Interactions of the vnd/NK-2 Homeodomain with DNA by Nuclear Magnetic Resonance Spectroscopy: Basis of Binding Specificity

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

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

Received August 12, 1996; Revised Manuscript Received December 17, 1996[⊗]

ABSTRACT: The interactions responsible for the nucleotide sequence-specific binding of the vnd/NK-2 homeodomain of *Drosophila melanogaster* to its consensus DNA binding site have been identified. A three-dimensional structure of the vnd/NK-2 homeodomain−DNA complex is presented, with emphasis on the structure of regions of observed protein−DNA contacts. This structure is based on protein−DNA distance restraints derived from NMR data, along with homology modeling, solvated molecular dynamics, and results from methylation and ethylation interference experiments. Helix III of the homeodomain binds in the major groove of the DNA and the N-terminal arm binds in the minor groove, in analogy with other homeodomain−DNA complexes whose structures have been reported. The vnd/NK-2 homeodomain recognizes the unusual DNA consensus sequence 5′-CAAGTG-3′. The roles in sequence specificity and strength of binding of individual amino acid residues that make contact with the DNA are described. We show, based primarily on the observed protein−DNA contacts, that the interaction of Y54 with the DNA is the major determinant of this uncommon nucleotide binding specificity in the vnd/NK-2 homeodomain−DNA complex.

The homeodomain, the product that results from encoding the homeobox, is a highly conserved DNA-binding domain of approximately 60 amino acid residues that is found in many proteins known to specify positional information and segmental identity in the commitment of embryonic cells to specific developmental pathways (Gehring *et al*., 1994; McGinnis *et al*., 1984; Scott & Weiner, 1984). The example that is of particular interest in this study comes from the *vnd* (*ventral nervous system defective*) gene (White, 1980), where mutations lead to loss of function in early neurogenesis (Jiménez & White, 1995). The *vnd* gene of *Drosophila melanogaster* is the earliest predominantly neural gene regulator found thus far that is expressed in part of the ventrolateral neurogenic anlage, which gives rise to part of the central nervous system of the embryo (Skeath *et al*., 1994; Jiménez & Campos-Ortega, 1990; Mellerick & Nirenberg, 1995). Strong evidence was presented recently (Jiménez & White, 1995) that the *vnd* gene is identical to the *NK-2* gene that has been studied extensively in our laboratories (Nirenberg *et al*., 1995). The 723 amino acid residue protein product of this gene contains the vnd/NK-2 homeodomain. The *NK* family of genes, to which *vnd/NK-2* belongs, was first described by Kim and Nirenberg (1989). The identification of a set of sequentially related genes, all of which are phylogenetically ancient, led to the establishment of the *NK-2* class of homeobox-containing genes (Bürglin, 1993; Harvey, 1996). A unique feature of homeodomains in the NK-2 class is the presence of tyrosine at position 54. The consensus DNA sequence recognized by the vnd/NK-2 homeodomain ($K_d = 1.9 \times 10^{-10}$ M) as well as other homeodomains in the NK-2 class is unusual and consists of the site 5′-CAAGTG-3′ (Damante *et al*., 1994; Chen *et al*., 1995; Harvey, 1996) rather than the canonical 5′-TAATGG-3′. One of the purposes of this paper is to describe the association of Y54 with the DNA and to demonstrate that it is the interaction primarily responsible for recognition of the unusual DNA sequence.

Homeodomains recognize specific nucleotide sequences in DNA and in many cases a single homeodomain is the sole DNA-binding domain for the entire regulatory protein (Scott *et al.*, 1989; Affolter *et al.*, 1990; Lin & McGinnis, 1992; Zeng *et al.*, 1993; Damante *et al*., 1994). Although the overall structures of proteins containing homeodomains most likely vary significantly, structures of the free homeodomains, including vnd/NK-2 (Tsao *et al*., 1994, 1995; Qian *et al*., 1994; Cox *et al*., 1995; Sivaraja *et al*., 1994; Morita *et al*., 1995), as well as those complexed with DNA studied thus far exhibit significant similarities (Billeter *et al*., 1993; Kissinger *et al*., 1990; Wolberger *et al*., 1991; Li *et al*., 1995; Klemm *et al*., 1994; Hirsch & Aggarwal, 1996; Wilson *et al*., 1995). Since homeodomains from a wide range of species recognize a limited variety of DNA sequences, an understanding of the contribution of the interaction of the homeodomain with DNA in exerting specific regulatory effects becomes important.

Our investigation of the vnd/NK-2 homeodomain and its unusual DNA binding behavior has motivated us to examine mechanisms of binding specificity with respect to vnd/NK-2 as well as to the homeodomains in general. The rules for specific protein/DNA interactions are still evolving and there are many open questions in this field. In this study we describe the site specific formation of the vnd/NK-2−DNA complex by combining the experimentally determined intermolecular protein−DNA distances from NMR data with molecular modeling computations and complementary results from methylation and ethylation interference experiments.

[‡] Laboratory of Biophysical Chemistry.
[||] Present address: Genetics Institute, 85 Bolton St., Cambridge, MA 02140.
[§] Laboratory of Biochemical Genetics.
[⊗] Abstract published in *Advance ACS Abstracts,* April 15, 1997.

S0006-2960(96)02006-5

We identify features of the homeodomain that are responsible for sequence-specific DNA binding, examining in terms of these features the roles played by various driving forces in DNA binding strength and specificity.

MATERIALS AND METHODS

Sample Preparation. The expression and purification of the ^{15}N-labeled vnd/NK-2 was as described previously (Tsao *et al.*, 1994). The ^{15}N/^{13}C-labeled vnd/NK-2 protein was purified in *Escherichia coli* grown in minimal medium M9 containing $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources. The vnd/NK-2 homeodomain protein was purified to apparent homogeneity as described previously (Tsao *et al.*, 1994). The molecular weight of the protein was confirmed by electrospray mass spectrometry.

The preparation of the 16-mer duplex DNA and the vnd/NK-2−DNA complex were described previously (Tsao *et al.*, 1994). The protein to DNA ratio of samples for NMR was 1.0:0.95. The concentration of the sample varied from 1.5 to 2.0 mM ^{15}N- or ^{15}N/^{13}C-labeled vnd/NK-2 and natural abundance DNA and contained 80 mM NaCl (pH 6.0) in 90% H_2O/10% D_2O or 100% D_2O. The ^{15}N/^{13}C vnd/NK-2−DNA complex in D_2O was prepared by lyophilizing the complex in water to dryness and redissolving the complex in D_2O.

NMR Spectroscopy. All NMR measurements were performed at 35 °C on a Bruker AMX 600 spectrometer. A variety of double- and triple-resonance experiments were conducted to help obtain the assignments for the protein in the complex, together with isotope edited or isotope-filtered nuclear Overhauser enhancement-(NOESY-) type experiments. For the protein, the spectra obtained were a 3D ^{15}N-edited NOESY (3D NOESY-HMQC) (Marion *et al.*, 1989) for the ^{15}N-labeled protein in H_2O (80 ms mixing time, 512 × 105 × 32 complex data points) and ^{13}C-edited NOESY for the ^{15}N/^{13}C-labeled protein in D_2O (120 ms mixing time, 512 × 150 × 32 complex data points) (Ikura *et al.*, 1990). Additionally, a 3D ^{12}C-filtered/^{13}C-edited NOESY spectrum (120 ms mixing time in D_2O, 512 × 150 × 32 complex data points) was measured in which only the NOE cross peaks between protein and DNA protons are observed (Ikura & Bax, 1992). For the DNA, assignments of exchangeable protons that have resonances between 12 and 14 ppm were obtained from a 2D NOESY experiment (150 ms mixing time in H_2O, 1024 × 390 complex data points) with a 1-1 semiselective excitation pulse (Plateau & Guéron, 1982). All other DNA protons were assigned from 2D ^{12}C-filtered NOESY, 100 and 160 ms mixing times (Ikura *et al.*, 1992), and TOCSY in D_2O, 1024 × 420 complex data points. Here protons attached to ^{12}C are observed, and protons attached to ^{13}C (from the protein) are suppressed (Bax *et al.*, 1994). The Felix software package of Molecular Systems, Inc. was used to process the spectra.

Homology Modeling. To generate an initial model structure for the vnd/NK-2−DNA complex, structural homology with other reported homeodomain−DNA complexes was assumed. All modeling was done using Quanta/CHARMm (Molecular Simulations, Inc.). The backbone of the structured core of unbound vnd/NK-2, residues 8−53 (Tsao *et al.*, 1995), was aligned with the Antennapedia (Antp) homeodomain core from the coordinates of the Antp/DNA complex (Billeter *et al.*, 1993). The 16 base pair deoxynucleotide sequence used in the NMR study (see Figure 1) was generated in B-DNA form with the 5′ ends hydroxylated and aligned to the coordinates of the double-stranded oligodeoxynucleotide of the Antp/DNA complex. The N-terminal arm backbone and side-chain torsional angles were adjusted to correspond to those found for engrailed (en) (Kissinger *et al.*, 1990) and MAT α2 complexes (Wolberger *et al.*, 1991). The backbone torsional angles of the C-terminal end of helix III (residues 53−60) were set at typical α-helix values ($\phi = -57°$, $\psi = -47°$, $\omega = -179°$), since these residues were shown previously to be helical in the vnd/NK-2−DNA complex (Tsao *et al.*, 1994). The unstructured N- and C-terminal ends of the protein were removed from the coordinate file to simplify the molecular dynamics simulations (see Figure 1). A total of 16 sodium ions were placed 3 Å from 16 of the 30 DNA phosphate groups. The ions were placed at those phosphates which were furthest for the vnd/NK-2−DNA interface. The net charge of the vnd/NK-2−DNA/sodium ion system is zero.

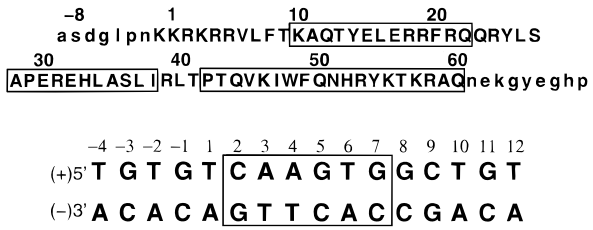

FIGURE 1: Sequences of the vnd/NK-2 homeodomain, plus flanking amino acid residues, and its consensus DNA binding site. The NMR and binding studies were done using the 77 residue protein, while the modeling was done with the 61 residue sequence shown in uppercase letters. The homeodomain is defined as amino acid residues 1−60. The three helical segments are boxed. The consensus DNA sequence was used for the NMR and modeling studies. The core DNA binding sequence is boxed.

Energy Minimization and Molecular Dynamics Simulations. The model of the vnd/NK-2−DNA complex was first solvated with an 8 Å shell of water, which corresponds to about 1000 water molecules. Initially the solvent was energy-minimized to a gradient of 0.42 kJ/(mol·Å) [0.1 kcal/(mol·Å)] holding the vnd/NK-2−DNA/sodium ion complex fixed. Subsequently, the protein, sodium ions, and solvent were energy-minimized to a gradient of 0.42 kJ/(mol·Å) [0.1 kcal/(mol·Å)], holding only the DNA fixed. The following restraints were imposed during this and all subsequent energy minimizations and restrained molecular dynamics simulations:

(1) The torsional angles for the three helical segments of the homeodomain were restrained to be within ±5° of their typical values with a quadratic restraining force for deviations beyond this range with force constant 71 000 kJ/(mol·rad) [300 kcal/(mol·deg)].

(2) The backbone hydrogen-bond lengths for helical regions of vnd/NK-2 were restrained to be between 1.8 and 2.4 Å with a quadratic restoring force beyond this range with force constant 420 kJ/(mol·Å) [100 kcal/(mol·Å)].

(3) Distance restraints for assigned protein to DNA NOE cross peaks were imposed with the distance ranges depending on the cross-peak intensity (see Table 1) using a quadratic restraining force with a force constant of 420 kJ/(mol·Å) [100 kcal/(mol·Å)]. These restraints were supplemented with additional distance restraints for several intraprotein NOE cross peaks of side chains at the DNA/protein interface.

Table 1: Protein−DNA and Selected Intraprotein NOESY Cross Peaks and Restraints

DNA or protein ^{1}H		protein ^{1}H		intensity	restraint (Å)	avg distance[a] (Å)
(+)A3	H1′	R5	Hδ1	very weak	3.5−5.0	4.3 (0.2)
(+)A3	H1′	R5	Hδ2	very weak	3.5−5.0	4.6 (0.2)
(+)A3	H5′ or ″ [b]	L7	CH_3δ1[c]	medium	2.5−6.4	3.8 (0.5)
(+)A3	H5′ or ″ [b]	L7	CH_3δ2[c]	medium	2.5−6.4	5.5 (0.6)
(+)A3	H8	I47	CH_3γ[c]	weak	3.1−6.0	6.2 (0.7)
(+)A3	H1′	I47	CH_3γ[c]	weak	3.1−6.0	5.7 (0.3)
(+)A4	H1′	K3	Hε^{1,2} [d]	medium	2.5−5.4	3.4 (0.2)
(+)A4	H2	K3	Hε^{1,2} [d]	weak	3.1−6.0	5.5 (0.2)
(+)A4	H3′	V6	CH_3γ2[c]	weak	3.1−6.0	5.6 (0.1)
(+)A4	H3′	F8	Hε1	weak	3.1−5.0	4.4 (0.2)
(+)A4	H3′	I47	CH_3γ[c]	weak	3.1−6.0	3.5 (0.1)
(+)A4	H1′	I47	CH_3γ[c]	very weak	3.5−6.0	5.6 (0.1)
(+)A4	H8	I47	CH_3γ[c]	medium	2.5−5.4	2.7 (0.1)
(+)A4	H8	I47	CH_3δ[c]	very weak	3.5−6.0	5.4 (0.1)
(+)G5	H1′	K3	Hε^{1,2} [d]	very weak	3.5−6.0	4.5 (0.1)
(−)T4	CH_3[c]	Y54	Hδ1	weak	3.1−6.0	5.8 (0.3)
(−)T4	CH_3[c]	Y54	Hε1	medium	2.5−5.4	4.2 (0.2)
(−)C5	H5	Y54	Hδ2	very weak	3.5−5.0	3.7 (0.1)
(−)C5	H5	Y54	Hε2	weak	3.1−5.0	3.2 (0.1)
(−)A6	H8	Y54	Hδ2	very weak	3.5−5.0	4.9 (0.1)
(−)A6	H8	Y54	Hε2	very weak	3.5−5.0	4.0 (0.1)
(−)A6	H2′ or ″ [e]	Y54	Hε2	weak	3.1−5.0	3.1 (0.1)
(−)A6	H2′ or ″ [e]	Y54	Hε2	weak	3.1−5.0	4.5 (0.2)
(−)A6	H3′	Y54	Hε2	medium	2.5−4.4	3.8 (0.2)
(−)C7	H5	Q50	Hα	very weak	3.5−5.0	3.8 (0.1)
R5	Hδ1	L7	CH_3δ1[c]	medium	2.5−6.0	5.2 (0.4)
R5	Hδ2	L7	CH_3δ2[c]	medium	2.5−6.0	3.8 (0.4)
R5	Hε	L7	CH_3δ1[c]	medium	2.5−5.0	4.6 (0.1)
R5	Hε	L7	CH_3δ2[c]	medium	2.5−5.0	3.3 (0.2)
K46	Hε1	I47	CH_3δ[c]	medium	2.5−5.0	4.9 (0.1)
K46	Hε2	I47	CH_3δ[c]	medium	2.5−5.0	3.4 (0.1)
K46	Hε1	I47	HN	very weak	3.5−5.0	4.8 (0.1)
K46	Hε2	I47	HN	very weak	3.5−5.0	4.0 (0.2)
Q50	HNε2	I47	CH_3δ[c]	very weak	3.1−5.6	4.2 (0.3)
N51	HNδ1	I47	CH_3γ[c]	weak	*f*	3.4 (0.3)
N51	HNδ2	I47	CH_3γ[c]	weak	*f*	4.1 (0.2)

[a] Average distance and rms deviation (in parentheses) during the 50 ps of restained dynamics. [b] The proton resonance is not stereospecifically assigned so 1.0 Å was added to the upper bound of the restraint. [c] The restraint is to the methyl carbon and 1.0 Å was added to the upper bound of the restraint. [d] The proton resonances of K3 Hε^{1,2} are degenerate, so the restraint was applied to both protons and 1.0 Å was added to the upper bound of the restraint. The shorter of the two average interproton distances is given in the last column. [e] While these proton resonances are not stereospecifically assigned, Y54 Hε2 has NOE cross peaks of approximately the same intensity to both the protons, so no adjustment to the restraint upper bound was needed. [f] No restraints for these cross peaks were employed since these cross peaks were not confirmed until after the time of the simulation.

With the coordinates of the DNA fixed, several simulation cycles of slow heating from 0 to 500 K for 5 ps, dynamics for 10 ps, and reminimization were performed in order to allow the protein structure to relax to a low-energy conformation without undue distortion of the DNA structure. Having achieved a stable, low-energy protein conformation, the constraint on the DNA was removed, and several additional heating−dynamics−minimization cycles were performed at 500 K, allowing both the DNA and protein to simultaneously adjust so as to best satisfy the protein−DNA distance restraints. For the final simulation cycle, the system was heated to 300 K and dynamics was performed for 50 ps.

RESULTS

Methylation and Ethylation Binding Interference. Two vnd/NK-2 binding sites in the 5′-upstream region of the *vnd/NK-2* genomic DNA were characterized by DNase footprinting and by methylation and ethylation binding interference studies; these results will be described elsewhere. The core of both of the high-affinity vnd/NK-2 binding sites matched the consensus sequence, the latter having been selected from a random pool of oligonucleotides.

The segment of the DNA major groove lined by ethylation binding interference sites (see Figure 2) corresponds to where the recognition helix contacts the major groove. In the methylation binding interference experiment, methylation can occur at the N7 atom of guanine (in the major groove) and the N3 atom of adenine (in the minor groove). Methylation binding interference observed at (+)G5 and (+)G7 in the major groove and (+)A3, (+)A4, and (−)A1 in the minor groove correspond to where the recognition helix and the N-terminal arm of the homeodomain contact the DNA, respectively. Methylation interference found at bases (−)G8, (−)A6, and (+)A9 not in direct contact with the homeodomain are difficult to interpret; methylation at these sites may alter DNA structure or charge distribution or even the flexibility of the core nucleotide sequence in a manner unfavorable for binding of the vnd/NK-2 homeodomain protein.

NMR Spectroscopy. The sequence-specific assignments for the vnd/NK-2 homeodomain are essentially complete and those for the DNA are approximately 90% complete. A total of 25 intermolecular NOE cross peaks have been assigned between eight amino acid residues in the protein and eight DNA bases using ^{13}C-edited, ^{15}N-edited, and ^{13}C-edited/^{12}C-

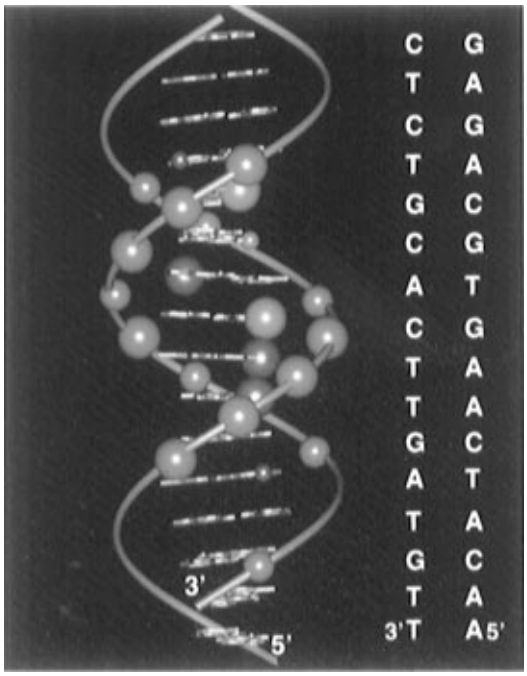

FIGURE 2: Methylation and ethylation binding interference sites for the vnd/NK-2−DNA complex. Methylation interference sites in the major groove (N7 of guanine) are shown as blue spheres, methylation interference sites in the minor groove (N3 of adenine) are shown as purple spheres, and ethylation interference sites (phosphates) are shown as green spheres. Larger spheres indicate a greater degree of binding interference at those sites.

filtered NOESY data. Protein−DNA cross peaks are observed between I47 and Y54 side chains and Q50 backbone and protons in the major groove of the DNA (see Figure 3). The side chains of V6, L7, and F8 have cross peaks with the phosphate backbone in the area where the N-terminal arm of vnd/NK-2 crosses from the minor to the major groove, and the side chains of K3 and R5 have cross peaks with protons in the minor groove (see Table 1). Also, a significant number of NOE cross peaks are observed between the side chains of R5 and L7 only when the protein is bound to DNA, which suggests some ordering of the N-terminal arm upon binding. The modeling suggests that the side chains of K46, Q50, and N51 potentially can contact the DNA, although no corresponding cross peaks were found. The side-chain resonances of these three amino acid residues show significant broadening, presumably due to fluctuations in the side-chain environments at the protein−DNA interface on a millisecond time scale. As a consequence of the broadening, the ability to find cross peaks involving these resonances is greatly reduced. In the analogous study on Antp, no direct protein−DNA cross peaks were reported for N51, although protein−DNA cross peaks were observed for the K46 and Q50 side chains (Billeter *et al.*, 1993). Intraprotein cross peaks between the side-chain resonances of K46, Q50, and N51 and the methyls of I47 were observed, and the corresponding distance restraints were included along with the protein−DNA restraints in the molecular dynamics in order to better characterize the conformations of these side chains with respect to the DNA.

Binding of vnd/NK-2 to the DNA results in large DNA chemical shift changes in the region of the core vnd/NK-2 binding site in DNA and negligible changes at the ends of the DNA. Most of the largest changes are attributed to ring current shifts associated with amino acid residues F8, Y25,

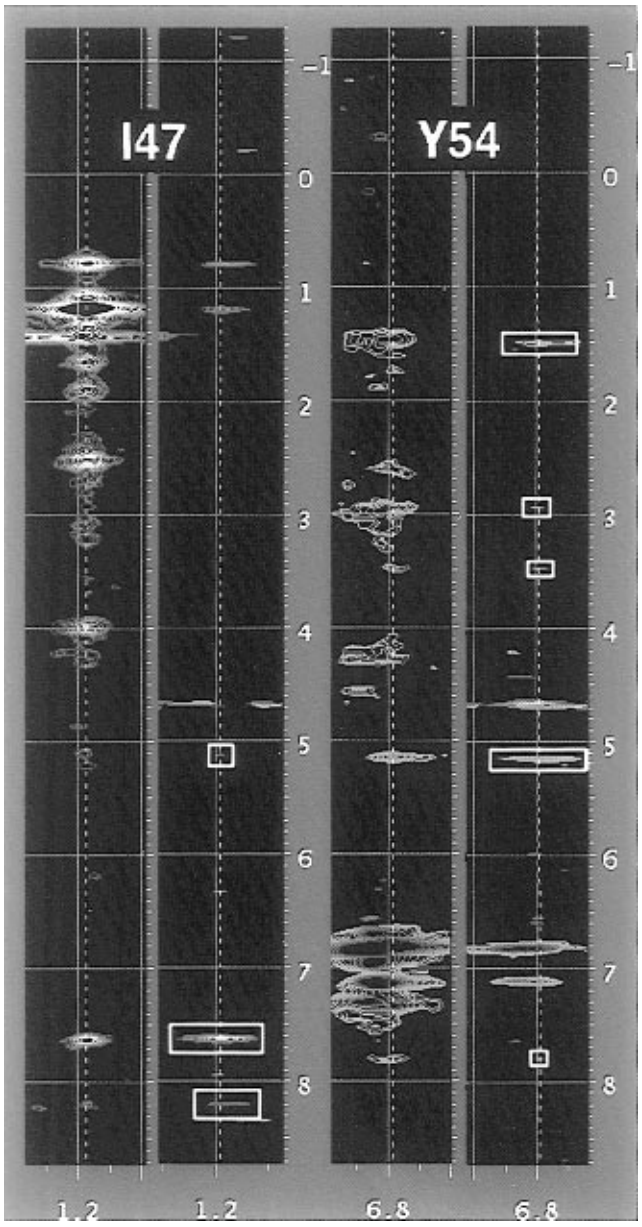

FIGURE 3: Slices from ^{13}C-edited and ^{12}C-filtered/^{13}C-edited 3D NOESY-HMQC spectra for the $CH_3\gamma$ methyl resonance of I47 and the Hϵ1,2 resonance of Y54. In the ^{12}C-filtered/^{13}C-edited slices the cross peaks from DNA are boxed. The orange and blue coloring of cross peaks indicates that the resonances of the attached ^{13}C atoms lie in different regions of the ^{13}C frequency range.

and Y54 that contact the DNA (see Figure 4). For example, the H5 proton of (−)C5 lies directly beneath the ring of Y54, accounting for the large, 2 ppm change in chemical shift. Another large chemical shift change is observed for the H1′ proton of (−)G2, a base that is in contact with the N-terminal arm of vnd/NK-2. This change from 5.47 ppm to an unusual value of 4.96 ppm [H1′ chemical shifts typically fall between 6.3 and 5.3 ppm (Wüthrich, 1986)] does not appear to be related to any ring current effect from the protein (the closest is F8 at ∼13 Å) and thus might be an indication of some distortion of the DNA structure in that region.

Molecular Modeling and Simulations. Homology modeling to other homeodomain−DNA complexes provided a reasonable starting structure for vnd/NK-2. Most of the NOESY-derived restraints were satisfied quite easily at a very early stage in the simulations, thereby lending validity to the homology modeling approach. The resulting structure

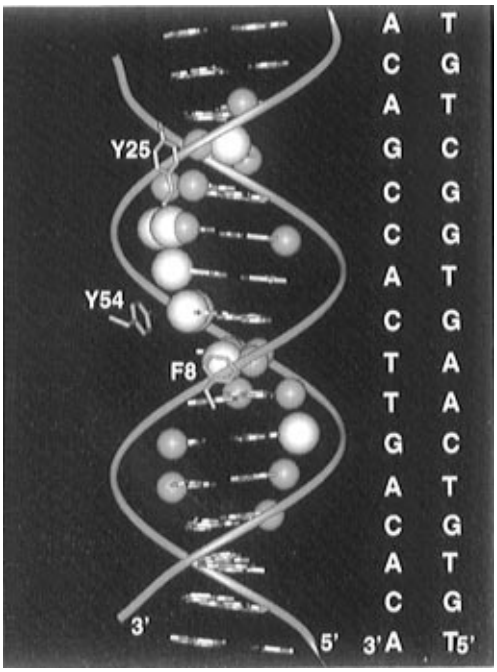

FIGURE 4: Chemical shift changes of the DNA upon binding vnd/NK-2. Base protons H6/H8, ribose H1′, cytosine H5, and thymine C5 methyl resonances are compared; the other proton resonances of the unbound DNA have not yet been assigned. Protons whose chemical shifts changed by greater than 0.2 ppm are shown as gray spheres, and those that changed by 0.1−0.2 ppm are shown as blue spheres. Aromatic rings of the protein that contact the DNA are also shown.

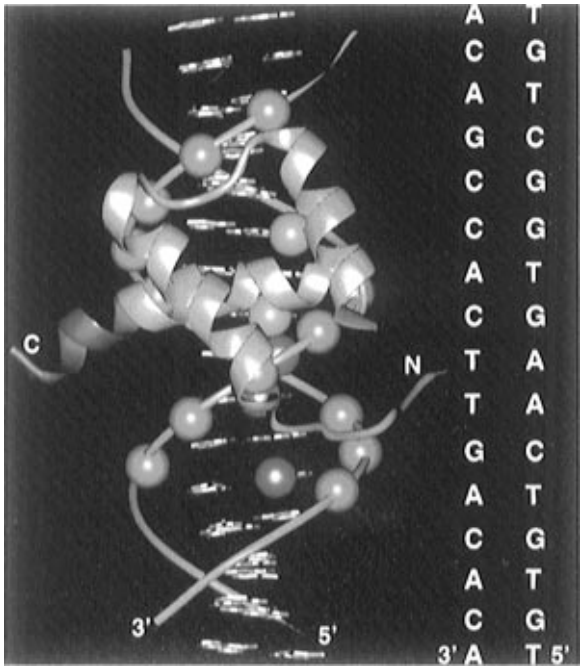

FIGURE 5: vnd/NK-2−DNA complex. The recognition helix is inserted into the major groove, while the N-terminal arm contacts the minor groove. The residues that form the recognition helix extension upon binding are shown in red. Potential methylation and ethylation binding interference sites (see Figure 2) contacted by the protein are shown as blue spheres (major groove), purple spheres (minor groove), and green spheres (phosphates).

is shown in Figure 5. In order to assess the extent to which the restrained dynamics simulation explored the allowed conformations of the protein side chains with observed DNA contacts, we compared the fluctuations in the restrained

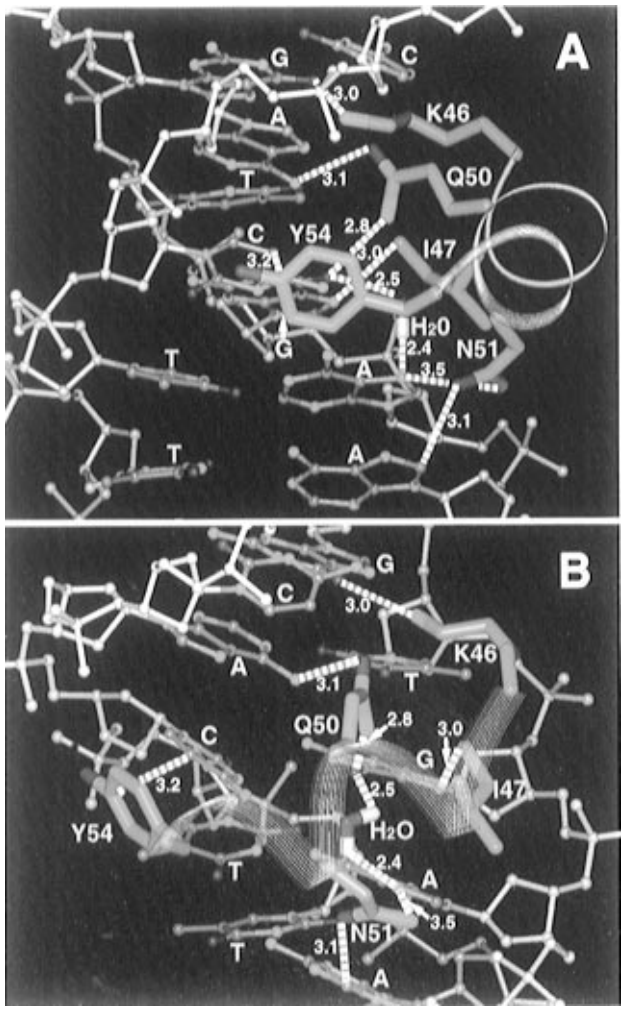

FIGURE 6: Major-groove base contacts in the vnd/NK-2−DNA complex. The five side chains of the recognition helix that contact bases in the major groove are shown with corresponding distances (in angstroms) averaged over the 50 ps simulation. The trapped water is also shown. All distances are heavy atom to heavy atom except for those of the water molecule.

protein−DNA distances during the simulation (Table 1) with those of similarly restrained protein−DNA distances in the set of 16 NMR structures of the Antp complex (Billeter *et al.*, 1993). The root mean square deviations of the vnd/NK-2 distances ranged from 0.1 to 0.7 Å overall, while those of Antp ranged from 0.2 to 0.8 Å for major-groove contact distances and from 0.2 to 1.1 Å for minor-groove and phosphate backbone contact distances. While the 50 ps simulation reproduced the distance variations in the more restrained and sterically hindered major-groove contacts, larger fluctuations were not reproduced for the less restrained, less sterically hindered minor-groove and phosphate backbone contacts.

Contacts between the protein and the major groove of the DNA are shown in Figure 6 and likewise summarized in Figure 7 with other homeodomain−DNA complexes given for comparison. Solvation and subsequent restrained dynamics of the complex revealed a water pocket at the protein−DNA interface in the vicinity of the side chains of N51, Q50, I47, and Y54, analogous to that observed for the Antp, even-skipped (eve), and paired (prd) homeodomain−DNA complexes (Billeter *et al.*, 1993; Hirsch & Aggarwal, 1996; Wilson *et al.*, 1995). Two water molecules were present after initial solvation of the complex, though one molecule

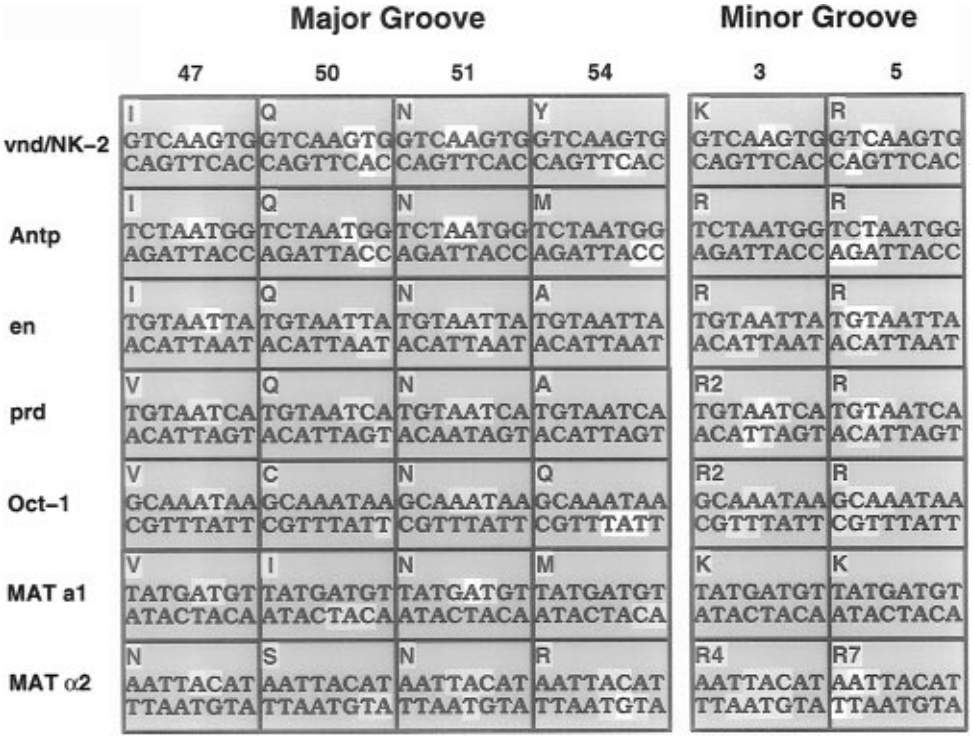

FIGURE 7: Comparison of DNA base contacts in the homeodomain−DNA complex structures. In order to focus on sequence-specific contacts, ribose and phosphate contacts were not included. Yellow indicates the closest contacts between homeodomain amino acid residues and DNA bases, with at least one heavy atom to heavy atom distance less than 3 Å. Green indicates a contact distance of 3−4 Å between heavy atoms, and blue indicates a contact distance of 4−5 Å. The gray regions show where modeling indicates close contact between the amino acid residue and the base, despite the inability to directly observe the contact experimentally. The amino acid residue/base contacts shown above do not necessarily include all such contacts in each individual structure. The coordinates of the recently published even-skipped homeodomain−DNA complex have not yet been examined.

was deleted during the course of the simulation since it interfered with the packing of Y54 and (−)C5. During the course of this simulation, the remaining water molecule was able to form various transient hydrogen bonds including ones between the side-chain amide oxygen atom of N51, the N4 amine of (−)C5, and the O6 atom of (+)G5.

Residues K46, Q50, and N51 lie in the middle of the major groove (see Figure 6) where they can interact with both hydrogen-bond donors and acceptors of the DNA, as well as with water trapped at the protein−DNA interface. The side chain of Q50 appears to form hydrogen bonds from the amide oxygen atom to the N6 amine of (−)A6 and from the amide NH_2 group to the O6 atom of (+)G5. The simulation suggests that hydrogen bonds involving Q50 also are possible with the N4 amine of (−)C7 and the O4 atom of (+)T6. Although no protein−DNA cross peaks are observed for residue K46, the NOE cross peaks between K46 and I47 have the consequence of directing the ϵ-NH_3 group of K46 toward the carbonyl groups of (+)T6 and (+)G7, as well as toward the side-chain carbonyl of Q50. This orientation of K46 is similar to that observed for en (Kissinger *et al.*, 1990) but significantly different from the conformation of K46 in Antp (Billeter *et al.*, 1993), where the ϵ-NH_3 group contacts a phosphate group in the DNA.

The side-chain conformation of N51 initially was set to the coordinates observed for en (Kissinger *et al.*, 1990), Oct-1 (Cox *et al.*, 1995), MATα2 (Wolberger *et al.*, 1991), MATa1 (Li *et al.*, 1995), eve (Hirsch & Aggarwal, 1996), and prd (Wilson *et al.*, 1995) with two hydrogen bonds from the N51 side chain to (+)A4 [one from N51 NH_2 to (+)A4 N7 and one from N51 CO to (+)A4 N6 amine]. The behavior of the invariant N51 in the homeodomain−DNA complex is of particular interest since its position is well-defined in crystal structures but poorly defined for the Antp−DNA complex studied in solution (Billeter *et al.*, 1993). For vnd/NK-2, only the hydrogen bond to the N7 nitrogen remained during the final dynamics simulation, despite the fact that several cycles of dynamics were run initially using distance restraints for the two hydrogen bonds (to allow the complex to relax and to accommodate the conformation imposed on N51). Note that the initial conformation with the two hydrogen bonds, and the subsequent N51 side-chain conformations, always had the side-chain NH_2 group well within 5 Å of the $CH_3\gamma$ methyl of I47, consistent with the observed NOE cross peaks between these two groups (Table 1). We observed identical behavior in modeling of en, where only the hydrogen bond to N7 remained, during a 10 ps solvated dynamics simulation of the en-DNA complex, again despite initial cycles of dynamics where restraints for the N51 hydrogen bonds were included. In a 2 ns, fully equilibrated, fully solvated simulation of the Antp−DNA complex, the two-hydrogen-bond conformation between N51 and adenine was not observed (Billeter *et al.*, 1996).

Residues I47 and Y54 lie at the more hydrophobic edges of the major groove. The side chain of I47 makes hydrophobic contacts with H8 base protons and the ribose rings of (+)A3, (+)A4, and (+)G5. Since the hydrophobic side chain of I47 is fully solvent-exposed in the unbound form, the burial of I47 at the interface in the protein−DNA complex no doubt enhances binding affinity. The interaction of I47 with the edge of the major groove could be important for the proper alignment of the helix III contacts with the DNA bases. Contact between I47 and a thymine methyl

Table 2: DNA Groove Widths

vnd/NK-2[a] major (Å)	MAT α2[a,b] minor[d] (Å)	major[c] (Å)	minor[d] (Å)
20.0	10.4	**16.8**	13.6
19.4	10.6	**18.1**	12.9
19.9	**9.6**	**18.7**	13.4
19.9	**11.9**	16.6	12.8
20.7	**10.5**	17.2	**12.4**
19.3	10.0	18.0	**10.2**
18.8	11.1	18.4	11.6
18.8	10.9	18.8	12.2
20.1	9.9	18.5	12.0
18.7	10.6	18.2	**10.5**
	10.3	16.8	**12.5**
	10.8	17.1	**12.5**
		18.4	12.6
		18.7	12.9
		16.5	13.4
		16.0	14.6

[a] The widths of the sections of the major and minor grooves contacted by the homeodomains are in boldface type and underlined. [b] MATα2 has two bound homeodomains in the X-ray structure. The DNA strands in the MATα2 X-ray structure have an extra 5′-end base, giving the same number of major- and minor-groove distances. [c] *i*, *i* + 5 cross-strand phosphorus−phosphorus distance. [d] *i*, *i* + 3 cross-strand phosphorus−phosphorus distance.

group, as seen in Antp (Billeter *et al.*, 1993) and en (Kissinger *et al.*, 1990), does *not* occur for vnd/NK-2. The hydrophobic interactions involving Y54 include contacts with the methyl group of (−)T4, the H5 and H6 base protons of (−)C5, and the ribose ring of (−)A6 (Figure 6).

The minor groove contacts for the vnd/NK-2−DNA and other homeodomain complexes are summarized in Figure 7. Both K3 and R5 of the vnd/NK-2 homeodomain are inserted into the minor groove in a manner which appears analogous to that found in en (Kissinger *et al.*, 1990). Even though we have three observed cross peaks between K3 and the DNA, a variety of conformations for the side chain of K3 are possible that satisfy the corresponding restraints. Although we have only one experimental restraint between R5 and the DNA, the packing of L7 with its own restraints above R5 serves to restrict the set of conformations accessible to R5. Hydrogen bonds were observed between R5 and (−)A1, (+)C2, and (+)A3.

Homeodomain−DNA Groove Widths. Major- and minor-groove widths are shown in Table 2 for the vnd/NK-2−DNA complex together with a corresponding set of widths from the MATα2 crystal structure (Wolberger *et al.*, 1991) for comparison. (MATα2 was chosen since the structure has two MATα2 homeodomain molecules bound at separate and nonidentical sites on the DNA, providing two sites for examination of groove widths.). For the vnd/NK-2−DNA complex, the distances shown are the averages from the 50 ps dynamics simulation. The groove widths of the Antp (Billeter *et al.*, 1993), en (Kissinger *et al.*, 1990), and Oct-1 (Klemm *et al.*, 1994) complexes also were examined, and in all cases the major groove narrows at the N-terminal side of the recognition helix−DNA interface and widens at the C-terminal side. The minor grooves tend to narrow in the region of the N-terminal arm−DNA interface, with the exception of vnd/NK-2, which shows a narrow−broad−narrow pattern. For vnd/NK-2 this broader distance corresponds to where L7 packs against the (+)A3 ribose ring and C5′ methylene.

DISCUSSION

vnd/NK-2−DNA Binding Specificity. The molecular dynamics simulations using NMR-derived experimental distance restraints show that the vnd/NK-2 homeodomain binds to its cognate DNA utilizing the helix−turn−helix binding motif in a manner similar to that found previously for other homeodomains. The results presented above on vnd/NK-2 and its interaction with DNA, taken together with a detailed examination of several biological studies on homeodomain binding specificities, suggest the following ranking of the importance of the various protein−DNA contacts with respect to binding specificity of vnd/NK-2:

(1) The invariant N51 side-chain amide always must interact with adenine. The interaction may be mediated by water in the protein−DNA interface. (2) The Y54 side chain must contact a cytosine base. The Y54 side chain, which is found thus far only in the NK-2 class of homeodomains, is a primary determinant of recognition of the unusual nucleotide sequence by the vnd/NK-2 homeodomain. (3) The somewhat variable Q50 prefers to interact with T or G bases. (4) The nonconserved L7 contacts the phosphate backbone in a way that favors G/C base pairs. (5) The N-terminal arm of the homeodomain binds in the minor groove of the DNA and interacts preferentially with A/T-rich regions.

Asparagine 51. The invariance of N51 implies a critical role for this residue in homeodomain−DNA recognition and binding. Mutation of (+)A4, the base nearest N51 in vnd/NK-2, results in at least a 50-fold reduction in binding affinity. Mutation of N51 typically eliminates binding and in the best cases reduces binding affinity by factors of 20−1000 with concomitant loss of specificity (Botfield *et al.*, 1994; Pomerantz & Sharp, 1994; Hanes & Brent, 1991). The observed cross peaks between N51 and I47 (Table 1) limit the allowable conformations of the N51 side chain to be within 3−4 Å from the (+)A4 base. However, perhaps due to the substantial broadening of the side-chain amide resonances, no cross peaks from the DNA to the side chain could be detected. In the Antp−DNA complex a similar broadening of the same resonances prevented the observation of N51/DNA contacts (Billeter *et al.*, 1993). This broadening of resonances in the vnd/NK-2 and Antp complexes indicates that the side chain experiences large-scale fluctuations in its environment on a time scale comparable to that of the NMR measurement [characteristically on the order of a millisecond (Billeter *et al.*, 1996)]. Motion of the N51 side chain on the NMR time scale would typically be the explanation for such line broadening. However, the six X-ray structures, en (Kissinger *et al.*, 1990), Oct-1 (Klemm *et al.*, 1994), MATα2 (Wolberger *et al.*, 1991), MATa1(Li *et al.*, 1995), eve (Hirsch & Aggarwal, 1996), and prd (Wilson *et al.*, 1995), all show that N51 makes two hydrogen bonds to the second adenine in their respective consensus DNA sequences. These observations suggest the side-chain resonance broadening might be due to some process other than side-chain motion. In this case water molecules whose interactions with the side chain vary significantly on the NMR time scale would seem the most likely explanation. Such water molecules near N51 could be important for the specificity of nucleotide sequences bound by homeodomains. Consequently, a full explanation of the specificity might ultimately entail

the detailed description of the ensemble of N51/water/DNA contacts.

Tyrosine 54. The amino acid residue in position 54 is variable, although it is usually conserved in evolutionarily related homeodomains (Gehring *et al.*, 1994). Tyrosine in position 54 has been found thus far only in the NK-2 class of homeodomains. The contact between Y54 and the (−)-C5 gives rise to the requirement for the guanine [the base paired with (−)C5] in the sequence 5′-CAA**G**TG-3′ and therefore is a primary determinant of the nucleotide sequence of the vnd/NK-2 homeodomain binding site in DNA. Mutation of this base pair reduces the protein−DNA affinity by a factor of 4−25. Guanine in this position is distinct from all other homeodomain−DNA complexes whose structures have been solved, with thymine being most common (see Figure 7). Mutation of this tyrosine to methionine (found in Antp) results in a 10-fold reduction in the binding affinity (Weiler *et al.*, 1996).

The hydrophobic effect [burial of hydrophobic surface from exposure to solvent (Ha *et al.*, 1989)] provides the principal mechanism favoring the contact between Y54 and (−)C5. In addition, the ηOH is well-positioned to form a hydrogen bond with the cytosine phosphate and thus may be compensated for any loss of enthalpy due to lack of solvation. Hydrophobic contacts between Y54 and the thymine and adenine neighboring (−)C5 suggest that the tyrosine might prefer these bases flanking the cytosine. However, most of the contact with adenine is with its ribose ring and therefore would seem to be a non-base-specific contact. In the case of (−)T4, this base is already required because it is paired with the adenine that interacts with N51, so that preference for this thymine by Y54 would be synergistic.

Glutamine 50. In vnd/NK-2, Q50 contacts the bases at the 3′ end of the binding site. While the NMR experiments detected cross peaks between the backbone protons of Q50 and (−)C7, the side-chain proton resonances were significantly broadened, indicating that this side chain experiences a disordered environment. Similar broadening was observed for the Antp−DNA complex (Billeter *et al.*, 1993), and multiple side-chain conformations for Q50 were observed in the eve homeodomain−DNA X-ray structure (Hirsch & Aggarwal, 1996). The entropy associated with the disordered Q50/DNA interactions could have important thermodynamic consequences favoring homeodomain−DNA complex formation.

A role for residue 50 in sequence specific binding has been demonstrated by mutation experiments. When Q50 is mutated to lysine, many homeodomains bind to sites with CC near the 3′ end of the binding sequence rather than TG or GG (Laughon, 1991). This lysine/glutamine effect on binding specificity has been shown for the bicoid (Hanes & Brent, 1989), prd (Treisman et al, 1989), fushi tarazu (ftz) (Percival-Smith *et al.*, 1990), and NKx-2.1 (TTF-1) (Damante *et al.*, 1994) homeodomains. From the structural standpoint, the preference for TG or GG, (+)T6(+)G7 for vnd/NK-2, appears to be due to the carbonyl oxygens of these bases, which are conveniently placed to form hydrogen bonds with the side-chain NH_2 protons of Q50; and, at the same time, the side chain C═O oxygen atom of Q50 can form hydrogen bonds to NH_2 protons of the paired bases. The Q50 amino acid residue is not as discriminating in its base contacts as Y54, since mutations of (+)T6 or (+)G7 lead to only a 2−7-fold reduction in binding of the vnd/NK-2 homeodomain, compared to a 4−25-fold reduction for mutations of the cytosine (−)C5 contacting Y54.

Leucine 7. Leucine 7 of the vnd/NK-2 homeodomain appears to be important in sequence specificity in a manner not observed in any other homeodomain complex. Leucine in this position is conserved for the NK-2 class of homeodomains but does not occur in any other homeodomain sequences known thus far except cut (Gehring *et al.*, 1994). Leucine 7 is located above the minor groove and above R5 and packs tightly against the (+)A3 ribose ring. This packing gives rise to strong NOE cross peaks to the H5′, H5″ and H4′ protons of the ribose ring. It is rather surprising to find a hydrophobic leucine side chain lying between negatively charged phosphate groups; one would expect a positively charged amino acid side chain. It is established that A/T tracts of DNA have narrow minor grooves but that upon mixing with G/C bases, the minor groove widens (Yoon *et al.*, 1988). The adjacent (+)C2 (−)G2 base pair could thus permit a wider minor groove, which would allow L7 deeper access to the groove and hence better shielding from solvent. This concept is supported by the observation that mutation of (+)C2 to T, a base more commonly found in this position, reduces the binding affinity of the vnd/NK-2 homeodomain to DNA by a factor of 3 (Damante *et al.*, 1994).

N-Terminal Arm. The interactions of the N-terminal vnd/NK-2 homeodomain residues −1 through 5, KKRKRR, with the DNA are characterized by strong electrostatic attraction between the positively charged side chains of amino acid residues and the negatively charged phosphate groups of DNA. Both K3 and R5 are inserted into the minor groove of DNA, while the other side chains are relatively unstructured and form nonspecific contacts with the phosphates. The minor-groove surface of AT base pairs is more electrostatically negative than that for GC base pairs; hence, the N-terminal arm of the vnd/NK-2 homeodomain should prefer A/T-rich binding sites. In addition, A/T-rich sequences have narrower minor grooves than mixed and G/C-rich sequences, bringing the phosphate groups of opposite strands closer together (Yoon *et al.*, 1988). More closely spaced phosphate groups should enhance longer range electrostatic interactions with the N-terminal arm.

CONCLUSION

The protein−DNA contacts observed by NMR, together with homology modeling and restrained molecular dynamics simulations, permit the identification of interactions in the vnd/NK-2−DNA complex that give rise to the unique binding specificity of the NK-2 class of homeodomains. Numerous contacts observed between Y54 and the DNA serve to place the tyrosine ring in direct contact with (−)-C5, and it is this base pair position that most distinguishes the NK-2 class binding-site sequence from those of other homeodomains. Although the importance of residue 50 for determining binding specificity is known, this study presents evidence that residue 54 can be an even more important determinant of specificity. In addition, the strong signals observed from L7 to the DNA phosphate backbone and the slight preference for G/C base pairs in this region suggest a mechanism of binding specificity for the N-terminal arm not seen thus far in any other known homeodomain−DNA

structure. Future studies on the role of waters in mediating the homeodomain−DNA interaction as well as the role of cofactors will provide important additional information regarding sequence specificity.

ACKNOWLEDGMENT

We thank Dr. Solly Weiler for many useful discussions, and the members of the laboratory of Professor Carl Pabo for supplying the coordinates of the Oct-1 homeodomain−DNA complex.

REFERENCES

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

Bax, A., Grzesiek, S., Gronenborn, A. M., & Clore, G. M. (1994) *J. Magn. Reson., Ser A 106*, 269−273.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J., & Wüthrich, K. (1993) *J. Mol. Biol. 234*, 1084−1097.

Billeter, M., Güntert, P., Luginbühl, P., & Wüthrich, K. (1996) *Cell 85*, 1057−1065.

Botfield, M. C., Jansco, A., & Weiss, M. A. (1994) *Biochemistry 38*, 8121−8133.

Bürglin, T. R. (1993) in *Guidebook to the Homeobox Genes* (Duboule, D., Ed.) pp 25−71, Oxford University Press, Oxford, England.

Chen, C. Y., & Schwartz, R. J. (1996) *Mol. Cell. Biol. 16*, 6372−6384.

Cox, M., van Tilborg, P. J. A., de Laat, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C., & Kaptein, R. (1995) *J. Biomol. NMR 5*, 23−32.

Damante, G., Fabbro, D., Pellizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M., Cauci, S., Quadrifoglio, F., Formisano, S., & Di Lauro, R. (1994) *Nucleic Acids Res. 22*, 3075−3083.

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

Ha, J.-H., Spolar, R. S., & Record, M. T., Jr. (1989) *J. Mol. Biol. 209*, 801−816.

Hanes, S. D., & Brent, R. (1989) *Cell 57*, 1275−1283.

Hanes, S. D., & Brent, R. (1991) *Science 251*, 426−430.

Harvey, R. P. (1996) *Dev. Biol. 178*, 203−216.

Hirsch, J. A., & Aggarwal, A. K. (1996) *EMBO J. 14*, 6280−6291.

Ikura, M., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 2433−2440.

Ikura, M., Kay, L. E., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 86*, 204−209.

Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., & Bax, A. (1992) *Science 256*, 632−638.

Jiménez, F., & Campos-Ortega, J. A. (1990) *Neuron 5*, 81−89.

Jiménez, F., Martin-Morris, L. E., Velasco, L., Chu, H., Sierra, J., Rosen, D. R., & White, K. (1995) *EMBO J. 14*, 3487−3495.

Kim, Y., & 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.

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

Lin, L., & McGinnis W. (1992) *Genes Dev. 6*, 1071−1081.

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

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., & Gehring, W. J. (1984) *Cell 37*, 403−408.

Mellerick, D., & Nirenberg, M. (1995) *Dev. Biol. 171*, 306−316.

Morita, E. H., Shirakawa, M., Hayashi, F., Imagawa, M., & Kyogoku, Y. (1995) *Protein Sci. 4*, 729−739.

Nirenberg, M., Nakayama, K., Nakayama, N., Kim, Y. S., Mellerick, D., Wang, L.-H., Webber, K., & Lad, R. (1995) *Ann. N.Y. Acad. Sci. 758*, 224−242.

Percival-Smith, A., Müller, M., Affolter, M., & Gehring, W. J. (1990) *EMBO J. 9*, 3967−3974.

Plateau, P., & Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.

Pomerantz, J. L., & Sharp, P. A. (1994) *Biochemistry 33*, 10851−10858.

Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W. J., & Wüthrich, K. (1994) *J. Mol. Biol. 234*, 1070−1083.

Scott, M. P., & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4115−4119.

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

Sivaraja, M., Botfield, M. C., Mueller, M., Jansco, A., & Weiss, M. A. (1994) *Biochemistry 33*, 9845−9855.

Skeath, J. B., Pangoniban, G. F., & Carroll, S. B. (1994) *Development 120*, 1517−1524.

Treisman, J., Gönczy, P., Malini, V., Harris, E., & Desplan, C. (1989) *Cell 59*, 553−562.

Tsao, D. H. H., Gruschus, J. M., Wang, L.-H., Nirenberg, M., & Ferretti, J. A. (1994) *Biochemistry 33*, 15053−15060.

Tsao, D. H. H., Gruschus, J. M., Wang, L.-H., Nirenberg, M., & Ferretti, J. A. (1995) *J. Mol. Biol. 251*, 297−307.

Weiler, S., Tsao, D. H. H., Gruschus, J. M., Yu, L., Wang, L.-H., Nirenberg, M., & Ferretti, J. A. (1996) *Biophys J. 70*, A342.

White, K. (1980) *Dev. Biol. 80*, 332−344.

Wilson, D. S., Guenther, B., Desplain, C., & Kuriyan, J. (1995) *Cell 82*, 709−719.

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.

Yoon, C., Prive, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332−6336.

Zeng, W., Andrew, D. J., Mathies, L. D., Horner, M. A., & Scott, M. P. (1993) *Development 188*, 339−352.

BI9620060