FEBS Letters 379 (1996) 148–152 FEBS 16577

Role of base–backbone and base–base interactions in alternating DNA conformations

Masashi Suzuki[a,*], Naoto Yagi[b], John T. Finch[a]

[a]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*
[b]*Tohoku University, School of Medicine, Seiryo-machi, Sendai, 980-77, Japan*

Received 20 November 1995; revised version received 14 December 1995

Abstract Sequence-specific conformational differences between dinucleotide steps are characterised using published crystal coordinates with special attention to steric hindrance of the methyl group of a T base to the neighbouring base, and, more importantly, to the sugar–phosphate backbone. The TT step is inflexible and B-like, as it has two methyl groups which interlock with each other and with the sugar–phosphate backbones. AT slides, or overtwists, so that the methyl groups move away from the backbones, both lead the step towards the A-conformation. TA is most flexible as it does not have such restriction. These characteristics are observed with other pyrimidine–pyrimidine, pyrimidine–purine, purine–pyrimidine steps, respectively, but to less extent, depending on the number of non-A:T basepairs in the steps.

Key words: Nucleic acid; Crystal structure; Sequence–structure correlation; Structural biology

1. Introduction

Possible sequence–structure correlation in DNA has been discussed by many scientists for years without coming to a simple and clear conclusion (see a review [1], and original reports [2–7]). However, all agree with one major point, that stacking of the two neighbouring basepairs in a dinucleotide step is important; a pyrimidine(Y)–purine(R) step has a poor stacking, while the base stacking at a purine(R)–pyrimidine(Y) step is tighter.

The limitation of earlier arguments might be due to the lack of understanding the interactions between bases and the sugar–phosphate backbone. For example, Hunter [5] stated that 'since the sugar–phosphate backbone appears to act as no more than a constraint on the ranges of conformational space accessible to the bases, the computational problem has been simplified by ignoring it completely'. As we discuss in this paper, the first half of his statement is untrue and the second half is misleading. It has been noted by many crystallographers that the methyl groove of a T base can approach very closely to the backbone, depending on the nucleotide sequence (for example, see Fig. 8 of Coll et al. [8]).

DNA molecules which are composed only of A:T and T:A basepairs have attracted the attention of many scientists, since these have 'unusual' characteristics [2,9–15]. A/T-rich sequences are indeed interesting, since the characteristics of a T base as a pyrimidine base are enhanced by its methyl group (see Fig. 9 of ref. 6); the major groove side of a basepair is tilted so that the bulky pyrimidine base, T or C, is pushed into the groove, and the methyl group is at the furthest end of the tilted major groove edge. Thus possible sequence dependent variety of DNA conformations, which originate from the Y–R antisymmetry, might be 'enhanced' in A/T-rich sequences. Indeed, such sequences can fold into a wide range of structures from a very rigid and B-like one [11] to a very distorted one characterised by largely untwisted helicity and a very wide minor groove [13,14].

In this paper we aim firstly to understand the sequence specific characteristics of DNA conformation by using published crystal coordinates, and secondly to explain the characteristics by focussing attention on the position of the methyl groups of T bases in the DNA structures.

2. Materials and methods

The database of DNA crystal structures described earlier [7] was used for this study. The database contains the following structures; ADH006 (NADB code name), ADH007, ADH030, ADH018, ADH012, ADH014, ADH020, ADH023, ADH041, ADH047, ADH038, ADH039, ADI0009, ADJ022, ADJ049, ADL045, ADL046, BDJ017, BDJ019, BDJ051, BDJ025, BDJ036, BDJ031, BDJ039, BDJ052, BDL006, BDL007, BDL015, BDL020, BDL028, BDL038, BDL047, BDL042, 2BOP (PDB code name), 1CGP, 1TRO, 1TRR, 1RPE, 2OR1, 1PER, 3CRO, 1LMB, 1HCR, 1DGC, 1YSA, 1GLU, 1HDD, 1ZAA, 1DRR, 2DRP, 1PAR, and 1CMA. Two sets of coordinates, DNA–GLI complex and DNA–TBP complex, were kindly provided by Prof. Burley.

Calculation of the dinucleotide step parameters [16] was carried out using a computer program [17,18]. The values were averaged separately for the ten types of dinucleotide steps found in A-DNA, B-DNA, and DNA in complexes with transcription factors (Table 1).

3. Results and discussion

3.1. Variable and invariable parameters

The geometry of two neighbouring basepairs in a dinucleotide step is described by six parameters; three rotational angles (helical twist, roll, and tilt) and three translational distances (rise, slide, and shift) ([16], see also Fig. 1 of this paper). Among the six parameters (Table 1) some are fairly constant, while the others vary from one type of step to another (this can be judged by comparing the averaged values) or within one type (this can be judged by the standard deviation).

The rise distance is most conserved not only at each type but also between the types. It does not change much even between A-DNA and B-DNA. This clearly indicates how important the base stacking is for the DNA structure. The GC step in B-DNA has the highest rise on average, which might be a consequence of overlap of the two negatively charged G bases [5], but the deviation from the other types of steps is small.

*Corresponding author.

SSDI 0014-5793(95)01506-X

The shift distance remains nearly zero at all the steps. A symmetric sequence such as TA has, by definition, no shift or tilt on average [16]. However, the standard deviation of shift is very small (less than 1 Å in all the steps and in most of them 0.5 Å or smaller) and thus the movement in this direction is very limited. The tilt angle scarcely exceeds 5 degrees anywhere. Higher tilting is found at the YY steps, which increases the separation of the two Y bases along the helix axis and decreases that of the partner R bases (this is likely to be due to the shape of a Y base, which can create a larger movement with the same roll angle, and thus it is necessary to keep more distance on the side of Y's).

It seems, therefore, reasonable to concentrate on the remaining three parameters, slide, roll, and helical twist.

3.2. Slide–twist correlation

As in our earlier papers [6,7] our aim is to understand the characteristics of the three parameters, slide, roll, and helical

Table I. Dinucleotide step parameters

Step	Roll(°)	Twist(°)	Tilt(°)	Rise(Å)	Slide(Å)	Shift(Å)
A-DNA						
YR						
TG(9)	10.6±2.2	29.1±4.7	-1.4±3.4	3.2±0.2	-1.2±0.4	0.3±0.2
CG(28)	7.2±7.7	28.8±5.1	0.0±2.5	3.3±0.1	-1.8±0.3	0.0±0.5
TA(24)	8.6±3.3	29.5±5.3	0.0±2.1	3.3±0.1	-1.2±0.2	0.0±0.2
YY						
TT(0)	ND	ND	ND	ND	ND	ND
TC(8)	11.0±7.2	32.9±6.4	0.3±3.3	3.3±0.1	-1.4±0.2	-0.1±0.5
CT(7)	3.9±5.4	38.9±10.9	0.7±2.1	3.3±0.1	-1.3±0.1	-0.1±0.5
CC(39)	6.0±4.9	31.4±3.9	-1.1±2.5	3.4±0.2	-1.7±0.4	0.0±0.4
RY						
GT(32)	5.7±4.2	32.7±2.3	-0.3±2.2	3.3±0.1	-1.3±0.2	-0.2±0.5
GC(18)	5.9±2.8	30.8±3.0	0.0±1.8	3.3±0.1	-1.2±0.2	0.0±0.3
AT(4)	6.3±1.8	29.3±5.8	0.0±5.2	3.3±0.1	-1.1±0.0	0.0±0.9
B-DNA						
YR						
TG(21)	0.5±7.2	39.2±9.8	-0.7±3.0	3.4±0.1	1.6±1.1	0.1±0.4
CG(76)	3.1±5.6	36.6±5.0	0.0±3.7	3.4±0.2	0.7±0.6	0.0±0.6
TA(14)	2.8±4.3	40.5±5.4	0.0±4.0	3.4±0.1	0.9±0.9	0.0±0.4
YY						
TT(42)	0.3±4.2	35.3±3.9	0.5±3.2	3.3±0.1	-0.1±0.4	0.0±0.4
TC(17)	-1.3±2.7	40.3±2.2	0.9±2.9	3.3±0.1	0.0±0.3	0.0±0.3
CT(7)	4.5±3.2	31.2±6.2	2.8±1.4	3.3±0.1	0.4±0.4	-0.4±0.4
CC(14)	6.0±2.8	33.3±5.5	2.7±2.4	3.4±0.1	0.8±0.3	0.0±0.5
RY						
GT(9)	0.5±5.4	32.6±4.7	0.1±4.5	3.3±0.1	-0.2±0.5	0.2±0.5
GC(40)	-6.2±7.0	37.3±3.5	0.0±4.1	3.5±0.2	0.4±0.5	0.0±0.8
AT(36)	-0.8±3.8	31.2±3.8	0.0±1.9	3.3±0.1	-0.4±0.3	0.0±0.4
COMPLEX						
YR						
TG(33)	6.4±8.5	35.9±6.8	0.3±3.3	3.4±0.2	0.4±0.8	0.0±0.6
CG(46)	6.5±6.1	34.9±4.3	0.0±3.6	3.4±0.1	0.7±0.6	0.0±0.8
TA(84)	2.7±7.1	39.5±4.8	0.0±3.6	3.4±0.2	0.1±0.8	0.0±0.4
YY						
TT(48)	0.8±4.0	35.6±4.2	1.9±3.7	3.3±0.2	0.1±0.5	0.1±0.5
TC(42)	2.4±4.7	35.7±4.9	1.7±4.2	3.4±0.1	0.1±0.6	0.3±0.4
CT(49)	5.6±3.8	31.9±5.0	1.3±3.9	3.4±0.2	-0.3±0.4	-0.2±0.6
CC(23)	3.3±5.4	33.3±5.2	1.0±5.6	3.4±0.1	-0.1±0.6	0.0±0.8
RY						
GT(67)	-0.2±4.1	31.1±3.9	-0.1±3.2	3.4±0.1	-0.6±0.4	-0.1±0.6
GC(18)	-2.0±4.2	34.6±4.9	0.0±3.2	3.4±0.1	-0.3±0.6	0.0±0.6
AT(54)	0.0±3.4	29.3±3.8	0.0±3.2	3.3±0.2	-0.7±0.4	0.0±0.4

The numbers of examples are shown in parentheses. The numbers discussed in the text are underlined. ND: not determined.

twist, in terms of changes of the slide and roll parameters with that of helical twist, keeping the important rise distance nearly constant.

The correlation of roll and helical twist has been analysed in some detail (see Fig. 4 of ref. 7, Fig. 9 of ref. 6) and can be explained as follows. A dinucleotide step is helically twisted as a result of the sugar–phosphate backbone distance being about twice the base-stacking distance (Calladine and Drew [19], see also Fig. 1a of this paper). If a step is untwisted (Fig. 1b), the basepairs are pushed apart and the rise distance increases. Such geometry can occur and is used for drug intercalation (see, for example, the DNA–actinomycinD complex [20]). To regain the stacking, i.e. to decrease the rise distance the step then rolls around the major groove (by upto about 45 degrees [6,7,21,22]; the positive rolling in Fig. 1c).

The rise distance can decrease by sliding as well (Fig. 1d,e). An RY step is more difficult to roll than a YR step because of its tighter stacking [6,7], and indeed the RY steps in B-DNA and in the protein complexes in particular, AT and GT adopt negative sliding (on average −0.2 to −0.7 Å, note also that the standard deviation of the parameter at the two steps is not high, 0.3 to 0.5 Å) together with a small helical twist (29.3–32.6 degrees on average). In this regard the AT and GT steps may be regarded as adopting an A-like conformation even in B-DNA. A-DNA is characterised by a negative slide distance, −1.1 to −1.8 Å, while most other steps in B-DNA have nearly zero or positive values (Table 1). Also, of course, steps in A-DNA have a smaller helical twist.

(Another XT step, CT, has a small helical twist angle. The slide parameter of CT in the protein complexes is negative but it is positive in B-DNA. Thus it seems to be intermediate between AT and TT, see sections 3.3 and 3.4.)

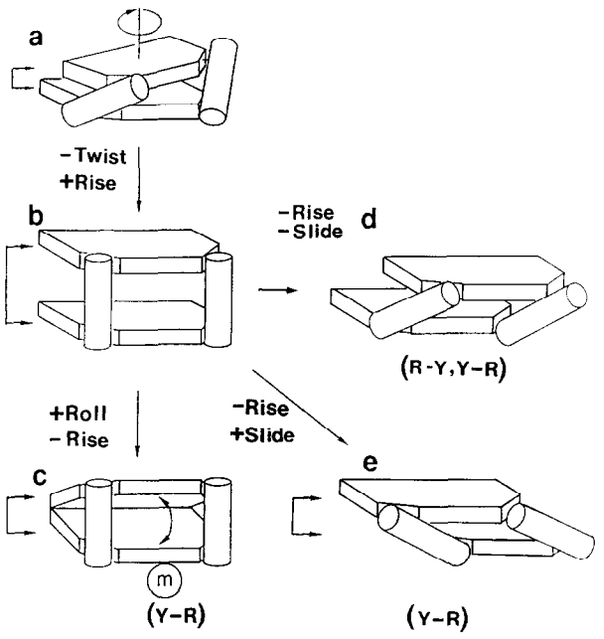

Fig. 1. Conformational changes of a dinucleotide step. The steps are looked from the minor groove side (indicated by 'm' in (c)). Base pairs are schematically drawn as rectangles. The sugar–phosphate backbones are schematically drawn as a pair of straight bars connecting the basepairs. The rise distance is indicated by a pair of arrows on the left of each step. RY, YR: purine–pyrimidine and pyrimidine–purine steps, respectively.

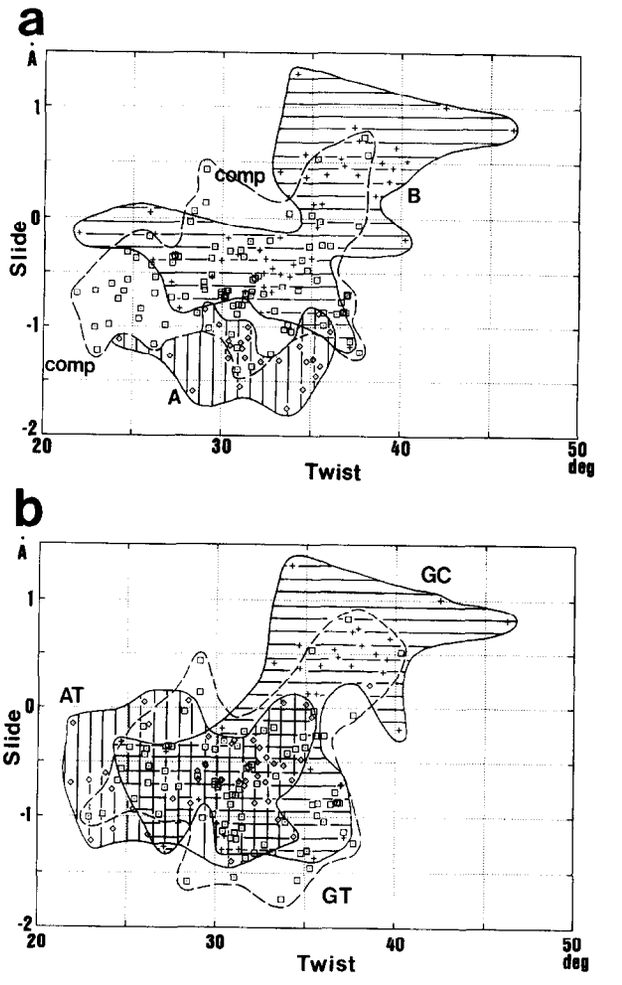

Fig. 2. Correlation of slide and helical twist parameters at purine–pyrimidine steps. (a) Those in A-DNA (◇), B-DNA (+), and in complexes with transcription factors (□). The region in which the entries from B-DNA are found is shadowed by horizontal lines, while the region in which the entries from A-DNA are found by vertical lines. Note that the entries from A-DNA and those from B-DNA distribute so that the two steps have almost no overlap, while those from the protein complexes fill the gap overlapping with those from A and B. (b) Those at AT (◇), GT (□), and GC (+). The region in which the entries fron GC steps are founs is shadowed by horizontal lines, while the region in which the entries from AT steps are found by vertical lines. The AT steps distribute more on the side with smaller helical twist and higher negative slide (A-like), GC is on the other side (B-like), and GT is intermediate.

The slide–twist correlation of RY steps and the importance of the A:T basepairs in this correlation can be shown by plotting the two parameters against each other (Fig. 2). The plot shows that AT, which has two A:T basepairs, is positioned on one side with larger negative slide and smaller helical twist (A-like), and GC, which has no A:T, on the other side (B-like), while GT, which has one A:T, is intermediate (Fig. 2b). In brief, the RY steps change along the slide–twist correlation from an B-like conformation to a A-like conformation (Fig. 2a) depending on the number of A:T basepairs in the steps.

Now we seem to be facing two important questions: (1) why AT and GT slide only in the 'negative' direction to decrease the rise distance created by untwisting of the helicity, and (2) why

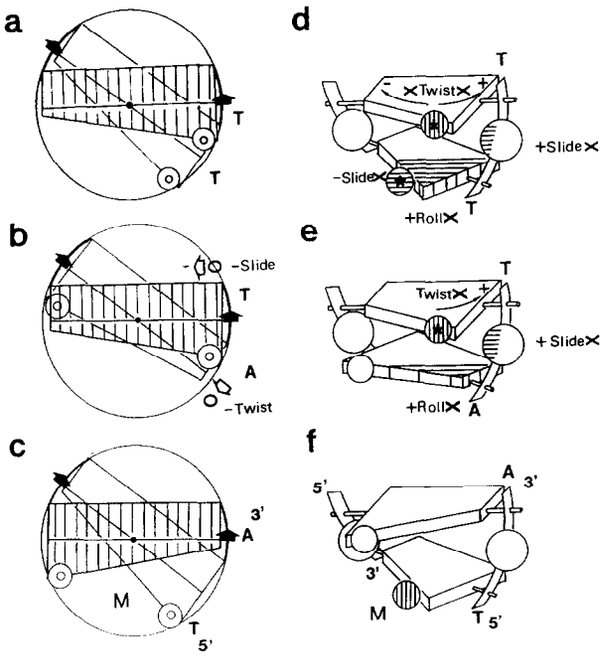

Fig. 3. Schematic drawings of TT (a,d), AT (b,e), and TA (c,f) steps. (a–c) The T base in an A:T basepair is pushed into the major (M) groove. The double circles show the positions of methyl groups in T bases on the major (M) groove side (see (c)). White arrows in (b) show the directions in which the steric hindrance can be relaxed (marked with single circles). The 5′–3′ directions are shown by black arrows. Compare these subfigures with Fig. 9 of ref. 6. The basepairs closer to the reader are shadowed. The centres of basepairs are shown by filled circles (the centre of a basepair is defined as that of the line which connects C6 of the pyrimidine and C8 of the purine). (d–f) The steps are seen from the major groove side (indicated by 'M' in (f)). The methyl groups are drawn as smaller speres. Those which are positioned from the strand drawn on the right are shadowed by vertical lines. The methyl groups which cause serious steric hindrance are marked (*). The bulky parts of the backbones are schematically drawn as larger spheres. In (d) and (e) the sugar–phosphate strand 5′-terminal to the T base on which a methyl, and the base 5′-terminal become close to the methyl group (shadowed by horizontal lines, the distance between the methyl group and C2′ sugar is about 3.4 Å in the standard B-DNA). The directions of the movements which would create collision with the sugar–phosphate backbone and/or the neighbouring base are indicated with crosses. The distance between the two methyl groups at a TT step is about 4.8 Å in the standard B-DNA.

the degree of sliding and untwisting depends on the number of A:T basepairs.

3.3. *Flexible TA, inflexible B-like TT, and inflexible A-like AT*

Before answering the questions of the RY steps, we discuss some interesting features of other types of steps shown in our statistics, which will be explained later in this paper by the same principle as that will be applied on understanding the RY steps.

The standard deviations found in the parameters at each type of YY step are generally similar to those found at each type of RY (Table 1), and thus the two groups have similar extent of flexibility. However, the average values themselves of YY steps are different from those of RY.

In particular, TT, CT and TC, are characterised by smaller slide distances on average. Also the averaged helical twist angles of these steps are closer to 36 degrees, meaning that these structures are closer to a standard B-conformation. It might be interesting to note that even in A-DNA, YY steps have a higher helical twist than the average (Table 1).

In contrast, the YR steps in B-DNA and in the protein complexes are characterised with high standard deviations in roll, helical twist, and slide parameters, and thus these are more flexible than the others. The slide parameter at YR steps is positive on the average, and TA and TG can slide in both directions (thus their standard deviations are larger than the averaged values).

In brief, AT can slide in only one direction or untwist (towards an A-like conformation), and TT is equally inflexible but stays near the standard B-conformation, while TA is flexible and can change in many different directions.

3.4. *Steric hindrance imposed by the methyl group in T*

The characteristics of dinucleotide steps described in the earlier paragraphs can be explained simply by focussing attention on steric hindrance imposed by the methyl groups of the T bases to the neighbouring bases, and, more importantly, to the nearest parts of sugar–phosphate backbones (Fig. 3). Thus, in the AT step the T base at the 3′-terminus of one of the strands (marked '*' in Fig. 3e) is physically close to the part of sugar–phosphate backbone between the T base and the 5′-terminal A base (Fig. 3b). The distance between the methyl of the T and C5′ of the 5′-terminal sugar is as close as 3.8 Å in the standard B-DNA. Therefore, to avoid possible collision between the methyl group and the backbone the step must either slide in the negative direction and/or untwist (shown by white arrows in Fig. 3b) but it cannot change in the opposite directions (shown by crosses in Fig. 3e), resulting a very A-like conformation (Fig. 1d).

The AT step also cannot roll in the positive direction or it would collide with the 5′-terminal A base and/or the sugar–phosphate backbone (Fig. 3e). Similarly the GT step needs to adopt a small twist angle and a negative slide distance, but at the AT step the same types of constraints arise from the other T base on the other strand and thus, being doubly constrained, the AT step has the smallest twist angle and the largest negative slide distance among the RY steps and indeed among all the steps (Table 1). This explains the differences between the RY steps shown in Fig. 2 (needless to say, GC has no constraint from a methyl group and remains more normal).

The TT step (Fig. 3a,d) has the same constraints as those of AT and, in addition, it cannot slide in the positive direction or it would collide with the methyl group of the other T base. The distance between the two methyl groups is about 4.8Å in the standard B-DNA. For the same reason it cannot appreciably untwist. This is probably the reason why the TT step remains very B-like (it might appear that the TT step has constraints more sevier than those of AT but it should be remembered that AT has constraints on both strands).

The TA step (Fig. 3c,f) and other YR steps do not have any of the above constraints. Thus these are most flexible (they can also roll [6,7]) and can adopt many different conformations (Fig. 1c,d,e).

The simple model discussed in this paper introduces a rather naive way of understanding the effects of interactions between bases and the backbone on DNA conformation. Nevertheless, the model can explain many sequence dependent characteristics of DNA conformation, which, to our knowledge, had not been

explained or even noticed in earlier work. Thus the interactions seem to be most important.

References

[1] Dickerson, R.E. (1992) Methods Enzymol. 211, 67–111.
[2] Klug, A., Jack, A., Viswamitra, M.A., Kennard, O., Shakked, Z. and Steitz, T.A. (1979) J. Mol. Biol. 131, 669–680.
[3] Calladine, C.R. (1982) J. Mol. Biol. 161, 343–352.
[4] Dickerson, R.E. (1983) J. Mol. Biol. 166, 419–441.
[5] Hunter, C.A. (1993) J. Mol. Biol. 230, 1025–1054.
[6] Suzuki, M., Yagi, N. and Gerstein, M. (1995) Protein Eng. 8, 329–338.
[7] Suzuki, M. and Yagi, N. (1995) Nucleic Acids Res. 23, 2083–2091.
[8] Coll, M., Aymami, J., van der Marel, G.A., van Boom, J.H., Rich, A. and Wang, A.H.-J. (1989) Biochemistry 28, 310–320.
[9] Kopka, M.L., Fratini, A.V., Drew, H.R. and Dickerson, R.E. (1983) J. Mol. Biol. 163, 129–146.
[10] Koo, H.-S., Wu, H.-M. and Crothers, D.M. (1986) Nature 320, 501–506.
[11] Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987) Nature 330, 221–226.
[12] Goodsell, D.S., Kaczor-Grzeskowiak, M. and Dickerson, R.E. (1994) J. Mol. Biol. 234, 79–96.
[13] Kim, Y., Greiger, J.H., Hahn, S. and Sigler, P. (1993) Nature 365, 512–520.
[14] Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) Nature 365, 520–527.
[15] Klug, A. (1995) Nature 365, 486–487.
[16] Dickerson, R.E., Bansal, M., Calladine, C.R., Diekmann, S., Hunter, W.N. and Kennard, O. (1989) EMBO J. 8, 1–4.
[17] Babcock, M.S., Pednault, E.P.D. and Olson, W.K. (1993) J. Biomol. Struct. Dynamics 11, 597–628.
[18] Babcock, M.S., Pednault, E.P.D. and Olson, W.K. (1994) J. Mol. Biol. 237, 125–156.
[19] Calladine, C.R. and Drew, H.R. (1992) Understanding DNA, Academic Press, London.
[20] Kamitori, S. and Takusagawa, F. (1992) J. Mol. Biol. 225, 445–456.
[21] Suzuki, M. and Yagi, N. (1994) Proc. Jpn. Acad. B71, 108–113.
[22] Suzuki, M., Loakes, D. and Yagi, N. (1996) Adv. Biophys., in press.
[23] Suzuki, M. and Yagi, N. (1996) J. Mol. Biol., in press.