

Stereochemical basis of DNA bending by transcription factors

Masashi Suzuki* and Naoto Yagi¹

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK and ¹Tohoku University, School of Medicine, Seiryō-machi, Sendai, 980-77, Japan

Received April 24, 1995; Accepted May 11, 1995

ABSTRACT

Superstructure-formation of DNA plays an important role in transcription regulation as well as in chromatin formation. To understand the stereochemical basis of DNA bending by proteins we analysed the structural characteristics of dinucleotide steps which occur at the site where DNA is bent upon binding a transcription factor. When DNA is considerably bent in a crystal structure the bending is not spread smoothly over a length, but the DNA is kinked at a pair of crucial steps which are highly rolled and untwisted. These rolled steps are spaced 6–10 bp apart and are predominantly occupied by pyrimidine–purine sequences. In association with another dinucleotide step at the centre, which combines 6 bp-spaced rolled steps towards the same side of the DNA, these produce two essentially different types of DNA bending.

INTRODUCTION

Superstructure-formation of DNA plays an important role in transcription regulation as well as in chromatin formation (1–6). In the nucleosome a series of bends positioned every 10 bp produces approximately two superhelical turns of DNA (7,8). In many crystal structures of DNA-transcription factor complexes, the DNA is not totally straight and in some it is considerably bent—for example, in the complexes with CAP (9), E2 (10) and the TATA-box-binding protein (TBP) (11,12).

DNA structures have been studied in detail by crystallography, NMR spectroscopy and other methods (see reviews 13–16) and there has been much discussion on their sequence dependent nature, particularly in relation to DNA bending (17–31). However, we are still far away from the final goal of describing clear sequence-structure correlation to predict particular DNA superstructures.

In this paper we aim to understand some structural features common to the bent DNA found in crystals of complexes with transcription factors. Some of the features of the DNA which are discussed in this paper are similar to those described earlier by other groups for DNA structures crystallised in the absence of a protein (see for example 26, 30, 31). However, the bending found in the complexes is notably larger than that found in the protein-unbound DNA crystal structures and one would expect

that the stereochemical basis of the DNA-bending would be correspondingly clearer. This study is based on statistics on crystal structures of 22 DNA-transcription factor complexes and 33 protein-unbound DNAs, which include as many as 909 dinucleotide steps.

MATERIALS AND METHODS

Crystal structures

The crystal co-ordinates (listed in Fig. 1) were taken from the Brookhaven Protein Data Bank (PDB, 32) and the Nucleic Acid Data Bank (NADB, 33). The co-ordinates of the GLI structure were given by Prof. Pabo. The co-ordinates taken from the NADB are those which do not contain a mismatched basepair, an unusual base, a non-Watson–Crick basepair or a nick. DNA oligomers of the same sequence which were crystallised in the same form (such as the Dickerson–Drew dodecamers determined by slightly different methods or determined in different environments) are represented by one of the structures.

Calculation of the dinucleotide step parameters

The six parameters were calculated by using a computer program (34,35). Altogether 909 dinucleotide steps were collected: 33 TG, 46 CG, 84 TA, 48 AA, 42 GA, 49 AG, 23 GG, 67 GT, 18 GC and 54 AT from PDB; 39 TG, 104 CG, 38 TA, 42 AA, 25 GA, 14 AG, 53 GG, 41 GT, 58 GC and 40 AT from NADB.

A dinucleotide step of a symmetric sequence such as GC provides two examples (note that the partner of GC is again GC, and the two conformations are usually different, while for example, that of GT is not GT but AC and therefore provides only one example). For such sequences the averaged tilt angle, by definition, becomes zero but the averaged roll angle does not become zero (36,37).

Many DNAs studied here have palindromic sequences. The two halves are related by crystallographic 2-fold symmetry in some examples, such as the two halves of the E2 binding site. Such identical steps related by symmetry were treated as two independent examples. This is because, some other structures such as the Gal4 binding site were determined without assuming such a crystallographic 2-fold symmetry, and thus two identical steps in the binding site must be treated as independent examples, although the two are, in fact, almost the same. Since whether the structures become crystallographically symmetric or pseudo-symmetric is, in

* To whom correspondence should be addressed

a DNA-transcription factor complexes

Code	name	angle	bending
---	TBP+TATA box	-90*	+++
2BOP	E2+DNA	62-51	+++
1CGP	CAP+DNA	58-67	+++
1TRO	Trp repressor+operator	41-24	++
1TRR	Trp repressor+operator	24-22	+
1RPE	434 repressor+OR2 operator	40-40	++
2OR1	434 repressor+OR1 operator	28-28	+
1PER	434 repressor+OR3 operator	27-29	+
3CRO	434 cro+OR1 operator	30-29	+
1LMB	Lambda repressor+operator	27-23	+
1HCR	Hin recombinase+operator	26-20	+
1DGC	GCN4+CREB site	33-21	+
1YSA	GCN4+AP1 site	28-20	+
1GLU	Glucocorticoid receptor+DNA	10-7	±
1HDD	Engrailed homeodomain+DNA	10-4	±
1ZAA	Zif268+DNA	10-8	±
---	GLI+DNA	1-7	±
1D66	GAL4+DNA	12-31	±
2DRP	Tramtrack+DNA	18-17	±
1PAR	Arc repressor+operator	0-7	±
1CMA	METJ repressor+operator	48-31	++

b DNA crystals

A-DNA				B-DNA			
code	bp	form	sequence	code	bp	form	sequence
ADH006	8	P 61	GGGGCOCC	BDJ017	10	C 2	CCAGGCCTGG
ADH007	8	P 61	GGGATCCC	BDJ019	10	C 2	CCAACGTTGG
ADH030	8	P 61	GGGTACCC	BDJ051	10	P 21 21 2	CATGGCCATG
ADH008	8	P 43 21 2	GCCCGGGC	BDJ025	10	P 21 21 21	CGATCGATCG
ADH012	8	P 43 21 2	CCCCGGGG	BDJ036	10	P 21 21 21	CGATATATCG
ADH014	8	P 43 21 2	GTGTACAC	BDJ031	10	P 21 21 21	CGATTAATCG
ADH020	8	P 43 21 2	CTCTAGAG				
ADH023	8	P 43 21 2	GTACGTAC	BDJ039	10	R 3	CCGGCGCCGG
ADH041	8	P 43 21 2	GTCTAGAC				
ADH047	8	P 43 21 2	GTGCGCAC	BDJ052	10	P 6	CCAAGCTTGG
ADH038	8	P 61 2 2	GTGTACAC (spermine)	BDL006	12	P 21 21 21	CGCAAAAAGCG
				BDL007	12	P 21 21 21	CGCATATATCGG
ADH039	8	P 61 2 2	GTGTACAC (spermidine)	BDL015	12	P 21 21 21	CGCAAAAATGCG
				BDL020	12	P 21 21 21	CGCGAATTCGCG
				BDL028	12	P 21 21 21	CGTGAATTCACG
ADI009	9	P 43	GGATGGGAG	BDL038	12	P 21 21 21	CGCAAATTTGCG
ADJ022	10	P 61 2 2	ACCGGCCGGT	BDL047	12	P 21 21 2	CGCGAAAAAACG
ADJ049	10	P 21 21	21CCCGCCGGG	BDL042	12	C 2	CGTAGATCTACG
ADL045	12	P 61 2 2	CCGTACGTACGG				
ADL046	12	P 61 2 2	GCGTACGTACGC				

Figure 1. Crystal structures which are studied in this paper. The code names are those of the PDB (a) and the NADB (b). (a) The bending angles were calculated by defining the helix axes at both ends, each with 4 or 5 bp and are shown as (that calculated with 4 bp) – (that calculated with 5 bp). The bending angle of the TBP binding site (*) was taken from the original report (11), as the co-ordinates have not been published yet. According to these numbers the DNAs can be roughly classified into four groups (±, shown in normal type; +, shown in italic; ++, shown in italic and underlined; +++, shown in bold and underlined). References to the structures are: TBP (11,12), 1CGP (9), 2BOP (10), 1TRO (58), 1TRR (59), 1RPE (60), 2OR1 (61), 1PER (62), 3CRO (63), 1LMB (64), 1HCR (65), 1DGC (66), 1YSA (67), 1GLU (68), 1HDD (89), 1ZAA (70), GLI (71), 1D66 (72), 2DRP (73), 1PAR (74) and 1CMR (75). (b) The crystal forms and the DNA sequences are shown. Some pyrimidine-purine steps mentioned in the text are shown in bold. The classification of A and B DNA is according to the NADB; the code names starting with A and B are those of A and B DNA, respectively. References to the structures are: ADH006 (76), ADH007 (77), ADH030 (78), ADH008 (79), ADH012 (80), ADH014 (81), ADH020 (82), ADH023 (83), ADH041 (84), ADH047 (85), ADH038 (86), ADH039 (87), ADI009 (88), ADJ022 (89), ADJ049 (90), ADL045 (91), ADL046 (92), BDJ017 (93), BDJ019 (94), BDJ051 (95), BDJ025 (96), BDJ036 (97), BDJ039 (98), BDJ052 (99), BDL007 (100), BDL006 (54), BDL015 (101), BDL020 (102), BDL028 (103), BDL038 (104) and BDL047 (105).

a sense, a matter of assumption, we felt it better to include all the symmetric or pseudo-symmetric DNA structures into the statistics rather than only half the structures. Some structures such as

in the tramtrack crystal contain two molecules of the complex per asymmetric unit and the two structures were treated as independent examples.

Some unusual steps found in the complexes were excluded from the statistics i.e. a non-Watson-Crick G-C basepair in 1D66, two nicked steps in 1CGP, three unpaired basepairs in 1YSA.

Calculation of the bending angle

The bending angle of each DNA structure was calculated as that between the two helix axes at the ends. The helix axes were defined by using a computer program, NEWHELIX (38) for either four consecutive basepairs at each end or five basepairs; both values are listed (Fig. 1a). The reason why we show the two values for each DNA structure is as follows.

If a DNA is zig-zagging, the bending angle of the DNA is dependent on the phases of the two ends. Also the bending angle might be affected by a small number of basepairs at the ends. If, for example, basepairs at the ends, which do not directly interact with the co-crystallised protein, are bent by the packing force, it might create an apparently large bending angle. We found that the above awkward influences can be avoided to some extent by choosing a reasonably long piece of DNA from each end and by comparing the numbers calculated with different lengths of the pieces. Since some DNAs are not much larger than 10 bp, it is difficult to use more than 5 bp for defining the axis.

RESULTS AND DISCUSSION

Parameters important for describing DNA-bending

Any DNA structure can be described as an accumulation of dinucleotide steps (36,37). The geometry of each dinucleotide step is characterised by three rotational angles (helical twist, roll, tilt) and three translational distances (rise, slide, shift) (Fig. 2). These parameters can be defined locally, independent of the overall DNA structure (34,35) and we use this type of definition in this paper. It is sometimes useful to define the parameters in reference to the overall DNA helix axis (38) but it is difficult to do so accurately, when the DNA is bent (see for example, 39). The six parameters were calculated by using a computer program (34,35) for the dinucleotide steps in the DNA-factor complexes (listed in Fig. 1a) and those in DNA structures crystallised in the absence of a protein (listed in Fig. 1b). The numbers calculated for the parameters here might be slightly different from those published earlier, depending on which calculation algorithm is used, in particular, which definition of the parameters, either the local definition or the axis-referred definition, is adopted.

Some parameters are directly related to bending (see the double helix axes drawn in Fig. 2). In particular, positive rolling changes the direction of the DNA double helix axis around the major groove (Fig. 2—note that positive rolling decreases the distance between the 2 bp on the major groove side and increases that on the minor groove side and thus bends the DNA helix axis around the major groove). But two parameters, helical twist and rise, are not directly related with bending. Sliding and shifting do not produce bending, but create a gap in the helix axis, unless these take place over several steps in a coherent manner (2). Thus two important parameters need to be analysed—roll and tilt.

Bent DNAs in crystals

Before describing our analysis in detail it might be relevant to describe which DNA is appreciably bent as a whole and which is not. It is not easy to do so accurately, no algorithm has been

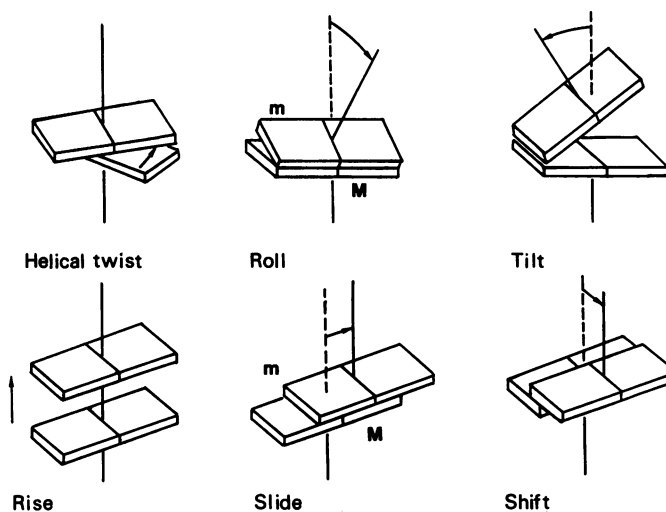


Figure 2. Geometrical parameters for describing a dinucleotide step. The lower basepair is fixed and the upper basepair is moved to the positive direction of the parameter. The change in the double helix axis is also shown. All the steps are viewed from the major (M) groove; 'm' shows the minor groove side.

provided so far to calculate the degree of bending. We have calculated the angle between the double helix axes at the two ends of the DNA by using the crystal co-ordinates of DNA-transcription factor complexes (Fig. 1, also see the figure legend for the references to the structures). According to our calculation in the three complexes with TBP, CAP and E2, the DNA is considerably bent (marked '+++ in Fig. 1a). In those with Trp repressor and MetJ repressor it is moderately bent (marked '++'), and in those with glucocorticoid receptor, Engrailed homeodomain, Zif268, GLI, Gal4, Tramtrack and Arc repressor, it is fairly straight (marked '±'), while the others are only slightly bent (marked '+').

The definition of the bending angle, adopted here, might not be the best (see Materials and Methods) and thus the individual numbers are not very important. However, these seem to be reasonable as they agree well with the overall views of the DNA structures. Whatever definition is used for the bending angle, it seems safe to state that in the complexes with TBP, CAP and E2, the DNA is considerably bent. In what follows we first concentrate on the three considerably bent DNA structures and then examine the mildly bent structures in comparison with them.

Similarities among considerably bent DNAs

In the binding sites of CAP, E2 and TBP, pairs of identical steps, which are related by a 2-fold symmetry (or a pseudo symmetry), adopt distinctively high roll angles (Fig. 3a, c and e)—the two TG steps in the CAP site, the two CG steps in the E2 site and the TA and AN steps in the TBP site—appreciably higher than the usual roll angle at dinucleotide steps of $0 \pm 10^\circ$ (Figs 3 and 4—note that a dinucleotide step can adopt a large positive roll but cannot roll in the opposite direction to a large extent, as shown in Fig. 4a). In what follows we mainly focus attention on these unusual steps and try to understand the mechanism of DNA bending in terms of them.

The spacing found between the unusual pairs is 10 (CAP) to 6 bp (E2, TBP). Most of the unusual steps are occupied by

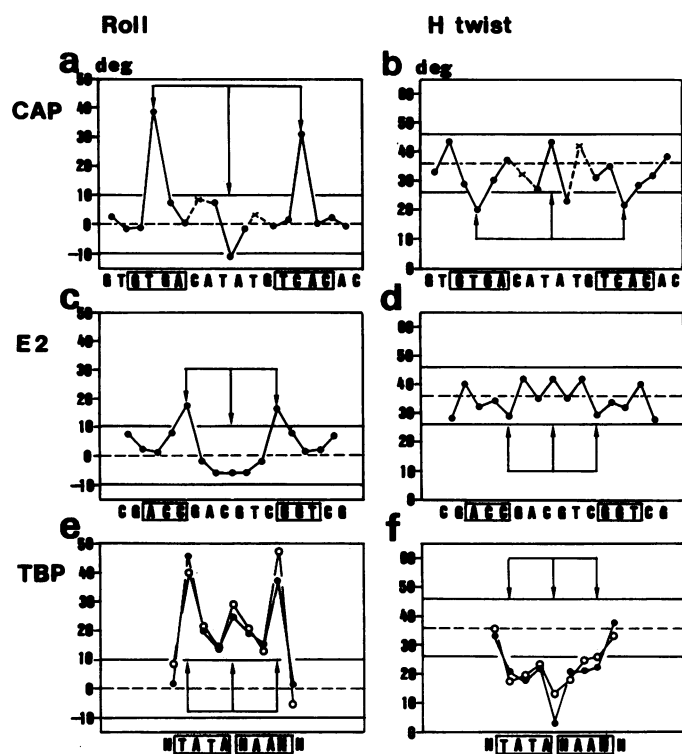


Figure 3. Roll and helical twist angles of dinucleotide steps in the considerably bent DNAs of the CAP (a, b), E2 (c, d) and TBP (e, f) binding sites. The angles were calculated using the co-ordinates deposited in the PDB (code names: 1CGP, 2BOP, 1DGC) except for those of the TBP site, which were taken from the original reports [those reported in (12) are shown with open circles and those reported in (11) are shown with closed circles]. Three steps in each structure, on which attention is focussed in the text, are indicated with arrows. Three or four basepair sites bound by the recognition helices (CAP, E2) or β -sheets (TBP) are boxed. In (a), (c) and (e) lines are drawn to show $0 \pm 10^\circ$. In (b), (d) and (f) lines are drawn to show $36 \pm 10^\circ$.

pyrimidine(Y)–purine(R) sequences—TG/CA, CG and TA. The tilt angle of the steps remains small (at the TG steps in the CAP binding site the roll angle is as high as $30\text{--}40^\circ$, while the tilt angle is as small as -1° , also see Figure 4c). At these steps, a high roll angle coincides with a small helical twist angle (Fig. 3b, d and f), although the features in the helical twist plots are less clear. In fact, the helical twist plots, if the \pm direction is reversed, resemble those of the roll angle [this has been noted by Yanagi *et al.* (26) for the DNAs unbound by a protein].

The roll angle of the steps at the centres of the binding sites also deviates from zero—either to negative (CAP and E2) or positive (TBP).

Comparison of mildly bent DNAs

The DNA in the complexes with MetJ repressor (Fig. 5a), Trp repressor (Fig. 5c and d), 434 repressor and 434 cro (Fig. 5b) and GCN4 (Fig. 5e) are bent but only mildly (Fig. 1a). The roll angle found in these structures is more normal but the roll angle plots (Fig. 5) do hint that some pyrimidine–purine steps adopt slightly high roll angles on both ends, spaced 6–10 bp apart [6-GCN4, TrpR(1TRP), 8-MetJ, TrpR(1TRO), 10-434R, 434C] and also at the centre (marked with arrows in Fig. 5). These and the three considerably bent structures can be classified into two groups

depending on whether the central YR step has a negative (CAP, E2, MetJ, 434R, 434C) or positive (TBP, TrpR, GCN4) roll angle (compare Fig. 3a and c with Fig. 5a and b, and Fig. 3e with Fig. 5c–e, for understanding the overall similarities in the two groups). [Note that Shakked *et al.* (40) have compared the Trp operator DNA bound and unbound by the protein and noticed the importance of the AT step at the centre for the bending.]

Even some DNA oligomers crystallised in the absence of a protein have pairs of rolled-untwisted steps spaced, again, 6 bp apart i.e. the dodecamers crystallised into the same packing form of the Dickerson–Drew dodecamer (41) (Fig. 6a and b). These dodecamers are slightly bent (by $15\text{--}13^\circ$) and are banana-shaped. To adopt the particular superstructure imposed by the packing, which is in a sense comparable with the superstructure imposed by a protein, again, YR sequences separated by 6 bp are used (except for one YY/RR in BDL006). [Also note that in relation to unusual electrophoretic behavior of A tracts, Goodsell *et al.* (30, 31), in the light of some oligonucleotide crystal structures, have discussed that a TA step in protein-unbound DNA can produce high rolling.]

In brief we have found that in the DNA-transcription factor complexes: (i) DNA is not smoothly bent over a length but in each bent DNA structure two steps are distinctively different from the rest and adopt high roll and small helical twist angles, (ii) the rolled steps are predominantly occupied by YR sequences, (iii) the high rolled steps are positioned 6–10 bp apart and (iv) these DNAs can be classified depending on whether the central steps adopt positive or negative rolling.

Rolling-untwisting correlation at a YR step

Why are the YR steps used for rolling-untwisting? It has been pointed out that the two basepairs in a YR step are poorly stacked onto each other (42–44). The pyrimidine bases are more bulged towards the major groove, while the purine bases are slimmer and thus the major groove edge of a basepair is tilted by $\sim 15^\circ$ from the line connecting the sugars. As a consequence, to roll the two Y bases in a YR step towards the major groove (Fig. 7a) is much easier than to roll the two Y bases in an RY step (Fig. 7b).

Coupling of rolling and helical twisting has been analysed in some detail (26,45). In brief, the repeat distance in the sugar–phosphate backbone is longer than the base stacking distance, so when a dinucleotide step is untwisted, it brings a longer component of the repeat distance parallel to the helix axis and thus creates a larger distance between the two basepairs. To maintain hydrophobic interaction, the basepairs approach each other on the major groove side by positive rolling. Thus untwisting can trigger rolling or *vice versa*. [Note that a similar explanation was given to slide-helical twist correlation by Calladine and Drew (2) and that some other correlations have been discussed by Bhattacharya and Bansal (46) and Sponer and Kypr (47).]

A plot of roll and helical twist angles of YR steps in DNA crystallised in the presence or absence of a protein (Fig. 4a) indeed shows the correlation. On the correlation curve (Fig. 4a) the parameters of YR steps in B DNA (here we use the A–B definition given by NADB) are confined to the left side and those in A DNA clustered in the centre, while those in DNA–protein complexes are scattered around (Fig. 4b). The ‘unusual’ steps (numbered in Fig. 4a) are found on the far right side. In this sense the ‘unusual’ steps may be regarded as ‘super A’ steps.

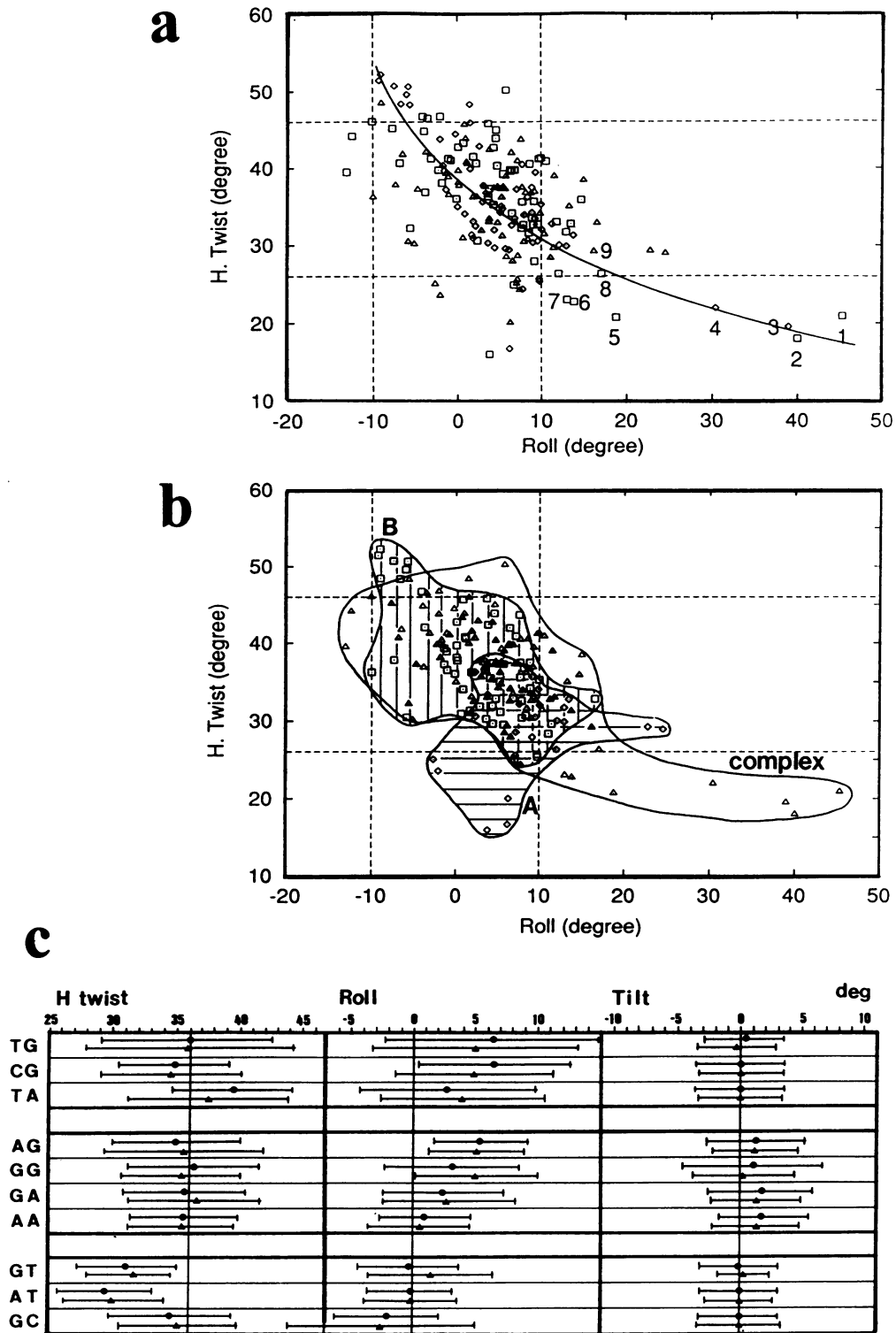


Figure 4. Roll-twist correlation at pyrimidine-purine steps. (a) and (b) The two angles of pyrimidine-purine steps found in DNA bound or unbound by a transcription factor are plotted. In (a) the entries are classified by the sequences: TG/CA (\diamond), TA (\square) and CG (\triangle) steps. In (b) they are classified to those found in DNA-transcription factor complexes (Δ), and in A (\diamond) and B (\square) DNA which were crystallised in the absence of a protein. 1, 2, 5, 6, 7: TA steps in the two TBP structures. 3, 4: the TG steps in the CAP binding site. 8: a TA step in the Arc repressor binding site. 9: the CG steps in the E2 binding site. (c) The averaged helical twist, roll and tilt angles and the standard deviations are shown. Circles are used for the steps found in DNA-factor complexes (listed in Fig. 1a) only, while triangles are used for those in all the DNA structures including A and B DNAs (listed in Figs 1a and b).

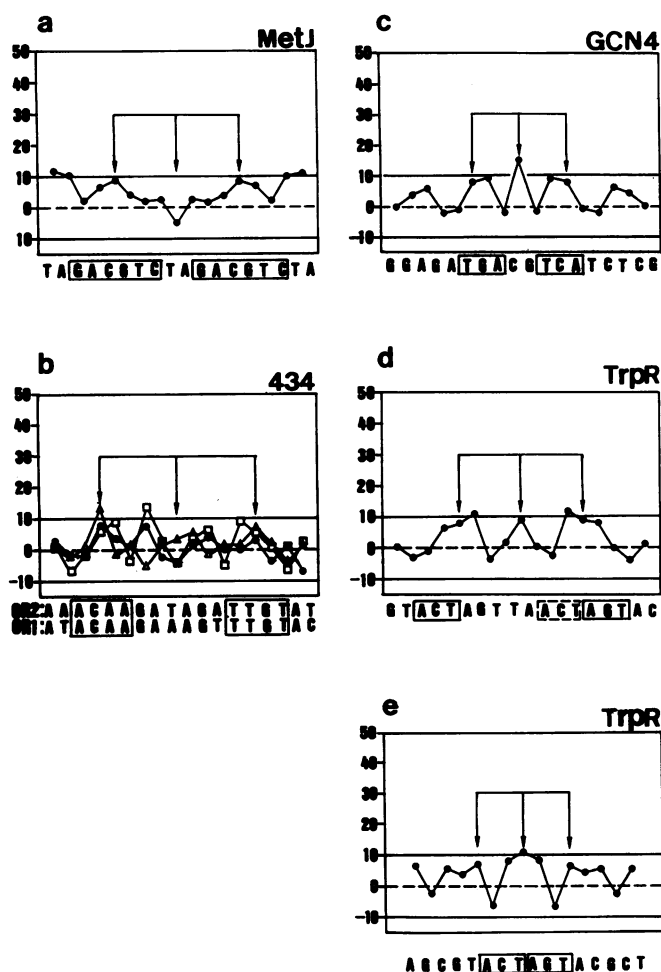


Figure 5. Helical twist angles of dinucleotide steps in the mildly bent DNA structures of the MetJ (a), 434 repressor (b), 434 cro (b), GCN4 (c) and Trp repressor (d and e) binding sites. The angles were calculated using the co-ordinates deposited in the Protein Data Bank (code names-1CMA, 1RPE, 2OR1, 3CRO, 1TRO, 1TRR). In (b) those of OR2 bound by 434 repressor (●), OR1 bound by 434 repressor (□) and OR2 bound by 434 cro (Δ) are shown. Three YR steps in each structure, on which attention is focussed (see text), are indicated with arrows. Three or four basepair sites recognised by the recognition helices or 6 bp sites recognised by β-sheets are boxed. Lines are drawn to show $0 \pm 10^\circ$. More than two molecules of Trp repressor can bind to the operator DNA and the two complexes (d and e) were crystallised with two molecules binding at different positions on the DNA. The equivalent position with the ACT sequence bound by Trp repressor in (e) are shown in (d) with a broken line.

Nekludova and Pabo (39) and Shakked *et al.* (40) have analysed overall features of some protein-bound DNA structures and noticed that these are in some sense intermediate between the standard A and B DNAs. We slightly modify this view to one in which a YR step is flexible and can adopt A, B or super A conformations upon binding a protein.

We note that high rolling can occur only around the major groove (to the positive direction) but not around the minor groove (to the negative direction) and that this direction coincides with that predicted by Sobell *et al.* (48) (also they predicted the rolling angle correctly as -40°) but is opposite to that predicted by Crick and Klug (17) [see also Fig. 1 of Zhurkin (49)].

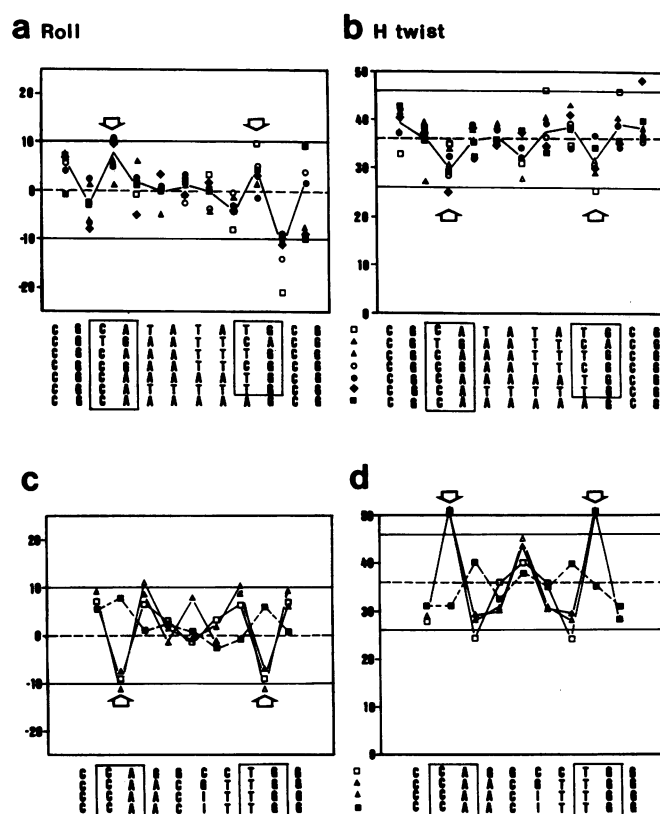


Figure 6. Roll and helical twist angles of dinucleotide steps in the dodecamers crystallised in the P 21 21 21 symmetry (a and b) and the decamers (c and d) crystallised in the C2 (shown with solid lines) or P 32 2 1 (shown with dotted lines) symmetries. The pyrimidine-purine steps (their sequences are shown in boxes) spaced 6 bp apart are indicated with arrows. In (a) and (c) lines are drawn to show $0 \pm 10^\circ$. In (b) and (d) lines are drawn to show $36 \pm 10^\circ$. The co-ordinates were taken from the NADB, CGCATATATGCG (BDL007), CGCAAAAAGCG (BDL006), CGCAAAAATGCG (BDL015), CGCATT-TTTGCG (BDL015), CGCGAATTCGCG (BDL020), CGTGAATTCACG (BDL028), CGCAAATTTGCG (BDL038), CCAGGCCTGC (BDJ017), CCAACGTTGG (BDJ019), CCAACITTTGG [BDJB43 (50)], and CCAAC-ITTTGG [BDJB44 (50)].

The correlation curve of YR steps does not necessarily have to be exactly the same as those of YY/RR or RY, as the chemical structures and the stacking profiles are different. The averaged roll angles and the averaged helical twist angles in the DNA crystal structures are correlated among the steps of the same type (Figure 4c, see for example, the two averaged angles of the RY steps shift in correlation from GT to GC) but the details of the correlation are different in different types (the averaged TG step has a helical twist angle similar to that of the averaged GC step but its roll angle is distinctively higher).

Flexible YR steps

The statistics of roll angles (Fig. 4c) have some more interesting features. The averaged roll angles of YR steps are higher than those of RY steps, while those of YY/RR are intermediate. The deviation in the angles for each of the individual YR steps tends to be larger than that of the others. In other words the YR steps can adopt greater positive rolling than the other steps and are more flexible.

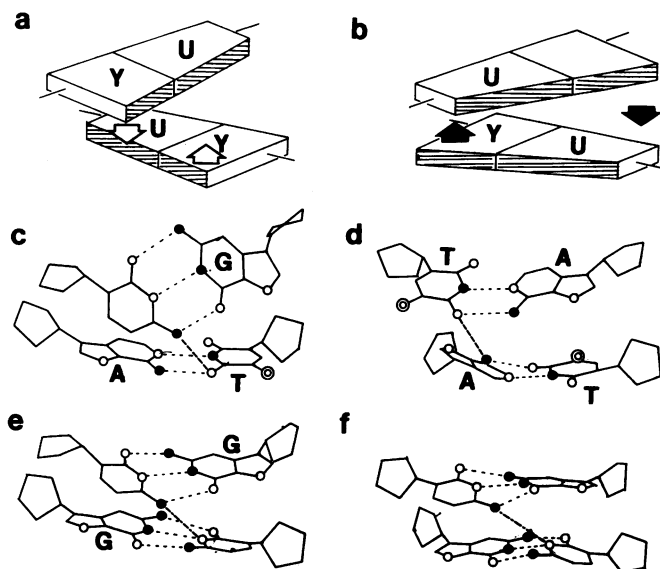


Figure 7. Rolling-untwisting of pyrimidine-purine steps. (a) and (b) Geometry of a YR step, (a) is compared with that of an RY step, (b) (both steps are fixed with the same helical twist angle). Note that high rolling at an RY step would cause awkward steric clashes between the pyrimidine bases and the bases in the neighboring basepairs (black arrows), while it will not cause such a clash at a YR step, permitting higher rolling (white arrows). This is because pyrimidine bases are bulkier on the major groove side than purines, as shown in the drawing. The basepairs are shown from the major groove side. (c)–(f) TG in the CAP binding site (c), TA in the TBP binding site (d), CG in the E2 binding site (e) and CG in the Dickerson–Drew dodecamer (f). Hydrogen bond donors and acceptors are shown with closed and open circles, respectively. Methyl groups are shown with double circles. In (c) and (d) inter-basepair hydrogen bonds are marked with broken lines (---). In (e) and (f) two features in the two C bases which are closer than those in the standard B DNA are marked (---, see text). All the views are the same as in (a).

Some of the central steps are negatively rolled and these are again predominantly occupied by YR sequences (Figs 3 and 5). The reason why the YR steps are used for negative rolling is less clear as the features of the four DNA bases on the minor groove side are not very different from each other, either physically or chemically (see also the ‘central step’ section) but it may be related to the flexibility of YR which apparently originates from the poor stacking.

It might be interesting to note that a protein-unbound DNA decamer, which has been crystallised in two forms, changes its structure by switching the conformations of two YR steps, which are spaced 6 bp apart, from large negative rolling to positive rolling [(50), see also Fig. 6c and d]. This is consistent with the idea of flexible YR steps.

Almost all drugs intercalate into DNA from the minor groove side. Such intercalation is likely to take place to a step which is widely opened on the minor groove side [thus to a positively rolled step, see also Sobell *et al.* (48)]. Indeed in all the intercalator–DNA complex structures collected in NADB (except for actinomycin D) intercalation takes place at YR steps (TG/CA and CG) suggesting again that the YR steps are flexible.

Two of the three steps highly rolled in each TBP binding site are occupied by non-YR steps (one of the two steps at the ends is occupied by AA or AG and the central step by AN, see Figure 3e). Although the conserved TATA half and the non-conserved A-rich half adopt very similar structures in the complex with a pseudo

dyad axis in the centre, somehow the differences between the two halves must be recognised by TBP as this determines the direction of transcription [see a short review by Klug (51)]. One possibility is that the DNA has an intermediate conformation in which only the conserved TATA half adopts an unusual structure and that this non-symmetric nature of the DNA determines the N–C direction of the TBP on the DNA.

Inter-basepair hydrogen bonds which can stabilise rolled steps

Heavy rolling-untwisting of the YR steps in the markedly bent DNAs are stabilised by interaction with the proteins (in particular hydrophobic residues of TBP open up the AT step from the minor groove side) as well as by additional hydrogen bonds between the two neighboring basepairs (inter-basepair hydrogen bonds). At the TG steps in the CAP binding site an inter-basepair H-bond is made between $O^4(T)$ and $N^4H_2(C)$ (Fig. 7c, see also 52,45). At the TA step in the TBP structures an inter-basepair H-bond has been reported (12) between $O^4(T)$ and $N^6H(A)$ (Fig. 7d). In our statistics we do not have an example of CG as highly rolled as the TG or the TA. In two examples, however, namely those in the E2 binding site and in the Dickerson–Drew dodecamer, two chemical features N^5 (of one C) and N^4H (of another C) approach fairly closely (Fig. 7e and f), and thus it would be possible that the two features are properly bridged when the step is more rolled-untwisted. (Recently the structure of PurR–DNA complex has been determined (53). Although the co-ordinates have not been published, it is stated in the original report that the DNA is bent towards the major groove at two CG steps by $\sim 40^\circ$ at each.)

The importance of a hydrogen bond formed between the neighboring basepairs was first noticed by Nelson *et al.* (54) in the context of understanding high propeller twisting in an A tract but such an inter-basepair hydrogen bond is not confined to propeller twisting and is found in many different DNA structures (55,52).

Importance of the central step

Combination of two rolling steps on the same phase i.e. with the separation of 10 bp as seen in the CAP binding site, obviously enhances the bending. However, if two such steps are separated 6 bp apart in the standard B-DNA, the two are placed on almost opposite sides and thus cancel each other’s effects to large extent (Fig. 8b). Thus the 6 bp separations found in the E2 and TBP binding sites appear to be puzzling. The clue to understand the separations lies in the steps at the centre.

The central step in the TBP binding site is positively rolled and untwisted, while that in the E2 binding site is negatively rolled and overtwisted (Fig. 3). By this untwisting or overtwisting at the centre, the major groove sides of the two steps separated 6 bp apart approach one side of the DNA (Fig. 8). When the DNA is untwisted at the centre (TBP), it pulls the major groove sides of the two steps back to one side of the DNA so that the three minor grooves face the same side (Fig. 8c). When overtwisted (E2), it pushes the two rolled steps towards the other side of the DNA so that the two major grooves and the other minor groove at the centre face the same side (Fig. 8a). The two types of bending are different i.e. since the two steps at the edges roll around the major groove, in the former type found in the complexes with CAP, E2, MetJ, 434R and 434C, the DNA is bent around the minor groove at the centre, while in the other type found in the complexes with TBP, TrpR and GCN4, the DNA is bent around the major groove

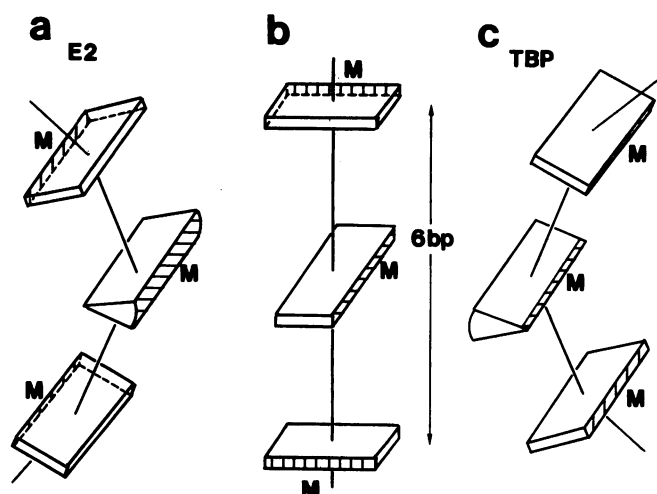


Figure 8. Bending of straight DNA (b) by E2 (a) and TBP (c). The major (M) groove sides of the dinucleotide steps are indicated. Note that the E2 binding site is bent around the minor groove at the centre, while the TBP binding site is bent around the major groove at the centre. The overtwisting (a)/ under-twisting (c) at the central step is coupled with the positive (a)/ negative (c) rolling of the step, so that in both cases the bending is towards the major grooves at the outer steps.

at the centre. The sum of the helical twist angles between the two rolled step in the E2 binding site is 226° , which is 46° pushed towards the far side, while that in the TBP binding site is 107° , which is 73° pulled to the near side.

The under- and overtwisting at the central step coincide with, respectively, positive and negative rolling of the step (Fig. 4), which itself facilitates even more bending of the DNA around the major and minor grooves, respectively, at the centre (Fig. 8). As a consequence, in both types, rolling of the three steps are nicely coupled for bending of the DNA in one direction; the two steps at the edges roll around the major groove and the central step rolls in support halfway between, though its sense (sign) depends on whether it is bending around the major or minor groove at the centre. In other words, the roll-helical twist correlation at the central step can bend the DNA effectively even when the other two rolled steps are spaced 6 bp apart.

Obviously a basepair is under structural influences of the neighboring basepairs on both sides and so a structural change at a step might propagate through 3 bp or so [see also (26)]. Therefore, two unusual steps might not be easily positioned closer than 3 bp and this might be one of the reasons why the three steps, two at the ends and the other at the centre, are incorporated into 6 or more basepairs [see also (17)].

CONCLUSIONS

In this paper we have discussed that bending of DNA by transcription factors can be understood in terms of a pair of highly rolled pyrimidine-purine steps, which bends the helix axis, and of the central pyrimidine-purine step, which changes the phase of the pair. Therefore, our view seems to have common basis with the earlier discussion (17) that DNA can be kinked at particular steps rather than smoothly bent (56). Most of the discussion by other groups has focussed on the deviation of parameters among

usual steps [for example, Hunter (28) has calculated the stacking effects of dinucleotide steps with the roll angle of $0 \pm 5^\circ$] and this might need some modification for applying to highly rolled steps.

Our results and discussion put more weight on the roll-centered view of DNA-bending [see, for example, (30)] than on the tilt-centred view [see classification of the two views in (24) and the references therein]. Another way of describing DNA bending is to do so by measuring the width of major and minor grooves of the DNA. If, for example, the major groove is compressed and the minor groove on the opposite side is widened, the DNA will be bent. We have analysed the correlation between the closing/opening of the grooves and the rolling of the step in the context of understanding DNA bending by a β -sheet (57). Further discussion in this direction will be given elsewhere.

ACKNOWLEDGEMENTS

We thank Drs A. Klug, J. Finch and D. Loaks for their critical reading of the manuscript. M.S. thanks Prof. C.R. Calladine and Dr M. ElHassan for their discussion at early stages of the study.

REFERENCES

- 1 Saenger W. (1983) *Principles of Nucleic Acid Structure*, Springer-Verlag, NY.
- 2 Calladine, C.R. and Drew, H.R. (1992) *Understanding DNA*, Academic Press, London.
- 3 Daune, M. (1993) *Biophysique Moléculaire*, Inter Editions, Paris.
- 4 Frank-Kamenetskii, M. (1993) *Unraveling DNA*, VCH, NY.
- 5 Travers, A. (1993) *DNA-Protein Interactions*, Chapman & Hall, London.
- 6 Neidle, S. (1994) *DNA Structure and Recognition*, IRL, Oxford.
- 7 Satchwell, S.E., Drew, H.R. and Travers, A.A. (1986) *J. Mol. Biol.* **191**, 659-675.
- 8 Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) *Nature* **311**, 532-537.
- 9 Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) *Science* **253**, 1001-1007.
- 10 Hegde, P.S., Grossman, S.R., Laimins, L.A. and Sigler, P.B. (1992) *Nature* **359**, 505-512.
- 11 Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) *Nature* **365**, 520-527.
- 12 Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P. (1993) *Nature* **365**, 512-520.
- 13 Van de Ven, F.J. and Hilbers, C.W. (1988) *Eur. J. Biochem.* **178**, 1-38.
- 14 Kennard, O. and Hunter, W.N. (1989) *Q. Rev. Biophysics* **22**, 327-379.
- 15 Crothers, D.M., Gartenberg, M.R. and Sharader, T.E. (1991) *Methods Enzymol.* **208**, 118-146.
- 16 Dickerson, R.E. (1992) *Methods Enzymol.* **211**, 67-111.
- 17 Crick, F.C. and Klug, A. (1975) *Nature* **255**, 530-533.
- 18 Rhodes, D. and Klug, A. (1981) *Nature* **292**, 378-380.
- 19 Calladine, C.R. and Drew, H.R. (1984) *J. Mol. Biol.* **178**, 773-782.
- 20 Rhodes, D. and Klug, A. (1986) *Cell* **46**, 123-132.
- 21 Koo, H.S., Wu, H.M. and Crothers, D.M. (1986) *Nature* **320**, 501-506.
- 22 Diekman, S. (1986) *FEBS Lett.* **195**, 53-56.
- 23 Zinkel, S.S. and Crothers, D.M. (1987) *Nature* **328**, 178-181.
- 24 Travers, A. and Klug, A. (1987) *Phil. Trans. R. Soc. Lond.* **B317**, 537-561.
- 25 Heinemann, U., Alings, C. and Lauble, H. (1990) *Nucleosides Nucleotides* **9**, 349-354.
- 26 Yanagi, K., Privé, G.G. and Dickerson, R.E. (1991) *J. Mol. Biol.* **217**, 201-214.
- 27 Travers, A. (1991) *Curr. Opin. Struc. Biol.* **1**, 174-122.
- 28 Hunter, C.A. (1993) *J. Mol. Biol.* **230**, 1025-1054.
- 29 Bruckner, I., Dlakic, M., Savic, A., Susuc, S., Pongor, S. and Suck, D. (1993) *Nucleic Acids Res.* **21**, 1025-1029.
- 30 Goodsell, D.S., Kaczor-Grzeskowiak, M. and Dickerson, R.E. (1994) *J. Mol. Biol.* **234**, 79-96.
- 31 Goodsell, D.S. and Dickerson, R.E. (1994) *Nucleic Acids Res.* **22**, 5497-5503.
- 32 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F.Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535-542.

- 33 Berman, H.M., Olson, W.K., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. and Beveridge, D. (1991) *The Nucleic Acid Database: A Guide to the Use of a Relational Database of Nucleic Acid Crystal Structures*. Technical Report.
- 34 Babcock, M., S., Pednault, E.P.D. and Olson, W.K. (1993) *J. Biomol. Struct. Dynamics* **11**, 597-628.
- 35 Babcock, M., S., Pednault, E.P.D. and Olson, W.K. (1994) *J. Mol. Biol.* **237**, 125-156.
- 36 Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N. and Kennard, O. (1989) *EMBO J.* **8**, 1-4.
- 37 Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N. and Kennard, O. (1989) *J. Mol. Biol.* **205**, 787-791.
- 38 Fratini, A.V., Kopka, M.L., Drew, H.R. and Dickerson, R.E. (1982) *J. Biol. Chem.* **257**, 14686-14070.
- 39 Nekudova, L. and Pabo, C.O. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6948-6952.
- 40 Shakked, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joakimiak, A. and Sigler, P.B. (1994) *Nature* **368**, 469-473.
- 41 Dickerson, R.E. and Drew, H. R. (1981) *J. Mol. Biol.* **149**, 761-768.
- 42 Klug, A., Jack, A., Viswamitra, M.A., Kennard, O., Shakked, Z. and Steitz, T.A. (1979) *J. Mol. Biol.* **131**, 669-680.
- 43 Calladine, C.R. (1982) *J. Mol. Biol.* **161**, 343-352.
- 44 Dickerson, R.E. (1983) *J. Mol. Biol.* **166**, 419-441.
- 45 Suzuki, M., Yagi, N. and Gerstein, M. (1995) *Prot. Engng* in press.
- 46 Bhattacharyya, D. and Bansal, M. (1992) *J. Biomol. Struct. Dynamic.* **10**, 213-226.
- 47 Sponer, J. and Kypr, J. (1993) *J. Biomol. Struct. Dynamic.* **11**, 027-041.
- 48 Sobell, H.M., Tsai, C.-C., Jain, S.C. and Gilbert, S.G. (1977) *J. Mol. Biol.* **114**, 333-365.
- 49 Zhurkin, V.B. (1985) *J. Biomol. Struct. Dynamic.* **2**, 785-804.
- 50 Lipanov, A., Kopka, M.L., Kaczor-Grzeskowiak, M., Quintana, J. and Dickerson, R.E. (1993) *Biochemistry* **32**, 1373-1389.
- 51 Klug, A. (1993) *Nature* **365**, 486-487.
- 52 Suzuki, M. and Yagi, N. (1994) *Proc. Japan Acad.* **B70**, 190-193.
- 53 Schmacher, M.A., Choi, K.Y., Zalkin, H. and Brennan, R.G. (1994) *Science* **266**, 763-770.
- 54 Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) *Nature* **330**, 221-226.
- 55 Timsit, Y., Vilbois, E. and Moras, D. (1991) *Nature* **354**, 167-170.
- 56 Trivonov, E.N. (1985) *CRC Crit. Rev. Biochem.* **19**, 89-106.
- 57 Suzuki, M. and Yagi, N. (1995) *Proc. Japan Acad.* **B71**, 108-113.
- 58 Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C., Joakimiak, A., Marmorstein, R.Q., Luisi, B.F. and Sigler, P.B. (1988) *Nature* **335**, 321-329.
- 59 Lawson, C.L. and Carey, J. (1993) *Nature* **366**, 178-182.
- 60 Shimon, L.J.W. and Harrison, S.C. (1993) *J. Mol. Biol.* **232**, 826-838.
- 61 Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988) *Nature* **242**, 899-907.
- 62 Rodgers, D.W. and Harrison, S.C. (1993) *Structure* **1**, 227-240.
- 63 Mondragón, A. and Harrison, S.C. (1991) *J. Mol. Biol.* **219**, 321-334.
- 64 Clarke, N.D., Beamer, J.L., Goldberger, H.R., Berkower, C. and Pabo, C.O. (1991) *Nature* **254**, 267-270.
- 65 Feng, J.-A., Johnson, R.C. and Dickerson, R.E. (1994) *Science* **263**, 348-355.
- 66 König, P. and Richmond, T.J. (1993) *J. Mol. Biol.* **233**, 139-154.
- 67 Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) *Cell* **71**, 1223-1237.
- 68 Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) *Nature* **352**, 495-505.
- 69 Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell* **63**, 579-590.
- 70 Pavletich, N.P. and Pabo, C.O. (1991) *Science* **252**, 809-817.
- 71 Pavletich, N.P. and Pabo, C.O. (1993) *Science* **261**, 1701-1707.
- 72 Marmorstein, R., Carey, M., Ptashne, M. and Harrison, S.C. (1992) *Nature* **356**, 408-414.
- 73 Fairall, L., Schwabe, J.W.R., Chapman, L., Finch, J.T. and Rhodes, D. (1993) *Nature* **66**, 483-487.
- 74 Raumann, B.E., Rould, M.A., Pabo, C.O. and Sauer, R.T. (1994) *Nature* **367**, -757.
- 75 Somers, W.S. and Phillips, S.E.V. (1992) *Nature* **359**, 387-393.
- 76 McCall, M., Brown, T. and Kennard, O. (1985) *J. Mol. Biol.* **183**, 385-396.
- 77 Lauble, H., Franj, R., Blocker, H. and Heinemann, U. (1988) *Nucleic Acids Res.* **16**, 7799-7816.
- 78 Eisenstein, M., Frolow, F., Shakked, Z. and Rabinovich, D. (1990) *Nucleic Acids Res.* **18**, 3185-3194.
- 79 Heinemann, U., Lauble, H., Frank, R. and Bloecker, H. (1987) *Nucleic Acids Res.* **15**, 9531-9550.
- 80 Haran, T.E., Shakked, Z., Wang, A.H.-J. and Rich, A. (1987) *J. Biomol. Struct. Dyn.* **5**, 199-217.
- 81 Jain, S.C., Zon, G. and Sundralingam, M. (1989) *Biochemistry* **28**, 2360-2364.
- 82 Hunter, W.N., Langlois D'Estaintot, B. and Kennard, O. (1989) *Biochemistry* **28**, 2444-2451.
- 83 Courseille, C., Dautant, A., Hospital, M., Langlois D'Estaintot, B., Precigoux, G., Molko, D., and Teoule, R. (1990) *Acta Crystallogr.* **A46**, 9-12.
- 84 Cervi, A., Langlois D'Estaintot, B. and Hunter, W.N. (1992) *Acta Crystallogr.* **B48**, 714-719.
- 85 Bingman, C., Li, X., Zon, G. and Sundralingam, M. (1992) *Biochemistry* **31**, 12803-12812.
- 86 Thota, N., Li, X.H., Bingman, C. and Sundralingam, M. cited as 'to be published (ADH038)' in Nucleic Acid Data Base.
- 87 Thota, N., Li, X.H., Bingman, C. and Sundralingam, M. cited as 'to be published (ADH039)' in Nucleic Acid Data Base.
- 88 McCall, M., Brown, T., Hunter, W.N. and Kennard, O. (1986) *Nature* **322**, 661-664.
- 89 Frederick, C.A., Quigley, G.J., Teng, M.-K., Coll, M., Van der Marel, G.A., Van Boom, J.H., Rich, A. and Wang, A.H.-J. (1989) *Eur. J. Biochem.* **181**, 295-307.
- 90 Ramakrishnan, B. and Sundralingam, M. (1993) *J. Mol. Biol.* **231**, 431-444.
- 91 Bingman, C.A., Zon, G. and Sundralingam, M. (1992) *J. Mol. Biol.* **227**, 738-756.
- 92 Bingman, C.A., Zon, G. and Sundralingam, M. (1992) *Nucleic Acids Res.* **20**, 6637-6647.
- 93 Heinemann, U. and Alings, C. (1989) *J. Mol. Biol.* **210**, 369-381.
- 94 Privé, G.G., Yanagi, K. and Dickerson, R.E. (1991) *J. Mol. Biol.* **217**, 177-199.
- 95 Goodsell, D.S., Kopka, M.L., Cascio, D. and Dickerson, R.E. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2930-2934.
- 96 Grzeskowiak, K., Yanagi, K., Prive, G.G. and Dickerson, R.E. (1991) *J. Biol. Chem.* **266**, 8661-8883.
- 97 Yuan, H., Quintana, J. and Dickerson, R.E. (1992) *Biochemistry* **31**, 8009-8021.
- 98 Heinemann, U., Alings, C. and Bansal, M. (1992) *EMBO J.* **11**, 1931-1939.
- 99 Grzeskowiak, K., Goodsell, D.S., Kaczor-Grzeskowiak, M., Cascio, D. and Dickerson, R.E. (1993) *Biochemistry* **32**, 8923-8931.
- 100 Yoon, C., Privé, G.G., Goodsell, D.S. and Dickerson, R.E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6332-6336.
- 101 DiGabriele, A.D., Sanderson, M.R. and Steitz, T.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1816-1820.
- 102 Westhof, E. (1987) *J. Biomol. Struct. Dynamic.* **5**, 581-600.
- 103 Narayana, N., Ginell, S.L., Russu, I.M. and Berman, H.M. (1991) *Biochemistry* **30**, 4449-4455.
- 104 Christin, M.C., Frederick, C.A., Wang, A.H.-J. and Rich, A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8385-8389.
- 105 DiGabriele, A.D. and Steitz, T.A. (1993) *J. Mol. Biol.* **231**, 1024-1039.