Faraday Discuss., 1996, **103**, 245–253

NMR studies of the hydration of biological macromolecules

Kurt Wüthrich,* Martin Billeter, Peter Güntert, Peter Luginbühl, Roland Riek and Gerhard Wider

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

The information on the hydration of proteins and nucleic acids in aqueous solution that can be obtained using high resolution nuclear magnetic resonance (NMR) spectroscopy is largely complementary to the information obtained by diffraction experiments with single crystals, which is the only other approach enabling studies of hydration at atomic resolution. The presently discussed NMR studies are primarily focused on the rate processes governing the exchange of water molecules between the bulk solvent and the hydration sites on the biological macromolecules. A novel NMR experiment for observation of protein and nucleic acid hydration is described, and long-time molecular dynamics (MD) simulations used to support the analysis of NMR data on hydration water in the intermolecular interface of protein–DNA complexes are presented.

During the period 1990–1994, X-ray diffraction with single crystals and nuclear magnetic resonance (NMR) spectroscopy in solution have provided more than 1200 new three-dimensional structures of proteins and nucleic acids at atomic resolution.[1] The resulting database includes a wealth of information on hydration water associated with the macromolecular structures and is therefore of high relevancy to the theme of these *Faraday Discussions.*

At first glance, the major impact of the availability of two techniques for atomic resolution structure determination of biological macromolecules lies in the fact that both X-ray crystallography and NMR contribute a wealth of new structures at a so far continuously increasing rate. However, the two techniques further provide complementary data for a given system. Complementarity of the two methods results from the facts that the measurements are made in crystals and in solution, respectively, and that the time-scales of the two experiments are widely different.[2,3] An NMR structure determination, representing atomic spatial resolution, can thus be supplemented with data on rate processes, with a temporal resolution extending over a wide range of frequencies.

Studies of hydration water afford a particularly impressive illustration of the complementarity of the two methods. Individual hydration water molecules continuously exchange in and out of particular hydration sites of the macromolecular structure, and NMR measures the average duration of these visits, in addition to the population of the individual sites. Diffraction experiments also probe the total fraction of time that a particular hydration site is occupied by a water molecule, but they are insensitive to the residence time at that site on any particular visit. Typically, water molecules in crystals of globular proteins are observed in a small number of solvent-inaccessible hydration sites in the interior of the molecular structure, as well as in discrete sites covering 30–60% of the molecular surface. Although they usually give rise to somewhat bigger *B* factors, the diffraction data on hydration waters are not qualitatively different from those of other parts of the structure. For studies in solution, considering the high water

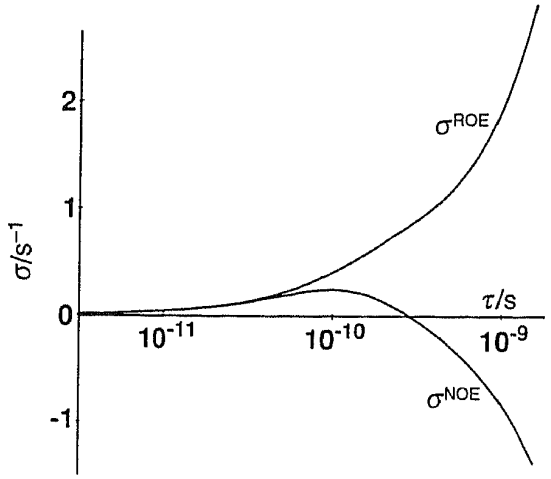

Fig. 1 Plot of the cross relaxation rates σ^{NOE} and σ^{ROE} between two protons *vs.* the effective correlation time τ for modulation of the dipole–dipole coupling by Brownian rotational tumbling. The curves were calculated for a ^1H frequency of 600 MHz and a ^1H–^1H distance of 2.0 Å. (Adapted from ref. 4).

concentration in aqueous protein solutions, one anticipates intuitively that the entire protein surface must be covered with water molecules. Water molecules in this first hydration layer are observable by NMR experiments, which further show that the residence times for typical non-specific surface hydration water molecules are very short, in the approximate range 20–200 ps at 10 °C.[4,5] For interior, solvent-inaccessible hydration water, observation of exchange with the bulk solvent manifests internal time fluctuations of the protein structure. A frequency range from *ca.* 50 s^{-1} to 2×10^9 s^{-1} has been reported for these rate processes.[4,6–8] Similar data to those for interior hydration of globular protein molecules in solution have been obtained from NMR studies of the intermolecular interface in complexes with proteins, in particular protein–DNA complexes.[9,10] In this paper we briefly review the basic principles underlying the use of high resolution NMR for studies of macromolecular hydration, and then report on new experimental and theoretical work with proteins and protein–DNA complexes.

NMR studies of protein and nucleic acid hydration in aqueous solution

Identification of individual molecules of hydration water by NMR in solution relies on the observation of nuclear Overhauser effects (NOEs) between hydrogen atoms of the polypeptide chain and water protons. These NOEs are due to time-dependent dipole–dipole coupling between nearby protons. The intensity of the NOE between two hydrogen atoms i and j is proportional to d_{ij}^{-6}, where d_{ij} is the distance between the two protons. Because of the strong distance dependence, NOEs can be observed only between spatially close protons, *i.e.* for $d_{ij} \lesssim 4.0$ Å. The NOE intensity is further related to a correlation function describing the stochastic modulation of the dipole–dipole coupling between the two protons. For studies of hydration, it is important that this correlation function may be governed either by the Brownian rotational tumbling of the hydrated protein molecule or by interruption of the dipolar interaction through translational diffusion of the interacting spins, whichever is faster.[4]

The NOEs can be measured either by NOE spectroscopy in the laboratory frame of reference (NOESY)[11] or in the 'rotating frame' (ROESY).[12] In suitably executed experiments, the measured NOE intensities reflect directly the cross-relaxation rates in the laboratory frame, σ^{NOE}, or in the rotating frame, σ^{ROE}, respectively. The two rates differ

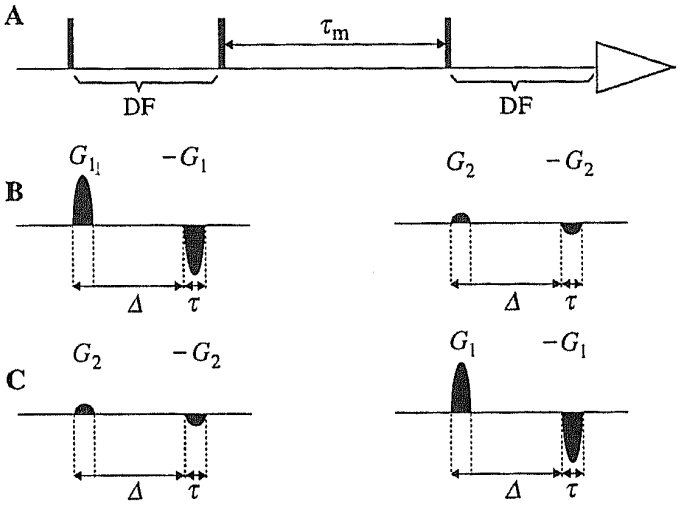

Fig. 2 (A) Schematic survey of a one-dimensional transient NOE experiment with diffusion filters. The three vertical bars represent 90° hard pulses at the ^{1}H frequency. The initial 180° pulse of the transient NOE experiment[31] is decomposed into two 90° pulses which enclose the first diffusion filter. The second diffusion filter is inserted between the observation pulse and the start of data acquisition. (B) and (C) Diffusion filters, which consist of two pulsed field gradients (PFG) of length τ spaced by the diffusion time Δ. The strength of the PFGs is represented by the height of the symbol. The HYDRA-N experiment for selective observation of hydration water molecules consists of the difference between two recordings of (A) which use the arrangements of the diffusion filters shown in (B) and (C), respectively, and otherwise identical conditions.

in their functional dependence on the spectral densities in such a way that studies of the sign and value of the ratio $\sigma^{NOE}/\sigma^{ROE}$ can be used for investigations of the rate processes that determine the effective correlation time for modulation of the dipole–dipole coupling. As mentioned above, such rate processes include the Brownian rotational tumbling of the molecules considered and translational diffusion of two interacting spins, for example, upon rapid dissociation of a bimolecular complex. As an illustration, the plots of σ^{NOE} and σ^{ROE} *vs.* the effective rotational correlation time (Fig. 1) show that positive values of $\sigma^{NOE}/\sigma^{ROE}$ are expected only for very short correlation times, *i.e.* shorter than *ca.* 0.3 ns at a proton frequency of 600 MHz (for modulation by translational diffusion, the sign change occurs at *ca.* 0.5 ns), and that the NOE and ROE intensities will increase rapidly for longer correlation times. NOEs with hydration water have been observed using homonuclear two-dimensional (2D) NOESY and 2D ROESY,[13] homonuclear three-dimensional (3D) TOCSY-relayed NOESY and ROESY, and 3D ^{15}N- or ^{13}C-correlated NOESY and ROESY.[9,14–16] An important and technically demanding feature of these NMR experiments is that they have to be performed in H_2O solution, and that the intense solvent resonance can be suppressed only at the very end of the NOESY or ROESY pulse sequences.

HYDRA experiments for identification of NOEs with hydration water

The goal of the HYDRA experiment[17] is to overcome a technical limitation which is quite ubiquitous in high resolution NMR studies of the hydration of macromolecules such as proteins: the water line and some protein resonances usually overlap,[18] which makes the distinction between water–protein interactions and intraprotein ^{1}H–^{1}H NOEs difficult. Even in 2D or 3D NMR experiments the unambiguous assignment of water–protein interactions is then often not possible in a single experiment. In HYDRA NMR pulse sequences for hydration studies, the first step consists of a 1D NMR experiment that selects strictly for intermolecular protein–water interactions. Thereby the

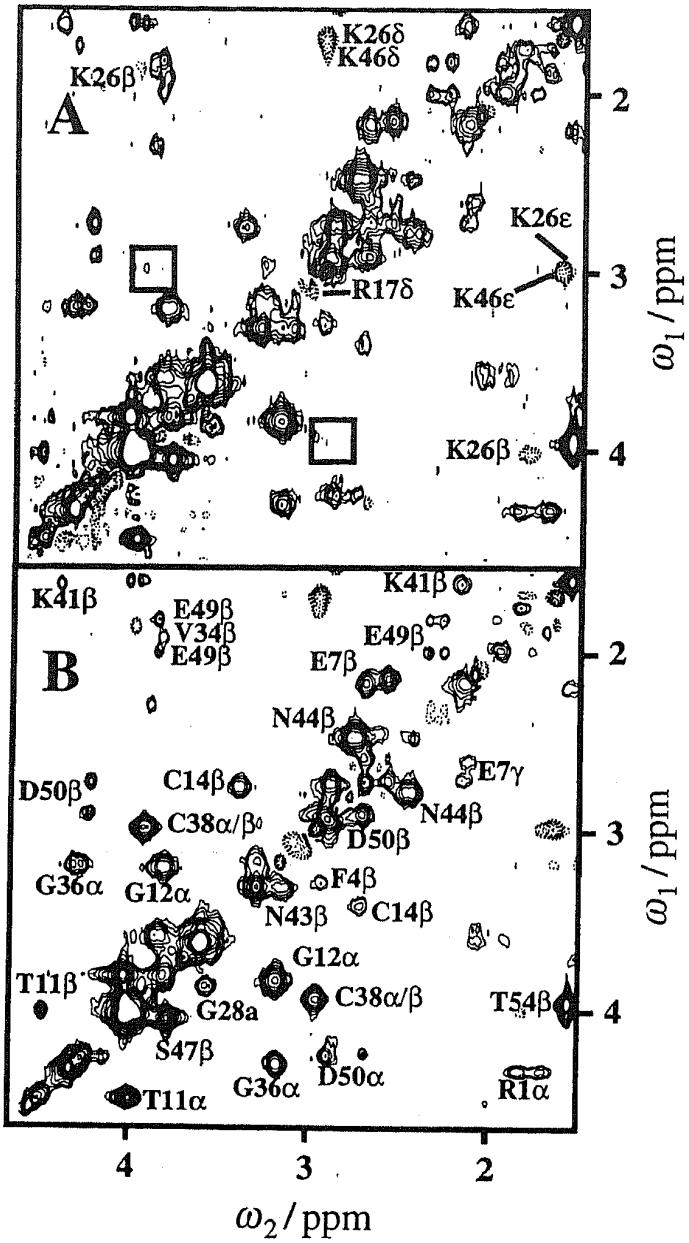

Fig. 3 Contour plots of the spectral region (ω_1 = 1.6–4.5 ppm, ω_2 = 1.6–4.5 ppm) of 2D TOCSY-relayed NOE difference experiments. The NOE mixing time was τ_m = 60 ms (Fig. 2) and the TOCSY mixing time was 27 ms. The sample used was 20 mM BPTI in 90% H_2O/10% D_2O at pH 3.5 and T = 277 K. The water resonance was suppressed by combining WATERGATE[32] with the diffusion filter before acquisition.[17] Positive cross peaks are drawn with solid lines and negative cross peaks with dotted lines. (A) 2D [^{1}H,^{1}H]-TOCSY-relayed HYDRA-N spectrum recorded with the diffusion filters of Fig. 2 (Δ = 7.8 ms, gradient strengths = 105 and 15 G cm^{-1}, respectively, τ = 1 ms, t_{1max} = 23.2 ms, 768 scans per t_1 increment). Negative cross peaks (representing positive values for σ_{NOE}, see Fig. 1) are identified with the protein proton interacting with the water, using the one-letter amino acid code, the sequence position, and greek letters to indicate the proton position. Two squares indicate the positions where the TOCSY-relayed NOE cross peaks Cys 38 H^α–Cys 38H^β would be observed. (B) 2D [^{1}H,^{1}H]-TOCSY-relayed NOE spectrum recorded using radiation damping for selection of the water resonance [t_{1max} = 30.7 ms, 320 scans per t_1 increment, otherwise same parameters as in (A)]. Positive water–protein cross peaks and the cross peak between H^α and H^β of Cys 38 are identified.

separation of intermolecular solvent–protein NOEs or chemical exchange effects from intramolecular NOEs is based on the different diffusion properties of individual water molecules and the biological macromolecules (Fig. 2). For practical applications the initial 1D HYDRA experiment can be used as a platform for relays by higher-

dimensional experiments with improved peak separation (Fig. 3). HYDRA experiments can be performed either in the laboratory frame of reference (HYDRA-N) or in the rotating frame (HYDRA-R).

In liquids, translational diffusion during a time interval Δ results in different NMR signal intensities depending on whether or not Δ is bounded by two identical pulsed magnetic field gradients (PFG), leading to signal intensities S with and S_0 without gradients according to[19]

$$\frac{S}{S_0} = \exp\left[-\gamma^2\tau^2G^2\left(\Delta - \frac{\tau}{3}\right)D\right] \quad (1)$$

D is the translational diffusion coefficient, γ the gyromagnetic ratio (26.7519×10^7 rad T^{-1} s^{-1} for protons), τ the length of the gradient, G the gradient strength, and Δ the time period between the two gradients (Fig. 2). As a rule, water molecules diffuse *ca.* 15–20 times faster than proteins in the molecular mass range 10–20 kDa, so that intermolecular water–protein NOEs can be distinguished from intramolecular NOEs with the use of diffusion filters.[20] A suitably designed experiment, which measures the difference between the signals obtained with strong and weak PFGs in a diffusion filter (Fig. 2), selectively records intermolecular protein–water interactions. This approach can be used for a wide range of NOE mixing times, including the very short values that are typically required for studies of surface hydration in the presence of spin diffusion from interior hydration water molecules and from exchanging protons of amino acid side chains or nucleic acid riboses and bases.[21,22]

As an illustration, Fig. 3 compares a 2D [^{1}H,^{1}H]-TOCSY-relayed HYDRA-N spectrum with a 2D [^{1}H,^{1}H]-TOCSY-relayed NOE experiment which uses radiation damping for the water selection (ref. 17 and 23). The spectra were recorded with the protein basic pancreatic trypsin inhibitor (BPTI), which contains four interior, solvent-inaccessible hydration water molecules.[13] This comparison documents that HYDRA-based higher-dimensional experiments can be applied simultaneously for unambiguous separation of water–protein interactions and intramolecular protein–protein NOEs, and for detection of very weak NOEs with surface hydration water. In Fig. 3A the diagonal comprises peaks which originate from water–protein NOEs or water–protein chemical exchange. The TOCSY-relay transfers magnetization from the diagonal peaks to scalar coupled protons. Protons coupled to the proton manifesting an NOE peak on the diagonal then show a cross peak at the same ω_1 frequency as the diagonal peak. All the cross peaks identified in Fig. 3A have negative sign and represent positive values of σ^{NOE} (Fig. 1), which manifest NOEs with short-lived surface hydration water molecules.[4] That the spectrum of Fig. 3A contains exclusively interactions with water protons is demonstrated by the absence of the TOCSY-relayed NOE between Cys 38 H^β and Cys^{38} H^α (squares in Fig. 3A), since the chemical shift of the Cys 38 H^α resonance coincides nearly perfectly with that of the solvent water resonance. The Cys 38 H^β–Cys 38 H^α peak is present in the spectrum of Fig. 3B, which shows that the clean separation of hydration interactions and intramolecular protein–protein NOEs by the HYDRA scheme implemented in the 2D [^{1}H,^{1}H]-TOCSY-relayed HYDRA-N experiment is superior to the performance of other schemes presently available for this purpose. In Fig. 3B, positive signals are identified that were previously assigned either to NOEs with the four interior water molecules in BPTI, or to exchange peaks or NOEs with rapidly exchanging side chain hydroxy protons of the protein.[13]

Complementation of NMR data on macromolecular hydration by MD simulations

For water molecules in hydration sites on the surface of proteins and in the major groove of DNA, lifetimes in the range 20–200 ps are typically observed.[4] Therefore,

using a simulation extending over a time span of one to several nanoseconds, which can be handled by present supercomputers for a protein in a water bath, a statistically significant comparison of the experiments with theoretical models can be achieved. Highly satisfactory agreement between measurements and MD simulations has actually been obtained.[5]

NMR studies of the rate processes with interior hydration water are incomplete in the sense that only bounds on frequency ranges rather than precise frequency values have been reported. For example, for the solvent-inaccessible interior hydration water molecules in the outstandingly stable protein basic pancreatic trypsin inhibitor (BPTI), upper bounds on the lifetimes with respect to exchange in and out of the hydration sites were determined with different NMR techniques,[6–8] and all the data are in agreement with an upper limit of 2 μs. The lower bound is about 1 ns,[4] so that the actual values may vary over several orders of magnitude. Similar information is available on the lifetimes of hydration water molecules in the intermolecular interface of protein–DNA complexes, except that with the data collected so far the upper limit of the lifetimes may be as long as 20 ms.[9] To obtain an indication of whether the actual lifetimes are nearer to the upper or lower limit of the experimentally defined range, a 2 ns MD simulation was performed for a DNA complex of the *Antennapedia* (*Antp*) homeodomain (Fig. 4) in explicit water.[24] The results thus obtained further supplement the measurement of lifetimes with a description of the pathways associated with the water exchange in and out of the hydration sites. For the MD simulation the program OPAL[25] was used. During

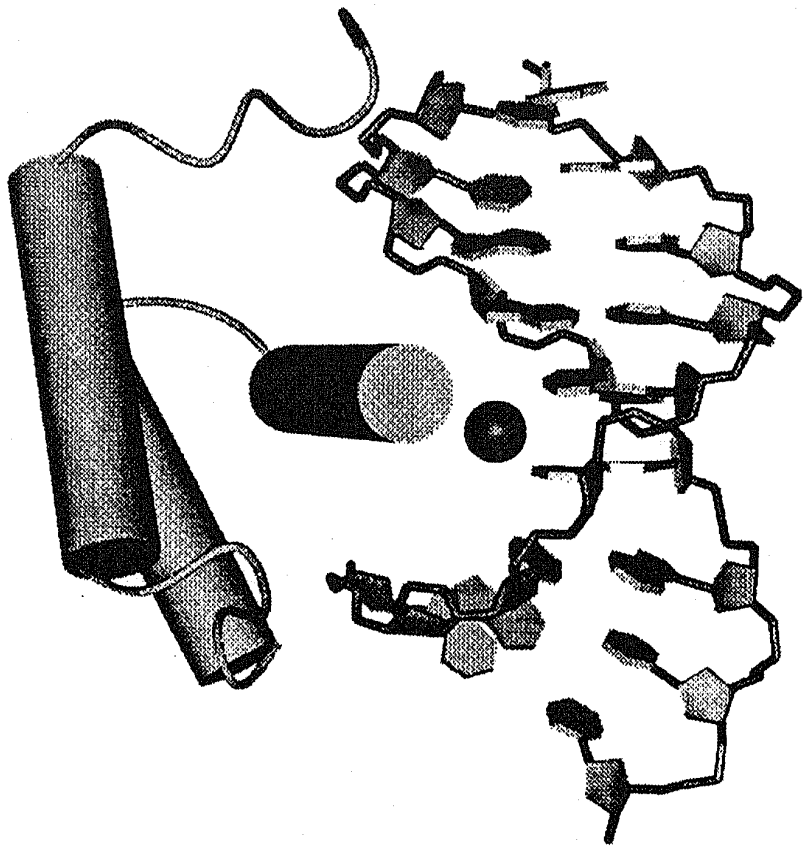

Fig. 4 The *Antennapedia* homeodomain–DNA complex used for the MD simulation of hydration. The protein backbone is shown by cylinders representing the helices and a thin ribbon for the chain ends and the connecting loops. The DNA is shown with bonds linking the backbone heavy atoms and planes representing the sugar rings and the bases. This complex was immersed in a water bath with 2714 water molecules. These were numbered 1–2714 at the outset, and the numbers were followed during the simulation. The sphere in the centre identifies the initial position of the water molecule 384 in the protein–DNA interface.

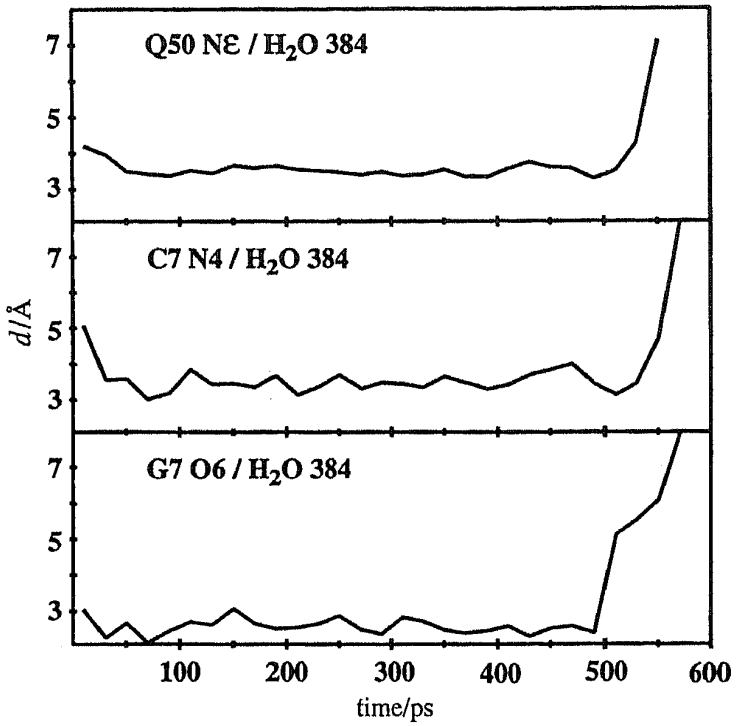

Fig. 5 Variation of the distances between the oxygen atom of the water molecule 384 (Fig. 4) and three hydrogen bonding partners in the homeodomain and the DNA during the initial 600 ps of the MD simulation. At the start of the MD simulation the water molecule 384 is in the protein–DNA interface, and it diffuses into the bulk solvent during the interval from *ca.* 500–550 ps. Distances near 3.0 Å indicate intermolecular hydrogen bond formation with this water molecule.

an MD simulation, OPAL computes with an average of 1.5 GFlops on a NEC SX-3 supercomputer. The AMBER force field[26] was used for the definition of the interaction energies. The coordinates of the starting structure of the MD calculation were taken from one of the 16 conformers that represent the NMR solution structure of the *Antp* homeodomain–DNA complex (Fig. 4).[10] An ellipsoid-shaped spatial boundary surrounds the 1007 protein atoms and 572 DNA atoms of the macromolecular complex, with an initial minimal distance of 6.0 Å from any atom of the macromolecules. The ellipsoid was filled with 2714 water molecules, and atoms were confined to the interior of the ellipsoid by a potential rising from zero to infinity within a boundary layer of 1.0 Å thickness. A trajectory covering a time range of 2 ns was then performed, with integration steps of 2.5 fs, and the coordinates of the protein, the DNA and the water atoms were recorded in intervals of 1 ps. Equilibration was reached within 100 ps, and thereafter the fluctuations about the mean positions of the protein and DNA heavy atoms rarely exceeded 1.0 Å.

In the MD simulation the water–water hydrogen bonds are very short-lived, in the range of the time resolution given by the sampling of the trajectory at 1 ps intervals. Different individual water molecules exhibit different behaviour depending on their interactions with the macromolecular complex. Most water molecules remain outside of the complex, with occasional contacts to the surface. In agreement with previous experiments[4,27] and MD simulations[5] with different proteins, these surface hydration water molecules have very short residence times of the order 10–60 ps. After the equilibration of the starting system (Fig. 4) during the initial 100 ps of MD simulation, the protein–DNA interface in the major groove contained several hydration water molecules at all instances. Overall, during the 2 ns of MD simulation *ca.* 5% of all water molecules penetrated the protein–DNA interface. These interfacial water molecules remain only for a limited time before leaving the interface either along the same way

they came in or by a different pathway. As an illustration, Fig. 5 shows the time evolution of the distances between one of the water molecules that was located in the interface at the start of the simulation and three hydrogen-bonded heavy atoms of the protein and the DNA. From the end of the equilibration period at 100 ps until *ca.* 500 ps, this water molecule forms a short hydrogen bond with the DNA base G7, and somewhat longer hydrogen bonds with the base C7 and Gln 50. The steep simultaneous increase of all three intermolecular distances between 500 and 600 ps reflects the fact that the water molecule 384 diffuses out of the protein–DNA interface into the bulk solvent. The top trace of Fig. 6 shows a similar plot for the non-bonding contact between the water molecule 384 and the δ-methyl carbon of Ile 47, which extends over the entire length of the MD trajectory. It shows that this water molecule stays in the bulk solvent during the period 600 ps to 2000 ps. The time dependence of the distances between the Ile 47 δ-methyl carbon and the water molecules 359, 509, 866, 903, 1269, 1400 and 2608 (Fig. 6) shows that the water 384 is replaced in its initial location in an interfacial hydration site by other water molecules which were located in the bulk solvent at the outset of the simulation. For all eight water molecules in Fig. 6, which are representative of the entire population of water molecules that were transiently located in the protein–DNA interface, the residence time in the interface is of the order 400–1000 ps. These model calcu-

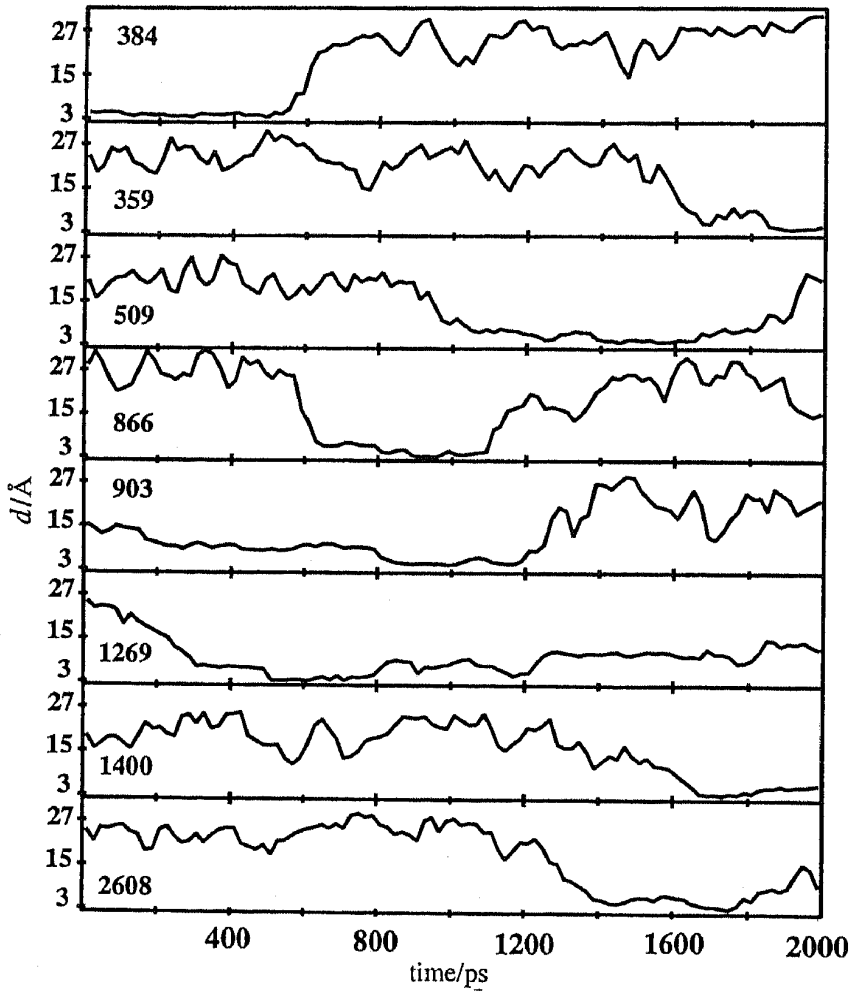

Fig. 6 Variation of the distances between the oxygen atoms of the eight water molecules identified by the numbers in each panel and the δ-carbon of Ile 47 during the entire 2 ns duration of the MD simulation of the *Antp* homeodomain–DNA complex in explicit water. Because Ile 47 is buried within the protein–DNA interface, distances shorter than *ca.* 5 Å indicate penetration of the water molecules into the interface, while distances of >15 Å identify snapshots in which the water molecules are clearly outside of the interface.

lations thus indicate that the actual lifetimes of solvent-in accessible hydration waters in the intermolecular interface in protein–DNA complexes are closer to the lower end of the range defined by NMR experiments.

Conclusions

The three currently available techniques for characterization of macromolecular structures at atomic resolution, X-ray diffraction with single crystals, NMR in solution and computational modelling, provide important complementary aspects of the hydration of proteins and nucleic acids. Advances in NMR methods for studies of hydration are still forthcoming, and the NMR studies provide novel insights into the physico-chemical properties of proteins and nucleic acids in solution. For the biologically highly relevant protein–DNA complexes, the combined information from NMR studies and MD simulations indicates new avenues for the functional interpretation of the rapidly growing data base of crystal structure information on hydration water in the protein–DNA interface, for example, ref. 28–30.

We thank the Centro Svizzero di Calcolo Scientifico for use of the NEC SX-3 computer, the Schweizerischer Nationalfonds for financial support (project 31.32033.91), and R. Marani for careful processing of the text.

References

1 W. A. Hendrickson and K. Wüthrich, *Macromolecular Structures 1995*, Current Biology, London, 1995.
2 K. Wüthrich, *Acta Crystallogr., Sect. D*, 1995, **51**, 249.
3 K. Wüthrich, *NMR in Structural Biology*, World Scientific, Singapore, 1995.
4 G. Otting, E. Liepinsh and K. Wüthrich, *Science*, 1991, **254**, 974.
5 R. M. Brunne, E. Liepinsh, G. Otting, K. Wüthrich and W. F. van Gunsteren, *J. Mol. Biol.*, 1993, **231**, 1040.
6 G. Otting, E. Liepinsh and K. Wüthrich, *J. Am. Chem. Soc.*, 1991, **113**, 4363.
7 V. Dötsch and G.Wider, *J. Am. Chem. Soc.*, 1995, **117**, 6064.
8 V. P. Denisov, B. Halle, J. Peters and H. D. Hörlein, *Biochemistry*, 1995, **34**, 9046.
9 Y. Q. Qian, G. Otting and K. Wüthrich, *J. Am. Chem. Soc.*, 1993, **115**, 1189.
10 M. Billeter, Y. Q. Qian, G. Otting, M. Müller, W. J. Gehring and K. Wüthrich, *J. Mol. Biol.* 1993, **234**, 1084.
11 Anil-Kumar, R. R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 1980, **95**, 1.
12 A. A. Bothner-By, R. L. Stephens, J. Lee, C. D. Warren and R. W. Jeanloz, *J. Am. Chem. Soc.*, 1984, **106**, 811.
13 G. Otting and K. Wüthrich, *J. Am. Chem. Soc.*, 1989, **111**, 1871.
14 K. Wüthrich and G. Otting, *Int. J. Quant. Chem.* 1992, **42**, 1553.
15 J. D. Forman-Kay, A. M. Gronenborn, P. T. Wingfield and G. M. Clore, *J. Mol. Biol.*, 1991, **220**, 209.
16 I. P. Gerothanassis, B. Birdsall, C. J. Bauer, T. A. Frenkiel and J. Feeney, *J. Mol. Biol.*, 1992, **226**, 549.
17 G. Wider, R. Riek and K. Wüthrich, *J. Am. Chem. Soc.*, in the press.
18 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
19 E. O. Stejskal and J. E. Tanner, *J. Chem. Phys.*, 1965, **42**, 288.
20 P. C. M. Van Zijl and C. Moonen, *J. Magn. Reson.*, 1990, **87**, 18.
21 G. Otting, E. Liepinsh, B. T. Farmer II and K. Wüthrich, *J. Biomol. NMR*, 1991, **1**, 209.
22 E. Liepinsh, G. Otting and K. Wüthrich, *J. Biomol. NMR*, 1992, **2**, 447.
23 G. Otting and E. Liepinsh, *J. Biomol. NMR*, 1995, **5**, 420.
24 M. Billeter, P. Güntert, P. Luginbühl and K. Wüthrich, *Cell*, in the press.
25 P. Luginbühl, P. Güntert, M. Billeter and K. Wüthrich, to be published.
26 S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *J. Comp. Chem.*, 1986, **7**, 230.
27 G. Otting, E. Liepinsh and K. Wüthrich, *J. Am. Chem. Soc.*, 1992, **114**, 7093.
28 Z. Otwinowski, R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi and P. B. Sigler, *Nature (London)*, 1988, **335**, 321.
29 J. A. Hirsch and A. K. Aggarwal, *EMBO J.*, 1995, **14**, 6280.
30 D. S. Wilson, B. Günther, C. Desplan and J. Kuriyan, *Cell*, 1995, **82**, 709.
31 S. L. Gordon and K. Wüthrich, *J. Am. Chem. Soc.*, 1978, **100**, 7094.
32 M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661.

Paper 6/01936F; *Received* 20*th March*, 1996