Biochem. J. (1991) 278, 1–23 (Printed in Great Britain)

REVIEW ARTICLE
Structural aspects of protein–DNA recognition

Paul S. FREEMONT,*§ Andrew N. LANE† and Mark R. SANDERSON‡
*Protein Structure Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX,
†National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, and
‡Cancer Research Campaign Biomolecular Structure Unit, Institute of Cancer Research, Cotswold Road, Sutton, Surrey, U.K.

INTRODUCTION

Protein–DNA interactions are involved in many of the fundamental processes that occur inside cells, including packaging, replication, recombination, restriction and transcription. It is therefore vitally important to understand the nature of protein–DNA interactions, as they are so intimately linked to the control of gene expression, cell division and differentiation. Many proteins, particularly those involved in packaging (e.g. the histones) or replication (DNA polymerases) have low or no sequence specificity, whereas others, such as repressors, transcriptional activators and restriction endonucleases, have extremely high specificity for their special target sites, and are able to find a single site of say 10–20 base pairs in a background of 10^6–10^9 or so base pairs. It is the latter group of proteins with which we are concerned in this Review.

It is particularly important to determine the molecular basis of specificity, which requires a characterization of the conformational properties of the protein (and protein–ligand complex if relevant), the DNA target site, and the changes that ensue as a consequence of the interaction. Any conformational differences that are observed in the specific complex need to be compared with any changes that occur in non-specific complexes to be able to determine what constitutes specific binding.

In recent years significant progress has been made in the detailed analysis of specific protein–DNA interactions, in part due to advances in recombinant DNA technology, but also to improvements in both X-ray crystallographic techniques and the use of two-dimensional n.m.r. (for recent reviews see Struhl, 1989; Brennan & Matthews, 1989*a*; Steitz, 1990; Harrison & Aggarwal, 1990). The most detailed information is available for phage and bacterial repressors (X-ray, n.m.r., biophysics and biochemistry) and this forms the core of the Review. Significant information is also available for numerous eukaryotic transcription factors (zinc fingers, leucine zippers and homeodomain proteins), which will be discussed in terms of the generality of structural motifs used in DNA recognition.

DNA-BINDING PROTEINS

Conceptually there are several ways in which DNA-binding proteins can be classified. For the purposes of this Review, we have chosen to classify the proteins in terms of structural classes (referring to the DNA-binding motif), and further subdividing proteins according to their functional properties (e.g. allosteric modulation of binding activity).

There are at least four structural classes of DNA binding proteins that have been characterised in some structural detail. These are shown in Table 1, which also includes divisions according to functional properties. Table 2 shows those proteins whose X-ray crystal structures are known, including structures of protein–DNA complexes.

Helix–turn–helix proteins

From the early crystal structures of λ phage *cro* repressor (λ *cro*) (Anderson *et al.*, 1981) and *Escherichia coli* catabolite gene activator protein (CAP) (McKay & Steitz, 1981), it became apparent that a structural motif consisting of two helices, related by the two-fold symmetry of the dimeric protein, was at the correct orientation (34 Å apart, equivalent to one turn of duplex DNA) to mediate protein–DNA interaction. It was proposed that one helix from each pair would lie in successive major grooves, with the two-fold symmetry of the protein dimer coincident with the two-fold symmetry of the operator sequence (Anderson *et al.*, 1981; Ohlendorf *et al.*, 1982; Sauer *et al.*, 1982; Steitz *et al.*, 1982; for a recent review see Harrison & Aggarwal, 1990). This structure of two helices (Fig. 1), now termed the helix–turn–helix DNA-binding motif, contains a tight β-turn resulting in an interhelical angle of nearly 90° and an amino acid sequence preference which has been extended to other DNA-binding proteins (Anderson *et al.*, 1982; White, 1987; Brennan & Matthews, 1989*b*; Dodd & Egan, 1990). One of the helices (Fig. 1) lies in the major groove, providing sequence-specific DNA interactions and is termed the recognition helix. The *N*-termini of both helices point towards the phosphate backbone, using the positive helix-dipole for the correct positioning of the recognition helix (Pabo & Sauer, 1984; Brennan & Matthews, 1989*a*). It is presumed that the amino acid side chains jutting out of the recognition helix are able to make sequence-specific interactions with exposed functional groups in the major groove of the DNA.

The helix–turn–helix structural motif has been found in many other prokaryotic repressors and activators (Table 1), although similar protein geometries have been found in unrelated proteins (e.g. L7/L12 ribosomal protein; Richardson & Richardson, 1988; Rice *et al.*, 1990). However, a three-dimensional comparison of all helix–turn–helix motifs has shown that these latter proteins are less similar to λ *cro* repressor than those proteins which are involved in DNA-binding and transcriptional regulation (Brennan & Matthews, 1989*b*). More recently, the helix–turn–helix motif has been found in homeodomain proteins from *Drosophila*. However, the length of the recognition helix is somewhat longer in homeodomain proteins as compared with the prokaryotic helix–turn–helix structures (see below).

The proteins listed in Table 1 can be placed into two groups, namely those that control transcription without the binding of an effector molecule (λCI, λ *cro*, P434 and P434 *cro* repressors) and those whose DNA-binding affinities are allosterically modulated by the binding of a small effector molecule (*trp*, *lac* and *metJ* repressors and CAP).

λCI and λ *cro* repressors. The repressor from bacteriophage λCI (236 amino acids) binds in direct competition with λ *cro* repressor (66 amino acids) to six approximately two-fold symmetric 17-bp operator sites in phage DNA, resulting in a regulation of gene expression which determines the growth characteristics of the phage (Ptashne, 1986). A series of bio-

Abbreviations used: rms, root mean square; CAP, catabolite gene activator protein; NOESY, nuclear Overhauser spectroscopy.
§ To whom correspondence should be addressed.

Table 1. Classes of DNA-binding proteins

Class	Examples	Subunits (kDa)	Allosteric effector	Terminal flexible extension (residues)	DNA target
Helix–turn–helix	λCI	α_2(26)	No	*N*(5)	TATCACCGCCAGTGGTA (OL1)*
	λ *cro*	α_2(7.5)	No	*C*(4)	TATCACCGCCAGTGGTA (OL1)*
	P434	α_2(7.5)	No	No	TACAAGAAAGTTTGTT (OR1)*
	P434 *cro*	α_2(7.5)	No	No	TACAAGAAAGTTTGTT (OR1)*
	trp	α_2(12.5)	L-Trp	*N*(10)	CGAACTAGTTAACTAGTACG
	CAP	α_2(22.5)	Cyclic AMP	No	TGTGANNNNNNTCACT
	lac	α_4(35)	*allo*-Lactose	No	GAATTGTGAGCGCTCACAATTC
	ant	α (7)	No	*N*(10)	ATTA
	eng	α (7)	No	*N*(10)	TAAT
	FIS	α_2(10)	No	*N*(24)	Not applicable
	γδ Resolvase	α_2(20.5)	No	No	CGTCCGAAATATTATAAATTATCGCACA (site 1)¶
Zinc finger (C2H2)	mKR2	α(3)†	No	No	Not known
	ADR1	α(3)†	No	No	Not applicable
	SW15	α(3)†	No	No	CCAGCATGCTATAATGC
	HEB	α(3)†	No	No	GGGGAATCCCC
	Xfin	α(3)†	No	No	Not known
Zinc finger (C4)	*glc*	α_2(7.6)‡	Glucocorticoid	No	AGAACANNNTGTTCT
	est	α_2(7.6)‡	Oestrogen	No	AGGTCANNNTGACCT
Leucine zipper	*fos*	α_2(8)§	No	No	GTGACTCAG
	jun	α_2(8)§	No	No	GTGACTCAG
	GCN4	α_2(8)§	No	No	ATGACTCTT
β Sheet	*arc*	α_2(8.5)	No	No	CATGATAGAAGCACTCTACTAT
	metJ	α_2(12)	*S*-Adenosylmethionine	No	AGACGTCGT
	HU	α_2(9.5)	No	No	Not applicable
Other	*Eco*RI	α_2(31)	No	*N*(14)	GAATTC
	*Eco*RV	α_2(28.5)	No	No	GATATC
	DNAase I	α (30.5)	No	No	Not applicable
	DNA polymerase I	α (103)	No	No	Not applicable

* Only one of the six operator sites is shown.
† Zinc-binding motif which is usually represented as more than two tandem repeats.
‡ DNA-binding domain.
§ Leucine zipper and basic region. *fos* and *jun* can form tight functional heterodimers.
¶ Only one of the three res sites is shown.

chemical and genetic studies (reviewed in Pabo & Sauer, 1984; Ptashne, 1986), have shown that the *N*-terminal domain (92 amino acids) of λCI repressor binds to its operator sites as a dimer and that DNA recognition and transcriptional regulation can be altered by mutations of amino acids within this domain. λ *cro* repressor also binds to the same operator sites as a dimer, although both proteins show little amino acid sequence similarity. The crystal structures of the λCI DNA binding domain and λ *cro* repressor show that their overall structures are quite different (Fig. 1; Anderson *et al.*, 1981; Pabo & Lewis, 1982). For example, λCI repressor is α-helical and the dimer is stabilized by hydrophobic helical contacts, whereas λ *cro* repressor contains both α-helix and β-sheet, with one β-strand stabilizing the λ *cro* repressor dimer. However, the helix–turn–helix structure in both proteins is related, although the relative orientation of helix 3 (recognition helix) to helix 2 in both cases is somewhat different (Fig. 1; Pabo & Lewis, 1982). The most striking difference between both repressors is the *N*-terminal extension of λCI repressor which is absent from λ *cro* repressor. Subsequent co-crystal studies have shown that this extension is an intimate part of the λCI protein–DNA complex. In contrast, λ *cro* repressor has a flexible *C*-terminal region, which has been implicated in making specific DNA contacts in the minor groove (Ohlendorf *et al.*, 1982; Leighton & Lu, 1987).

P434 and P434 *cro* repressors. The P434 *cro* repressor (71 amino acids) and P434 repressor (97 amino acids) from bacteriophage 434 bind in direct competition to a set of six related 14-bp operator sites, analogous to the λ *cro* and λCI repressors (Ptashne, 1986). The different relative affinities of the two proteins for the same operator sites determine whether the phage enters lytic or lysogenic growth and single operator base changes can result in altered binding for one or both of the repressors (Wharton & Ptashne, 1987). The crystal structures of both the *N*-terminal domain of P434 repressor (residues 1–69; Mondragon *et al.*, 1989*a*) and P434 *cro* repressor (Mondragon *et al.*, 1989*b*) show a high degree of structural similarity (0.77 Å rms differences for main chain atoms) as expected from the high amino acid sequence similarity (50% identity). Both proteins are α-helical, contain the helix–turn–helix DNA-binding motif, and show remarkable similarity to λCI repressor (e.g. P434 repressor superimposed with λCI repressor has 1.78 A rms differences for main chain atoms; Mondragon *et al.*, 1989*b*).

These similarities suggest that λCI, P434 *cro* and P434 repressors recognize and interact with their operator sites in a similar fashion. However, there are subtle differences in the mode of DNA recognition and interaction displayed by these repressors, the details of which will be discussed later. On the other hand, the three helix–turn–helix repressors listed in Table 1 that bind effector molecules (the *trp* and *lac* repressors and CAP) each differ both in mechanism of action and in three-dimensional structure.

***trp* repressor.** The *trp* repressor from *E. coli* (108 amino acids)

Table 2. X-ray crystal structures of DNA–binding proteins

The R factor is defined as $R = \Sigma|F_o - F_c| / \Sigma|F_o|$ where F_o and F_c are the observed and calculated structure factors respectively. The nucleotides which make up the specific DNA-binding site are shown in bold type. *, half site.

Protein	Resolution (Å)	R factor (%)	Co-crystallized DNA	Reference
λCI repressor	3.2	–	–	Pabo & Lewis (1982)
λCI repressor	2.5	24.2	TA**TATCACCAGTGGTA**T **TATAGTGGTCACCAT**AA	Jordan & Pabo (1988)
λ *cro* repressor	2.8	–	–	Anderson *et al.* (1981)
λ *cro* repressor	3.9	–	**TATCACCGCGGGTGATA** **ATAGTGGCGCCCACTAT**	Brennan *et al.* (1990)
P434 repressor	2.0	19.3	–	Mondragon *et al.* (1989*a*)
P434 repressor	3.2	44.0	**ACAATATATATTGT** **TGTTATATATAACA**	Anderson *et al.* (1987)
P434 repressor	2.5	17.9	TAT**ACAAGAAAGTTTGT**ACT TA**TGTTCTTTCAAACA**TGAA	Aggarwal *et al.* (1988)
P434 *cro* repressor	2.35	19.5	–	Mondragon *et al.* (1989*b*)
P434 *cro* repressor	3.2	36.6	**ACAATATATATTGT** **TGTTATATATAACA**	Wolberger *et al.* (1988)
trp repressor	2.6	27.0	–	Schevitz *et al.* (1985)
trp repressor	1.8	20.4	–	Zhang *et al.* (1987)
trp aporepressor	1.65	18.0	–	Lawson *et al.* (1988)
trp pseudorepressor	1.65	18.2	–	Lawson & Sigler (1988)
trp repressor	2.4	24.9	T**GTACTAGTTAACTAGTC** **CATGATCAATTGATCAG**T	Otwinowski *et al.* (1988)
CAP	2.9	–	–	McKay & Steitz (1981)
CAP	2.5	20.7	–	Weber & Steitz (1987)
CAP	3.0	–	**GCGAAAAGTGTGACATT*** **GCTTTTCACACTG**	Steitz (1990) Shultz *et al.* (1990)
Homeodomain *eng*	2.8	24.4	TTTTGCCATG**TAAT**TACCTAA AACGGTAC**ATTA**ATGGATTA	Kissinger *et al.* (1990)
FIS protein	2.0	25	–	Kostrewa *et al.* (1991)
metJ repressor	1.7	21.0	–	Rafferty *et al.* (1989)
metJ repressor	2.8	22.0	T**TAGACGTCTAGACGTCTA** **ATCTGCAGATCTGCAGA**TT	Phillips (1991)
Glucocorticoid receptor	2.9	19.6	CC**AGAACA**TCGA**TGTTCT**G G**TCTTGT**AGCT**ACAAGA**CC	B. F. Luisi *et al.*, unpublished work
Zif268	2.1	18.2	AGCGTGGGCGT CGCACCCGCAT	Pavletich & Pabo (1991)
DNAase I endonuclease	2.0	15.7	–	Oefner & Suck (1986)
DNAase I endonuclease	4.0	–	pdTp	Suck & Oefner (1986)
DNAase I endonuclease	2.0	20.7	GCGATCGC CGCTAGCG	Suck *et al.* (1988)
*Eco*RI endonuclease	3.0	–	TCGC**GAATT**CGCG GCG**CTTAA**GCGCT	McClarin *et al.* (1986) Kim *et al.* (1990)
*Eco*RV endonuclease	–	–	–	Winkler *et al.* (1991)
*Eco*RV endonuclease	–	–	GG**GATATC**CC CC**CTATAG**GG	F. Winkler (personal communication)
*Eco*RV endonuclease	–	–	CGAGCTCG GCTCGAGC	F. Winkler (personal communication)
DNA polymerase I	3.3	–	–	Ollis *et al.* (1985)
DNA polymerase I	2.6	18.0	–	Beese & Steitz (1991)
DNA polymerase I	2.8	–	pdT_4	Freemont *et al.* (1988)
DNA polymerase I	3.1	19.0	pdT_4	Beese & Steitz (1991)
DNA polymerase I	3.8	–	AGACCGCCCGG GGCGGGCC	Freemont *et al.* (1988)
recA	2.8	–	–	Steitz (1990)
HU protein	2.1	–	–	White *et al.* (1989)
γδ Resolvase	2.7	21.5	–	Sanderson *et al.* (1990)
Nucleosome core	3.3	–	–	Burlingame *et al.* (1985)
Nucleosome core	7.0	–	–	Richmond *et al.* (1984)

is involved in the regulation of tryptophan biosynthesis by binding as a dimer to three different operator sites in the presence of its corepressor L-tryptophan (Klig *et al.*, 1988). Binding in the absence of L-tryptophan is weak and non-specific. However, desamino derivatives of tryptophan bind with similar affinity, but do not activate the repressor, whereas decarboxy derivatives both bind the repressor, and activate it; clearly the ammonium group is essential for activation, but not for binding (Marmorstein *et al.*, 1987, Marmorstein & Sigler, 1989).

The crystal structures of the holorepressor and aporepressor show that the subunits contain six α-helices, two of which correspond to the helix–turn–helix structure as observed in other prokaryotic repressors, and an *N*-terminal arm of eleven residues (Schevitz *et al.*, 1985; Zhang *et al.*, 1987). The dimer contacts are

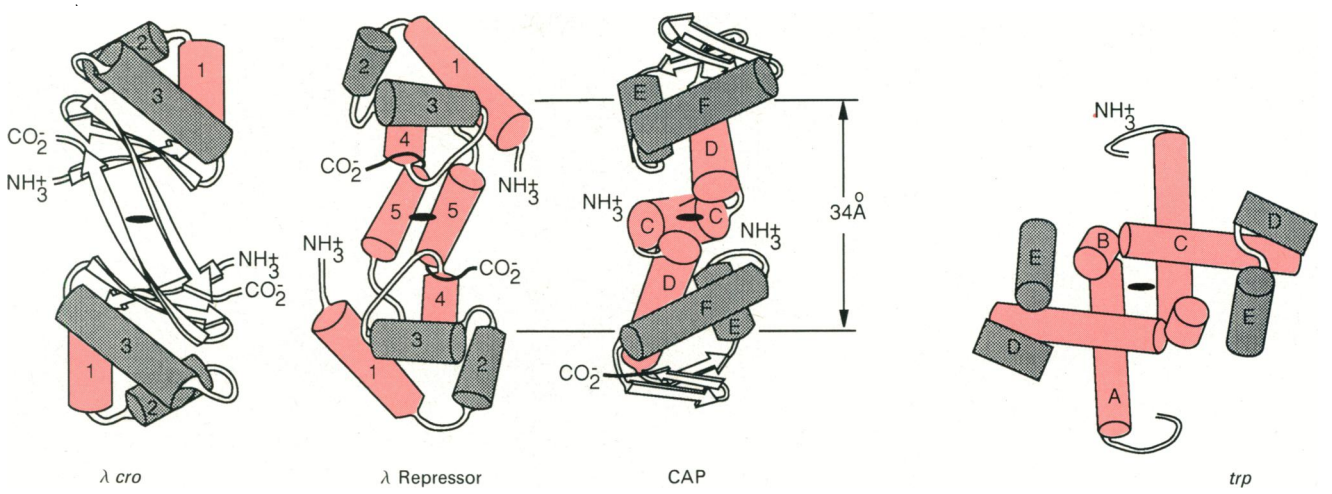

Fig. 1. Structures of helix–turn–helix proteins

The structures are based on crystallographic studies. Cylinders represent α-helices and arrows represent β-strands. The helix–turn–helix motif is shown shaded grey in each protein; ⬬ shows the two-fold symmetry axis.

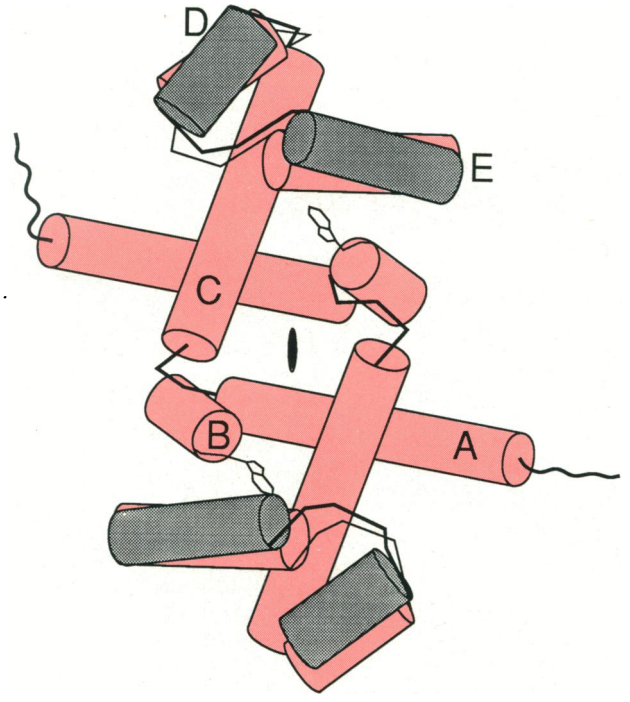

Fig. 2. Conformational changes in the *trp* repressor on binding L-tryptophan

The repressor is shown as the apo state and the holo-form (shaded grey) oriented so that the A, B and C helices coincide. Differences are evident particularly for the D and E helices, which form the helix–turn–helix motif. Adapted from Lawson *et al.* (1988).

formed by the interlocking of four α-helices, the conformation of which are identical in both aporepressor and repressor (Fig. 1; Zhang *et al.*, 1987). The binding of L-tryptophan to aporepressor, however, leads to a reorientation of the helix–turn–helix structure with the tryptophan molecule stabilizing the 'active' conformation (Zhang *et al.*, 1987; Otwinowski *et al.*, 1988). Interestingly, the differences between the two holorepressor structures, which are crystallized in different crystal space groups, are as large as the differences between the holo and apo repressors (Fig. 2; Lawson *et al.*, 1988). This indicates that there is considerable flexibility of the DNA binding motif, which could be related to the need for the repressor to recognize three similar operator sequences (Lawson *et al.*, 1988). Indeed, analysis of the n.m.r. signature of the helix–turn–helix indicated that the D helix may be less ordered in solution than in any of the crystal environments (Arrowsmith *et al.*, 1990). In addition, the *N*-terminus of the A helix showed substantial differences in the different crystal states, and the first 10 residues are disordered. N.m.r. experiments show that the disorder is dynamic (Arrowsmith *et al.*, 1989; Lane, 1989) i.e. these residues move with a large amplitude on a time-scale significantly faster than 1 ns. By analogy with λCI repressor and the homeodomains, it is possible that the *N*-terminal extensions make additional contacts with the DNA. However, recent evidence suggests that under near physiological conditions, the contribution of the first seven residues to the binding energy are rather small (Marmorstein *et al.*, 1991, and see below).

The structure of *trp* aporepressor bound to the tryptophan analogue, indole-3-propanoate, has also been solved. Despite the loss of the ammonium group, this analogue binds slightly more tightly than tryptophan. The crystal structure shows that the indole ring of indole-3-propanoate is rotated approximately 180° with respect to the position in the tryptophan complex, such that the indole NH group would not be correctly positioned for making a contact with the phosphate backbone in the complex with DNA. The lack of the ammonium group and the incorrect orientation is proposed to account for the antirepressor properties of indole-3-propanoate (Lawson & Sigler, 1988).

Cyclic AMP receptor protein (CAP). CAP is an α_2 dimer of molecular mass 2 × 22.5 kDa which binds to 22-bp sites in the operons which it regulates. In the cell, as the cyclic AMP concentration increases, CAP binds cyclic AMP and activates transcription through RNA polymerase. Apo-CAP has only a low non-specific affinity for DNA and binds cyclic AMP with a dissociation constant of about 5 μM (Donoso-Pardo *et al.*, 1987). Interestingly, while cyclic GMP competes effectively with cyclic AMP, it does not activate the protein. Indeed, activation is associated with a substantial conformational change, which has been observed both by small angle X-ray scattering (which indicated a 5% change in the radius of gyration; Kumar *et al.*, 1980) and by ^{19}F n.m.r. (Sixl *et al.*, 1990). However, cyclic GMP does not cause this change in conformation (Sixl *et al.*, 1990). CAP can also act as a positive regulator for a number of sugar operons, including lactose and galactose, and also acts as a repressor.

CAP was one of the first DNA-binding proteins exhibiting the helix–turn–helix motif whose structure was solved. The structure

consists of an *N*-terminal region consisting of a β-roll which binds cyclic AMP and a *C*-terminal domain consisting of three α-helices, two of which are the helix–turn–helix motif similar to helices 2 and 3 in λ *cro* repressor (Fig. 1). Important contacts with the 6-amino position of the adenine ring of cyclic AMP are made through residues Ser-128 and Thr-127 (McKay & Steitz, 1981). The dimer is held together by interactions between two α-helices at the interface region but there is no intertwining of structural units as observed in the *trp* and *metJ* repressors.

The question of how transcriptional activation is mediated through CAP's interactions with cyclic AMP, bound DNA and RNA polymerase has directed the course of the crystallographic and n.m.r. experiments. In the CAP structure the two dimer subunits are not superimposable by the dyad axis, but are offset by 18°. The conformational switch which is induced by cyclic AMP and results in the repositioning of the helix–turn–helix motif stabilizing a favourable interaction with DNA (McKay *et al.*, 1982) is as yet not fully understood as the crystal form of apo-CAP has not been determined (Steitz, 1990). The structures of a number of mutant forms of CAP (CAP* and CAP91) which activate transcription in the absence of cyclic AMP have been elucidated, in an attempt to further the understanding of the allostery. The CAP* mutations are located at the dimer interface and appear not to perturb significantly the CAP structure, whereas the CAP91 structure shows small changes in the small domain and hinge regions relative to wild-type CAP (Weber *et al.*, 1987). However, it is still not clear from the structures of these altered proteins how cyclic AMP affects CAP activation.

***lac* repressor.** The *lac* repressor is unusual in being an α_4 tetramer of total molecular mass 140 kDa. Each subunit consists of two autonomously folded domains, the larger of which contains the binding site for the inducer allolactose. The smaller *N*-terminal domain can be easily cleaved with different proteases to yield the monomeric 51–59-residue head-piece (Geisler & Weber, 1977) which has specific DNA binding activity (Scheek *et al.*, 1983), but which has lost its allosteric response to allolactose. However, n.m.r. studies of intact repressor show that the 'headpiece' is rather mobile with respect to the ligand-binding domain (Wade-Jardetzky *et al.*, 1979).

Recently, crystals suitable for structure determination (3.5 Å) have been obtained for the intact repressor, although no suitable co-crystals of the *lac* repressor–operator complex are available (Pace *et al.*, 1990). However, the headpiece is sufficiently small to be amenable to detailed characterisation by n.m.r. Thus, Kaptein's group were able to show that the headpiece consists of three α-helices (Zuiderweg *et al.*, 1983; Kaptein *et al.*, 1985). The first two helices were separated by a short linker, and the third helix by a longer loop. Detailed molecular dynamics calculations (de Vlieg *et al.*, 1988) showed that the first two helices had high conformational similarity to the helix–turn–helix motifs found in the phage and bacterial repressors, suggesting a similar mode of binding. However, as will be seen later, the mode of interaction of the *lac* repressor headpiece with the *lac* operator differs from that of the other helix–turn–helix proteins. Also because the inducer-binding domain is absent, no information is available on the mechanism of the reduction of affinity for the *lac* operator on binding the inducer allolactose.

Other helix–turn–helix proteins

Homeodomains. The homeobox genes involved in embryonic segmentation in for example *Drosophila* contain a 60-residue motif that codes for a DNA-binding protein called the homeodomain protein. The *antennapedia* homeodomain protein has been expressed in *E. coli* and purified to homogeneity (Müller *et al.*, 1988). Both biochemical and structural studies of the protein have been initiated. The homeodomain protein recognizes and binds with high affinity (K_d approx. 1 nM) the same DNA target site as the full homeotic protein (Affolter *et al.*, 1990), suggesting that it exists as an autonomously folded domain within the intact protein. The solution conformation has been determined using n.m.r. techniques, and was shown to contain four helices. The fourth helix could also be viewed as an extension of the third helix (Qian *et al.*, 1989; Billeter *et al.*, 1990). Comparing the helices with those in λ *cro* and *trp* repressor showed that there is indeed a helix–turn–helix motif, although the second helix is significantly longer than in the phage and prokaryotic repressors (Qian *et al.*, 1989). The first seven residues and last ten residues are not well-defined by the n.m.r. data, consistent with significant flexibility in solution. Recently, the crystal structure of the *engrailed* homeodomain protein complex with DNA has been solved and shows that a helix–turn–helix structure is again involved in DNA binding, although the helices are longer than those characterized for prokaryotic repressors (Kissinger *et al.*, 1990). A superimposition of C^α atoms of the helix–turn–helix (residues 33–52) of λCI repressor with that of the *engrailed* homeodomain protein (residues 31–50) shows an rms deviation of 0.84 Å. Furthermore, certain hydrophobic residues reside in similar environments in the two structures, suggesting that they have a conserved structural role (Kissinger *et al.*, 1990).

FIS protein. The crystal structure of the factor for inversion stimulation (FIS, 98 amino acids) from *E. coli* has recently been determined (Kostrewa *et al.*, 1991). FIS stimulates site-specific recombination by binding as a dimer to consensus 15-bp enhancer sequences. The structure of the FIS monomer comprises four α-helices, with the dimer stabilized by a tight intertwining of both subunits similar to that observed in *trp* and *metJ* repressors. However, the *N*-terminal 24 residues are disordered in the crystal structure, although they may be involved in DNA binding and recognition. The helix–turn–helix motif protrudes from each monomer and is related in sequence and structure to other prokaryotic helix–turn–helix-containing proteins (rms deviation of 0.51 Å on C^α atoms when compared with λ *cro* repressor; Kostrewa *et al.*, 1991). However, FIS has six positively charged residues in the recognition helix which is different to other helix–turn–helix proteins, and suggests that FIS recognizes DNA predominantly through non-specific interactions. Based on the separation distance between the FIS helix–turn–helix structures (24 Å) which is shorter than that required for binding to successive major grooves of ideal B-DNA (34 Å), a model of DNA interaction involving DNA bending has been proposed (Kostrewa *et al.*, 1991). A DNA bend of approx. 90° has been suggested for the FIS–DNA interaction, although the model does not exclude the possibility that the flexible *N*-terminal regions of FIS play a role in DNA recognition and binding, analogous to λCI repressor and the *C*-terminal region of λ *cro* repressor.

***γδ* resolvase.** Gene transfer from one plasmid to another, which occurs for instance in the transfer of antibiotic resistance, is catalysed by two enzymes. The first step, which results in the fusion of two plasmids to produce a cointegrated plasmid, is catalysed by transposase. The second step in the case of the transposon $\gamma\delta$ is catalysed by $\gamma\delta$ resolvase, and involves an intramolecular site-specific recombination event, resulting in two plasmids that each contain a copy of the gene. Resolvase also acts as a repressor of the expression of the resolvase and the transposase transcripts. The active recombination complex is thought to comprise three resolvase tetramers bound to the six *res* sites, three from each plasmid (Hatfull & Grindley, 1988).

Resolvase (183 amino acids) can be cleaved into a DNA-binding domain of 43 residues and a catalytic domain of 140

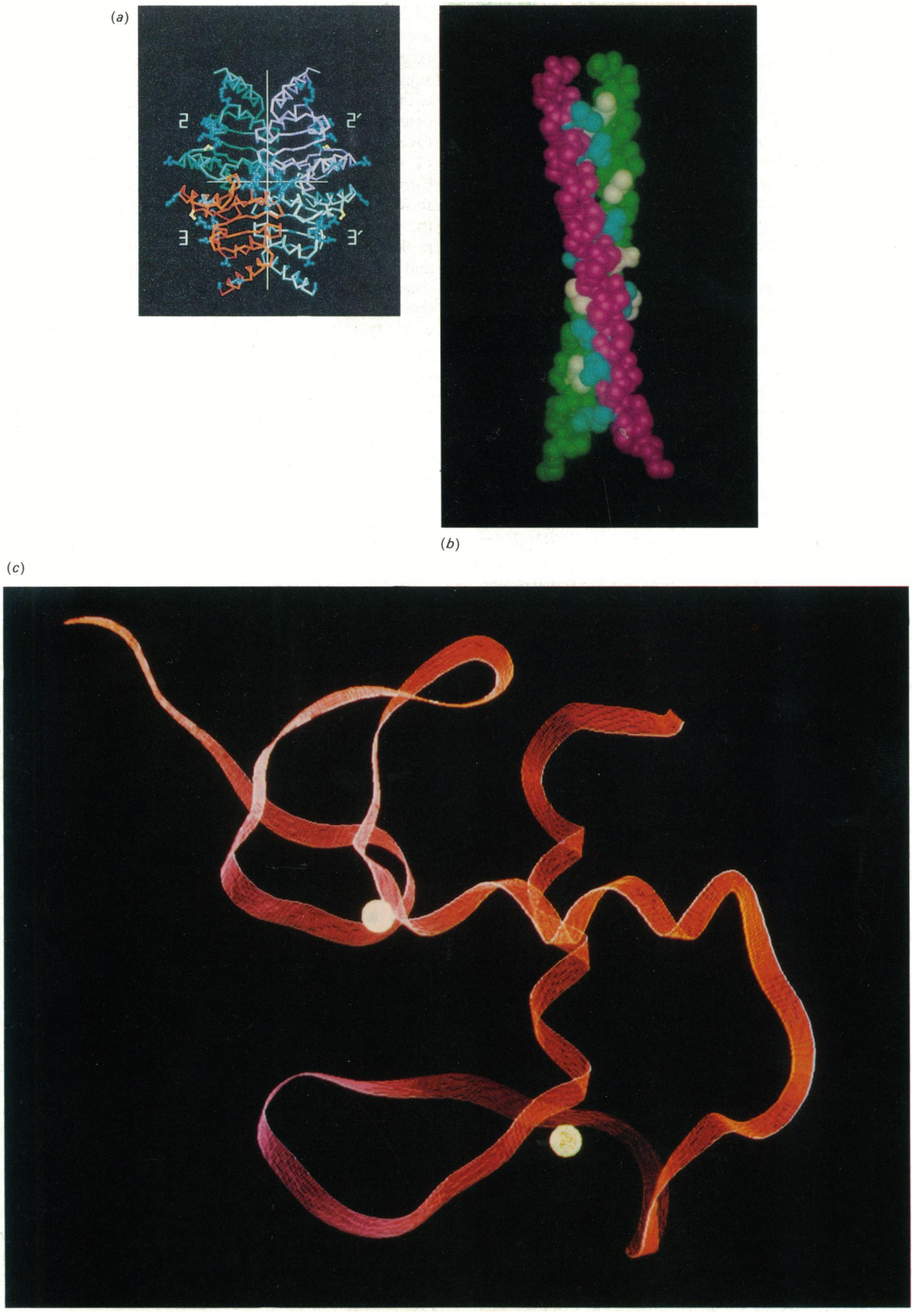

Fig. 3. Structures of DNA-binding motifs

(*a*) The crystallographic tetramer of $\gamma\delta$-resolvase is shown. The resolvase monomers are labelled 2, 2′, 3 and 3′ (2, green; 2′, purple; 3, red; 3′, silver) with the side chains of mutant proteins defective in catalysing the resolution reaction but not specific DNA binding shown in blue. The active site

residues (Abdel-Meguid *et al.*, 1984). Sequence alignment shows that the DNA-binding domain has a helix–turn–helix motif. The catalytic domain which catalyses site-specific recombination has been crystallized and solved (Sanderson *et al.*, 1990). The structure shows resolvase to be an α/β protein with a parallel five-stranded β-sheet surrounded by five α-helices. Biochemical experiments have implicated Ser-10 as being linked to the 5′-phosphate at the cleavage site during the recombination event. One pair of crystallographic dimers places Ser-10 at a separation of 13 Å across the dyad axis. Modelling studies in which the DNA for the recombination site is docked onto this dimer require the DNA to be highly twisted. Although this dimer has not been definitively confirmed to be the catalytic dimer, the side chains of protein mutants which retain DNA-binding ability but lack recombinase activity are close to Ser-10. Fig. 3(*a*) shows a possible tetrameric arrangement of resolvase subunits taking into account the location of the active site serine residue and side chains which seem important in catalysing the resolution reaction but not in specific DNA binding. One may note that resolvase has been shown to bend DNA significantly, similar to FIS, presumably in forming the recombination synatosome. At present the structural basis for the resolvase–*res* site interaction is not available.

In summary, the helix–turn–helix motif is a well-defined supersecondary structural element that has convenient geometric properties for DNA recognition. However, the remainder of the protein to which the helix–turn–helix motif is attached can vary widely among proteins of similar functional class. Thus, the dimerization interfaces can be a pair of nearly parallel α-helices (CAP), antiparallel helices (λCI repressor), antiparallel β-sheet (λ *cro* repressor) or an intertwined arrangement of helices (*trp* repressor) (cf. Fig. 1). Indeed, the DNA-binding mode may be in part dictated by structural or functional constraints other than the dimerisation interface. For example, the allosteric response requires separate effector-binding sites that can interact with the DNA-binding domain. However, the effector binding sites of CAP and *trp* repressor are very different. Moreover, as will be seen later, the helix–turn–helix motif is not an essential element for specific DNA recognition, as there are now many examples of other structural motifs that are involved in DNA binding and recognition which may not even involve α-helices at all (e.g. *metJ* repressor).

Zinc fingers

Zinc fingers are autonomously-folding domains that require zinc for DNA binding activity, and were called fingers because of the way the primary structure could be drawn on paper (Klug & Rhodes, 1987), with the zinc atom linking distant residues, and the intervening sequence written as a loop. Based on known zinc-binding motifs, tetrahedral co-ordination of the metal ion by the conserved cysteine and histidine residues and structure prediction, a model of a single finger was proposed by Berg (1988). This model consists of an antiparallel two-stranded β-sheet and an α-helix. These two structural elements are held together by the tetrahedrally co-ordinated Zn atom, whose ligands are two sulphur atoms provided by the cysteine residues in the β-sheet, and the two nitrogen atoms of two histidine side chains in the helix (see Fig. 4*a*).

The solution conformations of several zinc fingers have been determined using n.m.r. data and recently the crystal structure of a three-zinc-finger–DNA complex has been elucidated. All of the structures give essentially the same picture, namely a β-loop folded on an α-helix, with a hydrophobic core. The β-element is not a regular antiparallel sheet, although the precise description of this part of the structure differs from protein to protein. Also, the length of the helix varies somewhat in the different structures, and in one case a turn of 3_{10} helix was found (Lee *et al.*, 1990). The *N*-terminal end of the α-helix is responsible for making sequence-specific interactions with DNA (Nardelli *et al.*, 1991; Pavletich & Pabo, 1991). The two-finger motif of the human enhancer binding protein has been shown to bind specifically to a 16-bp DNA fragment containing the enhancer binding site (Sakaguchi *et al.*, 1991). The specific interactions appear to involve three or at most four base-pairs per finger, in agreement with Nardelli *et al.* (1991). Interestingly, Omichinski *et al.* (1990) report that the same basic fold is found in the third domain of the Japanese quail ovomucoid protease inhibitor (which does not contain zinc). This suggests that the zinc motif is not specifically a DNA-binding fold, but rather an intrinsically stable motif from which side chains protrude to provide the DNA-recognizing function. The zinc presumably stabilizes the fold, and this may explain why a folded structure was obtained for mKr2 in the absence of metal ions (Carr *et al.*, 1990), and which has the same overall features as the other sequences which contain zinc (Fig. 4*b*; Omichinski *et al.*, 1990).

Individual zinc motifs usually have low affinity for DNA; specific binding seems to require multiple, tandem motifs. In ADR1, the two motifs are independently folded domains (Klevit *et al.*, 1990), as are the three domains in SW15 (Neuhaus *et al.*, 1990). In the case of SW15, the n.m.r. results indicate that the domains interact with one another weakly if at all (Neuhaus *et al.*, 1990). A further interesting aspect of the SW15 protein is that the first zinc domain has an additional seven-residue strand of β-sheet starting at the first residue of the domain (D. Neuhaus, personal communication), to form a three-stranded β-sheet. The other two fingers, like all the other zinc domains so far examined in detail, do not have this additional strand, partly because the linker between the domains is too short. It remains to be seen whether this feature is specific to the SW15 family.

The steroid-receptor proteins in which all four Zn ligands are cysteinyl sulphur were also originally called zinc-finger proteins, by analogy with the C2H2 class. These proteins, which consist of at least three domains, are allosterically modulated by steroid hormones (Beato, 1989). The intact proteins appear to be dimeric in solution, and bind to DNA as dimers (Wrange *et al.*, 1989). N.m.r.-derived structures of the DNA-binding domains (containing two Zn-binding motifs) have appeared, for the glucocorticoid receptor (Hård *et al.*, 1990*a*,*b*) and for the oestrogen receptor (Schwabe *et al.*, 1990). Each motif consists of an ill-defined loop which may be relatively flexible, held at one end to the *N*-terminus of an α-helix. These two domains are packed together with the two α-helices perpendicular to one another, and crossing near their mid-points (Fig. 3*c*). Mutational data suggest that the *N*-terminal helix interacts with the DNA. However, the isolated DNA-binding domains are monomeric in solution, and bind to the target site with relatively low affinity (K_d of micromolar) compared with the intact proteins (K_d of nanomolar) (Hård *et al.*, 1990*c*). The principal role of the

serine is shown in yellow. The tetramer is formed from subunits 2 and 3 related by a vertical crystallographic two-fold axis (white line) to subunits 2′ and 3′. From Sanderson *et al.* (1990), with permission. © Cell Press. (*b*) The structure of the coiled coil motif in leucine zippers. The leucine zipper peptide from the yeast activator protein GCN4 forms a coiled coil of parallel α-helices. The two helices are shown in red and green, and the leucine residues in white and blue. The picture was kindly provided by Dr. A. Pastore, EMBL, Heidelberg, Germany. (*c*) Structure of the oestrogen receptor protein DNA-binding domain. The backbone is shown as a ribbon structure in red, and the zinc atoms are depicted as white spots. The picture was kindly provided by Drs. D. Neuhaus and J. Schwabe, LMB, Cambridge, U.K.

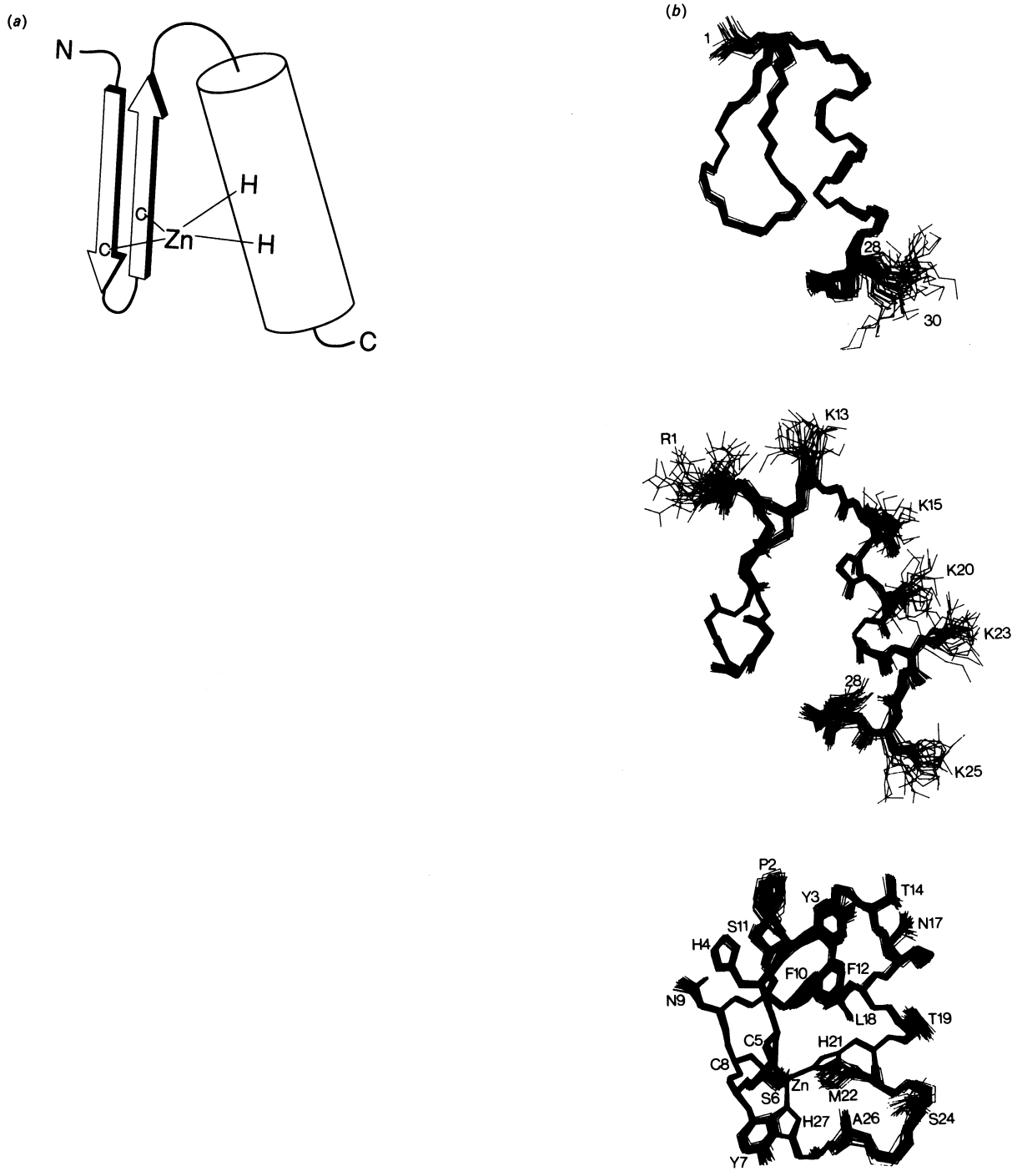

Fig. 4. Structure of zinc fingers

(*a*) Model of a single zinc-finger motif. Adapted from Berg (1988). (*b*) Structure of a zinc domain from the human enhancer binding protein (C2H2) derived from n.m.r. data, showing the backbone atoms only (top), backbone atoms plus Arg and Lys sidechains (centre) and all heavy atoms (bottom). Kindly provided by Drs. G. M. Clore and A. M. Gronenborn, and reprinted with permission from Omichinski *et al.* (1990) Biochemistry **29**, 9324–9334. © (1990) American Chemical Society.

hormone appears to be in activation of the cellular form of the receptor, which in the absence of ligand is complexed with a 90 kDa heat-shock protein or chaperone. Binding ligand dissociates the complex, whereupon the receptor–hormone complex can dimerize and migrate to the nucleus (Beato, 1989). The role of the hormone in modulating the affinity of the receptor for DNA, or the binding kinetics is not completely clear (Schauer *et al.*, 1989; Wrange *et al.*, 1989). It must be borne in mind that structures of DNA-binding domains of transcription factors as complex as these can provide only limited information about the mechanisms of *trans*-activation.

Leucine coiled-coils/basic region ('leucine zippers')

Leucine zipper proteins consist of a dimerization domain (the actual 'zipper') and a basic region *N*-terminal to the 'zipper'. The basic region has been shown to be responsible for DNA binding (Gentz *et al.*, 1989; Kouzarides & Ziff, 1989; Nakabeppu & Nathans, 1989). The leucine zipper motif was named after a

model derived to account for the known functional dimerization of several eukaryotic transcription factors, and the presence of an approximately 29-residue stretch of peptide containing four or five leucine residues periodically arranged every seventh residue (the heptad repeat). In an α-helix, residues spaced seven apart in the sequence will lie on the same face, every other turn of the helix. This suggests the possibility of two helices with the leucines interdigitating (i.e. like a zipper; Landschultz *et al.*, 1989). However, while it was rapidly confirmed by c.d. and n.m.r. (O'Shea *et al.*, 1989) that the leucine-zipper peptide alone does form an α-helix that self-dimerizes with high affinity, the orientation is parallel rather than antiparallel, and the dimer is completely symmetric as far as the n.m.r. spectra are concerned (Oas *et al.*, 1990; Saudek *et al.*, 1990). Recent X-ray diffraction studies of crystals of a 33-amino-acid peptide, corresponding to the zipper region of GCN4, are in agreement with the n.m.r. studies, in that the crystalline peptide adopts a parallel coiled-coil (Rasmussen *et al.*, 1991). The above studies are also consistent with the coiled-coil motif, which was recently observed in the X-ray crystal structure of Ser-tRNA synthetase (Cusack *et al.*, 1990).

Recent genetic studies aimed at analysing the sequence requirements for coiled-coil motifs have found that the most frequent functional mutations were hydrophobic by nature (Hu *et al.*, 1990). However, many combinations of different hydrophobic residues were non-functional, which suggests that a hydrophobic interface *per se* is insufficient for zipper dimerization and that leucines have been specifically selected.

The DNA-binding domain (basic region) is a contiguous stretch of about 30 residues *N*-terminal to the zipper domain; a high proportion of these residues are basic (Kouzarides & Ziff, 1989; Nakabeppu & Nathans, 1989). Indeed, Kim and coworkers have shown that the basic region of the yeast activator GCN4 alone can bind its target site with significant affinity when dimerized by a disulphide bond in place of the zipper domain, clearly demonstrating that the zipper domain is used primarily for dimerization (Talanian *et al.*, 1990). In contrast, the proto-oncogene products *fos* and *jun*, which both contain leucine zippers, are functionally active as the *fos–jun* heterodimer, indicating that the details of the dimerization surfaces are important, and that single gene regulation can be controlled by the expression of at least two proteins (see Jones, 1990).

Recent n.m.r. studies of the GCN4 peptide confirmed that the leucine zipper is indeed a parallel coiled-coil (Oas *et al.*, 1990; Saudek *et al.*, 1990) (see Fig. 3*b*), and that the basic region contains a significant amount of α-helix (Saudek *et al.*, 1990, 1991). However, the amount of helix is much greater at low temperatures than at room temperature, indicating a marginally stable α-helix in equilibrium with other (non-helical) conformations (Weiss, 1990; Weiss *et al.*, 1990; Patel *et al.*, 1990). This probably accounts for the observation of Talanian *et al.* (1990) that the disulphide-linked GCN4 basic regions have much greater affinity for DNA at low temperature. In the presence of DNA, a large increase in helicity of the basic region is detected, both for GCN4 (Weiss *et al.*, 1990) and for *fos* and *jun* (Patel *et al.*, 1990). This appears to be a general property of this class of proteins. The results are consistent with the scissors-grip model of binding proposed by Sigler and colleagues (Vinson *et al.*, 1989), in which the fork-like structure of the dimer sits astride the DNA, allowing contacts to be made between the basic regions and successive major grooves of the DNA.

Beta proteins

Although all of the proteins so far discussed contain α-helices that are presumed to interact with the major groove of DNA, recent work has shown that there is an important class of proteins that use a two-strand antiparallel β-sheet instead. The potential use of β-sheet structures as DNA-interacting motifs was first suggested by a number of groups, who noticed the structural complementarity between the grooves of DNA duplex and the right-handed twist of β-sheet structures (Carter & Kraut, 1974; Church *et al.*, 1977). The recent structure determinations of *arc* and *metJ* repressors from *E. coli* have confirmed the use of a β-ribbon structure for specific DNA interaction in the major groove (see below). The structure of *B. stearothermophilus* HU protein, a non-specific DNA-binding protein related to the specific DNA-binding proteins IHF and TF1, shows an alternative β-sheet–DNA interaction mainly involving recognition through the minor groove (for a recent review see Phillips, 1991).

***metJ* and *arc* repressors.** The *metJ* repressor from *E. coli* (104 amino acids) is involved in the control and regulation of methionine biosynthesis by binding co-operatively, as dimers, to tandem repeats of six specific operator sites (Phillips *et al.*, 1989). The crystal structures of the repressor in the presence and absence of its corepressor *S*-adenosylmethionine have been elucidated (Rafferty *et al.*, 1989). Each monomer of *metJ* repressor comprises three α-helices and one β-strand, with the dimer formed by the intertwining of both subunits, as observed for *trp* repressor. The β-strands from both subunits form an antiparallel β-ribbon which lies on one face of the dimer and which, by crossing each other, interlocks both subunits (Fig. 5). The dimer interface is formed mainly by the interaction of helix B from each subunit. However, unlike *trp* repressor and CAP, *metJ* repressor does not have a helix–turn–helix motif and does not undergo any significant conformational change upon binding of its corepressor *S*-adenosylmethionine (Rafferty *et al.*, 1989), even though the corepressor increases the affinity of the repressor for its operator sites by at least three orders of magnitude (Phillips *et al.*, 1989). The *metJ* operator consists of several tandem repeats of an octanucleotide (the *metJ* box), on which multiple copies of the *metJ* protein bind to form protein–protein complexes along the DNA helix. The protein–DNA interactions occur by the placement of two symmetrically equivalent β-strands in the major groove (Phillips *et al.*, 1989; Rafferty *et al.*, 1989). The *metJ* repressor–operator complex could therefore be considered as a dimer of dimers, as each dimer binds to a single *metJ* box. This type of protein–DNA interaction has not been observed in any other protein–DNA complex to date.

The *arc* protein from P22 is a small (53 amino acids) α_2 dimer that has significant sequence homology with *metJ* repressor. The secondary structure consists of an *N*-terminal β-strand followed by two α-helices (Breg *et al.*, 1989; Zagorski *et al.*, 1989). The *N*-terminal region was shown to contain the DNA-binding site. A complete n.m.r. structure determination, partly based on the homology with *metJ* repressor, showed that the β-strand forms an antiparallel β-sheet in the dimerization interface (Breg *et al.*, 1990). The exposed sheet is entirely analogous to that observed in *metJ* repressor (see Phillips, 1991). Furthermore, *arc* repressor has been shown to bind to its operator site co-operatively as a tetramer (Brown *et al.*, 1990), analogous to *metJ* repressor.

Protein HU. Protein HU (90 amino acids) binds non-specifically to DNA as a dimer and is involved in the formation of nucleoprotein structures (the nucleoid) analogous to the nucleosomes found in eukaryotes. The crystal structure has been determined and shows HU to be composed of two distinct structures, namely a *N*-terminal helical region containing two α-helices and a *C*-terminal sheet region consisting mainly of a three-stranded antiparallel β-sheet (Tanaka *et al.*, 1984; White *et al.*, 1989). The HU dimer is intertwined and forms a cleft bordered by two arm-like structures (one from each subunit) consisting of two antiparallel β-strands (Fig. 6). It is these structures, which together resemble a distorted β-ribbon, that

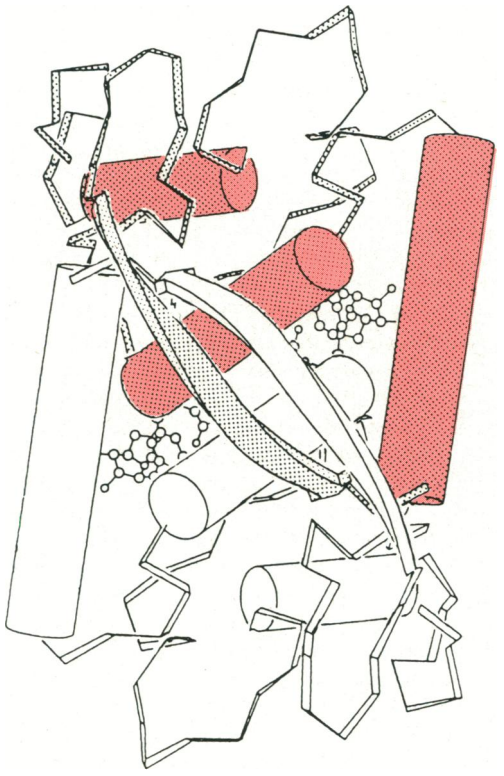

Fig. 5. Structure of *metJ* repressor

Schematic representation of the *metJ* repressor dimer viewed along the molecular two-fold axis. The corepressor is shown as a 'ball and stick' drawing. The two symmetrically related β-sheets on the surface of the dimer interact with the major groove of the operator DNA. Reprinted by permission from Rafferty *et al.* (1989). © 1989 Macmillan Magazines Ltd.

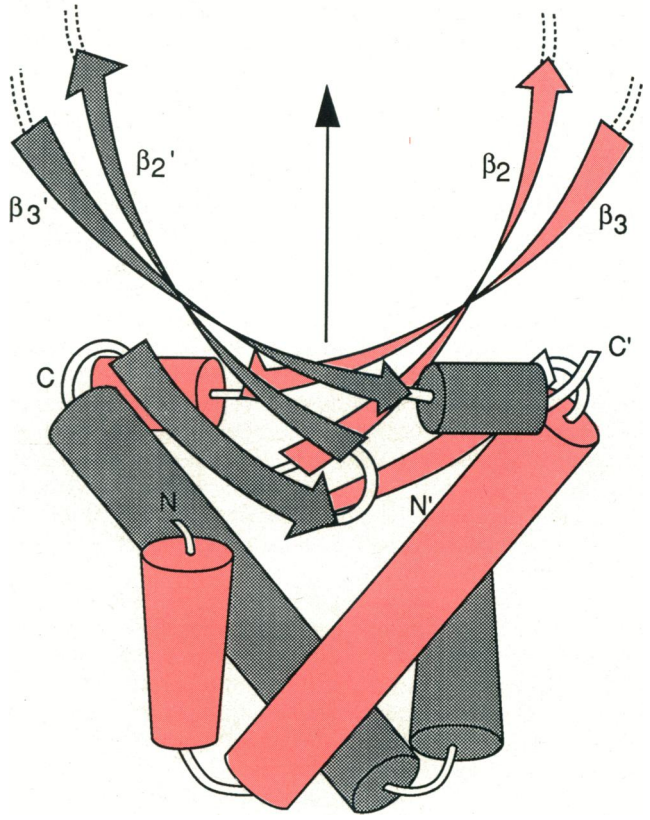

Fig. 6. Structure of HU protein

A schematic representation of the HU dimer with α-helices as cylinders and β-strands as arrows. One subunit is shaded grey and one pink; the arrow represents the molecular two-fold axis of the non-crystallographic dimer. The DNA is thought to interact with the distorted β-ribbon formed by the strands β_2 and 3 with each pair of strands inserted into successive major grooves. Adapted from Tanaka *et al.* (1984).

have been proposed to interact with the DNA by inserting into the minor groove, although a major groove model is also possible (White *et al.*, 1989). Although no co-crystal structure is available, HU has been shown to bend DNA, which, if the proposed model were correct, would result in a widening of the minor groove to accommodate the β-ribbon structure. The formation of the nucleosome-like particles would involve a number of protein–protein interactions and therefore DNA bending would be a result not only of protein–DNA interaction but also protein–protein interaction as well (see Phillips, 1991).

Other folds

There are several DNA-modifying enzymes whose structures have been solved crystallographically, and which do not fall into any of the above classes.

DNAase I. DNAase I is a glycoprotein endonuclease (257 amino acids) that catalyses the non-specific cleavage of double-stranded DNA through hydrolysis of a P–$O^{3'}$ bond yielding 5'-oligonucleotides (for a review see Moore, 1981). DNA cleavage proceeds one strand at a time and the enzyme shows no sequence specificity although it has been suggested that certain variations of groove width of the DNA can be detected (Drew & Travers, 1984). The crystal structure of bovine pancreatic DNAase I has been determined (Suck *et al.*, 1984; Oefner & Suck, 1986). The protein is an α/β protein with a core of two tightly packed six-stranded β-sheet structures surrounded by eight α-helices and several loop regions. The enzyme shows internal symmetry with an approximate two-fold axis relating two topologically equivalent structural elements, namely $\beta\alpha\beta\alpha\beta\beta\beta\beta$.

***Eco*RI and *Eco*RV endonucleases.** In contrast to DNAase I, *Eco*RI endonuclease (276 amino acids) and *Eco*RV endonuclease (244 amino acids) from *E. coli* are dimeric proteins that catalyse specific cleavages of double-stranded DNA by recognizing the hexanucleotides GAATTC and GATATC respectively (Rosenberg *et al.*, 1981; Taylor & Halford, 1989). Both enzymes require Mg^{2+} for activity and cleavage occurs between the G and A bases for *Eco*RI and the T and A bases for *Eco*RV. The crystal structure of *Eco*RI endonuclease complexed with the cognate oligonucleotide TCGCGAATTCGCG and in the absence of Mg^{2+} has been solved (Frederick *et al.*, 1984; McClarin *et al.*, 1986). A reinterpretation of the structure using more isomorphous derivatives has led to a different chain tracing and the new model suggests that *Eco*RI is an α/β protein consisting of a five-stranded β-sheet surrounded by α-helices (Kim *et al.*, 1990). However, a detailed description of the new structure is at present unavailable.

Recently, the crystal structures of *Eco*RV as the free protein and bound to cognate DNA (GGGATATCCC, one duplex per dimer) and non-cognate DNA (CGAGCTCG, two duplexes per dimer) have been elucidated (Winkler *et al.*, 1991; F. Winkler, personal communication). The structure of the enzyme itself shows *Eco*RV to be an α/β protein with protruding loops (Fig. 7). However, no structural similarity between *Eco*RV and *Eco*RI is observed, as expected from the lack of amino acid sequence similarity between both proteins.

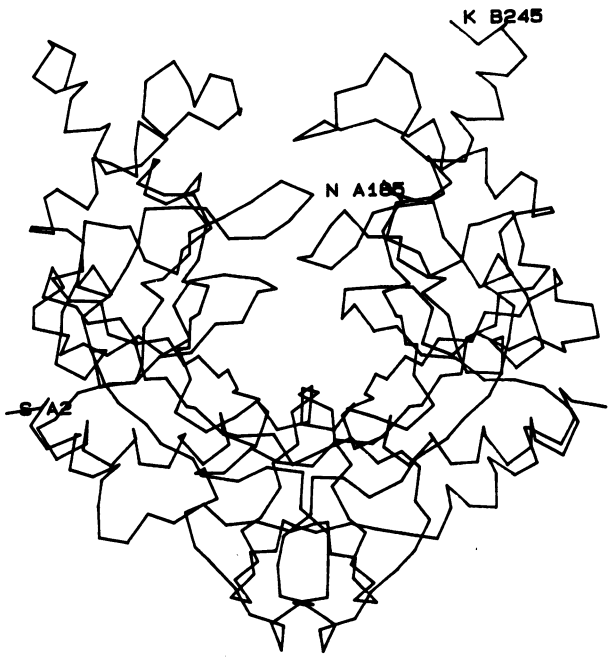

Fig. 7. Structure of *Eco*RV

An α-carbon drawing of the *Eco*RV endonuclease dimer. The *N*-terminus of subunit A and *C*-terminus of subunit B are labelled. The DNA interacts with loops which protrude from the sides of the 'U' shaped dimer. Sequence-specific contacts are made with the upper loop (residues 183–188) on the top surface of the molecule from this view. Kindly provided by Dr. F. Winkler, Hoffmann La Roche, Basel, Switzerland.

DNA polymerase I. DNA polymerase I in *E. coli* (928 amino acids) is a monomer with three enzymic activities: DNA polymerase, 3′–5′ exonuclease thought to edit out mismatched nucleotides, and a 5′–3′ exonuclease. Limited proteolysis cleaves the molecule into two fragments; the larger *C*-terminal domain (Klenow fragment, 605 amino acids) has both the DNA polymerase and 3′–5′ exonuclease activities, whereas the smaller *N*-terminal domain has the 5′–3′ exonuclease (for a recent review see Joyce, 1991). The crystal structure of the Klenow fragment complexed with thymidine monophosphate (dTMP) has been solved and shows that the protein is folded into two distinct domains (Ollis *et al.*, 1985; Beese & Steitz, 1991). The larger *C*-terminal domain forms a deep cleft, the bottom of which is made up by a six-stranded antiparallel β-sheet structure. The sides of the cleft are composed of α-helices and the whole cleft has dimensions which are suitable for binding duplex B-DNA. The smaller *N*-terminal domain has a central core of β-sheet with α-helices on both sides and in the crystal can bind dTMP and two divalent metal ions. Various experiments (reviewed in Joyce & Steitz, 1987; Freemont *et al.*, 1986; Derbyshire *et al.*, 1988) have shown that the *C*-terminal domain contains the active site for the polymerase reaction and the *N*-terminal domain catalyses the 3′–5′ exonuclease activity. DNA polymerase I has a strict requirement for base-paired primer-template DNA substrates and no sequence specificity, although under certain conditions single bases can be added to blunt-ended duplex (Clark *et al.*, 1987). It is immediately apparent from the structure of the Klenow fragment that the architecture of the protein is suitably designed to carry out 'error-free' DNA replication (see Fig. 12*c*). This is highlighted by the large cleft which binds double-stranded DNA, a domain that could clamp the DNA in place and an error-correcting domain with specificity for single-stranded DNA.

PROTEIN–DNA COMPLEXES

Even proteins that recognize specific target sequences of DNA have significant affinity for non-cognate sequences (or bulk DNA); the non-specific binding to DNA is probably an essential property of all DNA-binding proteins. Indeed, von Hippel *et al.* (1974) and Lin & Riggs (1975) have shown that, for simple bacterial repressors at least, the ratio of the affinity for specific binding to that of non-specific binding should be of the order 10^5 to allow control of expression in response to changes of allosteric effectors. DNA–protein interactions are in many instances highly dependent on the concentration of mono- and di-valent cations and pH. The phosphodiester backbone tends to attract positive ions, to the extent that up to 75% of the charge is neutralized. If a salt bridge is formed between a phosphate and a positively charged side chain in the protein, the counterion may be displaced. This process will be associated with a large increase in entropy as the small cation becomes able to move freely in bulk solution. For example, the *lac* repressor binds bulk DNA with the neutralisation of ten or eleven charges, whereas in the complex with the *lac* operator DNA, only seven or eight charges are neutralized by the protein, the extra free energy being supplied by a wealth of specific interactions that are not possible with the bulk DNA (Barkley & Bourgeois, 1980). However, because of the long range of the electrostatic potential of charges (it varies as $1/r$) and the complex arrangement of charges on proteins, electrostatic interactions are rather complex in practice, and what is important is the complementarity of the electrostatic potential surfaces of the protein and the DNA. Warwicker *et al.* (1987) have shown the importance of these considerations for the interaction of CAP with its operator site (see below).

It has been argued that the most important interactions for specificity are hydrogen bonding between protein side chains and functional groups exposed in the major groove of DNA (von Hippel *et al.*, 1974; von Hippel & Berg, 1986, 1989). As the differential free energy change for forming such a hydrogen bond may be small (as both groups may be hydrated in the free state), one contribution to specificity may be that in non-specific complexes many of the hydrogen bonds cannot be formed at all. The reason for concentrating on hydrogen bonding patterns derives from the considerations of Seeman *et al.* (1976), who showed that discrimination of base pairs by hydrogen bonding is possible in the major groove. Fig. 8 shows G·C and A·T base pairs and the functional groups that are exposed in the major and minor grooves. Clearly these base pairs could be distinguished by the formation of three hydrogen bonds, and, in the case of A·T, placement of an appropriate hydrophobic group next to the thymine methyl group. In fact, more than half of the amino acid residues are capable of hydrogen bonding, so that the repertoire of possible interactions for the protein is potentially quite large. However, the side chains of Asn and Gln are particularly versatile, because they contain both a hydrogen bond donor and acceptor, and the side chain is relatively free to rotate about the C^{β}–C^{γ} bond. It is therefore not surprising to find a high proportion of hydrophilic residues in DNA-binding domains (Table 3).

While hydrogen bonding is probably of great importance for sequence recognition, such simple schemes tend to emphasize the static view of both the protein and DNA; a favourable hydrogen bond can be made only if the groups in the DNA are in the appropriate position and orientation. In order to optimize the number of favourable contacts (or minimize the number of unfavourable contacts) it may be necessary for the DNA or the protein or both to change in conformation. The energy required to deform either or both molecules has to be paid for by the increased number of favourable interactions. In the following section, there are examples of all kinds of these interactions,

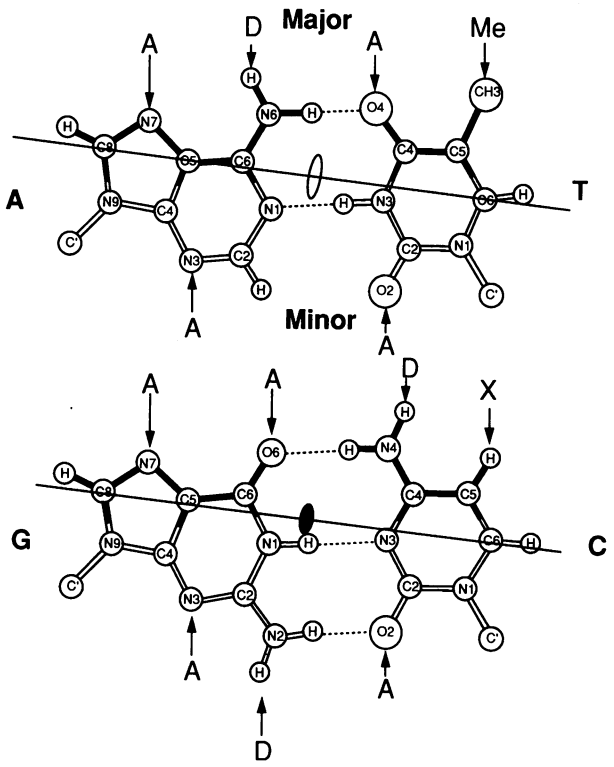

Fig. 8. Hydrogen-bonding patterns in protein–DNA interactions

The A·T and G·C base pairs are shown in the same relative orientation. A denotes an atom that can act as a hydrogen-bond acceptor, and D denotes an atom that can act as a hydrogen-bond donor. Major and minor refer to the two grooves of DNA. The pseudo two-fold axis of the base pairs is indicated by ●. X denotes no interaction at this position.

Table 3. DNA-binding protein recognition sequences

Bold type shows residues capable of participating in hydrogen bonding or other electrostatic interactions. In all proteins except the *metJ* repressor the residues shown are found in the recognition helix. In *metJ* repressor they form the β-strand.

Protein	Sequence
λ *cro* repressor	**YQSAINKAIHA**
P434 *cro* repressor	**KQQSIQLIEAG**
λCI repressor	**GQSGVGALFNG**
P434 repressor	**TQQSIEQLENGK**
trp repressor	GIATITRGSNSLKA
CAP	**SRETVGRILKM**
lac repressor	**SYQTVSRVVNQ**
ant homeodomain	**RQIKIWFQNRRMK**
eng homeodomain	**AQIKIWFQNKRAK**
Glucocorticoid receptor	**SCKVFFKRQVEQQH**
metJ repressor	**VKKITVSIP**

including variations in the apparent degree of distortion of both DNA and protein molecules.

λCI and P434 repressor–operator complexes

The structure of the DNA binding domain of λCI repressor complexed with a 17-bp consensus operator sequence has been determined (Jordan *et al.*, 1985; Jordan & Pabo, 1988). The structure shows that the protein binds to the operator site symmetrically, in accordance with the two-fold symmetry of the operator site, with one helix (helix 3) from each subunit fitting into the major groove of each half-site and the *N*-terminal arm of each monomer wrapping around the DNA making contacts with the central base-pairs of the operator site, as had previously been proposed (Fig. 9; Pabo & Sauer, 1984). The structure of the monomer in the co-crystal complex shows no significant differences to that of the monomer in the absence of DNA and the conformation of the DNA fits with the general geometric constraints of B-DNA with no noticeable distortions (Jordan & Pabo, 1988). However, the DNA is not uniform and local variations in geometry are observed from base to base. The extent to which this is due to either the binding of the repressor or crystal packing constraints or both is not known, as the structure of the DNA alone is not available. However, the solution structure of the OR3 half operator has been examined by n.m.r. (Baleja *et al.*, 1990), and is clearly in the B-family of conformations with small variations about the canonical structure. The major protein-DNA contacts are illustrated in Fig. 9 with the repressor making extensive contacts with DNA phosphate backbone and exposed edges of bases in the major groove. Jordan & Pabo (1988) suggest that the specificity for the interaction is due to the specific interactions of Gln-44 and Ser-45 at the amino end of the recognition helix and Lys-4 in the *N*-terminal arm and Asn-55 in a loop after the recognition helix. Furthermore, several side chains co-operate to recognize single bases. For example, Gln-33 is hydrogen-bonded to both the phosphate of adenine base-pair 2 and Gln-44, with Gln-44 directly interacting with the N-6 and N-7 of the adenine base itself. This type of co-operative interaction is also observed for Lys-4 and Asn-55 which both interact with the O-6 and N-7 of guanine base-pair 6 respectively.

Although these models give considerable insight into the role of different amino acid residues in making sequence-specific contacts, Senear & Ackers (1990) have shown that repressor–operator binding is accompanied by proton uptake, which involves acidic groups not involved in direct contacts with the DNA. As the proton uptake is different according to which of the three operators is bound, it appears that the binding specificity is determined at least in part by more global properties of the protein, including conformation in the bound state and the ionization state (p*K* values) of titratable groups elsewhere in the binding interface.

Recently, the structure of the *N*-terminal domain of P434 repressor complexed with a 20-bp DNA fragment has been solved (Aggarwal *et al.*, 1988). The structure shows that P434 repressor binds to its operator site in a symmetrical fashion, without significant structural changes in the protein, analogous to the λCI repressor–operator complex (Fig. 9). An α-helix (helix 3) from each monomer is fixed into the major groove and the *N*-termini of helices 2 and 4 are close to the phosphate backbone (Aggarwal *et al.*, 1988). However, the DNA, although B-type, is distorted by bending and twisting, giving rise to a compression of the minor groove (11.5 to 8.8 Å) at the centre of the complex, resulting in a network of bifurcated hydrogen bonds among non-coplanar base-pairs (Anderson *et al.*, 1987; Aggarwal *et al.*, 1988). This is in direct contrast to the λCI repressor–operator complex where the DNA is generally of B-type (see above). The major protein–DNA contacts show several co-operative networks of hydrogen bonds which are similar to those observed in λCI repressor (e.g. base-pair 1 in P434 repressor and base-pair 2 in λCI repressor). There are also a number of solvent-mediated protein–DNA interactions similar to those observed in the *trp* repressor–operator complex (Fig. 10). A detailed comparison of the λCI and P434 repressor–operator complexes shows that three conserved residues in the helix–turn–helix structure (Gln-33,

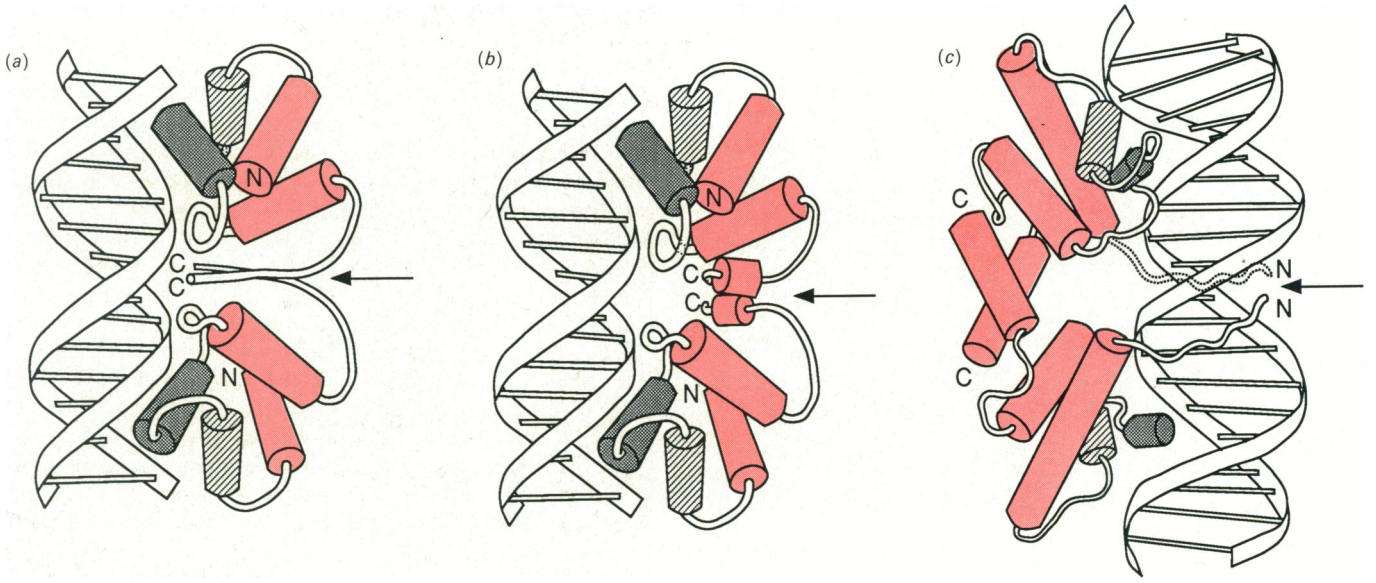

Fig. 9. Interaction of P434, P434 *cro* and λCI repressors with DNA

Schematic representation of P434, P434 *cro* and λCI repressor–operator complexes as derived from their crystallographic structures with α-helices represented as cylinders and the *N*- and *C*-termini labelled respectively (for details see the text). The molecular two-fold axis of each complex is indicated by an arrow and lies coincident with the dyad axis of the operator sequence. The DNA structures are represented as ribbons. (*a*) P434 repressor–operator complex (adapted from Aggarwal *et al.*, 1988). It should be noted that, on refinement, the repressor showed a short *C*-terminal helix similar to that in the P434 *cro* repressor (Harrison & Aggarwal, 1990). (*b*) P434 *cro* repressor–operator complex (adapted from Wolberger *et al.*, 1988). (*c*) λCI repressor–operator complex (adapted from Jordan & Pabo, 1988).

Table 4. Conformational changes in protein–DNA complexes

Complex	Protein	DNA	DNA interaction site
λCI repressor	No	No	Major groove
λ *cro* repressor	Yes (subunit rotation)	Yes (40° bend)	Major groove
P434 repressor	No	Yes (minor groove compression)	Major groove
P434 *cro* repressor	No	No	Major groove
trp repressor	Yes (rotation of helix–turn–helix unit)	Yes (60° bend, minor groove widening)	Major groove
lac repressor	Yes (rotation of helix)	No	Major groove
CAP	No?	Bend > 90° (two 45° kinks)	Major groove
ant homeodomain protein	(Yes)	No	Major groove, arm in minor groove
eng homeodomain protein	Not determined	Yes	Major groove, arm in minor groove
metJ repressor groove	No	Yes	(25° bend, major groove compression)
DNAase I endonuclease	No	Yes (minor groove widening)	Minor groove
*Eco*RI endonuclease	No	Yes (three kinks, major groove widening)	Major groove
*Eco*RV endonuclease	Yes	Yes	Major groove

Gln-44 and Asn-52 in λCI repressor; Gln-17, Gln-28, and Asn-36 in P434 repressor) make similar contacts with the DNA phosphate backbone (Pabo *et al.*, 1990; for a review see Harrison & Aggarwal, 1990). It is suggested that such contacts are important for orientating and positioning the helix–turn–helix structure, allowing sequence-specific interaction, and that conservative contacts may be a feature of helix–turn–helix protein families (Pabo *et al.*, 1990). However one noticeable difference between λCI and P434 repressors in terms of specific protein–DNA contacts is the interaction of Arg-43 of P434 repressor with phosphates straddling the minor groove near the central base pairs of the operator sequence. It has been suggested that these contacts are important for stabilizing the compressed minor groove and therefore bear relevance to operator specificity (Harrison & Aggarwal, 1990).

The major difference between λCI and P434 repressors in terms of repressor–operator recognition lies in the ability of P434 repressor to utilize different sequence-specific conformations of DNA for operator specificity (Kouldelka *et al.*, 1988). This is probably due to differences in the structures of the two repressors with λCI repressor using an extended 'arm' to contact central base-pairs in the operator (Jordan & Pabo, 1988) and P434 repressor relying on ease of interaction as a function of operator sequence (Aggarwal *et al.*, 1988). However, a recent analysis of DNA torsional constants from different DNA sequences (Fujimoto & Schurr, 1990), suggests that the binding of P434 repressor to its operator site could arise from variations in the equilibrium structures of the DNA rather than the deformability of the DNA, as proposed by Kouldelka *et al.* (1988).

In conclusion, the specificity of λCI and P434 repressors for

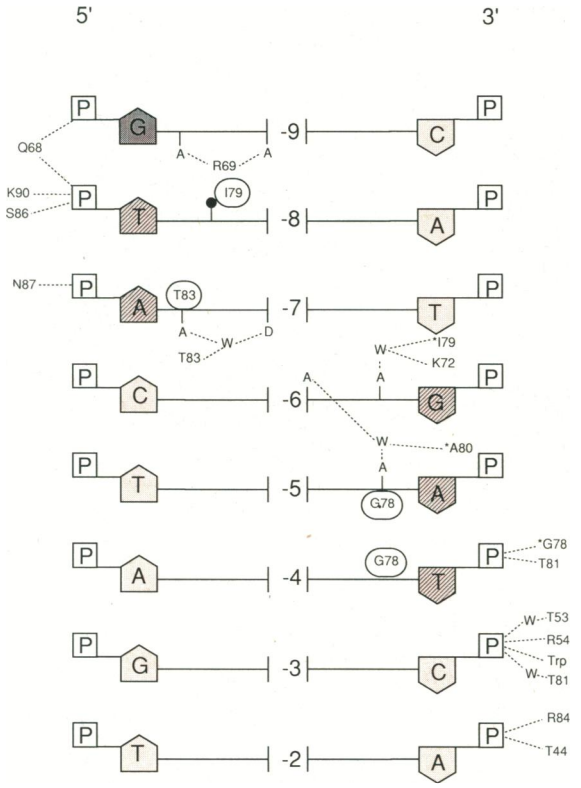

Fig. 10. Interaction of the *trp* repressor with DNA

A schematic representation of the specific protein–DNA contacts observed in the *trp* repressor–operator crystal structure (Otwinowski *et al.*, 1988). Amino acids are indicated by the single letter code. Amino acid interactions are shown as dotted lines. The phosphate (□), sugar (⬡) and base (⊣) make up one nucleotide. Hydrogen bond donors (D) and acceptors (A) are indicated on the bases involved in specific amino acid interactions. Non-polar interactions (-●) are also shown. Amino acids that form hydrogen bonds through main chain atoms are flagged with an asterisk (*). Interactions that occur in the major and minor groove are indicated by hatched and stippled shading of the specific bases respectively. The base numbering between the nucleotides is taken from the original references. W, water atoms.

their respective operator sites not only relies on specific amino acid–base contacts but on a network of non-specific phosphate backbone contacts which surround the 'recognition' helix, possibly enhancing specificity by positioning correctly the amino acids that contact bases. Furthermore the sequence-dependent flexibility of the DNA operator sites could also play a role in recognition and specificity.

λ *cro* and P434 *cro* repressor–operator complexes

The structure of λ *cro* repressor bound to a 17-bp operator site has recently been solved to moderate resolution (3.9 Å; Brennan *et al.*, 1990). In the λ *cro* repressor–DNA complex, a large conformational change of the protein dimer, relative to the structure of the protein alone, is observed. This movement is achieved by a twisting of the β-sheet strands that make up the dimer interface and results in a 40° rotation of one subunit relative to the other (Brennan *et al.*, 1990). The structure of the DNA within the complex is essentially B-form but is bent overall by 40°. The structure also shows that the recognition helices lie in successive major grooves of the DNA, as previously suggested, and that residues in the other helices determine the orientation and binding of the helix–turn–helix structure (Brennan *et al.*, 1990). The λ *cro* repressor–operator interaction has been subjected to extensive genetic analysis which has resulted in controversy surrounding the precise protein–DNA contacts inferred from these studies. Unfortunately, the side chains within the co-crystal structure cannot be unequivocally resolved (due to the moderate resolution of the structure) and therefore confirmation of specific protein–DNA contacts is difficult. N.m.r. experiments studying the binding of λ *cro* to its specific operator sequence have shown that there are differential changes of n.m.r. chemical shifts of the imino protons on forming the specific complex (Kirpichnikov *et al.*, 1984*a*,*b*) which are consistent with localized conformational change in the DNA. Further, the p*K* value of His-35 increases in the presence of DNA (Kirpichnikov *et al.*, 1984*a*), consistent with the residue approaching the phosphate backbone in a manner similar to that observed with *lac* repressor (see below). The interaction of His-35 is further supported by experiments that probe surface accessibility as His-35 becomes inaccessible to solvent in the presence of the operator (Shirakawa *et al.*, 1985). Further, the last four residues at the *C*-terminal end are dynamically disordered in the free protein, but become more ordered in the complex with DNA, suggesting that these residues make significant contacts with the DNA (Leighton & Lu, 1987).

The structure of the P434 *cro* protein complexed with a 14-bp fragment of DNA has been solved from anisotropically diffracting co-crystals (3.2 Å and 5.5 Å in perpendicular directions; Wolberger *et al.*, 1988). The structure of the complex is very similar to that of P434 repressor, with P434 *cro* repressor interacting with the operator through an α-helix (helix 3) in the major groove (Fig. 9; Wolberger *et al.*, 1988). It should be noted that due to the anisotropy of the crystals, the final electron density map showed density for only a few side chains and therefore made the interpretation of specific amino acid–DNA contacts difficult. A comparison of the P434 and P434 *cro* repressor–operator structures shows that the structure of the DNA in the P434 *cro* repressor complex is very different from that observed in the P434 repressor complex (see above). The DNA is B-type, straight and uniformly overwound with little variation in the widths of the major and minor grooves, whereas the DNA complexed to P434 repressor is bent, distorted and shows striking variation in minor groove width. Furthermore, the orientation of the monomers relative to the DNA helical axis of the operator sites are different in both cases, although a number of similar amino acid–DNA interactions are observed. Although P434 and P434 *cro* repressors seem to present to the DNA identical surfaces of interaction, a number of different protein–DNA interactions are observed between both proteins which are concentrated around base pairs 4 and 5. Therefore, P434 and P434 *cro* repressors recognise similar operator sites differentially by using a combination of specific amino acid–base interactions and the ability of each protein to bend specific operator sequences (Wolberger *et al.*, 1988). This argues for a more complex protein–DNA recognition code where specificity results from a combination of many factors, including DNA flexibility and protein orientation when complexed to DNA.

trp repressor–operator complex

The structure of *trp* repressor complexed to a symmetric 18-bp operator site has been solved (Otwinowski *et al.*, 1988). The repressor binds symmetrically to the operator site with the long axis of one helix of the helix–turn–helix motif pointing into the major groove of the DNA, almost perpendicular to the DNA helical axis (Otwinowski *et al.*, 1988). This is in contrast to the other repressor–operator complexes where the 'recognition' helix lies almost parallel to the major groove axis. The conformation of the central core of the repressor when bound to DNA is

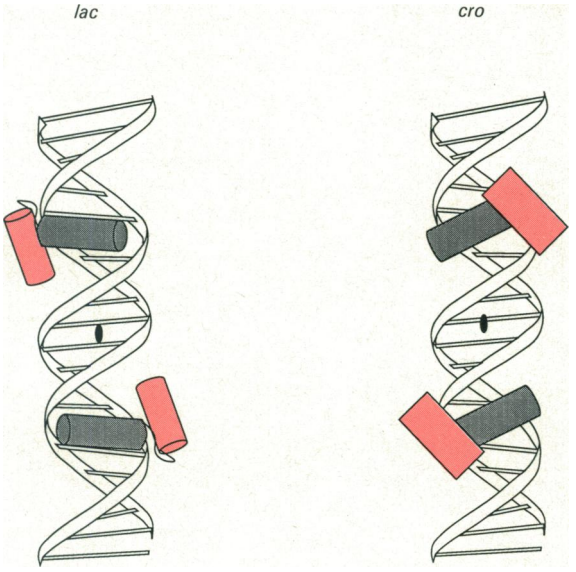

Fig. 11. Structure of the *lac* repressor headpiece–operator complex

The helix–turn–helix motif only is shown, with the recognition helix shaded grey. The relative orientation of the recognition helix of the *lac* repressor headpiece with respect to the DNA is compared with that in the λ *cro*–DNA complex. Adapted from Kaptein *et al.* (1989).

similar to the unbound form of the repressor, but the orientation of the helix–turn–helix motif shows significant differences (Otwinowski *et al.*, 1988). The *N*-terminal arm of the repressor, which has been recently suggested to increase the affinity of the repressor for its operator site (Carey, 1989), is poorly ordered in the co-crystal structure and therefore a structural role cannot be assigned. Moreover, repressor–operator binding studies performed at physiological ionic strength and pH indicate that removal of the first seven residues decreases specific operator affinity by only a factor of 2, indicating that these residues make no significant contribution to binding specificity (Marmorstein *et al.*, 1991). The structure of the operator in the complex is of B-type conformation, although deviations in local geometry of the base-pairs is observed, resulting in a widening of the minor groove (Otwinowski *et al.*, 1988). The DNA therefore has a significant bend (Table 4).

Surprisingly, the co-crystal structure shows no direct amino acid–base interactions, with most of the direct contacts made to the DNA phosphate backbone (Fig. 10). However, a number of solvent-mediated hydrogen bonds to individual bases are observed, which is in direct contrast to all of the other repressor–operator structures. Sigler and colleagues argue that DNA-sequence-specific recognition is achieved by the sequence-dependent conformation of the operator site, which allows the formation of 24 direct hydrogen bonds between the protein and the DNA phosphate backbone (Fig. 10; Otwinowski *et al.*, 1988). This they term the 'indirect readout mechanism', which is to be contrasted with the direct mechanism where hydrogen bonds are formed to functional groups on the bases directly by amino acids side chains. They also suggest that such specific conformations would increase the solvent-excluded contact surface and that non-specific sequences may achieve favourable conformations but at too high an energetic cost (Otwinowski *et al.*, 1988; Luisi & Sigler, 1990).

However, recent biochemical analyses of mutant *trp* repressors have suggested that Thr-81 is directly involved in sequence-specific recognition by making direct contacts with base-pairs (Bass *et al.*, 1988). In the *trp* repressor–operator structure, Thr-81 makes a hydrogen bond with the phosphate backbone but does not make any base-specific interactions (see Fig. 10). It has been suggested that the *trp* repressor–operator structure could represent a non-specific complex which results from the crystallization conditions (Matthews, 1988). In accordance with this view, the operator sequence used for co-crystallization may not allow specific complex formation if the sequence is considered in terms of tandem repeats (Phillips *et al.*, 1989; Staacke *et al.*, 1990). Nevertheless, direct binding studies with 20-bp operators and consensus operators embedded within 80-bp restriction fragments all point to specific 1:1 complexes with K_d in the range 0.5–5 nM in the presence of tryptophan, and much reduced affinity in the absence of tryptophan (Chandler & Lane, 1988; Klig *et al.*, 1988; Carey, 1989; Marmorstein & Sigler, 1989; Marmorstein *et al.*, 1991). In contrast, alternative operators, as discussed by Staacke *et al.* (1990), show only non-specific binding *in vitro* (P. Beckmann & A. N. Lane, unpublished work). Although these biochemical experiments indicate that the 20-bp consensus *trp* operator contains all the necessary information for recognition by the holorepressor, it is still unknown whether the effective stoichiometry for the *aroH* and *trpR* operators is 1:1 or higher.

In summary, the *trp* repressor–operator structure is unusual in that DNA sequence specificity is not achieved by direct amino acid–base interaction, as observed in all of the other repressor-operator complexes. It is suggested that sequence specificity for *trp* repressor may be achieved by the conformation of the operator site allowing favourable repressor–DNA interactions, with base specific contacts mediated by solvent.

lac repressor–operator complex

The structure of the isolated DNA-binding domain (headpiece) of the *lac* repressor was determined by n.m.r. (see above). The *lac* operator sequence has also been examined by n.m.r. The early work on the complex focussed on cataloguing the changes in the n.m.r. spectra of the repressor and operator fragments on forming the complex. For example, the p*K* of His-29 increases on forming the specific complex, consistent with proximity to a negative potential (Scheek *et al.*, 1983). Experiments were also carried out to probe the surface accessibility of aromatic residues, which showed that Tyr-7 and Tyr-17, which are in the recognition helix, and His-29, become buried in the protein–DNA complex, suggesting that these residues are in the protein–DNA interface (Stob *et al.*, 1988). The n.m.r. data also showed that these tyrosine residues do not intercalate, suggesting the possibility of hydrogen bonding either to the bases or to the phosphates. His-29 actually lies in the loop region between helices 2 and 3, which is also less well determined than the helix–turn–helix motif.

N.m.r. titrations with a 14-bp half operator and a 22-bp *lac* operator showed significant changes in chemical shifts for certain bases as follows:

```
1   5   10
GGAA**TT***G*TGAGCGG
CCTT**AA**C**A**CTCGCC
```

where bold face indicates changes for the imino protons, italic for the H8 (major groove), and underline for H1′ (minor groove). In the 22-bp operator, two *lac* repressor monomers were simultaneously bound, but the complex retains two-fold symmetry, indicating equivalent binding contacts for the two subunits. The absence of large chemical shift changes outside the stretch of five base pairs suggests the absence of large-scale conformational changes in the DNA (such as bending or global unwinding), although the interaction of the intact repressor with DNA is known to cause an unwinding of approx. 90° (Barkley &

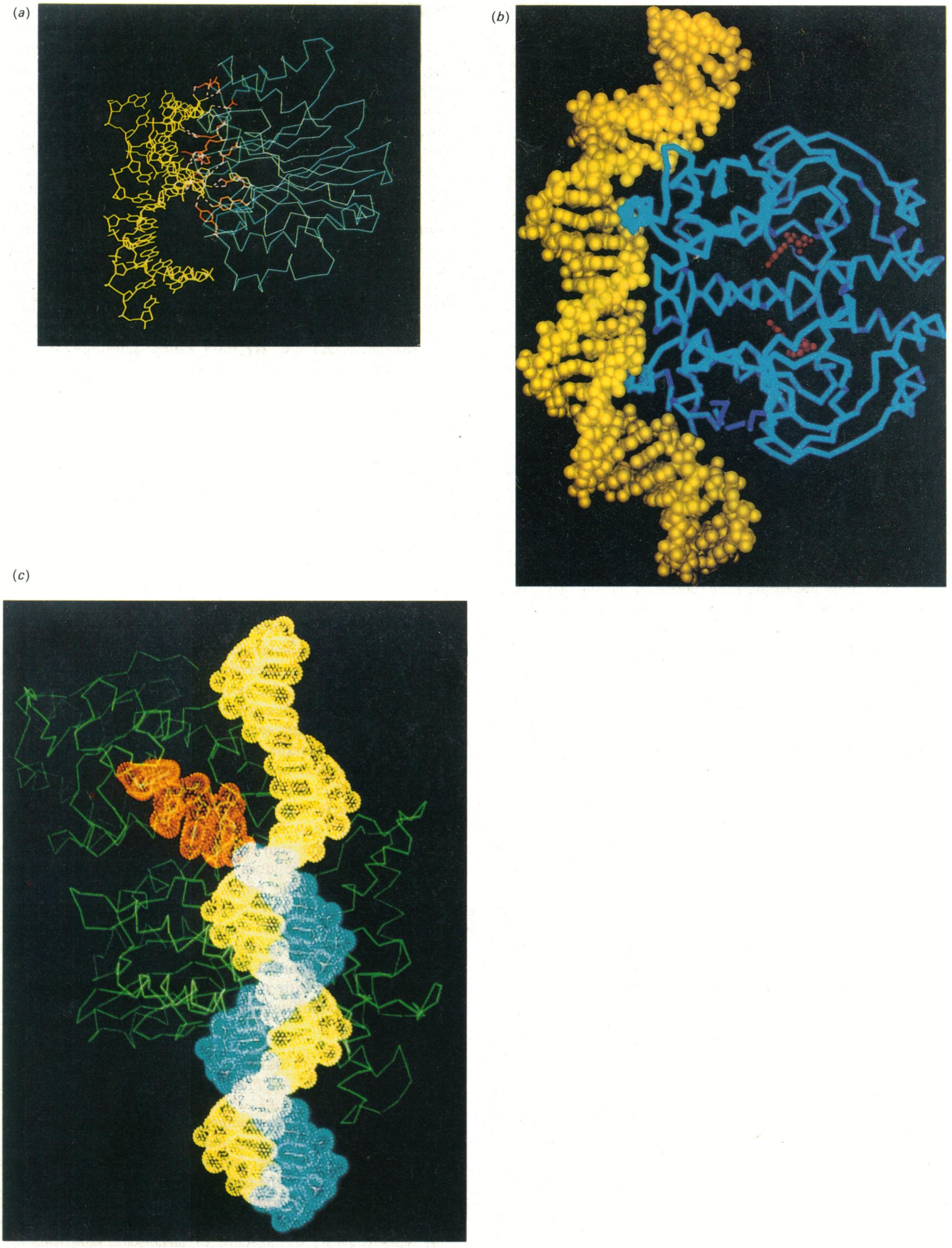

Fig. 12. Structures of protein–DNA complexes

(*a*) Interaction of DNAase I with DNA. The DNA is coloured yellow and the C^{α} carbon atoms of the protein blue. The residues which make specific DNA contacts (red) interact in the minor groove. Reprinted by permission from Suck *et al.* (1988). Copyright © 1988 Macmillan Magazines Ltd.

Bourgeois, 1980). However, the chemical shift changes of the imino protons, and also the 1′ protons in the minor groove, are consistent with local conformation changes in the DNA. Nevertheless, the DNA in these complexes clearly remains within the B-family of conformations.

Direct proximities between protein side-chains and the bases were established from NOESY experiments, which yield information about distances between pairs of protons of up to about 5 Å. Unambiguous NOE contacts were observed for Thr^5–G^{10}, Leu^6–C^9, Tyr^7–C^9 and G^{10} (all helix 1), Tyr^{17}–T^8 and C^9 (helix 2) and His^{29}–A^2 and T^3 (loop). Model building in which all the known constraints were simultaneously satisfied place the recognition helix (helix 2) in the major groove, essentially at right angles to the DNA helical axis, while the first helix has its *N*-terminus close to the phosphate backbone. Also, the loop containing His-29 is close to the phosphate backbone at the opposite side of the major groove. This model accounts for all the inferred contacts obtained from genetic experiments, and also for biochemical data obtained for the intact repressor–operator complex (Kaptein *et al.*, 1989). Measurements of changes in the phosphorus n.m.r. spectrum of the *lac* operator on repressor binding are also consistent with the TGTGA segment being close to the protein (Karslake *et al.*, 1990). The most recent analysis of the complex, in which many more NOEs were identified, showed that the structure of the protein also changes slightly in the complex, possibly due to a rotation of helix 2 (Lamerichs *et al.*, 1990).

A model of the complex that accounts for all of the n.m.r. and genetic observations is shown in Fig. 11. Interestingly, the relative orientation of the recognition helix is nearly 180° opposite to that found for λ *cro* repressor, in which the recognition helix also lies parallel to the major groove. It should also be mentioned that the biochemical data for *lac* repressor were originally interpreted in terms of a model built in the λ *cro* repressor orientation; clearly the interpretations of such experiments are at the mercy of the structural assumptions made.

CAP–operator complex

CAP when bound to its specific operator sequences has been shown to produce a large bend (greater than 90°) in the DNA, as detected by polyacrylamide gel electrophoresis (Wu & Crothers, 1984; Gartenberg & Crothers, 1988; Zinkel & Crothers, 1990). Warwicker *et al.* (1987) have used electrostatic calculations and model building to show that the region of positive electrostatic potential around the protein would be sufficient to stabilize a bent form of DNA in which contacts would be made in a site of about 28-bp, in agreement with biochemical protection data. The large bend would also account for the abnormal electrophoretic mobilities of CAP–DNA complexes (Gartenberg & Crothers, 1988). Recently, crystals have been grown of CAP complexed with cyclic AMP and a 30-bp DNA fragment having a 5′ base overhang, and nicks on the top and bottom strand two base-pairs either side of the dyad axis. The structure of this complex has been solved (Schultz *et al.*, 1990; Steitz, 1990). In the structure of the complex the F helices fit deeply into the major grooves of the DNA as observed in other repressor–DNA complexes. The DNA, however, is highly bent with an overall bend of about 90° in close agreement with the Warwicker model (Fig. 12*b*). The bend is produced by two 45° kinks at G·C and A·T base pairs 5 and 6 away from the dyad axis, i.e. three base-pairs removed from the nicks. Further, the DNA is bent in the same direction at each kink, indicating that the nicks are not directly involved in the bending.

engrailed and *antennapedia* homeodomain protein–DNA complexes

The structure of the *antennapedia* homeodomain protein complexed with a 14-bp DNA target has been solved by n.m.r. methods (Otting *et al.*, 1990), and a similar *engrailed* homeodomain protein complex with a 21-bp DNA fragment has been determined by X-ray crystallography (Kissinger *et al.*, 1990). The orientations of the proteins in the two complexes are extremely similar, in which the extended recognition helix (helix 3, see above) lies in the major groove perpendicular to the helix axis (Fig. 13*a*). However, the presumed sequence-specific contacts are towards the centre of the recognition helix, in contrast with the prokaryotic repressors where they tend to be toward the *N*-terminal end of this helix. Also, the relative placement of the first helix in the helix–turn–helix motif is different from any of the other complexes. Thus, all six independently-determined helix–turn–helix–DNA interactions use different orientations of the DNA-binding motif in their respective complexes.

In the *antennapedia* protein–DNA complex, the n.m.r. spectra of the DNA component are typical of B-conformation, suggesting that only small rearrangements of the DNA occur upon complex formation. Similarly, for the most part, the conformations of the helical regions of the protein are largely unchanged in the complex (and are indeed stabilized relative to the free protein); the exception was in parts of helix 2, which may become somewhat distorted in the complex (Otting *et al.*, 1990). Similarly, in the *engrailed* protein–DNA complex, the DNA has an overall twist of 34°, typical of B-DNA, though the base-pairs adjacent to the TAAT subsite show large (16–20°) base-pair tilts, which requires a certain degree of bending in this part of the DNA molecule.

Both complexes show that the *N*-terminal region, which is disordered in the free protein (see above), makes direct contacts, particularly via a conserved arginine residue, in the minor groove of the DNA. In contrast, the *N*-terminal extension of λCI repressor makes contacts in the major groove. In the *engrailed* protein–DNA complex, specific contacts are made in the major groove by Ile-47 which interacts with a thymine at position 4 (TAAT), Gln-50 which interacts with bases near the 3′ end (TAATNN) and Asn-51 which hydrogen-bonds to an adenine at position 3 (TAAT) (Fig. 13*b*). There are also extensive contacts with the phosphate backbone; however, the small number of specific base contacts suggests a moderate amount of sequence specificity. Therefore, other mechanisms of DNA specificity may be employed, including co-operativity of binding with other homeodomain proteins and/or regulatory proteins (Kissinger *et al.*, 1990). It is interesting to note that the n.m.r. and X-ray structure analysis of the *antennapedia* and *engrailed* protein–DNA complexes are consistent with each other. This is highlighted by the agreement between the twelve n.m.r. distances (from six amino acids) used to position *antennapedia* on the DNA in the n.m.r. analysis and the same distances as measured from the *engrailed* crystallographic model.

(*b*) Crystal structure of CAP complexed to a 30-bp DNA-binding site. The CAP α-carbon backbone is shown in blue, cyclic AMP is shown in red and the DNA as a space-filling representation in yellow. The DNA has an overall bend of about 90° resulting from two 45° kinks between base pairs 5 and 6 from the dyad. From Steitz (1990), with permission. (*c*) DNA polymerase I–DNA interaction. A model of the Klenow fragment bound to DNA showing the experimentally derived single-stranded DNA in the exonuclease active site (orange) and model-built DNA in the large cleft (3′ strand in blue, 5′ strand in yellow). The DNA is represented as a skeletal model superimposed on a van der Waals' dot surface. The C^{α} atoms of the Klenow fragment are shown in green. From Freemont *et al.* (1988), with permission.

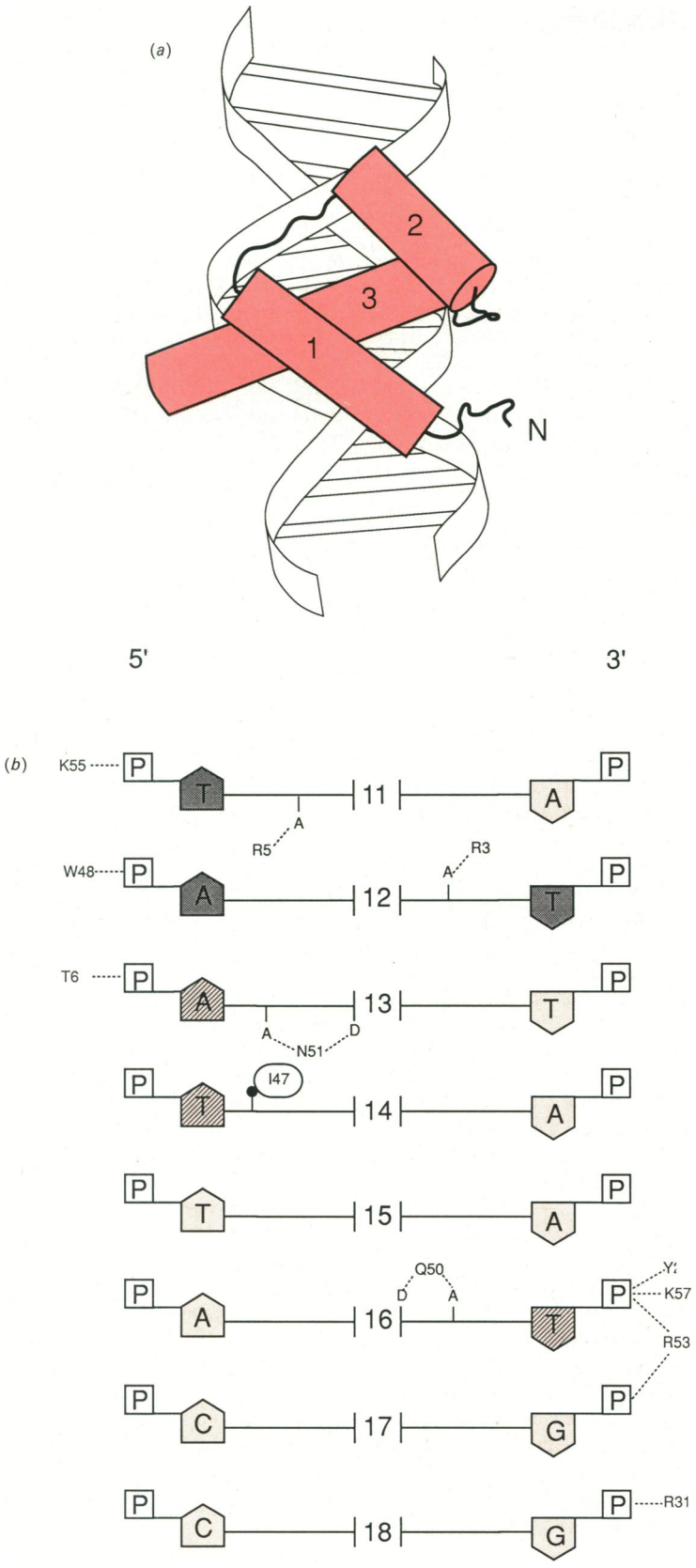

Fig. 13. Structure of the *engrailed* homeodomain protein–DNA complex

(*a*) The cartoon shows the orientation of the recognition helix (helix 3) in the major groove of the DNA. Adapted from Kissinger *et al.* (1990). (*b*) A schematic representation of the specific protein–DNA interactions observed in the *engrailed* homeodomain protein–DNA complex (Kissinger *et al.*, 1990). For a description of the symbols used, see legend to Fig. 10.

Glucocorticoid receptor–DNA complex

Recently, the crystal structure of the DNA-binding region of the glucocorticoid receptor (86 amino acids; residues 440–525) complexed to two palindromic 18-bp DNA duplexes have been solved to moderate resolution (2.9 and 4.0 Å; Luisi *et al.*, unpublished work). The DNA fragments differed only in the spacing between the hexamer half-sites which allowed the observation of specific and non-specific interactions within the higher resolution complex. The structure shows that the protein binds as a dimer to the DNA oligomer with each domain consisting of two zinc co-ordinating substructures of distinct conformation and function. The overall conformation of the bound subunits is very similar to those determined for the free DNA-binding domains determined by n.m.r. methods (see zinc fingers); the rms deviation between the structures is only 2 Å. The dimerization interface is made up of the two *C*-terminal zinc-motifs in which hydrophobic contacts and salt bridges form the major stabilizing interactions. Dimerization occurs even though only one subunit makes specific contacts with the DNA, indicating that the dimerization energy is significant even for this truncated protein (and see above).

The two protein domains bind on one side of the 18-bp fragment with the *N*-terminal α-helices of each domain lying in successive major grooves forming the base-specific contacts. This is yet another example of specific protein–DNA contact by an α-helix–major groove interaction, first highlighted in the prokaryotic repressor–DNA complexes. The specific interactions in one half-site involve hydrogen bonds between Arg-466 and G-4, a hydrophobic interaction between Val-462 and the methyl group of T-5 and a water-mediated hydrogen bond between Lys-461 and G-7. Interestingly, in the non-specific half-site, Arg-466 forms a hydrogen bond to a phosphate group and Val-462 is too far from T5. There is an extensive array of amino acid–phosphate contacts common to both the specific and non-specific complexes. The structure of the DNA within the complex adopts B-type conformation with no major deformations, although the major groove of the specific site is widened by 2 Å relative to the non-specific site.

Zif268–DNA complex

The first high-resolution crystal structure of a zinc finger–DNA complex has recently been determined using a complex containing the three fingers of Zif268 (89 residues; 349–421) a mouse immediate early protein, and an 11-mer consensus binding site (Pavletich & Pabo, 1991). The structure shows that the α-helix of each zinc finger fits into the major groove and that residues from the *N*-terminal part of each helix are responsible for making base-specific contacts to each 3-bp subsite. The complex also shows a simple periodicity in that each of the fingers is related in a way that reflects the periodicity of the 3-bp subsites (Fig. 14). Overall, the structures of the three fingers are very similar (rms deviation 0.45 Å), and finger 2 aligns best with the n.m.r. structure of Xfin31 (rms deviation 0.74 Å).

A number of base-specific interactions are observed and a simple pattern of DNA recognition can be deduced. Interestingly, fingers 1 and 3 recognize the same subsite (GCG) and have the same residues at critical positions. Finger 2, however, has different residues at these positions and recognises a different subsite (TGG). In summary, it appears that the residue immediately preceding the α-helix of each finger (a conserved Arg) contacts the third base on the primary stand of the subsite (5′--G). This interaction is stabilized by a side-chain-to-side-chain contact between the second residue of the helix (a conserved Asp) and the guanidinium group of the Arg. The third residue on the α-helix (finger 2, His) can contact the second base of the subsite (5′-G-) and the sixth residue (for fingers 1 and 3, Arg) can contact the

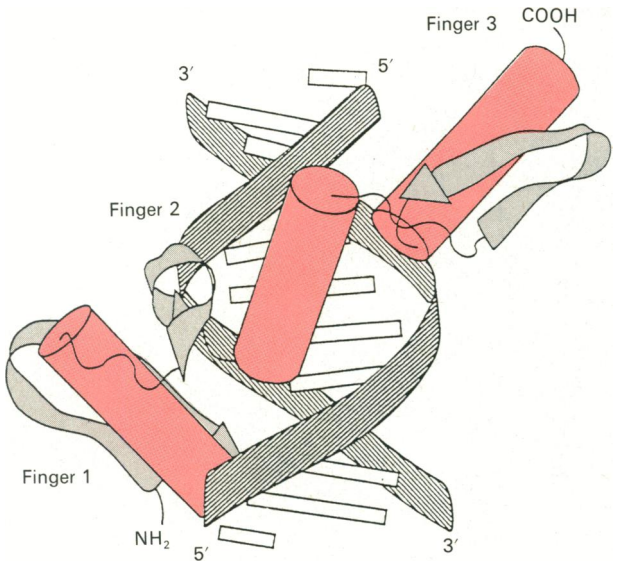

Fig. 14. Zif269 three zinc finger–DNA complex

A schematic representation of the crystal structure of Zif269 three zinc finger–DNA complex. The α-helices and β-strands are shown as cylinders and ribbons respectively. The three fingers wrap around the major groove with base specific contacts between side chains near the *N*-termini of each helix and 5′ 3-bp subsites (adapted from Pavletich & Pabo, 1991).

first base (5′ G--). It is noticeable that most of the DNA contacts are made to the G-rich strand and that the helix of each finger is antiparallel to this strand with the *N*-terminus near the 3′ end of each subsite (see Fig. 14). A number of phosphate DNA–backbone contacts are also observed between the zinc co-ordinating histidines and phosphate groups. Pavletich & Pabo (1991) suggest that these histidine–phosphate interactions will be widely conserved between zinc finger proteins and therefore could play a key role in positioning some of the base-specific contacts. The linker regions between each of the fingers make no important contacts with the DNA or the rest of the protein, but presumably are important in controlling the orientation and spacing of adjacent fingers. The structure of the DNA within the complex is of B-type with no major distortions or bends, though the helical twist angles alternate between high and low values regularly along the sequence.

To summarize, the Zif269 zinc finger–DNA complex again shows the importance of α-helix–major groove interactions as a motif for protein–DNA recognition. However, the use of α-helices by Zif269 is significantly different from other well-characterized protein–DNA complexes. Unlike the helix–turn–helix containing proteins, the zinc finger complex is based on modular recognition whereby individual zinc fingers recognize and contact 3-bp DNA sites, mainly on one strand, and that the periodicity of the protein structure is a function of the periodicity of the double helical DNA. There are also fewer protein–DNA backbone contacts and it appears that base contacts are more important for positioning and orienting the fingers onto the DNA. However, using the framework of the Zif269 three-finger structure as a basis for multiple finger–DNA interaction would be difficult although not unattractive, as the fingers could spiral around the major groove recognizing 3-bp subsites, as previously suggested by Berg (1988).

metJ repressor–operator complex

The crystal structure of the *metJ* repressor from *E. coli* bound to an 18-bp operator site containing two tandem *metJ* boxes (AGACGTCT) has recently been solved (Phillips, 1991). The structure shows two repressor dimers bound to the DNA with the β-ribbons of each dimer (two antiparallel β-strands, one from each subunit) lying in the major groove of the DNA (Fig. 15). The crystallographic dyad axes are coincident with the dyad axes of the *metJ* box operator sequences. The structure of the DNA within the complex shows a significant compression of the major groove compared to ideal B-DNA, which results in a 25° bend of the DNA-helix axis towards the protein (Phillips, 1991). Specific amino acid–base contacts are formed by hydrogen bonding between Lys-23 and G-2 and Thr-25 and A-3 from each β-strand, making a total of eight hydrogen bonds to bases within the complex. A superhelical array of repressors around the DNA is also observed with co-operative protein–protein contacts formed between antiparallel helices (helix A in Fig. 15) from each dimer unit. A number of phosphate contacts are made at the *N*-terminal region of helix B, with the remainder of the *metJ* repressor structure involved in the binding of corepressor. However, it is not clear from the structure of the complex as to the role of *S*-adenosylmethionine in terms of increasing the affinity of the *metJ* repressor for its operator sites.

This is the first detailed example of a β-sheet–DNA interaction and extends the diversity of protein structural units capable of binding specifically to DNA.

DNAase I–DNA complex

The crystal structure of DNAase I complexed with a nicked DNA octanucleotide has been determined (Suck *et al.*, 1988). The structure shows that DNAase I interacts with the octa-nucleotide in the minor groove, using a shallow cleft formed between the two-fold related equivalent halves of the protein structure. No significant conformational change occurs upon DNA binding (Fig. 12*a*; Suck *et al.*, 1988). The structure of the DNA in the co-crystal is of B-type conformation, although substantial distortions are observed, including bending and a widening of the minor groove by 5 Å, caused by the tight interaction of a small peptide loop (Suck *et al.*, 1988). Most of the protein–DNA contacts occur between DNA backbone phosphates and main-chain or side-chain groups of the protein, rather than hydrogen-bonding between the functional groups on the DNA bases and in the protein. This is reminiscent of the *trp* repressor–operator complex, though *Trp* binds to the major groove of the DNA. Although DNAase I displays no sequence specificity, it has been proposed that the enzyme is sensitive to local variations in DNA conformation which result from different nucleotide sequences (Drew & Travers, 1985; Nelson *et al.*, 1987; Suck *et al.*, 1988). Therefore, the observed bending in the DNAase I–DNA complex could provide a mechanism for hypersensitive cleavage sites which have sequences that bend, either intrinsically or by protein interaction, towards the major groove.

*Eco*RI–DNA complex

Recently, the structure of the *Eco*RI–DNA complex has been reinterpreted, although the essential features of the previous structure remain unaltered, but a detailed description of the structure is at present unavailable (Kim *et al.*, 1990). In summary, the *Eco*RI–DNA complex has two-fold symmetry which is related to the symmetry of the recognition sequence (McClarin *et al.*, 1986). The structure of the DNA in the complex is generally of B-form although there are three distinct areas where the DNA is kinked (McClarin *et al.*, 1986). The central kink results in an unwinding of the DNA by 25°, and overall a widening of the major groove and a longer phosphate–phosphate backbone distance between the bases that are hydrolysed is observed (McClarin *et al.*, 1986; Kim *et al.*, 1990). The major protein

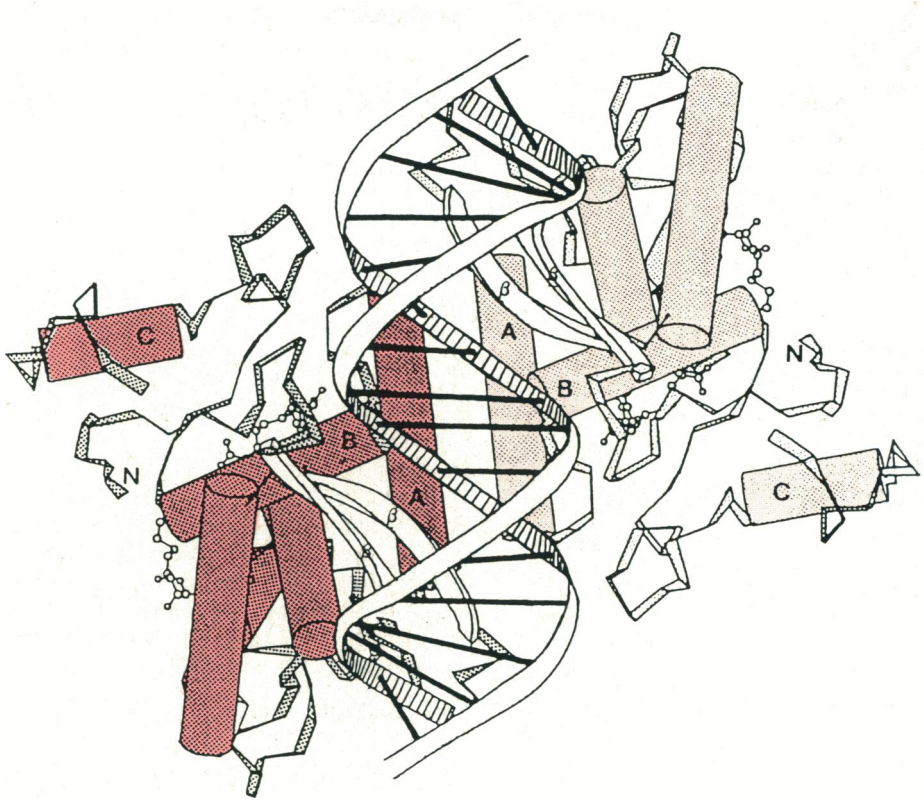

Fig. 15. *metJ* repressor–DNA complex

A schematic representation of the structure of *metJ* repressor bound to a symmetric 19-bp self-complementary oligonucleotide containing the sequence of two *metJ* operator boxes. The two repressor dimers are distinguished by shading and the corepressor *S*-adenosylmethionine is shown in 'ball and stick' representation. The β-strands (unshaded), which are centred on the dyads of each *metJ* box, lie in successive major grooves of the DNA. The co-operative protein–protein interactions occur through helix A (behind the DNA in this view). From Perutz (1990), with permission.

structural elements involved in specific DNA contacts are a parallel β-sheet, a novel β-bridge structure and two α-helices from each monomer, their orientation such that their *N*-terminal ends are pointing in the direction of the DNA–phosphate backbone, allowing favourable helix dipole–phosphate interactions (McClarin *et al.*, 1986; Kim *et al.*, 1990). Recent studies by Heitman & Model (1990) have shown that *Eco*RI recognises pyrimidine bases as well as purines. The new structure interpretation allows for such interactions. However, as the co-crystal complex does not represent a catalytically active complex, little information can be drawn on the mechanism of cleavage.

The crystal structures of *Eco*RV complexed to cognate and non-cognate DNA as well as the free enzyme itself (see above) have recently been elucidated (F. Winkler, personal communication). A comparison of all three structures shows that large conformational changes accompany the binding of *Eco*RV to its specific recognition sequence. Furthermore, the structure of the DNA in the complex is distorted and appears more like A-DNA. Although DNA deformations have been observed in the *Eco*RI–DNA complex, the *Eco*RV–DNA changes are somewhat different. In the complex with cognate DNA, a loop (residues 183–188) makes specific hydrogen bonds with bases in the major groove and three acidic residues, which may participate in Mg^{2+} binding and catalysis, are located in the vicinity of the scissile bond (F. Winker, personal communication). However, it appears that *Eco*RI and *Eco*RV are structurally unrelated both unbound and bound to their respective DNA cognate sequences. Therefore DNA recognition and cleavage by these two endonucleases appears to be different.

DNA polymerase I–DNA complex

Several co-crystal structures of the Klenow fragment complexed with both duplex and single-stranded DNA have been solved and show four nucleotides of single-stranded DNA bound to the 3′–5′ exonuclease active site (Fig. 12*c*; Freemont *et al.*, 1988; Beese & Steitz, 1991). The 3′–5′ exonuclease active site forms a shallow groove which is suitable for binding single-stranded DNA sequences and not duplex DNA, which relates to the proposed editing function of this activity (for a recent review see Joyce, 1991). All of the protein–DNA contacts in the 3′–5′ exonuclease active site are non-specific (summarized in Fig. 16), therefore allowing editing of any single-stranded DNA sequence as would be required for error-free DNA replication (Freemont *et al.*, 1988). Two divalent metal ions are also seen interacting with the phosphate of the 3′ terminal base and are thought to be involved in the positioning and cleavage of the phosphodiester bond (Derbyshire *et al.*, 1988; Freemont *et al.*, 1988). Recent site-directed mutagenesis studies of all the amino acids implicated by the crystal structure to be involved in substrate binding or catalysis are in agreement with the structural observations (Derbyshire *et al.*, 1991). The relationship between the polymerase and 3′–5′ exonuclease active sites, which are some 25 Å apart, remains unclear. However, a melting and sliding mechanism has been proposed for mismatch base cleavage (Freemont *et al.*, 1988). The editing reaction would therefore involve the melting of four base pairs so that the frayed 3′ terminus could be accommodated in the exonuclease active site and that mismatched bases would decrease the stability of the duplex, favouring

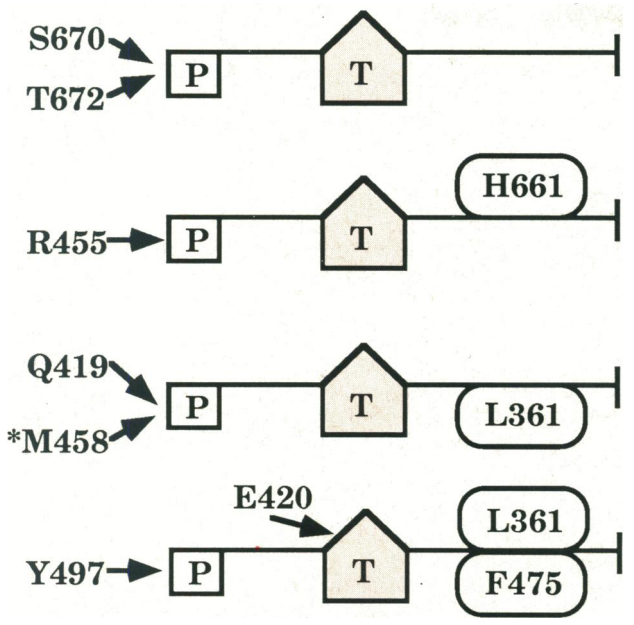

Fig. 16. DNA polymerase I–DNA interaction

A schematic representation of the specific protein–DNA contacts observed in the Klenow single strand (pdT_4) crystal structure (Freemont *et al.*, 1988; Beese & Steitz, 1991). For a description of the symbols used, see legend to Fig. 10.

melting and exonucleolytic cleavage. Recent experiments using covalently linked DNA substrates have confirmed the assertion that four base-pairs of single-stranded DNA are required for efficient exonuclease cleavage (Cowart *et al.*, 1989).

SUMMARY AND FUTURE DIRECTIONS

Numerous DNA-binding motifs have been characterized by structural methods and these motifs can interact with B-DNA in a wide variety of orientations, even those proteins that contain the highly conserved (structurally) helix–turn–helix motif. It appears therefore that the actual structural motif is in itself insufficient to impart DNA-binding properties on a protein, specific or otherwise. Indeed, the DNA-binding domain in some cases may well extend beyond such identified motifs, for example in the use of flexible terminal extensions as in λCI repressor and possibly *trp* repressor, and the general role of electrostatic complementarity as in CAP. Further, in many cases the DNA molecule in particular undergoes significant changes in conformation, such as unwinding of the helix as in *Eco*RI, *lac* repressor and CAP, as well as substantial bending or kinking of the helix axis as in CAP, P434 and *trp* repressors. In contrast, other apparently related repressors have much smaller effects on the DNA conformation, e.g. λCI repressor. Of the helix–turn–helix proteins so far examined, the degree of DNA distortion (bending) is roughly in the order CAP > P434 repressor ≈ λ *cro* repressor > *trp* repressor > *lac* repressor ≈ λCI repressor ≈ *antennapedia*, *engrailed* homeodomain proteins. Interestingly, the proteins that distort the DNA most appear not to use a terminal extension to make additional contacts with the DNA (e.g. CAP, *Eco*RI), whereas those that appear to have minor effects on the DNA conformation have the most extensive flexible extension–DNA interaction surface (e.g. the homeodomain proteins). However, the proteins of the *metJ* repressor class use the antiparallel β-sheet, which forms part of the dimerization interface. As DNA-binding sequences are continually being identified, it would not be surprising to find other folds that are involved in DNA sequence recognition.

It is also clear that biochemical experiments on their own can be misleading, as correct interpretation of the data requires some kind of structural model. For example, the mutation experiments on the *lac* repressor–operator system were originally interpreted, incorrectly, assuming an orientation of the helix–turn–helix motif the same as that presumed for the λ *cro*–DNA complex. However, because substantial conformational changes may occur on the formation of a specific protein–DNA complex (see Table 4), it may be insufficient to build models based on the structure of one of the components alone. Indeed, it is always necessary to compare the properties of the isolated components with those in the specific and non-specific complexes. To date, high-resolution information of this kind is not available for any particular system. Further, the simple idea that a point mutation produces only local structural changes can be misleading, especially as long-range contributions to binding specificity can occur, as emphasized by Senear & Ackers (1990), and implicit in the work of Warwicker *et al.* (1987) on the CAP–DNA complex. Thus, the mutation Glu-19 to Lys in the *trp* repressor increases operator binding (Kelley & Yanofsky, 1982), even though Glu-19 is far from the protein–DNA interface (Otwinowski *et al.*, 1988).

Therefore, in order to achieve a more detailed understanding of protein–DNA recognition, it will be necessary to solve a number of protein–DNA complexes for each protein, including non-specific complexes as well. So far the structures of 16 protein–DNA complexes have been solved, of which 14 are specific complexes. The glucorticoid receptor–DNA complex, however, contains both specific and non-specific interactions.

However, it is apparent that there is as yet a large number of undiscovered protein structural motifs that may specifically bind and recognize DNA (e.g. Freemont *et al.*, 1991). This is highlighted by the increasing numbers of amino acid sequence motifs that have now been characterized and for which no structural information is as yet available (e.g. eukaryotic transcription factors; Mitchell & Tjian, 1989; Latchman, 1990). Therefore, it seems likely that the database of protein structures which bind to DNA will increase rapidly in the coming years. Furthermore, attempts will be made to study more complex protein–DNA interactions including for example, eukaryotic transcriptional complexes. These complexes are made up of more than one protein component with protein–protein interactions (activator domains) equally important in determining transcriptional activation as protein–DNA interaction. Given the relatively small number of protein–DNA complexes that have been studied, it seems that there is probably no universal code for the use of structural motifs in DNA-binding proteins or even in the actual mode of recognition of specific sequences. Therefore, each case should be treated individually, and examined using as wide a variety of techniques as possible.

We thank Drs. K. Borden, D. Brown, M. Carr, R. W. King, B. Madden and S. Neidle for reading the manuscript and their comments. We thank too John Satchell for preparation of several figures. We are grateful to all authors who sent us reprints and preprints of their work, and particularly to Drs. G. M. Clore, A. M. Gronenborn, D. Neuhaus, A. Pastore, S. Phillips, J. Schwabe, T. Steitz, D. Suck and F. Winkler for providing us with original figures. This work was funded by the ICRF, the Cancer Research Campaign and the Medical Research Council of the U.K.

REFERENCES

Abdel-Meguid, S. S., Grindley, N. D. F., Smyth-Templeton, N. & Steitz, T. A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 2001–2005

Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. & Gehring, W. J. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 4093–4097
Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) Science **242**, 899–907
Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) Nature (London) **290**, 754–758
Anderson, W. F., Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. (1982) J. Mol. Biol. **159**, 745–751
Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) Nature (London) **326**, 846–852
Arrowsmith, C. H., Carey, J., Treat-Clemons, L. & Jardetzky, O. (1989) Biochemistry **28**, 3875–3879
Arrowsmith, C. H., Pachter, R., Altman, R. B., Iyer, S. B. & Jardetzky, O. (1990) Biochemistry **29**, 6332–6341
Baleja, J. D., Pon, R. T. & Sykes, B. D. (1990) Biochemistry **29**, 2828–2839
Barkley, M. D. & Bourgeios, S. (1980) in The Operon (Miller, J. H., & Reznikoff, W. S., eds.), pp. 177–220, Cold Spring Harbor Laboratory Press
Bass, S., Sorrells, V. & Youderian, P. (1988) Science **242**, 240–245
Beato, M. (1989) Cell **56**, 335–344
Beese, L. S. & Steitz, T. A. (1991) EMBO J. **10**, 25–33
Berg, J. M. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 91–102
Billeter, M., Qian, Y.-Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990) J. Mol. Biol. **214**, 183–197
Breg, J. N., Boelens, R., George, A. V. E. & Kaptein, R. (1989) Biochemistry **28**, 9826–9833
Breg, J. N., van Opheusden, J. H. J., Burgering, M. J. M., Boelens, R. & Kaptein, R. (1990) Nature (London) **346**, 586–589
Brennan, R. G. & Matthews, B. W. (1989*a*) Trends Biochem. Sci. **14**, 286–290
Brennan, R. G. & Matthews, B. W. (1989*b*) J. Biol. Chem. **264**, 1903–1906
Brennan, R. G., Roderick, S. L., Takeda, Y. & Matthews, B. W. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 8165–8169
Brown, B. M., Bowie, J. U. & Sauer, R. T. (1990) Biochemistry **29**, 11189–11195
Burlingame, R. W., Love, W. E., Wang, B. C., Hamlin, R., Xuong, N. H. & Moudrianakis, E. N. (1985) Science **228**, 546–553
Carey, J. (1989) J. Biol. Chem. **264**, 1941–1947
Carr, M. D., Pastore, A., Gausepohl, H., Frank, R. & Roesch, P. (1990) Eur. J. Biochem. **188**, 455–461
Carter, C. W. & Kraut, J. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 283–287
Chandler, L. R. & Lane, A. N. (1988) Biochem. J. **250**, 925–928
Church, G. M., Sussman, J. L. & Kim, S.-H. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 1458–1462
Clark, J. M., Joyce, C. M. & Beardsley, G. P. (1987) J. Mol. Biol. **198**, 123–127
Cowart, M., Gibson, K. J., Allen, D. J. & Benkovic, S. J. (1989) Biochemistry **28**, 1975–1983
Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N. & Leberman, R. (1990) Nature (London) **347**, 249–255
Derbyshire, V., Freemont, P., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M. & Steitz, T. A. (1988) Science **240**, 199–201
Derbyshire, V., Grindley, N. D. F. & Joyce, C. M. (1991) EMBO J. **10**, 17–24
de Vlieg, J., Scheek, R. M., van Gunsteren, W. F., Berendsen, H. J. C., Kaptein, R. & Thomason, J. (1988) Proteins **3**, 209–218
Dodd, I. B. & Egan, J. B. (1990) Nucleic Acids Res. **18**, 5019–5026
Donoso-Pardo, J. L., Turner, P. C. & King, R. W. (1987) Eur. J. Biochem. **168**, 687–694
Drew, H. R. & Travers, A. A. (1984) Cell **37**, 491–502
Drew, H. R. & Travers, A. A. (1985) J. Mol. Biol. **186**, 773–790
Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobson, L., Wang, B.-C., Greene, P., Boyer, H. W. & Rosenberg, J. M. (1984) Nature (London) **309**, 327–331
Freemont, P. S., Ollis, D. L., Steitz, T. A. & Joyce, C. M. (1986) Proteins **1**, 66–73
Freemont, P. S., Friedman, J. M., Beese, L., Sanderson, M. R. & Steitz, T. A. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 8924–8929
Freemont, P. S., Hanson, I. M. & Trowsdale, J. (1991) Cell **64**, 483–484
Fujimoto, B. S. & Schurr, J. M. (1990) Nature (London) **344**, 175–178
Gartenberg, M. M. & Crothers, D. M. (1988) Nature (London) **333**, 824–829
Geisler, N. & Weber, K. (1977) Biochemistry **16**, 938–943
Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) Science **243**, 1695–1699
Hård, T., Kellenbach, E., Boelens, R., Maler, B. A., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Kustafsson, J.-A. & Kaptein, R. (1990*a*) Science **249**, 157–160
Hård, T., Kellenbach, E., Boelens, R., Kaptein, R., Dahlman, K., Carlstedt-Duke, J., Freedman, L. P., Maler, B. A., Hyde, E. I., Yamamoto, K. R. & Gustafsson, J.-A. (1990*b*) Biochemistry **29**, 9015–9023
Hård, T., Dahlman, D., Carlstedt-Duke, J., Gustafsson, J.-Å. & Rigler, R. (1990*c*) Biochemistry **29**, 5358–5364
Harrison, S. C. & Aggarwal, A. K. (1990) Annu. Rev. Biochem. **59**, 933–969
Hatfull, G. F. & Grindley, N. D. F. (1988) in Genetic Recombination (Kucherlapati, R. & Smith, G., eds.), pp. 357–396, American Society for Microbiology, Washington DC
Heitman, J. & Model, P. (1990) Proteins **7**, 185–197
Hu, J. C., O'Shea, E. K., Kim, P. S. & Sauer, R. T. (1990) Science **250**, 1400–1403
Jones, N. (1990) Cell **61**, 9–11
Jordan, S. R. & Pabo, C. O. (1988) Science **242**, 893–899
Jordan, S. R., Whitcombe, T. V., Berg, J. M. & Pabo, C. O. (1985) Science **230**, 1383–1385
Joyce, C. M. (1991) Curr. Opinions Struct. Biol. **1**, 123–129
Joyce, C. M. & Steitz, T. A. (1987) Trends Biochem. Sci. **12**, 288–292
Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985) J. Mol. Biol. **182**, 179–182
Kaptein, R., Boelens, R. & Lamerichs, R. M. J. N. (1989) in Protein-Nucleic Acid Interaction (Saenger, W., ed.) chapter 3, Macmillan, London
Karslake, C., Schroeder, S., Wang, P. L. & Gorenstein, D. G. (1990) Biochemistry **29**, 6578–6584
Kelley, R. L. & Yanofsky, C. (1982) Proc. Natl. Acad. Sci. U.S.A. **82**, 483–487
Kim, Y., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990) Science **249**, 1307–1309
Kirpichnikov, M. P., Kurochkin, A. V., Chernov, B. K. & Skryabin, K. G. (1984*a*) FEBS Lett. **175**, 317–320
Kirpichnikov, M. P., Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Kurochin, A. V., Skryabin, K. G. & Bayev, A. A. (1984*b*) Nucleic Acids Res. **12**, 3551–3561
Kissinger, C. R., Liu, B., Martin-Blanco, E., Korberg, T. B. & Pabo, C. O. (1990) Cell **63**, 579–590
Klevit, R. E., Herriot, J. R. & Horvath, S. J. (1990) Proteins **7**, 215–226
Klig, L. S., Carey, J. & Yanofsky, C. (1988) J. Mol. Biol. **202**, 769–777
Klug, A. & Rhodes, D. (1987) Trends. Biochem. Sci. **12**, 464–469
Kostrewa, D., Granzin, J., Koch, C., Choe, H.-W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R. & Saenger, W. (1991) Nature (London) **349**, 178–180
Koudelka, P., Harbury, P., Harrison, S. C. & Ptashne, M. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 4633–4637
Kouzarides, T. & Ziff, E. (1989) Nature (London) **340**, 568–571
Kumar, S., Murphy, N. & Krakow, J. (1980) FEBS Lett. **109**, 121–123
Lamerichs, R. M. J. N., Boelens, R., van der Marel, G. A., van Boom, J. H. & Kaptein, R. (1990) Eur. J. Biochem. **194**, 629–637
Landschultz, W. H., Johnson, P. F. & McKnight, S. L. (1989) Science **243**, 1681–1688
Lane, A. N. (1989) Eur. J. Biochem. **182**, 95–104
Latchman, D. S. (1990) Biochem. J. **270**, 281–289
Lawson, C. L., Zhang, R.-G., Schevitz, R. W., Otwinowski, Z., Joachimiak, A. & Sigler, P. B. (1988) Proteins **3**, 18–31
Lawson, C. L. & Sigler, P. B. (1988) Nature (London) **333**, 869–871
Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1990) Science **245**, 635–637
Leighton, P. & Lu, P. (1987) Biochemistry **26**, 7262–7271
Lin, S.-Y. & Riggs, A. D. (1975) Cell **4**, 107–111
Luisi, B. F. & Sigler, P. B. (1990) Biochim. Biophys. Acta **1048**, 113–126
Marmorstein, R. Q. & Sigler, P. B. (1989) J. Biol. Chem. **264**, 9149–9154
Marmorstein, R. Q., Joachimiak, A., Sprinzl, M. & Sigler, P. B. (1987) J. Biol. Chem. **262**, 4922–4927
Marmorstein, R. Q., Sprinzl, M. & Sigler, P. B. (1991) Biochemistry **30**, 1141–1148
Matthews, B. W. (1988) Nature (London) **335**, 294–295
McClarin, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986) Science **234**, 1526–1541
McKay, D. B. & Steitz, T. A. (1981) Nature (London) **290**, 744–749

McKay, D. B., Weber, I. T. & Steitz, T. A. (1982) J. Biol. Chem. **257**, 9518–9524
Mitchell, P. J. & Tjian, R. (1989) Science **245**, 371–378
Mondragon, A., Subbiah, S., Almo, S. C., Drottar, M. & Harrison, S. C. (1989*a*) J. Mol. Biol. **205**, 189–200
Mondragon, A., Wolberger, C. & Harrison, S. C. (1989*b*) J. Mol. Biol. **205**, 179–188
Moore, S. (1981) The Enzymes **14**, 281–296
Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988) EMBO J. **7**, 4299–4304
Nakabeppu, Y. & Nathans, D. (1989) EMBO J. **8**, 3833–3841
Nardelli, J., Gibson, T. J., Vesque, C. & Charnay, P. (1991) Nature (London) **349**, 175–178
Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987) Nature (London) **330**, 221–226
Neuhaus, D., Nakaseko, Y., Nagai, K. & Klug, A. (1990) FEBS Lett. **262**, 179–184
Oas, T. G., McIntosh, L. P., O'Shea, E. K., Dahlquist, F. W. & Kim, P. S. (1990) Biochemistry **29**, 2891–2894
Oefner, C. & Suck, D. (1986) J. Mol. Biol. **192**, 605–632
Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) Nature (London) **298**, 718–723
Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) Nature (London) **313**, 762–766
Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990) Biochemistry **29**, 9324–9334
O'Shea, E. K., Rutowski, R. & Kim, P. S. (1989) Science **243**, 538–543
Otting, G., Qian, Y.-Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990) 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) Nature (London) **335**, 321–329
Pabo, C. O. & Lewis, M. (1982) Nature (London) **298**, 443–447
Pabo, C. O. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293–321
Pabo, C. O., Aggarwal, A. K., Jordan, S. J., Beamer, L. J., Obeysekare, U. R. & Harrison, S. C. (1990) Science **247**, 1210–1213
Pace, H. C., Lu, P. & Lewis, M. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 1870–1873
Patel, L., Abate, C. & Curran, T. (1990) Nature (London) **347**, 572–575
Pavletich, N. P. & Pabo, C. O. (1991) Science **252**, 809–817
Perutz, M. F. (1990) Mechanisms of Cooperative and Allosteric Regulation of Proteins, Cambridge University Press, Cambridge
Phillips, S. E. V. (1991) Curr. Opinions Struct. Biol. **1**, 89–98
Phillips, S. E. V., Manfield, I., Parsons, I., Davison, B. E., Rafferty, J. B., Somers, W. S., Magarita, D., Cohen, G. N., Saint-Girons, I. & Stockley, P. G. (1989) Nature (London) **341**, 711–715
Ptashne, M. (1986) A Genetic Switch, Cell Press, Cambridge MA
Qian, Y.-Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989) Cell **59**, 573–580
Rafferty, J. B., Somers, W. S., Saint-Girons, I. & Phillips, S. E. V. (1989) Nature (London) **341**, 705–710
Rasmussen, R., Benvegnu, D., O'Shea, E. K., Kim, P. S. & Alber, T. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 561–564
Rice, P. A., Goldman, A. & Steitz, T. A. (1990) Proteins **8**, 334–340
Richardson, J. S. & Richardson, D. C. (1988) Proteins **4**, 229–239
Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. (1984) Nature (London) **311**, 532–537
Rosenberg, J. M., Boyer, H. W. & Greene, P. J. (1981) Gene Amplification and Analysis, vol. 1, p. 131, Elsevier, Amsterdam
Sakaguchi, K., Appella, E., Omichinski, J. G., Clore, G. M. & Gronenborn, A. M. (1991) J. Biol. Chem., in the press
Sanderson, M. R., Freemont, P. S., Rice, P. A., Goldman, A., Hatfull, G. F., Grindley, N. D. F. & Steitz, T. A. (1990) Cell **63**, 1323–1329
Saudek, V., Pastore, A., Catiglione Morelli, M. A., Frank, R., Gausepohl, H., Gibson, T., Weih, F. & Roesch, P. (1990) Protein Eng. **4**, 3–10
Saudek, V., Pasley, H. T., Gibson, T., Gausepohl, H., Frank, R. & Pastore, A. (1991) Biochemistry **30**, 1310–1317
Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. (1982) Nature (London) **298**, 447–451
Schauer, M., Chalepakis, G., Willmann, T. & Beato, M. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 1123–1127
Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom, J. H., Kaptein, Rüterjans, H. & Beyreuther, K. (1983) Biochemistry **22**, 228–235
Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. (1985) Nature (London) **317**, 782–786
Schultz, S. C., Shields, G. C. & Steitz, T. A. (1990) J. Mol. Biol. **213**, 159–166
Schwabe, J. W. R., Neuhaus, D. & Rhodes, D. (1990) Nature (London) **348**, 458–461
Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 804–808
Senear, D. F. & Ackers, G. K. (1990) Biochemistry **29**, 6568–6577
Shirakawa, M., Lee, S. J., Akutsu, H., Kyogoku, Y., Kitano, K., Shin, M., Ohtsuka, E. & Ikehara, M. (1985) FEBS Lett. **181**, 286–290
Sixl, F., King, R. W., Bracken, M. & Feeney, J. (1990) Biochem. J. **266**, 545–552
Staacke, D., Walter, B., Kisters-Woike, B., Wilcken-Bergmann, B. & Müller-Hill, B. (1990) EMBO J. **9**, 1963–1967
Steitz, T. A. (1990) Q. Rev. Biophys. **23**, 205–280
Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3097–3100
Stob, S., Scheek, R. M., Boelens, R. & Kaptein, R. (1988) FEBS Lett. **239**, 99–104
Struhl, K. (1989) Trends Biochem. Sci. **14**, 137–140
Suck, D. & Oefner, C. (1986) Nature (London) **321**, 620–625
Suck, D., Oefner, C. & Kabsch, W. (1984) EMBO J. **3**, 2423–2430
Suck, D., Lahm, A. & Oefner, C. (1988) Nature (London) **332**, 464–468
Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) Science **249**, 769–771
Tanaka, I., Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. (1984) Nature (London) **310**, 376–381
Taylor, J. D. & Halford, S. E. (1989) Biochemistry **28**, 6198–6206
Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) Science **246**, 911–916
von Hippel, P. H. & Berg, O. G. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 1608–1612
von Hippel, P. H. & Berg, O. G. (1989) in Protein–Nucleic Acid Interaction (Saenger, W., ed.), chapter 1, Macmillan, London
von Hippel, P. H., Revzin, A., Gross, C. A. & Wang, A. C. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 4808–4812
Wade-Jardetzky, N., Bray, R. P., Conover, W. W., Jardetzky, O., Geisler, N. & Weber, K. (1979) J. Mol. Biol. **128**, 259–264
Warwicker, J., Engelman, B. P. & Steitz, T. A. (1987) Proteins **2**, 283–289
Weber, I. T. & Steitz, T. A. (1987) J. Mol. Biol. **198**, 311–326
Weber, I. T., Gilliland, G. L., Harman, J. G. & Peterkofsky, A. (1987) J. Biol. Chem. **262**, 5630–5636
Weiss, M. A. (1990) Biochemistry **29**, 8020–8024
Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990) Nature (London) **347**, 575–578
Wharton, R. P. & Ptashne, M. (1987) Nature (London) **326**, 888–891
White, S. W. (1987) Protein Eng. **1**, 373–376
White, S. W., Appelt, K., Wilson, K. S. & Tanaka, I. (1989) Proteins **5**, 281–288
Winkler, F. K., D'Arcy, A., Blocker, H., Frank, R. & van Boom, J. H. (1991) J. Mol. Biol. **217**, 235–238
Wolberger, C., Dong, Y., Ptashne, M. & Harrison, S. C. (1988) Nature (London) **335**, 789–795
Wrange, Ö., Eriksson, P. & Perlmann, T. (1989) J. Biol. Chem. **264**, 5253–5259
Wu, H. & Crothers, D. M. (1984) Nature (London) **308**, 509–513
Zagorski, M. G., Bowie, J. U., Vershon, A. K., Sauer, R. T. & Patel, D. J. (1989) Biochemistry **28**, 9813–9825
Zhang, R.-G., Joachimiak, A., Lawson, R. W., Otwinowski, Z. & Sigler, P. B. (1987) Nature (London) **327**, 591–597
Zinkel, S. S. & Crothers, D. M. (1990) Biopolymers **29**, 29–38
Zuiderweg, E. R. P., Kaptein, R. & Wüthrich, K. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 5837–5841