JOURNAL OF THE ROYAL SOCIETY Interface

Nuclear magnetic resonance analysis of protein–DNA interactions

S. Campagne, V. Gervais and A. Milon

J. R. Soc. Interface published online 9 March 2011
doi: 10.1098/rsif.2010.0543

JOURNAL OF THE ROYAL SOCIETY Interface

FirstCite® e-publishing

J. R. Soc. Interface
doi:10.1098/rsif.2010.0543
Published online

REVIEW

Nuclear magnetic resonance analysis of protein–DNA interactions

S. Campagne[1,2], V. Gervais[1,2] and A. Milon[1,2,*]

[1]*Université de Toulouse, UPS (Department of Structural Biology and Biophysics), F-31077 Toulouse, France*
[2]*CNRS; IPBS (Institute of Pharmacology and Structural Biology), 205 route de Narbonne, F-31077 Toulouse, France*

Recent methodological and instrumental advances in solution-state nuclear magnetic resonance have opened up the way to investigating challenging problems in structural biology such as large macromolecular complexes. This review focuses on the experimental strategies currently employed to solve structures of protein–DNA complexes and to analyse their dynamics. It highlights how these approaches can help in understanding detailed molecular mechanisms of target recognition.

Keywords: structural biology; protein–DNA complex; nuclear magnetic resonance

1. INTRODUCTION

Transcription, DNA replication and repair are key dynamic cellular processes that require extremely tight regulation by DNA-binding proteins. In particular, these proteins have to recognize their specific DNA targets in the presence of a large amount of non-specific DNA-binding sites.

In most cases, transcription factors expand a modular organization [1], encompassing a DNA-binding domain specialized in DNA recognition [2] and additional domains implicated in signal transduction via interaction with other protein partners [3]. The large amount of available structural data on DNA-binding domains has highlighted an important structural diversity that allows structural classifications to be proposed [4] and provides clues for understanding the molecular mechanisms involved in DNA recognition. The DNA-binding domain is mostly restricted to recognizing and targeting specific DNA in the genome, and strong binding affinities have been selected during evolution.

To gain insights into the mechanisms underlying specific DNA recognition, it is necessary to have detailed structural information on DNA–protein complexes. X-ray diffraction remains the most powerful method for determining the structure of large macromolecular complexes, as illustrated with the crystal structure of the complex formed by eight transcription factors bound to the interferon-β enhancer [5,6]. However, while X-ray crystallography is very useful for large molecular assemblies, its general application is limited by the difficulties in crystallizing highly dynamic and transient complexes. Over recent decades, liquid state nuclear magnetic resonance (NMR) has progressed in terms of sensitivity and spectral resolution with the production of very high magnetic field spectrometers together with important methodological developments, such as transverse relaxation optimized spectroscopy [7,8] (TROSY), opening up the way for NMR studies of larger macromolecules or complexes. In the case of protein–DNA complexes, most of the recent NMR advances can be applied, providing structural restraints that can be used to perform structure calculations. Two categories of information can be collected: ambiguous restraints generally derived from interaction surface mapping; and unambiguous restraints provided by intermolecular nuclear Overhauser effects (NOEs), residual dipolar couplings and paramagnetic relaxation enhancement (PRE) experiments. All these data can be combined together to compute high-quality structures of protein–DNA complexes.

Furthermore, in contrast to X-ray diffraction, NMR spectroscopy is performed at room temperature and gives access to the dynamic features of macromolecules. By studying ^{15}N relaxation rates of amide nitrogens, information on the overall diffusion and on the local behaviour of the proteins can be extracted from the analysis of spectral density functions.

The important work performed on the Lac operator [9–13] has provided keys to understanding the molecular mechanisms by which the Lac repressor binds to a specific target and a non-specific target. In the case of protein–DNA recognition, adaptation and plasticity appear to be the two major properties that confer specific recognition capacity. Furthermore, liquid state

*Author for correspondence (alain.milon@ipbs.fr).

Received 6 October 2010
Accepted 15 February 2011

NMR can also be used to have access to highly transient intermediates, and it is now possible to follow protein diffusion by NMR and to target the search mechanism along the DNA [14]. The recent work carried out on homeodomain HoxD9 DNA recognition has clearly demonstrated that HoxD9 binds to non-specific DNA with the same binding mode and orientation as that observed in the specific complex [15]. However, non-specific DNA binding is always associated with exchange processes, leading to different dynamic behaviours of side chains located at the interface with the DNA.

This paper reviews the current approaches in the field of NMR structural studies specific to protein–DNA complexes by NMR and focuses on new developments. All major stages from sample preparation, interface mapping and structural restraint collection to structure calculation and dynamic characterization are discussed.

2. SAMPLE PREPARATION

In general, when dealing with macromolecular complexes, different labelling strategies are employed. In most cases, only one partner is labelled while the other remains unlabelled, which has the advantage of allowing the number of observable NMR signals to be selected. Two samples can be prepared: in the first one, which is the most frequently encountered, the protein is ^{15}N and/or ^{13}C labelled while the DNA is unlabelled, and in the other one, which is optional, the DNA is ^{15}N and/or ^{13}C labelled and the protein unlabelled.

This provides a convenient means to follow the resonances during titration experiments and to use filtered experiments to select or filter out magnetization of specific proton sets. The advancements in biochemistry and dedicated-isotope labelling techniques allow labelled proteins to be produced, while samples of labelled DNA can be prepared by chemical or enzymatic methods, as previously well reviewed [16]. Double-stranded DNA is formed after annealing of the two strands at high temperature followed by slow cooling down. The DNA duplex may be further purified on a preparative ion-exchange column and desalted by dialysis before preparing the protein–DNA complex [17]. In order to mix the two strands in equimolar quantities, it is easiest to progressively add one of the two strands to a solution of the complementary strand, and NMR spectra are recorded after each addition step and integrated.

Preparing stable protein–DNA complexes for NMR, i.e. 0.1–1 mM solutions of complex, remains a difficult task because of the basic nature of DNA-binding domains (thus positively charged) contrasting with the negatively charged nucleic acid phosphate groups. A strong electrostatic interaction between the two partners occurs at high concentration and may give rise to sample precipitation. Different strategies have been described in the literature to address this issue: (i) increasing the salt concentration, (ii) reducing the number of DNA charges by shortening its length, and (iii) using site-directed mutagenesis to replace basic residues on the protein surface outside the DNA recognition interface. Other methods were shown to be useful in improving the solubility problem, as in the case of the Dread ringer protein–DNA complex, in which a C-terminal phenylalanine residue, responsible for the exchange process, was mutated to a leucine residue, thereby increasing the NMR spectra quality [18]. Mutation of cysteine residues to prevent oxidation has been shown to improve the stability of the protein–DNA complex, as in studies of the integrase–DNA complex, in which a cysteine was mutated to an alanine [19], or in the case of the THAP domain of hTHAP1, in which two free cysteines outside the C_2CH zinc coordination motif were mutated to serines [20].

The main problems in sample preparation remain the solubility and the stability of the complex, because, in order to identify intermolecular NOEs, the sample needs to be highly concentrated and stable at a reasonable temperature (between 25°C and 40°C) in order to obtain a sharp signal.

3. INTERACTION SURFACE MAPPING

3.1. Chemical shift mapping

A variety of techniques have been proposed to map out protein-binding interfaces and will be briefly described. The easiest NMR method is known as chemical shift mapping, and it is commonly based on the detection of the protein backbone ^{15}N and HN chemical shifts in heteronuclear single quantum coherence spectroscopy (HSQC) spectra. The development of TROSY makes it possible to record NMR spectra for large complexes with high spectral resolution. By recording ^{15}N HSQC spectra of the free and the bound ^{15}N-labelled protein, normalized chemical shift perturbations (CSPs) can be computed using

$$\mathrm{CSP} = \sqrt{\Delta\delta^2\mathrm{HN} + \left(\frac{\gamma_\mathrm{N}}{\gamma_\mathrm{H}}\right) \times \Delta\delta^2\mathrm{N}},$$

where $\Delta\delta$HN and $\Delta\delta$N are the observed chemical shift variations in parts per million and γ_H and γ_N are the gyromagnetic ratio of each nucleius.

Chemical shifts are very sensitive to subtle perturbations of the chemical environment. While binding-induced internal structural rearrangements of the protein may, in some cases, affect protein chemical shifts, generally it is the proximity of DNA phosphates or the ring currents induced by aromatic bases that are responsible for important CSPs. Therefore, NMR methods that monitor CSPs of protein amide groups upon DNA binding are widely used to rapidly map out the interaction surface [21–23].

For instance, in the case of the interaction between the C_2CH THAP zinc finger domain of human THAP1 and its specific DNA target, very small CSP values were observed for residues located in the α-helix region, while residues within the β-sheet or the N-terminus loop were strongly affected (figure 1*a*), providing evidence that the domain uses its β-sheet and surrounding loops to recognize the DNA instead of its α-helix [20].

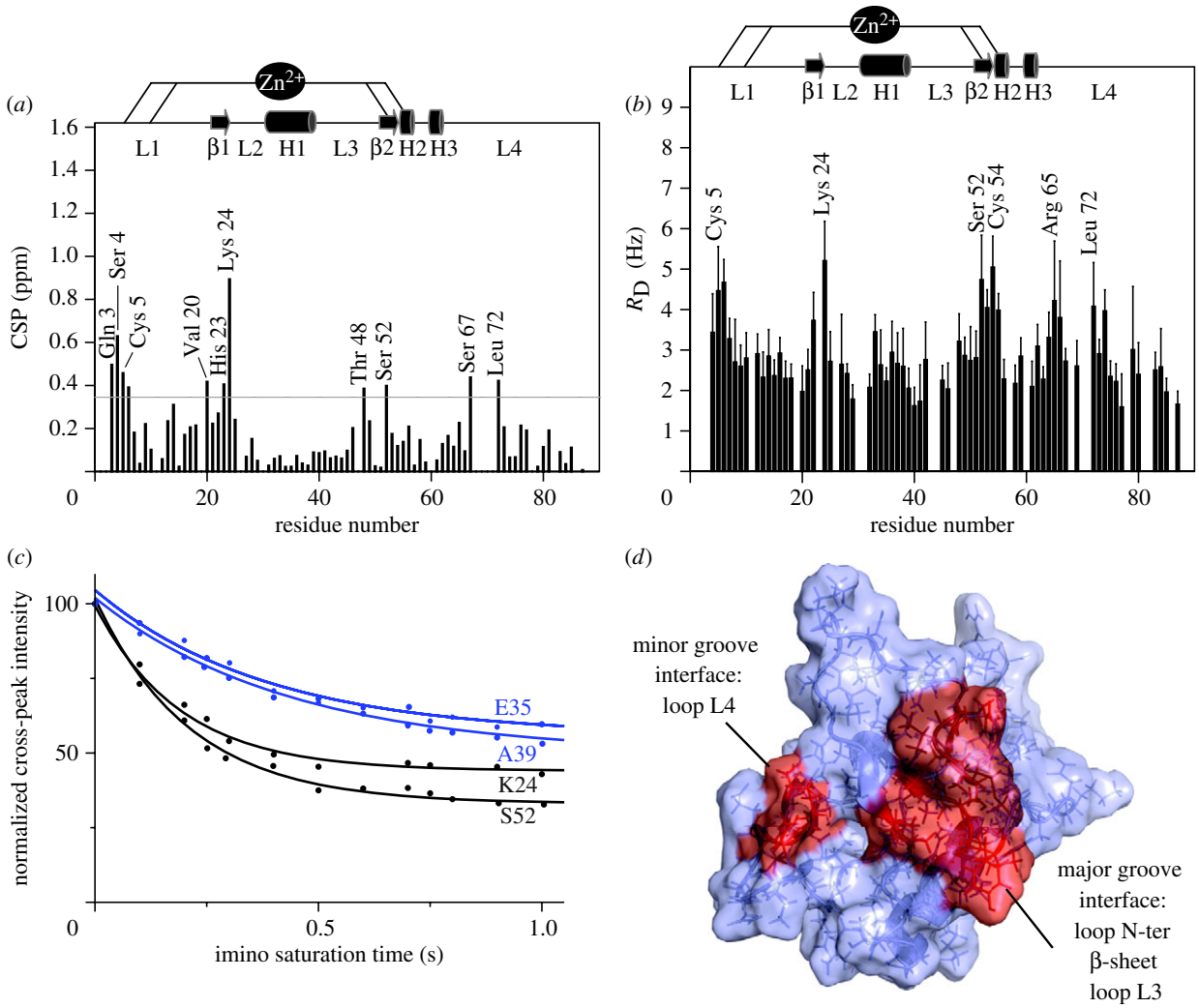


Figure 1. Interaction surface mapping by combining chemical shift perturbation (CSP) and imino cross saturation on the THAP domain of hTHAP1. (*a*) Histogram of the normalized CSP observed upon DNA binding as a function of the residue number. (*b*) Imino cross-saturation rates (R_D) as a function of residue number. (*c*) Examples of experimental points and fitted curves of the imino cross-saturation data. Experimental points and fitted curves are coloured in blue for the α-helical residues (away from DNA) and in black for β-sheet residues (close to DNA). (*d*) Mapping of the interaction surface on the solution structure of the THAP domain of hTHAP1.

It is noteworthy that the DNA-binding surface has to be located in a continuous protein interface area. In the case of the transcription factor Myb1, which contains two Myb-like DNA-binding motives, CSP experiments highlighted residues with large chemical shift changes that are located in several discontinuous faces of the free protein. The structure of the protein in DNA-bound conformation revealed conformational changes upon DNA binding, leading to a structural rearrangement of the domains that constitute a continuous interaction surface in the complex, in agreement with CSP data [24].

In a complementary approach, it is possible to record ^{15}N HSQC spectra on well-resolved DNA imino protons to assess local deformation of DNA base pairs as well as DNA bending and larger conformational transitions. In the latter case, it is common to observe significant chemical shifts extending to regions that are not part of the interface. Specific pulse programmes to detect nucleic acid imino protons have been recently developed [25]. However, the interpretation of CSPs in DNA is more difficult with respect to the protein owing to conformational dynamic behaviour found in DNA and imino proton exchange leading to severe line broadening. CSPs of imino protons have been used as a tool to identify regions in the DNA that undergo large conformational changes upon interaction with the protein. A recent interesting application of this approach was used to analyse the interaction of the rice telomere-binding protein, RTBP1, with its target DNA [26].

It is noteworthy that specific DNA-binding sites are often characterized using molecular biology methods such as *S*ystematic *E*volution of *L*igands by *EX*ponential enrichment (SELEX) and mutagenesis coupled to DNA-binding assays [27].

3.2. Imino protons cross saturation

Cross-saturation experiments were first introduced to identify the interfaces of large protein–protein complexes [28] and have been successfully applied to mapping out the binding surfaces in protein–nucleic acid complexes [29,30]. This method is based on

selective saturation of protons from one molecule partner that leads to a decrease in magnetization throughout the molecule, which is mostly controlled by spin diffusion. The saturation is transferred to the other partner, leading to a decreased NMR intensity. In the case of DNA–protein complexes, DNA imino proton resonances (of thymine H3 and guanine H1) are selectively saturated using a loop including an adiabatic band-selective inversion pulse centred in the imino frequency range (δ^1H, 12–15 ppm) [31]. During DNA saturation, imino proton relaxation processes perturb the thermodynamic equilibrium of neighbouring spins resulting in the transfer of saturation to the protein. By introducing the saturation period prior to a classical ^{15}N HSQC pulse sequence, imino cross-saturation experiments are recorded using several saturation times (ranging from 0.1 to 2 s). Then, the normalized intensity of the HSQC cross peaks is plotted as a function of the saturation time (figure 1*c*).

This method was recently used, in addition to chemical shift mapping experiments, to define the specific DNA-binding interface of the THAP domain of THAP1 [20]. By fitting experimental data with a mono-exponential equation (figure 1*c*), a characteristic decreasing time (T_D) was extracted for each non-proline residue and the R_D parameter ($1/T_D$) was plotted as a function of the protein sequence (figure 1*b*). Since spin diffusion is controlled by dipolar interaction (for small saturation times), the R_D parameter assesses DNA spatial proximity. As shown in figure 1, cross-saturation data are in good qualitative agreement with the CSP data, providing reliable identification of the DNA-binding interface (figure 1*d*).

3.3. Solvent accessibility

Amide hydrogen exchange experiments are commonly used to identify changes in solvent accessibility of interfacial residues and to provide additional evidence of the DNA–protein interface. Quantitative analysis of amide proton exchange rates for the free protein and the DNA-bound protein has been applied to provide information on the binding interface and also on protein folding upon DNA recognition [12].

An alternative approach is to use water-soluble relaxation agents to increase the relaxation process of the solvent-exposed atoms. Indeed, DNA binding will protect the atoms that are located at the binding interface, leading to different relaxation behaviours between the free and the bound forms [32].

All these techniques that allow interface mapping provide important information concerning the recognition mode and open the door to modelling approaches. Sometimes, when dealing with a family of homologous proteins for which a structure is already available, molecular modelling coupled with mutagenesis data can be enough to understand the mechanism of DNA recognition. But, when the structure of a homologous protein is not available, interface mapping techniques are not sufficient to understand the DNA recognition process and additional structural restraints need to be collected in order to get an accurate description of the DNA recognition process.

4. INTERMOLECULAR RESTRAINTS

4.1. Short-range intermolecular distance restraints: nuclear Overhauser effect

Spatial proximity of residues close to the protein–DNA interface can lead to an intermolecular dipole–dipole interaction sufficiently large for creating measurable NOEs. The resulting cross relaxation together with the intermolecular NOEs provide direct evidence of their spatial proximity. Obviously, resonances of both partners in the complex must be assigned prior to identification of the intermolecular NOEs. However, strong resonance overlapping and line broadening associated with the size of the complex and/or chemical exchange make it very difficult to unambiguously assign NOEs. Recent progress in different NMR processes, such as labelling strategies, pulse sequence developments and the availability of high magnetic fields, has largely contributed to solve this problem.

In most studies of DNA–protein complexes, the protein is uniformly ^{13}C/^{15}N labelled while the DNA remains unlabelled. Consequently, proton frequencies of the protein in the complex are generally assigned using a combination of classical triple-resonance experiments (HNCA, HNCACB) and TOCSY and nuclear Overhauser spectroscopy (NOESY) experiments (three-dimensional ^{15}N HSQC-NOESY, three-dimensional ^{15}N HSQC-TOCSY, three-dimensional HCCH-TOCSY, three-dimensional ^{13}C HSQC-NOESY). The development of isotope filtering and editing schemes has largely contributed to reducing the tedious task of DNA assignment. In particular, it is possible to select the ^{1}H atoms that are or are not one-bond scalar linked to a specific heteroatom, ^{15}N or ^{13}C. General considerations and experimental details on isotope-filtered NMR methods have been well reviewed by Breeze [33].

For instance, in the NMR studies of the complex formed by calmodulin and a peptide, peptide ^{1}H frequencies and intermolecular NOEs were assigned using a combination of two-dimensional NMR experiments [34,35]. This strategy has also been used for protein–DNA complexes ([18,36–38]; [39, p. 423]). Firstly, two-dimensional [F_1, F_2] ^{13}C-filtered NOESY experiments were recorded for the protein–DNA complex to observe NOEs between the ^{1}H atoms that are one-bond scalar coupled to ^{12}C spins, i.e. protons of the unlabelled DNA while the protein is uniformly ^{13}C- and ^{15}N-labelled. The two-dimensional [F_1, F_2] ^{13}C-filtered NOESY was first proposed using a filter based on the chemical shift-optimized adiabatic ^{13}C inversion pulse [40] and was then optimized using $^1J_{HC}$ coupling constant properties [15], providing a sensitivity gain of up to 40 per cent in the most favourable case. Once DNA proton frequencies are assigned, intermolecular NOEs can be detected using two-dimensional [F_2] ^{13}C-filtered NOESY experiments that provide correlations owing to cross relaxation between all protons (bound to DNA ^{12}C atoms and to protein ^{13}C atoms). The resulting spectrum corresponds to two-dimensional [F_1, F_2] ^{13}C-filtered NOESY plus intermolecular NOEs.

If resonance overlap persists, filtering and editing schemes in three-dimensional experiments may be

combined to increase resolution. Using three-dimensional [F_1] ^{13}C filtered–[F_3] ^{13}C edited–NOESY-HSQC experiments [40], it has been possible to detect intermolecular NOE positions on a three-dimensional experiment. Briefly, all the protons are excited but for ^{13}C-linked protons, then the mixing time permits DNA protons to perform cross relaxation with surrounding protons; finally, magnetization is transferred to ^{13}C atoms and back to ^{1}H before detection.

Obtaining intermolecular distance restraints remains difficult and time-consuming, but they permit the interface of protein–DNA complexes to be defined with high accuracy and precision to enable structure calculations. Gathering a significant number of intermolecular NOEs is certainly the most precise and efficient way to define the interface with high accuracy.

4.2. Angular intermolecular restraints: residual dipolar couplings

Structural determination of DNA using the NOE approach is more difficult than it is for proteins and has, for a long time, been severely hampered by a lack of precision and accuracy in the structures. The low proton density together with the elongated shape of the DNA molecule often lead to a small number of NOE restraints. The majority of observable ^{1}H–^{1}H contacts are between directly adjoining base pairs in the DNA sequence, leading to a lack of long-range structural restraints. Therefore, although the local structure is relatively well defined, the global NOE-derived structure of DNA suffers from a lack of long-range distances [41]. The introduction of additional orientational restraints using a small degree of alignment [42] turned out to be useful for DNA, improving both local and global structures [17,43]. The first example of a really extensive use of residual dipolar couplings (RDCs) for a protein–DNA structure determination was performed by Murphy and co-workers [17], who measured and included, in the structure calculation protocol, 274 protein RDCs and 46 DNA RDCs on the HMG-box domain of the human male sex-determining factor SRY, hSRYHMG, bound to a DNA 14-mer.

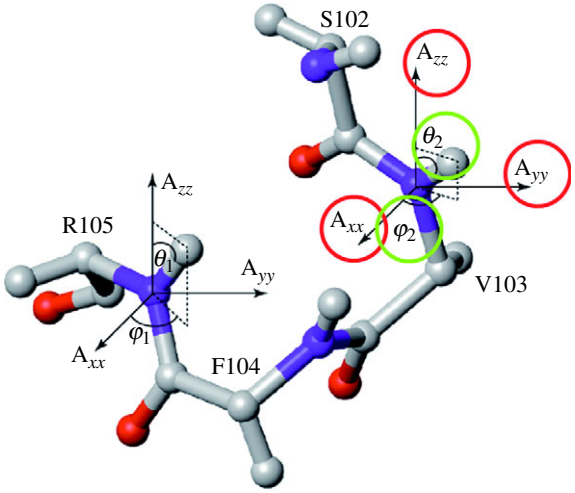
In solution, the dipolar coupling between two nuclei is averaged to zero because of the isotropic Brownian motion of molecules, unless the movement is restricted to a preferred direction relative to the magnetic field.

DNA molecules are naturally sensitive to an external magnetic field and are aligned partially with respect to the high magnetic field. In order to determine the orientation of the GATA-1 protein on its 16-base pair oligonucleotide, $^{1}D_{HN/N}$ and $^{1}D_{H\alpha/C\alpha}$ RDCs were collected on the protein backbone inside the protein–DNA complex at several external magnetic fields without adding liquid crystals [44]. The magnetic susceptibility tensor for the complex is dominated by contributions of the bases in the DNA duplex, yielding an axially symmetric susceptibility tensor with the main axis approximately parallel to the double helix. Using this information, a method for refining the solution structure of this protein–DNA complex using RDC was proposed. This approach, however, had to rely on very precise measurements of dipolar couplings of a fraction of a hertz, owing to a molecular order parameter of these complexes of about 10^{-4} or less.

A stronger partial alignment of molecules can be achieved by using dilute aqueous liquid crystalline media (phages Pf1, bicelles, PEG/alcohol, etc.). Thus, weak alignment (here weak alignment means molecular order parameter S of the order of 10^{-3}) of the molecule will lead to the loss of its purely isotropic motion, resulting in partial dipolar coupling between spins while retaining conditions of high-resolution solution NMR [42]. Pulse programmes have been specifically developed to measure coupling constants, allowing scalar couplings to be assessed under isotropic conditions and apparent scalar couplings (i.e. scalar coupling plus residual dipolar coupling) under anisotropic conditions. Several dipolar couplings can be extracted on the protein backbone bonds, such as $^{1}D_{HN/N}$, $^{1}D_{H\alpha/C\alpha}$, $^{1}D_{C\alpha/C'}$ and $^{1}D_{N/C'}$. Several methods have been proposed to determine the alignment tensor [45], allowing the transformation of RDC values into orientational restraints of the covalent bonds with respect to the alignment tensor (figure 2).

We have seen the importance of introducing RDCs in DNA structure determination to remedy the limited number of classical NOE restraints and to increase the accuracy of global DNA structures. The utility of employing RDCs was also established in the accurate determination of subtle structural features such as the DNA helix curvature, as in oligonucleotides containing A*n*T*n* segments [46]. Over the last decade, several pulse programmes have been developed for the measurement of residual dipolar couplings on DNA molecules [47–49]. However, in the simplest manner, RDCs can be extracted using two-dimensional IPAP ^{15}N/HN HSQC centred on the amide proton regions (for $^{1}D_{HN/N}$ RDCs). Once the alignment tensor has been determined, RDCs can be transformed into orientation restraints of covalent bonds in the molecular frame using the equations given in figure 2. Care should be taken to remove from the analysis the mobile loops (which do not follow such simple equations because of dipolar coupling averaging from internal motions), which can be clearly identified from ^{15}N relaxation analysis.

In the context of protein–DNA complexes with both labelled partners, RDCs can be measured for both molecules within the same sample. Alignment tensors are extracted for each molecule and, if they are identical, RDCs could be used differently. In fact, RDC restraints are transformed into purely intermolecular projection angle restraints [50]. This approach was recently developed and has been successfully applied to define protein–protein interactions [51] and the relative orientation of protein subdomains [52], and may be used for protein–DNA interactions. By collecting RDCs on each partner, it is possible to derive intermolecular projection angle restraints useful for orientating each partner with respect to the other. It represents a new class of intermolecular restraints that could be included in structure calculations of protein–DNA complexes. In particular, combination of these restraints with shape restraints derived from small-angle X-ray scattering (SAXS) represents a very powerful approach [53].



$$D_{IS}(\theta,\varphi) = \frac{-3\,\mu_0 h\,\gamma_I\,\gamma_S}{16\,\pi^3\,r_{IS}^3}\left(\frac{1}{2}\,A_{zz}\,(3\cos^2\theta - 1) + \frac{1}{2}\,(A_{xx} - A_{yy})\sin^2\theta\cos 2\varphi\right),$$

where

$D_{IS}\ (\theta, \varphi)$, residual dipolar coupling between *I* and *S* spins;
μ_0, vaccum permeability;
h, Planck constant;
γ_I, gyromagnetic ratio of spin *I*;
γ_S, gyromagnetic ratio of spin *S*;
r_{IS}, internuclear distance between *I* and S spins;
θ, Φ, polar angles defining the orientation of the IS bond in respect to alignment tensor;
A_{zz}, A_{xx}, A_{yy} are the three components of the alignment tensor.

This equation is usually rewritten as

$$D_{IS}\ (\theta,\varphi) = D_a((3\cos^2\theta - 1) + {}^3/_2\ R\sin^2\theta\cos^2 2\varphi),$$

where

$$D_a = -\left(\frac{3\,\mu_0\,h}{32\,\pi^3}\right)\gamma_I\,\gamma_S\,\langle r_{IS}^{-3}\rangle\cdot A_{zz}$$

$$R = {}^2/_3\,(A_{xx} - A_{yy})/A_{zz}$$

D_a corresponds to the magnitude of the dipolar coupling tensor and R to the rhombicity.

Figure 2. The residual dipolar coupling formalism, illustration and equations.

4.3. Long-range intermolecular distance restraints: paramagnetic relaxation enhancement

Magnetic dipolar interaction between a spin and an unpaired electron of a paramagnetic centre can result in chemical shift changes or in an increase in the relaxation rate of the nuclear magnetization, a phenomenon called PRE. Considering the distance r between the nucleus of interest and the paramagnetic centre, the magnitude of the PRE is proportional to r^{-6}. However, in contrast to the NOE effect that is limited to short distances (typically less than 5 Å, slightly more with fully deuterated proteins), the PRE effect can be observed up to 35 Å, depending on the magnetic moment of the paramagnetic centre. Therefore, paramagnetic NMR represents a valuable source of long-range distances that can complement NOE restraints for a complex in the slow-exchange regime. When the molecule does not contain an intrinsic paramagnetic centre, extrinsic spin labels need to be judiciously incorporated to the molecule of interest, usually by designing nitroxide spin labels and cysteine point mutations. In the case of protein–DNA complexes, a paramagnetic agent is commonly conjugated to the DNA. For instance, the use of DNA-containing EDTA-derivatized deoxythymidine chelated to Mn^{2+} in the SRY/DNA complex allowed the intermolecular PRE effects to be measured and the polarity of the SRY binding on the DNA to be defined [54]. More recently, the incorporation of deoxy-4-thiouracil into each end of the oligonucleotide reacting with 3-(2-iodoacetamido)-proxyl enabled label both 5′ DNA extremities to be spin labelled [55], and a protocol was described to incorporate nitroxide spin labels into specific 2′-sites within nucleic acids [56].

The PRE effect induced by the spin labelling leads to line broadening, the magnitude of which depends on the distance between the nucleus and the spin label. By combining different labelling sites, PRE data are extracted for different positions of the paramagnetic

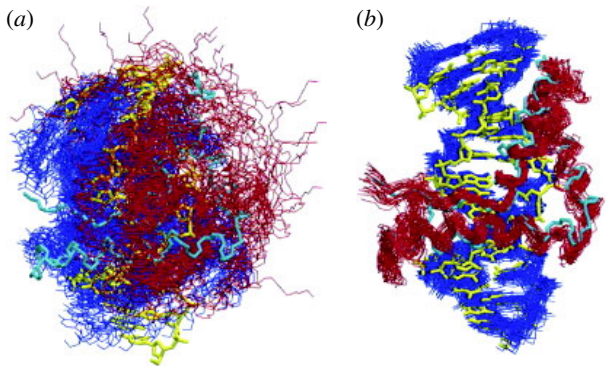


Figure 3. Long-range distance restraints from PRE, adapted from Iwahara *et al.* [58]. Impact of intermolecular PRE data on coordinate accuracy of the SRY–DNA complex when only a single intermolecular NOE restraint located at the centre of the protein–DNA interface is employed. Best-fit superposition of 30 simulated annealing structures (SRY, red; DNA, blue) calculated (*a*) without and (*b*) with 438 intermolecular ^{1}H-PRE restraints.

centre, allowing long-range distance restraint to be determined [57]. This strategy was used to compute solution structures of the SRY–DNA complex [58] (figure 3) and the Mrf2–DNA complex [55].

5. STRUCTURAL RESTRAINTS FOR DNA STRUCTURE CALCULATION

The determination of DNA structural restraints requires prior assignment of DNA resonances following standard procedure [59]. Briefly, sequential assignment of non-labile protons in B-DNA is accomplished by following short ^{1}H–^{1}H distance connectivities between the bases (purine H8/pyrimidine H6) and the H2′/H2″ and H1′ sugar protons in NOESY spectra recorded in $^{2}H_2O$. The intraresidue cross peaks may be independently confirmed by DQF-COSY. In B-DNA, connectivities are also observed between the base (purine H8/pyrimidine H6) and C H5 or T CH3 protons in the direction of 5′ to 3′. At the same time, the A H2 protons are identified by their weak inter- and intrastrand connectivities with H1′ sugar protons. Assignment of the H3′, H4′, H5′ and H5″ sugar protons is also based on the NOESY spectra following standard procedures. Assignment of the labile protons requires that a NOESY spectrum is recorded in H_2O buffer and at low temperature to ensure slowing down of proton exchange with the solvent. The main assignment pathway relies on detectable ^{1}H–^{1}H NOEs between the hydrogen-bonded imino protons of adjacent stacked base pairs [59].

The DNA double helix is a highly flexible structure that can adopt different classes of conformations, B, A and Z, and DNA local structural perturbations are often observed upon protein binding as well as larger conformational transitions. One should be particularly cautious in the description of the experimental restraints used in the structure calculation protocol as accurate DNA structure calculation needs to incorporate observables based on a combination of dedicated experiments. The resulting structure may range from a simple DNA classification (here, we are not talking about mere modelling in which the force field could play a major role) up to a precise description of the structure, including sugar puckering, base pair stacking and proper base pair parameters. A number of dedicated NMR experiments have been described, many of which are based on uniform ^{15}N–^{13}C labelling of the DNA. The sugar conformation can be estimated from the analysis of proton intraresidual NOEs within sugar rings in a NOESY spectrum recorded with short mixing time and, in particular, from the ratio of the cross-peak volumes for the intraresidue $H_{8/6}$–$H_{2'}$ and $H_{8/6}$–$H_{3'}$ interactions, which is higher than 1 in the presence of a predominant C2′ *endo* conformation [60]. Similarly, the orientation of the base relative to the deoxyribose ring (*syn* or *anti* conformation) can be determined from the ratio of $H_{8/6}$–$H_{2'}$ to $H_{8/6}$–$H_{1'}$, which should be higher than 1 in the B-DNA form. Information on both the sugar pucker and base torsion angles is needed to estimate the DNA conformation [59]. These measurements have been done in the NMR study of the THAP zinc finger in complex with its DNA target, and have confirmed that the DNA target adopts a B-DNA form in the complex (figure 4). Alternatively, the type and degree of DNA sugar puckering in protein–DNA complexes can be assessed by recording a NOESY sequence preceded by a constant-time scalar coupling period [61]. Notably, other methods have been developed to study sugar puckering of high molecular weight oligonucleotides based on the measurement of deoxyribose $^{3}J_{HH}$ scalar couplings [62] or cross-correlated relaxation rates involving ^{13}C chemical shift anisotropy (CSA) and ^{13}C–^{1}H dipolar interactions [63], making use of uniform ^{13}C labelling of the DNA. Alternatively, structural information and DNA conformational changes induced by protein binding can be provided by means of ^{31}P NMR [64,65]. It is noteworthy that ^{31}P NMR can be used to identify B_I/B_{II} DNA conformational transitions, which are likely to play a role in sequence-specific recognition [66,67].

Furthermore, without major conformational change, the DNA double helix has to accommodate the protein-binding surface by modifying base pair step parameters (i.e. twist, tilt, roll, shift, slide and rise [68,69]), which may result in DNA bending [70]. Assessing the global bending magnitude in DNA–protein complexes is of particular interest, but classical NMR methods are inefficient in capturing DNA bending. However, the use of RDCs proved their ability to give a reliable view of bending in intrinsically curved DNA [46] or changes in DNA bending induced by protein binding.

To experimentally estimate protein interaction with the DNA and DNA bending, electrophoretic mobility experiments [71,72] have been used based on the observation that bent DNA migrates more slowly in a gel matrix than straight DNA [73]. An alternative method called the circular permutation DNA-bending assay, which is based on differences in the electrophoretic mobility of protein–DNA complexes with DNA-binding sites located at different positions, may be used to provide information on protein-induced DNA bending [74,75].

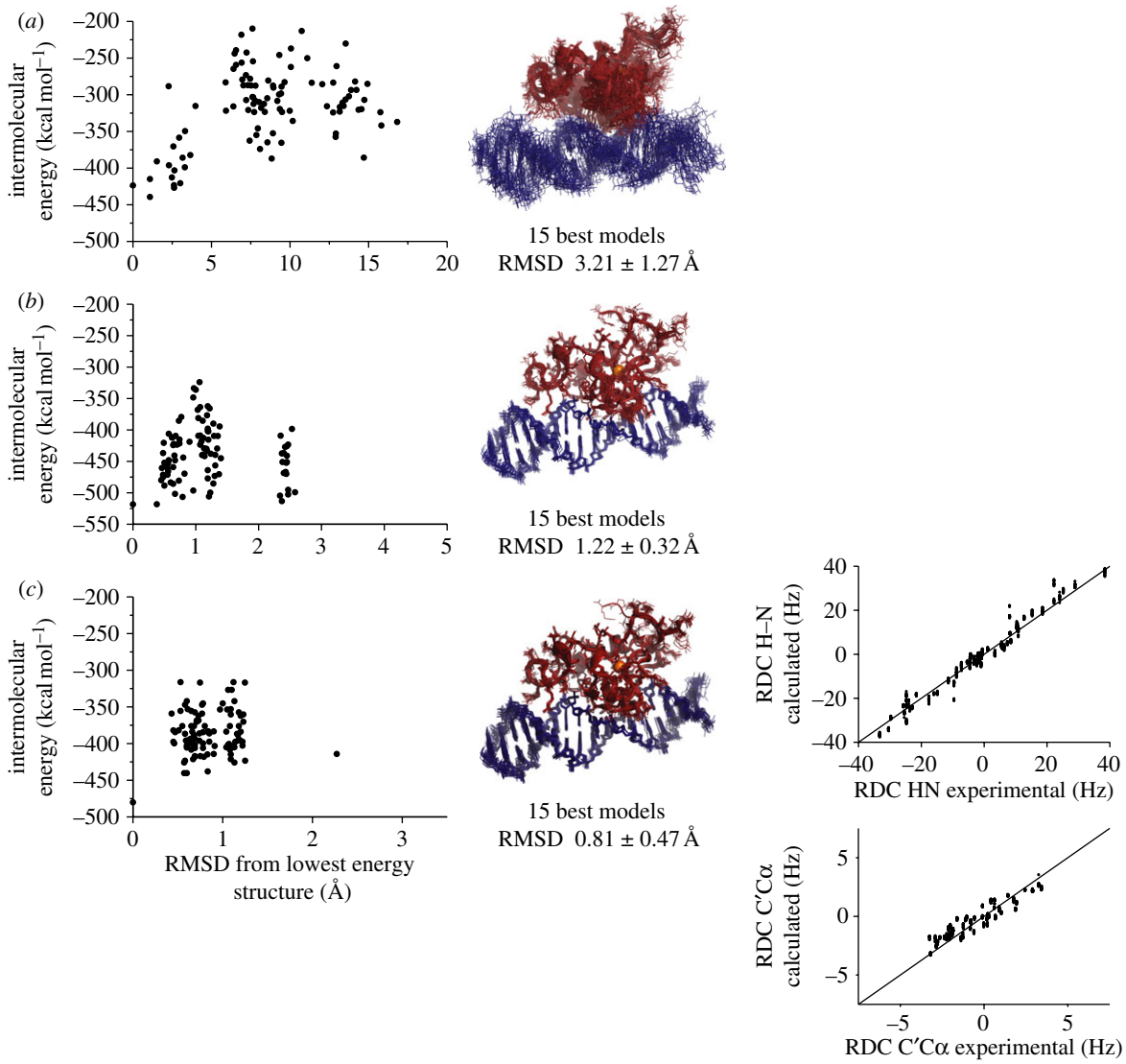


Figure 4. Computational strategies for calculation of protein–DNA complex structures using a data-driven docking protocol. (*a*) Modelling of the complex formed by the THAP domain of hTHAP1 and its DNA target, including solely interface mapping restraints as ambiguous interface restraints. On the graph, the intermolecular energy is plotted as a function of RMSD from the lowest energy structure in order to illustrate the structure calculation convergence. On the right-hand side, a superposition of the 15 best models is shown. (*b*) Modelling of the complex formed by the THAP domain of hTHAP1 and its DNA target, including interface mapping restraints as ambiguous interface restraints, 39 intermolecular NOEs and 49 $^{1}D_{HN\text{-}N}$ RDCs for refinement. In the graph, the intermolecular energy is plotted as a function of RMSD from the lowest energy structure in order to illustrate the calculation convergence. On the right-hand side, a superposition of the 15 best structures is shown (PDB ID 2ko0). (*c*) Modelling of the complex formed by the THAP domain of hTHAP1 and its DNA target, including interface mapping restraints as ambiguous interface restraints, 39 intermolecular NOEs and 49 $^{1}D_{HN\text{-}N}$ RDCs and 49 $^{1}D_{C'\text{-}C\alpha}$ RDCs for refinement. In the graph, intermolecular energy is plotted as a function of RMSD from the lowest energy structure in order to illustrate the calculation convergence. On the right-hand side, a superposition of the 15 best structures is shown and, in both graphs, a comparison between experimentally measured RDCs and backcalculated RDCs is shown for $^{1}D_{HN\text{-}N}$ RDCs and $^{1}D_{C'\text{-}C\alpha}$ RDCs. For each ensemble, pair-wise RMSD on backbone heavy atoms is shown.

In addition, several software programs have been developed to model DNA bending such as the 3DNA program, which allows analysis of DNA structural parameters and enables it to be rebuilt with customized DNA models [76]. Several Web servers have been created recently and provide interesting tools to analyse and rebuild DNA models [77,78].

6. STRUCTURE CALCULATION STRATEGIES

In most cases, several types of structural restraints (NOEs, RDCs, CSAs, PREs, low-resolution envelopes obtained from SAXS, small-angle neutron scattering (SANS) or cryoEM and so on) are combined to calculate protein–DNA structures, thus increasing precision and accuracy. Solution structures are commonly computed by simulated annealing protocols using CNS [79] or XPlor-NIH [19,36,38,80–82]. These protocols permit a complete exploration of the conformational space. Xplor-NIH includes a number of possibilities such as using a database potential of mean force describing base–base positional interactions, which has been shown to improve the structure quality, as viewed from an independent check with experimental observables (RDCs, NOEs) [83].

In some cases, the knowledge of both interaction surfaces is sufficient to orient the protein onto its DNA target. For this simpler purpose, several programs based on molecular docking have been developed. They use biophysical and/or biochemical interaction information, encoded into ambiguous interaction restraints (AIRs) [84], to drive the docking process. Provided that solution structures of the free partners are known and that no major conformational change occurs upon binding, the data-driven docking program HADDOCK can be applied to a large variety of biomolecular complexes [85,86] from protein–protein to protein–nucleic acid complexes [87]. In the context of full structure determination, and of including a wide range of experimental restraints, HADDOCK may be used as a python interface allowing a very simple and easy way to launch CNS for structure calculation of macromolecular complexes and analysis. The protocol includes three steps: a rigid body docking of the two partners based on interaction surface definition, a semi-flexible simulated annealing stage followed by a water-refinement step. The full HADDOCK program has been recently implemented into a HADDOCK Web server [88]. A specific protocol has been recently described for DNA–protein complex modelling using this approach, which highlights the importance of DNA flexibility to explore the conformational space [87,89].

The HADDOCK approach was used to gain insights into the DNA-binding mode by the THAP domain of hTHAP1. In the first step, AIRs derived from NMR chemical shift mapping and imino cross-saturation experiments were used to guide the docking, providing an ensemble of the 15 lowest energy structures with medium resolution (figure 4*a*). The structure calculation convergence is illustrated by plotting intermolecular energy as a function of the backbone RMSD from the lowest energy structure. In this case, several ensembles of solutions were generated.

This docking approach can take into account ambiguous as well as unambiguous intramolecular and intermolecular restraints such as intermolecular NOEs, RDCs as intermolecular projection angle restraints and long-range distances derived from PRE experiments [20,55]. It is noteworthy that the inclusion of a few unambiguous intermolecular data in the docking process can greatly increase the structure precision. In the case of the complex formed by the THAP domain and its DNA target, 39 intermolecular NOEs were included in the docking approach and the water-refinement step was done using $^{1}D_{HN\text{-}N}$ RDC restraints. Following this protocol, the structure calculation was highly convergent and an ensemble of 15 structures under high resolution is shown in figure 4*b*. By including additional $^{1}D_{C'\text{-}C\alpha}$ RDC restraints, an even better calculation convergence was obtained (figure 4*c*).

7. REFINED STRUCTURES USING SHAPE RESTRAINTS

Cryo-electron microscopy can provide low-resolution shape restraints, which may be used in conjunction with NMR data [90,91].

During the last decade, small-angle scattering of biomolecules in solution has gained popularity with the availability of powerful beamlines and the generation of new data analysis programmes [92–94]. There has been a growing interest in including low-resolution techniques such as SANS and SAXS to tackle challenging problems in structural biology [95–99]. X-ray scattering is generated during the interaction of light with electrons while neutron scattering is due to the interaction with the atom nuclei. The scattered intensity is a function of the macromolecule shape and of the contrast (the scattering density difference) between the macromolecule and the solvent [100]. In order to collect data, the sample ideally needs to be pure, highly homogeneous and monodisperse, without attractive or repulsive interaction among the scattering particles. The resulting one-dimensional diffusion curves contain information about the radius of gyration and the molecular mass, and the three-dimensional structural model of the shape of the particle can be extracted [101]. This information can be incorporated into structure calculation protocols in various ways: (i) for scoring the NMR-derived structures [102], (ii) directly in combination with NMR data (particularly RDCs) in order to perform a grid search using a rigid body protocol [103], or (iii) for direct refinement during NMR structure calculations [104–107]. An example of an application to DNA may be found in Schwieters & Clore [108], who have investigated by a combination of NMR and SAXS the structural and dynamic behaviour of the Dickerson DNA dodecamer. The resulting ensembles provided a detailed description of the conformational space sampled by the dodecamer in solution and of the fluctuations in helicoidal parameters, sugar puckers and B_I–B_{II} backbone transitions.

So far, very few examples have explored the potential of SANS for DNA–protein complexes. SANS is extremely sensitive to contrast variation and, by changing the solvent composition (typically by modifying the H_2O/D_2O ratio), it is possible to observe specifically one component of the macromolecular complex [101,109]. An early historical success of SANS in the area of protein–DNA complexes was the mapping of the relative topology of DNA and histone in the nucleosome [110]. A recent study by Falb and co-workers [111] has focused on the structure of the U4 RNA in solution. After NMR solution structure calculation, a SANS low-resolution model was constructed in order to validate the free U4 RNA structure. The authors concluded that the SANS data fully supported the NMR analysis and independently corroborate that the sharply bent K-turn motif (observed in the crystal structure of protein-bound U4 RNA) is not formed in solution in the absence of protein and divalent cations.

8. DYNAMIC STUDY OF PROTEIN–DNA INTERACTION USING NUCLEAR MAGNETIC RESONANCE

8.1. Backbone dynamic analysis in a protein–DNA complex

Protein and DNA motions are known to play important roles in the recognition mechanism and specificity [112,113]. In a number of cases, unfolded or partially folded protein domains may provide scaffolds for DNA

interaction, and intrinsically disordered protein domains or tails may facilitate DNA search efficiency as reviewed in Wright & Dyson [114] and Vuzman & Levy [115]. An extreme case of dynamic transition upon DNA binding has been described for the Brinker repressor, for which the protein domain Brk is unfolded and highly flexible throughout the entire backbone in the absence of DNA and folds into a helix–turn–helix motif upon DNA binding [116].

The dynamic behaviour of a protein in solution can be monitored by ^{15}N NMR relaxation studies of backbone amides. With the Lipari–Szabo formalism, the ^{15}N relaxation rates of individual amide ^{15}N nuclei (R_1, R_2, {HN–N} heteronuclear NOE) can be fitted to a generalized order parameter S^2, which reports amplitudes of ps–ns internal motions, to an internal correlation time (τ_e), to an exchange term (R_{ex}), caused by μs–ms exchange processes, and finally to a global isotropic correlation time (τ_c) [113]. In most cases, DNA-binding sites are flexible and undergo rapid μs–ms conformational fluctuations characterized by large values of R_{ex} for residues located at the target interface. This flexibility often decreases upon target binding. The structural study of the *lac* repressor system has provided keys to understanding the dynamic features of protein–DNA recognition [12,117]. The NMR solution structures of the *lac* DNA binding domain (DBD) in the free state or bound to non-specific or specific DNA targets have been solved and dynamic features of the protein in every state have been characterized. In the non-specific DNA-bound state, the Lac repressor displays a large number of sub-millisecond motions with large R_{ex} values that are observed at the DNA–protein interface, indicating a very transient and labile interface, whereas these motions are quenched upon binding of the protein to its cognate operator. These results have been confirmed using H/D exchange [12]. To conclude, the determination of the conformational exchange term for ^{15}N backbone atoms at the protein–DNA interface gives access to the interaction strength and stability of the DNA–protein complex.

8.2. Search mechanism of the DNA target site

Recognition and binding of specific sites on DNA by proteins are central for many cellular functions such as transcription, replication and recombination, as recently reviewed [118–120]. The proteins have to find their DNA target sites in the bulk of non-specific sequences present in the cell. Yet, it has been widely demonstrated that rate constants for association of proteins with their specific DNA targets can surpass the diffusion limit by almost two orders of magnitude. This observation can be explained by a kinetically optimized pathway involving a combination of three-dimensional diffusion and one-dimensional sliding diffusion along the DNA.

The one-dimensional diffusion rates of *Lac*I repressor proteins along elongated DNA have been measured using single-molecule imaging techniques [121,122]. This implies the existence of transient intermediates involving non-specific binding modes. PRE experiments allow the characterization of these transient intermediates and have been used to show diffusion of the DNA-binding domain along the non-specific DNA aiming to find its specific target [14]. Using different paramagnetic labelling positions on an oligonucleotide containing the HOXD9 homeodomain target site [15], PRE values could be extracted in various ionic strength conditions. At low ionic strength, PRE data are consistent with the previously described solution structure of the specific protein–DNA complex. However, at high ionic strength, several significant differences appear in the PRE data which indicate the existence of other modes of binding. This work provides for the first time NMR evidence of the one-dimensional diffusion along the DNA helix towards the recognition site. Interestingly, the PRE experiments give access to the dynamic of the DNA–protein interaction. Intermolecular jumping from one double-stranded DNA to another has also been observed using this method [14]. Theoretical studies and other evidence highlight the role of protein and DNA conformational flexibility on the rates of one-dimensional diffusion and hopping [123,124].

9. CONCLUSIONS AND PERSPECTIVES

Solution-state NMR developments have generated new methods to collect structural restraints on protein–DNA complexes. Mapping the interaction surfaces, using chemical shift mapping or cross-saturation experiments, gives answers with the lowest resolution and helps with modelling the complex. If detailed residue-specific knowledge is required, intermolecular restraints are needed. Various kinds of data can be extracted with different levels of difficulty. Intermolecular NOEs appear to be the most precise for interface definition but are time-consuming, requiring residue and atom-specific NOE assignments. Alternatively, intermolecular restraints can be represented by angular restraints with RDCs as intervector projection angles or by long-range distances with PRE. These approaches can be combined in order to increase structural accuracy. In some cases, combination of RDC and SAXS low-resolution shape restraints has been described as a powerful approach. NMR is limited by the size of the macromolecular edifice but offers useful possibilities in terms of dynamic characterization. During the last decade, dynamic descriptions of protein–DNA interactions have played an increasing role in understanding recognition specificity and kinetics.

REFERENCES

1 Ansari, A. Z. & Mapp, A. K. 2002 Modular design of artificial transcription factors. *Curr. Opin. Chem. Biol.* **6**, 765–772. (doi:10.1016/S1367-5931(02)00377-0)
2 Garvie, C. W. & Wolberger, C. 2001 Recognition of specific DNA sequences. *Mol. Cell* **8**, 937–946. (doi:10.1016/S1097-2765(01)00392-6)
3 Chen, L. 1999 Combinatorial gene regulation by eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **9**, 48–55. (doi:10.1016/S0959-440X(99)80007-4)
4 Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. 2000 An overview of the structures of protein–DNA complexes. *Genome Biol.* **1**, REVIEWS001. (doi:10.1186/gb-2000-1-1-reviews001)

5 Panne, D., Maniatis, T. & Harrison, S. C. 2004 Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-beta enhancer. *EMBO J.* **23**, 4384–4393. (doi:10.1038/sj.emboj.7600453)

6 Panne, D., Maniatis, T. & Harrison, S. C. 2007 An atomic model of the interferon-beta enhanceosome. *Cell* **129**, 1111–1123. (doi:10.1016/j.cell.2007.05.019)

7 Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. 1997 Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA* **94**, 12 366–12 371. (doi:10.1073/pnas.94.23.12366)

8 Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. 1998 TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95**, 13 585–13 590. (doi:10.1073/pnas.95.23.13585)

9 Kaptein, R., Slijper, M. & Boelens, R. 1995 Structure and dynamics of the lac repressor-operator complex as determined by NMR. *Toxicol. Lett.* **82–83**, 591–599. (doi:10.1016/0378-4274(95)03586-9)

10 Kalodimos, C. G., Folkers, G. E., Boelens, R. & Kaptein, R. 2001 Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. *Proc. Natl Acad. Sci. USA* **98**, 6039–6044. (doi:10.1073/pnas.101129898)

11 Kalodimos, C. G., Boelens, R. & Kaptein, R. 2002 A residue-specific view of the association and dissociation pathway in protein–DNA recognition. *Nat. Struct. Biol.* **9**, 193–197. (doi:10.1038/nsb763)

12 Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R. & Kaptein, R. 2004 Structure and flexibility adaptation in nonspecific and specific protein–DNA complexes. *Science* **305**, 386–389. (doi:10.1126/science.1097064)

13 Romanuka, J., Folkers, G. E., Biris, N., Tishchenko, E., Wienk, H., Bonvin, A. M., Kaptein, R. & Boelens, R. 2009 Specificity and affinity of Lac repressor for the auxiliary operators O_2 and O_3 are explained by the structures of their protein–DNA complexes. *J. Mol. Biol.* **390**, 478–489. (doi:10.1016/j.jmb.2009.05.022)

14 Iwahara, J. & Clore, G. M. 2006 Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227–1230. (doi:10.1038/nature04673)

15 Iwahara, J., Zweckstetter, M. & Clore, G. M. 2006 NMR structural and kinetic characterization of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl Acad. Sci. USA* **103**, 15 062–15 067. (doi:10.1073/pnas.0605868103)

16 Jamin, N. & Toma, F. 2001 NMR studies of protein–DNA interactions. *Prog. Nucl. Magn. Res. Sp.* **38**, 83–114. (doi:10.1016/S0079-6565(00)00024-8)

17 Murphy, E. C., Zhurkin, V. B., Louis, J. M., Cornilescu, G. & Clore, G. M. 2001 Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J. Mol. Biol.* **312**, 481–499. (doi:10.1006/jmbi.2001.4977)

18 Iwahara, J., Wojciak, J. M. & Clubb, R. T. 2001 Improved NMR spectra of a protein–DNA complex through rational mutagenesis and the application of a sensitivity optimized isotope-filtered NOESY experiment. *J. Biomol. NMR* **19**, 231–241. (doi:10.1023/A:1011296112710)

19 Wojciak, J. M., Connolly, K. M. & Clubb, R. T. 1999 NMR structure of the Tn916 integrase-DNA complex. *Nat. Struct. Biol.* **6**, 366–373. (doi:10.1038/7603)

20 Campagne, S., Saurel, O., Gervais, V. & Milon, A. 2010 Structural determinants of specific DNA-recognition by the THAP zinc finger. *Nucl. Acids Res.* **38**, 3466–3476. (doi:10.1093/nar/gkq053)

21 Yamasaki, K. *et al.* 2004 Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *Plant Cell* **16**, 3448–3459. (doi:10.1105/tpc.104.026112)

22 Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S. & Matsuo, H. 2008 Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119. (doi:10.1038/nature06638)

23 Veprintsev, D. B., Freund, S. M., Andreeva, A., Rutledge, S. E., Tidow, H., Canadillas, J. M., Blair, C. M. & Fersht, A. R. 2006 Core domain interactions in full-length p53 in solution. *Proc. Natl Acad. Sci. USA* **103**, 2115–2119. (doi:10.1073/pnas.0511130103)

24 Lou, Y. C., Wei, S. Y., Rajasekaran, M., Chou, C. C., Hsu, H. M., Tai, J. H. & Chen, C. 2009 NMR structural analysis of DNA recognition by a novel Myb1 DNA-binding domain in the protozoan parasite *Trichomonas vaginalis*. *Nucl. Acids Res.* **37**, 2381–2394. (doi:10.1093/nar/gkp097)

25 Farjon, J., Boisbouvier, J., Schanda, P., Pardi, A., Simorre, J. P. & Brutscher, B. 2009 Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *J. Am. Chem. Soc.* **131**, 8571–8577. (doi:10.1021/ja901633y)

26 Ko, S., Yu, E. Y., Shin, J., Yoo, H. H., Tanaka, T., Kim, W. T., Cho, H. S., Lee, W. & Chung, I. K. 2009 Solution structure of the DNA binding domain of rice telomere binding protein RTBP1. *Biochemistry* **48**, 827–838. (doi:10.1021/bi801270g)

27 Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F. & Girard, J. P. 2005 The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. *Proc. Natl Acad. Sci. USA* **102**, 6907–6912. (doi:10.1073/pnas.0406882102)

28 Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. 2000 A novel NMR method for determining the interfaces of large protein–protein complexes. *Nat. Struct. Biol.* **7**, 220–223. (doi:10.1038/73331)

29 Ramos, A., Kelly, G., Hollingworth, D., Pastore, A. & Frenkiel, T. 2000 Mapping the interfaces of protein–nucleic acid complexes using cross-saturation. *J. Am. Chem. Soc.* **122**, 11 311–11 314. (doi:10.1021/ja002233w)

30 Lane, A. N., Kelly, G., Ramos, A. & Frenkiel, T. A. 2001 Determining binding sites in protein–nucleic acid complexes by cross-saturation. *J. Biomol. NMR* **21**, 127–139. (doi:10.1023/A:1012486527215)

31 Allen, M. D., Grummitt, C. G., Hilcenko, C., Min, S. Y., Tonkin, L. M., Johnson, C. M., Freund, S. M., Bycroft, M. & Warren, A. J. 2006 Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J.* **25**, 4503–4512. (doi:10.1038/sj.emboj.7601340)

32 Moriya, J., Sakakura, M., Tokunaga, Y., Prosser, R. S. & Shimada, I. 2009 An NMR method for the determination of protein binding interfaces using TEMPOL-induced chemical shift perturbations. *Biochim. Biophys. Acta Gen. Subjects* **1790**, 1368–1376. (doi:10.1016/j.bbagen.2009.06.001)

33 Breeze, A. L. 2000 Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Magn. Reson. Sp.* **36**, 323–372. (doi:10.1016/S0079-6565(00)00020-0)

34 Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B. & Bax, A. 1992 Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632–638. (doi:10.1126/science.1585175)

35 Ikura, M. & Bax, A. 1992 Isotope-filtered 2d NMR of a protein peptide complex—study of a skeletal-muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.* **114**, 2433–2440. (doi:10.1021/ja00033a019)

36 Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. & Gronenborn, A. M. 1993 NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of Gata-1. *Science* **261**, 438–446. (doi:10.1126/science.8332909)

37 Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. & Wuthrich, K. 1993 Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. Mol. Biol.* **234**, 1084–1093. (doi:10.1006/jmbi.1993.1661)

38 Fadeev, E. A., Sam, M. D. & Clubb, R. T. 2009 NMR structure of the amino-terminal domain of the lambda integrase protein in complex with DNA: immobilization of a flexible tail facilitates beta-sheet recognition of the major groove. *J. Mol. Biol.* **388**, 682–690. (doi:10.1016/j.jmb.2009.03.041)

39 Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. 1995 Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795. (doi:10.1038/376791a0)

40 Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J., Konrat, R. & Kay, L. E. 1997 Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage lambda N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119**, 6711–6721. (doi:10.1021/ja970224q)

41 Lipsitz, R. S. & Tjandra, N. 2004 Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 387–413. (doi:10.1146/annurev.biophys.33.110502.140306)

42 Tjandra, N. & Bax, A. 1997 Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114. (doi:10.1126/science.278.5340.1111)

43 MacDonald, D. & Lu, P. 2002 Residual dipolar couplings in nucleic acid structure determination. *Curr. Opin. Struct. Biol.* **12**, 337–343. (doi:10.1016/S0959-440X(02)00328-7)

44 Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. 1997 Use of dipolar 1H–15N and 1H–13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Biol.* **4**, 732–738. (doi:10.1038/nsb0997-732)

45 Bryce, D. L. & Bax, A. 2004 Erratum: Application of correlated residual dipolar couplings to the determination of the molecular alignment tensor magnitude of oriented proteins and nucleic acids (vol. 28, p. 273, 2004). *J. Biomol. NMR* **29**, 219. (doi:10.1023/B:JNMR.0000019275.08261.18)

46 McAteer, K., Aceves-Gaona, A., Michalczyk, R., Buchko, G. W., Isern, N. G., Silks, L. A., Miller, J. H. & Kennedy, M. A. 2004 Compensating bends in a 16-base-pair DNA oligomer containing a T(3)A(3) segment: a NMR study of global DNA curvature. *Biopolymers* **75**, 497–511. (doi:10.1002/bip.20168)

47 Jaroniec, C. P., Boisbouvier, J., Tworowska, I., Nikonowicz, E. P. & Bax, A. 2005 Accurate measurement of N-15-C-13 residual dipolar couplings in nucleic acids. *J. Biomol. NMR* **31**, 231–241. (doi:10.1007/s10858-005-0646-2)

48 Boisbouvier, J., Bryce, D. L., O'Neil-Cabello, E., Nikonowicz, E. P. & Bax, A. 2004 Resolution-optimized NMR measurement of D-1(CH), D-1(CH) and D-2(CH) residual dipolar couplings in nucleic acid bases. *J. Biomol. NMR* **30**, 287–301. (doi:10.1007/s10858-005-1846-5)

49 Boisbouvier, J., Delaglio, F. & Bax, A. 2003 Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids. *Proc. Natl Acad. Sci. USA* **100**, 11 333–11 338. (doi:10.1073/pnas.1534664100)

50 Meiler, J., Blomberg, N., Nilges, M. & Griesinger, C. 2000 Corrigendum: A new approach for applying residual dipolar couplings as restraints in structure elucidation (vol. 16, p. 245–252, 2000). *J. Biomol. NMR* **17**, 185. (doi:10.1023/A:1017377906466)

51 van Dijk, A. D., Fushman, D. & Bonvin, A. M. 2005 Various strategies of using residual dipolar couplings in NMR-driven protein docking: application to Lys48-linked di-ubiquitin and validation against 15N-relaxation data. *Proteins* **60**, 367–381. (doi:10.1002/prot.20476)

52 Wiesner, S., Stier, G., Sattler, M. & Macias, M. J. 2002 Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J. Mol. Biol.* **324**, 807–822. (doi:10.1016/S0022-2836(02)01145-2)

53 Wang, J. *et al.* 2009 Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J. Am. Chem. Soc.* **131**, 10 507–10 515. (doi:10.1021/ja902528f)

54 Iwahara, J., Anderson, D. E., Murphy, E. C. & Clore, G. M. 2003 EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **125**, 6634–6635. (doi:10.1021/ja034488q)

55 Cai, S., Zhu, L., Zhang, Z. & Chen, Y. 2007 Determination of the three-dimensional structure of the Mrf2-DNA complex using paramagnetic spin labeling. *Biochemistry* **46**, 4943–4950. (doi:10.1021/bi061738h)

56 Edwards, T. E. & Sigurdsson, S. T. 2007 Site-specific incorporation of nitroxide spin-labels into 2′-positions of nucleic acids. *Nat. Protocol* **2**, 1954–1962. (doi:10.1038/nprot.2007.273)

57 Clore, G. M., Tang, C. & Iwahara, J. 2007 Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr. Opin. Struct. Biol.* **17**, 603–616. (doi:10.1016/j.sbi.2007.08.013)

58 Iwahara, J., Schwieters, C. D. & Clore, G. M. 2004 Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126**, 5879–5896. (doi:10.1021/ja031580d)

59 Wuthrich, K. 1986 *NMR of proteins and nucleic acids.* Chichester, UK: John Wiley & Sons.

60 Cuniasse, P., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A., Le Bret, M., Guschlbauer, W. & Fazakerley, G. V. 1989 A basic frameshift in DNA. Solution conformation determined by proton NMR and molecular mechanics calculations. *Biochemistry* **28**, 2018–2026. (doi:10.1021/bi00431a009)

61 Iwahara, J., Wojciak, J. M. & Clubb, R. T. 2001 An efficient NMR experiment for analyzing sugar-puckering in unlabeled DNA: application to the 26-kDa dead ringer-DNA complex. *J. Magn. Reson.* **153**, 262–266. (doi:10.1006/jmre.2001.2448)

62 Szyperski, T., Fernandez, C., Ono, A., Kainosho, M. & Wuthrich, K. 1998 Measurement of deoxyribose 3JHH

scalar couplings reveals protein binding-induced changes in the sugar puckers of the DNA. *J. Am. Chem. Soc.* **120**, 821–822. (doi:10.1021/ja973055i)
63 Boisbouvier, J., Brutscher, B., Pardi, A., Marion, D. & Simorre, J. P. 2000 NMR determination of sugar puckers in nucleic acids from CSA-dipolar cross-correlated relaxation. *J. Am. Chem. Soc.* **122**, 6779–6780. (doi:10.1021/ja000976b)
64 Kang, Y. M., Bang, J., Lee, E. H., Ahn, H. C., Seo, Y. J., Kim, K. K., Kim, Y. G., Choi, B. S. & Lee, J. H. 2009 NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z alpha domain of human ADAR1. *J. Am. Chem. Soc.* **131**, 11 485–11 491. (doi:10.1021/ja902654u)
65 Castagne, C., Murphy, E. C., Gronenborn, A. M. & Delepierre, M. 2000 31P NMR analysis of the DNA conformation induced by protein binding SRY/DNA complexes. *Eur. J. Biochem.* **267**, 1223–1229. (doi:10.1046/j.1432-1327.2000.01124.x)
66 Madhumalar, A. & Bansal, M. 2005 Sequence preference for BI/BII conformations in DNA: MD and crystal structure data analysis. *J. Biomol. Struct. Dyn.* **23**, 13–27.
67 Szyperski, T., Ono, A., Fernandez, C., Iwai, H., Tate, S.-I., Wüthrich, K. & Kainosho, M. 1997 Measurement of 3JC2ϕP scalar couplings in a 17 kDa protein complex with 13C,15N-labeled DNA distinguishes the BI and BII phosphate conformations of the DNA. *J. Am. Chem. Soc.* **119**, 9901–9902. (doi:10.1021/ja972290y)
68 Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. 1998 DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl Acad. Sci. USA* **95**, 11 163–11 168. (doi:10.1073/pnas.95.19.11163)
69 Olson, W. K. *et al.* 2001 A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313**, 229–237. (doi:10.1006/jmbi.2001.4987)
70 Dickerson, R. E. & Chiu, T. K. 1997 Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, 361–403. (doi:10.1002/(SICI)1097-0282(1997)44:4<361::AID-BIP4>3.0.CO;2-X)
71 Fried, M. & Crothers, D. M. 1981 Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucl. Acids Res.* **9**, 6505–6525. (doi:10.1093/nar/9.23.6505)
72 Jiang, D., Jarrett, H. W. & Haskins, W. E. 2009 Methods for proteomic analysis of transcription factors. *J. Chromatogr. A* **1216**, 6881–6889. (doi:10.1016/j.chroma.2009.08.044)
73 Harrington, R. E. 1993 Studies of DNA bending and flexibility using gel-electrophoresis. *Electrophoresis* **14**, 732–746. (doi:10.1002/elps.11501401116)
74 Wu, H. M. & Crothers, D. M. 1984 The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**, 509–513. (doi:10.1038/308509a0)
75 Papapanagiotou, I., Streeter, S. D., Cary, P. D. & Kneale, G. G. 2007 DNA structural deformations in the interaction of the controller protein C.AhdI with its operator sequence. *Nucl. Acids Res.* **35**, 2643–2650. (doi:10.1093/nar/gkm129)
76 Lu, X. J. & Olson, W. K. 2008 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protocol* **3**, 1213–1227. (doi:10.1038/nprot.2008.104)
77 van Dijk, M. & Bonvin, A. M. 2009 3D-DART: a DNA structure modelling server. *Nucl. Acids Res.* **37**, W235–W239. (doi:10.1093/nar/gkp287)
78 Zheng, G., Lu, X. J. & Olson, W. K. 2009 Web 3DNA–a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucl. Acids Res.* **37**, W240–W246. (doi:10.1093/nar/gkp358)
79 Brunger, A. T. *et al.* 1998 Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921. (doi:10.1107/S0907444998003254)
80 Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. 2003 The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73. (doi:10.1016/S1090-7807(02)00014-9)
81 Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. 2006 Using Xplor-NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Sp.* **48**, 47–62. (doi:10.1016/j.pnmrs.2005.10.001)
82 Wojciak, J. M., Iwahara, J. & Clubb, R. T. 2001 The Mu repressor-DNA complex contains an immobilized 'wing' within the minor groove. *Nat. Struct. Biol.* **8**, 84–90. (doi:10.1038/89582)
83 Kuszewski, J., Schwieters, C. & Clore, G. M. 2001 Improving the accuracy of NMR structures of DNA by means of a database potential of mean force describing base-base positional interactions. *J. Am. Chem. Soc.* **123**, 3903–3918. (doi:10.1021/ja010033u)
84 Nilges, M. 1995 Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660. (doi:10.1006/jmbi.1994.0053)
85 Dominguez, C., Boelens, R. & Bonvin, A. M. 2003 HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737. (doi:10.1021/ja026939x)
86 de Vries, S. J., van Dijk, A. D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. & Bonvin, A. M. 2007 HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733. (doi:10.1002/prot.21723)
87 van Dijk, M., van Dijk, A. D., Hsu, V., Boelens, R. & Bonvin, A. M. 2006 Information-driven protein–DNA docking using HADDOCK: it is a matter of flexibility. *Nucl. Acids Res.* **34**, 3317–3325. (doi:10.1093/nar/gkl412)
88 de Vries, S. J., van Dijk, M. & Bonvin, A. M. 2010 The HADDOCK web server for data-driven biomolecular docking. *Nat. Protocol* **5**, 883–897. (doi:10.1038/nprot.2010.32)
89 van Dijk, M. & Bonvin, A. M. 2010 Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucl. Acids Res.* **38**. 5634–5647. (doi:10.1093/nar/gkg222)
90 Tidow, H., Melero, R., Mylonas, E., Freund, S. M., Grossmann, J. G., Carazo, J. M., Svergun, D. I., Valle, M. & Fersht, A. R. 2007 Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex. *Proc. Natl Acad. Sci. USA* **104**, 12 324–12 329. (doi:10.1073/pnas.0705069104)
91 Llorca, O. & Pearl, L. H. 2004 Electron microscopy studies on DNA recognition by DNA-PK. *Micron* **35**, 625–633. (doi:10.1016/j.micron.2004.05.004)
92 Svergun, D. I. & Koch, M. H. J. 2003 Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735–1782. (doi:10.1088/0034-4885/66/10/R05)
93 Petoukhov, M. V. & Svergun, D. I. 2005 Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237–1250. (doi:10.1529/biophysj.105.064154)

94 Mertens, H. D. T. & Svergun, D. I. 2010 Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172**, 128–141. (doi:10.1016/j.jsb.2010.06.012)

95 Neylon, C. 2008 Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *Eur. Biophys. J.* **37**, 531–541. (doi:10.1007/s00249-008-0259-2)

96 Pretto, D. I., Tsutakawa, S., Brosey, C. A., Castillo, A., Chagot, M. E., Smith, J. A., Tainer, J. A. & Chazin, W. J. 2010 Structural dynamics and single-stranded DNA binding activity of the three N-terminal domains of the large subunit of replication protein A from small angle X-ray scattering. *Biochemistry* **49**, 2880–2889. (doi:10.1021/bi9019934)

97 Rambo, R. P. & Tainer, J. A. 2010 Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr. Opin. Struct. Biol.* **20**, 128–137. (doi:10.1016/j.sbi.2009.12.015)

98 Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M. & Fersht, A. R. 2008 Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl Acad. Sci. USA* **105**, 5762–5767. (doi:10.1073/pnas.0801353105)

99 Williams, R. S. *et al.* 2008 Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* **135**, 97–109. (doi:10.1016/j.cell.2008.08.017)

100 Svergun, D. I. 2010 Small-angle X-ray and neutron scattering as a tool for structural systems biology. *Biol. Chem.* **391**, 737–743. (doi:10.1515/BC.2010.093)

101 Jacques, D. A. & Trewhella, J. 2010 Small-angle scattering for structural biology: expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642–657. (doi:10.1002/pro.351)

102 Mattinen, M. L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J. & Annila, A. 2002 Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data. *Biophys. J.* **83**, 1177–1183. (doi:10.1016/S0006-3495(02)75241-7)

103 Wang, J. B. *et al.* 2009 Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J. Am. Chem. Soc.* **131**, 10 507–10 515. (doi:10.1021/ja902528f)

104 Grishaev, A., Wu, J., Trewhella, J. & Bax, A. 2005 Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J. Am. Chem. Soc.* **127**, 16 621–16 628. (doi:10.1021/ja054342m)

105 Grishaev, A., Ying, J., Canny, M. D., Pardi, A. & Bax, A. 2008 Solution structure of tRNA(Val) from refinement of homology model against residual dipolar coupling and SAXS data. *J. Biomol. NMR* **42**, 99–109. (doi:10.1007/s10858-008-9267-x)

106 Gabel, F., Simon, B., Nilges, M., Petoukhov, M., Svergun, D. & Sattler, M. 2008 A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. *J. Biomol. NMR* **41**, 199–208. (doi:10.1007/s10858-008-9258-y)

107 Schwieters, C. D., Suh, J. Y., Grishaev, A., Ghirlando, R., Takayama, Y. & Clore, G. M. Solution structure of the 128 kDa enzyme I dimer from *Escherichia coli* and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. *J. Am. Chem. Soc.* **132**, 13 026–13 045. (doi:10.1021/ja105485b)

108 Schwieters, C. D. & Clore, G. M. 2007 A physical picture of atomic motions within the Dickerson DNA dodecamer in solution derived from joint ensemble refinement against NMR and large-angle X-ray scattering data. *Biochemistry* **46**, 1152–1166. (doi:10.1021/bi061943x)

109 Madl, T., Gabel, F. & Sattler, M. 2010 NMR and small-angle scattering-based structural analysis of protein complexes in solution. *J. Struct. Biol* **173**, 472–482. (doi:10.1016/j.jsb.2010.11.004)

110 Pardon, J. F., Worcester, D. L., Wooley, J. C., Tatchell, K., Vanholde, K. E. & Richards, B. M. 1975 Low-angle neutron-scattering from chromatin subunit particles. *Nucl. Acids Res.* **2**, 2163–2176. (doi:10.1093/nar/2.11.2163)

111 Falb, M., Amata, I., Gabel, F., Simon, B. & Carlomagno, T. 2010 Structure of the K-turn U4 RNA: a combined NMR and SANS study. *Nucl. Acids Res.* **38**, 6274–6285. (doi:10.1093/nar/gkq380)

112 Mittag, T., Kay, L. E. & Forman-Kay, J. D. 2010 Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recogn.* **23**, 105–116.

113 Atkinson, R. A. & Kieffer, B. 2004 The role of protein motions in molecular recognition: insights from heteronuclear NMR relaxation measurements. *Prog. Nucl. Magn. Reson. Sp.* **44**, 141–187. (doi:10.1016/j.pnmrs.2004.01.001)

114 Wright, P. E. & Dyson, H. J. 1999 Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. *J. Mol. Biol.* **293**, 321–331. (doi:10.1006/jmbi.1999.3110)

115 Vuzman, D. & Levy, Y. 2010 DNA search efficiency is modulated by charge composition and distribution in the intrinsically disordered tail. *Proc. Natl Acad. Sci. USA* **107**, 21 004–21 009. (doi:10.1073/pnas.1011775107)

116 Cordier, F., Hartmann, B., Rogowski, M., Affolter, M. & Grzesiek, S. 2006 DNA recognition by the brinker repressor—an extreme case of coupling between binding and folding. *J. Mol. Biol.* **361**, 659–672. (doi:10.1016/j.jmb.2006.06.045)

117 Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R. & Kaptein, R. 2002 Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. *EMBO J.* **21**, 2866–2876. (doi:10.1093/emboj/cdf318)

118 Halford, S. E. & Marko, J. F. 2004 How do site-specific DNA-binding proteins find their targets? *Nucl. Acids Res.* **32**, 3040–3052. (doi:10.1093/nar/gkh624)

119 Slutsky, M. & Mirny, L. A. 2004 Kinetics of protein–DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035. (doi:10.1529/biophysj.104.050765)

120 Bonnet, I. *et al.* 2008 Sliding and jumping of single EcoRV restriction enzymes on non-cognate DNA. *Nucl. Acids Res.* **36**, 4118–4127. (doi:10.1093/nar/gkn376)

121 Wang, Y. M., Austin, R. H. & Cox, E. C. 2006 Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.* **97**, 048302. (doi:10.1103/PhysRevLett.97.048302)

122 Gorman, J. & Greene, E. C. 2008 Visualizing one-dimensional diffusion of proteins along DNA. *Nat. Struct. Mol. Biol.* **15**, 768–774. (doi:10.1038/nsmb.1441)

123 Hu, T., Grosberg, A. Y. & Shklovskii, B. I. 2006 How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys. J.* **90**, 2731–2744. (doi:10.1529/biophysj.105.078162)

124 Levy, Y., Onuchic, J. N. & Wolynes, P. G. 2007 Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **129**, 738–739. (doi:10.1021/ja065531n)