Journal of Biomolecular NMR, **20:** 111–126, 2001.
KLUWER/ESCOM

Quantitative measurement of water diffusion lifetimes at a protein/DNA interface by NMR

James M. Gruschus & James A. Ferretti*
Laboratory of Biophysical Chemistry, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892-0380, U.S.A.

Received 26 October 2000; Accepted 7 March 2001

Key words: B-factors, binding specificity and affinity, conserved water, cross-relaxation, homeodomain, hydrogen bond, relay magnetization transfer, ROE and NOE, temperature factors, water residence time, X-ray

Abstract

Hydration site lifetimes of slowly diffusing water molecules at the protein/DNA interface of the vnd/NK-2 homeodomain DNA complex were determined using novel three-dimensional NMR techniques. The lifetimes were calculated using the ratios of ROE and NOE cross-relaxation rates between the water and the protein backbone and side chain amides. This calculation of the lifetimes is based on a model of the spectral density function of the water-protein interaction consisting of three timescales of motion: fast vibrational/rotational motion, diffusion into/out of the hydration site, and overall macromolecular tumbling. The lifetimes measured ranged from approximately 400 ps to more than 5 ns, and nearly all the slowly diffusing water molecules detected lie at the protein/DNA interface. A quantitative analysis of relayed water cross-relaxation indicated that even at very short mixing times, 5 ms for ROESY and 12 ms for NOESY, relay of magnetization can make a small but detectable contribution to the measured rates. The temperature dependences of the NOE rates were measured to help discriminate direct dipolar cross-relaxation from chemical exchange. Comparison with several X-ray structures of homeodomain/DNA complexes reveals a strong correspondence between water molecules in conserved locations and the slowly diffusing water molecules detected by NMR. A homology model based on the X-ray structures was created to visualize the conserved water molecules detected at the vnd/NK-2 homeodomain DNA interface. Two chains of water molecules are seen at the right and left sides of the major groove, adjacent to the third helix of the homeodomain. Two water-mediated hydrogen bond bridges spanning the protein/DNA interface are present in the model, one between the backbone of Phe8 and a DNA phosphate, and one between the side chain of Asn51 and a DNA phosphate. The hydrogen bond bridge between Asn51 and the DNA might be especially important since the DNA contact made by the invariant Asn51 residue, seen in all known homeodomain/DNA structures, is critical for binding affinity and specificity.

Introduction

Along with X-ray and neutron diffraction, NMR spectroscopy is capable of detecting water molecules in the hydration shells of macromolecules. Unlike diffraction methods, NMR spectroscopy has the potential to quantitatively characterize dynamic timescales of these water molecules, such as water diffusion rates and hydration site lifetimes (Otting, 1997). NMR methods for measuring water diffusion and hydration site lifetimes include NMR dispersion (Halle and Denisov, 1999; Kiihne and Bryant, 2000), gradient refocusing (Kriwacki et al., 1993), and ROE and NOE experiments (Clore et al., 1990; Otting et al., 1991; Grzesiek and Bax, 1993; Wang et al., 1996; Karimi-Nejad et al., 1999; Phan et al., 1999; Melacini et al., 2000; Tsui et al., 2000).

The current study focuses on the role of slowly diffusing water molecules at a protein/DNA interface.

*To whom correspondence should be addressed. E-mail: jafer@helix.nih.gov

The macromolecular system is a 20 kDa complex of a 77 amino acid peptide encompassing the 60 amino acid vnd/NK-2 homeodomain and a 16 base pair DNA duplex containing the specific binding site sequence of the vnd/NK-2 homeodomain (Tsao et al., 1994). The method employed, i.e. taking the ratios of rates from ROE and NOE experiments to determine hydration site lifetimes, achieves optimal sensitivity and temporal accuracy in the 1 ns range (Otting, 1991). Fortunately, many of the slowly diffusing water molecules at protein/DNA interfaces diffuse at this timescale (Otting, 1997). Two related questions are addressed in this study: (1) Do these slowly diffusing water molecules play an important structural role, such as forming hydrogen bonding bridges between the protein and the DNA, and (2) What contribution, if any, might these water molecules make to the binding affinity and specificity of the complex formation?

Theory and methods

All spectra reported here are from the complex of the ^{15}N-labeled vnd/NK-2 homeodomain bound to its natural abundance cognate DNA at 1.3 mM concentration at 303 K and 309 K, 80 mM NaCl, pH 6, 20% $^{2}H_2O$/80% H_2O. The protein consists of a 77 amino acid peptide containing the 60 residue homeodomain plus 17 flanking residues, and the DNA is a duplex of 16 base pairs. Further sample preparation details have been reported previously (Tsao et al., 1994; Gruschus et al., 1997).

The 3D ^{15}N-edited NOESY-HSQC and ROESY-HSQC spectra were recorded with pulse sequences employing the 45° water flipback technique (Gruschus and Ferretti, 1999). These pulse sequences were specially designed for the quantitative measurement of water to ^{15}N-attached proton cross-relaxation rates. In particular, the water magnetization undergoing cross-relaxation is kept at near 100% of its equilibrium value during the full duration of the mixing time, and special measures are included in the pulse sequences to ensure that axial noise in the water cross-relaxation signals is kept to a minimum. For the ROESY spectrum, the two 25 kHz trim pulses at the beginning and end of the ROE mixing time were 0.5 ms each, rather than the 1 ms duration reported by Gruschus and Ferretti (1999), to reduce sample heating. Using a sample of TSP in a similar volume and NaCl concentration (80 mM) as the homeodomain/DNA complex sample, the ROESY sequence with mixing times of 1.0, 1.5 and 6 ms warmed the sample by approximately 1 °C in all three cases, compared to the same sequence with spinlock and trim pulses set to nearly zero (10 μs) duration. Because of this heating, the temperature was set 1 °C lower (308 K) for the ROESY experiment. A rather strong ROE mixing time spinlock field of 5 kHz was used to ensure that the magnetization not along the off-resonance effective spinlock field axis would be dephased by field inhomogeneity, even for short mixing times. All experiments were interleaved with two mixing times; that is, two FIDs were acquired sequentially for each pair of t_1 and t_2 evolution delays, the first FID for the first mixing time, and the second FID for the second mixing time. All spectra were recorded as $76^* \times 152 \times 512^*$ matrices (times 2 for the two mixing times), with t_1 13.3 ms (F_1 5700 Hz), t_2 50 ms (F_2 1520.5 Hz) and t_3 61.4 ms (F_3 8333.3 Hz). All spectra were linear predicted to double the size of t_1 and t_2, apodized with a shifted squared sinebell window function, and zero-filled in all dimensions. The spectra were recorded on a Bruker DRX600 spectrometer. All spectra were processed with NMRPipe, and the spectra were plotted using NMRDraw (Delaglio et al., 1995).

Two 3D NOESY spectra, one at 303 K and one at 309 K, and one 3D ROESY spectrum at 309 K were acquired and analyzed to obtain the rates of magnetization transfer from the water to the protein ^{15}N-attached protons. Each experiment was interleaved with two mixing times and was processed to yield three spectra for analysis: one from the FIDs recorded with the first mixing time, the second from the second mixing time, and the third from the sum of the FIDs from both mixing times. The sum spectrum provides a better signal to noise ratio than the two individual mixing time spectra, making the peak analysis more facile and robust, especially for weaker cross peaks. For each amide, the intensities (heights) of the diagonal peaks in all three spectra were measured. The water cross peak intensities were measured in the sum spectrum. The indirect ^{1}H dimension linewidths of the water cross peaks and diagonal peaks were measured in the sum spectrum. The rates of decay of the diagonal intensities as a function of mixing time are calculated first, assuming simple exponential decay. For the NOESY spectra the decay rate ρ is the sum of all the cross-relaxation rates from that amide to all other protons. For the ROESY spectra the decay rate ρ is the transverse relaxation rate, $1/T_{1\rho}$. In both NOESY and ROESY ρ also includes the relaxation due to chemical exchange with water.

The water cross-relaxation rates are calculated assuming the following rate equations

$$\dot{w} = \sigma - \rho w, \tag{1}$$

$$w(t) = \frac{\sigma}{\rho}(1 - e^{-\rho t}) \tag{2}$$

where w is the normalized water cross peak intensity, $\dot{w}$ is its derivative with respect to time, and σ is the NOE or ROE cross-relaxation rate plus the water chemical exchange rate. Equations 1 and 2 hold for mixing times much shorter than $1/\sigma$. The relaxation of the bulk water magnetization is negligible during the short mixing times employed in this study. The water cross peaks from the sum spectrum are analyzed because the better signal to noise ratio makes the peak height and linewidth analysis more reliable.

The water cross peak intensity in the sum spectra is normalized using the following expressions

$$w_{\mathrm{sum}} = \frac{y_{\mathrm{sum}}\Delta_{\mathrm{wat}}}{d(0)\Delta_{\mathrm{diag}}}, \tag{3}$$

$$d(0) = d_{\mathrm{sum}}(e^{-\rho\tau_{\mathrm{mix1}}} + e^{-\rho\tau_{\mathrm{mix2}}})^{-1}, \tag{4}$$

where y_{sum} and d_{sum} are the measured water cross peak and diagonal intensities, Δ_{wat} is the average water cross peak linewidth in the indirect ^{1}H dimension, which should be that of bulk water for all cross peaks from water diffusing faster than the NMR timescale (Otting et al., 1991), Δ_{diag} is the diagonal peak linewidth in the indirect ^{1}H dimension, and $d(0)$ is the diagonal peak intensity at zero mixing time. The multiplication in Equation 3 by the ratio of the linewidths accounts for the different relaxation rates of the water magnetization and the ^{15}N-attached proton magnetization during the t_1 evolution period. Using Equation 2, w_{sum} can be expressed as a sum of two terms, and the water rate can be determined using Equations 3 and 4

$$w_{\mathrm{sum}} = \frac{\sigma}{\rho}\left(1 - e^{-\rho\tau_{\mathrm{mix1}}}\right) + \frac{\sigma}{\rho}\left(1 - e^{-\rho\tau_{\mathrm{mix2}}}\right), \tag{5}$$

$$\sigma = \frac{y_{\mathrm{sum}}\Delta_{\mathrm{wat}}\rho(e^{-\rho\tau_{\mathrm{mix1}}} - e^{-\rho\tau_{\mathrm{mix2}}})}{d_{\mathrm{sum}}\Delta_{\mathrm{diag}}(2 - e^{-\rho\tau_{\mathrm{mix1}}} - e^{-\rho\tau_{mix2}})}. \tag{6}$$

In the ROESY experiment the measured water cross-relaxation is reduced by an off-resonance correction factor accounting for the projection of the transferred water magnetization along the effective spinlock axis of the observed amide resonance. The general expression for cross-relaxation in the rotating frame is (Cavanagh, 1996)

$$k = \sin(\theta_1)\sin(\theta_2)\sigma_{\mathrm{NOE}} + \cos(\theta_1)\cos(\theta_2)\sigma_{\mathrm{ROE}}, \tag{7}$$

$$\theta_1 = \arctan(\Omega_1/B_1),\ \theta_2 = \arctan(\Omega_2/B_1), \tag{8a,b}$$

where k is the general cross-relaxation rate, Ω_1 and Ω_2, and θ_1 and θ_2 are the off-resonance frequencies and corresponding angles, and B_1 is the frequency of the spinlock field. In our ROESY experiment, where the water magnetization is on-resonance, θ_1 is zero, thus

$$k = cos(\theta_2)\sigma_{\mathrm{ROE}}, \tag{9}$$

where $\cos(\theta_2)$ is due to the off-resonance frequency of the amide. The amide diagonal magnetization is also reduced by the same $\cos(\theta_2)$ factor due to its projection along its effective spinlock axis during the mixing. Were it not for the trim pulses during the ROE mixing time, these off-resonance correction factors would cancel in the calculation of σ_{ROE}, and the expression for the rate would be the same as for the NOE rate (Equation 6). Because of the trim pulses, which are important to obtain a correctly phased spectrum, the diagonal magnetization is projected first along the trim pulse spin lock axis, then along the ROE spin lock axis, then again along the trim pulse spin lock axis. The off-resonance correction factors for the water cross-relaxation during the trim pulses are different from the correction factors during the normal ROE mixing, and the transverse relaxation rates differ also. The water magnetization transferred during the first trim pulse, corrected for the subsequent projection and decay during the normal ROE mixing and second trim pulse, is

$$\begin{aligned} w_{\mathrm{trim1}} =& \frac{\sigma}{\rho_{\mathrm{trim}}}(1 - e^{-\rho_{\mathrm{trim}}\tau_{\mathrm{trim}}})\cos^2(\theta_{\mathrm{mix}} - \theta_{\mathrm{trim}}) \\ & \cos^2(\theta_{\mathrm{trim}})e^{-(\rho_{\mathrm{mix}}\tau_{\mathrm{mix}} + \rho_{\mathrm{trim}}\tau_{\mathrm{trim}})}, \end{aligned} \tag{10a}$$

$$w_{\mathrm{trim1}} = A\sigma, \tag{10b}$$

where θ_{trim} and θ_{mix}, and ρ_{trim} and ρ_{mix}, and τ_{trim} and τ_{mix} are the off-resonance angles, transverse relaxation rates, and the durations of the trim pulse and the normal ROE mixing period. For simplification of subsequent expressions, A represents all the factors multiplying σ. The corresponding expressions for cross-relaxation during the ROE mixing and second trim pulse are

$$\begin{aligned} w_{\mathrm{mix}} =& \frac{\sigma}{\rho_{\mathrm{mix}}}\cos(\theta_{\mathrm{mix}})(1 - e^{-\rho_{\mathrm{mix}}\tau_{\mathrm{mix}}}) \\ & \cos(\theta_{\mathrm{mix}} - \theta_{\mathrm{trim}})\cos(\theta_{\mathrm{trim}})e^{-\rho_{\mathrm{trim}}\tau_{\mathrm{trim}}}, \end{aligned} \tag{11a}$$

$$\begin{aligned} w_{\mathrm{trim2}} =& \frac{\sigma}{\rho_{\mathrm{trim}}}\cos(\theta_{\mathrm{trim}})(1 - e^{-\rho_{\mathrm{trim}}\tau_{\mathrm{trim}}}) \\ & \cos(\theta_{\mathrm{trim}}) \end{aligned} \tag{11b}$$

$$w_{\mathrm{mix}} = B\sigma,\ w_{\mathrm{trim2}} = C\sigma. \tag{11c,d}$$

The expression for $d(0)$, needed for the normalization of the cross-relaxation rate, is similar to Equation 4, but with the projection factors and corrected relaxation factors,

$$d(0) = d_{\text{sum}} / \left[\cos^2(\theta_{\text{trim}})\cos^2(\theta_{\text{mix}} - \theta_{\text{trim}}) \left(e^{-(2\rho_{\text{trim}}\tau_{\text{trim}} + \rho_{\text{mix}}\tau_{\text{mix1}})} + e^{-(2\rho_{\text{trim}}\tau_{\text{trim}} + \rho_{\text{mix}}\tau_{\text{mix2}})}\right)\right]. \quad (12)$$

The resulting expression for the ROE cross-relaxation rate is

$$\sigma = \frac{y_{\text{sum}}\Delta_{\text{wat}}}{d_{\text{sum}}\Delta_{\text{diag}}}\left[e^{-(2\rho_{\text{trim}}\tau_{\text{trim}} + \rho_{\text{mix}}\tau_{\text{mix1}})} + e^{-(2\rho_{\text{trim}}\tau_{\text{trim}} + \rho_{\text{mix}}\tau_{\text{mix2}})}\right]\frac{\cos^2(\theta_{\text{trim}})\cos^2(\theta_{\text{mix}} - \theta_{\text{trim}})}{A_1 + A_2 + B_1 + B_2 + 2C}, \quad (13)$$

where the subscript on A and B corresponds to the mixing time to be used in their expressions (C does not depend on τ_{mix}). The transverse relaxation rates during the trim pulses, needed in Equation 13, can be calculated from the ROE and NOE diagonal decay rates with the following expression

$$\rho_{\text{trim}} = \left[\frac{\rho_{\text{ROE}} - \rho_{\text{NOE}}\sin^2(\theta_{\text{mix}})}{\cos^2(\theta_{\text{mix}})}\right]\cos^2(\theta_{\text{trim}}) + \rho_{\text{NOE}}\sin^2(\theta_{\text{trim}}). \quad (14)$$

In practice, one can simplify Equation 13 by substituting ρ_{mix} (i.e., ρ_{ROE}) values for ρ_{trim} values, which results in less than a one percent difference in the ROE cross-relaxation rates calculated in this study, significantly less than experimental error.

The spectral density function used to calculate the ROE and NOE rate curves incorporates hydration site motions on three timescales (Halle and Wennerström, 1981; Lipari and Szabo, 1982). These motions are fast vibrational and re-orientational motions of the water protons and the protein protons with which the water interacts, diffusion of the water protons into and out of the hydration site, and macromolecular tumbling of the water/protein system. The expression for the spectral density function is

$$J(\omega) = \left[\frac{(1 - a^2)\tau_f}{(1 + \omega^2\tau_f^2)} + \frac{a^2\tau_d}{(1 + \omega^2\tau_d^2)}\right], \quad (15)$$

where ω is the frequency of the spectrometer, a^2 is the fast timescale order parameter, τ_f is the effective fast motion time and τ_d is the effective diffusion time. The macromolecular rotational correlation time, τ_c, enters into the expression for τ_d. In macromolecular systems,

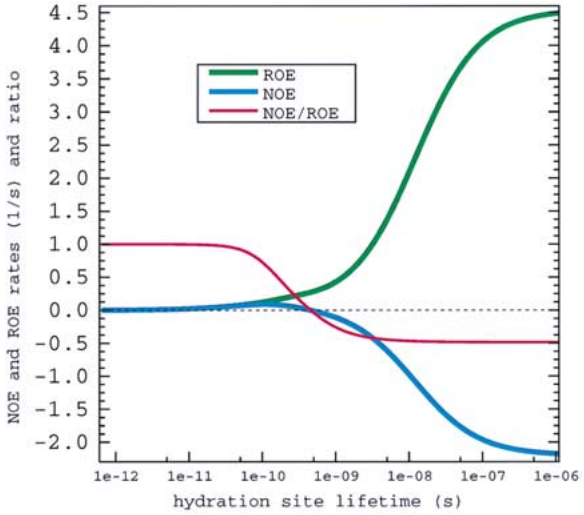

Figure 1. ROE and NOE rate curves as a function of water residence time, with the NOE/ROE rate ratio also shown. The curves were calculated assuming two water protons interacting with one protein proton both at an effective distance r_{eff} of 3.0 Å, using spectral density parameters of $\tau_{\text{fast}} = 20$ ps, $\tau_c = 12$ ns, and $a^2 = 0.3$.

it is possible for water molecules to reside in hydration sites for times longer than τ_c (>10 ns), water in buried sites, for example (Otting, 1997). For these hydration sites, the tumbling of the macromolecule can average the correlation function to zero before the diffusion into and out of the hydration site can have a significant impact on the dipole-dipole Hamiltonian correlation function. Mathematically, the smooth transition to this slow diffusion regime can be achieved by having τ_d asymptotically approach the value of τ_c for sites with residence times that exceed τ_c (Clore et al., 1990)

$$\tau_d = \left[\frac{1}{\tau_{\text{res}}} + \frac{1}{\tau_c}\right]^{-1}, \quad (16)$$

where τ_{res} is the residence time of the water. By the same logic as above, this smooth transition should be applied for τ_f if τ_{res} is shorter than τ_f

$$\tau_f = \left[\frac{1}{\tau_{\text{res}}} + \frac{1}{\tau_{\text{fast}}}\right]^{-1}, \quad (17)$$

where τ_{fast} is the fast timescale correlation time.

Using the spectral density function, the ROE and NOE rate curves can be expressed (Abragam, 1961; Cavanagh, 1996)

$$\sigma_{\text{ROE}} = \frac{\hbar^2\mu_o^2\gamma^4 b}{40\pi^2 r_{\text{eff}}^6}[3J(\omega) + 2J(0)], \quad (18)$$

$$\sigma_{\text{NOE}} = \frac{\hbar^2\mu_o^2\gamma^4 b}{40\pi^2 r_{\text{eff}}^6}[6J(2\omega) - J(0)], \quad (19)$$

where μ_0 is the permittivity of free space, γ is the magnetogyric ratio of the hydrogen nucleus, b is the average occupancy of the hydration site, and r_{eff} is the effective distance between the water and protein protons. Figure 1 shows the ROE and NOE rates that result from the equations above.

The ROE and NOE rates between an amide proton and neighboring water molecules can be analyzed without knowledge of r_{eff} and b in Equations 18 and 19 by taking the ratio of the rates. The resulting expression depends on the three correlation time constants τ_{fast}, τ_{res}, and τ_c, and the order parameter a^2. For τ_c a value of 12 ns is used, obtained from an analysis of ^{15}N T_1, $T_{1\rho}$ and ^{1}H-^{15}N heteronuclear NOE data of the core protein amides of the vnd/NK-2 DNA complex at 309 K (La Penna et al., 2000). For τ_{fast} a value of 20 ps is used and for a^2 a value of 0.3 is used. Appendix A gives details on the determination of the spectral density parameters τ_{fast} and a^2.

Results

Water cross-relaxation rates for the amide protons of the DNA-bound NK-2 homeodomain were obtained at 309 K from one ROESY spectrum and one NOESY spectrum. Rates from an additional NOESY spectrum at 303 K were also obtained in order to determine the dependence of the NOE rates on temperature. Figure 2 shows the water resonance planes from the ROESY sum spectrum and from the NOESY sum spectrum at 309 K. For the ROESY the two interleaved ROE mixing times were 4 ms and 6 ms, which includes the trim pulses at the beginning and end of the mixing times, and for both NOESY spectra the interleaved mixing times were 10 ms and 15 ms. The rates were calculated using Equations 6 and 13. Figure 3 shows the results for the backbone amide ROE and NOE rates. Full tables giving the backbone and side chain rates and rate ratios as well as the percent change in the NOESY spectra rates as a function of temperature can be found at the web site http://mariana.nhlbi.nih.gov/~lbcweb. The water cross-relaxation rates in the ROESY and NOESY spectra arise from the ROE and NOE dipole-dipole cross-relaxation rates, respectively, plus the chemical exchange rate with water

$$\sigma_R = \sigma_{\text{ROE}} + k_x, \tag{20}$$

$$\sigma_N = \sigma_{\text{NOE}} + k_x, \tag{21}$$

where σ_R and σ_N are the measured rates from the ROESY and NOESY spectra, σ_{ROE} and σ_{NOE} are the ROE and NOE dipole-dipole cross-relaxation rates, and k_x is the chemical exchange rate.

Errors were calculated for each rate using the rms amplitude of the background noise for each measured spectrum, and using an average rms deviation of diagonal peak indirect dimension linewidths measured in two NOESY spectra collected under similar conditions. The errors that affect the water cross peak amplitude, the diagonal peak amplitude, and the diagonal peak linewidth are assumed to be uncorrelated. For all rates in Table 1, the random noise in the water cross peak amplitude contributed the largest uncertainty, more than 80% of the calculated error. The water cross peak indirect dimension linewidth used in calculating the rates was taken as the average of all the measured water cross peak indirect linewidths, since, in theory, for water molecules with residence lifetimes much shorter than the NMR timescale, the water linewidth should be the same as bulk. Consequently, the uncertainty in the water cross peak linewidth should be comparatively small, and, hence, was not included in the error calculations. The errors calculated for the rates were propagated to give errors in the rate ratios and the percent change in NOE rate per degree, again assuming uncorrelated errors. Another source of error can arise from the water magnetization not fully returning to its equilibrium value by the beginning of the mixing time. The contribution of this type of error to the rates has been analyzed in detail by Grzesiek and Bax (1993).

The largest rates in Figure 3 are due to chemical exchange, primarily for amides in the unstructured N and C terminal segments of the homeodomain. The rates dominated by chemical exchange can be easily identified by their negative σ_R rate values. Large chemical exchange rates are also observed for the exposed amides at the N-termini of the three alpha helices, especially for Lys10, and also for Ala11, Ala28, and Thr43. Amides whose cross-relaxation is dominated by chemical exchange have positive NOE/ROE rate ratios greater than 1.0 (see Equations 20 and 21). The σ_N rates for all amides with large chemical exchange increase in magnitude as the temperature is increased, most by at least 7% per degree. In contrast, for amides whose relaxation is dominated by direct dipolar interactions with water, the σ_N rates decrease as temperature is increased because the diffusion and overall tumbling rates increase (see Figure 1). Thus, the temperature dependence of the σ_N rates enables one to distinguish between amide protons whose rates

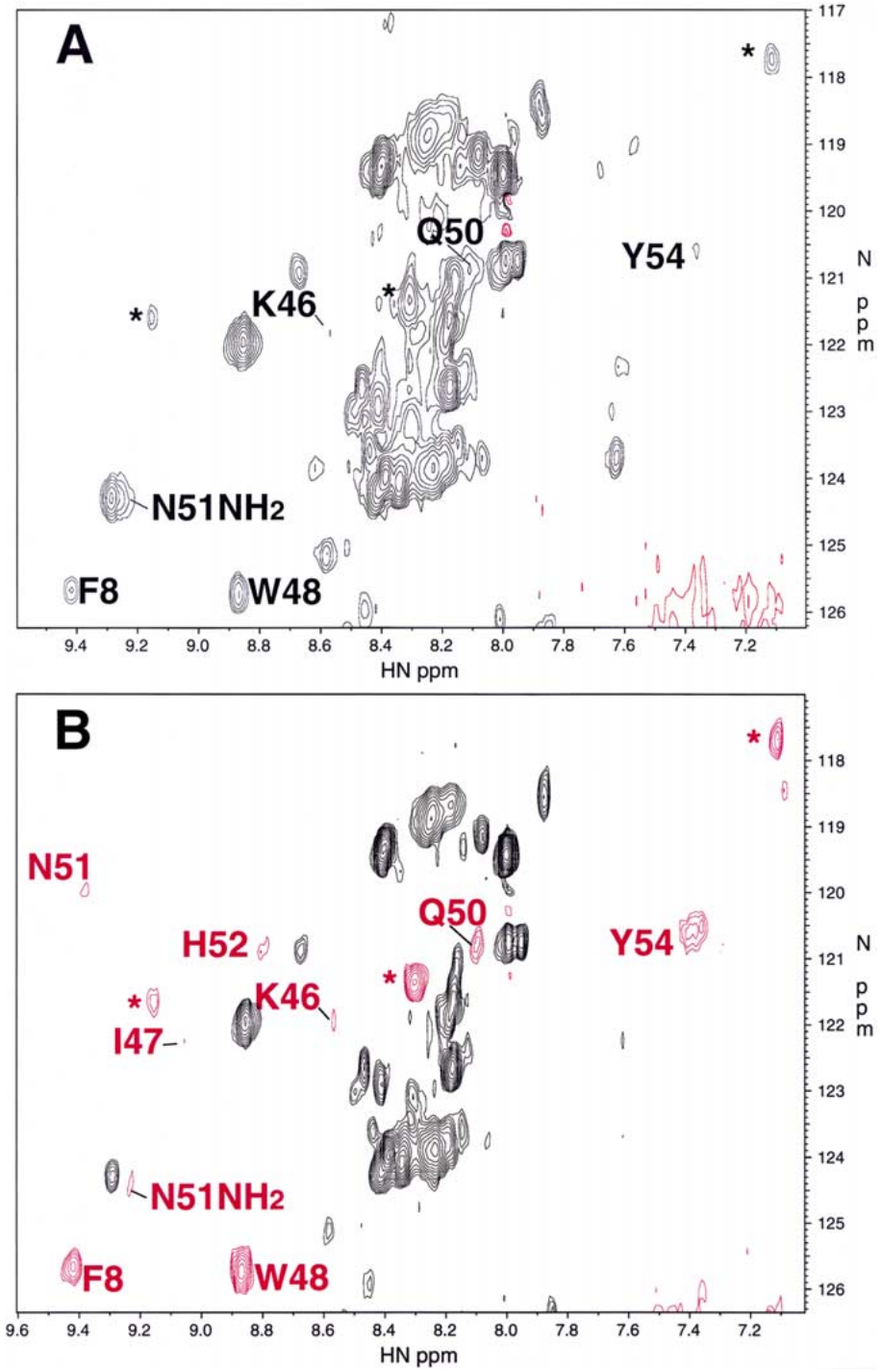

Figure 2. Water resonance planes from 3D NOESY (A) and ROESY (B) spectra at 309 K. The NOESY plane is from the sum spectrum of the 10 ms and 15 ms mixing times, and the ROESY plane is from the sum spectrum of the 4 ms and 6 ms mixing times. Each spectrum was collected in six days. Signals opposite the sign of the diagonal are shown in red. Backbone amide protons with ROE cross peaks due to direct water interactions have been labeled. The direct water interaction ROE cross peaks of Thr41 amide and the side chains of Trp48 and Arg53, whose rates are listed in Table 1, lie well outside the spectral range shown. Cross peaks labeled with asterisks have significant contributions from H^{α} resonances degenerate with the water resonance. The black cross peaks in the ROESY spectrum and the corresponding cross peaks in the NOESY spectrum are dominated by chemical exchange. The cross peaks appearing in the NOESY spectrum and not in the ROESY spectrum have chemical exchange contributions of the same order of magnitude as direct water interactions.

Table 1. Backbone amide and side chain direct water ROE and NOE rates

		ROE rate 309 K (s^{-1})	NOE rate 309 K (s^{-1})	NOE rate 303 K (s^{-1})	Surface area (Å^2)	Closest fast exchanging proton		Average distance (Å)
Phe8	HN	2.2 ±(0.2)	−0.9 ±(0.1)	−1.4 ±(0.1)	6.3	Thr13	$H^{\gamma 1}$	4.5
Gln12	HN	0.4 ±(0.2)	−0.1 ±(0.1)	−0.1 ±(0.1)	0.3	Ala11	HN	2.8
Glu15	HN	0.4 ±(0.1)	0.0 ±(0.1)	−0.1 ±(0.1)	0.3			
Thr41	HN	0.7 ±(0.2)	−0.2 ±(0.1)	−0.2 ±(0.1)	2.8			
Lys46	HN	0.7 ±(0.1)	−0.3 ±(0.1)	−0.2 ±(0.1)	0.1	Arg31	H^{ϵ}	4.5
Ile47	HN	1.0 ±(0.2)	−0.4 ±(0.1)	−0.3 ±(0.1)	0.9			
Trp48	HN	4.2 ±(0.2)	−2.0 ±(0.1)	−2.9 ±(0.1)	2.3			
Gln50	HN	1.6 ±(0.2)	−0.6 ±(0.1)	−1.1 ±(0.1)	0.4	Arg53	$NH_2^{\eta *}$	4.5
Asn51	HN	1.0 ±(0.2)	−0.4 ±(0.1)	−0.7 ±(0.1)	0.6			
His52	HN	1.1 ±(0.3)	−0.5 ±(0.1)	−0.5 ±(0.1)	0.9	His52	$H^{\delta 1}$	4.5
Tyr54	HN	2.7 ±(0.4)	−1.1 ±(0.2)	−1.5 ±(0.2)	4.9			
Tryp48	$H^{\varepsilon 1}$	1.2 ±(0.1)	−0.4 ±(0.1)	−0.6 ±(0.1)	7.5			
Asn51	NH_2(z)	2.3 ±(1.2)	−0.8 ±(0.5)	−2.2 ±(0.6)	0.2			
Arg53	Hε	6.0 ±(1.3)	−1.8 ±(0.7)	−3.5 ±(0.5)	3.5			
Arg53	$NH_2^{\eta 1,2}$	9.7 ±(4.8)	−3.3 ±(2.3)	−1.1 ±(3.0)	5.0			

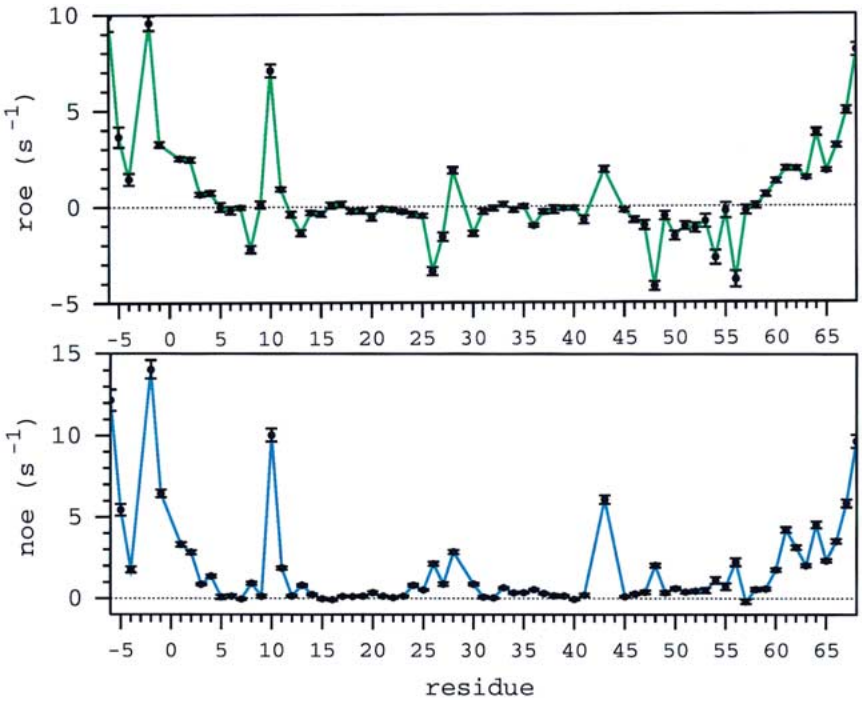

Figure 3. Water cross-relaxation rates for backbone amides measured in the ROESY and NOESY spectra.

are dominated by chemical exchange and those whose rates are dominated by direct dipolar cross-relaxation.

For amides with positive values of σ_R, greater than half of their cross-relaxation rates must arise from dipole-dipole cross-relaxation. There are three sources of dipole-dipole cross-relaxation that can give positive values of σ_R. The first source is direct cross-relaxation with water; the second is direct cross-relaxation with another protein resonance degenerate with the water resonance; and the third is indirect, or relayed, cross-relaxation through a neighboring proton with a large chemical exchange rate, that is, with a large negative value of σ_R (Otting et al., 1991; Wang et al., 1996). The two other sources of cross-relaxation for σ_R, chemical exchange and relay through a proton with a large positive value of σ_R, both make negative contributions to σ_R.

Direct cross-relaxation with water can be distinguished from the other two sources of dipole-dipole cross-relaxation by the value of rate ratio σ_N/σ_R. For direct cross-relaxation with water, the rate ratio σ_N/σ_R lies between -0.5 and 1.0 (see Figure 1). For the other two sources, the cross-relaxation involves a proton attached to the protein, and the rate ratio ideally would lie very near the slow diffusion limit -0.5 of the rate ratio curve. For cross-relaxation with a protein resonance degenerate with water, the ratio falls below -0.5 due to the non-negligible relaxation of the water-degenerate protein resonance during the mixing time. For the case of relayed cross-relaxation through a very fast exchanging proton, one additional step of ROE relay can add a non-negligible negative contribution to σ_R. Similarly, NOE relay would add a negative, though smaller, contribution to σ_N; the contribution should be considerably smaller since σ_{NOE} rates are less than half the magnitude of σ_{ROE} rates. As a result of the additional relay, the rate ratio can also fall below -0.5.

For the backbone amides with large positive values of σ_R, those most likely to be dominated by direct water cross-relaxation are Phe8, Thr41, Lys46, Ile47, Trp48, Gln50, Asn51, His52 and Tyr54, since all these amides have ratios σ_N/σ_R greater than -0.5. Table 1 lists the rates for these amides. The backbone amide of Trp48 has the largest values of k_R and k_N, 4.2 s^{-1} and -2.0 s^{-1}, respectively. A large, negative water NOE cross peak with this amide has also been observed in the Antennapedia homeodomain/DNA complex (Qian et al., 1993). In addition, the backbone amides of Gln12 and Glu15 also have positive values of σ_R and σ_N/σ_R ratios greater than -0.5, though their rates are much smaller, and hence, their ratios are less accurate. The NOE rates of all but three of the amides listed in Table 1 decrease as temperature is increased, consistent with direct dipole-dipole cross-relaxation with water. The three amides whose NOE rates increase with temperature, Gln12, Lys46, and Ile47, might have non-negligible contributions due to relayed magnetization through neighboring fast-exchanging protons. For instance, relay through the fast-exchanging amide of Ala11 could contribute to Gln12, and relay through the side chain of Arg31 could contribute to Lys46 (see Table 1). The change of the NOE rate with temperature is smaller than the calculated error in the NOE rates, so that the positive sign of the rate of change for these amides, as well as the negative sign for Thr41 HN and His52 HN, cannot be determined unambiguously.

All the backbone amides listed in Table 1 are known to exchange slowly with solvent, determined by previous deuterium exchange experiments (Tsao et al., 1994). For these slowly exchanging amides chemical exchange makes a negligible contribution to the measured rates ($k_x < 0.01$ s^{-1}); thus, the rates measured correspond to σ_{ROE} and σ_{NOE}. Table 1 also lists rate results for three side chains with large positive σ_R values and rate ratios greater than -0.5. The side chain amide of Trp48 is also known to exchange slowly with solvent, though the resonances of the side chain ^{15}N-attached protons of Asn51 and Arg53 were too weak to measure exchange rates in the deuterium exchange experiment (Tsao et al., 1994). However, since the side chain rate ratios for Asn51 and Arg53 are above -0.5, and since the changes in their σ_N rates are negative with rising temperature, it is unlikely that their rates have significant chemical exchange contributions.

Four residues with large positive values of σ_R, Phe20, Tyr25, Leu26, and Ser27, have alpha proton resonances degenerate or nearly degenerate with water. These four amides have σ_N/κ_R ratios below -0.5, consistent with σ_R being dominated by contributions from cross-relaxation with their degenerate $^{\alpha}H$ resonances. Three backbone amides, Thr13, Ser36, and Thr56, have large positive values of σ_R due to relay through their extremely fast exchanging (>500 s^{-1}) hydroxyl protons. These three amides also have σ_N/σ_R ratios slightly below -0.5.

Both Thr41 and Thr9 are very interesting in the vnd/NK-2 homeodomain because they are both involved in helix N-capping. Among other things, this means that their hydroxyl groups are hydrogen-bonded to glutamine backbone amides three residues

down in the sequence (residues Gln44 and Gln12), which slows the exchange of their hydroxyl protons enough so that their resonances are distinguishable from that of bulk water. The rates for Thr9 HN are too small to be measured; so there does not appear to be sufficient relay through its hydroxyl proton to give a large positive value of σ_R. In addition, the amide proton of Thr41 with a σ_N/σ_R ratio of -0.29 ± 0.17, significantly above the -0.5 limit, indicates that direct dipolar interaction with water, rather than relayed magnetization through its hydroxyl proton, dominates the cross-relaxation rate.

Using the 20 NMR structures obtained previously for the NK-2/DNA complex (PDB accession code 1NK2), the average exposed surface area for each amide proton was calculated. Also, for slowly exchanging amides within 5 Å of protons with large water cross-relaxation rates ($\sigma > 2$ s^{-1}), the proton is identified in Table 1 and the average proton-proton distance in the 20 NMR structures is given. The exposed surface area correlates only modestly with the rates overall (correlation coefficient of 0.52 for σ_N at 309 K), though the correlation is somewhat stronger for amides dominated by direct water cross-relaxation, listed in Table 1 (correlation coefficient of 0.70 for σ_N at 309 K). The majority of completely buried backbone amides have negligible σ_R and σ_N rates. In addition to the obvious cases of Thr13, Ser36, and Thr56, a few additional buried amides appear to have weak signals that could potentially be due to relay, based on their rate ratios and proximity to protons with large water cross-relaxation rates. One such case, Phe49 HN, which is sandwiched between Trp48 HN and Gln50 HN, both with large rates, is analyzed quantitatively in Appendix B.

For ^{15}N-attached protons with σ_N/σ_R ratios at 309 K between the limits of -0.50 and 1.0, listed in Table 1, corresponding diffusion lifetimes are listed in Table 2, estimated using the theoretically derived curve for σ_N/σ_R (Figure 1). The estimated lifetimes range from 0.4 ns to greater than 5 ns, with the majority in the 1–3 ns range. For Trp48 HN, the σ_N/σ_R ratio lies so close to the -0.5 limit that only the lower bound of the diffusion lifetime can be estimated. In fact, for most of the ^{15}N-attached protons, the accuracy of the ratio was inadequate to determine an upper bound to the diffusion lifetime, though lower bounds could be calculated for all the ^{15}N-attached protons. All but three of the ^{15}N-attached protons are found at the protein/DNA interface, as indicated in Table 2. The measured water cross-relaxation rates can arise from the interaction with one or more hydration site water molecules. Since the hydration site water molecules diffuse in times shorter than the NMR timescale, NMR cannot distinguish the signals of individual water molecules from that of bulk. For those cases where amides interact with more than one hydration site water molecule, the estimated diffusion lifetimes represent weighted averages of their respective lifetimes. Because the ROE and NOE cross-relaxation rates are higher for more slowly diffusing water molecules, the average will be weighted in favor of the more slowly diffusing water.

Discussion

In order to interpret the water diffusion lifetimes, the positions of the conserved water molecules observed in X-ray structures of homeodomain/DNA complexes were examined (Wilson et al., 1995; Jacobson et al., 1997; Tucker-Kellogg et al., 1997; Fraenkel et al., 1998; Li et al., 1998; Tan and Richmond, 1998; Passner et al., 1999; Piper et al., 1999). Water molecules lying within 5 Å of homeodomain amides (N to O distance) were tabulated. Those occurring in more than half the complexes studied are listed in Table 3. The percent occurrence of each conserved water molecule and the corresponding average temperature factor are also given. The σ_R and σ_N rates are listed for comparison. In nearly all cases, backbone and side chain amides with strong direct dipole-dipole interactions (as indicated by their large positive σ_R rates and lack of neighboring fast-exchanging protons) are located near conserved water molecules. The only amides with larger than average rates and no corresponding close conserved water molecules are the backbone amides of Thr41 and Lys46. In examining the individual X-ray structures, it was not at all obvious which waters were conserved until the full comparison was tabulated.

While there appears to be some measure of inverse correlation between the tabulated average distances and the σ_R and σ_N rates, the correlation is fairly modest (a simple linear correlation calculation gives correlations with shortest distance of -0.35 for σ_R and -0.51 for σ_N). For example, the σ_R and σ_N rates for the backbone amide of residue 5 are very much smaller than other amides with similar water distances. Of course, the average N to O distances are not exactly proportional to the actual proton-proton distances upon which σ_R and σ_N depend. Deviations from a simple distance-rate correlation are also ex-

Table 2. Water diffusion lifetimes

Residue		Ratio (error) [sign ΔNOE/°C]	Diffusion lifetime in ns (lower, upper bounds)[c]
Phe8	HN[a]	−0.42 (±0.06) [−]	3.1 (1.8, 23.)
Gln12	HN	−0.36 (±0.25) [+][b]	1.7 (0.7, n/d)
Glu15	HN	−0.09 (±0.21) [−][b]	0.4 (0.2, 0.6)
Thr41	HN	−0.29 (±0.17) [−][b]	1.1 (0.6, 6.5)
Lys46	HN[a]	−0.38 (±0.12) [+][b]	2.1 (1.0, n/d)
Ile47	HN[a]	−0.39 (±0.15) [+][b]	2.2 (0.9, n/d)
Trp48	HN[a]	−0.49 (±0.04) [−]	n/d (5.4, n/d.)
Gln50	HN[a]	−0.39 (±0.08) [−]	2.2 (1.3, 18.)
Asn51	HN[a]	−0.39 (±0.13) [−]	2.2 (1.0, n/d)
His52	HN[a]	−0.40 (±0.15) [−][b]	2.4 (1.0, n/d)
Tyr54	HN[a]	−0.42 (±0.09) [−]	3.0 (1.3, n/d)
Trp48	$HN^{\varepsilon 1}$ [a]	−0.35 (±0.06) [−]	1.6 (1.2, 2.8)
Asn51	NH_2 (z)[a]	−0.38 (±0.37) [−]	2.0 (0.5, n/d)
Arg53	$NH_2{}^{\varepsilon}$ [a]	−0.30 (±0.14) [−]	1.2 (0.7, 4.6)
Arg53	$NH_2{}^{\eta 1,2}$ [a]	−0.34 (±0.33) [−]	1.5 (0.5, n/d)

[a]These protons lie at the homeodomain/DNA interface.
[b]The sign of the NOE rate dependence on temperature is ambiguous due to the uncertainty in the measurement.
[c]For ratios with error ranges extending below the theoretical limiting rate ratio for direct dipolar cross-relaxation of −0.5, diffusion lifetime upper bounds could not be determined (n/d).

pected because the magnitudes of σ_R and σ_N also depend on the water diffusion lifetime, as well as on other parameters in the spectral density function. The σ_N for Arg5$NH_2^{\eta 1,2}$ and the average σ_N for the side chain amide of Gln12 are the only positive σ_N rates in the table, though both rates have large errors and Arg5 has spectral overlap problems, making the signs of these rates ambiguous.

The average temperature factors of the conserved waters show little inverse correlation with the water diffusion lifetimes. While the water molecules interacting with Trp48 HN have the longest diffusion lifetime (see Table 2) and also the lowest average temperature factors, the overall correlation between lifetime and temperature factor is only −0.26. For instance, Phe8 HN also interacts with slowly diffusing water, but the corresponding conserved water molecule has the highest temperature factor in the table. Since Phe8 is in the more flexible N-terminal arm region of the homeodomain (La Penna et al., 2000), the higher temperature factor might be due to the greater local flexibility of the region rather than larger motions of the water molecule itself. Averaged over all complexes, the average temperature factor of the conserved waters is 37, while the average temperature factor for all water molecules is 45. Given the large variability of temperature factors in different structures and the dependence of temperature factors on the method used to solve the X-ray structures, the difference between 37 and 45 might not be significant.

In order to visualize the conserved water molecules (Table 3) detected by NMR to diffuse slowly (Table 2), a vnd/NK-2 DNA homology model was created based on the X-ray structure observations. The water molecules included are those close to Phe8, Ile47, Trp48, Gln50, Asn51, His52, and Tyr54 backbone amide protons and the ^{15}N-attached side chain protons of Trp48, Asn51 and Arg53. The water molecule positions were obtained by alignment of the vnd/NK-2 DNA complex with that of engrailed (Fraenkel et al., 1998), whose interfacial water molecules correspond closely with those detected by NMR. An important function of the homology model is to test the compatibility of the conserved water positions in the context of the vnd/NK-2 DNA interface since several important residues at the interface, residues 47, 50, and 54, are variable in homeodomains and can have different conformations in different complexes. The conserved water positions were well accommodated in the vnd/NK-2 DNA interface, with only minor steric overlap. Retaining the relevant water molecules from the engrailed complex, hydrogen atoms were added to the water oxygens and

Table 3. Average water-amide distances in X-ray structures

Residue	σ_R[a] (s^{-1})	σ_N[a] (s^{-1})	X-ray av. water amide O-N distances Model water #: distance (% conservation) av. temp. factor[b]		
5	0.0 ±(0.2)	−0.1 ±(0.1)	3.7 Å (53%) 49		
8	2.2 ±(0.2)	−0.9 ±(0.1)	**1**: 3.6 Å (63%) 51		
25[a]	degen.	−0.1 ±(0.1)	4.5 Å (79%) 34		
26[a]	degen.	−0.9 ±(0.1)	3.1 Å (79%) 34	**6**: 4.5 Å (68%) 41	
27[a]	degen.	−0.9 ±(0.1)	3.5 Å (68%) 41		
45	0.2 ±(0.1)	−0.1 ±(0.1)	**2**: 4.9 Å (53%) 34		
47	1.0 ±(0.2)	−0.4 ±(0.1)	**2**: 4.5 Å (79%) 35		
48	4.2 ±(0.2)	−2.0 ±(0.1)	**2**: 3.5 Å (95%) 29	**3**: 3.8 Å(95%) 31	
49	0.5 ±(0.2)	−0.4 ±(0.1)	**6**: 4.7 Å (58%) 33		
50	1.6 ±(0.2)	−0.6 ±(0.1)	**6**: 4.1 Å (74%) 36		
51	1.0 ±(0.2)	−0.4 ±(0.1)	**7**: 4.3 Å (79%) 37	**8**: 4.6 Å(74%) 34	**4**: 4.7 Å(63%) 32
52	1.1 ±(0.3)	−0.5 ±(0.1)	**4**: 3.4 Å (74%) 36		
54	2.7 ±(0.4)	−1.1 ±(0.2)	**9**: 4.1 Å (47%) 37	**10**: 4.2 Å(47%) 37	
Arg5 $NH_2^{\eta 1,2}$	1.1 ±(0.8)	1.3 ±(0.6)	3.4 Å (53%) 45		
Gln12 NH_2	0.4 ±(0.6)	0.3 ±(0.3)	3.3 Å (55%) 51		
Gln44 NH_2	−0.2 ±(0.7)	−0.4 ±(0.3)	4.4 Å (57%) 41		
Trp48 $NH^{\varepsilon 1}$	1.2 ±(0.1)	−0.4 ±(0.1)	**5**: 3.4 Å (74%) 42	**4**: 3.6 Å(68%) 37	
Gln50 NH_2	1.8 ±(0.8)	−1.8 ±(0.4)	**7**: 3.1 Å (82%) 44		
Asn51 NH_2	3.1 ±(1.0)	−1.6 ±(0.5)	**3**: 3.1 Å (89%) 31	**7**: 4.0 Å(79%) 37	**4**: 4.5 Å(74%) 36
Arg53 NH^{ε}	6.0 ±(1.3)	−1.8 ±(0.7)	**6**: 4.4 Å (74%) 34		
Arg53 $NH_2^{\eta 1,2}$	9.7 ±(4.8)	−3.3 ±(2.3)	**6**: 3.0 Å (89%) 34	**9**: 3.3 Å(74%) 33	4.1 Å(58%) 32

[a] σ_N rates at 309 K, except for residues 25, 26 and 27 for which the σ_N rate at 303 K is given since the alpha resonances of these residues are degenerate with that of water at 309 K. The average of the rates of the two Gln and Asn side chain amide protons is given.

[b] Where appropriate, the homology model water corresponding most closely with the tabulated average distance is indicated in bold, followed by a colon. The number in parentheses gives the percentage of structures for which a water oxygen lies within 5 Å of the corresponding nitrogen atom, and the average distance is calculated using only the structures that satisfy this criterion. The last number is the average temperature factor for the conserved water molecule. A total of 19 crystallographically distinct homeodomain/DNA complexes from 9 structure files were analyzed (PDB accession codes 1B81, 1B72, 1MNM, 1AU7, 1FJL, 2HDD, 3HDD, 1AKH and 9ANT).

restrained minimization (InsightII/Discover, MSI, San Diego, CA) was performed on the vnd/NK-2 DNA water system to relax the minor steric overlap and to suggest plausible hydrogen atom orientations. Views of the water molecules in the homology model are shown in Figures 4 and 5. The homology model water molecules corresponding to the conserved water distances in Table 3 are indicated accordingly. Careful examination of Table 3 reveals that a minimum of seven out of the ten conserved water molecules is required for agreement with the NMR results in Table 2. In particular, water molecules 5, 8, and either 9 or 10 could be removed, while still retaining at least one conserved water molecule near each ^{15}N-attached proton observed to interact with slowly diffusing water. Of course, the NMR results are also consistent with the presence of all ten water molecules, since NMR cannot distinguish between cases of single and multiple water molecule interactions.

Two chains of four water molecules each form in the major groove where the recognition helix of the homeodomain crosses over the two phosphate backbones of the DNA duplex. The tendency of this set of water molecules to form hydrogen-bonded chains might be exaggerated due to the fact that no additional water molecules, which could compete for available water-water hydrogen bonds, were included during the minimization of the model. At the far ends of the chains, water molecules 2 and 10 can potentially form hydrogen bonds with the adjacent phosphate oxygen atoms. Water molecules 2 and 3 form a hydrogen-bonded pair bridging the side chain of Asn51 with a phosphate oxygen atom. Water molecules 2 and 3 are especially important since they appear to be invariant

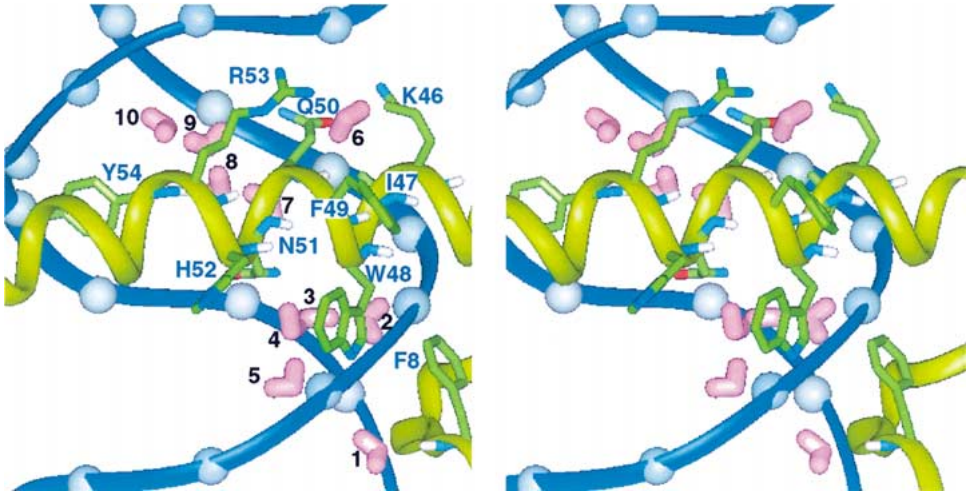

Figure 4. Stereoview of conserved water molecules at the vnd/NK-2 homeodomain DNA interface as determined from X-ray homeodomain/DNA complex structures. The conserved water molecules, labeled 1 through 10, correspond closely to the expected positions of slowly diffusing water molecules detected by NMR at the vnd/NK-2 DNA interface. The view shows the third helix of the homeodomain in the major groove of the DNA and the N-terminal arm in the minor groove.

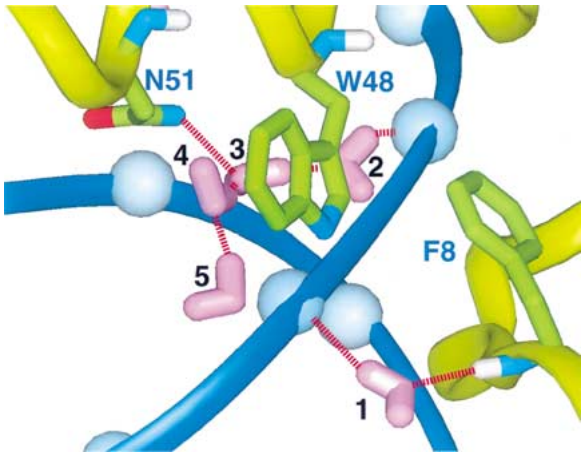

Figure 5. View of conserved water molecules focusing on the right side of the major groove and near the N-terminal arm, including water molecule 1 and the invariant water molecules 2 and 3 which can form hydrogen bond bridges between residues Phe8 and Asn51 and DNA phosphates. The hydrogen bonds suggested by the model structure are indicated by red dashed lines

in homeodomain/DNA complexes (see Table 3; in the one complex where this water pair was not observed, MAT α2 in the 1MNM structure (Tan and Richmond, 1998), the homeodomain is bound atypically at the end of the DNA duplex). Water molecule 1 can form a hydrogen bond bridge between the amide of Phe8 and a phosphate oxygen atom, and is almost close enough to form a hydrogen bond with the hydroxyl group of Thr13. This conformation of water 1 might explain the moderately slow deuterium exchange rate measured for the amide of Phe8 (Tsao et al., 1994), which is solvent exposed (Gruschus et al., 1999). Water 6 can possibly hydrogen bond to the guanidinium group of Arg53 and is also observed in many homeodomain/DNA X-ray structures to hydrogen bond to a phosphate oxygen atom, though the geometry is not ideal for these interactions in the homology model structure. Water molecule 7 lies between the side chains of Asn51 and Gln50. In the homology model the distances from water molecule 7 to both of these side chains, approximately 3.7 Å and 4.7 Å to Asn51 and Gln50 respectively, are too long for hydrogen bond formation; however, a hydrogen bond to one or the other of these side chains is observed in several X-ray homeodomain/DNA structures (Jacobson et al., 1997; Fraenkel et al., 1998).

Solvent accessible surfaces were calculated for the ten water molecules, and water molecules 2 and 9 have no accessible surface area to water molecules others than those included in the model. For the chain near Trp48 and Asn51, water molecule 2 is at the end of the chain, while for the chain near Gln50 and Tyr54, both ends of the chains are solvent accessible. In other words, the chain near Trp48 and Asn51 lies within a 'cave' at the protein/DNA interface with water molecule 2 at the back of the cave, while the chain near Gln50 and Tyr54 lies within a 'tunnel' with water molecule 9 in the middle of the tunnel. Water molecule 2 is closest to the backbone amide

of Trp48, which has the longest average diffusion time for its interacting water molecules. This behavior concurs with the observation that water molecule 2 appears to be the most inaccessible of the ten water molecules. This also agrees with the study of the Antennapedia homeodomain/DNA complex, where of all backbone amide/water cross peaks, only that of Trp48 was unambiguously determined to be due to interaction with slowly diffusing water (Qian et al., 1993). A molecular dynamics simulation of the Antennapedia complex showed similar results (Billeter et al., 1996). Water molecule 9 lies nearest Tyr54 amide, which has the second longest average diffusion time for its interacting water molecules.

In the model structure only a few of the conserved water molecules clearly have the potential to form hydrogen-bond bridges between the homeodomain and the DNA, those between Asn51 and DNA and between Phe8 and DNA. In several other potential water-mediated hydrogen-bond bridge locations, for example between DNA and the side chains of Arg31 and Lys46, both in the major groove, no slowly diffusing water or conserved water molecules from the X-ray structures were apparent. Given that water must, nevertheless, occupy these locations, any hydrogen-bond bridges that might be formed in these locations likely involve more rapidly diffusing water in more variable conformations. Thus, there does not appear to be a strong correlation between water forming hydrogen bonds in the interface and slowly diffusing, conserved interfacial water molecules.

Removal of the hydrogen bond bridge interactions of slowly diffusing water molecules could conceivably impact binding specificity as well as affinity. Consider the scenario where these waters are excluded by methylation of one of the chemical groups with which they interact, replacing the NH_2(e) proton of Asn51 or the backbone amide proton of Phe8, for instance. Placing a hydrophobic group at either of these sites would almost certainly reduce the free energy of protein/DNA binding since a methyl group instead of a water molecule would then lie adjacent a charged phosphate. In all homeodomain/DNA structures, the side chain of Asn51 assumes a well-defined conformation making specific hydrogen bonds with an adenine base. Removing the water hydrogen bond bridge interactions of the side chain of Asn51 by methylation at the NH_2(e) proton site could make this conserved side chain conformation less stable and, as a consequence, reduce the specificity of the Asn51 side chain for adenine. Alternatively, were one able to hydroxylate these same sites, they could then possibly form direct protein/DNA hydrogen bonds, and the entropic penalty of recruiting a water molecule to form a hydrogen bond bridge would be avoided. One question that remains is to what extent these arguments might also apply to more transient hydrogen bond bridges formed by faster diffusing water.

Conclusions

A strong correlation is observed between slowly diffusing water measured by NMR for the vnd/NK-2 DNA interface and the conserved interfacial water molecules, seen from X-ray structures. A few of these water molecules appear to form structurally important hydrogen bond bridges between the protein and the DNA. These water hydrogen bond bridges may play a significant role in homeodomain/DNA binding affinity and specificity.

Measuring the water ROE and NOE cross-relaxation rate ratios is an effective way of quantitatively probing the diffusion lifetimes of slowly diffusing water molecules at the homeodomain/DNA interface. The method proves most useful for water molecules with diffusion lifetimes of approximately 1 ns, and a large number of the interfacial water molecules, including the longest-lived ones, were observed to have lifetimes in this range. For lifetimes much greater than 1 ns, or more precisely, as lifetimes approach and exceed the macromolecular rotational correlation time, the rate ratios become increasingly insensitive to changes in the diffusion lifetimes. For lifetimes much shorter than 1 ns, the cross-relaxation rates themselves become too small to detect accurately. As the sensitivity of NMR spectrometers continues to improve, the range of lifetimes that can be accurately determined by the ROE NOE rate ratio method will steadily expand.

Appendix A. Spectral density function parameters

The fast timescale motions at a particular occupied hydration site will cause fluctuations around a time average value of the Hamiltonian of the dipole-dipole water-amide interaction, $h_{av}(\tau_d \gg \tau \gg \tau_f)$, where τ_f is the fast timescale correlation time and τ_d is the diffusion timescale correlation time. The Hamiltonian correlation function for a particular site will exponentially decay from an initial value of $bh^2(0)$ to a value

of $b\left(h_{av}(\tau_d \gg \tau \gg \tau_f)\right)^2$ with a time constant of τ_f, where the value b is the percentage occupancy of the particular site. Let a^2 represent the ensemble average of the ratio of the initial and final values

$$a^2 = \overline{\left(h_{av}(\tau_d \gg \tau \gg \tau_f)\right)^2 / h^2(0)} \tag{22}$$

The order parameter a^2 represents the factor reduction of $\overline{h(t)h(t+\tau)}$ due to fast timescale motions.

For diffusion into and out of the hydration site on a timescale faster than the overall macromolecular tumbling time, the diffusional motion will cause further fluctuations in the Hamiltonian. The protons in solution, including those of the water, have practically equal expectation values of spin in all directions; the very small Boltzmann difference in the z-component expectation value is negligible for the purposes of determining the dipole-dipole Hamiltonian fluctuations. Hence, the water protons diffusing into the site have effectively randomly oriented nuclear spins, and the Hamiltonian will fluctuate about an average value of zero. The Hamiltonian correlation function will decay to zero with a correlation time τ_d, the diffusion correlation time. This randomly fluctuating model of diffusion is equivalent to a two-state occupied/unoccupied hydration site system. Models also exist to describe the case of the interaction of water diffusing continuously past a protein proton (Ayant et al., 1977; Brüschweiler et al., 1994).

To obtain an estimate for a^2, the fast timescale order parameter, several computations of the dipole-dipole interactions between a (protein) proton and the protons of one to three water molecules were performed. The water oxygen atoms were placed at an average distance of 3 to 4 Å from the proton and this separation was allowed to fluctuate up to 1.5 Å with random reorientation of the water molecule. Values of a^2 were calculated for one, two and three water molecules in various geometries, resulting in values of a^2 between 0.0 and 0.6 (see table at the web site http://mariana.nhlbi.nih.gov/~lbcweb). Because the rates are measured in a sample containing 20% D_2O, a significant percentage (32%) of hydration sites are occupied by water molecules with just one proton, so a simulation was also done with a water molecule with just one proton. From these results the average value of 0.3 was chosen for a^2 for the diffusion lifetime calculations. In the future better estimates for a^2 might be obtained from analysis of molecular dynamics trajectories of water molecules in macromolecular hydration sites.

Various values of τ_{fast} can be found in the literature. In earlier ROE and NOE rate analyses of water protein interactions, a value of 4 ps can be found, though the origin of this value is not clear (Otting et al., 1991; Otting, 1997). In a study modeling the dielectric function of water at 298 K, a value of 8.5 ps has been used as the correlation time of bulk water (Kim et al., 1992). Experimentally, water (at NTP) has its first major absorption peak around 35 GHz, corresponding to a time of approximately 30 ps (Jackson, 1962). Though the Stokes equation is intended for solute molecules much larger than those of the solvent, it is interesting to note that the rotational correlation time one obtains for a water molecule from the Stokes equation is 55 ps at 309 K using an effective water radius of 2 Å. We have used a value of τ_{fast} of 20 ps in this study. It is possible that τ_{fast} for bulk water differs from that for water in typical hydration sites, and in the future better estimates for τ_{fast} might be obtained from analysis of molecular dynamics trajectories of water molecules in macromolecular hydration sites.

Appendix B. Relay analysis

Because the ROESY and NOESY spectra were each recorded with two mixing times, it is possible to further assess the impact of relay on the measured rates. Rates can be independently calculated for each mixing time using Equation 23 below. Up to this point, the intensities of the water cross peaks from the individual mixing time spectra have not been used in the calculation of the water rates. However, these intensities can also be used to calculate rates, though with less robust peak height determination due to the lower signal to noise ratio. For the NOE rates

$$\sigma = \frac{y(\tau_{mix})\Delta_{wat}\rho}{d(0)\Delta_{diag}(1 - e^{-\rho\tau_{mix}})}, \tag{23}$$

where $y(\tau_{mix})$ is the water cross peak intensity measured in the τ_{mix} spectrum. The ROE rate expression includes off-resonance correction factors. Rates calculated from the individual spectra can be useful in assessing whether relayed water cross-relaxation makes a significant contribution to the observed water-to-amide cross peaks.

The best-determined backbone amide rates (roughly the top 40% of all the rates), as judged by small error, were compared at the two mixing times. For the NOESY spectra, the rates were on average

1.06 ± 0.07 greater for the 15 ms mixing time compared to the 10 ms mixing time. Larger rates for the longer mixing time are expected since direct and relayed NOE cross-relaxation have the same sign. For the ROESY spectra, the rates were on average 1.03 ± 0.11 greater at 6 ms compared to 4 ms. Direct ROE relaxation has the opposite sign of relayed ROE cross-relaxation and cross-relaxation by chemical exchange. More than half of the top 40% of rates are dominated by chemical exchange, so relayed ROE cross-relaxation should enhance these rates at the longer mixing time. For those ROE rates dominated by direct cross-relaxation, relay should diminish the rates at the longer mixing time. This explains why the average ratio of ROE rates at the two mixing times is closer to 1.0 than the average NOE ratio. In addition, ROE rates are at least two times greater in magnitude than NOE rates; thus, relayed ROE cross-relaxation can build up much faster. Because of transverse relaxation during the ROESY mixing time, the ROESY signal to noise ratio is smaller than that of the NOESY, so the ROE rates are not determined as accurately. All of these factors lead to greater variation in the ROE rate ratio.

The backbone amide of Phe49 provides an example of how additional cross peak information in the three-dimensional spectra can be exploited for additional relay analysis. The percentage contribution to the water rate for Phe49 from relay can be crudely estimated using the water and diagonal peak intensities of Trp48, Phe49, and Gln50 amides

$$\sigma_{1w}t \cong \frac{\Delta_w y_1}{\Delta_{d1} d_1}, \quad \sigma_{12}t \cong \frac{\Delta_{d1} x_{12}}{\Delta_{d2} d_2}, \quad \sigma_{2w}t \cong \frac{\Delta_w y_2}{\Delta_{d2} d_2}, \tag{24}$$

$$w_{12} \cong \frac{\frac{1}{2}\sigma_{1w}\sigma_{12}t^2}{\sigma_{2w}t} \cong \frac{1}{2}\frac{x_{12} y_1}{d_1 y_2}, \tag{25}$$

where σ_{1w} and σ_{2w} are the water cross-relaxation rates to the first and second protons, σ_{12} is the cross-relaxation rate from the first to the second proton, t is the mixing time, Δ_w is the water indirect dimension linewidth, Δ_{d1} and Δ_{d2} are the first and second proton indirect dimension linewidths, y_1 and y_2 are the first and second proton water cross peak heights, d_1 and d_2 are the first and second proton diagonal peak heights, x_{12} is the height of the cross peak from the first to the second proton, and w_{12} is the approximate percentage contribution from relayed water cross-relaxation through the first proton to the second proton. These expressions are only approximate because relaxation of the magnetization of the two protons is not taken into account, nor are the differences in chemical shifts for ROE cross-relaxation, though the correction factors do cancel for the most part for w_{12}. Using these expressions, relay through Trp48 HN reduces the σ_R of Phe49 HN by approximately 7% and enhances the σ_N of Phe49 HN by approximately 8%, and Gln50 HN reduces σ_R by approximately 1% and enhances σ_N by approximately 1% . Adjusting σ_R and σ_N accordingly, the σ_N/σ_R ratio for Phe49 goes from -0.68 to -0.57. Since this ratio is still below the theoretical limit of -0.50, it is likely that additional water magnetization is relayed to Phe49 HN through neighboring ^{13}C-attached protons with large water cross-relaxation rates. Indeed, preliminary ^{13}C-edited NOESY and ROESY experiments show a strong interaction between water and the alpha proton of Trp48, which in turn interacts with Phe49 HN. Interestingly, this analysis also implies that Phe49 HN appears to have a fairly significant amount of direct cross-relaxation with water, presumably with water 6 (see Table 3), even though this amide is completely buried.

References

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford.

Ayant, Y., Belorizky, E., Fries, P. and Rosset, J. (1977) *J. Phys. (Paris),* **38**, 325–335.

Billeter, M., Guntert, P., Luginbuhl, P. and Wüthrich, K. (1996) *Cell*, **85**, 1057–1065.

Brüschweiler, R. and Wright, P.E. (1994) *Chem. Phys. Lett.*, **229**, 75–81.

Cavanagh, J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego, CA.

Clore, G.M., Bax, A., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 5671–5676.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277–293.

Fraenkel, E., Rould, M.A., Chambers, K.A. and Pabo, C.O. (1998) *J. Mol. Biol.*, **284**, 351–357.

Gruschus, J.M. and Ferretti, J.A. (1999) *J. Magn. Reson.*, **140**, 451–459.

Gruschus, J.M., Tsao, D.H.H., Wang, L.-H., Nirenberg, M. and Ferretti, J.A. (1997) *Biochemistry*, **36**, 5372–5380.

Gruschus, J.M., Tsao, D.H.H., Wang, L.-H., Nirenberg, M. and Ferretti, J.A. (1999) *J. Mol. Biol.*, **289**, 529–545.

Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 627–638.

Halle, B. and Wennerström, J. (1981) *J. Chem. Phys.*, **75**, 1928–1943.

Halle, B., Denisov, V.P. and Venu, K. (1999) In *Modern Techniques in Protein NMR*, Vol. 17 (Berliner, L.J. and Krishna, N.R., Eds.), Plenum, New York, NY.

Jacobson, E.M., Li, P., Leon-del-Rio, A., Rosenfeld, M.G. and Aggarwal, A.K. (1997) *Genes Dev.*, **11**, 198–208.

Jackson, J.D. (1962) *Classical Electrodynamics*, John Wiley and Sons, New York, NY.

Karimi-Nejad, Y., Lohr, F., Schipper, D., Ruterjans, H. and Boelens, R. (1999) *Chem. Phys. Lett.*, **300**, 706–712.

Kiihne, S. and Bryant, R.G. (2000) *Biophys. J.*, **78**, 2163–2169.

Kim, S.-H., Vignale, G. and DeFacio, B. (1992) *Phys. Rev.*, **A46**, 7548–7560.

Kriwacki, R.B., Flanagan, J.M., Caradonna, J.P. and Prestegard, J.H. (1993) *J. Am. Chem. Soc.*, **115**, 8907–8911.

La Penna, G., Fausti, S., Perico, A. and Ferretti, J.A. (2000) *Biopolymers,* **54**, 89–103.

Li, T., Jin, Y., Vershon, A.K. and Wolberger, C. (1998) *Nucleic Acids Res.*, **26**, 5707–5713.

Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546–4559.

Melacini, G., Bonvin, A.M.J.J., Goodman, M., Boelens, R. and Kaptein, R. (2000) *J. Mol. Biol.*, **300**, 1041–1046.

Otting, G., Liepinsh, E. and Wüthrich, K. (1991) *Science*, **254**, 974–980.

Otting, G. (1997) *Prog. NMR Spectrosc.*, **31**, 259–285.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S. and Aggarwal, A.K. (1999) *Nature*, **397**, 714–718.

Phan, A.T., Leroy, J.L. and Gueron, M. (1999) *J. Mol. Biol.*, **286**, 505–519.

Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L. and Wolberger, C. (1999) *Cell*, **96**, 587–592.

Qian, Y.Q., Otting, G. and Wüthrich, K. (1993) *J. Am. Chem. Soc.*, **115**, 1189–1190.

Tan, S. and Richmond, T.J. (1998) *Nature*, **391**, 660–665.

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

Tsui, V., Radhakrishnan, I., Wright, P.E. and Case, D.A. (2000) *J. Mol. Biol.*, **302**, 1101–1117.

Tucker-Kellogg, L., Rould, M.A., Chambers, K.A., Ades, S.E., Sauer, R.T. and Pabo, C.O. (1997) *Structure*, **5**, 1047–1056.

Wang, Y.X., Freedberg, D.I., Grzesiek, S., Torchia, D.A., Wingfield, P.T., Kaufman, J.D., Stahl, S.J., Chang, C.H. and Hodge, C.N. (1996) *Biochemistry*, **35**, 12694–12704.

Wilson, D.S., Guenther, B., Desplan, C. and Kuriyan, J. (1995) *Cell*, **82**, 709–712.