. Roth,§ and F. W. Dahlquist*

Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97404

Received May 9, 1994; Revised Manuscript Received August 31, 1994⊗

ABSTRACT: The homeodomain proteins, **a**1 and α2, act cooperatively to regulate cell-type specific genes in yeast. The basis of this cooperativity is an interaction between the two proteins, forming a heterodimer that binds DNA tightly and specifically. A fragment containing the homeodomain of **a**1, $\mathbf{a}1_{66-126}$, has been studied by NMR spectroscopy to gain secondary structure information and to characterize the changes in **a**1 upon heterodimerization with α2. Heteronuclear ($^1H-^{15}N$) NMR methods were used to assign backbone resonances of the 61 amino acid fragment. The $\mathbf{a}1_{66-126}$ secondary structure was determined using NOE connectivities, $^3J_{HN\alpha}$ coupling constants and hydrogen exchange kinetic data. NMR data identify three helical segments separated by a loop and a tight turn that are the characteristic structural elements of homeodomain proteins. The **a**1 fragment was titrated with $\alpha2_{128-210}$, the homeodomain-containing fragment of α2, to study changes in $\mathbf{a}1_{66-126}$ spectra produced by α2 binding. The $\mathbf{a}1_{66-126}$ protein was labeled with ^{15}N and selectively observed using isotope-edited NMR experiments. NMR spectra of bound $\mathbf{a}1_{66-126}$ indicate that residues in helix 1, helix 2, and the loop connecting them are directly involved in the binding of the α2 fragment. Relatively minor effects on the resonances from residues in helix 3, the putative DNA-binding helix, were noted upon α2 binding. We have thus located a region of the **a**1 homeodomain important for specific protein recognition.

Transcriptional factors containing the homeodomain DNA-binding motif have been found in many eukaryotic species, ranging from yeast to human [for reviews, see Hayashi and Scott (1990) and Laughon (1991)]. The regulatory function of the homeodomain derives from the specificity of its interactions with DNA and with other components of the transcriptional machinery. Our understanding of the homeodomain−DNA interaction has grown in the last several years as a result of X-ray and NMR studies of homeodomain−DNA complexes (Kissinger et al., 1990; Otting et al., 1990; Wolberger et al., 1991; Qian et al., 1993; Billeter et al., 1993). These structural studies have revealed a general model of recognition among homeodomains involving DNA contacts in the major and minor grooves via the third helix and an N-terminal arm, respectively. Despite conserved modes of DNA-binding, homeodomain proteins achieve a range of extremely specific regulatory actions. Interactions with other regulatory proteins may be largely responsible for the biological specificity of homeodomains. To date, there is little structural information about these protein−protein interactions.

Transcriptional regulators that contain homeodomain sequences often incorporate a dimerization region separate from the DNA-binding motif. For example, the mammalian tissue-specific activator HNF1 forms dimers via a myosin-like dimerization helix separate from the homeodomain (DeSimone & Cortese, 1991; Mendel & Crabtree, 1991). In a handful of cases, however, homeodomains themselves have been shown to mediate protein−protein interactions. The viral VP16 protein selectively associates with the mammalian Oct-1 homeodomain as part of a multiprotein−DNA complex (Stern et al., 1989; Kristie & Sharp, 1990). Recently, Andrés and co-workers (1994) demonstrated that the mClox homeodomain cooperatively interacts with cut II repeat to bind DNA. Perhaps the best understood interaction between two homeodomain-containing proteins involves the yeast transcriptional regulators α2 and **a**1.

Control of mating-type gene expression in the yeast *Saccharomyces cerevisiae* involves these regulators [for reviews, see Herskowitz (1989), Sprague (1990), Dolan and Fields, (1991), and Johnson (1992)]. The DNA-binding specificity of α2 depends on the presence of other transcriptional regulators in the cell. In both α haploid and **a**/α diploid cells, an α2 homodimer acts in combination with the non-cell-type specific protein MCM1 to bind DNA target sequences upstream of **a**-specific genes. This results in repression of these genes. The α2 protein has a second regulatory activity in diploid cells, where **a**1 is also present. In this case, α2 forms a heterodimer with **a**1 to repress haploid-specific gene expression (Goutte & Johnson, 1988; Dranginis, 1990; Goutte & Johnson, 1993). Full-length **a**1 and α2 bind DNA operators upstream of haploid specific genes (*hsg*) in a strongly cooperative fashion. The α2 protein, in the absence of **a**1, binds weakly to *hsg* operators, while **a**1 alone shows no detectable specific binding (Goutte & Johnson, 1993). Furthermore, the **a**1 and α2 proteins heterodimerize in the absence of DNA (Mak & Johnson, 1993). Recently, Phillips and co-workers (1994) demon-

† This work was supported by a National Research Service Award (GM15738) from the NIGMS to S.M.B. and by grants from the American Cancer Society and National Institutes of Health to F.W.D.

‡ Present address: Department of Biochemistry and Molecular Biology, Oregon Health Sciences University, Portland, OR 97201.

§ Present address: Department of Surgery, Wayne State University School of Medicine, Detroit, MI 48201.

⊗ Abstract published in *Advance ACS Abstracts,* December 1, 1994.

0006-2960/94/0433-15309$04.50/0

strated that fragments of the two proteins, $\alpha 2_{128-210}$ and $\mathbf{a}1_{66-126}$, which contain the homeodomains, are sufficient for cooperative binding to the *hsg* operator. The heterodimer formed by these fragments binds to target DNA with only a 10-fold reduction of affinity compared to the intact protein complex. This difference is attributed to the absence of protein−protein contacts made between **a**1 and the amino terminus of α2, which are deleted in the fragment. Importantly, Phillips and co-workers (1994) used NMR techniques to demonstrate that **a**1 and α2 fragments also heterodimerize in the absence of DNA.

The $\alpha 2_{128-210}$ fragment consists of the C-terminal 83 amino acids and has been well-characterized both structurally and functionally (Sauer et al., 1988; Phillips et al., 1991; Wolberger et al., 1991). This fragment contains the homeodomain and the adjacent 22 residues that comprise the C-terminal tail of the protein. NMR and X-ray studies have shown the fragment contains the characteristic three helices and two turns of the homeodomain proteins. NMR studies (Phillips et al., 1991) have shown the carboxy terminal tail is unstructured when the protein is free in solution. X-ray studies of the $\alpha 2_{128-210}$/DNA co-crystal also demonstrate that the C-terminal tail is disordered (Wolberger et al., 1991). Using NMR techniques to observe only the α2 fragment, Phillips and co-workers (1994) showed that the homeodomain proper of $\alpha 2_{128-210}$ is largely unaffected by the binding of the **a**1 homeodomain. Instead, the addition of the **a**1 homeodomain fragment to $\alpha 2_{128-210}$ induces the formation of structure in the previously unstructured C-terminal tail, suggesting that heterodimer formation between **a**1 and α2 is mediated by a region outside the α2 homeodomain region. On the basis of DNA protection assays, Mak and Johnson (1993) proposed that the C-terminal tail of α2 does not contact DNA, suggesting that the structure formed upon heterodimerization indirectly affects the heterodimer surface complementary to the DNA target site.

The object of the work presented here is to structurally characterize the **a**1 homeodomain-containing fragment $\mathbf{a}1_{66-126}$ and its interaction with $\alpha 2_{128-210}$. The **a**1 fragment contains the C-terminal 61 amino acids of the intact protein and corresponds to the putative homeodomain. NMR solution studies are used to provide structural information about the **a**1 fragment. NMR experiments using free $\mathbf{a}1_{66-126}$ demonstrate that the fragment contains three helical regions separated by two turns, characteristic of homeodomain proteins. Additionally, changes seen in the NMR spectra of **a**1 upon the addition of $\alpha 2_{128-210}$ enabled us to map the regions of the **a**1 protein important for α2 binding. By labeling $\mathbf{a}1_{66-126}$ with ^{15}N and using ^{15}N-edited experiments to filter out most signals from the unlabeled $\alpha 2_{128-210}$, structural changes in the **a**1 protein were clearly detected. Changes in the NMR spectra of bound $\mathbf{a}1_{66-126}$ indicate that residues in helix 1 and helix 2 and in the loop, connecting helix 1 and helix 2, are critical for the recognition of the C-terminal tail of $\alpha 2_{128-210}$.

MATERIALS AND METHODS

Protein Expression and Purification. The $\mathbf{a}1_{66-126}$ protein was purified from *E. coli* cells containing the plasmid (pCW/K66). This plasmid encodes residues 66−126 of the intact **a**1 protein under control of tandem P_{tac} promoters. The original $\mathbf{a}1_{66-126}$ plasmid, under control of the T7 promoter in the T7 expression system of Studier and Moffat, was generously provided by Martha Stark and A. D. Johnson. To achieve higher levels of protein expression, the **a**1 fragment was subsequently cloned into the *Nde*I and *Xba*I sites of pCWori+ (Muchmore et al., 1989; Gegner & Dahlquist, 1991). The construction of the resulting plasmid was confirmed by restriction digestion and double-stranded DNA sequencing.

The uniformly ^{15}N-labeled $\mathbf{a}1_{66-126}$ protein was purified to >95% homogeneity from a prototrophic *E. coli* strain, K38, transformed with the pCW/K66 plasmid. Each liter of M9T media with 100 mg/L ampicillin, containing 1.0 g/L $(^{15}NH_4)_2SO_4$ as the sole nitrogen source, was inoculated with 12.5-mL overnight culture of cells grown at 37 °C. The culture was grown at 37 °C to $A_{600} = 0.4$, at which time the $\mathbf{a}1_{66-126}$ expression was induced by the addition of IPTG to a concentration of 1.0 mM. After 12 h of induction, cells were harvested by centrifugation (10960*g*, 15 min at 4 °C). The overexpressed protein is quite susceptible to degradation, and the purification procedure was optimized to minimize exposure to proteases. The pellet was resuspended in lysis buffer (100 mM potassium phosphate, pH 6.5, 500 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 1 mM phenylmethanesulfonyl fluoride (PMSF), 250 μM E-64, 250 μM leupeptin, 0.01% NaN_3) using 25 mL of lysis buffer/L of harvested media. The cells were sonicated and spun at 31400*g* for 30 min. The clarified supernatant was desalted on a 100-mL G-25 column equilibrated with buffer A (25 mM potassium phosphate, pH 6.5, 0.01% NaN_3, 1 mM EDTA, 1 mM β-mercaptoethanol), 10 mM NaCl, and 1 mM PMSF. The desalted protein fractions were loaded directly onto a 30-mL CM Sepharose CL 6B (Sigma) column (100 × 30 mm) equilibrated with 10 mM NaCl and 1 mM PMSF in buffer A. The column was washed with 50 mL of 10 mM NaCl in buffer A with 1 mM PMSF. A 600-mL linear gradient of 10−600 mM NaCl in buffer A with 1 mM PMSF was applied to the column, and the protein eluted in a broad peak at 500 mM NaCl. The $\mathbf{a}1_{66-126}$ protein fractions were stable and protease free at this point. The pooled protein fractions were desalted again on a 100-mL G-25 column, equilibrated in 10 mM NaCl in buffer A, and loaded onto a second CM Sepharose CL 6B (25 mL) column (200 × 16 mm) equilibrated with 10 mM NaCl in buffer A to better purify the protein. A 400-mL linear gradient of 10−600 mM NaCl in buffer A was applied to the column, and the protein eluted at 500 mM NaCl. At this stage, the protein was free of protease inhibitors and found to be >95% pure by Coomassie-stained SDS−PAGE (19% polyacrylamide, 6 M urea). N-terminal sequencing analysis determined that the expression system did not process off the start codon, resulting in a 62 amino acid peptide with a methionine residue at the N-terminus. Electrospray and laser desorption mass spectroscopy confirmed the unlabeled protein mass at 7229 Da and the ^{15}N-labeled protein mass at 7366 Da, the expected mass for 100% ^{15}N enrichment. Protein concentrations were determined using a calculated molar extinction coefficient ($\epsilon_{280} = 5750\ M^{-1}\ cm^{-1}$). The pooled fractions were dialyzed overnight against 25 mM deuterated sodium acetate ($NaCD_3COOH$), pH 4.5, 100 mM KCl, and 0.01% NaN_3. The $\mathbf{a}1_{66-126}$ protein was concentrated to approximately 0.1 mM using Centriprep (Amicon) centrifugal concentrators with a 3000 molecular weight cutoff, followed by further concentration to final NMR concentration using

3000 MWCO concentrators (Filtron). Five percent D_2O was added for the spectrometer lock.

Selective ^{15}N-labeling of $\mathbf{a}1_{66-126}$ was carried out using *E. coli* strains, media, and methods described previously (Muchmore et al., 1989). We selectively enriched $\mathbf{a}1_{66-126}$ with [^{15}N]glycine and [^{15}N]serine using an auxotrophic *SerA* strain. [^{15}N]Alanine and [^{15}N]valine were incorporated using the auxotrophic *E. coli* strain DL39 *avtA::Tn10*. [^{15}N]-Leucine and [^{15}N]glutamine were incorporated using DL39. The auxotrophic strains were grown in synthetic rich media containing these pairs of isotopically labeled amino acids (Muchmore et al., 1989).

The $\alpha 2_{128-210}$ protein was purified to >95% homogeneity and concentrated as described (Phillips et al., 1991). For NMR experiments, the protein was dialyzed against 25 mM CD_3COOH, pH 4.5, 100 mM KCl, and 0.01% NaN_3 with 5% D_2O added for the spectrometer lock.

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 3-(trimethylsilyl)-propionate-*2,2,3,3-d*$_4$ 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 6024 Hz in the proton dimensions. The recycle times were usually 1 s. Water suppression was achieved by presaturation. Spectra were processed and analyzed using FELIX software from Hare Research and Biosym Technologies.

Conventional proton DQF-COSY, NOESY, RELAY, and TOCSY spectra for free and bound $\mathbf{a}1_{66-126}$ in H_2O and D_2O were recorded and processed as described (Wüthrich, 1986; Oppenheimer & James, 1989). ^{15}N-Edited COSY and NOESY spectra of uniformly labeled $\mathbf{a}1_{66-126}$ were recorded and processed as described by McIntosh and co-workers (1990). The HSMQC (Zuiderweg, 1990) experiment was used to measure single-bond $^1H-^{15}N$ correlations. HSMQC-NOESY, a variation of HMQC-NOESY (Gronenborn et al., 1989; Muchmore et al., 1989) for which an HSMQC $^1H-^{15}N$ pulse sequence is used to generate $^1H-^{15}N$ coherence, was collected to detect sequential amide connectivities. For the heteronuclear experiments, a delay of 4.8 ms was used as the nominal $(2J_{NH})^{-1}$ time period to reduce the loss of signal due to relaxation. TOCSY spectra were recorded with a mixing time of 70 ms using a MLEV-17 pulse train. The NOESY-type spectra were recorded with 100 and 200 ms mixing times. 2D heteronuclear correlation experiments were recorded for free and bound $\mathbf{a}1_{66-126}$ with a nitrogen spectral width of 3333 Hz and the ^{15}N carrier in the center of the amide region.

Chemical shift and line width changes were monitored during the titration of uniformly and specifically ^{15}N-labeled $\mathbf{a}1_{66-126}$ with $\alpha 2_{128-210}$ using HSMQC experiments. The HSMQC experiments were collected with 1024 complex points in t_2 and 512 increments in t_1. Spectra were processed with a shifted sine bell apodization and zero-filled to 2K and 1K in the proton and nitrogen dimensions, respectively.

The $^3J_{HN\alpha}$ coupling constants for free and bound $\mathbf{a}1_{66-126}$ were measured using a $^1H-^{15}N$ HMQC-*J* experiment (Kay et al., 1989a; Forman-Kay et al., 1990; Kay & Bax, 1990) and a set of nine 2D $^1H-^{15}N$ *J*-modulated HSQC experiments (Billeter et al., 1992; Neri et al., 1990). The HMQC-*J* data were processed and analyzed as described for $\alpha 2_{128-210}$ (Phillips et al., 1991). *J*-modulated HSQC experiments were recorded for free and bound $\mathbf{a}1_{66-126}$ with 1024 complex points in t_2 and 128 increments in t_1 and 128 scans per t_1 experiment. The proton spectral width was 6024.1 Hz, and the nitrogen spectral width was 1041.7 Hz with the ^{15}N carrier placed in the center of the amide region. Delay times (τ_2) ranging from 9.2 to 125 ms were chosen for the set of HSQC experiments. The set of HSQC experiments was processed identically with shifted sine bell apodization and zero-filled to 1K and 512 points in the proton and nitrogen dimensions, respectively. The volumes of the *J*-modulated HSQC signals were measured and plotted as a function of the delay time. Nonlinear least-squares fits to the τ_2-dependent cross-peak volume

$$V(\tau_2) = V_0 [\cos(\pi J\tau_1) \cos(\pi J\tau_2) - 0.5 \sin(\pi J\tau_1) \sin(\pi J\tau_2)] e^{(-\tau 2/T'2)} \quad (1)$$

were used to obtain $^3J_{HN\alpha}$, the apparent relaxation time T'_2, and V_0, the cross-peak volume at $\tau_2 = 0$, where τ_1 is a short INEPT transfer delay. A grid search, for which *J* values were incremented by 0.1 Hz steps from 0 to 11 Hz and V_0 and T'_2 were optimally fit with eq 1, was used to generate plots of *J* vs χ^2, where χ^2 is the sum of the squares of the deviations between the experimental cross-peak volumes, $V(\tau_2)$, and the corresponding volumes calculated from eq 1. Errors attributed to the *J* value measurements (Billeter et al., 1992) are estimated by lower and upper limits given by the two values of *J* where the $\chi^2(J)$ curves have values of $2(\chi^2_{min})$. Hydrogen exchange measurements were carried out as described for $\alpha 2_{128-210}$ (Phillips et al., 1991).

3D NOESY-HSMQC experiments, a variation of the 3D HMQC-NOESY (Kay et al., 1989b; Zuiderweg et al., 1989) for which an HSMQC $^1H-^{15}N$ pulse sequence is used to generate $^1H-^{15}N$ coherence, were collected for free and bound ^{15}N-labeled $\mathbf{a}1_{66-126}$. 3D NOESY-HSMQC data sets were recorded using 100- and 200-ms mixing times. Spectral widths used were 6024.1 Hz (^1H, F1), 1305 Hz (^{15}N, F2), and 6024.1 Hz (^1H, F3). The arginine N^ϵ protons were folded in the spectrum. Quadrature detection was obtained using time-proportional phase incrementation (TPPI) in the F1 and F2 dimensions. The 3D data matrices contained 256 × 64 × 1024 points. Sixteen scans were collected for each t_1 increment. The pulse delay between experiments was 800 ms. 3D clean TOCSY-HSMQC (70-ms mixing time) data sets were also collected similarly for free and bound ^{15}N-labeled $\mathbf{a}1_{66-126}$. The final 3D matrices were zero-filled to 1024 × 256 × 128 (real) points.

RESULTS

The $\mathbf{a}1_{66-126}$ fragment was overexpressed using the pCW/K66 expression system in *E. coli*, producing about 5 mg/L in minimal media. Polyacrylamide gel electrophoresis shows that the protein migrates as a single band. The presence of a methionine residue at the N-terminus was confirmed by N-terminal amino acid sequencing. Additionally, mass spectroscopy results indicate that the protein exists as a single species without cysteine oxidation or other covalent modifications. The $\mathbf{a}1_{66-126}$ fragment is stable over a pH range of 4.5−7.0 and gives good NMR spectra at pH 4.5, 100 mM KCl, and 298 K. No line width or chemical shift changes are seen over a concentration range, 0.3−4 mM.

Sequential Resonance Assignments, Free $\mathbf{a}1_{66-126}$. Due to poor spectral dispersion in the H^α region of the $\mathbf{a}1_{66-126}$ 2D

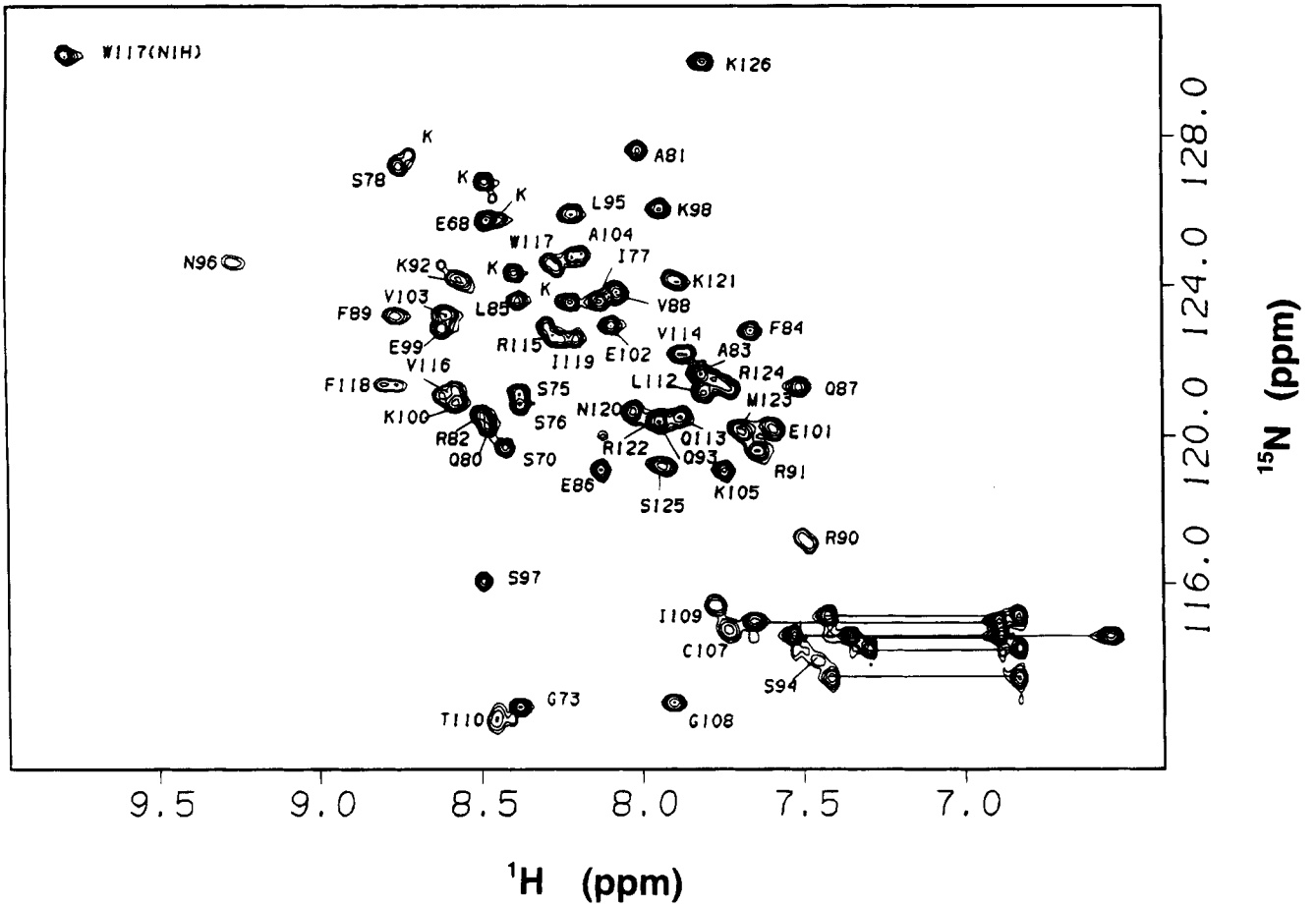

FIGURE 1: HSMQC spectrum of fully ^{15}N-labeled **a**1_{66-126} (1.6 mM). The spectrum was taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. Each peak in the spectrum represents a proton directly bound to ^{15}N. The primary amide groups of the glutamine and asparagine residues are represented by the cross-peaks pairs connected by lines in the lower right corner of the spectrum. Assigned backbone amide protons are labeled. Lysines in the N-terminal segment of the protein are not assigned sequentially and are labeled with K.

NOESY and 2D COSY spectra, the sequential resonance assignments were largely made using 2D and 3D heteronuclear NMR experiments. A fully ^{15}N-labeled **a**1_{66-126} sample was used to define all amide resonance positions in a fingerprint 2D HSMQC experiment. The ^{1}H$-^{15}$N HSMQC spectrum of the protein fragment is shown in Figure 1 together with the assignments made as described below. The resonances are numbered from 66 to 126 so that they correspond to their positions in the intact protein. Each resonance represents a proton directly attached to an ^{15}N atom. The **a**1_{66-126} fragment contains 62 amino acids, including three prolines. Excluding the N-terminal methionine, for which the amino group is generally not observed in protein spectra, 58 ^{1}H$-^{15}$N backbone amide resonances are expected. The single tryptophan indole proton resonance is evident in the upper left corner of the spectrum. Six pairs of resonances, connected by a line at a common nitrogen chemical shift, represent the four glutamine and two asparagine side-chain amide groups. The remaining resonances represent the backbone amides, and 54 cross-peaks can be identified. Four resonances are lost to overlap, which was confirmed in subsequent experiments at different temperature and pH conditions. Close inspection of the spectrum reveals that a number of the peaks are broadened and even doubled.

Heteronuclear 3D spectra were used to identify individual spin systems and sequential connectivities. Backbone amide resonances were correlated to their C$^{\alpha}$H resonances using a ^{1}H$-^{15}$N 3D clean TOCSY-HSMQC experiment. Where possible, the spin system assignments were extended further out the side chains using this experiment and a homonuclear 2D TOCSY experiment. Reference points along the backbone sequence were established using three specifically labeled **a**1_{66-126} samples: [^{15}N]glycine + [^{15}N]serine; [^{15}N]alanine + [^{15}N]valine; and [^{15}N]leucine + [^{15}N]asparagine. The ^{1}H$-^{15}$N 3D NOESY-HSMQC experiment identified connectivities between sequential backbone amides and between C$^{\alpha}$H and amide resonances. Strong sequential (*i*,*i*+*1*) amide$-$amide connectivities (d_{NN}) were seen between most of the backbone amides. Amide$-$amide connectivities were confirmed and extended using a 2D HSMQC$-$NOESY data set. Due to poor dispersion in the H$^{\alpha}$ region of the spectra, many sequential H$^{\alpha}-$amide connectivities were obscured. An example of the resulting sequential assignment is illustrated in Figure 2 for residues Glu80$-$Phe89. For simplicity, only narrow strips of the proton$-$proton planes at the ^{15}N frequency of the amide nitrogen are selected. Sequential amide$-$amide (d_{NN}) connectivities are traced in the figure. Lines highlight sequential NOEs between amide protons and H$^{\alpha}$ protons identified in COSY and TOCSY experiments. The two alanine and one valine resonances were identified using the specifically labeled [^{15}N-alanine, valine]**a**1_{66-126} sample. The sequence Q80-A81-R82-A83 is elucidated primarily from the d_{NN} connectivities. The overlap in the H$^{\alpha}$ region obscures the sequential ordering of the sequence. H$^{\beta}$ to amide NOEs verify the Q80-A81 and A81-R82 sequential assignments. By these methods, se-

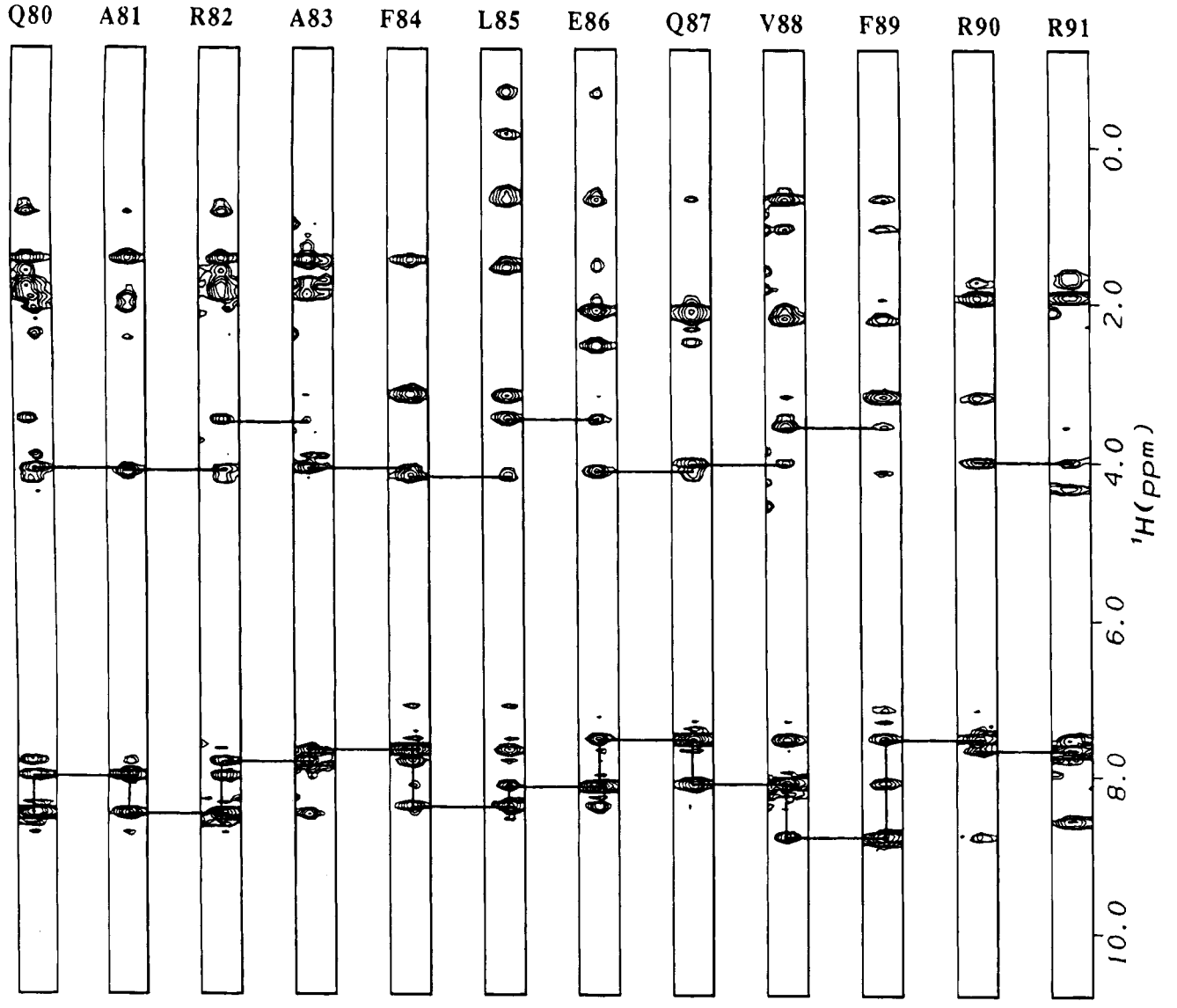

FIGURE 2: Amide proton strip plots from 3D HSMQC-NOESY (τ_M = 200 ms) of $\mathbf{a1}_{66-126}$ which show sequential connectivities in the peptide segment (Glu80−Arg91). These residues comprise helix 1 of the homeodomain structure. The spectrum was taken on a 1.8 mM $\mathbf{a1}_{66-126}$ sample in 95% H_2O/5% D_2O at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3. The individual strips were taken from proton−proton planes at different nitrogen chemical shifts and centered about the amide proton chemical shift of the residue indicated at the top of the strip. The strips are ordered according to amino acid sequence. Lines indicate the $d_{NN(i,i+1)}$ and $d_{\alpha N(i,i+1)}$ connectivities. Extensive overlap in the H^α region of the spectra prevents identification of long-range $(i,i+3)$ connectivities in many cases.

quential assignments for most of the backbone resonances except the prolines were obtained. K106 is tentatively assigned because of overlap with two other residues. Additionally, the lysines in the amino terminal segment of the protein are not assigned sequentially and are labeled with K in Figure 1. A listing of amide chemical shifts is given in Table I.

The sequential assignment of the $\mathbf{a1}_{66-126}$ fragment confirms that a number of the resonances are doubled or broadened due to some slow conformational equilibrium. Mass spectroscopic data and N-terminal sequencing confirm that the extra peaks are not due to the heterogeneity of the amino terminus or covalent modification. NOE connectivity patterns suggest that extra resonances in all cases were doubled peaks and not impurities in the protein preparation. Indeed, the same resonances were doubled in several protein preparations expressed in different *E. coli* strains and grown under slightly different conditions. The relative intensities of the component peaks varied slightly with protein preparation. The doubled resonances, split by greater than 10 Hz in the proton dimension, were assigned as A104, V114, R115, V116, W117, F118, and I119.

Aromatic Resonance Assignments, Free $\mathbf{a1}_{66-126}$. Aromatic resonances for the tryptophan and phenylalanine side chains were identified in 2D DQF-COSY and 2D TOCSY experiments. The spin system of the unique tryptophan W117 was identified in the COSY spectrum by three cross-peaks connecting the four ring protons. The ring connectivities were confirmed and extended through NOE connectivities to each other and from H7 to the indole proton and H2. The H4 ring proton displayed intraresidue NOE cross-peaks to the β protons. The remaining aromatic resonances were assigned to phenylalanines. Three separate spin systems were found in the COSY spectrum corresponding to the three phenylalanine residues. Aromatic protons of Phe84 and Phe118 were identified on the basis of intraresidue NOE cross-peaks to the corresponding β and α protons. The aromatic ring protons of Phe84 show overlap in chemical shifts so that only two aromatic resonances are observed. The remaining three aromatic resonances, linked by two

Table 1: Proton and Nitrogen Chemical Shifts (in ppm) for Backbone Amides in $\mathbf{a}1_{66-126}$, Free and Bound to $\alpha2_{128-210}$, at 25 °C

	free		bound			free		bound	
residue	HN	^{15}N	HN	^{15}N	residue	HN	^{15}N	HN	^{15}N
K66					S97	8.50	116.10	8.61	115.26
K67					K98	7.95	126.13	7.56	126.25
E68	8.47	125.50	8.47	125.50	E99	8.62	122.88	8.63	122.36
K69					K100	8.58	120.91	8.67	119.98
S70	8.43	119.70	8.43	119.62	E101	7.60	120.21	7.65	120.29
P71					E102	8.09	122.94	8.29	121.44
K72					V103	8.61	123.18	8.80	124.01
G73	8.38	112.68	8.39	112.68	A104	8.20	124.80	8.26	124.47
K74					K105	7.75	119.07	7.79	119.13
S75	8.38	121.16	8.38	120.82	K106[a]	7.80	121.50	7.80	121.50
S76	8.41	121.16	8.41	121.16	C107	7.74	114.78	7.69	113.76
I77	8.13	123.63	8.15	123.88	G108	7.90	112.77	7.94	112.70
S78	8.75	127.23	8.80	127.63	I109	7.78	115.42	7.78	115.07
P79					T110	8.46	112.31	8.48	112.15
Q80	8.48	120.30	8.59	120.08	P111				
A81	8.01	127.62	8.04	127.39	L112	7.81	121.20	7.79	120.98
R82	8.50	120.50	8.56	120.84	Q113	7.88	120.57	7.88	120.33
A83	7.83	121.70	7.83	121.41	V114	7.88	122.17	7.86	121.95
F84	7.66	122.79	7.64	123.06	R115	8.30	122.84	8.35	122.52
L85	8.39	123.61	8.50	123.53	V116	8.60	121.20	8.63	121.18
E86	8.13	119.09	7.96	118.67	W117	8.28	124.52	8.26	124.47
Q87	7.51	121.30	7.47	121.39	F118	8.78	121.35	8.89	121.46
V88	8.08	123.76	8.09	122.93	I119	8.23	122.57	8.28	122.54
F89	8.76	123.20	8.89	122.80	N120	8.03	120.64	7.99	120.31
R90	7.51	117.24	7.49	116.31	K121	7.89	124.11	7.93	124.27
R91	7.64	119.61	7.57	118.92	R122	7.95	120.42	7.98	120.40
K92	8.56	124.15	8.59	125.20	M123	7.70	120.18	7.66	119.77
Q93	7.95	120.42	7.83	121.48	R124	7.74	121.33	7.68	120.92
S94	7.51	114.12	7.35	113.62	S125	7.94	119.19	7.84	119.09
L95	8.21	125.94	8.45	126.78	K126	7.81	130.01	7.79	130.01
N96	9.27	124.66	9.47	124.84					

[a] Tentatively assigned.

COSY cross-peaks, were assigned to Phe89. An NOE connectivity between this spin system and the α proton is obscured because the α proton resonates at the water frequency. Aromatic residues which display mutual NOEs are Trp117, Phe118, and Phe89.

Secondary Structure Determination, Free $\mathbf{a}1_{66-126}$. Once sequential assignments for $\mathbf{a}1_{66-126}$ were completed, further experiments were collected to provide secondary structure information relating to backbone conformation and amide proton solvent accessibility. Protection from hydrogen exchange is an indicator of hydrogen-bonded secondary structure. $^3J_{HN\alpha}$ coupling constants are dependent on the dihedral angle ϕ of the protein backbone. Sequential stretches of resonances with $^3J_{HN\alpha}$ coupling constants less than 6.0 Hz are typical of helical structures, whereas stretches of $^3J_{HN\alpha}$ coupling constants larger than 8.0 Hz indicate extended structure. Intermediate values are possibly due to conformationally averaged backbone conformations. Using the HMQC-*J* experiment, we observed $^3J_{HN\alpha}$ coupling constants greater than 5 Hz; splittings less than 5 Hz were not resolved in the $\mathbf{a}1_{66-126}$ HMQC spectra, which had ^{1}H−^{15}N multiple quantum line widths of about 10 Hz. To more accurately quantitate the $^3J_{HN\alpha}$ coupling constants, a set of *J*-modulated HSQC experiments was performed. The dependence of the HSQC cross-peak volumes on the mix time, τ_2, were fit to eq 1 to extract *J* values. Analysis of this data (Figure 3) provided information about the secondary structure of the $\mathbf{a}1_{66-126}$ peptide. Data are missing for Lys66, three prolines, and four residues for which resonances are overlapped. The precision of this measurement for smaller *J* values is limited by the dominance of the short T'_2 (Billeter et al., 1992). Consequently, lower error limits could not be determined for Ser78, Leu85, Val116, and Phe118. These couplings are smaller than 4.5 Hz, and the cross-peak volumes recorded at all chosen delay times are positive. The experimental data for Phe89 and Ala104 were not fit adequately by eq 1 because the cross-peak volumes are far from the point of inversion and are reported as 4.0 Hz, a semiquantitative value. No significant lower limit for the Ser94 *J*-value was determined because of relatively rapid T'_2 relaxation time, resulting in poor signal to noise. Gly73 and Cys107 were also fit poorly and are estimated from the point of sign inversion, τ_2^0, where $^3J_{HN-H\alpha} = 1/2(\tau_2^0)$ (Neri et al., 1991).

Short-range and medium-range NOE data reveal three separate regions of strong d_{NN} and contiguous $d_{\alpha N(i,i+3)}$ connectivities. Missing connectivities in these regions are due to overlap and not absent NOEs. This combination of NOE connectivities is diagnostic of helical regions (Wüthrich, 1986). These NOE patterns coincide with regions of slowly exchanging amide protons and small coupling constants (<6.0 Hz), suggesting a helical structure for these segments. NOE connectivities, hydrogen exchange data, and $^3J_{HN-H\alpha}$ coupling constants are summarized in Figure 4. Protection from hydrogen exchange typically begins at the fourth residue of an α-helix due to intrahelical (*i*,*i*+4) hydrogen bonds. Using this indicator of helical start points, combined with other NMR data, we estimate that helix 1 extends from Gln80 to Arg91, helix 2 extends from Ser97 to Lys105, and helix 3 extends from Leu112 to Met123. Close examination of the hydrogen exchange data suggests that helix 1 and helix 3 are more uniformly protected compared to helix 2. This pattern of irregular protection from exchange was also noted in helix 2 of $\alpha2_{128-210}$ (Phillips et al., 1991). A number of residues in the amino terminal segment of helix 3 resonate as doubled peaks. Helical NOE patterns are maintained for

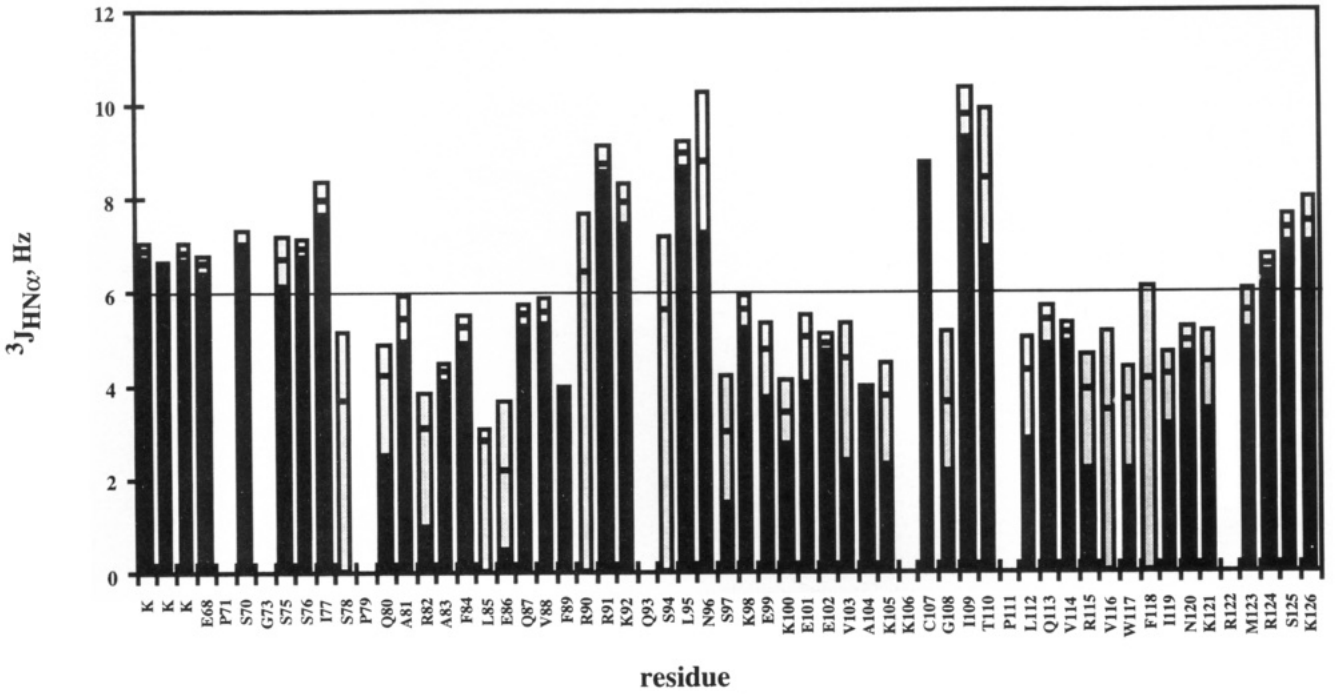

FIGURE 3: $^3J_{HN-H\alpha}$ coupling constants for **a**1$_{66-126}$ determined from the series of 2D $^1H-^{15}N$ *J*-modulated HSQC experiments. Asymmetric error ranges are shown for the $^3J_{HN-H\alpha}$ values determined from these experiments. Upper and lower limits were determined from a plot of χ^2 vs *J* where $\chi^2(J) = 2\chi^2_{min}$ (Billeter et al., 1992).

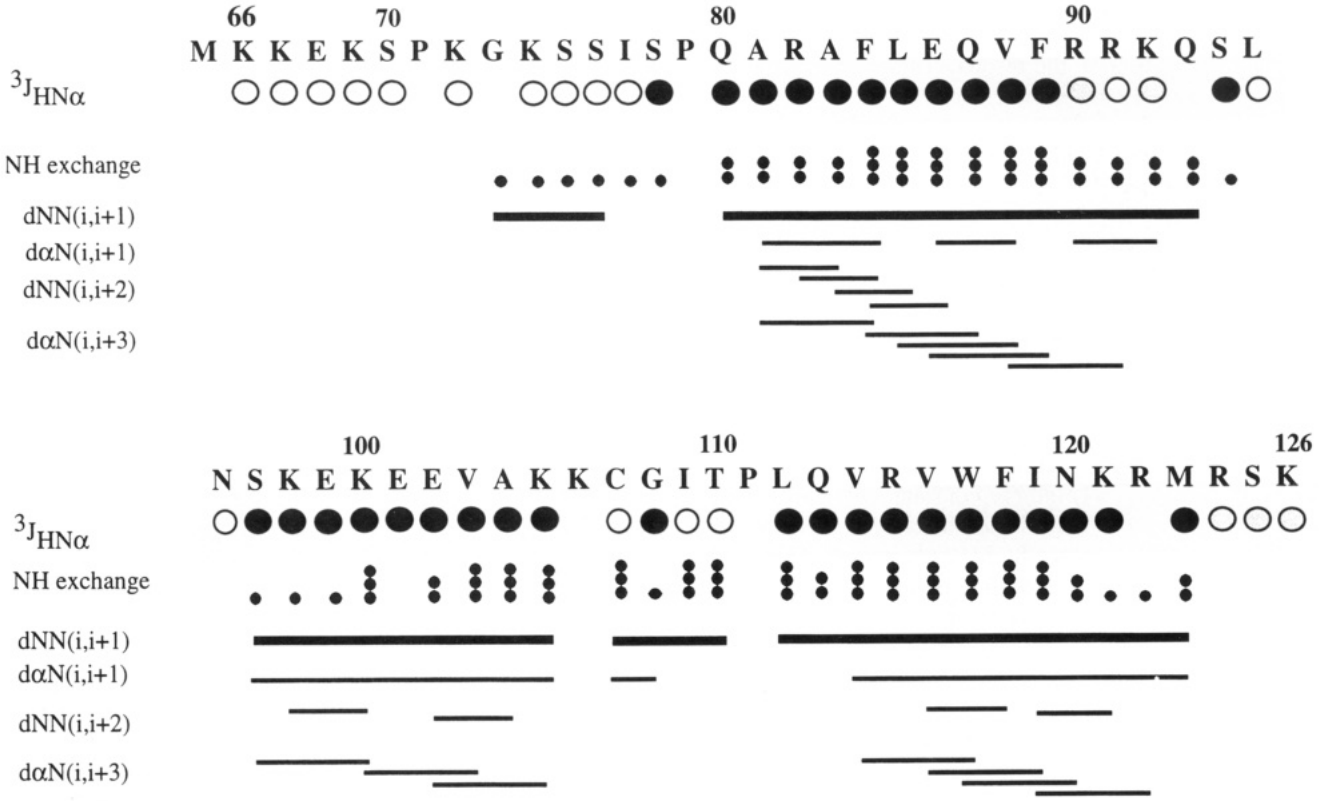

FIGURE 4: **a**1$_{66-126}$ amino acid sequence with $^3J_{HN-H\alpha}$ coupling constants, hydrogen exchange data and sequential and medium-range NOE connectivities. For $^3J_{HN-H\alpha}$ coupling constants, large filled circles represent coupling constants of less than 6.0 Hz, gray circles represent intermediate *J* values (6.0 Hz < *J* < 8.0 Hz), and open circles represent coupling constants greater than 8.0 Hz. For the hydrogen exchange data, 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 dots indicate a rate of hydrogen exchange slower than 0.05 min^1. Broad lines represent $d_{NN(i,i+1)}$ sequential connectivities; narrower lines indicate $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+2)}$, and $d_{\alpha N(i,i+3)}$ connectivities. Line thickness is not intended to imply NOE intensities.

both conformers, represented by the doubled resonances, suggesting that the slow conformational equilibrium represented is not a result of changes in the overall helical structure of this segment and that the helical structure is not perturbed by the observed heterogeneity.

Very few NOE connectivities are observed in the N-terminal region (residues 66−78) of the peptide. Intermediate (6.0 Hz < *J* < 8.0 Hz) $^3J_{HN\alpha}$ coupling constants are observed for most of these residues. In the HSMQC spectrum, the N-terminal amide resonances are sharp, and

their proton chemical shifts are consistent with a random coil structure. These observations lead to the conclusion that the N-terminal region of the peptide is unstructured and flexible.

Breaks in the helical connectivity patterns are found at two other regions, encompassing residues 93−96 and residues 106−110. In particular no $i,i+3$ connectivities are observed in these regions. The lack of NOE connectivities between amino acids 93−95 is accompanied by generally large (>6.0 Hz) $^3J_{HN\alpha}$ coupling constants and relatively fast amide hydrogen exchange rates. Hydrogen exchange rate could not be accurately measured for Asn96, which is downfield shifted at 9.25 ppm, because it is broadened and weak in intensity. This resonance is at the amino terminal end of helix 2. We believe this segment, residues 93−96, is largely unstructured, linking helix 1 and helix 2. Weak amide−amide NOE cross-peaks and a few H^{α}−amide connectivities link residues 106−110. The $^3J_{HN\alpha}$ coupling constants for three of these residues (C107, I109, and T110) are relatively large (>8.0 Hz), but the hydrogen exchange data suggest that these amides are somewhat protected from solvent. These observations suggest that this segment forms a tight turn, linking helix 2 and helix 3.

Formation of Heterodimer Composed of $\mathbf{a}1_{66-126}$ *and* $\alpha2_{128-210}$. Figure 5 shows a series of HSMQC spectra of ^{15}N-[Ala,Val]$\mathbf{a}1_{66-126}$ as it is titrated with unlabeled $\alpha2_{128-210}$. Since only the alanines and valines of the **a**1 fragment are labeled with ^{15}N, we observe only six resonances. Over the course of the titration, the resonances are affected in three ways. Two of the ^{1}H−^{15}N resonances, A81 and A83, are not noticeably changed as $\alpha2_{128-210}$ is added to the solution. One resonance, V103, shifts steadily to a new, bound position and broadens somewhat over the course of the titration. Two resonances, V116 and A104, that are doubled at the beginning of the titration coalesce as $\alpha2_{128-210}$ is added. These three types of changes upon heterodimerization are observed throughout the titration of uniformly ^{15}N-labeled $\mathbf{a}1_{66-126}$ with unlabeled $\alpha2_{128-210}$. The total number of HSMQC resonances remain the same. The gradual shifting of resonances from a free to a bound position indicates that heterodimerization is fast relative to the chemical shift time scale.

From the resonances that shift during the $\alpha2_{128-210}$ titration, we can calculate a binding constant for the complex formation. Figure 6 shows the chemical shift of a representative resonance, Lys98, as a function of $\alpha2_{128-210}$ concentration. Nonlinear least-squares fitting to a single binding-site model to the data produces a smooth curve predicting a dissociation constant (K_d) of 0.32 mM over a total shift of 223 Hz (0.45 ppm). Similar fits to other α2-sensitive resonances gave an average binding constant of 0.3 mM. This determination of the heterodimer dissociation constant agrees within experimental error with the K_d calculated from shifts seen in α2 upon **a**1 binding (Phillips et al., 1994). Subsequent NMR experiments were performed on a sample containing 2.0 mM $\mathbf{a}1_{66-126}$ and 4.0 mM $\alpha2_{128-210}$, for which about 90% of the $\mathbf{a}1_{66-126}$ protein is bound to $\alpha2_{128-210}$.

The behavior of the doubled peaks in the $\mathbf{a}1_{66-126}$ spectra was carefully examined during the $\alpha2_{128-210}$ titration. The doubled amide resonances are split into two peaks of roughly the same intensity. For most of the doubled resonances, the two peaks move toward each other and coalesce as the $\alpha2_{128-210}$ concentration is increased. The relative intensity of the peaks does not change noticeably, suggesting that the

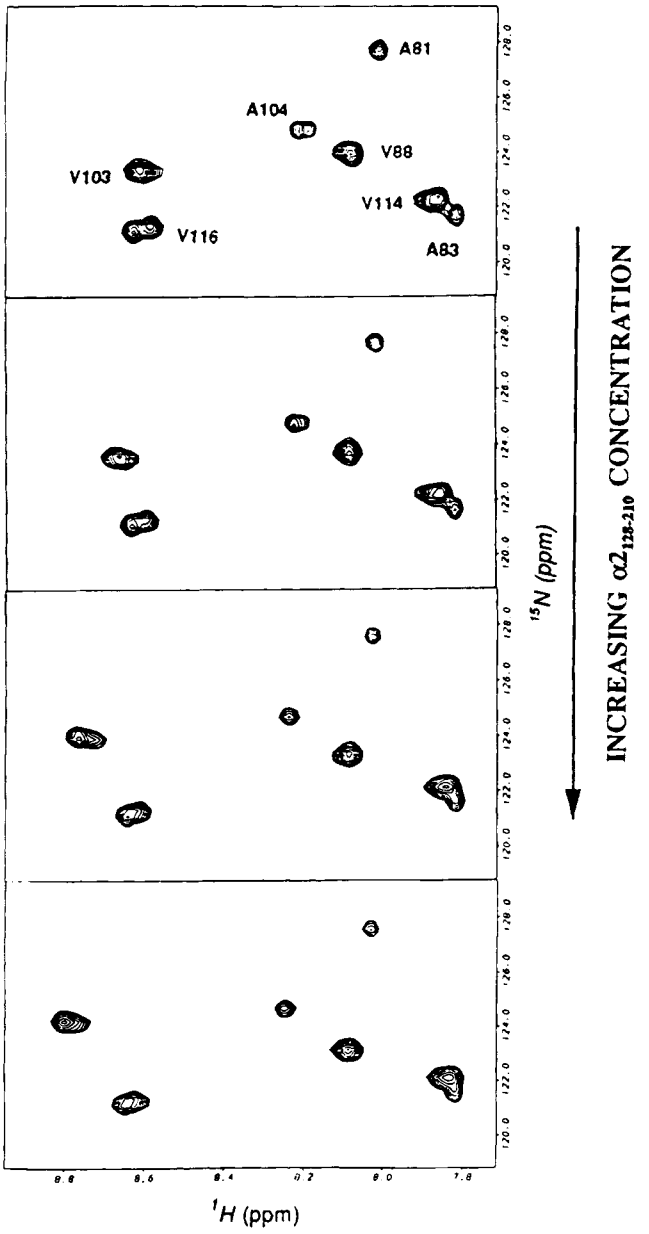

FIGURE 5: HSMQC spectra of $\mathbf{a}1_{66-126}$, selectively ^{15}N-labeled at alanine and valine residues during titration with $\alpha2_{128-210}$. The $\mathbf{a}1_{66-126}$ concentration was kept constant at 0.5 mM. The spectra were taken in H_2O, at 25 °C, in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3, with 5% D_2O for the lock. The spectra are apodized identically and are drawn at the same contour level. The ratio of $\alpha2_{128-210}$ to $\mathbf{a}1_{66-126}$ is 0 (a), 0.15 (b), 0.4 (c), 1 (d), 1.7 (e), and 3 (f).

conformers represented by the doubled peaks bind $\alpha2_{128-210}$ equally. Only one amide resonance, I119, is clearly doubled in the bound $\mathbf{a}1_{66-126}$ spectra.

Resonance Assignment and Secondary Structure Determination of $\mathbf{a}1_{66-126}$, *Bound to* $\alpha2_{128-210}$. Figure 7 shows the HSMQC fingerprint spectrum of ^{15}N-labeled $\mathbf{a}1_{66-126}$ bound to unlabeled $\alpha2_{128-210}$. Amide resonances were assigned by taking advantage of the previously determined assignments for free $\mathbf{a}1_{66-126}$. Due to spectral overlap, however, many resonances could not be tracked over the course of the titration. Sequential backbone resonance assignments were completed using a 3D HSMQC-NOESY spectrum (data not shown) of the heterodimer sample, in which $\mathbf{a}1_{66-126}$ was uniformly labeled with ^{15}N and $\alpha2_{128-210}$ was unlabeled. Two other spectra, 2D DQF-COSY and 3D TOCSY-HSMQC (data not shown), were used to confirm

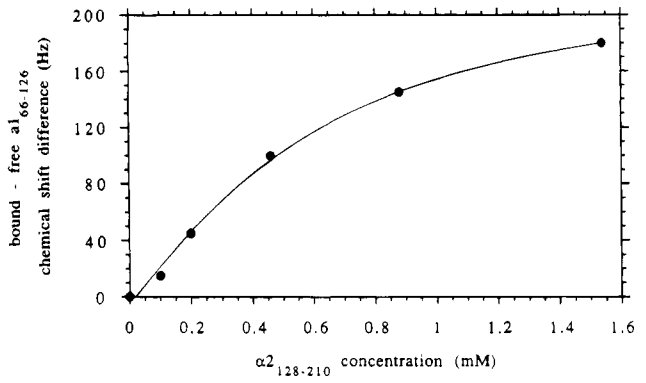

FIGURE 6: Proton chemical shift difference for a representative resonance, Lys98, as a function of $\alpha 2_{128-210}$ concentration. The $\mathbf{a}1_{66-126}$ concentration was held constant at 0.5 mM. The smooth line drawn through the data points is a theoretical curve based on a least-squares fit of a single binding site. This procedure yields a K_d of 0.32 mM and a total shift of 223 Hz (0.44 ppm).

the α proton assignments and to assign aromatic resonances and many side-chain protons. The amide resonance assignments for bound **a**1 are listed in Table 1.

These resonance assignments allowed the identification of NOE cross-peak patterns that describe secondary structure. An HMQC-*J* spectrum (not shown) and a series of *J*-modulated HSQC spectra of the bound complex were collected so that $^3J_{HN\alpha}$ coupling constants could be measured. Data are missing for K66, three prolines, and Asn96, which was too weak in signal to analyze. There are only three residues that have measurable changes in the $^3J_{HN\alpha}$ coupling constants from the values determined for the free protein. The $^3J_{HN\alpha}$ values for Lys100 and Lys98 change only marginally outside experimental error. These measurements imply helical backbone dihedral angles at these residues in both the free and the bound protein. The value for Leu95 changes from 8.9 (+0.3/−0.3) Hz to 7.2 (+0.9/−1.7) Hz. The coupling constant measurement for the Leu95 residue is different enough to suggest real backbone changes. The measured coupling constants and observed NOE connectivities are summarized in Figure 8. As a result of the fast exchange between the free and bound **a**1 forms on the NMR time scale, interpretation of hydrogen exchange data for the heterodimer was complex and is not included in Figure 8.

In free $\mathbf{a}1_{66-126}$, three helical regions are identified from the short- and medium-range NOE connectivities and the $^3J_{HN\alpha}$ coupling constants. There is only one $d_{\alpha N(i,i+3)}$ connectivity reported in helix 1, but this lack of NOEs is attributed to H^α overlap in this helix, which prevents unequivocable identification of these medium-range NOEs. Medium-range NOEs, including $d_{\alpha N(i,i+3)}$ are observed between K98 and E101, suggesting that this stretch of residues retains helical character. Medium-range NOE connectivities involving Ser125 and Met123 suggest that helix 3 extends further than observed for the free $\mathbf{a}1_{66-126}$. Coupling constants for the last three carboxy terminal residues are still greater than 6.0 Hz, though suggesting that the structure may be conformationally averaged and not an ideal α-helix. Breaks in the NOE patterns and measured coupling constants indicate that the loop and the turn seen in the free structure are present in the bound $\mathbf{a}1_{66-126}$. New NOE connectivities are seen between residues in the loop connecting helix 1 and helix 2. Additionally the $^3J_{HN\alpha}$ coupling constant for Leu95, contained in this loop, changes significantly upon $\alpha 2_{128-210}$ binding. The $^3J_{HN\alpha}$ values for residues in the tight turn leading into helix 3 do not change significantly upon heterodimer formation. The NOE patterns observed in this turn for free $\mathbf{a}1_{66-126}$ do not change when $\alpha 2_{128-210}$ binds. A general lack of short- and medium-range NOEs in the amino terminal region (residues 66−78) combined with intermediate $^3J_{HN\alpha}$ coupling constants (6.0 Hz $< J <$ 8.0 Hz) suggests that this segment is still relatively flexible, disordered, and unaffected by heterodimer formation. The mutual NOEs between aromatic side chains of Phe89, Trp117, and Phe118 are maintained in the bound form of $\mathbf{a}1_{66-126}$. Thus, it appears that there are no gross changes in the secondary structure or core of the $\mathbf{a}1_{66-126}$ protein upon $\alpha 2_{128-210}$ binding.

Figure 9 is a plot of the amide proton chemical shift changes in $\mathbf{a}1_{66-126}$ upon binding of $\alpha 2_{128-210}$ as a function of residue position. Amide proton chemical shifts are extremely sensitive to environment and secondary structure. The residues most affected by $\alpha 2_{128-210}$ binding as judged by proton chemical shift changes greater than 0.1 ppm are Glu86, Gln93, Ser94, Leu95, Asn96, Ser97, Lys98, Glu102, and Val103. The only residue to show significant change in measured $^3J_{HN\alpha}$ coupling constant upon heterodimer formation is Leu95. Thus residues showing significant spectral changes upon complex formation reside mostly in the loop between helix 1 and helix 2 and in the amino terminus of helix 2. Glu86 is near the center of helix 1. The amide resonances of the amino terminal region (residues 66−78) are largely unchanged. Relatively small changes in chemical shift are seen for the helix 3 resonances, although resonances that were doubled in the free $\mathbf{a}1_{66-126}$ spectra coalesce into one peak in the bound $\mathbf{a}1_{66-126}$ spectra. The amide resonance for Ile119 is an exception to this observation because it remains doubled in spectra of the complexed $\mathbf{a}1_{66-126}$. The H6/2 protons of Phe118, in helix 3, shift 0.12 ppm upon heterodimer formation, but other aromatic ring protons shift less than 0.1 ppm.

DISCUSSION

We have assigned nearly all of the backbone amide nitrogen and proton resonances of $\mathbf{a}1_{66-126}$ and have identified sequential and medium-range NOE connectivities involving the amide and α protons. These NOEs roughly define the secondary structure of $\mathbf{a}1_{66-126}$ as three helices separated by two turns. Measurements of $^3J_{HN\alpha}$ coupling constants and hydrogen exchange data also support the presence of three helices, joined by a loop and a tight turn. The amino terminal region of the protein appears to be unstructured. These secondary elements comprise the classical homeodomain structure, corroborating earlier predictions that $\mathbf{a}1_{66-126}$ is a homeodomain protein based on sequence homology with the *Antp*, *en*, and α2 homeodomains. The **a**1 amino acid sequence is closely related to the *pbx/exd* family of homeodomains (Monica et al., 1991; Rauskolb et al., 1993). The *pbx1* homeodomain is notable because of its involvement in pre-B acute lymphoblastic leukemias (ALL) (Nourse et al., 1990; Kamps et al., 1990). The **a**1 protein has 63% sequence identity to the *pbx1* homeodomain not only in helix 3 but also in helix 2.

NMR data reported here suggest that the third helix in $\mathbf{a}1_{66-126}$ is not a rigidly structured helix. Structural studies of other homeodomain proteins show that this helix binds in the major groove of the target DNA and is generally called the recognition helix. Despite a high degree of homology

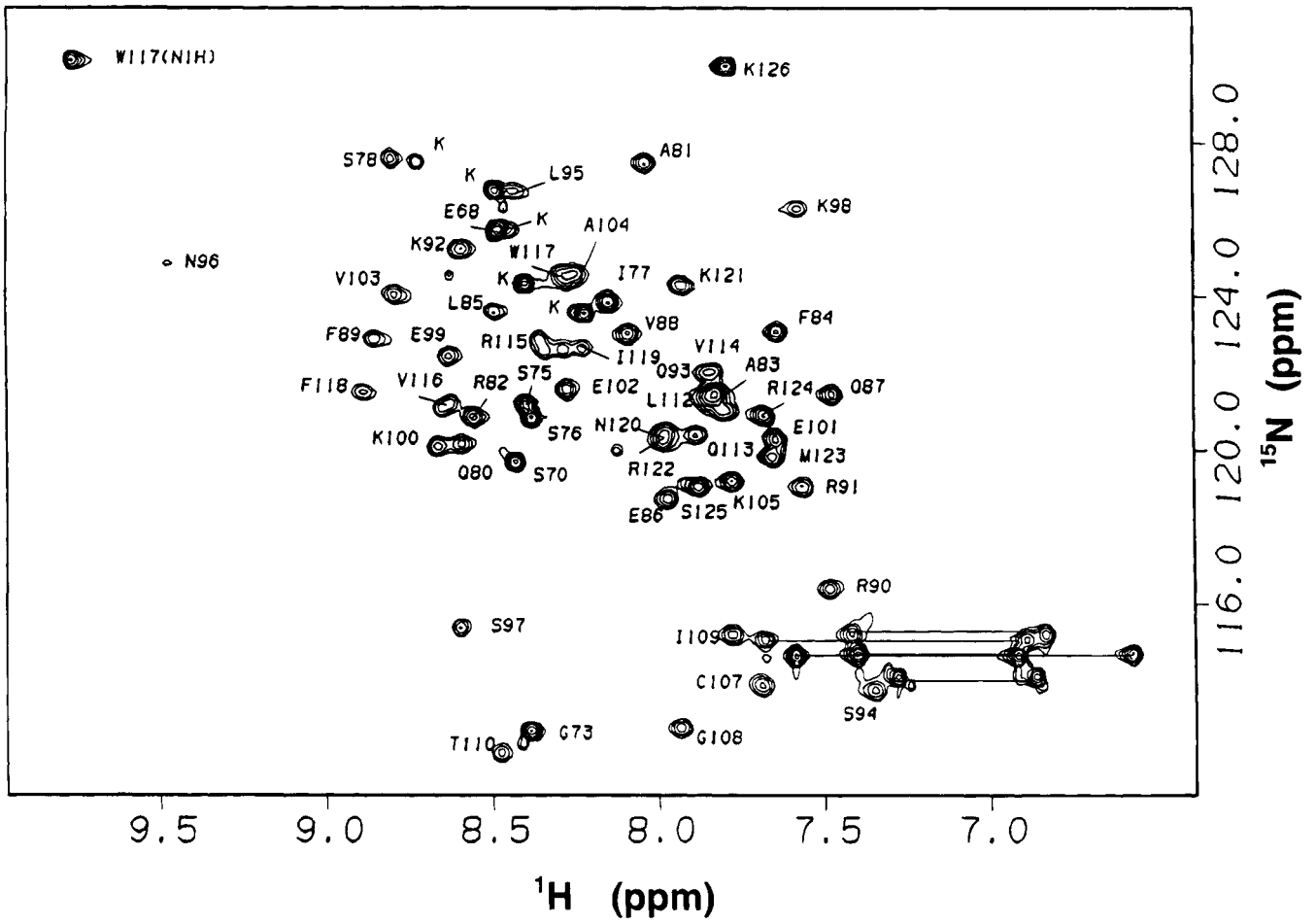

FIGURE 7: HSMQC spectrum of fully ^{15}N-labeled **a**l$_{66-126}$ (0.5 mM) bound to unlabeled α2$_{128-210}$ (1.8 mM). The spectrum was taken in 95% H_2O/5% D_2O at 25 °C in 25 mM deuterated sodium acetate, pH 4.50, 100 mM KCl, and 0.01% NaN_3. The resonance assignments are indicated for each cross-peak.

to the "classical" homeodomain sequence, the **a**l protein has not been shown to bind DNA on its own. Our NMR studies show that the homeodomain recognition helix is formed in the free **a**l$_{66-126}$ fragment. Nonetheless, an abrupt break in NOE connectivities, accompanied by larger coupling constants and faster amide exchange data, suggests that the three residues at the C terminus of the third helix are less regular and perhaps more flexible. Sequence alignment with other homeodomain proteins predicts that the third helix of the **a**l homeodomain extends to the carboxy terminus. Flexibility has also been observed in analogous regions of the *Antp* and α2 homeodomains, even though several conserved residues in this region of helix 3 are thought to be important DNA contacts. Additionally, the amino terminal residues in the third helix of **a**l$_{66-126}$ are affected by some slow conformational equilibrium, such as proline isomerization, as evidenced by peak doublings in the NMR spectra. A proline residue begins the third helix, and two other prolines are located in the unstructured amino terminal arm of the protein. We have not completed unambiguous identification of the source of the doubled peaks.

Similar to results found for other homeodomain proteins, the amino terminal region of **a**l$_{66-126}$ is unstructured. This region of the protein contains a high proportion of charged residues, primarily lysines. X-ray and NMR data of homeodomain−DNA complexes have shown that this flexible arm makes contacts in the minor groove of the DNA target. The interhelical linkages, a loop connecting helix 1 and helix 2 and a tight turn between helix 2 and helix 3, are also characteristic of the homeodomain structures determined previously. The NOE patterns observed in these regions of **a**l$_{66-126}$ do not suggest any regular turn structures. These turns bring hydrophobic residues in helix 1 and helix 3 in close contact at the core of the protein. Trp117, contained in helix 3, is a highly conserved residue in all homeodomains, and NOE cross-peaks suggest that it is close in space to Phe89 in helix 1. This is consistent with the tertiary structures of the *Antp*, *en*, and α2 homeodomains (Otting et al., 1989; Kissinger et al., 1990; Phillips et al., 1991) in which helix 1 and helix 3 cross each other in the vicinity of these residues. The irregular amide protection pattern observed for helix 2 implies that this short helix is relatively solvent exposed as predicted by previous structural studies of homeodomains. Therefore, we conclude that the overall fold of the **a**l fragment is similar to those of the *Antp*, *en*, and α2 homeodomains.

Residue-specific chemical shift changes in **a**l$_{66-126}$ during titration with α2$_{128-210}$ imply that the two proteins associate in a specific manner. Importantly, these results indicate that the heterodimer formation is directed by residues in the homeodomain proper of the **a**l protein. We determined that the two protein fragments associate with a K_d of 0.3 mM, which agrees with previous NMR titration studies of α2$_{128-210}$ (Phillips et al., 1994). The resonances that are most perturbed upon complex formation belong to the loop joining helix 1 and helix 2 and portions of helix 1 and helix 2. Subtle changes in the NOESY spectra of the bound **a**l$_{66-126}$ suggest that the carboxy terminal end of helix 3 is more regular than

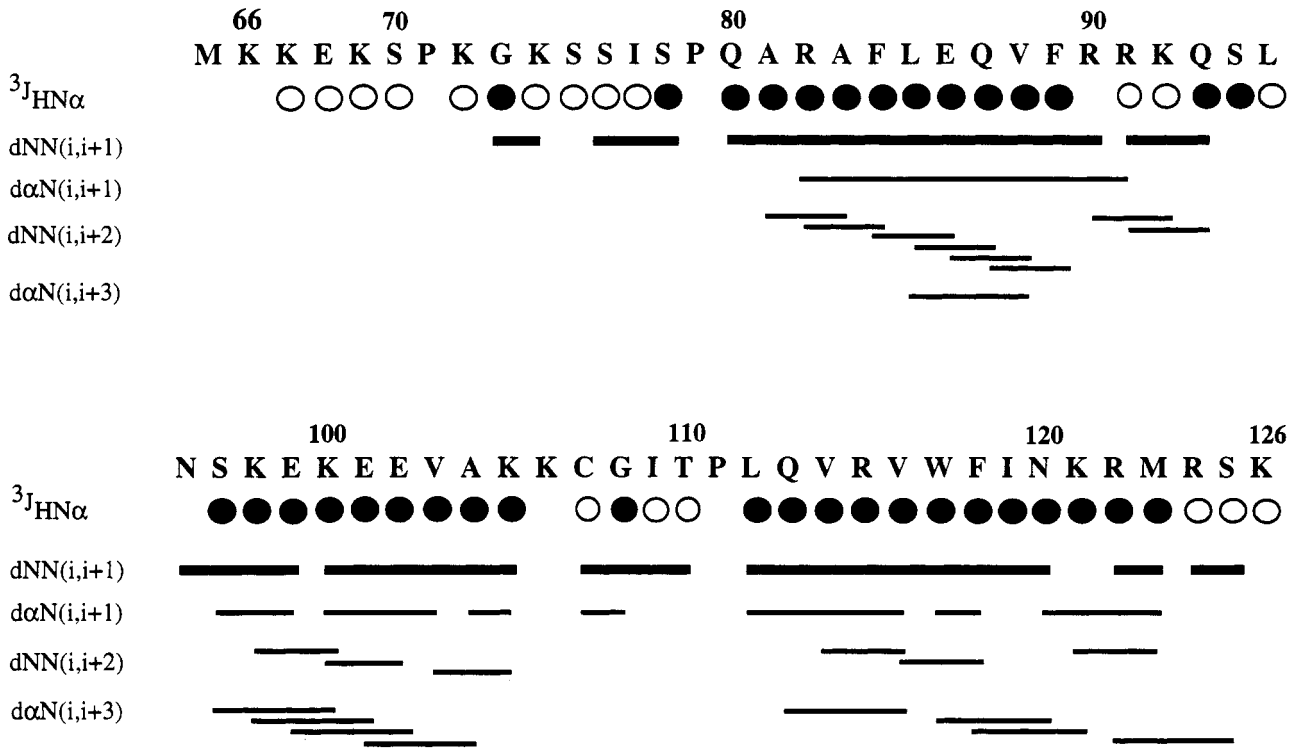

FIGURE 8: $\mathbf{a}1_{66-126}$ amino acid sequence with sequential and medium-range NOE connectivities and $^3J_{HN-H\alpha}$ coupling constants determined for the **a**1 peptide bound to $\alpha2_{128-210}$. Below the sequence, large filled circles represent $^3J_{HN-H\alpha}$ coupling constants of less than 6.0 Hz, gray circles represent intermediate J values (6.0 Hz < J < 8.0 Hz), and open circles represent coupling constants greater than 8.0 Hz. Lines of varying widths indicate NOE connectivities. Broad lines represent $d_{NN(i,i+1)}$ sequential connectivities; narrower lines indicate $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+2)}$, and $d_{\alpha N(i,i+3)}$ connectivities. Line thickness is not intended to imply NOE intensities.

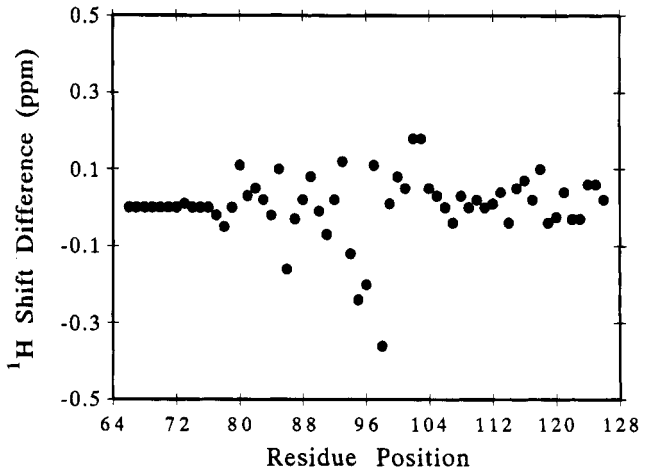

FIGURE 9: Amide proton chemical shift difference between the bound $\mathbf{a}1_{66-126}$ and the free $\mathbf{a}1_{66-126}$ as a function of residue position. The largest chemical shift perturbations are seen for residues in the second helix and residues in the turn between helix 1 and helix 2. Data are taken from the ^{15}N-HSMQC spectra collected as fully ^{15}N-labeled $\mathbf{a}1_{66-126}$ was titrated with $\alpha2_{128-210}$.

free $\mathbf{a}1_{66-126}$, but the proton chemical shift changes for the resonances in this helix are relatively small. Secondary structure determination of the bound $\mathbf{a}1_{66-126}$ shows that the structural elements of the homeodomain fold are largely preserved in the bound form. NOE connectivities between the aromatic side chains at the core of the protein are maintained in the bound form and suggest that the overall protein fold is maintained.

We conclude that the C-terminal tail of $\alpha2_{128-210}$ does not induce gross structural changes in $\mathbf{a}1_{66-126}$, such as the uncoiling or formation of a helix. The region of the homeodomain most affected by heterodimerization is centered around the loop connecting helices 1 and 2. No new regular structure in this loop is suggested by the NMR data although new NOE connectivities are observed. The only significant differences in measured $^3J_{HN\alpha}$ coupling constants between free and bound protein are observed in residues contained in or neighboring this loop. Indeed, the proton chemical shift changes observed are small compared to shift changes in the $\alpha2_{128-210}$ tail upon heterodimerization. The ordering of the tail in $\alpha2_{128-210}$ resulted in chemical shift changes greater than 0.5 ppm for the C-terminal resonances involved, probably due to the formation of hydrogen bonds. These observations suggest that relatively subtle structural changes occur in $\mathbf{a}1_{66-126}$ upon heterodimer formation.

The **a**1/α2 heterodimer is very different from the complexes formed between transcriptional factors of the leucine zipper and helix−loop−helix families (Lamb & McKnight, 1991). Dimerization and DNA-binding functions are segregated in these families. Furthermore, the leucine zipper and HLH proteins dimerize with pseudo 2-fold symmetry with respect to their common domains. Previous NMR studies of the $\alpha2_{128-210}$ protein showed that the association between the $\mathbf{a}1_{66-126}$ and $\alpha2_{128-210}$ proteins is not a symmetric interaction (Phillips et al., 1994). The $\alpha2_{128-210}$ residues responsible for heterodimerization are not part of the α2 homeodomain, rather they are carboxy terminal to it. The **a**1 homeodomain proper on the other hand contains the dimerization and DNA-binding functions. The pattern of **a**1 amino acids affected most by heterodimer formation describes a surface similar to the regions of Oct-1 homeodomain implicated in VP16 (αTIF) recognition (Pomerantz et al., 1992; Lai et al., 1992). Single amino acid substitutions on the surface of helix 1 or helix 2 of the Oct-1 POU homeodomain decreased the ability of Oct-1 to dimerize *in vitro* and *in vivo* (Lai et al., 1992). The similarities between the two systems suggest that homeodomain proteins have a

general mechanism for protein recognition mediated perhaps by the helix 1−loop−helix 2 region.

Indeed the lack of gross changes in the DNA recognition helix of $\mathbf{a}1_{66-126}$ reinforces the notion that biological specificity of the homeodomains many reflect a combination of DNA binding specificity and structurally specific protein−protein interactions. The canonical TAATNN DNA sequence (Scott et al., 1989; Treisman et al., 1992) recognized by a single homeodomain seems too small to allow unique recognition of important regulatory sites. By combining two homeodomains in a heterodimer, the effective size of the DNA target is doubled, and this could lead to much greater potential selectivity in its recognition. The heterodimer structure results in specific spacing of DNA recognition helices, further limiting the DNA sequences that could be recognized by the **a**1 protein. At the same time, we have shown (Phillips et al., 1994) that heterodimer formation results in the ordering of several disordered residues in the tail of $\alpha 2_{128-210}$. Thus, one role of heterodimer formation may be to "prepay" the entropy cost associated (Spolar & Record, 1994) with the ordering of residues in the DNA−heterodimer complex.

ACKNOWLEDGMENT

We thank Sandy Johnson and David Lowry for valuable discussions during the course of this work, Martha Stark for the gift of the original plasmid containing $\mathbf{a}1_{66-126}$, Karen Light-Wahl and Brian Arbogast for mass spectroscopic analyses, and Aida de la Cruz for valuable technical assistance. We also thank Doug Barrick and Janet Schultz for helpful comments on the manuscript.

REFERENCES

Andrés, V., Chiara, M. D., & Mahdavi, V. (1994) *Genes Dev. 8*, 245−257.
Billeter, M., Neri, D., Otting, G., Qian, Y. Q., & Wüthrich, K. (1992) *J. Biomol. NMR 2*, 257−274.
Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1097.
DeSimone, V., & Cortese, R. (1991) *Curr. Opin. Cell Biol. 3*, 960−965.
Dolan, J. W., & Fields, S. (1991) *Biochim. Biophys. Acta 1088*, 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.
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.
Gronenborn, A. M., Wingfield, P. T., & Clore, G. M. (1989) *FEBS Lett. 243*, 93−98.
Hayashi, S., & Scott, M. P. (1990) *Cell 68*, 283−302.
Herskowitz, I. (1989) *Nature 342*, 749−757.
Johnson, A. D. (1992) *Transcriptional Regulation* (McKnight, S. L., & Yamamoto, K. R., Eds.) pp 975−1006, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Kamps, M. P., Murre, C., Sun, Y., & Baltimore, D. (1990) *Cell 60*, 547−555.
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.
Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.
Kristie, T. M., & Sharp, P. A. (1990) *Genes Dev. 4*, 2383−2396.
Lai, J. S., Cleary, M. A., & Herr, W. (1992) *Genes Dev. 6*, 2058−2065.
Lamb, P., & McKnight, S. L. (1991) *Trends Biochem. Sci. 16*, 417−422.
Laughon, A. (1991) *Biochemistry 30*, 11357−11367.
Levy, G. C., & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy*, John 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.
Mendel, D. B., & Crabtree, G. R. (1991) *J. Biol. Chem. 266*, 677−680.
Monica, K., Galili, N., Nourse, J., Saltman, D., & Cleary, M. L. (1991) *Mol. Cell. Biol. 11*, 6149−6157.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44−73.
Neri, D., Otting, G., & Wüthrich, K. (1990) *J. Am. Chem. Soc. 112*, 3663−3665.
Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stanbridge, E., Smith, S. D., & Cleary, M. L. (1990) *Cell 60*, 535−545.
Oppenheimer, N. J., & James, T. L. (1989) *Methods Enzymol. 176/177*.
Otting, G., Qian, Y. Q., Muller, M., Gehring, W. J., & Wüthrich, K. (1989) *EMBO J. 7*, 4305−4309.
Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085−3092.
Phillips, C. L., Vershon, A. K., Johnson, A. D., & Dahlquist, F. W. (1991) *Genes Dev. 5*, 764−772.
Phillips, C. L., Stark, M. R., Johnson, A. D., & Dahlquist, F. W. (1994) *Biochemistry 33*, 9294−9302.
Pomerantz, J. L., Kristie, T. M., & Sharp, P. A. (1992) *Genes Dev. 6*, 2047−2057.
Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1989) *Cell 59*, 573−580.
Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W. J., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1070−1083.
Rauskolb, C., Peifer, M., & Wiechaus, E. (1993) *Cell 74*, 1101−1112.
Scott, M. P., Tamkun, J. W., & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta 989*, 25−48.
Spolar, R. S., & Record, T. M., Jr. (1994) *Science 263*, 777−784.
Sprague, G. F., Jr. (1990) *Adv. Genet. 27*, 33−62.
Stern, S. A., Tanaka, M., & Herr, W. (1989) *Nature 341*, 624−630.
Studier, F. W., & Moffat, B. A. (1986) *J. Mol. Biol. 189*, 113−130.
Treisman, J., Harris, E., Wilson, C., & Desplan, C. (1992) *Bioessays 14*, 145−150.
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.