Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA

G Marius Clore*, Ad Bax, James G Omichinski and Angela M Gronenborn*

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Background: The erythroid specific transcription factor GATA-1 is responsible for the regulation of transcription of erythroid-expressed genes and is an essential component required for the generation of the erythroid lineage. GATA-1 binds specifically as a monomer to the asymmetric consensus target sequence (T/A)GATA-(A/G) found in the *cis*-regulatory elements of all globin genes and most other erythroid specific genes that have been examined. We have previously determined the solution structure of the complex of the zinc-containing DNA-binding domain of chicken GATA-1 with its cognate DNA target site by multidimensional heteronuclear NMR. From previous studies of complexes between proteins and DNA, water appears to play an important role in DNA–protein recognition by mediating bridging hydrogen bonds between functional groups on the protein and DNA bases. Solvation free energy calculations, however, suggest that hydrophobic interactions should exclude water from parts of the GATA-1:DNA interface.

Results: Using water-selective two-dimensional heteronuclear magnetic resonance spectroscopy, we have identified the location of bound water molecules in the specific complex of chicken GATA-1 with DNA. A number of water molecules could be detected between the protein and the phosphate backbone, as well as at the solvent exposed surface of the protein. However, no water molecules could be observed at the interface of the protein with the bases of the DNA. With only one exception, the bound water molecules have a residency time >200–300 ps.

Conclusions: Unlike other protein–DNA complexes, the majority of specific interactions between GATA-1 and the DNA bases in the major groove are hydrophobic in nature. The exclusion of water from the protein–DNA base interface in the major groove supports the view that the specific binding energy is indeed dominated by hydrophobic effects.

Structure 15 February 1994, **2**:89–94

Key words: bound water, GATA-1–DNA complex, NMR solution structure, protein hydration

Introduction

A number of recent high resolution crystal structures of protein–DNA complexes have suggested that bound water molecules serve to bridge hydrogen bonds between functional groups on the protein and the DNA bases [1–3]. This may play an important part in the recognition process. In addition, a bound water molecule has been detected by NMR spectroscopy at the interface of the complex between the *Antp*(C39S) homeodomain and a 14 base-pair (bp) duplex in solution [4]. In a recent paper, we presented the solution structure of a specific complex between the zinc-containing DNA-binding domain of the erythroid transcription factor GATA-1 and a 16 bp duplex comprising its cognate recognition site [5]. The structure reveals interactions in both the major and minor grooves of the central 8 bp recognition motif. A helix and loop, which comprise part of the globular domain of the protein, contact the major groove, while the carboxy-terminal tail wraps around the DNA and is located in the minor groove. In contrast to other protein–DNA complexes, in which the majority of specific interactions involve hydrogen bonds between the protein and the DNA bases [6], the specific interactions between GATA-1 and the major groove of the DNA are dominated by hydrophobic contacts. We have suggested, on the basis of solvation free energy calculations, that hydrophobic effects constitute a large proportion of the specific binding energy [5]. We would therefore predict that water of hydration would be excluded from the interface between GATA-1 and the DNA bases in the major groove, but would be present at the interface between GATA-1 and the sugar-phosphate backbone, as well as at the solvent exposed surface of the protein. To test this hypothesis we have made use of nuclear Overhauser (NOE) and rotating frame Overhauser (ROE) enhancement measurements to detect through-space (<4 Å) contacts between bound water and protons of the protein.

Results and discussion

The complex we investigated comprises the zinc-containing DNA-binding domain of chicken GATA-1 (66 residues, numbers 158–223 of the intact protein) uni-

*Corresponding authors.

formly labeled (>95%) with ^{15}N and ^{13}C, and a 16 bp duplex [first strand 5′d(GTTGCAGATAAACATT); second strand 5′d(AATGTTTATCTGCAAC)] at natural isotopic abundance [5]. In the discussion that follows we identify DNA bases in this duplex by numbering from the left of the sequence as it is printed above. The first strand comprises G1–T16, the second strand A17–C32. NOEs and ROEs between bound water and protein protons attached to ^{13}C or ^{15}N were observed by recording ^{12}C-filtered two-dimensional (2D) H_2O–ROE/NOE–^1H–^{13}C or H_2O–ROE/NOE–^1H–^{15}N heteronuclear single quantum coherence (HSQC) spectra [7,8]. These two-dimensional (2D) experiments employ low power ^1H pulses (radiofrequency, r.f., field strength of 70–100 Hz) for selective inversion of the H_2O resonance, and are particularly sensitive as the slowly relaxing water is never saturated [7,8]. The spectra are obtained by recording two data sets in an interleaved manner such that the water magnetization lies either parallel (+z) or antiparallel (−z) to the magnetization of all other ^1H resonances (which are aligned along +z), prior to the NOE/ROE mixing period. The difference between the two data sets yields an NOE/ROE difference spectrum between water protons and protein, while the sum of the data sets yields the reference spectrum. Hence, in the difference spectrum ^1H–^{13}C or ^1H–^{15}N correlation peaks are only observed for protein resonances that interact with water. The sum reference spectrum, on the other hand, has the appearance of a regular ^1H–^{15}N or ^1H–^{13}C correlation spectrum. Thus the ^1H–^{15}N/^1H–^{13}C correlation peaks in the sum and difference spectra are equivalent to the diagonal peaks and cross-peaks, respectively, in conventional 2D ^1H–^1H NOE/ROE experiments.

The NOE/ROE interactions involving protein protons attached to ^{13}C which resonate in the vicinity of the water resonance, are effectively suppressed by means of a double ^{12}C-filter to select protons not attached to ^{13}C [8], and (in the case of the ROE experiments) by exploiting the fact that the spin-locked relaxation time of water is about two orders of magnitude longer than for protein protons [7]. DNA protons in the vicinity of the water resonance, however, are not suppressed in this manner as they are attached to ^{12}C. Any potential interactions between DNA and protein, however, are easily identified by recording a control spectrum in which the water resonance is presaturated with a weak r.f. field (~10 Hz) followed by a 200 ms delay prior to the first selective water pulse. Because the water relaxes slowly, the water magnetization remains attenuated by a factor of about 20 relative to its equilibrium value at the end of the 200 ms delay, while DNA protons attached to ^{12}C (as well as any protein protons attached to ^{12}C due to incomplete labeling) have returned to their equilibrium values due to very effective cross-relaxation with off-resonance protein and DNA protons [9]. As a result, all interactions with water arising either from NOE/ROEs or from chemical exchange are suppressed by a factor of about 20, whereas NOE/ROEs to protons attached to ^{12}C which resonate in the vicinity of the water resonance are essentially unaffected.

The methyl region of the H_2O–NOE ^1H–^{13}C HSQC reference spectrum is illustrated in Fig. 1a, together with the corresponding region of the H_2O–NOE and ROE ^1H–^{13}C HSQC difference spectra in Figs 1b and 1c. The H_2O–NOE ^1H–^{15}N HSQC difference spectrum is shown in Fig. 1d. All experiments were carried out at 25 °C as at higher temperatures the complex is unstable and at lower temperatures the spectra rapidly deteriorate due to increased linewidths associated with the reduction in the rotational correlation time. Control experiments indicated that all the interactions illustrated in the H_2O ROE/NOE HSQC difference spectra shown in Figs 1b, 1c and 1d arise from water. ROE/NOE cross peaks in these spectra can arise from two sources: directly by NOE/ROEs to bound water, and indirectly by NOEs to rapidly exchanging protons followed by chemical exchange with water [10,11]. Such rapidly exchanging protons comprise the hydroxyl protons of serine and threonine, the guanidino group of arginine, the $N^{\zeta}H_3$ group of lysine, some NH_2 groups of asparagine and glutamine, and some backbone NH groups. Hence, the presence of an NOE/ROE from a particular proton to bound water can only be unambiguously established if the distance from that proton to any potential rapidly exchangeable proton is greater than 4 Å in the structure of the complex [10,11]. In addition, peaks arising directly from rapid chemical exchange between exchangeable backbone amide protons and water are present in the H_2O NOE/ROE ^1H–^{15}N HSQC difference spectra. These are readily identified, as cross peaks due solely to chemical exchange with water are of opposite sign to ROE peaks in the ROE spectrum [10–12].

On the basis of the structure of the complex of GATA-1 with DNA [5], we were able to unambiguously identify 17 methyl protons and three backbone NH protons in close proximity to bound water. The location of the protein residues near bound water are indicated in the structure of the complex of GATA-1 with DNA shown in Fig. 2.

An approximate estimate of the residency times of these bound water molecules can be obtained from the sign of the NOE [13,14]. If a water molecule is bound to a macromolecule with a correlation time in the spin-diffusion limit (i.e. greater than about 2 ns), the NOE between water and a protein proton will be zero for a residency time of ~300 ps, while for residency times shorter and longer than 300 ps the NOE will be positive and negative, respectively. In contrast, the ROE remains positive for all residency and correlation times. The NH protons of Ala30, Tyr34 and Tyr35 and the methyl goups of Ala3, Ala30, Met46 and Ile51 (γ_m and δ_m) exhibited negative NOEs to water, indicating a residency time greater than about 500 ps. The methyl protons of Met23 exhibited a positive NOE to water, indicative of a residency time of 100–300 ps. For the remaining surface methyl protons, the NOEs

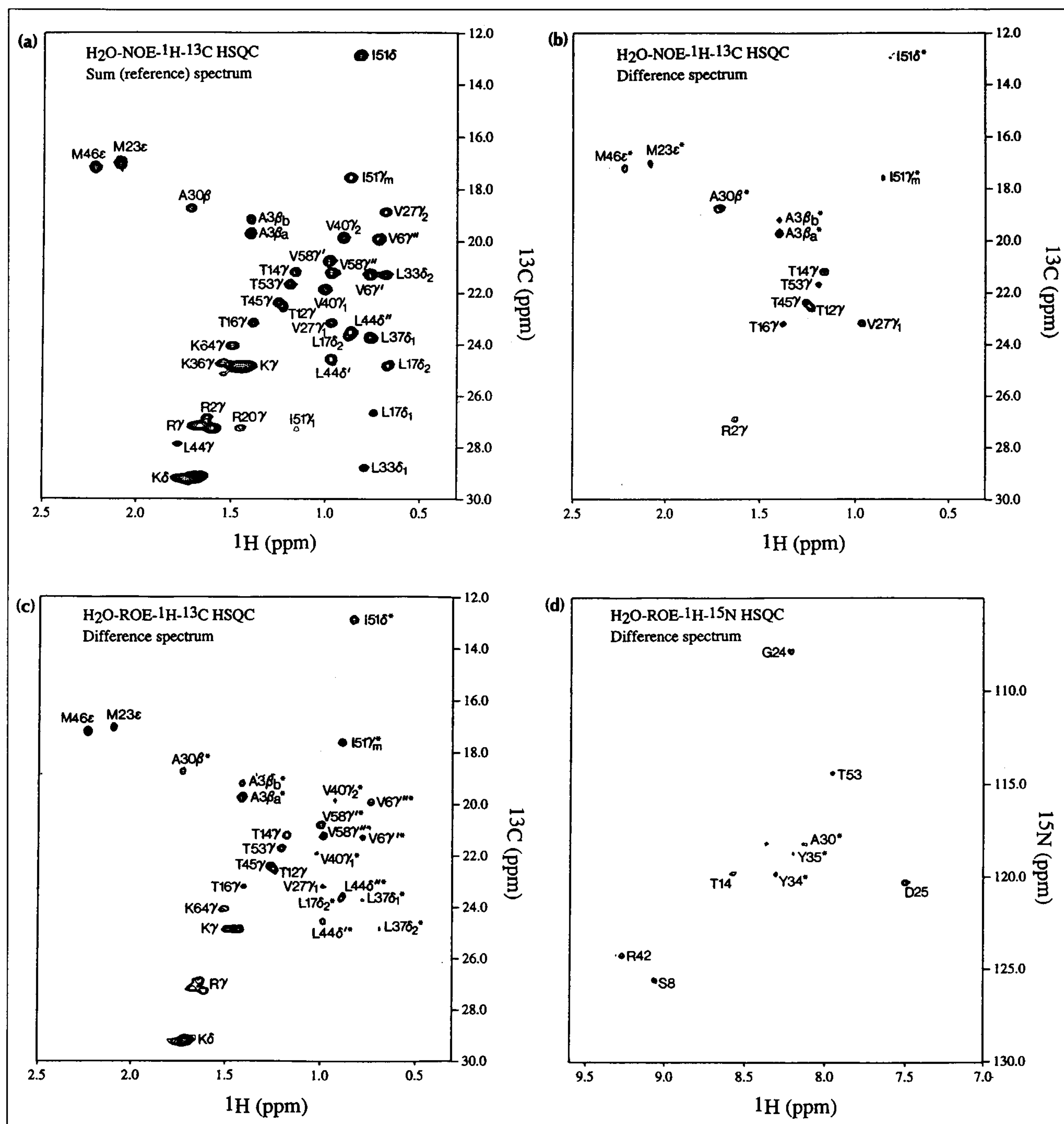

Fig. 1. Portions of the NOE and ROE spectra of the specific chicken GATA-1–DNA complex recorded with NOE/ROE mixing times of 60 ms at 25 °C. **(a)** The reference (i.e. sum) H_2O–NOE–^1H–^{13}C HSQC spectrum. **(b)** The difference H_2O–NOE–^1H–^{13}C HSQC spectrum. **(c)** The difference H_2O–ROE–^1H–^{13}C HSQC spectrum. **(d)** The difference H_2O–ROE–^1H–^{15}N HSQC spectrum. Positive and negative contours are shown as continuous and dashed lines, respectively. Cross-peaks that can be attributed unambigously to an NOE/ROE to bound water are indicated by an asterisk in panels (b) to (d); for all other peaks an indirect mechanism involving an NOE/ROE to a rapidly exchangeable proton followed by chemical exchange with water cannot be excluded. In (a), peaks involving a carbon atom attached to an even or odd number of neighboring carbon atoms are of opposite sign and are displayed as negative and positive contours, respectively. The sign of the ROE is always positive and opposite to that of the peaks observed in the reference spectrum (which are equivalent to the diagonal in a regular homonuclear 2D ^1H–^1H NOE spectrum). The sign of the NOE depends on the residency time of the water. With the exception of the methyl group of Met23, all the NOEs peaks are of the same sign as the peaks in the reference spectrum, i.e. the sign is negative, and hence the residency times of the bound water molecules are greater than 300 ps. In (d), only positive contours corresponding to ROEs are displayed; the spectrum also contains negative contours which arise from chemical exchange between rapidly exchanging amide groups and water. In panels (a), (b) and (c), note that there are two resonances for the methyl protons of Ala3 due to amino-terminal heterogeneity; in addition, methyl groups of valine and leucine that are not stereo-assigned (due to the presence of multiple rotamer conformations) are denoted by the superscripts prime and double prime. The reference (sum) spectrum in (a) is plotted at a 37.5 times higher contour level than the corresponding difference spectrum in (b).

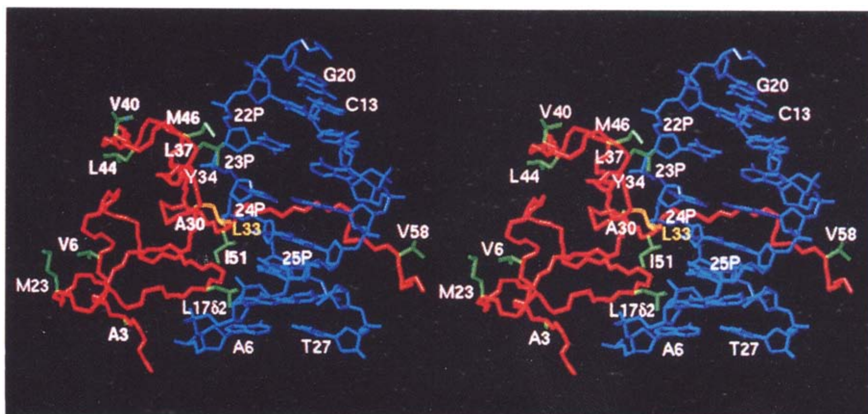

Fig. 2. Stereo view of the solution structure of the specific complex of GATA-1 and DNA, illustrating the location of groups in close proximity to bound water. The DNA is displayed in blue, the protein backbone in red, and the side chains in close proximity to bound water detected in the H_2O–ROE/NOE–HSQC experiments in green. The side chain of Leu33 which interacts directly with the bases of the DNA in the major groove and is not close to bound water is shown in yellow. In the case of Leu17, only the surface exposed $C^{\delta2}H_3$ methyl group gives rise to a ROE with bound water; the $C^{\delta1}H_3$ methyl group which interacts directly with the DNA bases is not close to bound water. The model is taken from the solution NMR structure of [5] (PDB accession number 1GAT) and was generated with the program VISP [23].

(at a mixing time of 60 ms) to water were too weak to observe and only ROEs were seen, indicating residency times in the range 200–500 ps.

All methyl groups that are exposed to solvent are associated with bound water molecules. These include Ala3, Val6 (γ_1,γ_2), Leu17 (δ_2 only), Met23, Val40 (γ_1,γ_2), Leu44 (δ_1,δ_2) and Val58 (γ_1,γ_2). Bound water molecules in contact with exposed hydrophobic surfaces are rarely seen in protein crystal structures as the water molecules are only well ordered (high occupancy and low thermal factor) if the protein donates anchor points in suitable hydrogen-bonding positions [15]. In the absence of hydrogen bonding, low occupancies and high thermal B-factors render such water molecules unobservable [15]. Indeed, there is only one crystallographic example where water surrounding an exposed hydrophobic group has been seen in the absence of hydrogen bonding, namely the water pentagons surrounding Leu18 in the 0.88 Å resolution crystal structure of crambin [16]. In contrast with the crystallographic case, bound water can be readily detected by NMR over a very wide range of residency times (with a lower limit of 50–100 ps) and is not eliminated by rapid independent rotation of a water pentagon around a rotating methyl group. Moreover, NMR experiments are not dependent on uniform ordering of the bound water which can be in somewhat different positions in different molecules. In the X-ray diffraction experiment, on the other hand, bound water could be detected only if it occupied the same position in different protein molecules; that is to say there has to be uniform ordering of the bound water throughout the protein molecules of the crystal. Consequently, rapid rotation of a water pentagon about a methyl group in the absence of direct hopping would result in smearing of the electron density beyond the limit of detectability.

The ubiquitous presence of bound water molecules close to exposed hydrophobic groups observed by NMR is completely consistent with the thermodynamic results of Privalov and Makhatadze [17, 18], who showed that the hydration of methyl groups is associated with a large negative entropy ($-40.3\,J\,mol^{-1}K^{-1}$), a negative enthalpy ($-8.28\,kJ\,mol^{-1}$) and only a small positive Gibbs free energy ($3.72\,kJ\,mol^{-1}$). Indeed, the entropy of hydration of methyl groups is comparable to that of polar groups (-41 to $-50\,J\,mol^{-1}K^{-1}$), indicating that their ordering power for water is comparable.

All methyl groups that are in close proximity to the phosphodiester backbone but do not interact with the bases, are also associated with bound water molecules. In these cases, it is clear that the water is stabilized by hydrogen bonding interactions with either the phosphate group or the O5′ and O3′ oxygen atoms. Interestingly, similarly located water molecules have also been observed in crystal structures of DNA–protein complexes [19,20]. The present results indicate the presence of a cluster of water molecules in the vicinity of the sugar-phosphate backbone of T22, T23, A24 and T25. In the minor groove there is a bound water molecule (or molecules) between the $C^{\varepsilon}H_3$ of Met46, the sugar of T22 and the phosphate of T23. In the major groove, four clusters of bound water molecules are observed: between the $C^{\delta2}H_3$ methyl group of Leu37 and the sugar and phosphate of T22; between the $C^{\delta1}H_3$ of Leu37, the backbone NH protons of Tyr34 and Ty35, and the phosphate of T23; between the methyl group of Ala30, the NH proton of Ala30 and the phosphate of A24; and, finally, between the $C^{\gamma m}H_3$ and $C^{\delta m}H_3$ methyl groups of Ile51, the sugar of A24 and the phosphate of T25. In all likelihood, the water molecule(s) associated with the NH protons Tyr34 and Tyr35, and with the NH proton of Ala30, participate in bridging hydrogen bonds between the relevant amide group of the protein and the sugar-phosphate backbone of the DNA.

In contrast to the above methyl groups, no NOE/ROEs to water were observed for either of the two Leu33 methyl groups or the $C^{\delta1}H_3$ methyl group of Leu17, all of which are involved in hydrophobic interactions

with the bases [5]. This indicates that water is excluded from the interface between the protein and the DNA bases in the major groove. In addition, complementary conventional 2D ^1H–^1H NOE experiments with ^{12}C filtering failed to detect any bound water close to the thymine methyl groups of the DNA, providing further evidence for the exclusion of water at the interface of the protein and the DNA bases. (Note that the methyl group of thymine is further than 5 Å away from the adjacent phosphate groups so that the water molecules at the interface of the phosphate groups and the protein discussed above would be too far away to give rise to NOEs with the thymine methyl groups of T22, T23 and T25.)

Biological implications

The transcription factor GATA-1 is responsible for the regulation of erythroid-expressed genes and is an essential component required for the generation of the eythroid lineage [24]. Recently, we solved the structure of the complex of the DNA-binding domain of chicken GATA-1 with its cognate DNA target site [5]. In contrast to other protein–DNA complexes, the majority of specific interactions with the DNA bases are hydrophobic in nature. Water has been implicated in the recognition process of a number of protein–DNA complexes [1–4]. We therefore set out to determine the locations of bound water molecules in the solution structure of the GATA-1–DNA complex.

We have shown that while water molecules are clustered around all surface exposed methyl groups, as well as around methyl groups in the neighborhood of the sugar-phosphate backbone, they are excluded from the interface between GATA-1 and the DNA bases in the major groove. The detection of bound water around solvent exposed methyl groups in the absence of any hydrogen-bonding interactions with protein acceptor or donor groups demonstrates the ability of hydrophobic methyl groups to order water. This is consistent with the large negative entropy of hydration of the methyl group [17,18]. The observation of bound water around protein atoms that are close to the sugar-phosphate backbone suggests that water molecules participate in bridging hydrogen bonds between functional groups on the protein (in this case the backbone amide protons of Ala30, Tyr34 and Tyr35) and the sugar-phosphate backbone. These hydrogen-bonding interactions, however, do not confer specificity. In contrast to other DNA–protein complexes which exhibit numerous hydrogen bonds between the protein and the bases of the DNA [6], the complex between GATA-1 and DNA is characterized by only two such hydrogen-bonding interactions: one between the side chain of Asn29 and the N6 atom of A8 in the major groove, and the other between Lys57($N^{\zeta}H_3$) and the O2 atom of T9 in the minor groove [5]. The remaining interactions between the protein and the DNA bases involve hydrophobic contacts between the DNA bases and the methyl groups of the protein. The absence of water at the interface between the protein and the DNA bases in the major groove lends further credence to the importance of hydrophobic interactions in stabilizing the specific interaction between chicken GATA-1 and DNA. Thus these results indicate that hydrophobic interactions can play an important role in protein–DNA recognition and specificity.

Materials and methods

Sample preparation

The expression, purification and $^{15}N/^{13}C$ labeling of the 66 residue, zinc-containing DNA-binding domain of chicken GATA-1 (residues 158–223 of the intact protein), the synthesis and purification of the 16mer duplex DNA, and the preparation of the DNA–protein complex were as described previously [5]. The sample for NMR comprised a 1:1 protein–DNA complex containing 2 mM uniformly $^{15}N/^{13}C$-labeled (> 95 %) GATA-1, 2 mM duplex DNA at natural ^{14}N and ^{12}C isotopic abundance, 2.2 mM $ZnCl_2$, 10 mM NaCl, pH 6.5 in 90 % H_2O/10 % D_2O.

NMR spectroscopy

All NMR experiments were carried out on a Bruker AMX600 600 MHz spectrometer equipped with a triple resonance self-shielded z-gradient probe. The pulse sequences employed for the 2D H_2O–NOE–HSQC and 2D H_2O–ROE–HSQC experiments were similar to those described previously [7,8] with two minor differences. First, in the ^{15}N correlated spectra ^{12}C filtering during selective excitation of the water resonance was accomplished by two ^{13}C 90° purge pulses instead of one as described in the original sequence [7]. Second, in the ^{13}C correlated spectra water suppression during the detection period was achieved by incorporating the WATERGATE water suppression scheme [21] instead of a 90_x–1.5ms–180_y–1.5ms–90_x ^1H pulse train as part of the last INEPT transfer as originally described in [8]. The mixing times employed were as follows: 60 ms for the H_2O–ROE–^1H–^{13}C HSQC, H_2O–NOE–^1H–^{13}C HSQC and H_2O–ROE–^1H–^{15}N HSQC experiments, and 100 ms for the H_2O–NOE–^1H–^{15}N HSQC experiment. Control spectra using weak presaturation followed by a 200 ms delay prior to the first selective ^1H pulse were also recorded for the H_2O–NOE–^1H–^{13}C HSQC and H_2O–NOE–^1H–^{15}N HSQC spectra.

Each difference spectrum results from two interleaved datasets recorded with $100^*(t_1) \times 1024^*(t_2)$ complex points and acquisition times of 65.4 ms (t_1) and 108.5 ms (t_2) for the H_2O–NOE/ROE–^1H–^{15}N HSQC spectra, and with $128^*(t_1) \times 512^*(t_2)$ complex points and acquisition times of 25.6 ms (t_1) and 63.5 ms (t_2) for the H_2O–NOE/ROE–^1H–^{13}C

HSQC spectra. The delay time between scans was 2.5 s, and the measuring times were 19 hours for the H_2O–NOE/ROE–^1H–^{15}N HSQC and H_2O–NOE–^1H–^{13}C HSQC spectra and 38 hours for the H_2O–ROE–^1H–^{13}C HSQC spectrum.

Spectra were processed using the in-house nmrPipe processing package (F Delaglio, unpublished program), and displayed and analyzed with the program PIPP [22]. Data were apodized with a 60° shifted sine-bell filter in the t_1 dimension, and with a 60° shifted squared sine-bell filter in the t_2 dimension prior to zero-filling and Fourier transformation.

Acknowledgements: This work was supported by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (GMC, AMG and AB). We thank Dan Garrett and Frank Delaglio for software development, S Grzesiek for development and initial parameterization of several of the pulse schemes used in the present study, Rolf Tschudin for developing and building the pulsed field gradient accessory, E de Castro and S Edelstein for the program VISP, and G Felsenfeld and C Trainor for useful discussions.

References

1. Otwinowski, Z., *et al.*, & Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321–329.
2. Lawson, C.L. & Carey, J. (1993). Tandem binding in co-crystals of a trp repressor/operator half-site complex. *Nature* **366**, 178–182.
3. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420.
4. Qian, Y.Q., Otting, G. & Wüthrich, K. (1993). NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. *J. Am. Chem. Soc.* **115**, 1189–1190.
5. Omichinski, J.G., *et al.*, & Gronenborn, A.M. (1993). NMR structure of a specific DNA complex of a Zn-containing DNA binding domain of GATA-1. *Science* **261**, 438–446.
6. Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.
7. Grzesiek, S. & Bax, A. (1993). Measurement of amide proton exchange rates and NOEs with water in $^{13}C/^{15}N$-enriched calcineurin B. *J. Biomol. NMR* **3**, 627–638.
8. Grzesiek, S. & Bax, A. (1993). The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.*, in press.
9. Brown, S.C., Weber, P.L. & Mueller, L. (1988). Toward complete ^1H NMR spectra of proteins. *J. Magn. Reson.* 77, 166–169.
10. Otting, G. & Wüthrich, K. (1989). Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Am. Chem. Soc.* **111**, 1871–1875.
11. Clore, G.M., Bax, A., Wingfield, P.T. & Gronenborn, A.M. (1990). Identification and localization of bound internal water in the solution structure of interleukin-1β by heteronuclear three-dimensional ^1H rotating frame Overhauser ^{15}N–^1H multiple quantum coherence NMR spectroscopy. *Biochemistry* **29**, 5671–5676.
12. Bothner-By, A.A., Stephens, R.L., Lee, J.T., Warren, C.D. & Jeanloz, R.W. (1984). Structure determination of a tetrasaccharide: transient nuclear Overhauser effect in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813.
13. Ernst, R.R., Bodenhausen, G. & Wokaun, A. (1987). *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Clarendon Press, Oxford.
14. Otting, G., Liepinsh, E. & Wüthrich, K. (1991). Protein hydration in aqueous solution. *Science* **254**, 974–980.
15. Jeffrey, G.A. & Saenger, W. (1991). *Hydrogen Bonding in Biological Structures.* Springer-Verlag, Berlin.
16. Teeter, M.M. (1984). Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. *Proc. Natl. Acad. Sci. USA* **81**, 6014–6018.
17. Makhatadze, G.I. & Privalov, P.L. (1993). Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **232**, 639–659.
18. Privalov, P.I. & Makhatadze, G.I. (1993). Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **232**, 660–679.
19. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907.
20. Hegde, R.S., Grossman, S.R., Laimins, L.A. & Sigler, P.B. (1992). Crystal structure at 1.7Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359**, 505–512.
21. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* **2**, 661–665.
22. Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. (1991). A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson.* **95**, 214–220.
23. de Castro, E., & Edelstein, S. (1992). *Visp 1.0 User's Guide,* University of Geneva.
24. Orkin, S.H. (1992). GATA-binding transcription factors in hematopoietic cells. *Blood,* **80**, 575–581.

Received: **3 Dec 93**; revisions requested: **13 Dec 93**; revisions received: **16 Dec 93**. Accepted **16 Dec 93**.