Elongation of Helix III of the NK-2 Homeodomain upon Binding to DNA: A Secondary Structure Study by NMR

Désirée H. H. Tsao,‡ James M. Gruschus,‡ Lan-Hsiang Wang,§ Marshall Nirenberg,§ and James A. Ferretti*,‡

Laboratory of Biophysical Chemistry and Laboratory of Biochemical Genetics, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-0380

Received August 8, 1994; Revised Manuscript Received October 4, 1994®

ABSTRACT: The secondary structure of the homeodomain encoded by the *NK-2* gene from *Drosophila melanogaster*, in both the free and DNA-bound states, was determined in solution using two- and three-dimensional (2D and 3D) NMR spectroscopy. Proton and ^{15}N studies were carried out on a 77 amino acid residue protein that contains the homeodomain, which was synthesized in *Escherichia coli*. On the basis of NOE connectivities, vicinal coupling constants, and proton−deuterium exchange behavior, three helical segments were found that consist of homeodomain amino acid residues 10−22, 28−38, and 42−52 for the protein in the absence of DNA. The major structural differences between free NK-2 and other homeodomains are the increased internal mobility of the second helix and the shorter length of the third helix, also termed the recognition helix. Despite this shorter helix, NK-2 exhibits high-affinity binding to DNA compared to other homeodomains ($k_D = 2.0 \times 10^{-10}$ M; L.-H. Wang and M. Nirenberg, unpublished results). The formation of the complex of NK-2 with the duplex DNA (TGTGTCAAGTG-GCTGT) significantly increases the thermal stability of the protein. The T_m increases from 25 °C (free NK-2) to >47 °C (DNA-bound NK-2). Also, a dramatic increase in the length of helix III is observed. In the absence of DNA, the DNA recognition helix is 11 amino acid residues long (residues 42−52), whereas in the presence of DNA, the length of this helix extends to 19 amino acids (residues 42−60). The exchange rates of the amide protons in the proton−deuterium exchange experiments are slower in general for the DNA-bound protein than for the free NK-2 homeodomain, which is indicative of a stable conformation for NK-2. A preliminary model of the unbound NK-2 tertiary structure is presented.

Homeobox genes encode a family of proteins that bind to specific nucleotide sequences in DNA and thereby regulate the expression of the corresponding genes (Gehring, 1987; Gehring *et al.*, 1994). The similarity between homeobox genes often is limited to 180 nucleotide residues, the homeobox, which encodes a 60 amino acid residue segment of each protein, termed the homeodomain. The homeodomain functions as the DNA binding site of the protein (Scott *et al.*, 1989; Affolter *et al.*, 1990b) and exhibits amino acid sequence homologies ranging from 20% to 100% (Laughon, 1991). The homeodomains investigated thus far also show structural homologies, for each has at least three α-helices and a flexible arm in the N-terminal portion of the homeodomain (Qian *et al.*, 1988, 1994; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Klemm *et al.*, 1994), and the second and third helices of the homeodomain form a helix−turn−helix DNA binding motif (Aggarwal & Harrison, 1990). The third α-helix, which is termed the DNA recognition helix, and the N-terminal region of the homeodomain are the primary regions of the protein that interact with DNA. The highest degree of amino acid sequence homology is found in the DNA recognition helix, where many of the amino acids that are involved in critical contacts with DNA are conserved. However, there are differences in the amino acid sequence of the recognition helix and the N-terminal region of the homeodomain which, perhaps in conjunction with subtle structural changes of the homeodomain, determine the DNA binding specificity of the protein.

Recently, attention has been given to the length of the recognition helix in various homeodomains studied in the absence of DNA (Qian *et al.*, 1994). A short recognition helix III of about 11 amino acid residues in length (Morita *et al.*, 1993; Qian *et al.*, 1994) and a longer extended helix III of up to 19 residues in length (Qian *et al.*, 1988; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Viglino *et al.*, 1993; Klemm *et al.*, 1994) have been found. Homeodomains with a short or long recognition helix III can bind to DNA with high affinity. The relative importance of both amino acid sequence and conformation of the homeodomain for DNA recognition, as well as the nature of the homeodomain−DNA contacts, are subjects of considerable interest.

In this investigation, we describe the secondary structure of the NK-2 homeodomain protein (Kim & Nirenberg, 1989; Nirenberg *et al.*, 1994) determined by NMR, both in the absence of DNA and bound to a 16-mer DNA segment. The *NK-2* gene is expressed by nuclei in the ventral half of the ventrolateral neurogenic anlagen (that have undergone commitment to the neuroectodermal pathway of differentiation) at the syncytial blastoderm stage of embryonic development (stage 4) (K. Nakayama, N. Nakayama, Y. S. Kim, K. Webber, R. Lad, and M. Nirenberg, unpublished results). The neuroectodermal cells that continue to express the *NK-2* gene give rise to medial neuroblasts that generate neurons in the ventral nerve cord and subesophageal ganglion of

* Author to whom correspondence should be addressed: Building 3, Room 418, NIH, Bethesda, MD 20892-0380.
‡ Laboratory of Biophysical Chemistry.
§ Laboratory of Biochemical Genetics.
® Abstract published in *Advance ACS Abstracts,* November 15, 1994.

Drosophila embryos. The consensus deoxynucleotide sequence recognized by the NK-2 homeodomain is T(T/C)-AAGT(G/A)G (L.-H. Wang, R. Chmelik, and M. Nirenberg, unpublished results). The mouse genome contains six genes that apparently originated by duplication of an ancestral *NK-2* gene during evolution (Price *et al.*, 1992; Lints *et al.*, 1993). These genes have homeodomain amino acid sequences that are 70−95% homologous to the *Drosophila* NK-2 homeodomain. The core consensus deoxynucleotide sequence recognized by one of the mammalian *NK-2*-like genes, rat *TTF-1*, is TCAAGTGTT (Guazzi *et al.*, 1990; Fogolari *et al.*, 1993); therefore, the DNA binding site for an NK-2-like homeodomain has been highly conserved during evolution.

A 77 amino acid residue segment containing the 60 residue homeodomain of *NK-2* was synthesized in *Escherichia coli* and purified to essential homogeneity. Two- and three-dimensional ^{1}H and ^{15}N NMR spectra were used to determine the elements of secondary structure of NK-2 and the complex of NK-2 bound to the 16 base pair DNA duplex (5′-TGTGTCAAGTGGCTGT-3′), as well as a preliminary three-dimensional structure for free NK-2. The secondary structure of NK-2 contains three helical segments analogous to those of the previously studied homeodomains: en (Kissinger *et al.*, 1990), mat-α2 (Wolberger *et al.*, 1991; Phillips *et al.*, 1991), Antp (Qian *et al.*, 1988), the Oct-1 homeodomain portion of the POU domain (Klemm *et al.*, 1994), Oct-3 (Morita *et al.*, 1993), TTF-1 (Viglino *et al.*, 1993), and ftz (Qian *et al.*, 1994). The relative positions and the packing of the three helices, which form a hydrophobic core, are similar in these homeodomains. However, NK-2 in the free state has a lower helix content and is thermally less stable than either Antp or mat-α2, which were both also studied by NMR. This unusual thermal instability was also observed for TTF-1 and *ftz* homeodomains. In addition, helix II in NK-2 and the preceding loop region show a higher degree of flexibility, and helix III is considerably shorter than was found in all of the other homeodomains, except for *ftz* and Oct-3, which in the free state also have relatively short DNA recognition helices. It is of interest to relate the importance of the size of this helix to DNA binding and recognition. We find that the interaction of the NK-2 homeodomain with DNA causes a change in the conformation of the protein, where the length of the third helix is dramatically increased from 11 to 19 residues. Furthermore, the thermal stability of the protein is also increased. The interactions between the three helical segments are stabilized upon binding to the DNA.

MATERIALS AND METHODS

Protein Preparation and Purification. A 77 amino acid residue fragment (NK-2 homeodomain protein) contains Ala at the N-terminus and amino acid residues 538−613 of the NK-2 full-length protein (K. Nakayama, N. Nakayama, Y. S. Kim, K. Webber, R. Lad, and M. Nirenberg, unpublished results), which correspond to amino acid residues −7 to 69 of the NK-2 homeodomain and flanking sequences. The NK-2 homeodomain region was cloned into expression vector pET11d, overexpressed in *E. coli*, and purified to homogeneity as described (L.-H. Wang and M. Nirenberg, unpublished results). All protein preparations were purified by ammonium sulfate fractionation, S-sepharose chromatography, and hydroxylapatite chromatography. The amino acid composition of the purified protein and part of the amino acid sequence were determined. The ^{15}N-labeled NK-2 homeodomain protein was isolated from bacteria grown in 25 L of minimal medium M9 containing 37 mM $^{15}NH_4Cl$ with excess glucose in a fermentor. Molecular weights (averages) of the unlabeled (9354) and ^{15}N-labeled (9488) proteins were confirmed by electrospray mass spectrometry.

DNA. The 16-mer single strands (5′-ACAGCCACTTGA-CACA-3′ and 5′-TGTGTCAAGTGGCTGT-3′) were synthesized by the Midland Certified Reagent Company (Midland, Texas) and were purified by gel filtration and HPLC to 95% purity. The DNA was annealed by dissolving each strand (12 mg) in water (0.5 mL) with 100 mM NaCl, mixing equimolar amounts, heating the solution to 80 °C for 10 min, and letting the solution cool slowly to room temperature. The formation of the double-stranded deoxynucleotide also was checked by 1D NMR.[1]

Circular Dichroism. All CD spectra were recorded for the free NK-2 homeodomain on a Jasco J-600 spectropolarimeter at a protein concentration of about 24 μM, both in water and in buffer (20 mM Tris-HCl (pH 7.3) and 25 mM NaCl). The pH ranges studied varied from 4.0 to 7.3. The pH values of the water solutions were adjusted by the addition of small amounts of HCl.

NMR. All of the NMR measurements were performed on a Bruker AMX600 spectrometer. The free NK-2 samples were dissolved in water to final concentrations of 2.5 (^{15}N-labeled) and 8 mM (natural abundance). The pH values were adjusted to 4.4 by the addition of small amounts of HCl. Spectra for the free NK-2, including experiments carried out in D_2O, were acquired at 12 °C. The NK-2/DNA complex was prepared according to the method used by Omichinski *et al.* (1993) by titrating the protein to a dilute DNA solution until a ratio of 1:0.95 of DNA to protein was reached. The solution containing the NK-2/DNA complex was concentrated by centrifuging in a Centricon 3 (Amicon) to give a final concentration of about 1.5 mM (pH 6.0, 80 mM NaCl). The NMR experiments were carried out at 35 °C. The protein concentration was determined using an extinction coefficient of 10 810 M^{-1} cm^{-1}.

Homonuclear 2D DQF-COSY (Rance *et al.*, 1983), TOCSY (70-100 ms mixing times) (Bax & Davis, 1985), and NOESY (50−150 ms mixing times) (Jeener *et al.*, 1975) were recorded with a spectral width of 7240 Hz for both dimensions, with 1024 complex points and at least 32 scans per FID. Typically, 1024 t_1 values were acquired for each 2D spectrum. Solvent suppression was achieved either by mild presaturation (~35 Hz field strength) during the relaxation delay or by using a spin-lock pulse (Messerle *et al.*, 1989). Spectra were processed with both Bruker UXNMR and Felix software (Biosym Technologies Inc.). Phase sensitive spectra were recorded according to the method of either States (States *et al.*, 1982) or States-TPPI (Marion *et al.*, 1989a).

Both 2D $^{1}H-^{15}N$ HMQC (Bax *et al.*, 1983) and HSQC (Bodenhausen & Ruben, 1980) spectra were recorded with

[1] Abbreviations: 1D, one-dimensional; 2D, two-dimensional; 3D, three-dimensional; NOE, nuclear Overhauser effect; NOESY, 2D NOE spectroscopy; NMR, nuclear magnetic resonance spectroscopy; DQF-COSY, double quantum filtered correlated spectroscopy; TOCSY, total correlated spectroscopy; HMQC, heteronuclear multiple quantum coherence; HSQC, heteronuclear single quantum coherence; FID, free induction decay; TPPI, time-proportional phase increment; CD, circular dichroism; ppm, parts per million.

an ^{15}N spectral width of 1520 Hz, without folding of resonances, and with either 128 or 400 t_1 increments. Broadband decoupling was achieved by the use of a Waltz-16 sequence. The $^3J_{NH-C\alpha H}$ coupling constants were measured with the HMQC-J experiment, according to Kay and Bax (1990).

Both 3D TOCSY-HMQC and 3D NOESY-HMQC on the ^{15}N-labeled protein were performed as described previously (Marion *et al.*, 1989b). DIPSI-2 (Shaka *et al.*, 1988) was used for isotropic mixing (40 ms) in the 3D TOCSY-HMQC experiments. In the NOESY-HMQC experiment, a 100 ms mixing time was used. All 3D experiments were performed with spectral widths of 7246, 1520, and 5434 Hz, with 512 (t_3), 32 (t_2), and 85 (t_1) complex points, respectively. The corresponding acquisition times were 60, 21, and 15.6 ms, respectively. The data were processed on an SGI Indigo computer, using Felix software. The final matrix consisted of 1024 × 64 × 256 real points in the ω_3, ω_2, and ω_1 dimensions. The residual water signal was removed by convolution of the time-domain data according to Marion *et al.* (1989c). Due to an increase in the overlap of resonances, only ^{15}N-filtered experiments were conducted for the NK-2/DNA complex sample. They were carried out in the same way as they were for the free protein, except that the mixing time in the 3D NOESY-HMQC was 80 ms.

For the proton−deuterium exchange experiments, the free protein or NK-2/DNA complex samples were lyophilized and redissolved in D_2O. The concentrations for both D_2O samples were the same as for the water samples. Immediately after the samples were dissolved in D_2O, a series of 2D-HMQC spectra was recorded at 12 °C for the free protein and at 35 °C for the complex. The first experiment started 11 min after the samples were dissolved. The collection time for each experiment was about 47 min. The final HMQC spectrum was taken after 26 h. A total of seven 2D spectra was recorded during this 26 h period.

The three-dimensional structures of NK-2 were generated with the program Dspace (Biosym Technologies Inc.), where 1100 NOESY cross peaks representing conformational constraints from the entire protein were used.

RESULTS

Free NK-2. Initially, circular dichroism (CD) studies were carried out on the NK-2 homeodomain to probe the conformational stability of the protein. The CD spectrum does not change with pH over the range 4.0−7.9. The CD measurements at different temperatures (at pH 4.4 and 7.3) show that the protein undergoes a reversible order−disorder transition with a melting temperature (T_m) of ~25 °C. Analogous behavior is observed in the 1D proton spectra, where peak shifts characteristic of a transition from a folded to an unfolded conformation occur with increasing temperature. Consequently, all subsequent NMR studies were carried out at 12 °C and pH 4.4, where the global fold of the protein is maintained. Under these conditions, the mean residue ellipticity of the protein at 222 nm is observed to be −8000 deg·cm^2/dmol.

Proton assignments of many of the amide, α-, and β-protons were attempted in 90% H_2O/10% D_2O on the natural abundance 77 amino acid residue NK-2 protein by means of the usual 2D TOCSY, COSY, and NOESY type experiments employing the accepted strategies (Wüthrich, 1986). However, significant overlap of cross peaks both in the NH−NH and NH−C$^\alpha$H regions, combined with the absence of some expected cross peaks because of conformational mobility, complicated the assignment process. The presence of eleven arginine and seven lysine residues in the NK-2 homeodomain represents the major contribution to a serious overlap problem. In addition, the four proline residues lacking NH make the sequence assignment problematic. As an assignment aid, a model of NK-2 was simulated using the coordinates of Antp. This simulation allowed us to identify numerous potential close interresidue distances. This approach proved to be quite helpful in the identification of a number of residues. As an example, the model suggested that the ϵ-amino proton of Trp 48 might be near β-protons and one of the δ-methyl groups of Leu 16. This suggestion allowed us to assign the Leu 16 resonances and identify the corresponding sequential cross peaks.

A uniformly ^{15}N-enriched protein was synthesized, and the assignments of the backbone amide resonances were made with the help of 2D H−^{15}N HMQC and HSQC, as well as 3D H−^{15}N NOESY-HMQC and TOCSY-HMQC experiments. Presented in Figure 1A is the HSQC of NK-2. The NK-2 protein synthesized contains three Asn and six Gln residues; eight of the side chain amides were resolved and assigned. The side chain amide protons for Gln 23 were not observed, probably due to conformational exchange between different hydrogen-bonded states. Of the eleven Arg ϵ-protons, eight were resolved; the remaining three are also overlapped and could not be assigned. Assignments of the protons associated with the four Pro residues were attempted in the usual way, by identifying cross peaks involving either the δ-CH_2 (trans) or α-CH (cis) protons with their respective preceding α-CH or NH protons. For Pro −3, Pro 29, and Pro 42, cross peaks with the δ-CH_2 resonances were found. Pro 69 was only identifiable by a process of elimination since the sequential connectivity with α-CH of His 68 is not observed due to its position directly under the residual water resonance at 12 °C. Additional weaker cross peaks were observed in the 3D NOESY-HMQC spectra, which are associated with the residues adjacent to Pro −3 and Pro 69. We attribute their presence to a small population of these two prolines in the cis configuration. No other cross peaks of this type were observed for Pro 29 and Pro 42.

After completion of the assignments from the H−^{15}N correlation spectrum, intraresidual and sequential NOEs involving these amide protons were connected by selecting the corresponding strip through the 3D ^{15}N separated NOESY spectrum (Marion *et al.*, 1989b). In Figure 2A, a strip plot is shown of amino acid residues 44−53. Here, strong $d_{NN(i,i+1)}$ connectivities, along with contiguous $d_{\alpha N(i,i+3)}$ and weaker $d_{\alpha N(i,i+1)}$ cross peaks, are observed. These connectivities are characteristic of regions with helical secondary structure (Wuthrich, 1986). In addition, $d_{\alpha N(i,i+4)}$ cross peaks are also observed, indicating the presence of helices of the α kind. Three stretches of α-helices were observed. The strip for Trp 48 in Figure 1 shows many cross peaks, some of very strong intensity, because of two additional overlapping resonances with close ^{15}N and NH chemical shift values. The His 52 β-proton resonances show an unusually low field shift (1.77 and 1.42 ppm) as a result of ring current effects from the nearby aromatic rings of Phe 20 and Trp 48.

A summary of the short range NOE connectivities is presented in Figure 3A. Between the first and second helices,

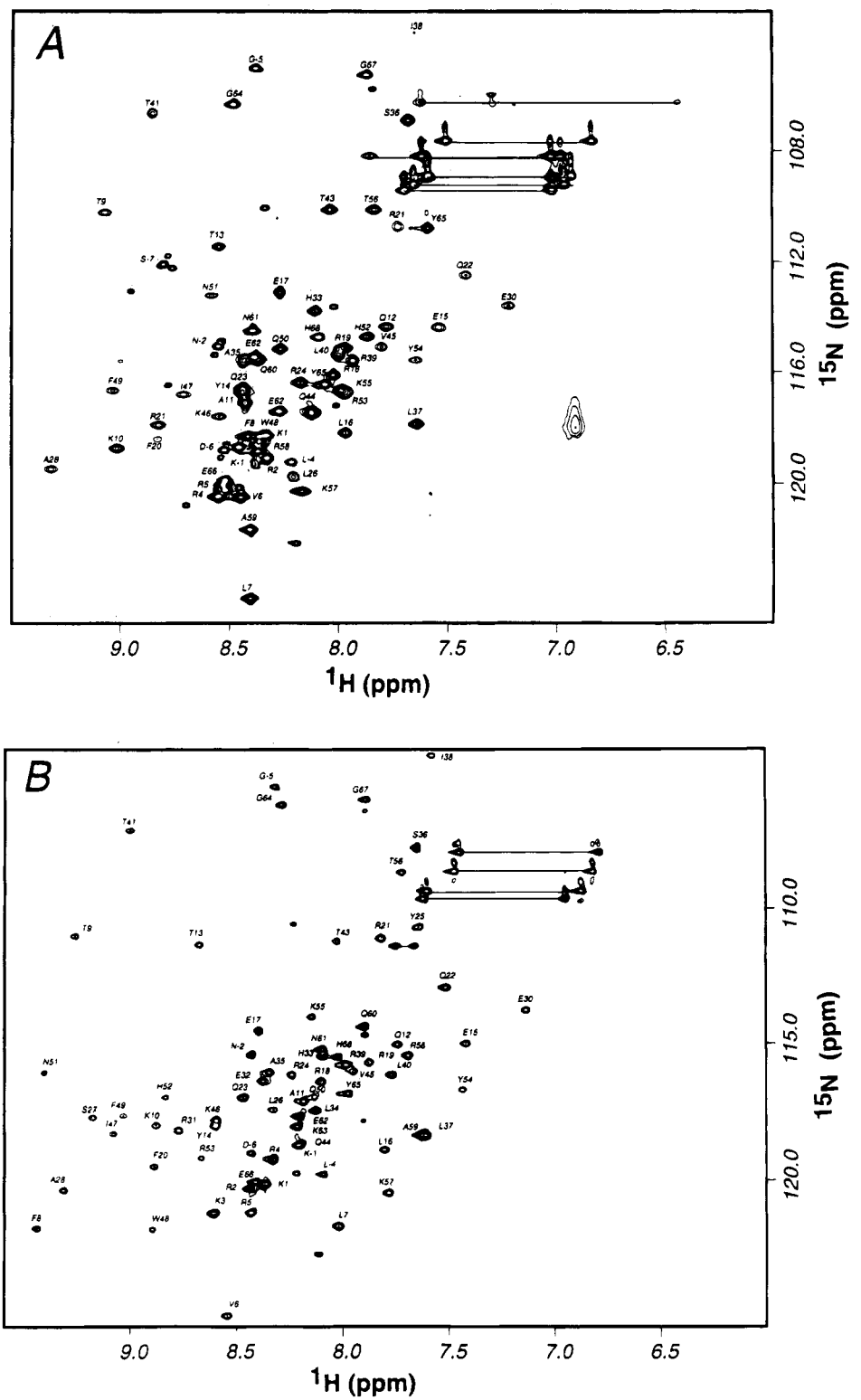

FIGURE 1: Region of the H−^{15}N HSQC spectra of (A) free NK-2 (2.5 mM) in H_2O at 12 °C, pH 4.4, and (B) DNA-bound NK-2 (1.5 mM) in H_2O at 35 °C (80 mM NaCl, pH 6.0). The amide side chains of Gln and Asn are connected by lines. Some amide side chains are of weaker intensity and do not show in the spectra. The ϵ-proton of Trp 48 does not appear in the region shown.

the NOE connectivity pattern indicates a loop region. Between helix II and helix III, we observe NOE signals that are characteristic of a turn. Thus, helices II and III form the well-known helix−turn−helix motif, which was first identified in prokaryotic repressor proteins (Aggarwal & Harrison, 1990). Additional cross peaks not representative of an α-helical conformation were observed in the three helical segments of the protein, such as weak $d_{\alpha N(i,i+2)}$ connectivities. The N-terminal region of the protein (−8 to 9) shows moderate $d_{\alpha N(i,i+1)}$ cross peaks together with weak or absent $d_{NN(i,i+1)}$ cross peaks, which is suggestive of an extended chain conformation. The C-terminal segment (55−69) shows both strong $d_{NN(i,i+1)}$ and $d_{\alpha N(i,i+1)}$ cross peaks, which are interpreted as arising from a mixture of conformational states.

The results of amide hydrogen−deuterium exchange experiments carried out at pH 4.4 and 12 °C are summarized in Figure 2. Typically, amide protons found in unstructured regions of the protein with no intramolecular hydrogen bonds undergo rapid exchange, whereas corresponding amide

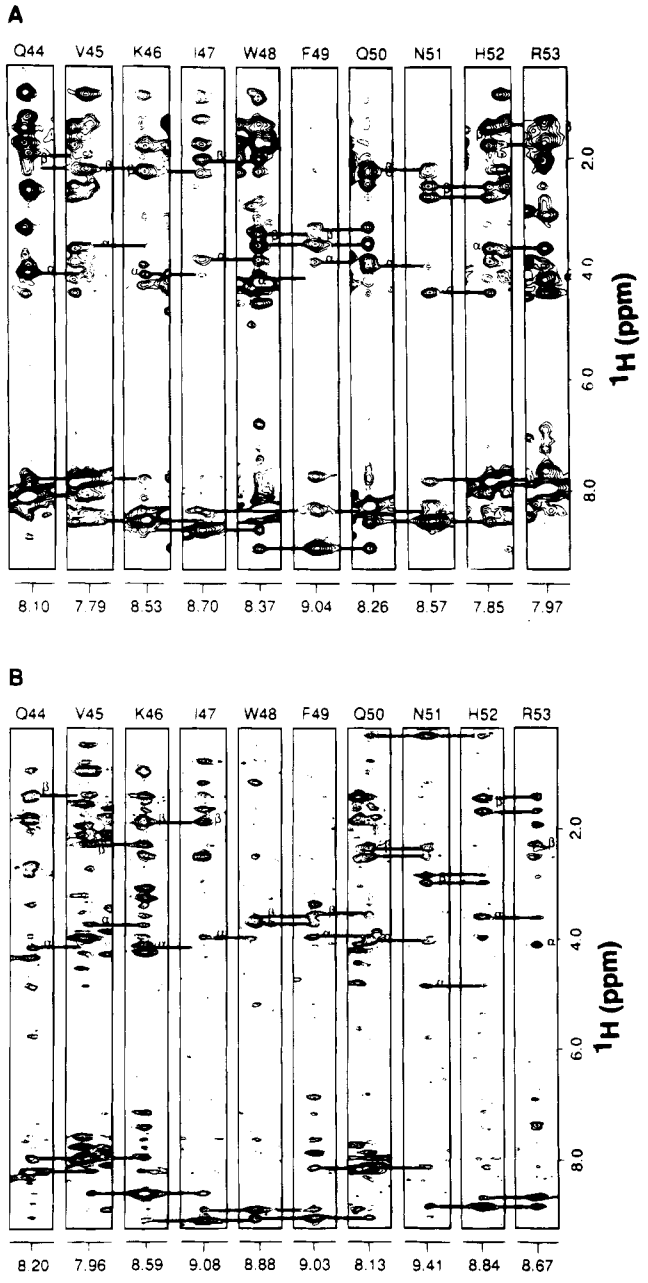

FIGURE 2: Slices of $\omega_1-\omega_3$ planes of the 3D NOESY-HMQC spectra of (A) free NK-2 and (B) DNA-bound NK-2. The slices were taken at the backbone ^{15}N (ω_2) frequencies of residues Gln 44–Arg 53. The intraresidual cross peaks were identified from the TOCSY spectra. The sequential connectivities are indicated with straight lines across the strips.

protons in helical regions exchange slowly with the solvent. After 11 min of dissolution, 29 amide protons remained in the spectrum, whereas after 26 h, the exchange appeared complete. Three regions in the protein were found where the backbone amide NH proton exchange rates were slow, which is consistent with the presence of three helical segments. However, the slow exchange rates are not uniform over the second helical region. Surprisingly, the rate of exchange of the backbone amide proton of His 33, which is in the middle of the putative helix II region, was too fast to be observed. In addition to the slowly exchanging backbone amide protons, the side chain amide NH_2 protons of both Gln 12 and Gln 44 were still present after 11 min, and those for Gln 12 were still observed after 105 min.

Supplementary evidence for the presence of the three helical regions is provided by the values of the $^3J_{NH-C\alpha H}$ coupling constants (see Figure 2). Typical values of 6 Hz or less are observed for those residues having cross peaks corresponding to helix formation. The N-terminal and C-terminal ends of the protein chain show vicinal coupling constant values of 7 Hz or larger, which are consistent with an extended chain conformation or a random coil.

In addition to the sequential assignments that determine the secondary structure of NK-2, a large number of long range NOEs have also been identified for the free NK-2. NOE cross peaks between the first and third helices have been assigned, such as between Trp 48 and Leu 16, Glu 17 and Phe 20, Phe 49 and Leu 16, Arg 24 and Leu 26, etc. On the basis of these long distance NOEs, a diagramatic representation of a preliminary structure of NK-2 is presented in Figure 4, which was generated by the program Dspace. Helices II and III form the helix–turn–helix motif, and the hydrophobic core of the protein is formed with helices I and III packing against each other. This structure is similar to those of Antp (Qian *et al.*, 1988), en (Kissinger *et al.*, 1990), mat α-2 (Wolberger *et al.*, 1991), oct-1 POU (Klemm *et al.*, 1994), and *ftz* (Qian *et al.*, 1994).

NK-2/DNA Complex. Upon binding to DNA, the thermal stability of NK-2 increased significantly. From the 1D NMR spectrum, there was no evidence of denaturation of the complex up to 47 °C. We did not observe peak shifts in the 1D spectrum, which were present in the case of the free protein as the temperature increased, indicating that the protein was losing conformation. All of the NMR experiments for the complex were performed at 35 °C.

Presented in Figure 1B in the H–^{15}N HSQC spectrum of the complex, which shows a better dispersion of peaks when compared to the spectrum of the free NK-2. The sequential assignments for the DNA-bound NK-2 were obtained from 3D TOCSY-HMQC and 3D NOESY-HMQC. Although many resonances in the protein shifted upon binding to DNA, the resonance assignments already determined for the free protein proved to be a valuable aid in confirming the DNA-bound assignments. The assignment of the backbone resonances was straightforward, and most side chain resonances were also assigned. Of the nine Asn and Gln residues, eight side chain amides were assigned. The ninth side chain amide, that from Asn 51, is missing. Explanations for this include the possibility that the amide side chain of Asp 51 may exchange rapidly with the solvent, or the side chain of Asn 51 may slowly alternate between metastable conformations. Either case will result in line broadening. For the arginines, only three ε-protons have been observed. Similar to the free protein, additional cross peaks were observed in the 3D NOESY-HMQC spectrum, which are associated with residues adjacent to Pro −3 and Pro 69.

From the 3D NOESY-HMQC spectrum, strip plots were constructed, and Figure 2B shows the α-helix-type connectivities for residues 44–53. Figure 3B shows a summary of all of the short range NOEs observed for the NK-2/DNA complex. Three stretches of α-helices are observed, analogous to the free protein, except for helix III. Interestingly, the DNA recognition helix III, which in the free NK-2 extends from Pro 42 to His 52, extends from Pro 42 to Gln 60 in the presence of the nucleotide sequence that is recognized by the NK-2 homeodomain.

The results of the hydrogen–deuterium exchange experiments are also presented in Figure 3B. Amide protons belonging to residues in the three α-helices exchange slowly with deuterium. There are several differences in the

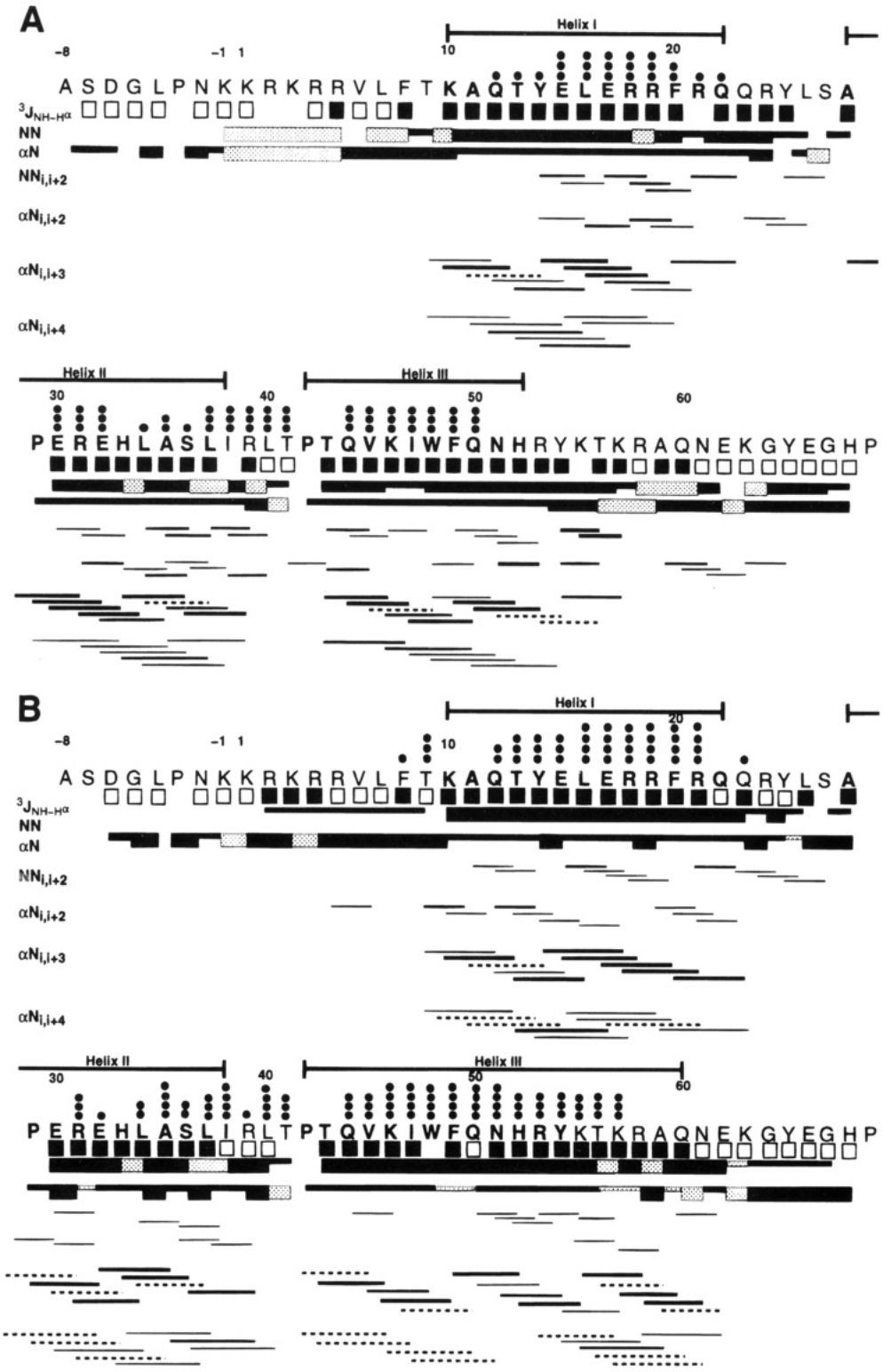

FIGURE 3: Amino acid sequence and summary of NOE contacts observed in the 2D NOESY and 3D NOESY-HMQC spectra of (A) NK-2 and (B) NK-2/DNA. The intensities of the NOEs are characterized as strong and medium for the sequential connectivities, represented by bar thickness. Gray bars indicate cases where the resonances were degenerate. In the case of nonsequential connectivities, the NOEs were categorized as medium, weak, and very weak. Dotted lines indicate cases where the resonances were degenerate and are shown for $\alpha N_{(i,i+3)}$ and $\alpha N_{(i,i+4)}$. The amide protons that exchange slowly are also indicated by solid black dots: one dot represents a lifetime greater than 11 min but less than 58 min; two dots represent a lifetime greater than 58 min but less than 105 min; three dots represent a lifetime greater than 105 min; and four dots represent a lifetime greater than 26 h, for B. All $^3J_{NH-C\alpha H}$ coupling constants less than 6.5 Hz are represented by filled squares, and $^3J_{NH-C\alpha H}$ greater than 6.5 Hz are represented by open squares. For amide protons with no squares, it was not possible to determine the coupling constants, since the cross peak gave an unresolvable splitting. Amino acids that form helix I, helix II, and helix III are indicated by bars above the sequence.

exchange rates when compared to the free protein. Immediately after dissolution of the complex in D_2O, 37 amide protons remained (compared with 29 for free NK-2), and after 26 h, 16 amide protons were present (compared to none for free NK-2), mostly from residues belonging to helix I and helix III. In addition, Phe 8, Thr 9, and Gln 23 exhibit slow backbone amide exchanges, whereas Glu 30 shows fast exchange. The values of the $^3J_{NH-C\alpha H}$ coupling constants

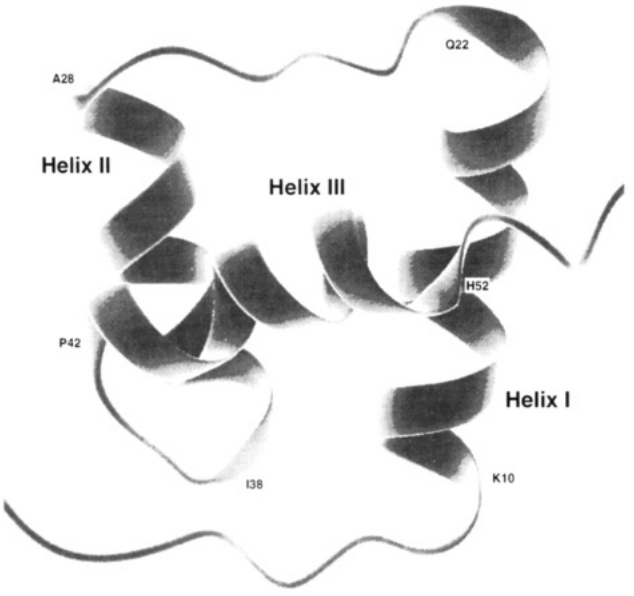

FIGURE 4: Schematic diagram illustrating the preliminary structure of the NK-2 homedomain, obtained using NMR data and the program Dspace (Biosym Technologies, Inc.). The ends of the N- and C-termini are not shown in this view.

(Figure 3B) also support the presence of three α-helices and the elongation of helix III upon binding to DNA.

DISCUSSION

The locations of the three helical segments that define the secondary structure of free and DNA-bound NK-2 are shown in Figure 3. Evidence for the presence of these helical segments comes from the observation of a sequence of relatively intense $d_{NN(i,i+1)}$ connectivities coupled with the presence of $d_{\alpha N(i,i+3)}$ and $d_{\alpha N(i,i+4)}$ cross peaks. Additional data supporting the location of the three helices come from the measurement of $^3J_{NH-C\alpha H}$ coupling constants, which are generally less than 6 Hz for torsional angles corresponding to the α-helical configuration. These NOE and coupling constant values define the presence of these helical segments: residues 10−22 for helix I, 28−38 for helix II, and 42−52 in the free protein case and 42−60 in the complex for helix III. Supplementary evidence for the presence of these three helical segments comes from the backbone amide proton−deuterium exchange data, where the presence of intramolecular hydrogen bonds associated with helix formation results in slower exchange rates.

From the proton−deuterium exchange experiment, slow exchange in a helix normally is observed starting at the fourth residue. However, for NK-2, as with Antp, only the first two backbone amide protons for helices I, II, and III exchange rapidly. Preliminary modeling studies support the interpretation, first proposed by Otting *et al.* (1988), that the slow exchange of the third helical residue arises from hydrogen bonding to the side chain of the residue immediately preceding the helix, which is either threonine or serine. In the NK-2/DNA complex, this also appears to be true, except for Glu 30.

There are several differences in the features of the D_2O exchange experiment in the NK-2/DNA complex when compared to the unbound NK-2. First, the amides of Phe 8 and Thr 9 are protected from exchange in the complex, possibly due to the formation of hydrogen bonds with the DNA. Second, amide protons that belong to residues 51−57, which in the DNA-unbound protein exchange very quickly, are protected from exchange in the NK-2/DNA complex. This agrees with the observed extension of helix III. Third, the overall exchange rates for the complex are much slower than for the free protein. After 26 h, 16 amides are still observed, suggesting that the stability of the protein is increased remarkably. The NK-2/DNA complex is also very stable. For instance, the NK-2/DNA complex has a much lower k_D than the Antp/DNA complex [2.0×10^{-10} M (L.-H. Wang and M. Nirenberg, unpublished results) and 1.6×10^{-9} M (Affolter *et al.*, 1990a) for NK-2 and Antp, respectively]. This suggests that NK-2 should form a more stable complex with DNA than with the Antp/DNA complex.

Although the NOE connectivities, coupling constants, and amide exchange rates are consistent with the formation of three helical regions in positions analogous to those found in other homeodomains, the results also demonstrate that the free NK-2 has some internal mobility and segmental helical stability. While our results apparently show NOEs between protons up to ~4.5 Å apart, several $d_{\alpha N(i,i+2)}$ and $d_{NN(i,i+2)}$ have stronger intensities than expected. In addition, the exchange rates in helix II are not uniform (Figure 2), and His 33 does not appear to be protected at all from exchange. Taken together, these results suggest that helix II has more internal mobility than either helix I or helix III. This helical flexibility has also been observed with the TTF-1 homeodomain, which is highly homologous with the NK-2 homeodomain (Viglino *et al.*, 1993), where helix II in TTF-1 shows a similar nonuniform exchange pattern and His 33 is not protected from exchange. Viglino *et al.* (1993) proposed that the correct geometry required for hydrogen bonding to the His 33 backbone amide is distorted by Pro 29.

The stability of the TTF-1 (Viglino *et al.*, 1993) and *ftz* homeodomain structures free in solution (with 82% and 38% amino acid sequence homologies to NK-2, respectively) is similar to that of NK-2. The NMR studies for TTF-1 and *ftz* proteins were performed at approximately 15 °C to maintain the protein in the folded state. The lack of the third helix extension in free NK-2 is most intriguing when compared to these two homeodomains. Despite the high degree of amino acid sequence homology between TTF-1 and NK-2, the extension of the helix III is present in free TTF-1. On the other hand, *ftz* shows much less sequence homology to NK-2, and the extension is absent. Qian *et al.* (1994) proposed that *ftz* amino acid residue 56 may determine whether or not the extension forms for the unbound homeodomain. In TTF-1 and in Antp, whose third helix also is extended in the unbound form, residue 56 is bulky and hydrophobic (Met for TTF-1 and Trp for Antp). A bulky and hydrophobic residue 56 is needed to interact with the hydrophobic core in an energetically favorable manner. Amino acid residue 56 of the NK-2 and *ftz* homeodomains (Thr and Ser, respectively) is small and hydrophilic. Thus, the free NK-2 structure seems to support the connection between residue 56 and the helix extension.

However, before concluding that the identity of residue 56 alone determines whether or not the extension of helix III is present, one must consider the case for mat-α2. This homeodomain has an extended third helix with Arg, which is not hydrophobic, in position 56. We propose that, in addition to residue 56, one must consider whether or not residue 54 is hydrophobic. This residue usually contacts the major groove of the DNA binding site, and it is exposed to solvent in the free form of the protein. In Antp, residue 54

is Met, whereas for mat-α2, it is Arg. Possibly to keep Met 54 in Antp on the side of helix III that is solvent exposed, the packing of Trp 56 against the core is required to stabilize the helix. For mat-α2, no additional helix stabilization is required. Thus, a bulky, hydrophobic residue 56 may be needed to form the extension of the third helix only when residue 54 is hydrophobic.

Upon binding to DNA, the end of helix III in NK-2 extends from residue 52 to 60. It is rather remarkable that the binding to DNA can cause such a structural change in the protein. Despite the entropic cost involved, NK-2 binds to its site with an affinity ($k_D = 2.0 \times 10^{-10}$ M; L.-H. Wang and M. Nirenberg, unpublished results) comparable to, if not higher than, the other homeodomains mentioned, and furthermore, the NK-2 homeodomain/DNA complex is stable up to at least 47 °C. Given these observations, the actual conformation of residues 53−60 in the unbound structure might not be relevant for binding to or recognition of the DNA. On the other hand, the enhanced flexibility of these residues may readily accommodate the structural changes induced by the interaction with the DNA. The binding affinity of *ftz* to its DNA site is also comparable to those of other homeodomains, which raises the possibility that helix III of other homeodomains similarly may be extended in length upon interacting with an appropriate DNA binding site. Verification of this hypothesis awaits the determination of the DNA-bound structure of *ftz*, as well as other homeodomains such as POU oct-3 that have a relative short third helix in the absence of DNA.

The ability of a homeodomain protein to bind a specific DNA sequence involves a complex interplay between sequence, structure, and solvation. The effects of flexibility and the orientation of the helical segments on DNA recognition clearly are important. To help answer these important questions, the determination of the three-dimensional structure of NK-2 both free in solution as well as bound to DNA is underway.

ACKNOWLEDGMENT

The authors thank Drs. Alice R. Ritter and James Omichinski for useful conversations, Drs. Stephan Grzesiek and Andy Wang for helpful discussions regarding the experimental details, Dr. Joseph Shiloach for growing the *E. coli* cells, and Dr. Henry M. Fales for the mass spectrometry experiments.

REFERENCES

Affolter, M., Schier, A., & Gehring, W. J. (1990a) *Curr. Opin. Cell Biol. 2*, 485−495.

Affolter, M., Percival-Smith, A., Muller, M., Leupin, W., & Gehring, W. J. (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4093−4097.

Aggarwal, A. K., & Harrison, S. C. (1990) *Annu. Rev. Biochem. 59*, 933−969.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.

Bax, A., Griffith, R., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301−315.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.

Fogolari, F., Esposito, G., Viglino, P., Damante, G., & Pastore, A. (1993) *Protein Eng. 6*, 513−519.

Gehring, W. J. (1987) *Science 230*, 1245−1252.

Gehring, W. J., Affolter, M., & Burlin, T. (1994) *Annu. Rev. Biochem. 63*, 487−526.

Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, M.-G., & Di Lauro, R. (1990) *EMBO J. 9*, 3631−3639.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1975) *J. Chem. Phys. 71*, 4546−4553.

Kay, L. E. & Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.

Kim, Y. S., & Nirenberg, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7716−7720.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., & Pabo, C. O. (1990) *Cell 63*, 579−590.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., & Pabo, C. O. (1994) *Cell 77*, 21−32.

Laughon, A. (1991) *Biochemistry 30*, 11357−11367.

Lints, T. J., Parsons, L. M., Hartley, L., Lyons., I., & Harvey, R. P. (1993) *Development 119*, 419−431.

Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A., & Clore, G. M. (1989b) *Biochemistry 28*, 6150−6156.

Marion, D., Ikura, M., & Bax, A. (1989c) *J. Magn. Reson. 84*, 425−430.

Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608−613.

Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., & Kyogoku, Y. (1993) *FEBS Lett. 2*, 107−110.

Nirenberg, M., Nakayama, K., Nakayama, N., Kim, Y. S., Mellerick-Dressler, D., Wang, L.-H., Webber, K., & Lad, R., (1994) *Ann. N.Y. Acad. Sci.* (in press).

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Sthal, S. J., & Gronenborn, A. M. (1993) *Science 261*, 438−446.

Otting, G., Quian, Y. Q., Muller, M., Affolter, M., Gehring, W., & Wüthrich, K. (1988) *EMBO J. 7*, 4305−4309.

Phillips, C. L., Vershon, A. K., Johnson, A. D., & Dahlquist, F. W. (1991) *Genes Dev. 5*, 764−772.

Price, M., Lazzaro, D., Pohl, T., Mattei, M.-G., Ruther, U., Olivo, J.-C., Duboule, D., & Di Lauro, R. (1992) *Neuron 8*, 241−255.

Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., & Wuthrich, K. (1988) *Cell 59*, 573−580.

Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W. J., & Wüthrich, K. (1994) *J. Mol. Biol. 238*, 333−345.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.

Scott, M. P., Tamkum, J. W., & Hartzell, G. W. (1989) *Biochim. Biophys. Acta 989*, 25−48.

Shaka, A. J., lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.

Viglino, P., Fogolari, F., Formisano, S., Bortolotti, N., Damante, G., Di Lauro, R., & Esposito, G. (1993) *FEBS Lett. 3*, 397−402.

Wolberger, C., Bershon, 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 and Sons, New York.