Quarterly Reviews of Biophysics **32**, 1 (1999), pp. 57–98 Printed in the United Kingdom

© 1999 Cambridge University Press

NMR studies of protein–nucleic acid complexes: structures, solvation, dynamics and coupled protein folding

Torleif Härd

Department of Biotechnology, Royal Institute of Technology (KTH), CSB, Novum, S-141 57 Huddinge, Sweden

1. Introduction 58

2. Historical recapture 59

3. Inventory of high-resolution structures 64

3.1 Sequence-specific DNA-binding proteins 64

3.1.1 The *Antennapedia* homeodomain – refinement with NOEs and role of water 64

3.1.2 The *trp* repressor – a large DNA–protein complex 65

3.1.3 Sequence-specific HMG box binding in the minor groove 66

3.1.4 The GATA and AREA DNA-binding domains and the role of tight packing with DNA 67

3.1.5 DNA recognition by classical zinc finger domains and inter-domain interactions 67

3.1.6 Coupled protein folding and conformational changes in NFATC1–DBD–DNA complexes 69

3.1.7. DNA major groove recognition by a solid three-stranded β-sheet 71

3.2 Nonsequence-specific DNA-binding proteins 71

3.2.1 The Sso7d protein kinks and unwinds the DNA helix 71

3.2.2 The NHP6A HMG box protein 72

3.3 Protein/peptide–RNA complexes 72

3.3.1 Protein-like fold of the RNP–RNA complex 73

3.3.2 The HIV-1 nucleocapsid protein bound to a Ψ-RNA recognition element 74

3.3.3 RNA helix major groove binding by lentiviral Tat and Rev and lambdoid phage N peptides 74

4. Methodology for structural studies by NMR 75

4.1 Production and isotope labelling of DNA and RNA 76

4.2 Isotope labelling of proteins 76

4.3 Obtaining complexes at NMR conditions 76

4.4 NMR at weak binding equilibrium conditions 77

4.5 NMR spectroscopy – which experiments are needed? 78

4.6 Structure determinations – how is precision obtained and quality judged? 80

5. Molecular biophysics: what do structures and NMR spectra tell us? 80

5.1 Solvation and desolvation – where do the waters go and for how long do they stay there? 81

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

5.2 Dynamics in a complex – where does it occur and at what time scales? 83
5.2.1 NMR studies of dynamics – time scales and techniques 83
5.2.2 Side chain and loop motions at protein–DNA interfaces 84
5.2.3 Implications for complex stability and biomolecular recognition 85
5.3 Coupled folding: the formation or stabilization of protein structure upon binding to DNA 86
5.3.1 DNA–nucleic acid interactions involving conformational changes probed by NMR 86
5.3.2 Function of coupled folding in biomolecular recognition 89

6. Concluding remarks 90

7. Acknowledgements 90

8. References 91

1. Introduction

Many basic events which concern management and manipulation of genes within a cell – for instance transcription, replication and recombination – rely on specific interactions between proteins and nucleic acids.[1] Such interactions are also essential for many house-keeping functions, like packing and unpacking of DNA in chromatin and assembly of ribosomes. Moreover, the details of protein–nucleic acid interplay is essential for understanding the action of viruses. The list of functional mechanisms in biology that rely on protein–DNA and protein–RNA interactions can be made much longer, but these examples represent some of the topics which motivated structural biologists to study complexes between proteins and nucleic acids as a first step beyond structure determinations of individual biomolecules.

Studies of DNA-binding proteins entered early into the NMR community following the break-through period around 1980–83 when the essential methodology for structure determination in solution was developed in the laboratory of Kurt Wüthrich and his collaborators. One can distinguish at least two reasons for this. First, many bacterial repressors and DNA-binding domains of eukaryotic transcription factors fall within a size range amenable for NMR. There are situations where DNA-binding proteins resisted crystallization and biologists turned to the alternative technique. Second, sequence-specific protein–nucleic acid interactions represent a very attractive system for detailed studies of biomolecular recognition in solution. In particular, NMR can be used to study many of the coupled processes that accompany recognition, such as solvation, changes in dynamics, and coupled folding/conformational change. Such studies are as important as structure

[1] Important abbreviations: NMR, nuclear magnetic resonance; dsDNA/dsRNA, double-stranded DNA or RNA; ssDNA/ssRNA, single-stranded DNA or RNA; snRNA, small nuclear RNA; DBD, DNA-binding domain; M_r, molecular weight; NOE, nuclear Overhauser enhancement; ROE, rotating frame Overhauser enhancement; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum coherence; SA, simulated annealing; HIV-1, human immunodeficiency virus type 1; BIV, bovine immunodeficiency virus; NMRD, nuclear magnetic relaxation dispersion; BPTI, bovine pancreatic trypsin inhibitor; MAD, multi-wavelength anomalous dispersion; PDB, protein data bank; MSD, macromolecular structures database; CIDNP, chemically induced dynamic nuclear polarization; bp, base pair; p.p.m., parts per million; RNP, ribonucleoprotein; NC, nucleocapsid; DTT, dithiothreitol; RDC, residual dipolar coupling; SAR by NMR, structure-activity relationship by NMR.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

determinations *per se* and they cannot be assessed by any other high-resolution method. Moreover, these subjects are frequently of concern for physical chemists and spectroscopists interested in the behaviour of molecules in solution. Both the structural and biophysical facets of biological NMR have been developed to provide the most powerful experimental alternatives for solution studies of proteins and nucleic acids. To give a single important example: NMR methods play a major role in elucidating the structural mechanisms for HIV function (Turner & Summers, 1999).

The intention of this review is to assess the current status of the field: what has been achieved in terms of atomic-resolution structures of complexes in solution, how do these structures contribute to our understanding of biological mechanisms and biophysical issues, and what methodology is available for successful studies. We begin with a review of the earlier (pre-PDB) history. The historical recapture is by no means intended to be comprehensive, but it is formulated to put later work in perspective of the time scale of development. We then focus on recent high-resolution structures, that is those published since 1993. This inventory of the data banks will form the basis for a discussion of the current state-of-the-art technology and methodology in a section in which also rather practical aspects are covered. Next, we turn to the biophysical face of the subject and review how NMR has contributed to our understanding of biomolecular recognition and how NMR data relate to, and complement, crystallographic studies. Finally, we briefly discuss future prospects.

The review is restricted to complexes between proteins, protein domains or functional peptide fragments and polynucleotides in solution. The related areas of DNA–small molecule and RNA–aptamer complexes are not covered, but the interested reader may find introductions to these fields in reviews by Geierstanger & Wemmer (1995), Shafer (1998), and Patel (1997) and Ramos *et al.* (1997), respectively. There is also a recent review on NMR studies of transcription factors which treats many of the subjects discussed here (Bagby *et al.* 1998).

2. Historical recapture

It has been known for a long time that certain entities of the cell, such as chromatin and ribosomes, consist of both protein and nucleic acid. However, it is still probably appropriate to date the entry of protein–nucleic acid interactions into molecular biology to the formulation of the operon hypothesis (Jacob & Monod, 1961) and the establishment that transcription regulators are protein molecules (Gilbert & Müller-Hill, 1966; Ptashne, 1967). These key discoveries sparked an interest in structural mechanisms of gene regulation. However, a significant time period of at least 15 years passed before detailed structural data could be obtained. During this time biophysical research developed around questions concerning information content, location and diffusion of proteins to their cognate sites on DNA (reviewed by von Hippel (1994)), possible interaction mechanisms (von Hippel & McGhee, 1972; Seeman *et al.* 1976; Yarus, 1969), binding thermodynamics (reviewed by Härd & Lundbäck (1996)) or low-resolution studies on small-molecule models (Hélène & Lancelot, 1982).

NMR spectroscopy contributed to the early data, in particular on small molecule–DNA interactions (Hélène & Lancelot, 1982), but also for protein–nucleic acid complexes. To our knowledge, the first published NMR spectrum recorded on a protein–DNA complex is that reported by Boublik *et al.* (1971) who studied histone–DNA interactions (Fig. 1*a*). This group

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

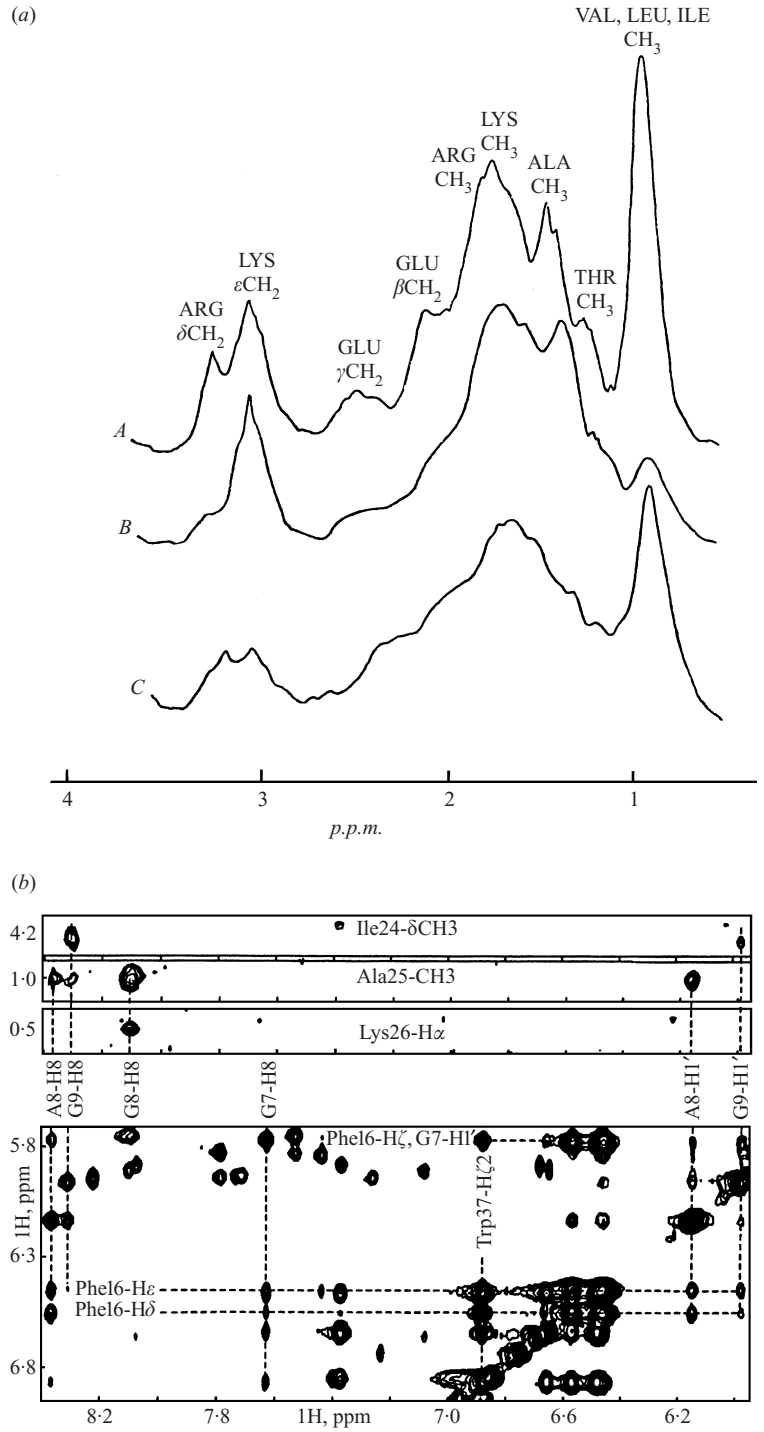


Fig. 1. (*a*) An early example of a DNA–protein NMR spectrum reproduced from Boublik *et al*. (1971). 100 MHz spectra of a histone protein (0.7 % w/v) in (top) D_2O, pH 3·2; (middle) 1 M NaCl in D_2O, pH 6·5; and (bottom) with DNA (1:1 w/w) in 1 M NaCl in D_2O, pH 6·5. (Figure reprinted with *Nature*'s copyright permission.) (*b*) Section of a high-field (800 MHz) NMR NOESY spectrum of a the HIV-1 nucleocapsid domain–RNA complex studied by De Guzman *et al*. (1998). (top) Selected strips

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

used low-field NMR to deduce several important structural properties. For instance, they found that the N- and C-terminal regions of histone protein H2A interact with DNA, whereas the portion of the peptide chain involved in inter-histone interactions does not (Bradbury *et al*. 1975). They also conducted similar studies on DNA-binding HMG1 and HMG2 proteins (Cary *et al*. 1976). An early landmark paper is that by Schulman *et al*. (1974) which includes 300 MHz NMR spectra of the impressively large EF-Tu/aminoacyl-tRNA complex. Another group that applied NMR in studies of protein–DNA interactions in the 1970s was that of David Kearns who studied the imino proton resonances of double-stranded DNA to conclude that nucleosomal DNA adopts essentially the same B-form structure as free DNA fragments (Feigon & Kearns, 1979).

The first more detailed assessment of protein–nucleic acid interactions concerned the possibility that aromatic side chains on the surface of the bacteriophage M13 gene-5 protein (G5P) interact with ssDNA nucleotides by means of intercalation (Coleman & Armitage, 1978). The G5P protein was also used to test more advanced NMR technology such as photo-CIDNP (Garssen *et al*. 1978) and, in particular, inter-molecular cross-relaxation between resonances in protein and DNA (Alma *et al*. 1981). The work by Alma *et al*. is significant in that it is, to our knowledge, the first paper where intermolecular NOEs in a complex between two large biomolecules are described. Another important paper showed that the NMR spectrum of the *lac* repressor headpiece was different in complexes with DNA fragments containing *lac* operator and nonspecific sequences (Nick *et al*. 1982). With these studies it became evident that NMR would have a role to play in structural investigations of protein–DNA recognition.

However, several groups working with X-ray crystallography began to deliver high-resolution structures of DNA-binding proteins. The first to be solved were those of the *E. coli* catabolite gene activator protein (CAP) (McKay & Steitz, 1981), λ Cro (Anderson *et al*. 1981), and λ repressor (Pabo & Lewis, 1982). The structures allowed for modeling of protein–DNA complexes and the concept of a 'recognition helix' was corroborated. Still, some models came out wrong: partly because few scientists at this time realized what Nature can come up with in terms of conformational changes.

The centre of gravity for methodology development in biomolecular NMR was at this time located in the laboratory of Kurt Wüthrich at ETH in Zürich, where the essentials for structure determinations of proteins in solution by NMR methods were developed (Wüthrich, 1986, 1989); see Wüthrich (1996) for a historical account. One of the first proteins for which the ^{1}H NMR spectrum was assigned and the secondary structure deduced was that of the *lac* repressor headpiece (Zuiderweg *et al*. 1983a, 1983b, 1985). These data, as well as a model based on NOE connectivities (Kaptein *et al*. 1985), suggested an arrangement of helices in a helix-turn-helix motif very similar to that found in for instance CAP and λ Cro.

---

from a 3D ^{13}C-filtered,^{12}C-selected HMQC-NOESY spectrum of the complex in D_2O; and (bottom) portion of a 2D NOESY spectrum of the same sample. Unambiguously identified intermolecular NOE cross-peaks have been indicated. (Reprinted with permission from De Guzman *et al*. *Science* **279**, 387–388 (1998). Copyright 1998 American Association for the Advancement of Science. Readers may view, browse, and/or download this material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or in part, without prior written permission from AAAS.)

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

As for protein–DNA complexes, the first 2D NOESY spectra in which intermolecular protein–DNA NOE connectivities could be identified were published in 1987 (Boelens *et al*. 1987; King & Coleman, 1987). In particular, Kaptein and his collaborators deduced the binding topology of the *lac* repressor headpiece to *lac* operator DNA (Boelens *et al*. 1987). They found that the orientation of the recognition helix in the DNA major groove was opposite to that proposed in an earlier model based on the dimeric structures CAP and Cro repressor (Matthews *et al*. 1982). The discussion was eventually resolved by additional NMR data (Chuprina *et al*. 1993b; Lamerichs *et al*. 1989) as well as the crystal structure of a complete *lac* repressor–DNA complex (Lewis *et al*. 1996), which confirmed the first NMR model.

The prospect of obtaining protein–DNA structures by NMR was exciting, but the advance was slowed by frustrating overlap of resonances in any two-dimensional ^{1}H NMR spectrum. For a few years NMR studies, although successful, were limited to ^{1}H NMR measurements on the protein component. The studies relied on biochemical and genetic data, and later NMR chemical shift changes, for modeling of complexes with DNA. Examples of such studies are structure determinations of the Arc repressor (Breg *et al*. 1990), and the glucocorticoid (Härd *et al*. 1990) and estrogen receptor (Schwabe *et al*. 1990a) DNA-binding domains (DBDs). It was evident that improved technology and methodology was needed to obtain high-resolution structures of complexes as well as of individual components larger than approximately 60–70 amino acid residues or 20 nucleotides. Enter uniform isotope ^{15}N and ^{13}C labelling and multi-dimensional heteronuclear NMR spectroscopy as it was perceived and developed into practical methodology in the laboratories of Ad Bax and Marius Clore and their collaborators at NIH (Clore & Gronenborn, 1992; Ikura *et al*. 1990; Marion *et al*. 1989a, 1989b); see Bax (1996) for a historical account. The possibility of using ^{15}N-editing and filtering techniques allowed Otting *et al*. (1990a) to obtain almost complete sequence-specific assignments of protein and DNA resonances in their *Antennapedia* homeodomain–DNA complex and calculate (rather than model) a structure. The structure was later refined to high resolution using ^{13}C- and ^{15}N-labelled protein (Billeter *et al*. 1993; Qian *et al*. 1993a).

The possibility of having the protein, but not the DNA, of a complex uniformly isotope-labelled is a major advantage in the crucial identification of protein–DNA NOE connectivities. This procedure, sometimes complemented with other labelling combinations, has, with two exceptions, been used in all NMR work on protein–nucleic acid complexes that have resulted in atomic resolution structures (Table 1). The basic technology and methodology necessary to determine structures of complexes were adapted by the leading groups in 1991 and 1992 and the first high-resolution NMR structures were soon to follow. Structures of sequence-specific DNA–protein complexes (and deposition of coordinates) appeared in 1993 (Billeter *et al*. 1993; Chuprina *et al*. 1993a; Omichinski *et al*. 1993). The first NMR RNA–peptide structures were published in 1995 (Puglisi *et al*. 1995; Ye *et al*. 1995), and the first NMR structure of a nonsequence-specific DNA–protein complex was published in 1998 (Agback *et al*. 1998).

Crystallographic methodology was also sharpened, although the first published structure of a protein–DNA complex (McClarin *et al*. 1986) turned out to need correction (Kim *et al*. 1990). This attempt was followed by three atomic-resolution structures of DNA complexed with helix-turn-helix motif proteins in 1988 (Aggarwal *et al*. 1988; Otwinowski *et al*. 1988; Wolberger *et al*. 1988); reviewed by Aggarwal (1990). By the end of 1993 when the first three NMR structures were deposited, there were 18 unique crystallographic structures of

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

Table 1. *Inventory of complexes in PDB and/or MSD as of August 1999*[a]

Category/complex	PDB	Main references
Protein/ssDNA		
Nucleocapsid N-terminal zinc finger (HIV-1)	1HVN, 1HVO[b]	South & Summers, 1993
Nucleocapsid protein NCp7 (HIV-1)	1BJ6	Morellet *et al*. 1998
Protein/dsDNA		
GATA-1, DNA-binding domain (chicken)	1GAT, 1GAU	Omichinski *et al*. 1993; Tjandra *et al*. 1997
Antennapedia homeodomain (*D. melanogaster*)	1AHD	Billeter *et al*. 1993; Billeter & Wüthrich, 1993; Qian *et al*. 1993a,b
lac repressor headpiece (*E. coli*)	1LCC, 1LCD	Chuprina *et al*. 1993a,b
Fli-1, ets domain (human)	1FLI[b]	Liang *et al*. 1994
c-Myb, DNA-binding domain (mouse)	1MSF, 1MSE	Ogata *et al*. 1994
trp repressor (*E. coli*)	1RCS	Zhang *et al*. 1994
SRY protein, DNA-binding domain (human)	1HRY, 1HRZ	Werner *et al*. 1995
Lymphoid enhancer-binding factor, HMG domain (mouse)	(2LEF)	Love *et al*. 1995
HMG-I(Y) DNA-binding domain 2 (human)	2EZD, 2EZE	Huth *et al*. 1997
HMG-I(Y) DNA-binding domain 3 (human)	2EZEF, 2EZG	Huth *et al*. 1997
ETS1, DNA-binding domain (human)	2STT, 2STW	Werner *et al*. 1997
GAGA factor, DNA-binding domain (*D. melanogaster*)	1YUI, 1YUJ	Omichinski *et al*. 1997
TFIIIA, zinc fingers 1–3 (frog)	1TF3	Foster *et al*. 1997; Wuttke *et al*. 1997
AtERF1 protein, GCC-box-binding domain (*A. thaliana*)	1GCC	Allen *et al*. 1998
AREA protein, DNA-binding domain (*A. nidulans*)	4GAT, 5GAT[c]	Starich *et al*. 1998a
NFATC1, DNA-binding domain (human)	1A66	Zhou *et al*. 1998
Hnf3/Fh homolog Genesis winged helix motif (rat)	2HDC	Jin *et al*. 1999
vnd/NK-2 homeodomain (*D. melanogaster*)	1NK2, 1NK3	Gruschus *et al*. 1999
Integrase, DNA-binding domain (Tn*916* transposon)	1B69, 1TN9	Wojciak *et al*. 1999; Connolly *et al*. 1999
Protein/nonspecific dsDNA		
Sso7d (*S. solfataricus*)	1BBX	Agback *et al*. 1998
NHP6A HMG box protein (*S. cervisiae*)	1CG7	Allain *et al*. 1999
Peptide/ssRNA		
Tat peptide (BIV)	1BIV	Ye *et al*. 1995
Tat peptide (BIV)	1MNB	Puglisi *et al*. 1995
Rev peptide (HIV-1)	1ULL	Ye *et al*. 1996
Rev peptide (HIV-1)	1ETF, 1ETG	Battiste *et al*. 1994, 1995
N peptide (bacteriophage λ)	n.a.	Legault *et al*. 1998
N peptide (bacteriophage P22)	1A4T	Cai *et al*. 1997; 1998
N peptide (bacteriophage λ)	1QFQ	
Protein/ssRNA		
U1A (human)	1AUD	Allain *et al*. 1996; Allain *et al*. 1997; Howe *et al*. 1998
Nucleocapsid protein (HIV-1)	1A1T	De Guzman *et al*. 1998

[a] Structures containing both protein or peptide and RNA or DNA polynucleotides.
[b] Structure of bound protein, model of complex based on intermolecular NOEs.
[c] Structure of complex with single-mutation protein also determined (Starich *et al*. 1998b) and deposited with PDB access codes 6GAT and 7GAT.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

protein–nucleic acid complexes in the PDB and the number has increased considerably since then. It has become clear that NMR at present cannot compete with crystallography in terms of the size of the complexes that are studied and the speed at which structures can be determined. Still, already the first NMR structures clearly indicated that NMR would provide information not directly available from crystallography.

3. Inventory of high-resolution structures

As of August 1999 there are 32 NMR structures of protein or peptide–nucleic acid complexes in the Protein Data Bank (PDB) or Macromolecular Structures Database (MSD). The count includes similar structures solved in different laboratories, but not variations such as structures with mutant proteins solved within a single lab. The count also includes complexes in which the structure of the nucleic acid component was modeled, rather than calculated, based on reasonable NOE-based assumptions about the polynucleotide conformation and protein binding topology. The inventory is shown in Table 1. Most structures are complexes between double-stranded DNA oligomers and domains of transcription factors or bacterial repressors. Many of the others involve model systems of retroviral structures, e.g. peptides from larger proteins, or systems where single-stranded DNA is used to emulate single-stranded RNA. It is notable that there are only two structures of protein–RNA complexes. The fairly large number of structures prohibits a more detailed assessment of each one. We therefore restrict the survey below to describe some of them in families and with regard to what we consider to be interesting or important aspects of the structure determination methodology, biophysical properties, or biological implications.

3.1. *Sequence-specific DNA-binding proteins*

This family of structures is the largest. There may be several reasons for the predominance, such as the historical development of protein–DNA studies (the quest to find the basis for recognition), the relatively simple means to establish and check strong specific binding, the abundance of various transcription factors and their involvement in particular physiological disorders, and the fact that some groups have specialized in this area and determined several structures.

3.1.1 *The* Antennapedia *homeodomain – refinement with NOEs and role of water*

The structure calculated by Billeter and co-workers in Wüthrich's laboratory (Billeter *et al*. 1993; Billeter & Wüthrich, 1993; Qian *et al*. 1993a) is interesting not only because it is the first for which an atomic-resolution model was presented (Otting *et al*. 1990a). The structure is also significant because of the way it is calculated and because of biophysical implications regarding water and dynamics at the interface. The *Antennapedia* homeodomain–DNA structure is refined based on NOE distance restraints alone, i.e. without restraints deduced from *J*-couplings, imposed hydrogen bonds, chemical shifts, residual dipolar couplings etc,

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

which are now commonly used in refinement. The refinement statistics indicate that the NOEs-only approach can yield structures which are not compromised in overall precision and accuracy. (This is not to say that the structure would not benefit from inclusion of other types of restraint data in a further refinement.) The precision is acceptable as shown by a protein heavy atom r.m.s.d. of 1·3 Å compared to the mean structure and a lower r.m.s.d. of 0·72 for protein–DNA interface atoms. The accuracy can be judged by a maximum distance restraint violation $< 0{\cdot}7$ Å, 76% of backbone dihedrals in most favoured regions of the Ramachandran diagram, and a good agreement of observed vs. expected NOEs derived from a simulation based on X-ray structures of related complexes (Billeter & Wüthrich, 1993).

An important aspect of this study is the observation and discussion of cavities at the protein–DNA interface. Complementary data show that these are occupied with water molecules (Qian *et al*. 1993b). Moreover, the residence time of bound water coincides with time scales for fluctuations of conserved side chains in the binding site. The concept of water diffusion and a dynamic side chain–water–DNA network at the interaction interface was corroborated by molecular dynamics simulations, which show that the hydration waters locate to the expected positions and also diffuse through the interface on a nanosecond time scale (Billeter *et al*. 1993, 1996). The role of water in this and other protein–DNA complexes is discussed further in Section 5.1.

3.1.2 *The* trp *repressor – a large DNA–protein complex*

The *trp* repressor binds DNA via a helix-turn-helix motif, as do many other bacterial repressors and eukaryotic homeodomains. Its DNA-binding is allosterically regulated by binding of L-tryptophan. There are two crystallographic structures of *trp* repressor–DNA complexes. However, NMR studies of a dimeric *trp* repressor–DNA complex (Zhang *et al*. 1994) are motivated despite of the scale of the undertaking – the molecular weight of the complex is in the order of 37 kDa – because they relate to a significant body of data on the free-state holo- and aporepressor structures, biochemistry and genetics, and on protein dynamics and binding thermodynamics in solution.

Many different isotope labelling schemes allowed Zhang *et al*. (1994) in Jardetzky's laboratory to collect an extensive distance restraint data set, which was complemented by hydrogen bond and dihedral angle restraints. The structure calculation was not straightforward, presumably due to the size of the complex. Different simulated annealing (SA) protocols were used to establish that the final set of SA structures were independent of starting structures. The precision of the calculated structures is high, albeit with a larger average number of distance and dihedral angle restraint violations than in most other structures in Table 1.

The overall topologies of the solution and crystal structures (for instance by Otwinowski *et al*. 1988) of the *trp* repressor–consensus operator complexes are the same, but they differ in an important detail. The protein–DNA interface side chains in the solution structure can make direct hydrogen bonds with DNA without violating NOE data, whereas the conclusion of the crystallographic work is that all but one hydrogen bond at the interface is water-mediated. Without drawing further conclusions based on these differences it is worth noting that the solution structure explains very well genetic (mutation) and biochemical (protection) data on the protein–DNA interaction.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

This and other studies by the same group have also yielded perspectives on the dynamical features of the *trp* repressor and its different binary and ternary complexes. Hence the aporepressor is found to be less stable with rapid exchange of many backbone amide protons. Binding of the co-factor makes the holorepressor more stable as evidenced by slower amide exchange rates. The stabilization is not only located to the binding site region, but throughout the structure, except for the two DNA-binding (D and E) helices of the helix-turn-helix motif (Arrowsmith *et al*. 1991). Still, the bound co-factor is in rapid equilibrium with the free state on the NMR time scale. Binding of holorepressor to DNA then stabilizes the formation of the D and E helices and also slows down the kinetics of co-factor binding. It is interesting to compare time scales for the three processes: repressor–operator binding, co-factor binding, and D and E helix unfolding (Zhang *et al*. 1994). The lifetime of bound co-factor is about one order of magnitude shorter than that of the complex, which is similar to that of the D and E helices. The authors suggest that complex dissociation depends on unfolding of the D and E helices, which, in turn, depends on dissociation of both co-factors from the dimer, by a concerted action mechanism. Moreover, it appears that the N-terminal 15 residues of the repressor undergo dynamic conformational averaging, possibly due to transient and nonspecific interactions with DNA.

3.1.3 *Sequence-specific HMG box binding in the minor groove*

The large family of eukaryotic DNA-binding proteins known as the HMG-1/HMG-2 family can be broadly divided into sequence-specific transcription factors, such as SRY, lymphoid enhancer factor 1 (LEF-1), T-cell factor 1 and SOX5, and nonspecific ubiquitous DNA-binders such as HMG-1, HMG-2, and NHP6A (see below). The former are interesting because sequence specificity is achieved by minor groove binding, and because their binding bends the DNA (Bewley *et al*. 1998). In addition, HMG box proteins are known to bind four-way DNA junctions. The solution structures of human SRY (Werner *et al*. 1995) and mouse LEF-1 (Love *et al*. 1995) HMG domains in complexes with cognate DNA sites reveal important details on HMG box–DNA interactions. These structures were determined in the laboratories of Clore and Gronenborn (SRY) and Wright (LEF-1).

HMG box proteins are characterized by three helices in an L-shaped conformation, where the short arm consists of two N-terminal helices and the long arm consists of a 22-residue C-terminal helix and the N-terminal tail. The DNA minor groove associates with the concave surface resulting in a highly distorted DNA conformation: the minor groove in widened, the major groove is narrowed and deepened, the helix is unwound and bent by 70–80° (SRY) to $117 \pm 10°$ (LEF-1). Still, Watson–Crick base pairing appears to be intact. There is an extensive protein–DNA interaction surface as the proteins cover eight or nine base-pairs of the DNA. Sequence specificity in the SRY complex is argued to be of the induced-fit type, i.e. it depends on the conformational adaptability of a particular DNA sequence, whereas the LEF-1 HMG box also make specific interactions with the DNA major groove *via* a C-terminal basic region.

The LEF-1-DNA structure is based on an impressive number of 332 (!) intermolecular NOE distance restraints. This is by far the largest number used for any of the structures in Table 1.

The similarity of binding topology and DNA distortion by the general transcription factor TFIID subunit TBP (Patikoglou & Burley, 1997) and HMG-box transcription factors might

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

be significant, since the bending of DNA provides a conformational DNA switch that can bring other DNA-bound components of the transcription initiation machinery in contact with each other (Bewley *et al*. 1998). The work on SRY is also important in that it provides plausible molecular explanations for how clinically manifested mutations result in 46X, Y-type male to female sex reversal.

3.1.4 *The GATA and AREA DNA-binding domains and role of tight packing with DNA*

The GATA transcription factors play important roles in cell differentiation in a range of organisms. They contain a conserved DNA-binding domain with a Cys_2–Cys_2 zinc finger. The structure of a GATA-1 C-terminal DNA-binding domain (cGATA-1) in complex with DNA solved by Omichinski *et al*. (1993) in the laboratory of Clore and Gronenborn is one of the first protein–DNA structures solved by NMR. It reveals a binding topology based on interactions by a helix and a loop in the DNA major groove complemented by interactions by a basic C-terminal tail in the DNA minor groove. The interactions in the major groove are predominantly hydrophobic. This particular protein–DNA interface can therefore be expected to exclude water molecules and this has also been verified experimentally using NMR techniques (Clore *et al*. 1994b).

As argued by Starich *et al*. (1998a,b), considerable work has been focused on identifying cellular targets of GATA family transcription factors, whereas the molecular mechanisms which determine specificity of regulatory proteins that bind to similar but different DNA sites is much less well understood. To bring further insight into the recognition issue they compared the structures of cGATA-1 DBD bound to a AGATAA target DNA sequence to the binding of the related AREA DBD bound to the CGATAG target and a mutant AREA DBD bound to a TGATAG target (Fig. 2). They find that the AREA DBD assumes the same fold as cGATA-1, but that the two proteins interact with DNA in distinctly different ways with regard to the conformation of the C-terminal: in the AREA DBD complex it runs parallel with the sugar phosphate backbone rather than wrapping around the backbone to interact in the DNA minor groove as in the cGATA-1 complex (Starich *et al*. 1998a). The implications on binding specificity are not immediately evident, but the authors provide a plausible scenario.

The second study addresses the wild-type AREA DBD preference for TGATAG sites compared to the preference for (A/C)GATAG sites by a leucine-to-valine mutant (Starich *et al*. 1998b). This point mutation results in dramatic changes in the expression profile of genes controlled by AREA. The structure nicely demonstrates the importance of a tightly packed hydrophobic interaction: the valine in the mutated protein is shortened by one methylene moiety compared to the leucine and the cavity thus created at the protein surface is then filled by a thymine methyl at the target DNA recognized by the mutant protein.

3.1.5 *DNA recognition by classical zinc finger domains and inter-domain interactions*

Wuttke, Foster, and collaborators in Wright's laboratory have determined a high-resolution structure of a complex between a three-zinc-finger fragment of TFIIIA (zf1–3) and a 15 bp DNA oligomer containing the cognate binding site (Foster *et al*. 1997; Wuttke et al., 1997).

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

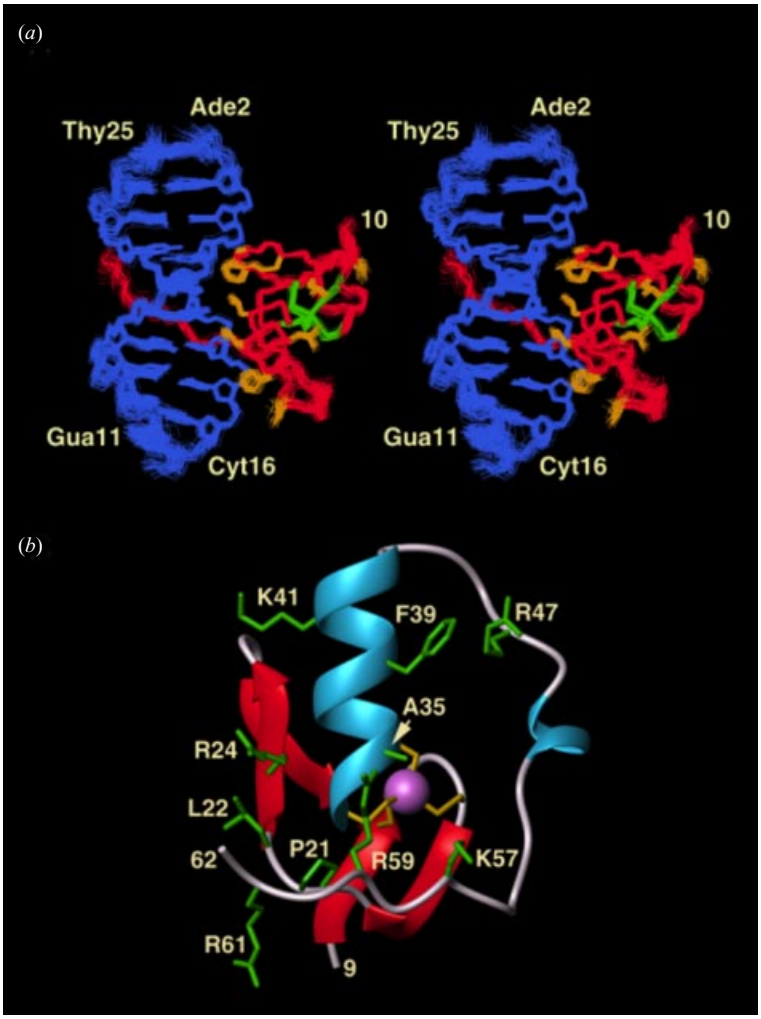


Fig. 2. Structure of the AREA DBD–DNA complex studied by Starich *et al.* (1998a). (*a*) Stereoview showing the best fit superposition of the final 35 simulated annealing structures of the complex; and (*b*) ribbon diagram showing the restrained regularized mean structure of the AREA DBD (residues 9–62) in its complexed form. (Reprinted from *J. mol. Biol.* **277**, 605–620 (1998), Starich *et al.* 'The solution structure of a fungal AREA protein–DNA complex: An alternative binding mode for the basic carboxyl tail of GATA factors', by permission of the publisher, Academic Press.)

This work provides information which is complementary to crystallographic studies of a similar complex (Pavletich & Pabo, 1991). The basic concept of zinc finger–DNA interactions, i.e. binding topology and certain direct side chain–DNA contacts, are observed in both structures. However, the crystal structure suggests a modular mode of zinc-finger–DNA interactions – each finger behaves independently and recognizes three to four DNA base pairs. The solution structure, on the other hand, shows that this perception is an over-simplification and that the repertoire of interactions is more complex and varies substantially between the zinc finger domains in zf1–3 (Wuttke *et al.* 1997).

There are two salient features of the solution zf1–3–DNA structure. First, there is tight

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

inter-domain packing of adjacent zinc fingers. In the free state, zf1–3 is a flexible molecule in which the finger domains behave like rigid bodies connected by flexible linkers (Brüschweiler *et al*. 1995). This flexibility is lost upon binding and the linker backbones assume defined conformations with extensive packing interactions between side chains within the linkers and with neighboring zinc fingers. The packing serves to fix relative orientations of the domains in the major groove. Second, the NMR data provide direct evidence for lysine side chain dynamics on microsecond to millisecond time scales at the protein–DNA interface (Foster *et al*. 1997). This is also discussed in Section 5.2.2.

The work on zf1–3 includes an analysis of protein and DNA chemical shifts and their relation to the structure. No direct correspondence between chemical shift changes and structure was identified, as the chemical shift changes upon complex formation appear to arise from a number of local direct and indirect effects (Foster *et al*. 1998).

3.1.6 *Coupled protein folding and conformational changes in NFATC1–DBD–DNA complexes*

The nuclear factor activated T-cell family (NFAT) transcription factors induce cytokine gene expression by binding to regulatory elements of antigen-responsive genes. The regulation mechanism depends on the binding of another transcription factor to a nearby AP-1 site (Rao *et al*. 1997). Zhou *et al*. (1998) in the laboratories of Wagner and Verdine determined the solution structure of the NFATC1 DBD in complex with a target site at the interleukin-2 promoter and compared it to the structure of uncomplexed NFATC1 determined by Wolfe *et al*. (1997).

The NFATC1–DBD protein fragment used for these studies contains 176 amino acid residues. With the 12 bp-DNA binding site, the NMR spectrum of the complex (M_r = 27.3 kDa) is the most complicated of all complexes in the inventory in terms of number of NMR resonances. The structure was solved using state-of-the-art methodology including many differently deuterated and reverse-labelled protein samples (additional details are given in Section 4.2). The specific labelling aided in the identification of a tyrosine hydroxyl resonance which is shifted down-field to below 10.6 p.p.m. due to hydrogen bonding in the DNA major groove. The DNA strands were joined by a hexaethylene glycol linker, which stabilizes duplex DNA without altering its structure. Initial NMR spectra recorded with the wild-type protein–DNA complex were not of sufficient quality. The problem was identified as too weak DNA binding by the wild type protein fragment. Inspection of the realated NF-κB p50 revealed that an arginine residue in the DNA-binding loop is replaced by a histidine residue in NFATC1. The corresponding His → Arg mutation in NFATC1–DBD* was found to increase the DNA-binding affinity considerably and yield a much improved NMR spectrum of the DNA-complex.

The fold of the NFATC1-DBD is characterized by a 10-stranded anti-parallel β-barrel similar to that of NF-κB p50 and s-type domains of the immunoglobulin super-family. DNA contacts are made by a large loop and a short α-helix. The corresponding fragments of the free protein are unstructured (Wolfe *et al*. 1997). The DNA-binding site recognition therefore relies on a coupled protein folding event. Moreover, a comparison with the ternary NFATC2/AP-1/DNA complex reveals that the orientation of the NFATC–DBDs cannot be the same in the two complexes without violating several observed NOE distance restraints. Hence, there must be a second conformational reorganization in order to form the

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

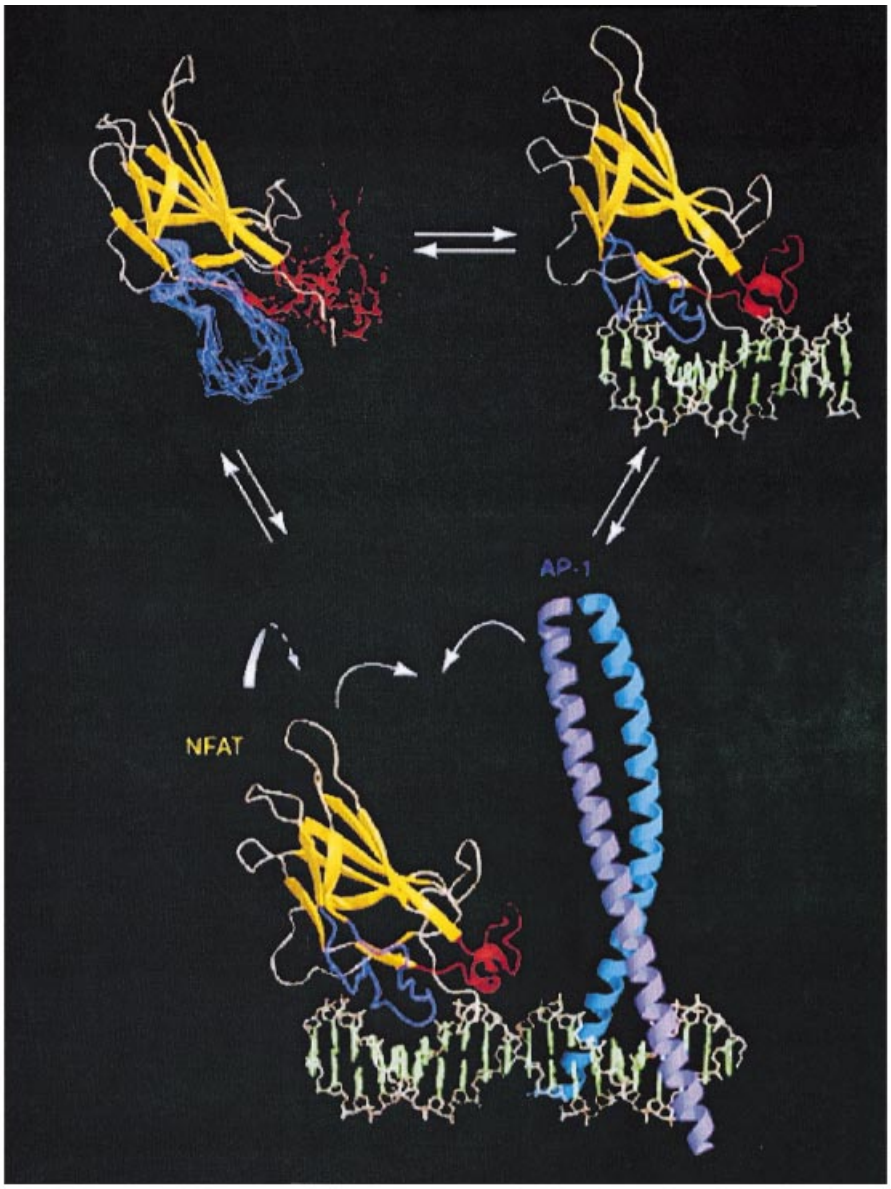


Fig. 3. Coupled folding and conformational change. Top left: solution structure of NFATC1–DBD (Wolfe *et al.* 1997), illustrating disorder in the DNA recognition loop (magenta) and insert region (red). Top right: solution structure of the NFATC1–DBD*–DNA complex in which the previously disordered regions adopt ordered structures. Bottom: model of the conformational change that is suggested to occur upon interactions between NFATC1–DBD and the AP-1 DBD bound to nearby sites on DNA (Zhou *et al.* 1998). (Figure reprinted with copyright permission by Cell Press.)

energetically most favourable contacts between NFATC1 and AP-1, as illustrated in Fig. 3. The authors discuss the conformational changes in relation to enhancement of transcription initiation by concerted action of many DNA-bound proteins.

Zhou *et al.* also point out the possibility to use the ^{1}H–^{15}N HSQC spectrum of the

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

NFATC1–DBD*–DNA complex for a direct screening of compound libraries to identify candidates for novel immuno-suppressing drugs.

3.1.7 *DNA major groove recognition by a solid three-stranded β-sheet*

Two entries in Table 1 represent complexes in which DNA-binding topology is based on a triple-stranded β-sheet interacting with the DNA major groove. These proteins are the GCC-box binding domain of a protein from *Arabidopsis thaliana* studied by Allen *et al*. (1998) in Yamasaki's and Suzuki's laboratory, and the N-terminal DBD of the Tn*916* conjugative transposon studied by Wojciak *et al*. (1999) in Clubb's laboratory. The two structures and that of the I-*Ppo*I homing endonuclease appear to represent a family in which the residue composition and position on the surface of the three-stranded β-sheet is varied to alter specificity (Wojciak *et al*. 1999). This mode of binding is distinct from the two-stranded β-sheet–DNA interaction found in the MetJ-Arc family (Allen *et al*. 1998). Neither the GCC-box binding domain nor the Tn*916* DBD undergo any major structural changes upon DNA-binding, but they induce a modestly bent DNA conformation.

3.2 *Nonsequence-specific DNA-binding proteins*

Two NMR structures of nonspecific protein–DNA complexes have been determined. These differ in that one is a low-affinity complex determined with resonances in the rapid-exchange regime and the other, which is much stronger, show slow-exchange kinetics. Furthermore, the two proteins are likely to perform distinctly different biological functions. We therefore describe both structures in some detail.

3.2.1 *The Sso7d protein kinks and unwinds the DNA helix*

The dimeric complex between the Sso7d protein from the hyperthermophilic archaeon *Sulfolobus solfataricus* and a 12 bp palindrome DNA duplex studied by Agback *et al*. (1998) in our laboratory has a molecular weight of about 22 kDa. NMR spectra and equilibrium data were used to establish a 2:1 protein:DNA binding stoichiometry and the complex was refined as a symmetric dimer. The weak-binding equilibrium required special considerations with regard to solution and concentration conditions, as discussed further in Section 4.4.

Sso7d binds DNA by interactions in and across the DNA minor groove. The observed topology is such that it is possible to understand why the association is nonspecific, yet sufficiently strong to be biologically relevant. The interaction involves electrostatic salt-bridges, nonpolar contacts with deoxyribose units and DNA bases and hydrogen bonds from protein donors to DNA base acceptors. The latter are nonspecific in this context, because the pattern of hydrogen bond acceptors exposed by the minor groove is independent of DNA sequence. Furthermore, the surface area that becomes buried from solvent exposure upon association is smaller than in more tightly packed sequence-specific complexes. The change in accessible surface area (ΔASA) in the Sso7d–DNA complex is approximately 1000 Å^2 per bound monomer. The corresponding value for sequence-specific binding normally is at least

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

twice as large with proteins of similar size, for instance for GATA-1 (Omichinski *et al*. 1993). The ΔASA value correlates with the effect of water activity on the binding constant, that is, the extent of desolvation of surfaces upon complex formation (Parsegian *et al*. 1995). Hence, the small ΔASA found in the Sso7d–DNA complex is consistent with thermodynamic data indicating a negligably small dependence of the binding constant on osmotic stress (water activity) (Lundbäck *et al*. 1998).

The function of Sso7d and other related abundant proteins in *Sulfolobus* species is not known, but they are thought to be involved in DNA packing in the cell. The Sso7d–DNA complex reveals a structural distortion of the B-form DNA, in particular significant unwinding of the double helix. This is consistent with biochemical work and suggests that Sso7d may act to compact DNA by introduction of negative super-helicity.

3.2.2 *The NHP6A HMG box protein*

The yeast HMG box protein NHP6A (93 residues) is present in the cell at concentrations similar to those of other nonspecific HMG box proteins. It binds DNA without apparent sequence specificity and binding induces DNA bending. The yeast protein forms a more stable complex with DNA than does mammalian HMG1/2. Allain *et al*. (1999) in Feigon's laboratory have presented an NOE-based model of a NHP6A–DNA complex.

The large size of the DNA (15 non-equivalent DNA bp) in combination with the nonspecific nature of the complex required special measures to assign the NMR spectrum of DNA. Hence, the investigation included four complexes with different isotope labelling: one with $^{13}C,^{15}N$-labelled protein and unlabelled DNA and three with $^{13}C,^{15}N$-labelled protein and $^{13}C,^{15}N$-labelled DNA, i.e. on the two complementary strands and on both strands, respectively. This method was successful in that it allowed for the assignment of a large number of resonances and inter-molecular NOEs in the crowded H4′, H5′–H5′ region of the spectrum (Masse et al., 1999).

The structure determination of the protein component in combination with several protein–DNA NOEs enabled the authors to propose a high-resolution model of the complex. The DNA duplex contains an SRY recognition sequence (see above). The model places NHP6A with a shifted binding site and a reverse orientation as compared to SRY. The DNA structure is distorted as in other HMG-box complexes and the N-terminal tail of NHP6A is wrapped around the major groove in a manner similar to that of the LEF-1 C-terminal tail (Love *et al*. 1995). Allain *et al*. provide a plausible explanation for the different binding modes of sequence-specific and nonspecific HMG-box domains: conserved nonpolar surface side chains make specific contacts with deoxyribose units on opposite strands in the specific SRY and LEF-1 recognition sequences, compared to that observed in the complex between NHP6A and a SRY recognition sequence. A possible and attractive implication is that the opposite orientation serves to repress transcription (by nonspecific HMG-box proteins) rather than activate the promoter (by sequence-specific transcription factors SRU and LEF1) (Allain *et al*. 1999).

3.3 *Protein/peptide–RNA complexes*

Progress on RNA–protein structures has been slower than on DNA–protein structures. RNA and RNA complexes are more difficult to crystallize, presumably due to larger conformational heterogeneity of RNA polynucleotides. RNA complexes also presented a challenge for NMR

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

because of their large sizes in combination with degeneracy of the NMR spectrum of a long RNA fragment. This latter obstacle has recently been partly removed by development of effective techniques for isotope labelling of RNA (Varani *et al*. 1996).

3.3.1 *Protein-like fold of the RNP–RNA complex*

The ribonucleoprotein (RNP) domain is one of the most common eukaryotic protein sequence motifs. The human U1A protein participates in the ribonucleoprotein particle complex in which splice-points in pre-mRNA are recognized. The U1A also controls its own expression by binding to a sequence within the 3′ untranslated region of its own mRNA thereby preventing polyadenylation (Nagai *et al*. 1995). The U1A RNP–RNA complex is therefore an attractive target for structural studies aimed at understanding protein–RNA interactions and mRNA splicing mechanisms.

RNP domains consist of a compact α/β fold with a four-stranded β-sheet packed against two α-helices. RNA binding occurs primarily on the surface of the β-sheet, as demonstrated in the (crystallographic) structure of the N-terminal RNP domain of U1A in complex with a snRNA fragment (Oubridge *et al*. 1994). Subsequent NMR studies on a complex between the same U1A fragment and an RNA polyadenylation inhibition element (PIE) have resulted in a more detailed understanding of the RNP domain–RNA recognition process (Allain *et al*. 1996, 1997; Howe *et al*. 1998).

The complex studied by Allain, Howe and co-workers in the laboratories of Neuhaus and Varani consists of an N-terminal 101-residue fragment of U1A and a 30-nucleotide internal RNA loop derived from the PIE (M_r approximately 22 kDa). Initial precipitation problems were overcome by using a doubly mutated U1A protein. The studies involved a large number of different NMR experiments and special techniques for assignments and identification of intermolecular NOEs (Howe *et al*. 1998). The refined structure is based on an extensive set of 2581 geometry restraints including 125 intermolecular distances. The precision of atom positions at the protein–RNA interface is sufficient for a detailed assessment of the interactions (the heavy atom r.m.s.d. compared to the mean structure is 1.08 Å).

The recognition involves folding and structural rearrangements in both RNA and protein components. Some of the RNA nucleotides become more ordered in the complex and the protein binding loop is opened. A C-terminal helix which partly covers the β-sheet surface in free U1A adopts another orientation in the complex. The bases of single-stranded RNA nucleotides pack against hydrophobic side chains at the surface of the β-sheet to form a hydrophobic core whereas negatively charged backbone phosphates of the RNA are on the surface. The authors note that this picture of RNA–protein recognition is more similar to the general picture of protein–protein interactions than that of protein–DNA interactions. The surface of the β-sheet is very similar in all RNP domains so this part of the interaction cannot be responsible for discrimination between different RNA targets. RNA-binding specificity is instead achieved through interactions between more variable loop regions of U1A and RNA phosphates. An electrostatic basis for binding specificity may appear counter-intuitive in view of the current understanding of nonspecific interactions in protein–DNA complexes (Härd & Lundbäck, 1996). However, the basic side chains that contact the RNA are specific to U1A. Moreover, it has been demonstrated that binding specificity of another RNP domain is attenuated at increased salt concentrations (Abdul-Manam *et al*. 1996) and this is consistent with binding site discrimination based on electrostatic interactions.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

3.3.2 *The HIV-1 nucleocapsid protein bound to a Ψ-RNA recognition element*

The HIV-1 nucleocapsid (NC) protein contains two CCHC-type zinc knuckle domains. It participates directly in viral RNA recognition and encapsidation by binding a ~120 nucleotide region of the unspliced RNA known as the Ψ-site (Turner & Summers, 1999). The Ψ-site contains four stem-loop structures of which one, SL3, is interesting since it is conserved in different HIV-1 strains. De Guzman and co-workers (1998) in Summers' laboratory have determined a high-resolution NMR structure of an NC-SL3 complex using 800 MHz NMR data (Fig. 1*b*).

The structure reveals a topology based on insertion of an N-terminal peptide 3_{10} helix in the major groove of the RNA stem. Additional contacts are made with the SL3 tetra-loop and both zinc domains are consistent with conservation of residues in protein and RNA. Interestingly, both zinc domains bind guanosine bases in a hydrophobic cleft with contacts that are essentially identical to those observed with a guanosine in an NC zinc domain–ssDNA complex (South & Summers, 1993), suggesting that zinc knuckle domains function by binding specifically to guanosine. As with the classical zinc fingers (described above) complex formation results in a packing of the zinc domains which is not observed in the uncomplexed protein. The authors speculate how the inherent NC flexibility might be utilized to adopt different conformations required to form stable complexes with other stem loops within the Ψ-RNA fragment.

This particular undertaking merits some comments on how a well-conceived structure determination of a protein–nucleic acid complex can be carried out by (i) a judicious choice of target RNA fragment – the SL3 was stabilized by the introduction of four extra base pairs at the stem not contacted by NC; (ii) obtaining both protein and RNA components as ^{13}C,^{15}N-labelled samples, (iii) checking the molecular weights by mass spectrometry to a precision of 0·03%; (iv) verifying by gel filtration that the complex stoichiometry is an unambiguous 1:1, (v) measuring the complex stability (K_d of ~100 nM) using electrophoretic mobility shift assays: (vi) collecting extensive data sets at 600 and 800 MHz spectrometers; (vii) carrying out the structure determination in such a way that high precision is obtained (heavy-atom r.m.s.d. of 0·93 Å for structured residues) without compromising distance restraints (maximum violation < 0·2 Å); and (viii) confirming disorder in the uncomplexed NC by NMR relaxation measurements.

3.3.3 *RNA helix major groove binding by lentiviral Tat and Rev and lambdoid phage N peptides*

The structures of peptide–stem loop RNA complexes reveal a large variability of binding topology within the RNA major groove, as illustrated by the following examples.

The *trans*-activating protein Tat from bovine immunodeficiency virus (BIV), and also from HIV, binds to the 5′ end of the viral RNA thereby enhancing transcription (Turner & Summers, 1999). The RNA binding site is known as the *trans*-activation response element (TAR). Two groups have determined high resolution structures of Tat peptide–TAR RNA complexes (Puglisi *et al*. 1995; Ye *et al*. 1995). The RNA-binding domain peptide of Tat is very abundant in arginine residues: seven of 14 residues in the fragment used by Puglisi *et al*. (1995) are arginines. The topology of the complexes reveal that the peptide forms a β-hairpin, which binds in the RNA major groove. Two bulged uracil bases act to induce a 40º bend in the RNA stem resulting in a deep major groove and the peptide is completely

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

enclosed within the diameter of the RNA helix. The resolution at the RNA–peptide interface is somewhat more reliable in the structure determined by Ye *et al*. (1995) in Patel's laboratory because they were able to identify many more intermolecular NOEs. A total of 102 intermolecular distance restraints was used in the structure determination compared to 26 used by Puglisi *et al*. Sequence-specific recognition is accomplished by several hydrogen bonds involving backbone and side chain groups and hydrophobic interactions, and the N-terminus of the Tar peptide interacts with the hairpin loop of the TAR RNA.

The HIV Rev protein binds to unspliced or incompletely spliced viral mRNA and modulates the proportion of spliced and unspliced mRNA exported to the cytoplasm. The Rev response element (RRE) is 234 nucleotides large. Rev oligomerizes on RRE following the recognition of a high-affinity stem loop site (Turner & Summers, 1999). The RNA-binding site on Rev is localized to an arginine-rich fragment containing 16 amino acid residues. Two articles describe the structures of this peptide fragment in complexes with the high-affinity RRE site (Battiste *et al*. 1996) and an RNA aptamer that binds the Rev peptide with even higher affinity (Ye *et al*. 1996). The structures were determined in the laboratories of Williamson and Patel, respectively. In both cases the Rev peptide forms an α-helix, which binds in a widened RNA major groove. The interactions involve a large number of hydrogen bonds and electrostatic contacts. The structures of the peptide and RNA appear to be mutually stabilizing as in a single folding unit (Tan & Frankel, 1994). The overall topology and peptide–RNA interactions in the Rev–RNA aptamer complex are similar to those in the Rev-RRE complex, but there are differences in the intra-RNA interactions that lead to the formation of the bound-state RNA conformation (Ye *et al*. 1996).

A third variation of peptide–RNA major groove binding is observed in complexes between N-protein peptides and phage P22 (Cai *et al*. 1998) and phage λ (Legault *et al*. 1998) *boxB* RNA. The N-protein functions as an anti-terminator of phage RNA transcription by binding to the N-utilization (*nut*) site on the nascent transcript. Specifically, an arginine-rich domain at the N-terminus recognizes a small RNA hairpin (*boxB*) within *nut*. The peptide forms a bent α-helix which is positioned in the major groove of the A-form RNA stem. The GAAGA loop in phage λ *boxB* adopts a GNRA (tetraloop) fold in which the fourth guanine nucleotide is excluded from the loop. In phage λ the tetraloop is further stabilized by stacking interactions between an adenosine base and an N-peptide tryptophan side chain. The paper published by Greenblatt's laboratory (Legault *et al*. 1998) provides a detailed discussion of this interaction compared to tetra-loop interactions found with Group I introns. They also conclude that the guanine which is excluded from the tetra-loop acts to recruit the NusA host factor to the complex. The phage P22 *boxB* tetraloop also adopts a GNRA fold in the complex by excluding the third nucleotide (a cytosine), which in this case makes hydrophobic contacts with the P22 N-peptide. This and some other conformational details are different in the two complexes as discussed by Patel and co-workers (Cai *et al*. 1998).

4. Methodology for structural studies by NMR

In this section we describe aspects of NMR structure determination that are specific to protein–nucleic acid complexes in some more detail, whereas methodology that is commonly used also in other areas of biomolecular NMR is reviewed more briefly.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

4.1 *Production and isotope labelling of DNA and RNA*

Unlabelled DNA oligomers are most conveniently obtained using commercial DNA synthesizers, followed by removal of the 5′-dimethoxytrityl group. DNA duplex stability can be improved by linking the oligomers (Altmann *et al*. 1995). DNA purification protocols involving HPLC, ion exchange, polyacrylamide gel eletrophoresis, or gel filtration have been used in the different articles reviewed here. One of the more recent high-resolution structures in the inventory was determined using ^{13}C,^{15}N-labelled DNA (Allain *et al*. 1999; Masse *et al*. 1999). Enzymatic methods are most likely a much cheaper alternative to automatic methods for DNA labelling (Zimmer & Crothers, 1995; Masse *et al*. 1998). Enzymatic synthesis of region-specific isotope-labelled DNA oligomers has also been described (Mer & Chazin, 1998).

Isotope labelling of RNA has been more common since the methodology is well established. Most often, RNA oligomers are prepared by *in vitro* transcription using synthetic DNA templates and T7 RNA polymerase (Milligan & Uhlenbeck, 1989) and purified by single-nucleotide resolution gel electrophoresis (Wyatt *et al*. 1991). The *in vitro* transcription protocols allow for use of isotope-enriched nucleoside triphosphates, which can be obtained commercially or from *Methylophilus methylotropus* grown with ^{13}C-methanol and ^{15}N-ammonium sulfate (Batey *et al*. 1995). Protocols for obtaining random fractional deuteration (Batey *et al*. 1996) or segmental isotope labelling (Xu *et al*. 1996) of RNA have also been described.

4.2 *Isotope labelling of proteins*

Isotope labelling of proteins is an area of continuous development. Several recent review articles address uniform, fractional, or site-specific isotope labelling with ^{2}H, ^{13}C and/or ^{15}N (Dötsch & Wagner, 1998; Gardner & Kay, 1998; Kainosho, 1997). The most advanced labelling schemes used to date in studies of protein–nucleic acid complexes are those applied for the structure determinations of the *trp* repressor-operator and the core NFATC1–DNA complexes. Zhang *et al*. (1994) employed semi-selective protein deuteration schemes to obtain seven different selectively deuterated *trp* repressor analogues, and also ^{13}C-, ^{15}N-, and ^{13}C,^{15}N-labelled protein samples. The NFATC1 DBD samples used by Zhou *et al*. (1998) included perdeuterated/^{15}N-labelled, fractionally deuterated ^{13}C,^{15}N-labelled, Arg(^{13}C,^{15}N)-selectively labelled, and fully deuterated/Phe(^{1}H)/Tyr(^{1}H)-labelled protein, which allowed them to obtain nearly complete ^{1}H, ^{13}C, and ^{15}N resonance assignments despite the large size of the complex.

A novel method with high potential is segmental labelling either by trans-protein splicing (Otomo *et al*. 1999; Yamazaki *et al*. 1998) or chemical ligation (Xu *et al*. 1999). However, these methods need to be developed further before they can be put to efficient use.

4.3 *Obtaining complexes at NMR conditions*

Many frustrated NMR spectroscopists have been witnessing protein–nucleic acid complexes precipitate before their eyes. As with most protein chemistry at high concentrations there are no general recipes on how to solve a solubility problem, if it is possible at all. Still, an

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

important project should be worth additional trials and the literature reveals some guidelines on how to proceed.

- Standard variables are buffer composition, pH, salt concentration and temperature. Low salt concentration conditions as well as use of EDTA are common in the literature.
- Variation of nucleic acid oligomer length and sequence might be tried to increase stability (De Guzman *et al*. 1998) and avoid unfavourable charge compensation scenarios.
- One may consider introduction of judicious point mutations in the protein that could improve stability. Single cysteine residues may for instance be mutated to serines to prevent oxidation. This methodology has been used in several projects, for instance in the *Antennapedia* homeodomain (Otting *et al*. 1990a). Cysteine oxidation can also be prevented by reducing agents such as DTT, but DTT is not very stable at high pH ($> 7{\cdot}0$–$8{\cdot}0$).
- Protein mutations may improve DNA-binding kinetics to move it out of the line-broadening regime and into the slow-exchange regime of the NMR time scale. This strategy was applied in studies of NFATC1–DBD. Here, preliminary studies yielded poor ^{1}H–^{15}N HSQC spectra and the project was deemed unsuitable for NMR. However, a histidine to arginine mutation improved DNA-binding affinity and resulted in a successful completion of the project (Zhou *et al*. 1998).
- Protein mutants with significantly improved solubility have also been encountered. For instance, nonpolar side chains on the surface might be replaced with charged side chains as in the work on U1A (Howe *et al*. 1998). (No guarantees for success are given.)
- The DNA–nucleic acid mixing conditions can be optimized to disfavour kinetic trapping of aggregates. Hence, it is a good idea to mix dilute solutions under stirring and then concentrate the complex (Omichinski *et al*. 1993). Ideally dilute solutions are impractical and aggregation kinetics might be effective also at intermediate concentrations. Thus, if addition of protein to nucleic acid results in precipitation then this does not rule out that the opposite procedure will work. The protein-to-nucleic acid titration has the advantage that the chemical shifts of DNA or RNA imino proton resonances can be used to monitor the binding stoichiometry.
- In studies of the ETS1-DBD the complex was formed in two ways, which both involved initial denaturation of the protein in 6 M guanidine hydrochloride followed by renaturation in the presence of cognate DNA oligomers (Werner *et al*. 1997).

An innovative approach to accelerate the initial assessment of protein NMR projects, which also can be used to screen the suitability of different protein constructs of target DNA oligomers under NMR conditions has been described (Huth *et al*. 1997). The method relies on expression vectors encoding the immunoglobulin domain of streptococcal protein G (GB1 domain) linked to the N-terminus of the desired protein. The fusion strategy takes advantage of the stability, solubility and high expression yield of the GB1 domain.

4.4 *NMR at weak binding equilibrium conditions*

The kinetics of a protein–nucleic acid binding equilibrium present no special problem for a stable complex in slow exchange at the NMR time scale. More rapid binding kinetics result in resonance line broadening and this situation must be dealt with in some way. Rapid

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

exchange in weak complexes again yield sharp NMR resonances, but this situation might require precautions as discussed in this section. At least two structures in Table 1 were solved under conditions of rapid exchange at the NMR time scale: the (sequence-specific) complex between the *lac* repressor headpiece and an 11 bp *lac* operator (Chuprina *et al*. 1993a) and the nonspecific Sso7d–DNA complex (Agback *et al*. 1998).

The work on Sso7d made evident some difficulties involved in solving solution structures of weak complexes. The structure was solved at 50 °C using a 12-bp DNA oligomer and about 60 mM total salt concentration. The binding affinity for Sso7d under these conditions is low (only about $1{\cdot}6 \times 10^4$ M^{-1}) and the life time of the bound state could be estimated to be in the order of milliseconds. With total protein or DNA concentrations in the millimolar range one is then faced with two possible relative concentration ratios: (i) a situation in which the protein is in excess resulting in a saturation of available DNA sites, but also a significant amount of free protein; or (ii) a situation in which the DNA is in excess and most protein exists in the complexed state, but where a significant proportion of DNA sites are not occupied. The largest chemical shift changes were observed for protein proton resonances ($>$ 1 p.p.m.) whereas the DNA chemical shifts were less affected ($<$ 0·3 p.p.m.). With these different shift changes it is advantageous to keep the DNA in access so that the protein resonances are not broadened, as the most shifted protein resonances will approach intermediate exchange if a significant amount of free protein is present. The NOE data was collected under these conditions. However, there is a potential possibility that cross-relaxation and *J*-coupling averaging at uncomplexed DNA sites cause artifacts in the sense that the determined DNA structure is an average of free and bound conditions. This possibility was largely avoided by the high temperature at which non-complexed DNA is highly flexible and cross-relaxation is minimized. NMR spectra at the opposite concentration conditions (excess protein) were also recorded. In these, the most shifted protein resonances were in the intermediate exchange regime on the NMR time scale and therefore not observable. However, an identical set of NOE connectivities was otherwise observed, compared to samples with excess DNA.

4.5 *NMR spectroscopy – which experiments are needed?*

The composition of an optimal set of NMR experiments very much depends on the isotope labelling of available samples. The following selection would constitute a basic portfolio. Additional examples of useful experiments can be obtained from the articles describing structure determinations in Table 1, or from recent reviews (Bax, 1994; Kay, 1995). Many of the experiments mentioned below have been described in detail by Cavanagh *et al*. (1996) and the reader is referred to this text for all experiments for which an explicit reference is not given. The details of different assignment procedures are not covered in this review. Moreover, several of the experiments below are being re-designed to include the resolution enhancement provided by TROSY (Pervushin *et al*. 1997; Salzmann *et al*. 1999). TROSY is an area in rapid development which is likely to have an impact on the methodology used with large complexes.

Needless to say, the two-dimensional ^{1}H–^{15}N HSQC experiment is the most convenient way to assess the suitability for NMR of a protein or complex preparation. As for backbone protein resonance assignments these are typically made using HNCA and HN(CO)CA

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

experiments with ^{13}C,^{15}N-labelled protein samples, which preferably also are fractionally deuterated. An alternative common variant is to utilize both C^{α} and C^{β} chemical shift information afforded by the CBCA(CO)NH experiment, which may be combined with CBCANH to secure intra-residue J connectivities. If chemical shifts are to be used in the refinement then it may be worthwhile also to assign carbonyl shifts using HNCO and/or HN(CA)CO.

The favourite methodology for assignments of side chain protons varies in the literature. However, most studies include HCCH-TOCSY and HCCH-COSY (in D_2O) and ^{15}N TOCSY-HSQC, in different combinations with amide proton-detected carbon TOCSY experiments and HBHA(CO)NH (Kay, 1995). The number of experiments needed here and also for nucleic acid assignments strongly depends on the degree of spectral overlap. One study employ more modern residue-specific labelling and spectroscopy for side chain residue assignments (Zhou *et al*. 1998).

Most structure determinations rely to some extent on J coupling measurements to determine ϕ backbone dihedral angles using the HNHA experiment. Alternative methods to determine ϕ and other dihedral angles are also available (Bax, 1994).

DNA resonances in non-labelled DNA are commonly assigned using for instance a 2D NOESY spectrum and a 2D TOCSY spectrum. These experiments require heteronuclear filters depending on the isotope label status of the protein. Care must be taken not to mistake exchangeable protons in protein and DNA. More complete assignments of DNA resonances (and protein–DNA NOEs) can be obtained from HCCH-TOCSY and HCCH-COSY experiments on ^{13}C-labelled DNA in complex with non-labelled protein. Experiments and procedures for assignments of resonances in isotope-labelled RNA fragments have, for instance, been described by Gubser & Varani (1996).

Distance restraints are obtained from NOESY experiments with ^{15}N and/or ^{13}C-editing or filtering. The most extensive intermolecular restraint data set reported for the structures in Table 1 were obtained from a 3D $^{13}C(\omega_1,\omega_2)$-edited, $^{13}C(\omega_3)$-filtered NOESY spectrum (Love *et al*. 1995) and the results – 332 intermolecular restraints – make this experiment appear superior to the alternatives. An easy procedure to identify the corresponding NOEs, albeit with a much larger overlap possibility, in a 2D ^{1}H,^{1}H-NOESY is to record this without heteronuclear decoupling. Protein-to-DNA NOEs then appear as doublets in only one of the frequency dimensions (Allen *et al*. 1998).

Residual dipolar couplings (RDCs) provide long-range, as opposed to local, structural restraints (Tolman *et al*. 1995). The presence of RDCs require some degree of molecular alignment, either by interactions between an anisotropic molecular susceptibility tensor χ and the magnetic field (Bothner-By, 1996) or by means of an anisotropic medium (Tjandra & Bax, 1997). The advantage for studies of protein–DNA complexes is that the anisotropic part of the tensor, $\Delta\chi$ becomes large and to a good approximation axially symmetric due to the stacking of aromatic bases in the double helix (Tjandra *et al*. 1997). Hence, RDCs can be measured accurately without resorting to orienting media. Tjandra *et al*. (1997) used the previously studied GATA-1–DNA complex to demonstrate that introduction of RDC restraints can improve the structural precision of DNA-complexed proteins, especially in loop regions. RDC restraints were also used in studies of the AREA protein in the same group (Starich *et al*. 1998a,b) and the method is likely to gain a wide-spread use.

A recently introduced methodology makes it possible to directly observe hydrogen bonds in biomolecules (Pervushin *et al*. 1998; Cordier & Grzesiek, 1999; Cornilescu *et al*. 1999). The

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

methodology is in principle applicable to protein–nucleic acid complexes, for which it might become extremely valuable to identify the presence of direct hydrogen bonds and to resolve hydrogen bond dynamics issues (see the discussion of the *Antennapedia* homeodomain in Section 5.2.2 below).

4.6 *Structure determinations – how is precision obtained and quality judged?*

Surveys of the most modern methods for protein structure determination have been published by Clore & Gronenborn (1998a,b). In addition to classical NOE-based distance restraints and dihedral restraints based on three-bond J coupling constants, they describe the use of secondary chemical shifts, residual dipolar couplings, R_2/R_1 relaxation rate ratios, and conformational data base potentials.

Hydrogen bonds are frequently used in the refinement in order to fix DNA base pairing and protein secondary structures. In these studies, corroborating NOE evidence is used to deduce the presence of hydrogen bonds, but this procedure should become unnecessary as experiments designed to detect hydrogen bonds directly become available, as described above.

Clearly, different groups favour different sets of restraints: some groups are more orthodox and rely on direct experimental evidence, whereas other groups make extensive use also of indirect and empirical data. There is to the best of our knowledge no indication that judicious use of all available data and refinement with conformational data base potentials should compromise the accuracy of structures.

The most common method to determine structures is restrained molecular dynamics and simulated annealing – either in Cartesian space (Brünger, 1992; Brünger & Nilges, 1993) or with torsion angle dynamics (Güntert *et al*. 1998). The annealing procedure is more critical for molecular complexes than for individual molecules owing to the larger scale of the conformational sampling problem. Howe *et al*. (1998) devised a very successful combination of standard RNA and protein annealing protocols which yielded a high convergence rate in their U1A–RNA structure calculations.

The most common measures of accuracy used in the literature are the appearance of Ramachandran diagrams, deviations from ideal geometry, number and magnitude of violations of experimental restraints, and potential energies such as the van der Waals interaction energy. Common programs to evaluate the quality of a structure are WHAT IF (Vriend, 1990), WHAT_CHECK (Hooft *et al*. 1996) and PROCHECK (Laskowski *et al*. 1996), which to a large part rely on comparisons with high-resolution crystallographic structures. Without discussing the problem in more detail, we note that the WHAT_CHECK scores of the structures in Table 1 on average are rather poor also for more recent structures. There is clearly more work to done in the structure calculation and validation field and such studies are also being pursued (Linge & Nilges, 1999).

5. Molecular biophysics: what do structures and NMR spectra tell us?

In this section we come back to the biophysical aspects of nucleic acid binding by proteins and address these in some more detail by recapturing evidence obtained also from other studies than those resulting in the high-resolution structures in the inventory in Section 3.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

5.1 *Solvation and desolvation – where do the waters go and for how long do they stay there?*

About 10 years ago Otting and Wüthrich discovered that NMR can be used to identify specific hydration sites in proteins and on protein surfaces. They developed methods based on comparison of cross-relaxation rates of longitudinal magnetization (NOE) and transverse magnetization in the rotating frame (ROE) to estimate average residence times of waters at their specific sites. The residence times can be quantified rather precisely if they fall in the picosecond to low-nanosecond time range and a lower limit (depending on the rotational correlation time of the molecule or complex) can be estimated for more slowly exchanging waters. Principles and applications have been reviewed by Otting *et al*. (1990b) and Gerothanassis (1994), and additional references may be found in more recent literature (Sobol *et al*. 1998). In general, water residence times are found to be less than 1 ns for a surface water and more than 1 ns, but less than 10 ms, for waters in protein interiors or deep surface cavities. A different approach for determination of residence times of water or other solutes that lie outside the range accessible by high-resolution NMR has been taken by Halle and Denisov. Their methodology is based on nuclear magnetic relaxation dispersion (NMRD) measurements. It has, for instance, been used to determine residence times of surface and internal waters in proteins (Denisov & Halle, 1995) and study the conformational fluctuation landscape in BPTI (Denisov *et al*. 1996). An often-quoted result is the residence time of an internal water in BPTI: 170 ± 20 μs at 300 K (Denisov *et al*. 1996).

In terms of protein–DNA interactions the NOE/ROE method has been used to study hydration at the *Antennapedia* homeodomain-DNA complex (Qian *et al*. 1993b; Wüthrich, 1993) and the GATA-1 DBD-DNA complex (Clore *et al*. 1994a), for which the structures have been described above. A large number of hydration sites with residence times longer than 1 ns was identified at the homeodomain–DNA interface, whereas the different NMR structures seem to allow room for only a few waters at a time. The apparent discrepancy would be consistent with a continuous diffusion of waters in and out of the interface. To investigate this possibility Billeter *et al*. (1996) conducted a molecular dynamics simulation for 2 ns of the solvated complex. The MD trajectory provides very convincing evidence for a rapid water-diffusion scenario (Fig. 4) and allows for assessment of diffusion pathways and residence times, which are concluded to be close to the lower end (nanoseconds) of the range given by the experiments.

No waters are found at the interface between protein and the DNA major groove in the GATA-1 DBD–DNA complex (Clore *et al*. 1994a) and this is consistent with close packing and nonpolar character of that region. Several waters were on the other hand found to reside in the more hydrophilic regions of the interface close to DNA phosphates. With only one exception these waters have residence times longer than 200–300 ps.

An issue which does not explicitly involve hydration of a protein–DNA interface studied by NMR, but which nonetheless is relevant to the subject is the role of DNA hydration and binding site recognition. This discussion has concerned the *trp* operator and its recognition by the *trp* repressor. It was sparked by the first crystal structure of the complex in which nearly all contacts between DNA and protein were mediated by waters (Otwinowski *et al*. 1988). These waters are present at equivalent positions in different repressor/operator complexes and mutations designed to remove water sites result in severely reduced protein–DNA binding affinities (see e.g. Smith *et al*. 1994). In particular, it was suggested that DNA major groove waters constitute integral elements of the DNA structure recognized by

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

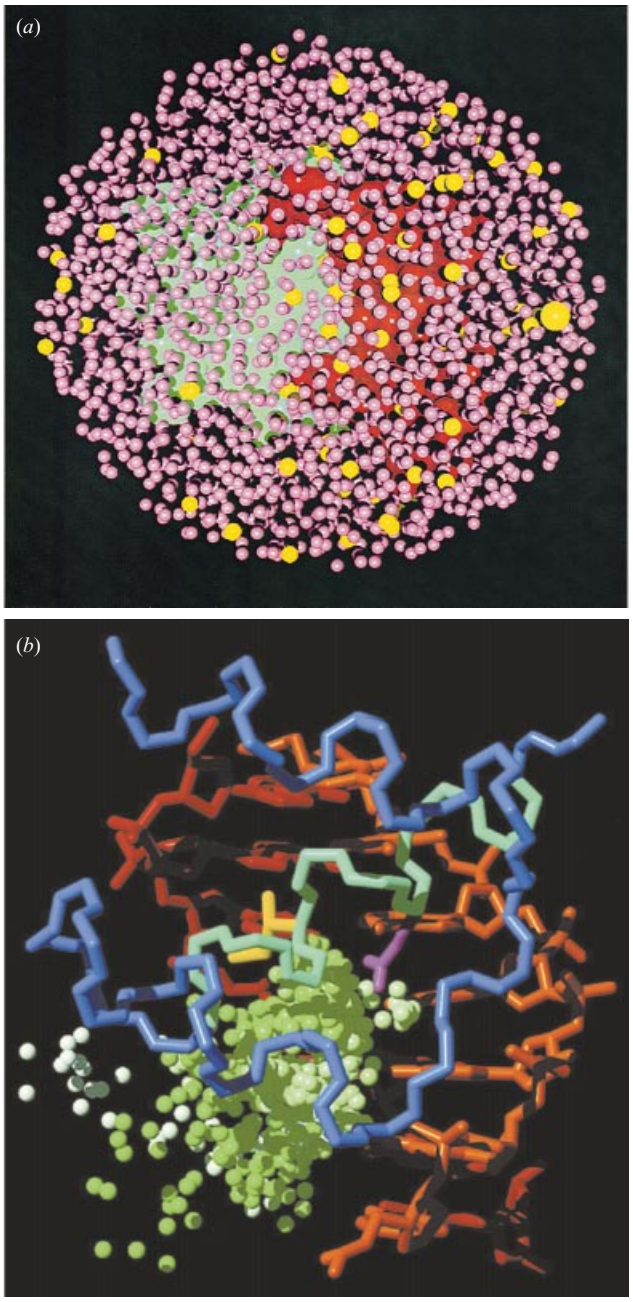


Fig. 4. Water diffusion in and out of a protein–DNA complex interface. (*a*) Starting configuration for the molecular dynamics simulation of the *Antennapedia* homeodomain–DNA complex surrounded by 2714 water molecules reported by Billeter *et al*. (1996). The large yellow sphere to the right indicates the starting position for the water followed in (*b*). (*b*) Trajectory for a water molecule: a 660 ps time period of the 2 ns MD trajectory. The water sphere is colored as a function of the simulation time from green at 840 ps to white at 1500 ps with a time resolution of one sphere per picosecond. (Figure reprinted with copyright permission by Cell Press.)

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

the repressor (Shakked *et al*. 1994). These data called for more detailed studies of the positions and residence times of waters in the major groove of the *trp* operator in solution. Using both NOE/ROE and NMRD measurements, Otting's laboratory found no evidence for residence times longer than sub-nanoseconds in the DNA major groove (Sunnerhagen *et al*. 1998). They also observe that the location of waters in two different *trp* repressor–operator complexes are much more similar than in the complexed compared to uncomplexed operators. Sunnerhagen *et al*. therefore find it unlikely that the hydration of the uncomplexed DNA target is a major determinant for recognition by the *trp* repressor, but that water instead diffuses into appropriate sites as part of the recognition. The conclusion regarding occupancies were further corroborated using MD simulations (Bonvin *et al*. 1998).

Finally, it is important to note that the inclusion of a few or several waters at a protein–DNA interface is not inconsistent with the understanding that surface desolvation effects dominate the thermodynamic profile for complex formation (Lundbäck & Härd, 1996). This is because several hundred waters are expected to leave the interacting surfaces as part of the protein–DNA association process and this results in the large compensatory enthalpic and entropic effects that signify desolvation (hydrophobic effect). The few remaining waters might on the other hand act to tip the compensation balance towards a stable complex by forming bridging hydrogen bonds.

5.2 *Dynamics in a complex – where does it occur and at what time scales?*

Studies of rapidly exchanging molecular conformations – be they more extensive as in loop rearrangements or more localized as in side chain rotamer inter-conversions – is an area where NMR provides superior information in comparison to X-ray crystallography. Still, this is an area in which discussions between representatives of the two fields sometimes have become infected. The most decisive evidence for dynamics in protein–nucleic acid complexes come from studies of complexes with known structures. The main conclusion is that side chain dynamics at the DNA–protein interface, as opposed to a packing in a protein core, is a reality in many, although perhaps not all, protein complexes. Below we describe the evidence in some more detail, following a brief introduction to the methodology.

5.2.1 *NMR studies of dynamics – time scales and techniques*

The NMR methodologies available to monitor biomolecular dynamics essentially fall into three categories. First, to assess motions that occur on time scales that are more rapid than overall molecular tumbling, the spectroscopist can measure and analyse a set of different magnetization and coherence relaxation rates. These give information on time scales (picosecond or nanosecond motions) and, in particular, average order parameters. The methodology has recently been reviewed by Palmer *et al*. (1996). The most common applications are measurements of backbone and side chain dynamics using ^{15}N nuclei as probes, but many other nuclei are also amenable for relaxation analysis.

Second, to assess motions that occur on longer time scales – microseconds to milliseconds – it is frequently sufficient to monitor NMR resonance line widths because these are inherently sensitive to changes in magnetic environment on intermediate time scales. The analysis can be made more quantitative by direct measurements of transverse magnetization relaxation rates (R_2) under different conditions; see Akke *et al*. (1998) and work cited therein. A convincing quantitative analysis must in this case consider the effects of rapid motions

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

described above and anisotropic tumbling (Tjandra *et al*. 1995). A common result is a conclusion that slow motions occur. (This is frequently referred to as 'exchange on the NMR time scale'.) A major drawback is that it is less easy to deduce precisely which the involved conformational states are. Another drawback is that it is frequently difficult to distinguish intra-molecular from inter-molecular exchange. For instance, the effects of fairly weak protein dimerization on R_2 rates can be large (Pfuhl *et al*. 1999). Without pointing out specific cases we propose that there are several studies in the literature in which weak transient binding by small-molecule impurities or transient protein dimerization have been mistaken for intra-molecular conformational exchange.

The third time range accessible for NMR encompass even slower motions for which the major detection strategy is backbone hydrogen exchange measurements (see e.g. Bai *et al*. (1995) and work cited therein). These can be made either indirectly by water saturation transfer for lifetimes in the range 0·1–10 s (Neuhaus & Williamson, 1989), or directly for lifetimes longer than several minutes. In favourable cases it is possible to directly determine the structures of conformational states that inter-convert on long time scales.

5.2.2 *Side chain and loop motions at protein–DNA interfaces*

The solution structure of an *Antennapedia* homeodomain–DNA complex, detailed in Section 3.1.1, provide evidence that the invariant Asn51, and possibly also Gln50, side chains in the protein are in a slow equilibrium between two or more different DNA-contacting conformations. The evidence is partly direct in that resonance line broadening is observed (Qian *et al*. 1993a), and partly indirect in that multiple conformations are consistent with NOE data (Billeter *et al*. 1993). Other side chains at the interface do not show line broadening and the effect cannot be inter-molecular because the lifetime of the complex is within the minutes time range. The dynamic scenario is strengthened by the evidence for rapid water diffusion at the interface (Section 5.1).

The dynamic view of the *Antennapedia* homeodomain–DNA interface has been contested in an article describing the structure determination of a very similar complex by X-ray crystallography (Fraenkel & Pabo, 1998). These authors claim that both NMR and X-ray data are consistent with a single average conformation. Although it may be true that the average structures are the same it cannot be ruled out, in our opinion, that conformational substates are populated. The distribution of substates might also be different in solution and crystal conditions. In fact, Fraenkel & Pabo do find two different conformations of the Gln50 side chains in one of the two complexes in the asymmetric unit. Moreover, the crystallographic structure relies on a model-dependent calculation of crystallographic phases. It might be worthwhile to also address the issue using multi-wavelength anomalous dispersion (MAD) phasing. The crystallographic structure of the *Antennapedia* homeodomain–DNA complex may be put in perspective considering the results from a high-resolution MAD phasing structure of rat mannose-binding protein A (Burling *et al*. 1996). In this latter study, an accurate image was obtained without resort to model-dependent phase determinations. The structure reveals multiple side chain conformations, ordered as well as disordered water, partly reduced disulphide bonds, etc., showing that such phenomena can be detected also in crystallographic structures.

A similar scenario appears to exist in the complex between the N-terminal three zinc fingers of TFIIIA (zf1–3) and DNA (Section 3.1.5). Here, NMR line broadening is observed for

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

several, but not all, lysine side chains that make specific contacts at the protein–DNA interface (Foster *et al*. 1997). The spectra shown in the paper by Foster *et al*. provide direct evidence for dynamic averaging on microsecond to millisecond time scales. Furthermore, simulated annealing does not provide well-defined conformations, but the different conformations that are observed are consistent with several possible hydrogen bond combinations. These data add to the body of evidence indicating a dynamic protein–DNA recognition interface. Resonance line-broadening for DNA-interacting side chains was also observed and in the Ets-1 DBD–DNA complex (Werner *et al*. 1997).

The line-broadening evidence imply motions that are slow compared to rotational tumbling of protein–DNA complexes in solution. Two studies address rapid motions at the interface by means of quantitative relaxation measurements. Berglund *et al*. (1995) studied arginine side chain dynamics at the interface of the nonspecific Sso7d–DNA complex. They found that dynamics of an arginine side chain at the interface of the DNA-complex is restricted compared to the free state of Sso7d, but not at all as restricted as in an arginine side chain which is hydrogen bonded within the core of a reference protein. They calculated that a full reduction of mobility for this single side chain would result in an entropic cost of 0·5–1·0 kcal mol^{-1} (at room temperature) which, in turn, corresponds to a reduction in the binding affinity by a factor of two to five. Side chain motions on both short (picosecond to nanosecond) and longer (microsecond to millisecond) time scales have also been concluded to occur in a *lac* repressor headpiece-DNA complex (Slijper *et al*. 1997).

The literature provides examples of dynamics that encompass larger segments of DNA-interacting peptides. Such motions have for instance been detected in studies of the Genesis (HNF-3/fkh) DBD–DNA complex (Jin *et al*. 1998). The HNF-3/fkh domains adopt a mixed α/β fold and bind DNA by several interactions involving helix side chains and the DNA major groove. In addition there are two loop regions (wings) that interact with the DNA minor groove. Wing 1, which is conserved within the family, was found to only make a single DNA contact whereupon it extends outside the DNA binding site in a crystal structure of the HNF-3γ DBD–DNA complex (Clark *et al*. 1993). Jin *et al*. (1998) used a four-base pair longer DNA fragment and find that wing 1 makes several contacts in the DNA minor groove, which explains why the loop is conserved. Still, relaxation data and in particular ^{1}H–^{15}N NOEs imply a large mobility of the loop in the complex. Protein mutagenesis followed by binding assays show that wing 1 contributes to binding affinity, but that the effect of mutations is not dramatic. Thus, it appears that wing 1 is a weak DNA binder which makes transient contacts with DNA in the complex.

5.2.3 *Implications for complex stability and biomolecular recognition*

Hence, the evidence for protein–nucleic acid interface dynamics in several complexes is firm. This does not necessarily mean that it occurs in all complexes. It may appear contradictory that a highly specific interface should be dynamic and/or conformationally flexible. There are, on the other hand, powerful arguments to why it should. Calculations show that loss of conformational freedom upon complex formation inevitably means loss of binding entropy (Akke *et al*. 1993; Yang & Kay, 1996). This loss might be compensated by favourable interaction energy contributing to binding enthalpy. Dunitz (1995) has put forward simple thermodynamic arguments to suggest that enthalpy–entropy compensation is a general

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

property of weak intermolecular interactions, and that the two contributions to the (binding) free energy should nearly balance each other for a hydrogen bond at 300 K. By the same arguments, the possibility of forming many different hydrogen bonds at a protein–DNA interface ought to reduce the entropic cost of binding while the enthalpic gain from a favourable hydrogen bond is present at all times. Furthermore, it is not inconceivable that a fine balance between rigidity and flexibility needs to be obtained in order to achieve specificity (at some entropic expense) without loosing too much affinity. This possibility is discussed by Neuhaus and co-workers (Allain *et al*. 1997) on basis of their own observations on the U1A–RNA complex and those of others on peptide–protein complexes (Kay *et al*. 1996). An interesting observation which relates to this discussion is that the inherent flexibility of the DNA phoshpodiester backbone is retained in specific tight-binding *lac* repressor headpiece–operator complexes whereas it is restricted in weak-binding protein–operator complexes (Botuyan *et al*. 1993; Karslake *et al*. 1992).

5.3 *Coupled folding: the formation or stabilization of protein structure upon binding to DNA*

An issue that is related to the dynamics in a complex, yet distinct from that discussion, concerns coupled folding and conformational adaptation upon biomolecular recognition and complex formation. It is becoming increasingly evident that Nature uses coupled folding and/or conformational dynamics as a tool to achieve specificity in various recognition processes. The generic coupled folding reaction is a protein peptide fragment which is completely unstructured in the free state, but which adopts a well-defined fold (frequently a helix) in contact with DNA in the complex. This is also the scenario in the earliest described cases: DNA-binding by the GCN4 (Weiss *et al*. 1990), C/EBP (Shuman *et al*. 1990) and Fos/Jun (Patel *et al*. 1990) transcription factors. However, in actuality there are normally many more subtle details to the story, as will become evident below.

NMR is particularly well suited to study these processes for several reasons: (i) there are techniques available with which site-specific dynamics can be monitored on at least three time scales, as described above; (ii) a partly unfolded peptide is not prone to crystallization; and (iii) if a protein in which for instance a loop region assumes multiple exchanging conformations indeed does crystallize, then it is not unlikely that the crystallization process or conditions for it alter the solution equilibrium of accessible states.

5.3.1 *DNA-nucleic acid interactions involving conformational changes probed by NMR*

Evidence for a dynamic disorder in the free state of a protein but a well-defined conformation in the complexed state, have for example been found in the following cases.

- The binding of the estrogen receptor (ER) DBD to cognate DNA sequences. Here, the solution NMR structure suggested that a large portion of a zinc binding subdomain was disordered in the free state (Schwabe *et al*. 1990b, 1993b) and this was later confirmed by backbone dynamics measurements (Wikström *et al*. 1999). The crystal structure contains two different DNA–DBD complexes in the asymmetric unit: one in which the disordered region is folded and makes DNA and protein–protein contacts, and one in which it forms a loop (Schwabe *et al*. 1993a). The latter was interpreted as a folding intermediate. The same observation was made in a complex between the ER DBD and an alternative cognate DNA sequence (Schwabe *et al*. 1995).

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

- The binding of the retinoid X receptor (RXR) DBD to DNA. The RXR DBD is homologous to the ER DBD. A similar scenario of folding of the second zinc domain was concluded to occur based on dynamics measurements on free and DNA-bound forms (Holmbeck *et al*. 1998a), and comparison of the NMR solution structure (Holmbeck *et al*. 1998b) to the crystallographic structures of a DNA complex (Rastinejad *et al*. 1995). Notably, the NMR data also show that a C-terminal portion of the RXR DBD *unfolds* upon DNA complex formation (Holmbeck *et al*. 1998a,b).
- Dimeric DNA binding by the *lac* repressor headpiece to the full *lac* operator results in structuring of C-terminal hinge helices (Lewis *et al*. 1996; Spronk *et al*. 1996). These are stabilized by mutual interactions within the bound dimer, because monomeric binding does not induce helix formation. The data add to the understanding of how an inducer molecule (e.g. IPTG) causes the full *lac* repressor–operator complex to dissociate. Binding of inducer to the repressor core makes the hinge helices move apart by about 3.5 Å. The favourable dimerization interactions are then lost and the overall DNA affinity is attenuated, which makes the repressor come off the operator. The re-association process might then be less efficient in competition with other DNA-binding proteins (Spronk *et al*. 1996).
- In the DNA-binding of the N-terminal three zinc fingers of TFIIIA, stable inter-domain interactions are found in the bound state (Foster *et al*. 1998; Wuttke *et al*. 1997), as discussed above. NMR data provide strong evidence that the finger domains behave as independent dynamic units in the free state, i.e. that they are mutually disordered (Brüschweiler *et al*. 1995). A similar situation exists for the free and bound states of the HIV-1 NC protein (De Guzman *et al*. 1998).
- NMR structure determinations of the free and DNA-bound states of the NFATC1 DBD provides a classical and well-documented picture of both a disorder–order transition and a stabilization of a partially populated conformation. In the free state two long loops which connect β-strands are either partly ordered or disordered, respectively (Wolfe *et al*. 1997). In the DNA complex the first of these assume a well-defined DNA-interacting conformation, whereas the second loop forms a compact fold containing an α-helix (Fig. 3). The helix is stabilized directly by interactions involving N-terminal amide groups, additionally supported by side chain–DNA interactions, and oriented so that the helix dipole–DNA conformation is electrostatically favourable (Zhou *et al*. 1998).
- A short, but rigid, helix is found at the C-terminal of the Genesis (HNF-3/fkh) DBD complexed with DNA (Jin *et al*. 1999), whereas the corresponding fragment of the free protein is a relatively mobile coil (Marsden *et al*. 1998). In this case the newly folded unit does not contact the DNA, but the following C-terminal (wing 2) residues do. Thus it appears that the wing 2–DNA interactions and the helix stabilize each other in the DNA-bound state. This is consistent with very elegant mutational data which show that helix-disrupting mutations (to Pro or Gly residues) accelerate dissociation from DNA, while helix-stabilizing mutations (to Ala) have the opposite effect (Shiyanova & Liao, 1999). The authors suggest that protein–DNA dissociation rates are encoded in the sequence of the helix-forming peptide.
- The recognition mechanism for U1A binding to its RNA target is imagined to involve two conceptually different contacts which may occur concurrently. The first recognition contact is made by variable loops in the U1A RNP domain and well-structured helical regions of RNA. The second contact requires a reorganization of the C-terminal region of

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

the protein so that the full RNA-interacting site is exposed as well as folding of a single-stranded RNA loop (Allain *et al*. 1996, 1997). These two recognition events explain both the requirement for a basic RNA structural integrity and that nucleotides in a disordered RNA fragment are required for a fine-tuning of the recognition (Allain *et al*. 1996).

There are cases where the nature of interaction-induced change is more subtle as it involves stabilization of a weakly folded domain or transiently populated state, or slowly interconverting states:

- In the *trp* repressor there are two helical segments (the D and E) helices which adopt the typical helix-turn-helix motif in tryptophan-repressor–DNA complexes (Zhang *et al*. 1994). It has been concluded that the free holorepressor D helix, although less stable than in the protein core, assumes two or more different conformational states in which the helical conformation is retained (Gryk *et al*. 1995). (See also the discussion of off-rates, and ligand binding and destabilization kinetics in Section 3.1.2.)
- The DNA-binding domain of the proto-oncogene product c-Myb protein contains three imperfect tandem repeats (R1, R2 and R3) which contain variations of the helix-turn-helix motif. Solution structures of the c-Myb DBD has been determined in both free and DNA-complexed states (Ogata *et al*. 1992, 1994). It has been noted that the R2 repeat is much less stable than the R1 and R3 repeats (Sarai *et al*. 1993). The destabilization is due to a cavity in the free state of R2, which is closed by a conformational change upon DNA-binding (Ogata *et al*. 1995). In some very elegant experiments Ogata *et al*. demonstrate that the free R2 region is dynamic on the microsecond to millisecond time scale (Ogata *et al*. 1996). A cavity-filling mutation stabilizes the structure, but the mutant protein binds less well to DNA than wild type c-Myb DBD. Ogata *et al*. conclude that the cavity is reserved for DNA recognition and that the core is optimized for the DNA-bound rather than free protein.
- Slow conformational dynamics might vary in the free states of homologous DNA-binding domains, and such a variation can parallel DNA binding specificity. In addition to the rapid motions and coupled folding in RXR and ER DBDs described above, there is a connection between the presence of slow conformational mobility or time scales of milliseconds to microseconds, or longer, in nuclear receptor DBDs and the DNA sequences which they recognize. The glucucorticoid receptor (GR) DBD is quite rigid in the free state and it binds as to an AGAACA site on DNA, whereas the homologous ER, RXR and retinoic acid receptor (RAR) DBDs which all show substantial picosecond and/or microsecond to millisecond mobility bind to the related AGGTCA site; see Wikström *et al*. (1999) for complementary references. An intermediate between these two classes is a triple GR DBD mutant which prefers a AGGTCA site. This mutant shows lower stability in the same region of the molecule that are flexible in the ER, RXR, and RAR DBDs, but not in the wild type GR DBD (Berglund *et al*. 1997). Moreover, despite structural and dynamical differences in the free states of the nuclear receptor DBDs they all adopt nearly identical structures in the bound state. Wikström *et al*. (1999) propose that a key to the DNA recognition in this case is the ability of DNA to allosterically induce the optimal bound state conformation.

This compilation, which is not comprehensive, illustrates that conformational changes and changes in dynamics are extremely common in protein–nucleic acid interactions. A similar

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

inventory of crystallographic structures would add many additional examples to the list, although changes in dynamics are less easily observed with crystallography, as mentioned above. Still, there are also examples of proteins studied by NMR that do not change their conformations upon DNA binding. The more rigid DNA-binding domains include the nonspecific DNA-binding proteins Sso7d (Agback *et al*. 1998) and NHP6A (HMG domain) (Allain *et al*. 1999). These do, on the other hand, instead induce conformational changes in the DNA.

5.3.2 *Function of coupled folding in biomolecular recognition*

Hence, it appears that folding and/or conformational change is a rule rather than an exception in sequence-specific protein–nucleic acid interactions. In fact, a similar situation appears to exist also with eukaryotic transcription activation domains, which typically are unstructured or partially unstructured in solution, but fold upon interaction with the correct target; see references given in Bagby *et al*. (1998). The general functional significance of conformational changes in recognition processes is not understood, although insightful explanations have been given in many of the specific cases, such as with U1A (Allain *et al*. 1996, 1997) and the Genesis DBD (Jin *et al*. 1999; Shiyanova & Liao, 1999). Additional folding and conformational change are processes which are unfavourable in terms of free energy of uncomplexed proteins. This free energy loss is (over)compensated by additional interactions that occur over large protein–nucleic acid interaction surfaces. Still, one would expect a complex between pre-formed surfaces to be stronger than one in which one of the surfaces must be made in the interaction process. An answer to why surfaces are not pre-formed then presumably requires arguments based on kinetics, as described in the following.

A more general mechanism that could account for the omnipresent occurrence of coupled folding processes in DNA-binding and activation of transcription is that Nature might not benefit from having too many fully functional surfaces present in the cell at any one time, but that the correct surface should be formed only upon allosteric action by the correct effector. One can envision two practical consequence in terms of protein–nucleic acid interactions. First, there is normally a very large number of potential binding sites available for a nucleic acid-binding protein. There is therefore a need to efficiently explore these in a non-committal manner in order to find the correct one. This might be accomplished by a partly folded protein in which small number of exposed DNA-interacting groups ensure a rapid off-rate. The protein attempts to fold also at these sites, but the additional structure is not sufficiently stabilized. However, the appropriate DNA site, and only this site, will provide a sufficiently matching surface to stabilize the fully folded protein. Some part of that interaction free energy gain is then also used to obtain a tighter and more long-lived complex.

Second, the allosteric induction of additional structure might form a new surface which, in turn, can perform some function that is not required, and even unwanted, at any other site than at the correct DNA promoter. For instance, newly made surfaces can act to recruit other factors to the complex or induce a conformational distortion of DNA or RNA. Recruited proteins can be a second monomer in order to form a dimer on DNA, or other factors of a transcription initiation complex.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

6. Concluding remarks

An inventory of structures of protein–nucleic acid complexes shows that this is an area of structural biology where NMR, after a slow start, is now ready to make substantial contributions. This is, in particular, due to the many methodological improvements made in the past 10 years, which have made it possible to study also quite large complexes. Novel methodology such as TROSY, long-range relative orientations measured by dipolar couplings and direct evidence for hydrogen bonds can be expected to increase precision and size of the studied systems even more. The structures in the inventory contain many protein–DNA and peptide–RNA complexes, but protein–RNA and nonspecific protein–DNA complexes are under-represented. The former might be the area in which NMR will make the largest contributions. This is because protein–RNA are more difficult to crystallize using conventional methods, whereas RNA fragments are rather easily obtained in isotope enriched forms to facilitate NMR measurements.

We have reviewed the biophysical evidence for solvation, protein side chain and loop dynamics, and coupled folding and conformational stabilization and conclude that NMR methods in many cases have provided profound insights into these issues. We find that coupled folding or conformational change in some form contributes to almost all protein–nucleic acid recognition events in solution. We propose that coupled folding might be devised by Nature to increase kinetics for dissociation of nonspecific complexes and also achieve a precise localization of functional surfaces, i.e. at a certain DNA promoter region.

As for the future, it can be expected that biology and biotechnology should continue to provide the objectives and priorities for research in NMR. There will be three important areas for biomolecular NMR in the post-Structural Genomics era when all individual protein structures, or at least their folds, will be known: structures of complexes, mechanisms of recognition, and drug design. For efficient and competitive structure determination there is a need for optimization of the current methodology. Ideally, the number of necessary experiments should be minimized to obtain a smaller number of high-information content spectra from which structures can be determined using automated or semi-automated procedures. To understand more about the physical events that accompany molecular recognition it might be advisable to focus on a few model systems, which is now also what is being done in groups working on for instance the Genesis, c-Myb and nuclear receptor DBDs. The last issue – drug design – has great promise, given novel methods like SAR-by-NMR (Shuker *et al.* 1996).

7. Acknowledgements

Dr Helena Berglund is acknowledged for constructive comments on the manuscript. Dr Mary Starich and Dr Pei Zhou provided high-resolution images of Figs 2 and 3, respectively. The author also acknowledges sponsors of his own research mentioned in this review: the Center for Structural Biochemistry (CSB) at Novum, the Swedish Natural Sciences Research Council (NFR), the Swedish Strategic Research Foundation (SSF), and the Knut and Alice Wallenberg Foundation.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

8. References

Abdul-Manam, N., O'Malley, S. M. & Williams, K. R. (1996). Origins of the binding specificity of the A1 heterogenous ribonucleoprotein. *Biochemistry* **35**, 3545–3554.

Agback, P., Baumann, H., Knapp, S., Ladenstein, R. & Härd, T. (1998). Architecture of nonspecific protein–DNA interactions in the Sso7d–DNA complex. *Nature Struct. Biol.* **5**, 579–584.

Aggarwal, A. K. (1990). Crystallization of DNA binding proteins with oligodeoxynucleotides. *Methods* (compl. to *Meth. Enzymol.*) **1**, 83–90.

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907.

Akke, M., Brüschweiler, R. & Palmer, A. G. (1993). NMR order parameters and free energy: an analytical approach and its application to cooperative Ca2+ binding by calbindin D_9k. *J. Am. Chem. Soc.* **115**, 9832–9833.

Akke, M., Liu, J., Cavanagh, J., Erickson, H. P. & Palmer, A. G. (1998). Pervasive conformational fluctuations on microsecond time scales in a fibronectin type III domain. *Nature Struct. Biol.* **6**, 55–58.

Allain, F. H., Yen, Y. M., Masse, J. E., Schultze, P., Dieckmann, T., Johnson, R. C. & Feigon, J. (1999). Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding. *EMBO J.* **18**, 2563–2579.

Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**, 646–650.

Allain, F. H.-T., Howe, P. W. A., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **18**, 5764–5772.

Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. & Suzuki, M. (1998). A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* **17**, 5484–5496.

Alma, N. C., Harmsen, B. J., Hull, W. E., van der Marel, G., van Boom, J. H. & Hilbers, C. W. (1981). Double-resonance experiments at 500 MHz on gene-5 protein and its complex with octadeoxyribo-adenylic acid. *Biochemistry* **20**, 4419–4428.

Altmann, S., Labhardt, A. M., Bur, D., Lehmann, C., Bannwarth, W., Billeter, M., Wüthrich, K. & Leupin, W. (1995). NMR studies of duplexes singly cross-linked by different synthetic linkers. *Nucl. Acids Res.* **23**, 4827–4835.

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290**, 754–758.

Arrowsmith, C. H., Czaplicki, J., Iyer, S. B. & Jardetzky, O. (1991). Unusual dynamic features of the trp repressor from *Escherichia coli*. *J. Am. Chem. Soc.* **113**, 4020–4022.

Bagby, S., Arrowsmith, C. H. & Ikura, M. (1998). New perceptions of transcription factor properties from NMR. *Biochem. Cell Biol.* **76**, 368–378.

Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192–197.

Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995). Preparation of isotopically enriched RNAs for heteronuclear NMR. *Meth. Enzymol.* **261**, 300–322.

Batey, R. T., Cloutier, N., Mao, H. & Williamson, J. R. (1996). Improved large scale culture of Methylophilus methylotrophus for 13C/15N labelling and random fractional deuteration of ribonucleotides. *Nucl. Acids Res.* **24**, 4836–4837.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1994). Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine–purine base pairs. *Biochemistry* **33**, 2741–2747.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1995). Assignment and modelling of the Rev Response Element RNA bound to a Rev peptide using ^{13}C-heteronuclear NMR. *J. biomol. NMR* **6**, 375–389.

Battiste, J. L., Mao, H., Rao, S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α-helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE complex. *Science* **273**, 1547–1551.

Bax, A. (1994). Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. struct. Biol.* **4**, 738–744.

Bax, A. (1996). NMR of ethanol and interferon-g. In Encyclopedia of NMR, vol. 1, (eds. R. K. Harris & D. M. Grant), pp. 202–207, Chichester: John Wiley & Sons.

Berglund, H., Baumann, H., Knapp, S., Ladenstein, R. & Härd, T. (1995). Flexibility of an arginine side chain at a DNA-protein interface. *J. Am. chem. Soc.* **117**, 12883–12884.

Berglund, H., Wolf-Watz, M., Lundbäck, T., van den Berg, S. & Härd, T. (1997). Structure and dynamics of the glucocortiocid receptor DNA-binding domain: comparison of wild type and a mutant with altered specificity. *Biochemistry* **36**, 11188–11197.

Bewley, C. A., Gronenborn, A. M. & Clore, G. M.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

(1998). Minor groove-binding architectural proteins: structure, function and DNA recognition. *Ann. Rev. Biophys. biomol. Struct.* **27**, 105–131.

Billeter, M., Güntert, P., Luginbühl, P. & Wüthrich, K. (1996). Hydration and DNA recognition by homeodomains. *Cell* **85**, 1057–1065.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. mol. Biol.* **234**, 1084–1093.

Billeter, M. & Wüthrich, K. (1993). Model studies relating nuclear magnetic resonance data with the three-dimensional structure of protein–DNA complexes (Appendix to *J. mol. Biol.* **234**, 1084–1093 (1993)). *J. mol. Biol.* **234**, 1094–1097.

Boelens, R., Scheek, R. M., van Boom, J. H. & Kaptein, R. (1987). Complex of lac repressor headpiece with a 14 base-pair lac operator fragment studied by two-dimensional nuclear magnetic resonance. *J. mol. Biol.* **193**, 213–216.

Bonvin, A. M. J. J., Sunnerhagen, M., Otting, G. & van Gunsteren, W. F. (1998). Water molecules in DNA recognition II: a molecular dynamics view of the structure and hydration of the trp operator. *J. Mol. Biol.* **282**, 859–873.

Bothner-By, A. A. (1996). Magnetic field induced alignment of molecules. In Encyclopedia of NMR, (eds. R. K. Harris & D. M. Grant), pp. 2932–2938, Chichester: John Wiley & Sons.

Botuyan, M. V., Keire, D. A., Kroen, C. & Gorenstein, D. G. (1993). 31P nuclear magnetic resonance spectra and dissociation constants of lac repressor headpiece-duplex operator complexes: the importance of phosphate ester backbone flexibility in protein–DNA recognition. *Biochemistry* **32**, 6863–6874.

Boublik, K., Bradbury, E. M., Crane-Robinson, C. & Rattle, H. W. (1971). Proton magnetic resonance studies of the interactions of histones F1 and F2B with DNA. *Nature New Biol.* **229**, 149–150.

Bradbury, E. M., Cary, P. D., Crane-Robinson, C., Rattle, H. W., Boublik, M. & Sautière, P. (1975). Conformations and interactions of histone H2A (F2A2, ALK). *Biochemistry* **14**, 1876–1885.

Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R. & Kaptein, R. (1990). Structure of Arc repressor in solution: evidence for a family of b-sheet DNA-binding proteins. Biochemistry 28, 9826–9833.

Brünger, A. (1992). *X-PLOR version 3.1: A system for X-ray crystallography and NMR*. Yale University Press.

Brünger, A. T. & Nilges, M. (1993). Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Quart. Rev. Biophys.* **26**, 49–125.

Brüschweiler, R., Liao, X. & Wright, P. E. (1995). Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886–889.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science* , 72–77.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.* **4**, 567–577. [Published erratum appears in *Nature Struct. Biol.* (1997) **10**, 839–840.]

Cai, Z., Gorin, A., R., F., Ye, X., Hu, W., Majumdar, A., Kettani, A. & Patel, D. J. (1998). Solution structure of P22 transcriptional antitermination N peptide-box B RNA complex. *Nature Struct. Biol.* **5**, 203–212.

Cary, P. D., Crane-Robinson, C., Bradbury, E. M., Javherian, K., Goodwin, G. H. & Johns, E. W. (1976). Conformational studies of non-histone chromosomal protein and their interactions with DNA. *Eur. J. Biochem.* **62**, 583–590.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. (1996). *Protein NMR spectroscopy: Principles and practice*. London: Academic Press.

Chuprina, V. P., Rullmann, J. A., Lamerichs, R. M., van Boom, J. H., Boelens, R. & Kaptein, R. (1993a). Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. mol. Biol.* **234**, 446–462.

Chuprina, V. P., Rullmann, J. A., Lamerichs, R. M., van Boom, J. H., Boelens, R. & Kaptein, R. (1993b). Structure of the complex of *lac* repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. mol. Biol.* **234**, 446–462.

Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). Cocrystal structure of the HNF-3/fork head DNA recognition motif resembles histone H5. *Nature* **364**, 412–420.

Clore, G. M., Bax, A., Omichinski, J. G. & Gronenborn, A. M. (1994a). Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. *Structure* **2**, 89–94.

Clore, G. M., Bax, A., Omichinski, J. G. & Gronenborn, A. M. (1994b). Localization of bound water in the solution structure of the erythroid transcription factor GATA-1 with DNA. *Structure* **15**, 89–94.

Clore, G. M. & Gronenborn, A. M. (1992). Struc-

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

tures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399.

Clore, G. M. & Gronenborn, A. M. (1998a). New methods of structure refinement for macromolecular structure determinations by NMR. *Proc. natl. Acad. Sci. USA* **95**, 5891–5898.

Clore, G. M. & Gronenborn, A. M. (1998b). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2**, 564–570.

Coleman, J. E. & Armitage, I. M. (1978). Tyrosyl-base-phenylalanyl intercalation in gene 5 protein–DNA complexes: proton nuclear magnetic resonance of selectively deuterated gene 5 protein. *Biochemistry* **17**, 5038–5045.

Connolly, K. M., Wojciak, J. M. & Clubb, R. T. (1999). Resonance assignments of the Tn*916* integrase DNA-binding domain and the integrase: DNA complex. *J. biomol. NMR* **14**, 95–96.

Cordier, F. & Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J. Am. Chem. Soc.* **121**, 1601–1602.

Cornilescu, G., Ramirez, B. E., Frank, M. K., Clore, G. M., Gronenborn, A. M. & Bax, A. (1999). Correlation between (3h)J(NC′) and hydrogen bond length in proteins. *J. Am. Chem. Soc.* **121**, 6275–6279.

De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Θ-RNA recognition element. *Science* **279**, 384–388.

Denisov, V. P. & Halle, B. (1995). Protein hydration dynamics in aqueous solution: a comparison of bovine pancreatic trypsin inhibitor and ubiquitin by oxygen-17 spin relaxation dispersion. *J. mol. Biol.* **245**, 682–697.

Denisov, V. P., Peters, J., Hörlein, H. D. & Halle, B. (1996). Using buried water molecules to explore the energy landscape of proteins. *Nature Struct. Biol.* **3**, 505–509.

Dötsch, V. & Wagner, G. (1998). New approaches to structure determination by NMR spectroscopy. *Curr. Opin. struct. Biol.* **8**, 619–623.

Dunitz, J. D. (1995). Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions. *Chem. & Biol.* **2**, 709–712.

Feigon, J. & Kearns, D. R. (1979). 1H NMR investigation of the conformational states of DNA in nucleosome core particles. *Nucl. Acids Res.* **6**, 2327–2337.

Foster, M. P., Wutke, D. S., Radhakrishnan, I., Case, D. A., Gottefeld, J. M. & Wright, P. E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature Struct. Biol.* **4**, 605–608.

Foster, M. P., Wuttke, D. S., Clemens, K. R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J. & Wright, P. E. (1998). Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *J. biomol. NMR* **12**, 51–71.

Fraenkel, E. & Pabo, C. O. (1998). Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. *Nature Struct. Biol.* **5**, 692–697.

Gardner, K. H. & Kay, L. E. (1998). The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. biomol. Struct.* **27**, 357–406.

Garssen, G. J., Kaptein, R., Schoenmakers, J. G. & Hilbers, C. W. (1978). A photo-CIDNP study of the interaction of oligonucleotides with gene-5 protein of bacteriophage M13. *Proc. natl. Acad. Sci. USA* **75**, 5281–5285.

Geierstanger, B. H. & Wemmer, D. E. (1995). Complexes of the minor groove of DNA. *Annu. Rev. biomol. Struct.* **24**, 463–493.

Gerothanassis, I. P. (1994). Multinuclear and multidimensional NMR methodology for studying individual water molecules bound to peptides and proteins in solution: principles and applications. *Prog. NMR Spectr.* **26**, 171–237.

Gilbert, W. & Müller-Hill, B. (1966). Isolation of the lac repressor. *Proc. natl. Acad. Sci. USA* **56**, 1891–1898.

Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1999). The three-dimensional structure of the vnd/NK-2 homeodomain-DNA complex by NMR spectroscopy. *J. Mol. Biol.* **289**, 529–545.

Gryk, M. R., Finucane, M. D., Zheng, Z. & Jardetzky, O. (1995). Solution dynamics of the trp repressor: a study of amide proton exchange by T1 relaxation. *J. mol. Biol.* **246**, 618–627.

Gubser, C. C. & Varani, G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry* **35**, 2253–2267.

Güntert, P., Mumenthaler, C. & Wüthrich, K. (1998). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. mol. Biol.* **273**, 283–298.

Härd, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J.-Å. & Kaptein, R. (1990). Solution structure of the gluco-

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

corticoid receptor DNA-binding domain. *Science* **249**, 157–160.

Härd, T. & Lundbäck, T. (1996). Thermodynamics of sequence-specific protein–DNA interactions. *Biophys. Chem.* **62**, 121–139.

Hélène, C. & Lancelot, G. (1982). Interactions between functional groups in protein–nucleic acid associations. *Prog. Biophys. molec. Biol.* **39**, 1–68.

von Hippel, P. H. (1994). Protein–DNA recognition: new perspectives and underlying themes. *Science* **263**, 769–770.

von Hippel, P. H. & McGhee, J. D. (1972). DNA–protein interactions. *Annu. Rev. Biochem.* **41**, 231–294.

Holmbeck, S. M. A., Dyson, H. J. & Wright, P. E. (1998a). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J. mol. Biol.* **284**, 533–539.

Holmbeck, S. M. A., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J. & Wright, P. E. (1998b). High resolution solution structure of the retinoid X receptor DNA binding domain. *J. mol. Biol.* **281**, 271–284.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). Errors in protein structures. *Nature* **381**, 272.

Howe, P. W., Allain, F. H., Varani, G. & Neuhaus, D. (1998). Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J. biomol. NMR* **11**, 59–84.

Huth, J. R., Bewley, C. A., Jackson, B. M., Hinnebusch, A. G., Clore, G. M. & Gronenborn, A. M. (1997). Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* **6**, 2359–2364.

Ikura, M., Kay, L. E. & Bax, A. (1990). A novel approach for sequential assignment of H-1, C-13, and N-15 spectra of larger proteins – heteronuclear triple-resonance 3-dimensional NMR-spectroscopy – application to calmodulin. *Biochemistry* **19**, 4659–4667.

Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. mol. Biol.* **3**, 318–356.

Jin, C., Marsden, I., Chen, X. & Liao, X. (1998). Sequence specific collective motions in a winged helix DNA binding domain detected by ^{15}N relaxation NMR. *Biochemistry* **37**, 6179–6187.

Jin, L., Fukayama, J. W., Pelczer, I. & Carey, J. (1999). Long-range effects on dynamics in a temperature-sensitive mutant of *trp* repressor. *J. mol. Biol.* **285**, 361–378.

Kainosho, M. (1997). Isotope labelling of macromolecules for structural determinations. Nature Struct. *Biol. Suppl.* **4**, 858–861.

Kaptein, R., Zuiderweg, E. R., Scheek, R. M., Boelens, R. & vanGunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J. Mol. Biol.* **182**, 179–182.

Karslake, C., Botuyan, M. V. & Gorenstein, D. G. (1992). 31P NMR spectra of oligodeoxyribonucleotide duplex lac operator–repressor headpiece complexes: importance of phosphate ester backbone flexibility in DNA–protein recognition. *Biochemistry* **31**, 1849–1858.

Kay, L. E. (1995). Pulsed field gradient multidimensional NMR methods for the study of protein structure and dynamics in solution. *Prog. Biophys. molec. Biol.* **63**, 277–299.

Kay, L. E., Muhandiram, D. R., Farrow, N. A., Aubin, Y. & Forman-Kay, J. D. (1996). Correlation between dynamics and high affinity binding in an SH2 domain interaction. *Biochemistry* **35**, 361–368.

Kim, Y. C., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990). Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science* **249**, 1307–1309.

King, G. C. & Coleman, J. E. (1987). Two-dimensional 1H NMR of gene 5 protein indicates that only two aromatic rings interact significantly with oligodeoxynucleotide bases. *Biochemistry* **26**, 2929–2937.

Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H., Kaptein, R., Buck, F., Fera, B. & Rüterjans, H. (1989). ^{1}H NMR study of a complex between the lac repressor headpiece and a 22 base pair symmetric *lac* operator. *Biochemistry* **28**, 2985–2991.

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. biomol. NMR* **8**, 477–486.

Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. (1998). NMR structure of the bacteriophage l N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**, 289–299.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Shumacher, M. A., Brennan, R. G. & Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254.

Liang, H., Mao, X., Olejniczak, E. T., Nettesheim, D. G., Yu, L., Meadows, R. P., Thompson, C. B. & Fesik, S. W. (1994). Solution structure of the *ets* domain of Fli-1 when bound to DNA. *Nature Struct. Biol.* **1**, 871–875.

Linge, J. P. & Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures:

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

A new force field for NMR structure calculation. *J. biomol. NMR* **13**, 51–59.

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.

Lundbäck, T., Hansson, H., Knapp, S., Ladenstein, R. & Härd, T. (1998). Thermodynamic characterization of nonsequence specific DNA-binding by the Sso7d protein from Sulfolobus solfataricus. *J. mol. Biol.* **276**, 775–786.

Lundbäck, T. & Härd, T. (1996). Sequence specific DNA-binding dominated by dehydration. *Proc. natl. Acad. Sci. USA* **93**, 4754–4759.

Marion, D., Driscoll, P., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989a). Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann–Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry* **28**, 6150–6156.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b). Three-dimensional heteronuclear NMR of ^{15}N-labelled proteins. *J. Am. Chem. Soc.* **111**, 1515–1517.

Marsden, I., Jin, C. & Liao, X. (1998). Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. *J. Mol. Biol.* **278**, 293–299.

Masse, J. E., Allain, F. H.-T., Yen, Y.-M., Johnson, R. C. & Feigon, J. (1999). Use of ^{13}C,^{15}N-labelled DNA in a non-sequence-specific protein–DNA complex resolves ambiguous assignments of intermolecular NOEs. *J. Am. Chem. Soc.* **121**, 3547–3548.

Masse, J. E., Bortmann, P., Dieckmann, T. & Feigon, J. (1998). Simple, efficient protocol for enzymatic synthesis of uniformly ^{13}C,^{15}N-labelled DNA for heteronuclear NMR studies. *Nucl. Acids Res.* **26**, 2618–2624.

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. & Takeda, Y. (1982). Structure of the DNA-binding region of lac repressor inferred from its homology with the cro repressor. *Proc. natl. Acad. Sci. USA* **79**, 1428–1452.

McClarin, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986). Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å. *Science* **234**, 1526–1541.

McKay, D. B. & Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* **290**, 744–749.

Mer, G. & Chazin, W. J. (1998). Enzymatic synthesis of region-specific isotope-labelled DNA oligomers for NMR analysis. *J. Am. Chem. Soc.* **120**, 607–608.

Milligan, J. F. & Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.* **180**, 51–62.

Morellet, N., Deŕméné, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournié-Zaluski, M. C. & Roques, B. P. (1998). Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J. Mol. Biol.* **283**, 419–434.

Nagai, K., Oubridge, C., Ito, N., Avis, J. & Evans, P. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *TIBS* **20**, 235–240.

Neuhaus, D. & Williamson, M. (1989). *The nuclear Overhauser effect in structural and conformational analysis*. New York, NY: VCH Publishers, Inc.

Nick, H., Arndt, K., Boschelli, F., Jarema, M. A., Lillis, M., Sadler, J., Caruthers, M. & Lu, P. (1982). *lac* repressor-*lac* operator interaction: NMR observations. *Proc. natl. Acad. Sci. USA* **79**, 218–222.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S. & Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. natl. Acad. Sci. USA* **89**, 6428–6432.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y., Ishii, S. & Sarai, A. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature Struct. Biol.* **3**, 178–187.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S. & Nishimura, Y. (1995). Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nature Struct. Biol.* **2**, 309–319.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. & Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell* **79**, 639–648.

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.

Omichinski, J. G., Pedone, P. V., Felsenfeld, G.,

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

Gronenborn, A. M. & Clore, G. M. (1997). The solution structure of a specific GAGA factor–DNA complex reveals a modular binding mode. *Nature Struct. Biol* **4**, 122–132.

Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. (1999). Improved segmental isotope labelling of proteins and application to a larger protein. *J. biomol. NMR* **14**, 105–114.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990a). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990b). Protein-DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **335**, 321–329.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438.

Pabo, C. O. & Lewis, M. (1982). The operator-binding of l repressor: structure and DNA recognition. *Nature* **298**, 443–447.

Palmer, A. G., Williams, J. & McDermott, A. (1996). Nuclear magnetic resonance studies of biopolymer dynamics. *J. phys. Chem.* **100**, 13293–13310.

Parsegian, V. A., Rand, R. P. & Rau, D. C. (1995). Macromolecules and water: Probing with osmotic stress. *Meth. Enzymol.* **259**, 43–94.

Patel, L., Abate, C. & Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347**, 572–575.

Patel, D. J. (1997). Structural analysis of nucleic acid aptamers. *Curr. Opin. Chem. Biol.* **1**, 32–46.

Patikoglou, G. & Burley, S. K. (1997). Eukaryotic transcription factor–DNA complexes. *Ann. Rev. Biophys. biomol. Struct.* **26**, 289–325.

Pavletich, N. P. & Pabo, C. O. (1991). Zinc-finger-DNA recognition: crystal structure of a Zif268–DNA complex at 2·1 Å. *Science* **252**, 809–817.

Pervushin, K., Ono, A., Fernandez, C., Szyperski, T., Kainosho, M. & Wüthrich, K. (1998). NMR scalar couplings across Watson-Crick base pair hydrogen bonds observed by transverse relaxation optimized spectroscopy. *Proc. natl. Acad. Sci. USA* **95**, 14147–14151.

Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates and avenue to NMR structures of very large biological macromolecules in solution. *Proc. natl. Acad. Sci. USA* **94**, 12366–12371.

Pfuhl, M., Chen, H. A., Kristensen, S. M. & Driscoll, P. C. (1999). NMR exchange broadening arising from specific low affinity protein self-association: Analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *J. biomol. NMR* **14**, 307–320.

Ptashne, M. (1967). Specific binding of the λ phage repressor to l DNA. *Nature* **214**, 232–234.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide–RNA complex. *Science* **270**, 1200–1203.

Qian, Y. Q., Otting, G., Billeter, M., Muller, M., Gehring, W. & Wüthrich, K. (1993a). Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly labelled Antennapedia homeodomain and structure of the DNA-bound homeodomain. *J. mol. Biol.* **234**, 1070–1083.

Qian, Y. Q., Otting, G. & Wüthrich, K. (1993b). NMR detection of hydration water in the intermolecular interface of a protein–DNA complex. *J. Am. Chem. Soc.* **115**, 1189–1190.

Ramos, A., Gubser, C. C. & Varani, G. (1997). Recent solution structures of RNA and its complexes with drugs, peptides and proteins. *Curr. Opin. struct. Biol.* **7**, 317–323.

Rao, A., Luo, C. & Hogan, P. G. (1997). Transcription factors of the NFAT family: regulation and function. *Annu. Rev. Immunol.* **15**, 707–747.

Rastinejad, F., Perlmann, T., Evans, R. M. & Sigler, P. B. (1995). Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature* **375**, 203–211.

Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. (1999). TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J. Am. Chem. Soc.* **121**, 844–848.

Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y. & Ishii, S. (1993). Thermal stability of the DNA-binding domain of the Myb oncoprotein. *Biochemistry* **32**, 7759–7764.

Schulman, R. G., Hilbers, C. W. & Miller, D. L. (1974). Nuclear magnetic resonance studies of protein–RNA interactions. I. The elongation factor Tu-GTP aminoacyl–tRNA complex. *J. mol. Biol.* **90**, 601–607.

Schwabe, J. W., Neuhaus, D. & Rhodes, D. (1990a). Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* **348**, 458–461.

Schwabe, J. W. R., Chapman, L., Finch, J. T. &

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

Rhodes, D. (1993a). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements. *Cell* **75**, 567–578.

Schwabe, J. W. R., Chapman, L., Finch, J. T., Rhodes, D. & Neuhaus, D. (1993b). DNA recognition by the oestrogen receptor: from solution to the crystal. *Structure* **1**, 187–204.

Schwabe, J. W. R., Chapman, L. & Rhodes, D. (1995). The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain comformation. *Structure* **3**, 201–213.

Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. (1990b). Solution structure of the oestrogen receptor. *Nature* **348**, 458–461.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. natl. Acad. Sci. USA* **73**, 804–808.

Shafer, R. H. (1998). Stability and structure of model DNA triplexes and quadroplexes and their interactions with small ligands. *Prog. Nucl. Acids Res. & mol. Biol.* **59**, 55–94.

Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. & Sigler, P. B. (1994). Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* **368**, 469–473.

Shiyanova, T. & Liao, X. (1999). The dissociation rate of a winged helix protein–DNA complex is influenced by non-DNA contact residues. *Arch. Biochem. Biophys.* **362**, 356–362.

Shuker, A. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.

Shuman, J. D., Vinson, C. R. & McKnight, S. L. (1990). Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* **249**, 771–774.

Slijper, M., Boelens, R., Davis, A. L., Konings, R. N. H., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1997). Backbone and side chain dynamics of *lac* repressor headpiece (1–56) and its complex with DNA. *Biochemistry* **36**, 249–254.

Smith, S. A., Rajur, S. B. & McLaughlin, L. W. (1994). Specific purine N7-nitrogens are critical for high affinity binding by the *trp* repressor. *Nature Struct. Biol.* **1**, 18–22.

Sobol, A. G., Wider, G., Iwai, H. & Wutrich, K. (1998). Solvent magnetization artifacts in high-field NMR studies of macromolecular hydration. *J. magn. Reson.* **130**, 262–271.

South, T. L. & Summers, M. F. (1993). Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Science* **2**, 3–19.

Spronk, C. A. E. M., Slijper, M., van Boom, J. H., Kaptein, R. & Boelens, R. (1996). Formation of the hinge helix in the lac repressor is induced upon binding to the *lac* operator. Nature Struct. Biol. **3**, 916–919.

Starich, M. R., Wikström, M., Arst, H. N., Clore, G. M. & Gronenborn, A. M. (1998a). The solution structure of fungal AREA protein–DNA complex: an alternative binding mode for the basic carboxyl tail of GATA factors. *J. mol. Biol.* **277**, 605–620.

Starich, M. R., Wikström, M., Schumacher, S., Arst, H. N., Clore, G. M. & Gronenborn, A. M. (1998b). The solution structure of the Leu22 → Val mutant AREA DNA binding domain complexed with a TGATAG core element defines a role for hydrophobic packing in the determination of specificity. *J. mol. Biol.* **277**, 621–634.

Sunnerhagen, M., Denisov, V. P., Venu, K., Bonvin, A. M. J. J., Carey, J., Halle, B. & Otting, G. (1998). Water molecules in DNA recognition I: hydration lifetimes in *trp* operator DNA in solution measured by NMR spectroscopy. *J. mol. Biol.* **282**, 847–858.

Tan, R. & Frankel, A. D. (1994). Costabilization of peptide and RNA structure in an HIV Rev peptide–RRE complex. *Biochemistry* **33**, 14579–14585.

Tjandra, N. & Bax, A. (1997). Direct measurements of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114.

Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562–12566.

Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997). Use of dipolar ^{1}H-15N and ^{1}H-^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Struct. Biol.* **4**, 732–738.

Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. natl. Acad. Sci. USA* **92**, 9279–9283.

Turner, B. G. & Summers, M. F. (1999). Structural biology of HIV. *J. Mol. Biol.* **285**, 1–32.

Varani, G., Aboul-ela, F. & Allain, F. H.-T. (1996). NMR investigations of RNA structure. *Prog. NMR Spectrosc.* **29**, 51–127.

Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.

Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990). Folding

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press

transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347**, 575–578.

Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J. & Gronenborn, A. M. (1997). Correction of the NMR structure of the ETS1/DNA complex. *J. biomol. NMR* **10**, 317–328.

Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995). Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY–DNA complex. *Cell* **81**, 705–714.

Wikström, A., Berglund, H., Hambraeus, C., van den Berg, S. & Härd, T. (1999). Conformational dynamics and molecular recognition: Backbone dynamics of the estrogen receptor DNA-binding domain. *J. mol. Biol.* **289**, 963–980.

Wojciak, J. M., Connolly, K. M. & Clubb, R. T. (1999). NMR structure of the Tn916 integrase–DNA complex. *Nature Struct. Biol.* **6**, 366–373.

Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S. C. (1988). Structure of a phage 434 Cro/DNA complex. *Nature* **335**, 789–795.

Wolfe, S. A., Zhou, P., Dötsch, V., Chen, L., You, A., Ho, S. N., Crabtree, G. R., Wagner, G. & Verdine, G. L. (1997). An unusual Rel-like architecture in the DNA-binding domain of NFATc. *Nature* **385**, 172–176.

Wüthrich, K. (1986). *NMR of protein and nucleic acids*. New York: John Wiley & Sons.

Wüthrich, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**, 45–50.

Wüthrich, K. (1993). Hydration of biological macromolecules in solution: surface structure and molecular recognition. *Cold Sping Harbor Symp. Quant. Biol.* **58**, 149–157.

Wüthrich, K. (1996). NMR structures of biological macromolecules. In *Encyclopedia of NMR*, vol. 1, (eds. R. K. Harris & D. M. Grant), pp. 710–719, Chichester: John Wiley & Sons.

Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M. & Wright, P. E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J. mol. Biol.* **273**, 183–206.

Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991). Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**, 764–769.

Xu, J., Lapham, J. & Crothers, D. M. (1996). Determining RNA solution structure by segmental isotopic labelling and NMR: application to *Caenorhabditid elegans* spliced leader RNA 1. *Proc. natl. Acad. Sci. USA* **93**, 44–48.

Xu, R., Ayers, B., Cowburn, D. & Muir, T. W. (1999). Chemical ligation of folded recombinant proteins: segmental isotopic labelling of domains for NMR studies. *Proc. Natl. Acad. Sci. USA* **96**, 388–393.

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. & Nakamura, H. (1998). Segmental isotope-labelling using peptide splicing for protein NMR. *J. Am. Chem. Soc.* **120**, 5591–5592.

Yang, D. & Kay, L. E. (1996). Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. mol. Biol.* **263**, 369–382.

Yarus, M. (1969). The recognition of nucleotide sequences. *Annu. Rev. Biochem.* **38**, 841–880.

Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide–RNA aptamer complex. *Nature Struct. Biol.* **3**, 1026–1033.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide–TAR RNA complex. *Chem. & Biol.* **2**, 827–840.

Zhang, H., Zhao, D., Revington, M., Lee, W., Jia, X., Arrowsmith, C. & Jardetzky, O. (1994). The solution structures of the *trp* repressor–operator complex. *J. mol. Biol.* **238**, 592–614.

Zhou, P., Sun, L. J., Dötsch, V., Wagner, G. & Verdine, G. L. (1998). Solution structure of the core NFATC1/DNA complex. *Cell* **92**, 687–696.

Zimmer, D. P. & Crothers, D. M. (1995). NMR of enzymatically synthesized uniformly 13C15N-labelled DNA oligonucleotides. *Proc. natl. Acad. Sci. USA* **92**, 3091–3095.

Zuiderweg, E. R., Kaptein, R. & Wüthrich, K. (1983a). Secondary structure of the *lac* repressor DNA-binding domain by two-dimensional ^{1}H nuclear magnetic resonance in solution. *Proc. natl. Acad. Sci. USA* **80**, 5837–5841.

Zuiderweg, E. R., Kaptein, R. & Wüthrich, K. (1983b). Sequence-specific resonance assignments in the ^{1}H nuclear-magnetic-resonance spectrum of the *lac* repressor DNA-binding domain 1–51 from Escherichia coli by two-dimensional spectroscopy. *Eur. J. Biochem.* **137**, 279–292.

Zuiderweg, E. R., Scheek, R. M. & Kaptein, R. (1985). Two-dimensional 1H-nmr studies on the *lac* repressor DNA binding domain: further resonance assignments and identification of nuclear Overhauser enhancements. *Biopolymers* **24**, 2257–2277.

https://doi.org/10.1017/S0033583599003509 Published online by Cambridge University Press