PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps

M. A. El Hassan and C. R. Calladine

Phil. Trans. R. Soc. Lond. A 1997 **355**, 43-100
doi: 10.1098/rsta.1997.0002

Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps

By M. A. El Hassan and C. R. Calladine

Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

Contents

This paper is concerned with an investigation of the geometry and structure of DNA as revealed by X-ray diffraction of single crystal oligomeric structures. A database of atomic coordinates of 60 *naked* (i.e. not bound to any protein or drug) DNA oligomers (25 dodecamers, 18 decamers, 16 octamers and 1 tetramer) is set up and carefully described. An extensive empirical study of the geometries of DNA dinucleotide steps in the database, involving only unmodified Watson–Crick base pairs (A–T and G–C), is reported, and a number of new correlations and classifications are described in

Phil. Trans. R. Soc. Lond. A (1997) **355**, 43–100
Printed in Great Britain

detail. The main conclusions include the kinematic classification of dinucleotide steps into two main classes: *rigid* and *loose* (or *flexible* or *bistable*). 'Continuously flexible' steps are shown to exercise their flexibility along a well-defined single-degree-of-freedom pattern, with roll, slide and twist all correlated linearly. The rigid steps are AA/TT, AT and GA/TC, and the loose (bistable) steps are GG/CC, GC, CG while the loose (continuously flexible) steps are CA/TG and TA. AC/GT is the least clear of all steps and it is perhaps best described as neither a rigid nor a loose step but rather an 'intermediate' step. The base-pair parameters are also carefully examined and the resulting pivotal correlation between the average propeller and the flexibility of the step (equals the standard deviation of slide), that we have recently described elsewhere (El Hassan & Calladine 1996), is examined in some detail.

A simple two-parameter scheme for the description of the conformation of the sugar phosphate backbone is given and used to classify the sugar phosphate backbones in all entries of our database into A-backbone and B-backbone conformations. The role of the backbone in determining the conformational preferences of the dinucleotide steps is examined by demonstrating that whereas the B-backbone conformation permits a fairly narrow channel in the roll/slide/twist conformational space, with all three parameters linearly correlated, the A-backbone allows only a small 'box' region near the high-roll, low-twist, low-slide end of the space.

Finally, the empirically determined conformational characteristics of the various dinucleotide steps are accounted for in terms of (*a*) mechanical stacking effects associated with propeller-twisting of constituent base pairs (the propeller-flexibility correlation), (*b*) chemical stacking effects associated with the special electrostatic charge distributions and π–π effects in homogeneous G|C steps (Hunter 1993), and (*c*) backbone-dictated effects that govern in the absence of (*a*) and (*b*).

1. Introduction

The crucial step that transformed the subject of DNA structure into an extremely active area of enquiry was Watson and Crick's discovery of the double helix in 1953. Although the thirty years or so that followed saw a flurry of research in the subject area, questions relating to the fine structure of DNA at the base-step level remained unclear. The main impediment to unravelling such questions was the inability of X-ray diffraction studies on fibres to determine more than an 'average' structure for the double helical structure. By the early 1980s a technique had been developed whereby it became possible to produce specific DNA oligomers in sufficiently pure form so as to form single crystals. This opened up the possibility of detailed near-atomic resolution structural studies of DNA. The first such right-handed structure to be solved was the dodecamer d(CGCGAATTCGCG), to a resolution of 2.5 Å, by Dickerson and co-workers (Drew *et al.* 1981). This 'parent' dodecamer, as it came to be known, displayed a number of important features relating to the fine structure of DNA. No longer could DNA be thought of as a regularly stacked molecule that is rigidly defined by the backbone structure. Rather, it became clear that there are some definite sequence-dependent structural perturbations at the dinucleotide step level. Thus, the geometric step parameters helical twist, roll and slide (Dickerson *et al.* 1988) varied from 32° to 45°, −9° to 5° and −0.6 Å to 0.8 Å respectively in this

particular oligomer. Furthermore, the base pairs were not co-planar; the bases were rotated relative to each other about their long axes, or *propeller-twisted*.

Following the solution of the parent dodecamer, a number of other dodecamers were solved. Initially these were mainly sequence-variants of the parent dodecamer. Later, a new class of oligomers whose backbone had a distinctly different pattern appeared. These were 'A' form oligomers, and they were mainly octamers (McCall *et al.* 1985). Rather than having a mean helix axis passing through the base pairs as in existing dodecamers, these octamers had the base pairs displaced from the mean axis in a way that would follow if the consecutive base pairs did not stack in a parallel fashion. The database of solved oligomers, mainly dodecamers and octamers, continued to grow until a new class of very high resolution decamers appeared in the early ninties (Dickerson *et al.* 1991). In many ways, these resembled the earlier dodecamers; but they had crystal packing schemes that were distinct from those of most of the dodecamers.

The growing availability of high resolution data prompted several workers in the field to carry out comparative studies with the general aim of providing some structural classification of the 10 independent types of dinucleotide step† (Dickerson 1992; Dickerson *et al.* 1991, 1994; Grzeskowiak *et al.* 1991; Yanagi *et al.* 1991). Although some sound conclusions have been reached in these studies, some questions still remain unresolved; and these can be attributed mainly to the general approach followed within these comparative studies.

Before discussing such problems, let us summarize the main conclusions of the studies cited above. First, it has been proposed that the proper unit of DNA structure description is not the simple dinucleotide step but a *tetranucleotide unit*. Second, while DNA is not randomly polymorphous, it is nearly so but with *some* sequence-specific features. These features are (i) the straightness of the adenine-tracts, (ii) the flexibility of the junctions between adenine tracts and mixed-sequence DNA, (iii) the preference of the TA step and the GGCC tetramer for positive roll (i.e. roll that compresses the major groove); and (iv) the propensity of Y-C-A-R sequences to adopt very high twists and slides. Here Y = pyrimidine (= T or C) and R = purine (= A or G).

One of the problems with the general approach taken in such studies has been the use of the NEWHELIX package for assessing the dinucleotide step geometry. NEWHELIX, as we have explained elsewhere (El Hassan & Calladine 1995) is a software package that starts the computation of the step parameters from given atomic positions by fitting a *global helix axis* to the oligomer; and then it defines most of the computed parameters with respect to this global reference frame. Now while this is acceptable when dealing with a single oligomer, it can lead to artificial differences when dinucleotide steps from different oligomers are compared. Another problem that can be identified in these earlier studies is the bias towards a certain subset of the available coordinate-data. Thus, all A-form oligomers have conventionally been

† There are *ten* independent types of dinucleotide step by sequence: AA/TT, AC/GT, AG/TC, AT/AT, CA/TG, CG/CG, GA/TC, GC/GC, GG/CC and TA. The terms 'step' or 'dinucleotide step' will be used to refer *either* to a certain step from a specific oligomer e.g. the fifth step in the oligomer d(CGCGAATTCGCG), *or* to a step type as a whole e.g. AA/TT step in general. Where an ambiguity might arise, the terms 'step type' or 'dinucleotide step type' will be used to refer to a step type as a whole. In general, it should be clear from the context whether the term 'step' refers to a step type or to a specific examples from a certain oligomer.

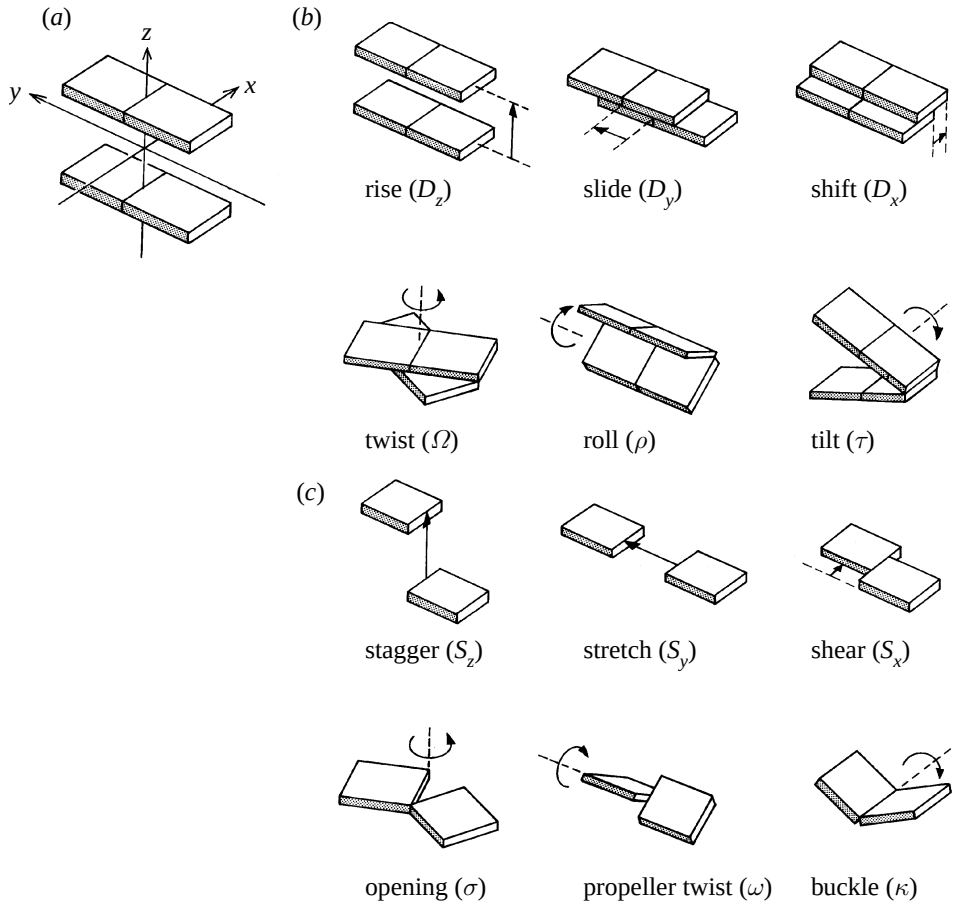

Figure 1. (a) Two consecutive base-pairs, showing the mid-step x, y, z coordinate system. The minor-groove edges of the base pairs are shaded. (b) A schematic display of the six independent dinucleotide step parameters. The relationship of the two base pairs and the coordinate system (not shown) are as in (a). In each case, the arrow shows the movement of the upper base pair relative to the lower. (c) A schematic display of the six independent base-pair parameters. In each case the arrow shows the movement of the base having the higher value of y relative to its pair.

discarded in such studies. The reasons given for this move stem from the claim that there is a general tendency for A-form oligomers to exist only in low-humidity conditions. The counter argument that we invoke here is that an aim of comparative studies is to draw a comprehensive conformational map of all possible right-handed DNA. The question of biological significance is important but it must be remembered that to date no evidence has been presented that eliminates the possibility of A-form DNA in solution (Biburger *et al.* 1995; see also § 5 *a* below). Another vexed issue is the vagueness inherent in the use of the terms A-DNA and B-DNA. Historically, these terms were coined in the early 1950s for the two forms of DNA fibre seen in low- and high-humidity conditions, respectively. Until the early 1980s the term B-form was used to describe DNA conformations where the base pairs stacked in a parallel fashion, perpendicular to the mean helix axis, with approximately zero roll and 36° helical twist; while A-DNA was reserved for DNA conformations with the

base pairs tilted with respect to the mean helix axis, and generally with higher roll and lower helical twist. Nowadays we know that the actual situation is considerably more complicated than this simple picture would suggest.

Thus B-DNA oligomers are known to allow a wide range of conformations. Moreover, many biological processes seem to involve deformation of the DNA that would certainly qualify the constituent steps to be classed as A-DNA or at least non-B-DNA. It is for all of these reasons that we adopt the view here that A-DNA oligomers are important and worthy of investigation. Another key problem with existing studies concerns the completely systematic numerical approach that has been followed. Thus Yanagi *et al.* (1991) list all relevant parameters from a small selection of 11 oligomers (eight dodecamers and three decamers), and try to correlate each parameter with all others. Although this approach is capable of extracting certain correlations, it tends to obscure the physical nature of the problem, which may well be too subtle to be captured by such a systematic approach. The final problem concerns the 'tetrad' hypothesis. An obvious difficulty here is the lack of data on all of the 136 possible tetranucleotides by sequence. Moreover, as we shall discuss below (§ 2), it is not at all obvious that we actually need a tetrad to describe the conformational characteristics in most cases. But, we ought to point out here that, according to Goodsell *et al.* (1993) the GGCC sequence appears to have special structural features that need a tetrad for their proper description. We shall address this issue in some detail in § 7 and § 13 below.

A most promising step towards a better understanding of the stacking preferences of dinucleotide steps was made when Hunter (1993) proposed a chemical model for the prediction of the preferred geometries of the 10 types of dinucleotide step. Hunter's model involves calculating the π–π interactions between the two base pairs and minimizing this interaction energy with respect to the stacking geometry. Those calculations gave results that agreed well with much of the available empirical data in a broad and qualitative sense, despite the limited number of degrees of freedom that were included in the study. Hunter's conclusions will be examined in more detail in § 8, and compared with the empirical findings from the crystal data.

The aim of the present paper is therefore to examine the available database of naked oligonucleotide structures so that a map of the conformational preferences of the 10 types of dinucleotide step can be drawn. Here, the dinucleotide step geometry will be described by the geometric parameters of the *Cambridge Accord* (Dickerson *et al.* 1988) (figure 1), and these will be calculated according to the Cambridge University Engineering Helix Calculation Scheme (CEHS) scheme (El Hassan & Calladine 1995).

The layout of the paper is as follows. First we outline our general approach to the problem, and its underlying philosophy. Next we present a survey of the available data on solved oligomeric crystal structures, and we also examine the distribution of the 400 available step examples among the 10 types of dinucleotide step. We then present a global analysis of the database, with the aim of identifying which of the six step parameters and six base-pair parameters can be safely ignored on account of their lack of sequence-dependent variability. The remaining significant step and base-pair parameters are then examined in detail with the aim of uncovering any underlying correlations. Based on these correlations, we present a kinematic classification of each of the 10 types of dinucleotide step. We then compare our findings with the predictions of Hunter's chemical model, before examining the relationship between the conformation of the backbone and the dinucleotide step geometry. Next

we present a partly mechanical and partly chemical hypothesis that accounts for much of the observed behaviour, before we finally present a summary-map of the conformational preferences of the DNA dinucleotide step types. We conclude with a general discussion.

2. Underlying philosophy

The conformations of dinucleotide steps as observed in single-crystal structures are determined partly by sequence and partly by crystal packing effects (Dickerson *et al.* 1991, 1994). In the present study we regard a given step as having one or several preferred conformations on account of its sequence. The crystalline environment is regarded here as a 'test bed' that determines which of these conformations is adopted, rather than an obstacle in the problem of determining the sequence-specific conformational characteristics. Dickerson *et al.* (1994) have emphasized the need to regard the crystalline environment as having a positive and important role to play in determining the observed conformations, as distinct from being a merely negative artefact.

An implication of the way in which we think of the crystalline environment, is that some steps that are particularly deformable may be seen in many different conformations; whereas steps that are not so easily deformable should persist in more or less a single conformation. In other words, if we encounter a step that is known to be capable of adopting several conformations, then the particular conformation adopted might be dictated by the overall global constraint rather than only the two flanking steps, as implied in the tetrad hypothesis proposed by Yanagi *et al.* (1991). Therefore, by examining each step in several global sequence contexts and under different crystal-packing schemes, we hope to make statements about the conformational characteristics of all types of dinucleotide step. One obvious advantage of such an approach is its ability to give a more complete picture. Another is its direct relevance to the very important phenomenon of DNA bending around proteins, such as the histone octamer (Satchwell *et al.* 1986; Calladine & Drew 1992; Travers 1995), and the bending of DNA around the catabolite gene activator protein, CAP (Schultz *et al.* 1991).

The most immediate consequence of our general philosophy of not excluding segments of the available data is that DNA oligomers that crystallize into the A-form such as McCall's octamer (McCall *et al.* 1985), are included. The key point here is that the A-form is one possible right-handed DNA conformation, and indeed in order for DNA to wrap around nucleosomes, it must clearly adopt 'non-B' conformations. Moreover, as we shall see below, despite the 'backbone-switching' involved in the B to A transition (Calladine & Drew 1984), the step-conformations as described by roll, slide and twist that are accommodated by these two backbone configurations are not severely discontinuous, and in many respects, they form a continuum of conformations that are available for right-handed DNA.

3. Database of oligonucleotide structures

(*a*) *Some general considerations*

There are presently over a hundred solved naked oligomer crystal structures available in public databases such as the Nucleic Acids Database (NDB) at Rutgers and

the Cambridge Crystallographic Data Centre (CCDC) at Cambridge. But before making statements about the conformational characteristics of dinucleotide steps found in those structures, we need to pay careful attention to which oligomer data should be included and which should not. Table 1 gives details of **60** oligonucleotide structures that we have 'short-listed' from public databases. The table gives: (1) journal reference and public database code, (2) sequence/complementary sequence (both in the $5' \longrightarrow 3'$ direction), (3) crystallographic space group, (4) resolution and (5) R factor. The column headed 'T' (type) gives the overall classification of the oligomer as 'A-DNA' or 'B-DNA' in the original database. Structures that have been obtained from the CCDC (WWW URL: `http://csdvx2.ccdc.cam.ac.uk/`) are denoted by their respective (upper case) CCDC-codes. The remaining structures which have been obtained from the NDB (WWW URL: `http://ndbserver.rutgers.edu:80`) are referred to by their (lower-case) NDB codes. Modified (e.g. methylated) base pairs are *italicized*. Unusual (e.g. inosines and uracils) and mispaired bases are highlighted in bold type. Oligomers that have a one-strand asymmetric unit, i.e. those that have only half the oligomer as independent, are indicated by †. Two dodecamers can be seen to appear more than once: bdl015/1/2 (diGabrielle *et al.* 1989) and bdl047/1/2/3 (diGabrielle & Steitz 1993). In both cases the crystallographic asymmetric unit contains multiple dodecamers; 2 for bdl015 and 3 for bdl047.

We shall now discuss the basic considerations that were taken into account in setting up this working database.

The most obvious constraint that we need to think about is that of the resolutions and *R*-factors of the structures. Our approach has been to confine the data broadly within accepted limits on resolution and *R*-factors (around 2.5 Å and 20% respectively), but not to impose stricter constraints which would make the database smaller and thus less useful. The averages (and standard deviations) of the resolution and *R*-factors of all Watson–Crick base pairs in our database are 2.17 Å (0.33 Å) and 17.58% (2.52%) respectively, while the peak values are 2.6 Å and 23.2% respectively. Only two oligomers – the two dodecamers (codes bdl015/1/2) of diGabrielle *et al.* (1989) – have a resolution worse than 2.5 Å, (actually equal to 2.6 Å), and only four have an *R*-factor worse than 20%: three dodecamers (codes bdl047/1/2/3) due to diGabrielle & Steitz (1993) (*R*-factor = 23.2%), and an octamer (code = adh041) due to Cervi *et al.* (1992) (*R*-factor = 21.1%).

Other considerations taken into account in building the database outlined in table 1 were mainly concerned with the uniqueness of the structures. In general, of course, we distinguish two different oligomers by their sequences. However, following from our general philosophy, the crystallographic packing scheme is also important. Thus two oligomers that are identical with respect to sequence but which crystallize into different space groups are here taken to be distinct. According to these rules, then, the oligomers listed in table 1 are all distinct either with respect to sequence or to crystallographic packing scheme. We should mention here that two sequences that only differ with respect to a single base-pair mismatch/modification are taken to be distinct, without reference to the crystallographic packing scheme. The point here is that although the difference in sequence might not cause a drastic enough change such as for example changing the crystallographic space group, it might still cause enough perturbation to warrant regarding the two oligomers in question as structurally distinct. Needless to say, we have eliminated all duplicates in all cases where multiple solutions of the same structure have been deposited by different research groups – who perhaps used different numerical programs to refine their

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1. *A listing of all analysed oligomers*

((*a*) Dodecamers, (*b*) decamers, (*c*) octamers and (*d*) tetramers. Full journal references for all entries are given in the bibliography with the corresponding database code except for the decamer (‡) which has been obtained from A. A. Lipanov, M. L. Kopka, M. Kaczor-Grzeskowiak and R. E. Dickerson (personal communication).)

reference/code	sequence	type	space group	res. (Å)	R (%)
(*a*) **Dodecamers**					
Bingman *et al.* 1992*a*/adl045†	CCGTACGTACGG/CCGTACGTACGG	**A**	$P6_122$	2.5	15.0
Bingman *et al.* 1992*b*/adl046†	GCGTACGTACGC/GCGTACGTACGC	**A**	$P6_122$	2.55	14.0
Brown *et al.* 1986/DODBAH	CGC**G**AATT**A**GCG/CGC**G**AATT**A**GCG	**B**	$P2_12_12_1$	2.5	17.0
Brown *et al.* 1989/DODBAD	CGC**A**AATT**G**GCG/CGC**A**AATT**G**GCG	**B**	$P2_12_12_1$	2.25	16.0
Corfield *et al.* 1987/bdlb10	CGC**I**AATT**A**GCG/CGC**I**AATT**A**GCG	**B**	$P2_12_12_1$	2.5	19.0
diGabrielle *et al.* 1989/bdl015/1	CGCAAAAATGCG/CGCATTTTTGCG	**B**	$P2_12_12_1$	2.6	20.1
diGabrielle *et al.* 1989/bdl015/2	CGCAAAAATGCG/CGCATTTTTGCG	**B**	$P2_12_12_1$	2.6	20.1
diGabrielle & Steitz 1993/bdl047/1	CGTTTTTTCGCG/CGCGAAAAAACG	**B**	$P2_12_12$	2.3	23.2
diGabrielle & Steitz 1993/bdl047/2	CGTTTTTTCGCG/CGCGAAAAAACG	**B**	$P2_12_12$	2.3	23.2
diGabrielle & Steitz 1993/bdl047/3	CGTTTTTTCGCG/CGCGAAAAAACG	**B**	$P2_12_12$	2.3	23.2
Drew *et al.* 1981/DODBAJ	CGCGAATTCGCG/CGCGAATTCGCG	**B**	$P2_12_12_1$	1.9	17.8
Edwards *et al.* 1992/bdl038	CGCAAATTTGCG/CGCAAATTTGCG	**B**	$P2_12_12_1$	2.2	18.1
Fratini *et al.* 1982/bdlb04	CGCGAATT*C*GCG/CGCGAATT*C*GCG	**B**	$P2_12_12_1$	2.3	17.3
Frederick *et al.* 1988/bdlb13	CGCGA*A*TTCGCG/CGCGA*A*TTCGCG	**B**	$P2_12_12_1$	2.0	16.9
Hunter *et al.* 1987*a*/DODBAC	CGC**A**AATT**C**GCG/CGC**A**AATT**C**GCG	**B**	$P2_12_12_1$	2.5	19.0
Hunter *et al.* 1987*b*/bdl009	CGC**G**AATT**T**GCG/CGC**G**AATT**T**GCG	**B**	$P2_12_12_1$	2.5	18.0
Larsen *et al.* 1991/bdl029	CGTGAATTCACG/CGTGAATTCACG	**B**	$P2_12_12_1$	2.5	15.8
Leonard *et al.* 1990/bdlb26	CGC**G**AATT**T**GCG/CGC**G**AATT**T**GCG	**B**	$P2_12_12_1$	2.0	18.5
Leonard *et al* 1992/bdlb41	CGC**A**AATT**I**GCG/CGC**A**AATT**I**GCG	**B**	$P2_12_12_1$	2.5	15.8
Leonard & Hunter 1993/bdl042	CGTAGATCTACG/CGTAGATCTACG	**B**	$C2$	2.25	13.8
Nelson *et al.* 1987/DODBAA	CGCAAAAAAGCG/CGCTTTTTTGCG	**B**	$P2_12_12_1$	2.5	20.0

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1. *Cont.*

reference/code	sequence	type	space group	res. (Å)	R (%)
Skelly *et al.* 1993/bdl046	CGC**G**AATT**G**GCG/CGC**G**AATT**G**GCG	**B**	$P2_12_12_1$	2.2	18.8
Webster *et al.* 1990/bdl022	CGC**A**AGCT**G**GCG/CGC**A**AGCT**G**GCG	**B**	$P2_12_12_1$	2.5	19.3
Xuan & Weber 1992/bdlb40	CGC**I**AATT**C**GCG/CGC**I**AATT**C**GCG	**B**	$P2_12_12_1$	2.4	17.4
Yoon *et al.* 1988/DODBAF	CGCATATATGCG/CGCATATATGCG	**B**	$P2_12_12_1$	2.2	18.7
(*b*) **Decamers**					
Baikalov *et al.* 1993/bdjb48	CGATCG*A*TCG/CGATCG*A*TCG	**B**	$P3_221$	2.0	17.2
Egli *et al.* 1992/ahj040	GGGTATACGC/GCGTATACCC	**A**	$P2_12_12_1$	2.0	15.6
Egli *et al.* 1993/ahj043	*G*CGTATACG*C*/*G*CGTATACG*C*	**A**	$P2_12_12_1$	2.25	17.8
Egli *et al.* 1993/ahj044	GCGT*A*TACGC/GCGT*A*TACGC	**A**	$P2_12_12_1$	2.0	19.5
Frederick *et al.* 1989/adj022†	ACCGGCCGGT/ACCGGCCGGT	**A**	$P6_122$	2.0	18.0
Goodsell *et al.* 1993/bdj051	CATGGCCATG/CATGGCCATG	**B**	$P2_12_12_1$	2.0	19.6
Grzeskowiak *et al.* 1991/bdj025	CGATCGATCG/CGATCGATCG	**B**	$P2_12_12_1$	1.5	16.1
Heinemann & Alings 1989/bdj017†	CCAGGCCTGG/CCAGGCCTGG	**B**	$C2$	1.6	16.9
Heinemann *et al.* 1992/bdj039	CCGGCGCCGG/CCGGCGCCGG	**B**	$R3$	2.2	16.7
Heinemann & Hahn 1992/bdjb27	CCAGGC*C*TGG/CCAGGC*C*TGG	**B**	$P6$	1.75	17.4
Lipanov *et al.* 1993/bdjb43	CCAA**CI**TTGG/CCAA**CI**TTGG	**B**	$P3_221$	2.2	16.2
Lipanov *et al.* 1993/bdjb44†	CCAA**CI**TTGG/CCAA**CI**TTGG	**B**	$C2$	1.3	15.2
Lipanov *et al.* (‡)	CCAA**IA**TTGG/CCAA**IA**TTGG	**B**	$C2$	2.0	13.4
Privé *et al.* 1987/bdj008†	CCAA**GA**TTGG/CCAA**GA**TTGG	**B**	$C2$	1.3	17.8
Privé *et al.* 1991/bdj019†	CCAACGTTGG/CCAACGTTGG	**B**	$C2$	1.4	16.0
Quintana *et al.* 1992/bdj031	CGATTAATCG/CGATTAATCG	**B**	$P2_12_12_1$	1.5	15.7
Wang *et al.* 1982/ahj015	*GCG*TATACGC/*GCG*TATACGC	**A**	$P2_12_12_1$	2.0	16.0
Yuan *et al.* 1992/bdj036	CGATATATCG/CGATATATCG	**B**	$P2_12_12_1$	1.7	17.8

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Table 1. *Cont.*

reference/code	sequence	type	space group	res. (Å)	R (%)
(*c*) **Octamers**					
Cervi *et al.* 1992/adh041†	GTCTAGAC/GTCTAGAC	**A**	$P4_32_12$	2.5	21.1
Cruse *et al.* 1989/adhb17	GG**I**GC**T**CC/GG**I**GC**T**CC	**A**	$P6_1$	1.7	14.0
Eisenstein *et al.* 1990/OCTAAR	GGGTACCC/GGGTACCC	**A**	$P6_1$	2.4	11.9
Haran *et al.* 1987/OCTAAA†	CCCCGGGG/CCCCGGGG	**A**	$P4_32_12$	2.25	16.0
Heinemann *et al.* 1987/adh008†	GCCCGGGC/GCCCGGGC	**A**	$P4_32_12$	1.8	17.1
Heinemann *et al.* 1991/adhp36†	GCCCGGGC/GCCCGGGC	**A**	$P4_32_12$	2.12	16.0
Hunter *et al.* 1986/OCTAAJ	GG**G**GC**T**CC/GG**G**GC**T**CC	**A**	$P6_1$	2.25	13.6
Hunter *et al.* 1989/adh020†	CTCTAGAG/CTCTAGAG	**A**	$P4_32_12$	2.15	14.7
Jain *et al.* 1989/adh014†	GTGTACAC/GTGTACAC	**A**	$P4_32_12$	2.0	11.5
Kennard *et al.* 1986/adhb11	GG**UAUA**CC/GG**UAUA**CC	**A**	$P6_1$	1.7	14.0
Kneale *et al.* 1985/OCTAAK	GGG**GT**CCC/GGG**GT**CCC	**A**	$P6_1$	2.1	14.0
Lauble *et al.* 1988/OCTAAG	GGGATCCC/GGGATCCC	**A**	$P6_1$	2.5	17.2
McCall *et al.* 1985/OCTAAI	GGGGCCCC/GGGGCCCC	**A**	$P6_1$	2.5	20.0
Shakked *et al.* 1989/OCTAAH†	GGGCGCCC/GGGCGCCC	**A**	$P4_32_12$	1.7	16.2
Takusagawa 1990/adh024†	GTACGTAC/GTACGTAC	**A**	$P4_32_12$	2.25	18.4
Thota *et al.* 1993/adh038†	GTGTACAC/GTGTACAC	**A**	$P6_122$	1.4	19.8
(*d*) **Tetramers**					
Conner *et al.* 1984/addb01	CCGG/CCGG	**A**	$P2_1$	2.0	19.9

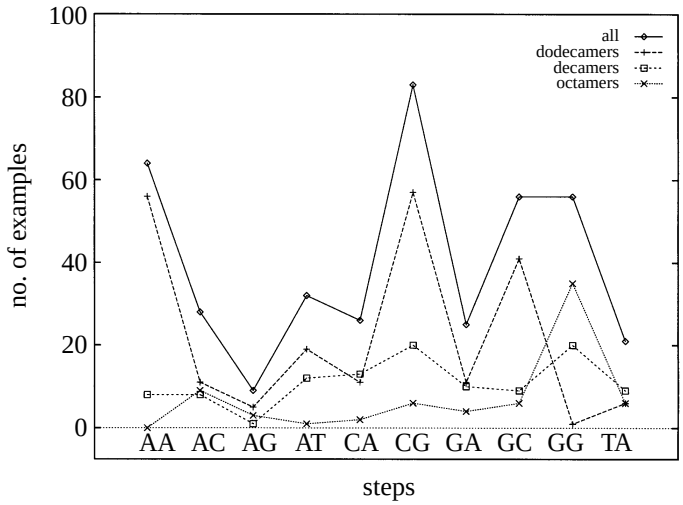

Figure 2. The distribution of the numbers of available examples of all types of step in naked DNA oligomers. There are two GG/CC and one CG step that come from a 4mer; these are not included in this plot. The step types are laid out here in alphabetic order.

structures, or solved the same structures at different temperatures or by the use of different heavy metal derivatives while still ending up with structures that are identical both with respect to structure and sequence. In general, whenever there was a choice, the oligomer with the better resolution and *R*-factor was chosen. Finally, we should note that, in all cases where the asymmetric unit consisted of only one strand, only the independent half was, in effect, included.

(*b*) *Sequence analysis*

Before we begin to make statements regarding the conformational characteristics of the 10 types of step, we need to examine carefully the distribution of the available data among these types of step. Now, since we are concerned with steps that involve only Watson–Crick base pairs, we need to eliminate all steps that are directly involved in a base-pair mismatch and/or modification. However, unlike other workers in the field (e.g. Gorin *et al.* 1995), we shall not exclude dodecamer-end sequences simply on the grounds that they are involved in more direct crystal-packing effects. Indeed, we shall pay particular attention to any differences that may exist between their conformations and those of similar sequences that occur elsewhere in the oligomers in question.

After the elimination of all unwanted steps from the 60 chosen oligomers, a total of 400 steps remain. The distribution of these examples among the 10 types of step is shown in figure 2. An ideal distribution would give equal representation of all step types. However, as is clear from figure 2, there are wide differences in availability of data on the various step types. Thus while there are 64 examples of AA/TT, 83 examples of CG and 56 examples of GG/CC, there are only 25 examples of GA/TC, 21 examples of TA and 26 examples of CA/TG. The most under-represented step is AG/CT: there are only 9 examples of it; and therefore we shall discuss this step only briefly. Thus for the most part we shall discuss the conformational properties of the *nine* types of dinucleotide step for which there are adequate data.

As we have already explained in §§ 1 and 2, we shall treat data from A-form

oligomers and B-form oligomers equally in the present paper; and unless stated otherwise, all plots, tables, etc., will include data from both A-form and B-form DNA. However, it might be useful at this stage to state how are the data of figure 2 divided among A-form and B-form oligomers (as given by the authors of the respective papers) (see table 1). It turns out that none of the 64 examples of AA/TT come from A-form oligomers, while very few of the 32 examples of AT and the 25 examples of GA/TC, and 26 examples of CA/TG come from A-form oligomers (only 4 examples in the cases of GA/TC and AT and 2 examples in the case of CA/TG). Of the 28 examples of AC/GT 20 come from A-form oligomers, and of the 21 examples of TA 14 come from A-form oligomers. The GG/CC, GC and CG steps are disaggregated among A-form and B-form oligomers as follows: 40 A-form and 16 B-form in the case of GG/CC; 11 A-form and 45 B-form in the case of GC; and 16 A-form and 67 B-form in the case of CG.

4. Overall behaviour

(*a*) *Step parameters*

Our general approach in this paper is to start with all six of an Eulerian set of geometrical step parameters as unknown variables, and then to look for evidence from the database that warrants the elimination of some of these parameters on account of their insignificance. Here an insignificant parameter is one that shows very little variation in a clearly *sequence-independent* manner. Figure 3 shows histograms or 'frequency diagrams' of the step parameters: helical twist, roll, tilt, rise, slide and shift, for all 400 steps taken together. It is immediately obvious that whereas helical twist, roll and slide have flat broad-peaked frequency plots, tilt, shift and rise all have narrow frequency plots with very sharp peaks (typically greater than 0.7) at 0°, 0 Å and 3.4 Å respectively. Hence roll, slide and twist are clearly significant as variables while tilt, rise and shift are not. Although figure 3 indicates clear qualitative differences between roll, slide and twist on the one hand and shift, tilt and rise on the other, it is important to examine such frequency diagrams for each step type individually so as to eliminate any bias due to the uneven representation of the dinucleotide step types that we have already described above. This has indeed been done (plots not shown) and it appears that the frequency diagrams for the individual step types all reflect the same features seen in the overall plot of figure 3, i.e. little or no variation in shift, tilt and rise but some variation in the other parameters. The only special feature that emerged was the higher than usual values of shift (greater than 0.5 Å) that are sometimes adopted by GC and AC/GT.

To conclude: apart from a few cases where shift assumes some significance, the conformational space of any dinucleotide step may be taken to be more or less completely defined by the three parameters: roll, slide and helical twist.

(*b*) *Base-pair parameters*

In order to see whether some base-pair parameters can similarly be discarded on account of their invariance, we have constructed frequency diagrams for propeller, buckle, opening, shear, stagger and stretch for all base pairs taken together (figure 4). Due to their sign ambiguity (El Hassan & Calladine 1995), only the absolute values of buckle and shear have been plotted. It is at once apparent from the plots that whereas all the translational parameters, i.e. shear, stagger and stretch show

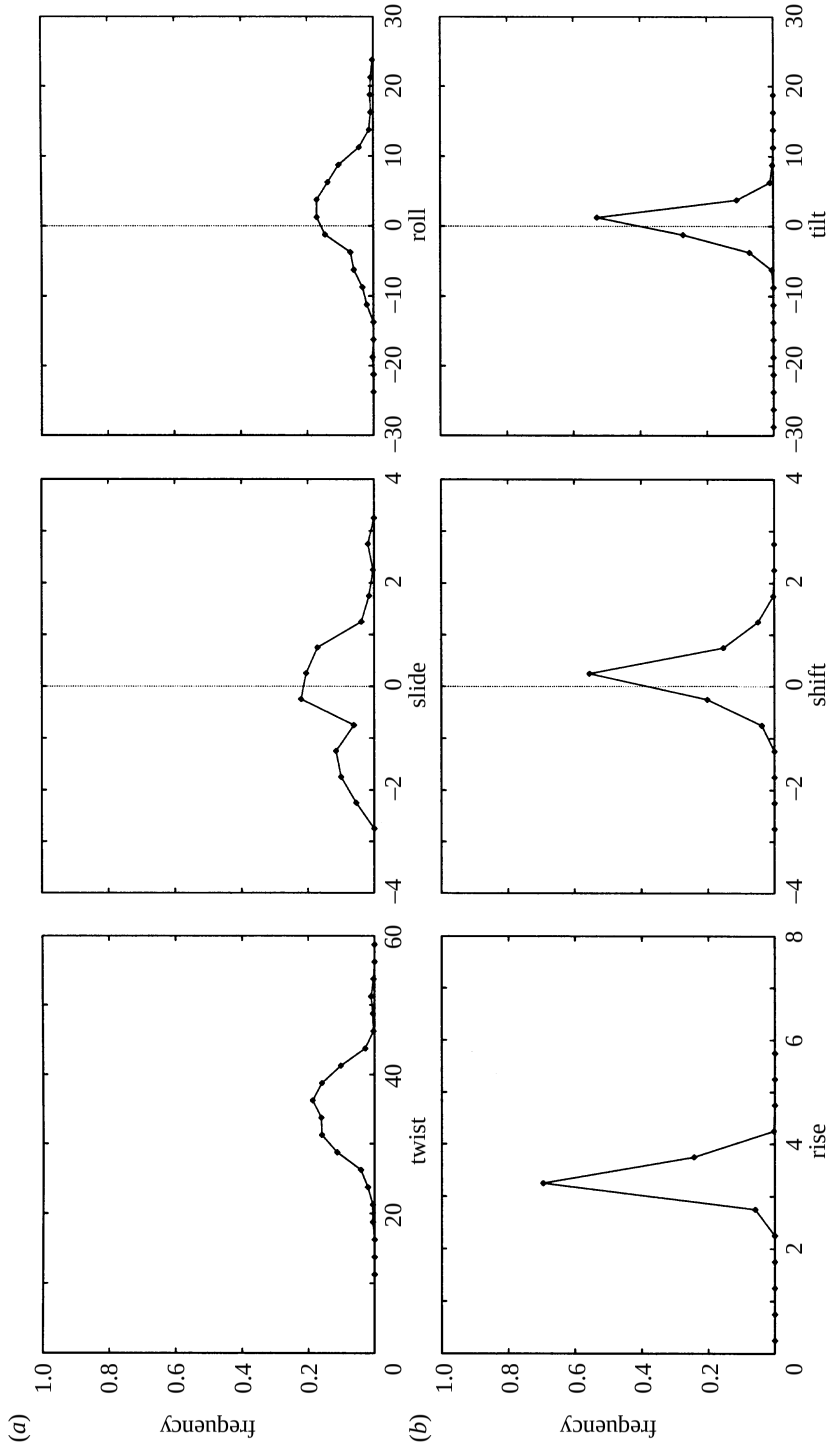

Figure 3. 'Frequency diagram' for roll, slide, helical twist, tilt, shift and rise of all steps taken together. A frequency diagram is essentially a histogram with 'bars' replaced by a line joining the mid points of the peaks of the bars. The total ranges that were used are: −3 to 3 Å for slide and shift, 0 to 6 Å for rise, −25° to 25° for roll and tilt, and 10° to 60° for helical twist. Twelve increments have been used for all translational parameters (rise, slide and shift), and 20 increments for all angular parameters (helical twist, roll and tilt). Note that, by definition, the area underneath a frequency plot is equal to unity.

very little variation, with very strong peaks at 0 Å, 0 Å and 5.5 Å respectively, the rotational parameters, particularly propeller and buckle, show considerable variation with very broad flat-peaked frequency diagrams. It is therefore clear that we

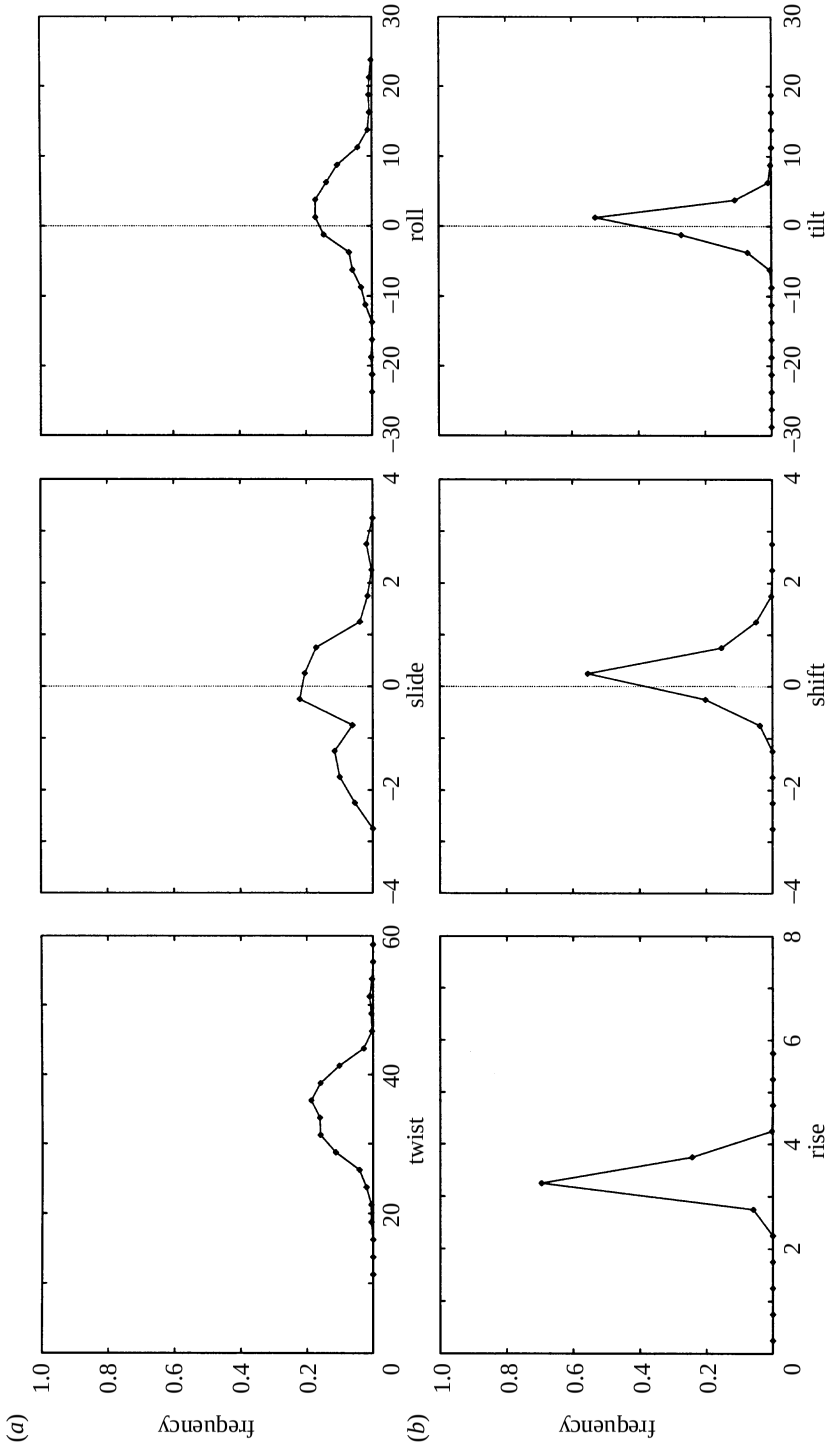

Figure 4. Frequency diagram for propeller, |buckle|, opening stretch, |shear| and stagger of all base pairs taken together. The frequency diagrams are constructed in a way similar to that of figure 3. The total ranges of the parameters are: −35° to 15° for propeller, −20° to 30° for opening, 0° to 50° for buckle, −3 to 3 Å for stagger, 0 to 6 Å for shear and 2 to 8 Å for stretch. Twelve increments have been used for all translational parameters, and 20 increments for all angular parameters.

may take all *translational* parameters as insignificant and concentrate only on the *rotational* parameters, i.e. propeller, buckle and opening.

5. The roll/slide/twist conformational space; behaviour of dinucleotides

(*a*) *Overall behaviour*

Figure 5 shows roll/slide, roll/twist and twist/slide plots for all 400 available steps taken together. The range of values is −2.3 Å to 2.8 Å for slide, −21° to 25° for roll and 20° to 54° for helical twist. Within these ranges, the plots of figure 5 are very fuzzy; but we can still see a very broad positive correlation between twist and slide, a negative correlation between roll and twist, and a resulting negative correlation between roll and slide. The overall means (and standard deviations) of helical twist, roll and slide are approximately 35°(5°), 2.5°(7°) and 0.0 Å(1.0 Å) respectively. These averages and standard deviations, which we shall refer to hereafter as the *global* values, will provide good landmarks with which to compare the corresponding averages and standard deviations of individual step types.

We have mentioned in § 1 the inherent vagueness of the terms 'A-form DNA' and 'B-form DNA' as used by many workers in the field. Now figure 5 shows roll, slide and twist data on all steps from oligomers that have been designated 'A-DNA' and oligomers designated 'B-DNA' by the authors of the respective reports. The data from the different oligomer classes have been plotted using different symbols. It is clear from the plots that whereas there is a slight discontinuity of data in slide, neither roll nor twist show any discontinuity at all. Indeed whereas A-form steps, in general, occupy the low-slide, low-twist, high-roll end of the figures, they generally form, with B-form steps, a continuum of conformations that are available for right-handed DNA. Moreover, there is no evidence to exclude all possibility of the existence of these oligomeric A-form conformations in solution. Thus Biburger *et al.* (1995) have demonstrated that oligo[d(C).(G)] runs in solution exhibit a helical repeat of just over 11, which would correspond to helical twists values similar to those seen in 'A-form' oligomers. We therefore conclude that A-DNA is equally important as B-DNA; and we shall therefore include both sets of data in our analysis. We shall however, where appropriate, take note of whether the oligomer concerned is in the A-form or B-form; and we shall describe, in § 9, explicitly an indicator that is defined with respect to the backbone, and which classifies a given dinucleotide step unambiguously as an A-form or a B-form step.

(*b*) *Behaviour of individual steps*

(i) *Purine–purine (RR/YY) steps*

AA/TT Step. Figure 6*a* shows roll/slide, roll/twist and twist/slide diagrams for all available examples of AA/TT (= 64 examples). (The 'boxes' in the roll/slide plots of figures 6–8 will be explained and discussed in § 8.) The averages (and standard deviations) of twist, roll and slide for this step and all other steps are given in table 2. For this particular step, these values are 35.9°(3.3°), 1.25°(3.6°) and −0.16 Å(0.3 Å) respectively.

As is evident from the plots and the values of the means and standard deviations of roll, slide and twist, most of the AA/TT steps cluster very tightly around a classical 'B' conformation of zero slide and roll and canonical 36° helical twist. The standard deviations are much smaller than those for the global set of values; and, as is evident from table 2, they are indeed smaller than those of any other step except for the standard deviation of roll of GG/CC. But despite the tightly clustered conformation

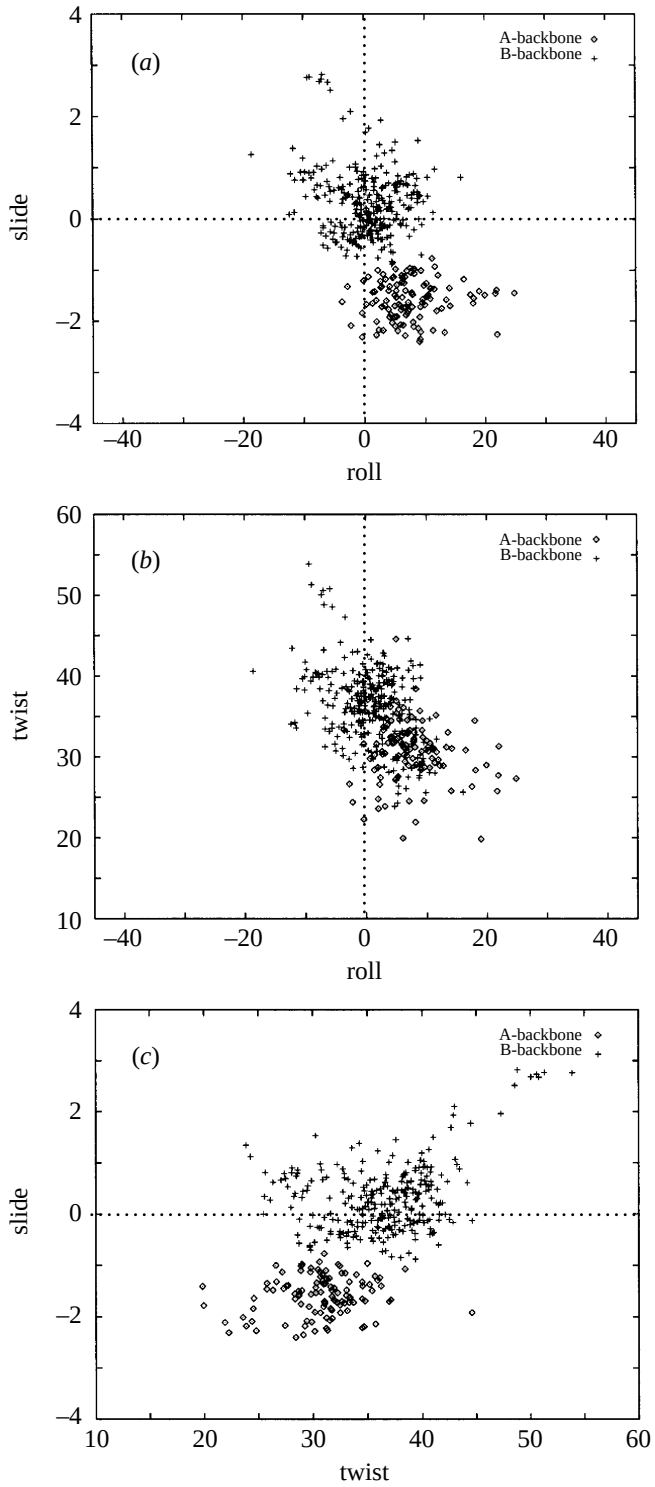

Figure 5. Roll/slide, roll/twist and twist/slide scatter plots for all steps taken together. Steps from A-backbone oligomers are marked differently from those from B-backbone oligomers.

Table 2. *Averages (and standard deviations) of the leading step parameters, helical twist, roll and slide*

step	$\overline{\text{helical twist}^\circ}$	$\overline{\text{roll}^\circ}$	$\overline{\text{slide}}$ (Å)
AA/TT	35.9(3.3)	1.25(3.6)	−0.16(0.28)
AC/GT	32.9(3.8)	3.3(4.5)	−0.89(0.71)
high-slide (scattered)	33.4(4.5)	0.3(4.7)	0.08(0.47)
low-slide (clustered)	32.7(3.6)	4.5(4.0)	−1.28(0.30)
AT	32.4(2.8)	−2.4(4.4)	−0.44(0.48)
CA/TG	37.4(9.5)	2.0(6.9)	1.18(1.23)
CG	35.1(5.3)	3.1(5.4)	0.00(1.02)
high-slide	36.8(4.1)	2.1(4.3)	0.45(0.46)
low-slide	28.4(4.7)	7.3(7.5)	−1.86(0.36)
GA/TC	37.8(3.8)	4.4(3.8)	−0.35(0.69)
GC	37.4(4.0)	−2.5(6.8)	0.25(0.86)
high-slide	38.9(2.6)	−4.7(5.4)	0.63(0.39)
low-slide	31.1(2.7)	6.5(3.5)	−1.29(0.35)
GG/CC	31.9(3.7)	6.3(2.7)	−1.06(1.17)
high-slide	32.3(5.6)	5.5(2.7)	0.71(0.34)
low-slide	31.8(2.6)	6.5(2.7)	−1.77(0.32)
TA	30.6(6.7)	11.8(7.1)	−0.80(1.09)
ALL	34.8(5.2)	2.5(6.2)	−0.24(1.05)

of AA/TT, we can still see a somewhat wide range (*ca.* 15°) in helical twist. This range is primarily due to the following few examples:

(i) C-G-C-A-A-A-A-A-T-G-C-G. Helical twist of underlined AA/TT = 29.7°.
(ii) C-C-A-A-C-G-T-T-G-G. Helical twist of underlined AA/TT = 28.4°.
(iii) C-C-A-A-G-A-T-T-G-G . Helical twist of underlined AA/TT = 25.6°.
(iv) C-C-A-A-I-A-T-T-G-G. Helical twist of underlined AA/TT = 25.6°.

The most unusual values of helical twists can be seen to belong to the AA/TT steps from oligomers (iii) and (iv), in both of which the step is flanked by the mispairs G–A and I–A respectively. Oligomer 1 has in fact a relatively low resolution (2.6 Å). Oligomer 2 has neither a poor resolution/*R*-factor (1.4 Å/16%) nor a mispair. However, the AA/TT step in this oligomer is flanked by a CA/TG with a very high helical twist (*ca.* 51°). The apparent undertwisting of the AA/TT step might well be forced onto the step in order to immediately compensate for the severe overtwisting of the flanking CA/TG. As we shall explain below (§ 11), it is highly significant that the propeller values in these particular examples of AA/TT are in fact well below the average value for AA/TT in general.

We can therefore conclude that AA/TT in general has a very well defined *single* conformation, characteristic of the classical B-DNA form. However, the step appears to be able to adopt slight variants of the observed average conformations in certain extreme cases.

GA/TC Step. Figure 6*b* shows the roll/slide/twist characteristics for the GA/TC step. It can be seen that this step is not very different from AA/TT. There is per-

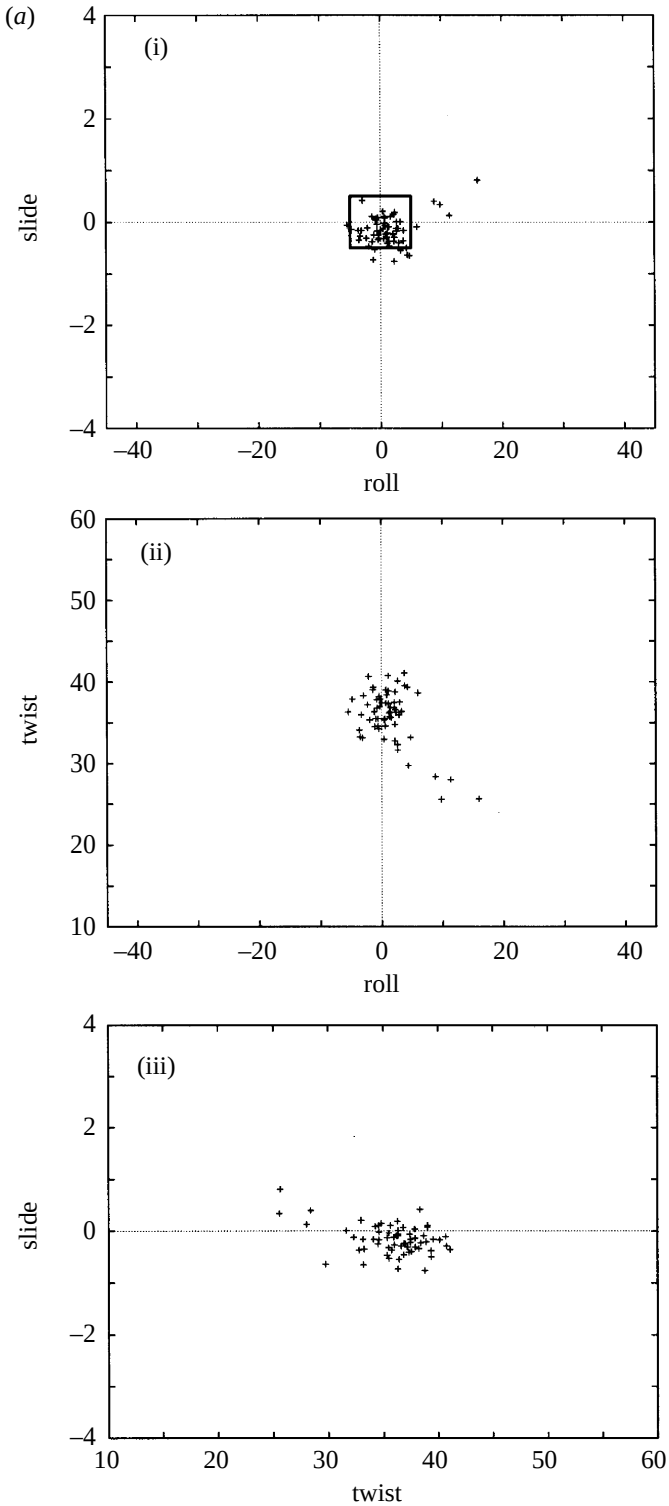

Figure 6. Roll/slide, roll/twist and twist/slide scatter plots for **RR/YY** steps: (*a*) AA/TT. The boxes indicate the roll/slide region predicted by the electrostatic analysis of Hunter (1993).

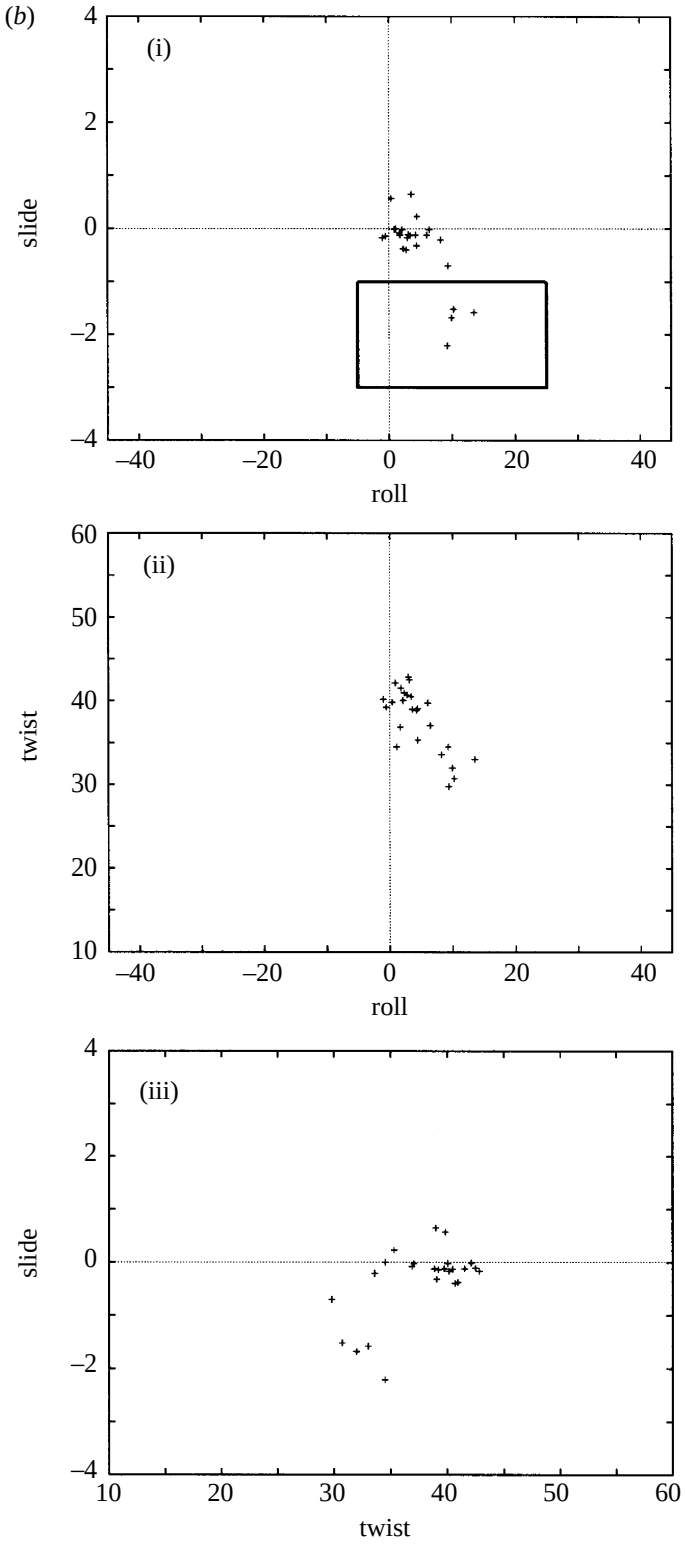

Figure 6. *Cont.* (*b*) GA/TC.

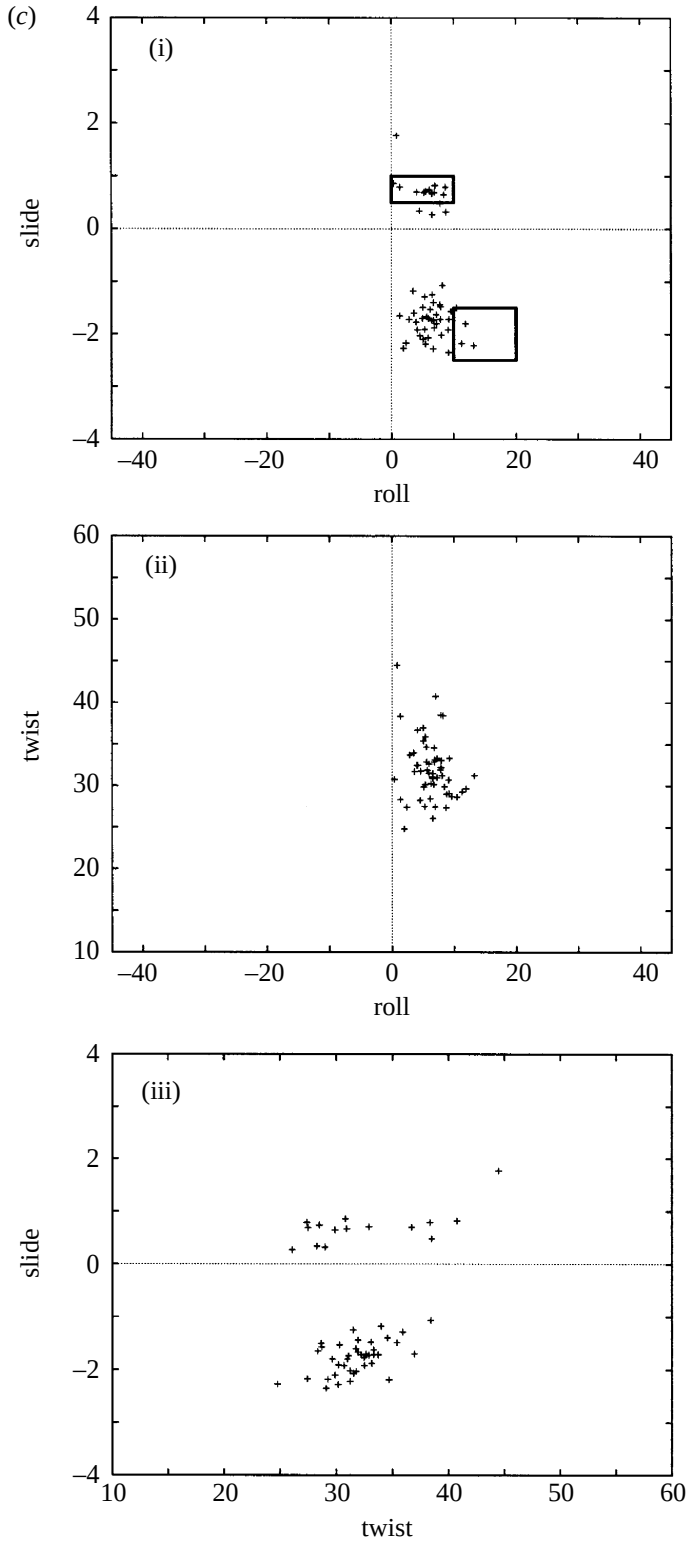

Figure 6. *Cont.* (*c*) GG/CC.

Table 3. *Roll, slide and helical twist values of all available examples of AG/CT steps*
(The data have been separated into low-slide and high-slide classes.)

oligomer-code	sequence	helical twist°	roll°	slide (Å)
		high-slide		
DODBAA	CGCAAAAA**AG**CG	33.50	4.29	0.85
bdj017	CC**AG**GCCTGG	23.73	5.40	0.90
bdl022	CGCA**AG**CTGGCG	30.92	15.71	0.32
bdl022	CGCAAG**CT**GGCG	37.82	7.78	0.42
bdl042	CGT**AG**ATCTACG	25.25	5.19	0.55
bdl042	CGTAGAT**CT**ACG	37.00	2.66	0.11
mean (standard deviation)		31.4(5.9)	6.8(4.7)	0.52(0.31)
		low-slide		
adh020	**CT**CTAGAG	32.42	7.11	−1.49
adh020	CT**CT**AGAG	40.23	5.05	−1.40
adh041	GT**CT**AGAC	35.21	1.48	−1.48
mean (standard deviation)		36.0(4.0)	4.6(2.9)	−1.46(0.05)

haps more scatter, but the majority of the examples prefer a single conformation which is only slightly different from that of the AA/TT step. The actual averages and standard deviations of the twist, roll and slide are 37.8°(3.8°), 4.4°(3.8°) and −0.3 Å(0.7 Å) respectively; see table 2. These values indicate that the average helical twist of GA/TC steps is only slightly higher than the canonical 36° value, and that it prefers a slightly more positive roll angle and a slightly negative slide. The standard deviations are all lower than the global values. In summary, this step can be classed as another *single*-conformation step that is again more or less in the 'B' form. It is worth noting that the four GA/TC steps that appear to adopt large negative slide values (between −1.5 Å and −2.5 Å) are all very similar and they all come from oligomers whose backbones are in the A-conformation (Calladine & Drew 1984). These steps are: (see table 1 for details of the oligomers)

(i) G-G-G-A-T-C-C-C; OCTAAG (Lauble *et al.* 1988).
(ii) C-T-C-T-A-G-A-G; adh020 (Hunter *et al.* 1989).
(iii) G-T-C-T-A-G-A-C; adh041 (Cervi *et al.* 1992).

Note that in oligomers adh020 and adh041, the asymmetric unit consists of one strand and hence the GA/TC steps in the second half are duplicates of the underlined ones.

If we exclude these steps, the standard deviations of twist, roll and slide reduce to 3.2°, 2.7° and 0.3 Å respectively, while the averages change to 38.8°, 3.2° and −0.1 Å respectively. As can be seen, except perhaps for slide, the averages change very little while the deviations all reduce slightly. Thus, although the majority (21 out of 25 examples) of the data suggest a single conformation for the GA step, there is some evidence to suggest that another conformation is possible.

AG/CT step We have already noted that AG/CT is very poorly represented

in the database. We can nevertheless make some tentative statements regarding the behaviour of the available nine examples. Of these, six come from oligomers whose backbones are in the B-conformation and three from oligomers whose backbones are in the A-conformation. Table 3 gives values of roll, slide and helical twist of all nine examples of AG/CT. There is a conformational bimodality, with all A-form steps adopting low-slide (*ca.* −1.5 Å) conformations and all B-form steps adopting high-slide conformations (*ca.* 0.5 Å). It should be noted that low-slide steps do not adopt correspondingly high roll and low twist as might be expected from the overall roll/slide/twist correlation that we have already discussed (see figure 5). Indeed, on average the high-slide steps adopt higher roll and lower helical twist than the low-slide ones. We should emphasize, however, that due to the lack of data, no firm conclusions can be drawn regarding this step, and any tentative remarks may well need to be modified as more data become available.

GG/CC Step. Figure 6*c* shows the roll/slide/twist behaviour of GG/CC. This step is, obviously, quite different from the AA/TT and GA/TC steps. A distinct bimodal behaviour can be seen, especially with respect to slide: both roll/slide and twist/slide plots show two distinct clusters of points. One mode is characterized by high slide, 0.7 Å(0.3 Å) and the other by low slide, −1.8 Å(0.3 Å). The roll angles are more or less the same, 5.5°(2.7°) and 6.6°(2.6°) for the high- and low-slide modes respectively. The twists are again not too dissimilar for the two modes but the high-slide mode appears to have a more variable helical twist; the means and standard deviations for the high slide and low slide modes are 32.3°(5.6°) and 31.8°(2.6°) respectively. Table 2 summarizes the means and standard deviations of twist, roll and slide for each mode of GG/CC and for the step taken as a whole; that is why GG/CC and all other bimodal steps (to be described presently) are represented by three rows in the table. The standard deviations of both roll and slide for each mode taken separately, and of the helical twist of the low-slide mode are significantly smaller than the overall deviations, indicating that each mode can be described as a 'single' conformation with the high-slide mode being somewhat mobile with respect to helical twist. It is worth emphasizing the *rigidity* of the GG/CC step with respect to roll. Indeed, the averages and standard deviations for roll of the two modes are almost identical, and the standard deviations are not only smaller than the global value, but are smaller than those of any of the other steps.

(ii) *Purine–pyrimidine (RY/RY) steps*

AT/AT Step. The roll/slide/twist characteristics are plotted in figure 7*a*. Like AA/TT and GA/TC, this step appears to have a very well-clustered single conformation. The means and standard deviations of twist, roll and slide are 32.4°(2.8°), −2.4°(4.4°) and −0.4 Å(0.5 Å) respectively. The standard deviations are still less than the global values, and are indeed comparable to those of AA/TT and GA/TC. This step, however, prefers a significantly low-twist conformation, together with a corresponding negative slide; whereas roll is more or less zero. This tendency for AT to adopt low twist was in fact first noted by Klug *et al.* (1979), who suggested low-twist for AT and high-twist for TA in the alternating 'TATA' context. As with GA/TC, there are four examples that are very similar to each other and yet are rather different from the general trend. They all adopt the A-backbone conformation. These steps are (see table 1 for details of the oligomers)

(i) G-G-G-A-T-C-C-C; OCTAAG (Lauble *et al.* 1988).
(ii) *G-C-G*-T-A-T-A-C-G-C; ahj015 (Wang *et al.* 1982).

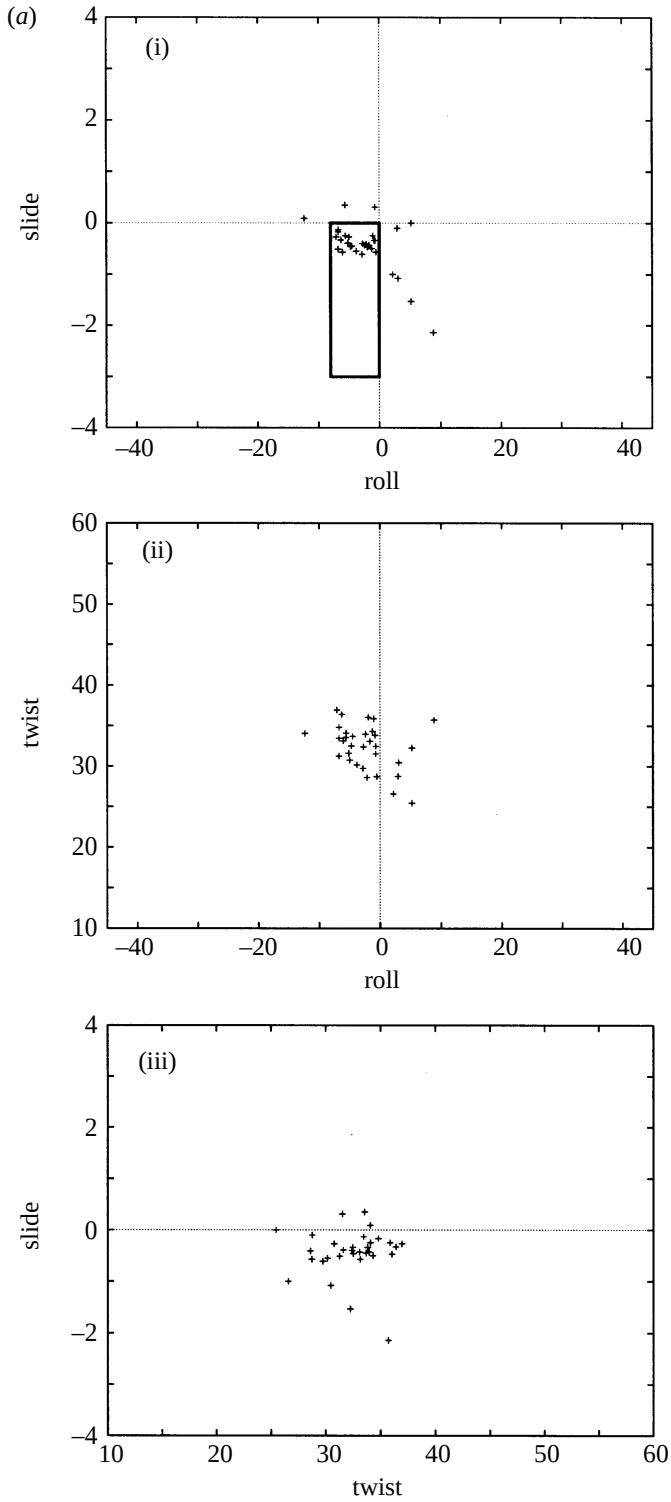

Figure 7. Roll/slide, roll/twist and twist/slide scatter plots for **RY** steps: (*a*) AT. Boxes as in figure 6.

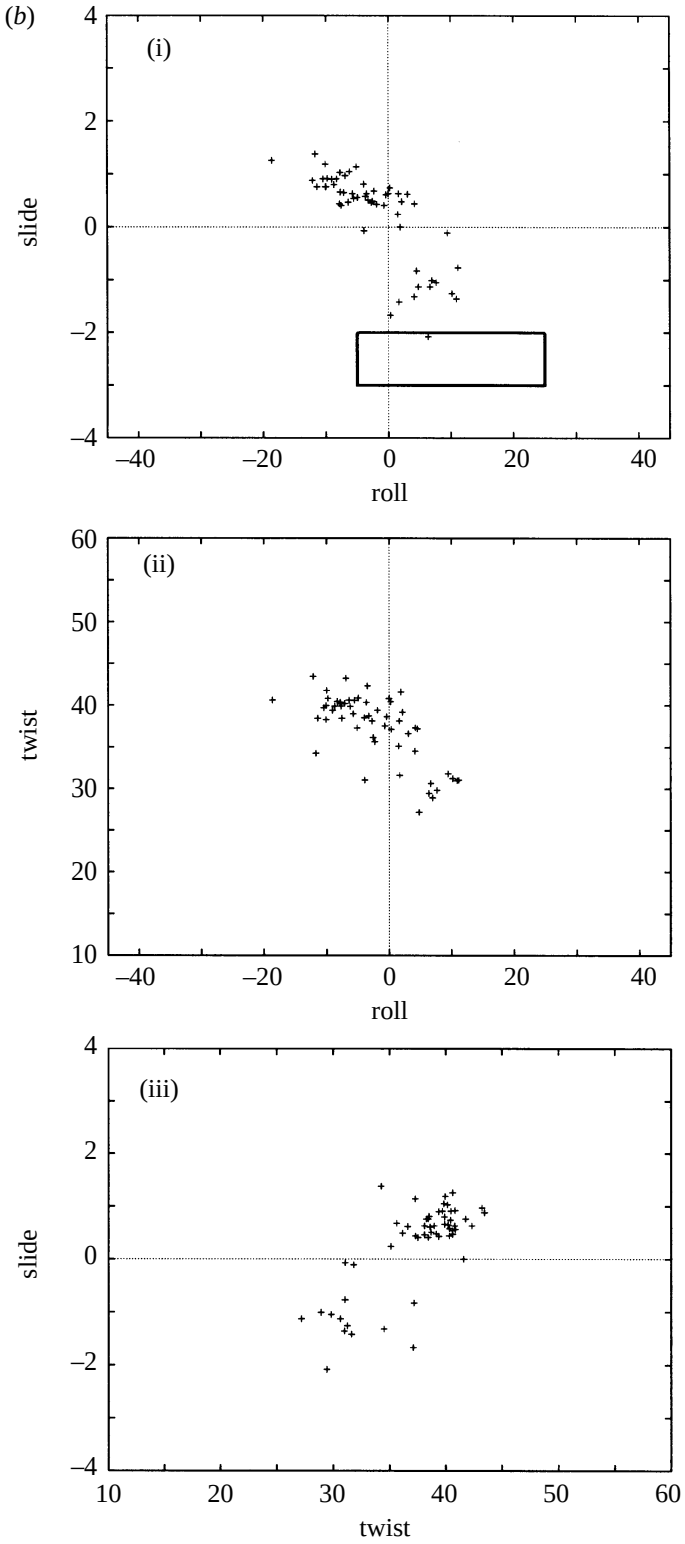

Figure 7. *Cont.* (*b*) GC.

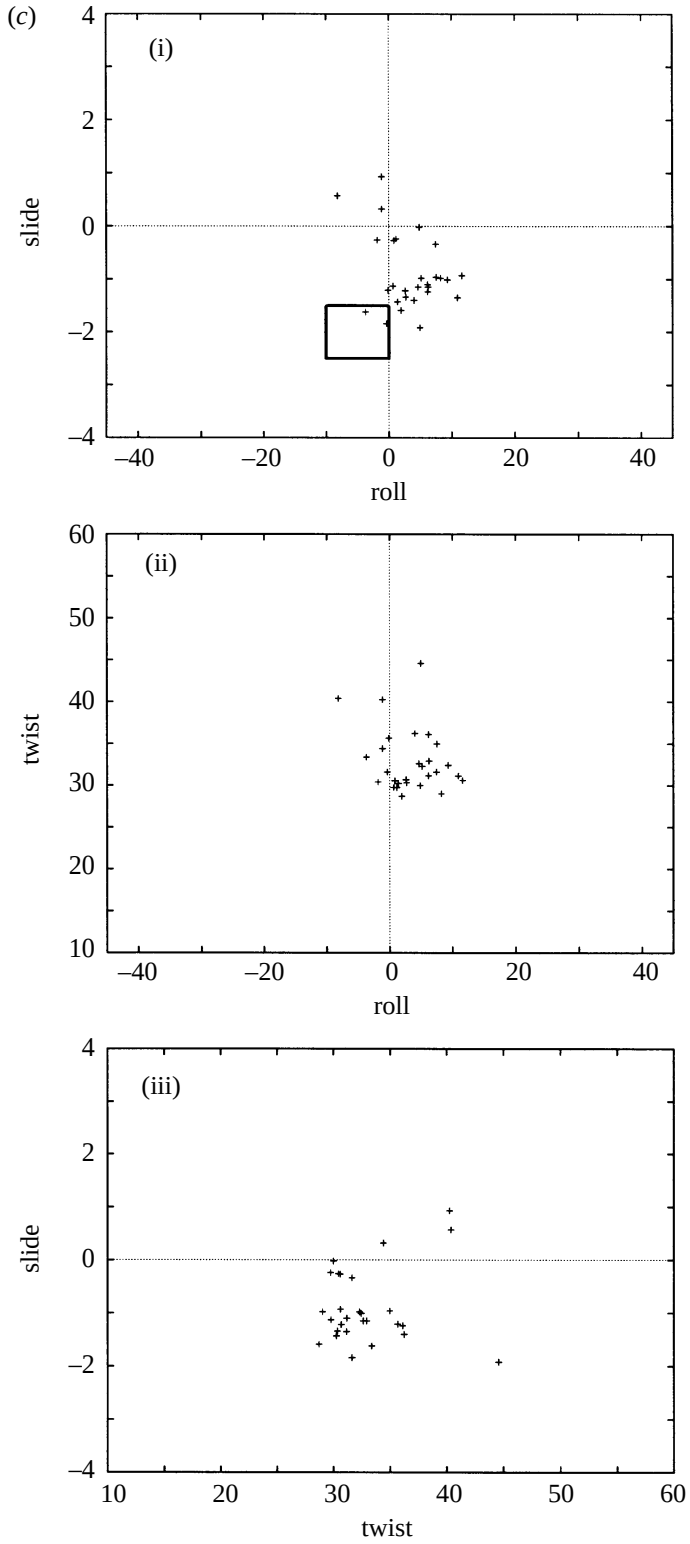

Figure 7. *Cont.* (*c*) AC/GT.

(iii) G-G-G-T-A-T-A-C-G-C; ahj040 (Egli *et al.* 1992).
(iv) *G*-C-G-T-A-T-A-C-G-*C*; ahj043 (Egli *et al.* 1993).

If we exclude these four steps, the standard deviations of twist, roll and slide reduce to 2.7°, 3.5° and 0.25 Å respectively, while the averages change to 32.5°, −3.5° and −0.30 Å respectively. Thus, as in GA/TC, the averages change a little and the deviations all decrease slightly. We can therefore conclude that whereas the majority of AT examples suggest a 'single-conformation' step, there is evidence that another low-slide conformation is possible. We should note, however, that these low-slide steps, in both AT and GA/TC, are in a clear minority, and hence they do not display as clear a bistability as can be seen in, say, GG/CC.

GC/GC Step. Figure 7*b* gives the roll/slide/twist characteristics of this step. A significant degree of bistability can be seen, mainly with respect to slide. The two modes are broadly similar to those of GG/CC, but they are less clearly divided. Also, the roll angles of the two modes are distinct with the high-slide mode having a highly variable roll (standard deviation = 5.4°). Table 2 summarizes the means and standard deviations of the two modes and of the step in totality. Except for the roll of the high-slide mode, all other parameters have standard deviations that appear to be small enough in comparison to the overall deviations to warrant classifying each mode as a 'single' conformation.

AC/GT Step. The roll/slide/twist characteristics of this step are shown in figure 7*c*. This is perhaps the least clear of all steps. The majority of individual examples of AC/GT appear to prefer a low slide conformation (*ca.* −1.5 Å). There are eight significantly more scattered examples that lie outside the strong low-slide cluster. However, even among these scattered points, there is a tendency for low-slide/high roll/low helical twist conformations. We may therefore conclude that AC/GT is a bistable step with one well defined low-slide cluster and another less clear one. All low-slide examples come from 'A-form' oligomers while those in the scattered cluster come from 'B-form' oligomers. Table 2 summarizes the means and standard deviations of the leading step parameters, roll, slide and twist for all AC/GT examples, together with those of each subset. It is clear from the standard deviations that the A-backbone gives a closer clustering of points than does the B-backbone. As far as average conformations are concerned, there is little difference between the two modes with respect to helical twist. However, the two modes do exhibit significant differences with respect to both roll and slide.

(iii) *Pyrimidine–purine (YR/YR) steps*

CG/CG Step. The roll/slide/twist characteristics of this step are shown in figure 8*a*. There is a clear bistability, mainly with respect to slide. One mode prefers positive slide (*ca.* 0.5 Å), and the other prefers negative slide (*ca.* −1.9 Å). Although the averages of the rolls and twists of the two clusters are distinct, their standard deviations are high enough to make them almost indistinguishable; and this is reflected in the more or less single cluster that can be seen in the roll/twist plot. Table 2 summarizes the means and standard deviations of the helical twist, roll and slide for each cluster and for the step in totality. We should point out here that although two strong clusters (with respect to slide) can be seen, there are three points (slide = −0.6, −0.72 and −0.85 Å) that appear to lie between the two regions. These points correspond to the first step in each of the three independent oligomers in the asymmetric unit of the dodecamer d(C-G-T-T-T-T-T-T-C-G-C-G) by diGabrielle &

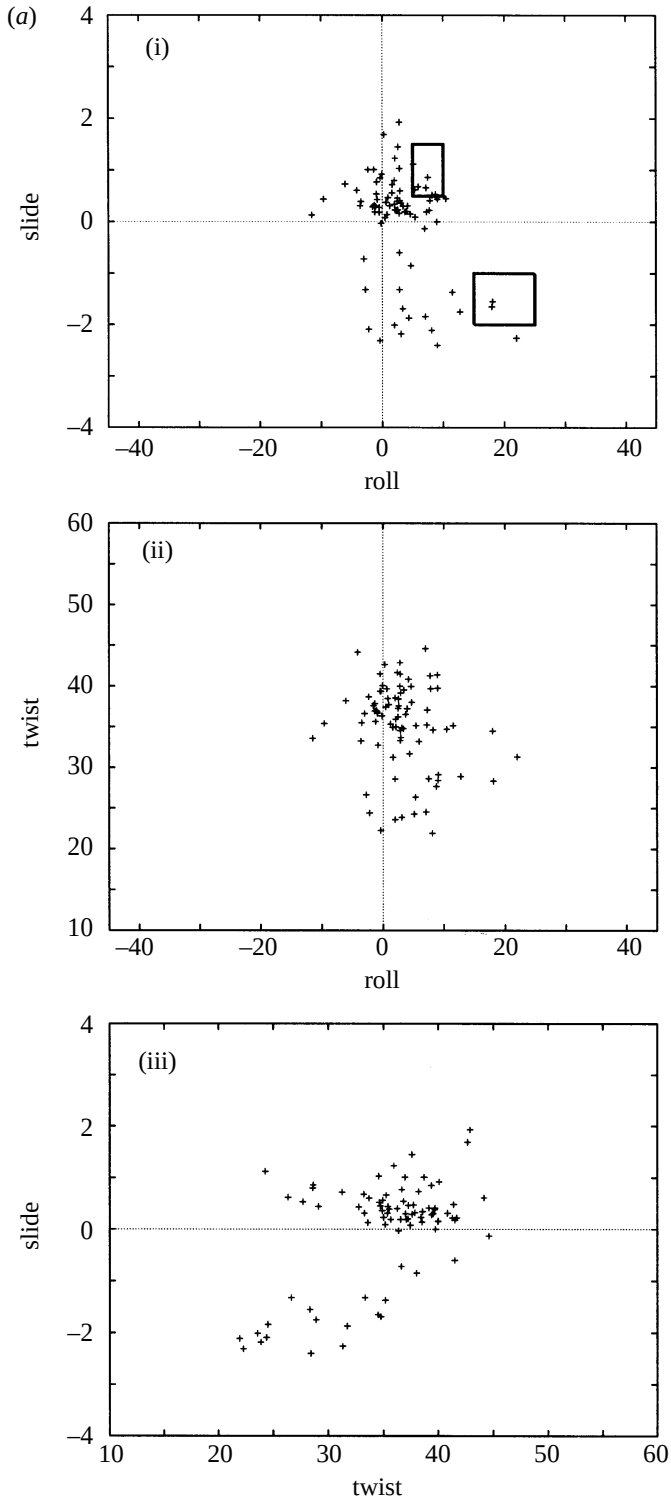

Figure 8. Roll/slide, roll/twist and twist/slide scatter plots for **YR** steps: (*a*) CG. Also shown are best-fit lines in the cases of CA/TG and TA. Boxes as in figure 6.

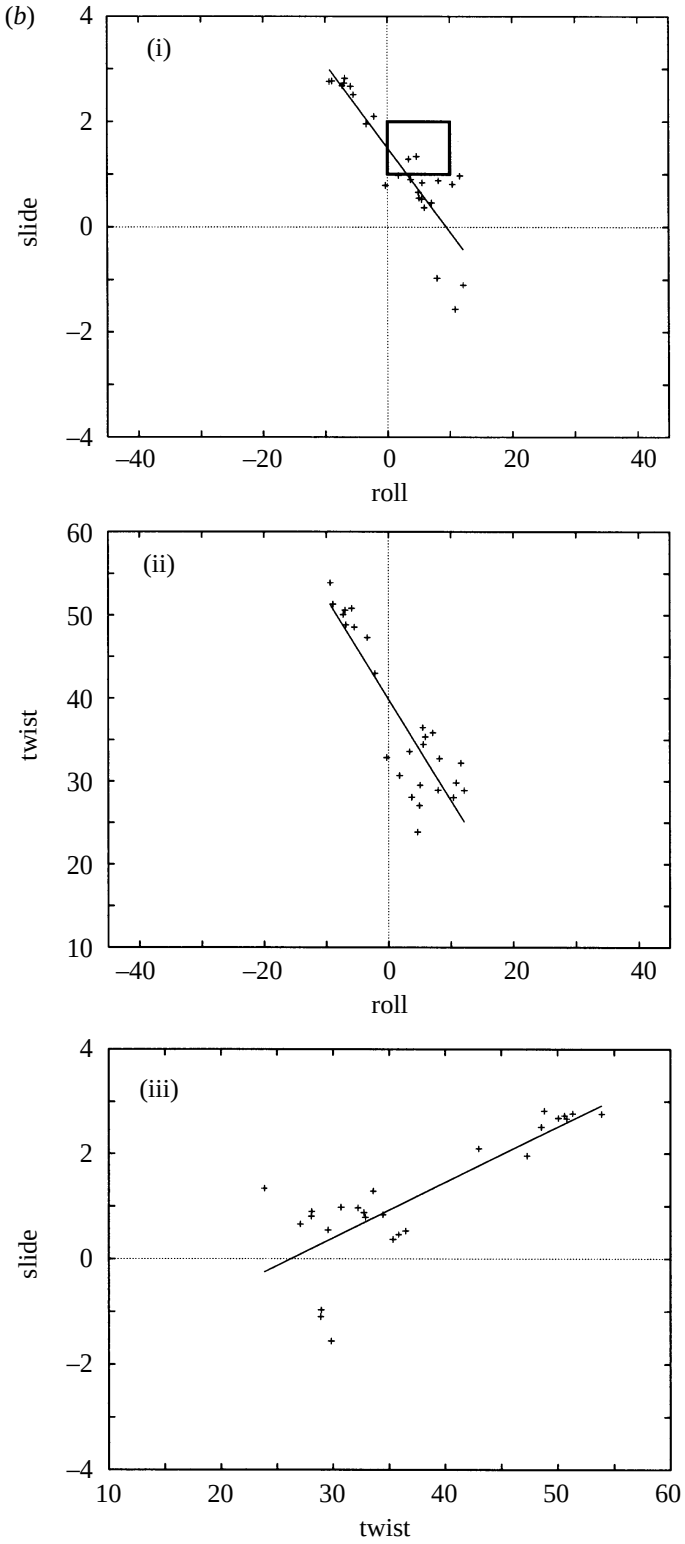

Figure 8. *Cont.* (*b*) CA/TG.

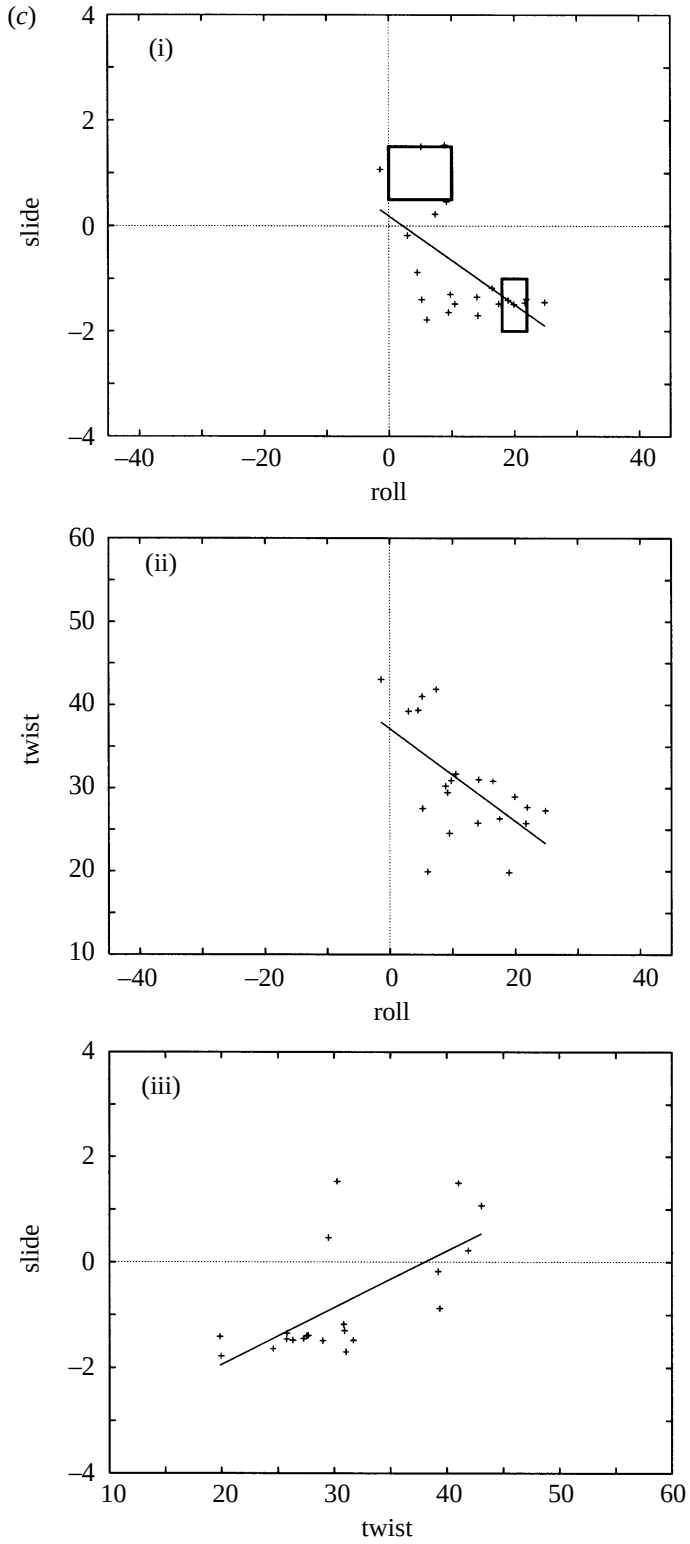

Figure 8. *Cont.* (*c*) TA.

Steitz (1993); (bdl047/1/2/3 (see table 1)). However, these intermediate points are too few to warrant a description of this step as a continuously mobile one.

CA/TG Step. Figure 8*b* shows the roll/slide/twist characteristics of this step. Several striking features are immediately obvious. First, the standard deviations of all parameters are the highest of all the steps; and they are also higher than the global ones: $\mathrm{sd}(\Omega(\mathrm{CA/TG})) \approx 9.5°(> \mathrm{sd}(\Omega(\mathrm{ALL})) \approx 5.2°)$, $\mathrm{sd}(\rho(\mathrm{CA/TG})) \approx 6.9°(> \mathrm{sd}(\rho(\mathrm{ALL})) \approx 6.2°)$ and $\mathrm{sd}(D_y(\mathrm{CA/TG})) \approx 1.2$ Å$(> \mathrm{sd}(D_y(\mathrm{ALL})) \approx 1.0$ Å). Second, these deviations do not reflect a bimodal behaviour but a continuously mobile one. Third, all parameters are linearly correlated, with a positive correlation between twist and slide and negative ones between twist and roll and roll and slide. In summary, this step appears to constitute a more or less continuous 'single-degree-of-freedom' kinematic mechanism. With slide measured in angströms and roll and twist in degrees, the correlations are:

$$D_y(\mathrm{CA}) = -0.16\rho(\mathrm{CA}) + 1.50 \qquad (R = -0.88), \tag{5.1}$$

$$\Omega(\mathrm{CA}) = -1.22\rho(\mathrm{CA}) + 39.85 \qquad (R = -0.88), \tag{5.2}$$

$$D_y(\mathrm{CA}) = 0.11\Omega(\mathrm{CA}) - 2.77 \qquad (R = 0.81). \tag{5.3}$$

Note that the values of the correlation coefficient R presented above, and in correlations throughout the paper, are computed from the standard relationship:

$$R = \frac{\mathrm{COV}(X,Y)}{\sqrt{\mathrm{VAR}(X)\,\mathrm{VAR}(Y)}} = \frac{\sum_1^N (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_1^N (X_i - \overline{X})^2 \ \sum_1^N (Y_i - \overline{Y})^2}.$$

Here, R is the correlation coefficient of the fit ($Y = mX + c$). COV and VAR are the covariance and variance respectively. N is the number of pairs of (X, Y) values, and $\overline{X}$ and $\overline{Y}$ are the averages of X_i and Y_i respectively.

The conformational flexibility of the CA/TG step has been noted by many workers. Thus the sequence-positioning data of Satchwell *et al.* (1986) indicated that CA/TG is the only dinucleotide step that has a preference for lying with its minor groove facing *either* the inside *or* the outside of a circle. diGabrielle *et al.* (1989) also noted that the CA/TG step in the dodecamer d(CGCAAAAATGCG) is capable of bending in two opposite directions. Moreover CA/TG and TA (see below) appear frequently in DNA sequences bound to protein, and particularly in the occasional site of severe kinking and untwisting as in CAP (Schultz *et al.* 1991) and the TATA-box (Y. Kim *et al.* 1993; J. L. Kim *et al.* 1993). Indeed, on the basis of such lines of evidence, Travers (1991, 1995) pointed out the inherent conformational bimodality of CA/TG. But note that whereas we propose a continuously flexible CA/TG step based on figure 8*b*, and in particular the roll-slide plot of the figure, the data show a certain extent of bimodality, particularly in the twist-slide and the roll-twist plots. This 'bimodality' is mainly with respect to helical-twist, which appears to be less continuous than either roll or slide. However it is obvious that there is a very clear difference between CA/TG and (say) GG/CC, which is much more strongly bimodal, with two clear 'single-conformation' clusters. Thus we might perhaps classify CA/TG as a weakly bistable step with two very broad clusters. We prefer to classify CA/TG as a *continuously flexible* step in order to emphasize the difference between the roll/slide/twist pictures of CA/TG and (say) GG/CC.

TA/TA Step. Figure 8*c* shows the roll/slide/twist characteristics of this step. Some similarity can be seen between this and the CA/TG step. Like CA/TG, this

step exhibits large standard deviations: 6.7°, 7.1° and 1.1 Å for helical twist, roll and slide respectively; and these are significantly higher than the global values.

A (weak) continuously flexible single-degree-of-freedom mechanism can be seen but this is clearly not as well defined as that of CA/TG. With slide measured in angströms and roll and twist in degrees, the correlations are:

$$D_y(\mathrm{TA}) = -0.08\rho(\mathrm{TA}) + 0.19 \qquad (R = -0.55), \tag{5.4}$$

$$\Omega(\mathrm{TA}) = -0.56\rho(\mathrm{TA}) + 37.15 \qquad (R = -0.59), \tag{5.5}$$

$$D_y(\mathrm{TA}) = 0.11\Omega(\mathrm{TA}) - 4.09 \qquad (R = 0.66). \tag{5.6}$$

The correlation coefficients are all inferior to those for CA/TG. This can be seen from the plots to be due to the clustering of a significant portion of the data in the low-slide region. In fact, we might arguably describe TA as a bistable step; and this is indeed what Hunter (1993) concludes from his base-base energy interaction calculations. But note that despite the fact that a certain degree of bistability can be seen, it is obscured by considerable scatter, which in turn helps to give TA the appearance of a continuously mobile step. TA is thus a somewhat inconclusive step (cf. AC/GT above), particularly on account of its relative under-representation: there are only 21 examples (figure 2). Here, we shall adopt the classification of TA as a continuously flexible step; but we note nevertheless a 'bistable' tendency.

(*c*) *Overall classification*

We have now considered individually the conformational characteristics of each type of dinucleotide step. In order to obtain an insight into the relative characteristics of the steps, it is useful to have a diagram that summarizes the behaviour of all steps. We shall now consider a single diagram which epitomizes the performance of all nine steps (i.e. all steps except AG, for which there are few data). Figure 9 is a plot of the standard deviations of roll, slide and twist *versus* step type. The steps along the horizontal axis have been arranged in the following order: CA/TG, TA, GG/CC, CG, GC, AC/GT, GA/TC, AT, AA/TT. This order has been set up so that the more variable steps precede the less variable. Slide gives a very clear descending plot. Roll and twist give similar trends, but not as clear as slide. The GG/CC step appears to be responsible for disturbing the general trend by showing especially low deviations of both roll and twist: the rigidity of this step with respect to roll has already been noted.

From our discussion thus far, it appears that slide is emerging as the single parameter that best discriminates between the various steps: if one seeks a single parameter to describe the various conformations, then slide appears to be the best candidate. We have presented and discussed a (3×3) comparative matrix of slide frequency plots elsewhere (figure 2 of El Hassan & Calladine 1996). From that figure and from figures 6–9 we can summarize the following kinematic classification of the dinucleotide steps:

Rigid. These occupy a specific region of the roll/slide/twist conformational space. This class includes AA/TT, AT and GA/TC. All members of this class show low standard deviations of roll, slide and twist.

Loose. These are all non-rigid steps, namely GG/CC, GC, CG, CA/TG, TA and AC/GT – which is the least-clear of the steps. A single feature common to all steps in this class is the high standard deviations of roll, slide and twist compared to those of the rigid steps. The only exceptions to this general rule are the standard deviations

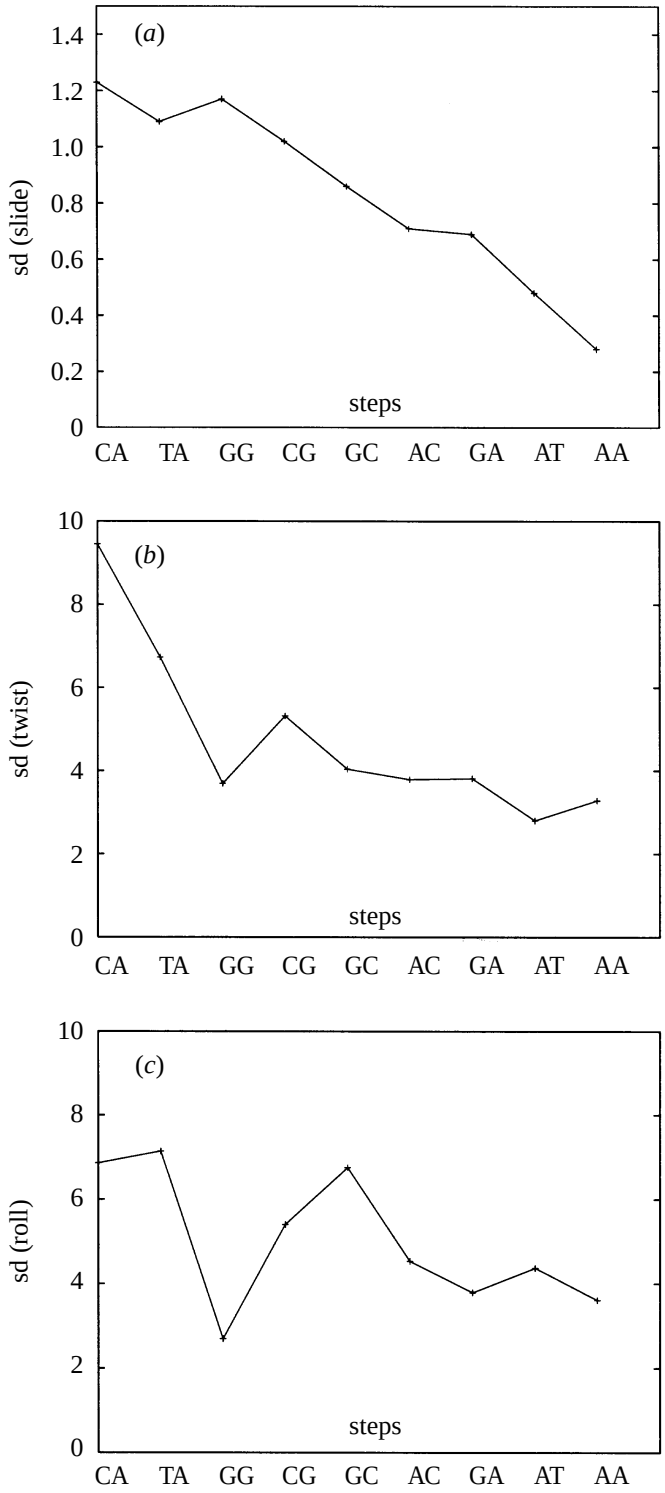

Figure 9. Standard deviations of (*a*) slide, (*b*) helical twist and (*c*) roll plotted *vs* step type. The steps are here arranged broadly in order of increasing *rigidity* along the horizontal axis.

of roll and twist of GG/CC, because the lack of rigidity of this step is mainly with respect to slide, as we have already seen. Another exception is AC/GT which has standard deviations of step parameters that are not as high as those of the other loose steps. The reason for this is that in addition to the fact that the majority of examples lie in a reasonably tight low-slide cluster, the average step parameters of the few remaining scattered examples are not very different from those of the low-slide cluster. Thus it is perhaps more appropriate to classify AC/GT as neither loose nor rigid but rather an 'intermediate' step. Apart from that step, loose steps can be further classified into:

Bistable. These are steps that mainly appear to belong to one of two regions in the conformational space. Members of this class are all homogeneous G|C steps.

Flexible. This class consists of the CA/TG and TA steps. The conformational characteristics within this class are highly variable, and they score highest in terms of standard deviations of roll, slide and twist. Moreover this variation appears to take place along a single-degree-of-freedom path with roll, slide and twist all approximately linearly correlated; but recall the 'bistable' tendency of TA.

6. Base-pair parameters

(*a*) *General*

We have already concluded that of all the base-pair parameters, only the angular parameters, namely propeller, buckle and opening are significant in the sense that, on the whole, they show considerable variation that warrants further investigation. In what follows, we shall examine the variation of opening, propeller and buckle in more detail, paying particular attention to their sequence-context. In order to do this, we need somehow to relate the base-pair parameters to the dinucleotide-step context in question. The most straightforward way to proceed is perhaps to take the average of the two values of the base pairs making up the step as a base-pair parameter value for the step. We shall indeed use this approach for propeller. Thus, from now onwards, we shall use the term 'propeller' to refer *either* to the propeller of a particular base pair *or* to the step-propeller, i.e. average of propellers of the two constituent base pairs. It should be clear from the context which definition is being used. Opening, on the other hand, will not be pursued in such detail for reasons that will be discussed shortly. The sign ambiguity of buckle implies that it cannot be dealt with in the same manner as propeller. Yanagi *et al.* (1991) have suggested a very good way of defining a step parameter that carries some information about the buckling of the constituent base pairs, while eliminating the sign ambiguity. This parameter, which is illustrated in figure 10*c*, is cup; it is defined simply as the difference between the values of buckle (κ) for the two consecutive base pairs. Thus

$$\text{cup} = \kappa_{i+1} - \kappa_i.$$

In summary, step-propeller and cup will be used to define step versions of base-pair propeller and buckle respectively. No claim is made for any equivalence between 'average buckle' and cup. The point here is that some means for assessing step-buckling is required, and this should be done in a way that would result in a single value for the dinucleotide step in question, while being independent of the way the step is reckoned. We can now examine the behaviour of the angular base-pair parameters with respect to their sequence context.

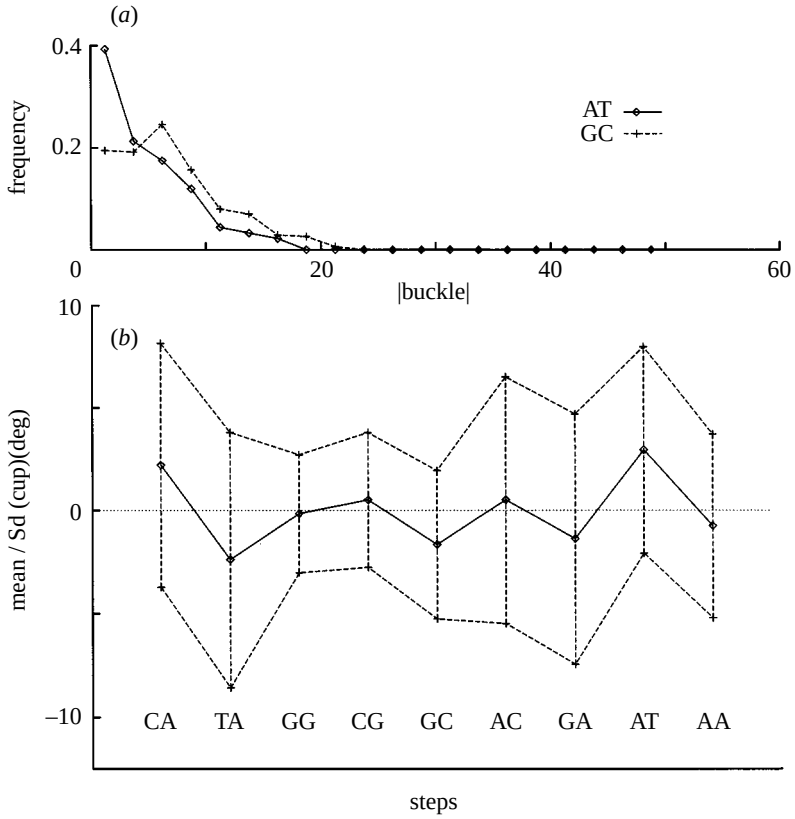

Figure 10. (*a*) Overall behaviour of base-pair buckle of AT and GC base pairs (only the absolute values are plotted). (*b*) Overall behaviour of cup. The plot gives a central line through the average values as well as the standard deviations bars and envelope. The steps are arranged as in figure 9.

(*c*)

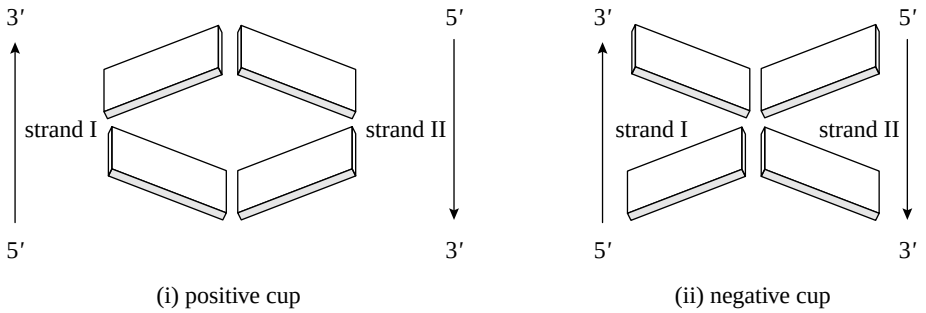

Figure 10. (*c*) Definition of cup.

(*b*) *Opening*

Opening, as defined by the Cambridge Accord (Dickerson *et al.* 1988), essentially describes the relative extension or compression of the major-groove and minor-groove Watson–Crick H-bonds. The right-handed sign convention of the Cambridge Accord is such that positive opening implies an extended major-groove H-bond with a compressed minor-groove H-bond.

It turns out that from an examination of opening of all A–T and G–C base pairs that A–T base pairs prefer slightly higher opening and show more spread about the mean than G–C base pairs. The average (and standard deviations) of opening are *ca.* 8°(8°) for A–T and *ca.* 3°(6°) for G–C. This difference between G–C and A–T base pairs is consistent with the *three* H-bonds of G–C compared with the *two* of A–T; thus we might expect G–C base pairs to be more resistant to opening than A–T base pairs. However, it is clear that, unlike propeller, opening may perhaps play little part in the interaction between the base pairs making up the dinucleotide step; all of these deformations are in the plane of the base pairs. Thus we shall not pursue any sequence-dependent trends in the opening variable.

(*c*) *Buckle*

Figure 10*a* shows a frequency plot of buckle for all G–C base pairs and all A–T base pairs available in the database. The striking feature of this plot is the similarity in performance between the two types of base pair. Unlike opening (and propeller, as will be seen shortly), there is no tendency for A–T base pairs to show more deformation and variability with respect to buckle when compared to G–C base pairs: indeed, G–C base pairs display more spread with respect to buckle. The average and standard deviations of buckle are *ca.* 1°(6°) for A–T and *ca.* 1°(8°) for G–C. This is somewhat surprising; and it suggests that, energetically, buckling might not be very costly and that it might therefore be less sensitive to the base sequence. This is confirmed when the plot of averages and standard deviations of cup *versus* step type of figure 10*b* is examined. In this plot, step type is plotted along the horizontal axis in the same order as in figure 9. The average values of cup for the nine dinucleotide steps are plotted along the vertical axis together with 'error-bars', each of which has a length twice the standard deviation of cup for the step in question. An 'envelope' enclosing all 'error-bars' is also included. As is clear from the plot, cup shows little sequence-dependence and somewhat large deviations in most of the steps.

(*d*) *Propeller*

Figure 11*a* gives overall frequency plots of propeller for A–T and G–C base pairs. Two smooth (nearly Gaussian) plots can be seen for the A–T and G–C base pairs, with the peak of the curve for A–T base pairs displaced by around −6° relative to that of the G–C base pairs. The averages (and standard deviations) of propeller for A–T and G–C base pairs are −15.4°(5.8°) and −9.2°(6.7°) respectively. The greater levels of the propeller average for A–T reflect a tendency of the A–T base pairs to show a greater extent of deformation when compared to G–C. This is compatible with the A–T base-pair's missing minor-groove H-bond in comparison with G–C, as noted above.

Figure 11*b* shows an average/standard deviation plot for propeller, along the same lines as the plot of figure 10*b* for cup. As in that plot, and indeed all plots where the nine dinucleotide steps are arranged along the horizontal axis, we have arranged the steps along the horizontal axis in order of increasing rigidity, just as in figure 9. It is clear that the standard deviations are very similar for most steps, and that there appears to be a very good correlation between propeller and rigidity of the nine dinucleotide steps. Thus AA/TT, AT, GA/TC steps show generally high levels of propeller while the loose steps, CA/TG, TA and the homogeneous G|C steps all show low levels of propeller. The AC/GT step, which has been classified as an intermediate

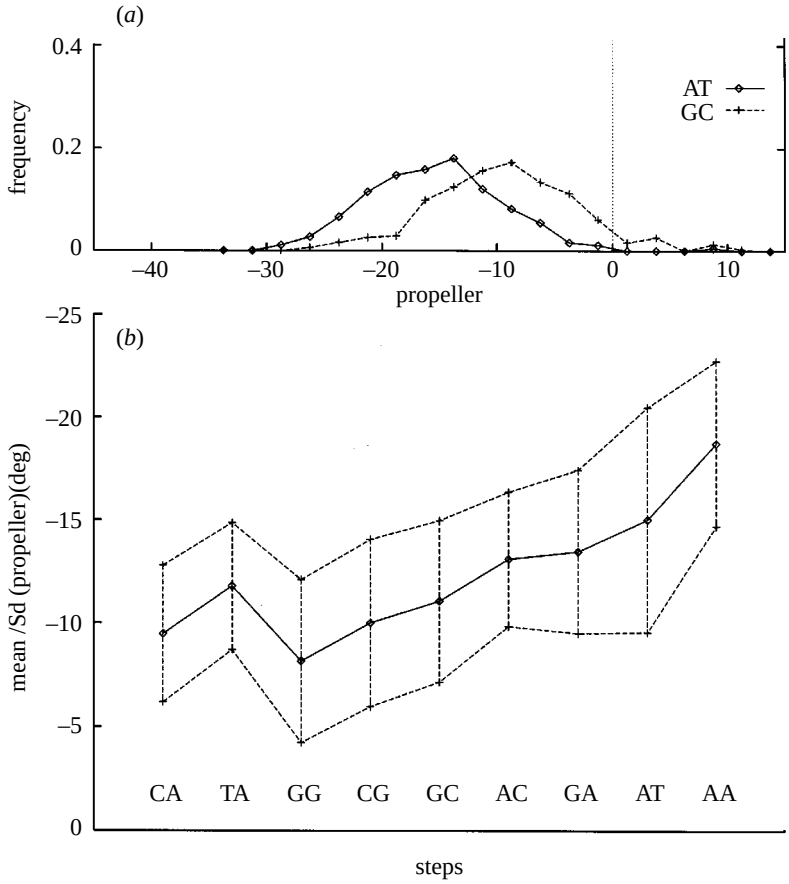

Figure 11. (*a*) Overall behaviour of base-pair propeller of AT and GC base pairs. (*b*) Overall behaviour of step-propeller. The plot gives a central line through the average values as well as the standard deviations bars and envelope. The steps are arranged as in figure 9.

one between loose and rigid steps, also adopts intermediate propeller levels. We shall elaborate on this propeller/flexibility correlation later on in this paper (§ 11).

7. Dodecamer end-sequences of the type CGCG

The crystal-packing effects on the dodecamer-end sequences, usually of the type CGCG, are well recognized; and indeed, these regions are sometimes argued to be responsible for the poorer resolution to which dodecamer-structures can be solved by X-ray methods when compared to decamers (Dickerson *et al.* 1991). Previous comparative studies (e.g. Gorin *et al.* 1995) have ignored these sequences on the basis that they carry very little information on the sequence-structure relationships. As we have already remarked above our viewpoint here is different. We treat the conformational characteristics of these end-sequences in much the same way as we treat other sequences from different regions of the various oligomers in the database. In all cases, the adopted conformation is affected both by sequence and by the crystalline environment. However, we acknowledge the special significance of the crystal-packing effects in these particular sequences, and since the majority of the high-slide

Table 4. *A comparison between CG/CG and GC/GC steps as found in CGCG 'dodecamer end-sequences' and in 'other' high-slide examples of the steps*

class	helical twist°	roll°	slide (Å)	shift (Å)	step propeller°
(1) **CG**; 67 examples overall: 46 'end-CGCG' and 21 'other'					
(ALL)	36.75(4.06)	2.06(4.27)	0.45(0.46)	0.41(0.33)	−9.94(3.92)
(END-CGCG)	38.08(2.62)	1.43(4.31)	0.33(0.48)	0.41(0.35)	−10.30(4.16)
(OTHER)	33.85(5.10)	3.44(3.94)	0.71(0.31)	0.41(0.26)	−9.14(3.29)
(2) **GC** 45 examples overall: 39 'end-CGCG' and 6 'other'					
(ALL)	38.93(2.56)	−4.66(5.42)	0.63(0.39)	0.75(0.32)	−11.42(3.31)
(END-CGCG)	39.43(1.12)	−5.70(4.78)	0.74(0.25)	0.81(0.28)	−11.08(3.21)
(OTHER)	35.71(3.93)	2.13(4.56)	−0.06(0.43)	0.35(0.30)	−13.67(3.33)

GC/GC and CG/CG examples come from these dodecamer-end regions, we need to pay particular attention to them. We shall therefore attempt here to examine more closely these end-sequences, and compare their characteristics with those of similar sequences derived from other regions in the oligomers included in our database.

Table 4 summarizes the means and standard deviations of twist, roll, slide, shift and step-propeller, for the GC and CG steps. Three rows are included for each step: the first gives the averages and standard deviations for all high-slide examples; the second gives the corresponding parameters for examples from ends of dodecamers; and the third gives the data for all 'other' examples. Excluding shift for the time being, we can see that there are, in fact, some subtle differences between the end-sequences and 'other' examples, for both step types. As far as the CG step is concerned, the main differences are in the values of helical twist and, to a lesser extent, slide. However, there appear to be more noticeable differences in the case of GC. Whereas GC steps from dodecamer-end regions appear to prefer negative roll, high slide and high twist, GC steps from 'other' regions appear to prefer positive roll, low slide and low twist. Note that five of the six 'other' high-slide GC steps are in the context G<u>GC</u>, which has been noted in several studies to prefer positive roll (Calladine & Drew 1986; Brukner *et al.* 1993; Travers 1995). Indeed the only 'other' GC step that is not found in the context G<u>GC</u>, adopts a negative roll angle *ca.* −4°.

Shift in CG from the dodecamer end sequences can be seen from table 4 to be similar to that for 'other' CG steps. The situation is rather different in the case of GC. End-sequence GC steps appear to adopt significantly higher shifts than 'other' GC steps. This effect will in fact become useful below (§ 10) when we attempt qualitatively to modify Hunter's prediction of the conformational preference of GC, and to suggest that a high-slide conformation is possible for GC steps provided that shift attains relatively higher values than normal. It is interesting to note that of the 'other' GC examples, there are two particular steps – from the bdj039 decamer – that adopt relatively higher slide values (*ca.* 0.25 and 0.4 Å) than the remaining 'other' GC steps; and indeed these two examples adopt appropriately high shift values (*ca.* 0.8 and 0.6 Å respectively).

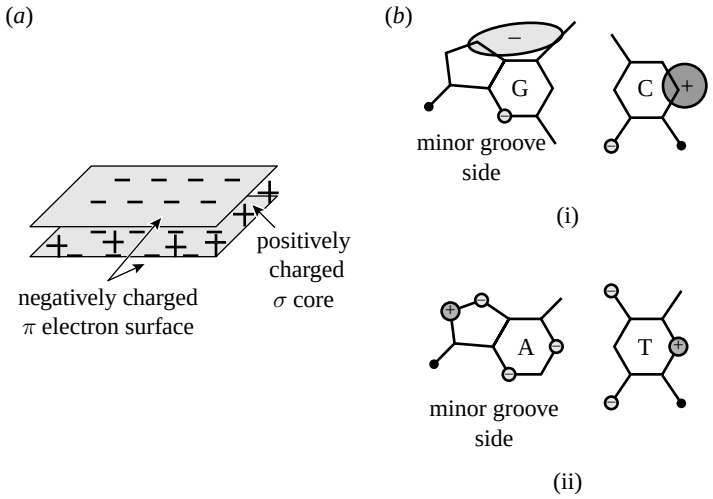

Figure 12. Basic elements of Hunter's π–π interaction model. (*a*) Schematic representation of an aromatic molecule with a positively charged σ core sandwiched between two π electron clouds. (*b*) The partial atomic charge distributions in A-T and G-C base pairs.

The high-shift property is not in fact unique to the GC step, but is also seen in some examples of AC/GT. The average shift for all AC/GT steps is actually low, *ca.* 0.2 Å but the standard deviation (*ca.* 0.6) is much higher than values shown by other steps, which are in the range 0.2–0.3 Å. There are 12 AC/GT examples that exhibit high shifts, i.e |shift| $>$ 0.5 Å: 9 from the low-slide cluster and 3 from the scattered cluster. What is striking is that the only common feature among these steps is the high magnitude of shift; all other parameters show normal scatter within the conformational space of AC/GT. This tendency for AC/GT to adopt high shift values is a purely empirical observation, and we have no clear explanation for it. It seems possible that high-shift is an intrinsic property of AC/GT that can perhaps be demonstrated by some theoretical approaches such as the π–π energy calculations of Hunter (1993).

8. Hunter's predictions and comparison

Hunter (1993) proposed a theoretical scheme for calculating preferred conformations of DNA dinucleotide steps. He considered each of the 10 dinucleotide step types and minimized the interaction energies between two consecutive base pairs with respect to some selected degrees of freedom. Thus roll and slide were varied from $-25°$ to 25° and -2.5 Å to 2.5 Å respectively, and twist and propeller were set to some discrete values: 31°, 36° and 41° for helical twist and 0° and $-15°$ for propeller.

For a given step, Hunter's interaction energies are due to two main effects. First there are *steric effects*. These are due to atoms, not in a bonding situation, approaching each other to within van der Waals contact. Second there are *electrostatic effects*. An aromatic molecule, such as a DNA base, can be modelled as consisting of a positively charged σ core sandwiched between two π electron surfaces (figure 12*a*). The resulting electrostatic charge distribution is, however, not uniform. The reason for this is that certain atoms, that are more electronegative or electropositive than others, tend to polarize the uniform charge distribution and produce an overall non-uniform distribution (figure 12*b*). All of this leads to three electrostatic contri-

butions to the overall interaction energy, as follows, (1) atom–atom effects, which are due to the interactions between the partial atomic charges; (2) $\pi\sigma$–$\pi\sigma$ effects, which are due to the electrostatic interactions between the σ core of one base pair and π electron surface of the other; and (3) $\pi\sigma$–atom effects, which are due to the interactions between the partial atomic charges of one base pair and the π electron cloud of the other.

The main conclusions of Hunter can be summarized as follows. First, for thymine (T) bases, there is a clash between its methyl group and the sugar-phosphate backbone, which blocks positive roll and induces high propeller in steps of the type AX/XT. Note that in his calculation Hunter removes the backbone and schematically replaces it by a methyl group in each base: the carbon of the methyl group is situated at the C1′ location and the rest of the methyl group is constructed using standard C–H bond lengths and H–C–H bond angles. Second, as shown in figure 12*b*, the G–C base pair is electrostatically very different from the A–T base pair. Thus, while the A–T base pair has relatively small and well spread-out positive and negative charges, the G–C base pair is characterized by a large positive partial atomic charge in the cytosine and a large negative partial atomic charge in the guanine. This would, in general, result in the GG/CC, GC and CG steps preferring *offset* (i.e. non-zero slide) geometries as opposed to the well-stacked B-DNA conformations. It may be inferred from figure 12*b* that the partial atomic charges on their own might explain the bistability of GG/CC but not that of GC and CG. A more complete discussion of this particular aspect will be given below in § 10.

In addition to these broad and qualitative generalizations, Hunter suggested preferred roll/slide conformations in greater detail for each of the 10 types of dinucleotide steps. The 'boxes' in the roll/slide plots of figures 6, 7 and 8 represent Hunter's predicted conformations. It is immediately obvious that with the exception of AA/TT, and, to a lesser extent GG/CC and CG, where there is excellent agreement, there is in general poor agreement between Hunter's predictions and the actual data. Significantly, Hunter's calculations seem to miss the somewhat continuous single-degree-of-freedom mechanism of the CA/TG step.

We should emphasize, however, that in a broad and a qualitative sense these theoretical predictions are consistent with the data from X-ray studies. Thus, Hunter's suggestion that AX/XT steps should prefer higher levels of propeller than the other steps is largely confirmed. The aversion of homogeneous G|C steps to zero-slide geometries and their preference for offset geometries, is reflected in figures 6, 7 and 8, and also in the strong bistable trend down the homogeneous G|C column of figure 2 of El Hassan & Calladine (1996). Finally, Hunter's predicted rigidity of the AA/TT step in a classical 'B' form conformation is also consistent with what is observed. Another feature is the bistability of TA predicted by Hunter's calculations. We have noted this tendency earlier in § 5 *b*.

Possible reasons for the lack of agreement between Hunter's predictions and the actual data follow from the underlying hypotheses of the calculations. In addition to performing all calculations in the absence of surrounding water, Hunter also removed the sugar-phosphate backbones. Although the variability of the backbone (Hunter 1993; Calladine & Drew 1992) indicates that its effect might perhaps not be very pronounced, we shall argue in § 9 that the backbone may play some role in determining the preferred dinucleotide step conformations. Another point is the limited number of degrees of freedom that were considered by Hunter, namely roll and slide, and to a lesser extent twist and propeller. Remembering that the electrostatic effects are very

nonlinear, it might be possible that some of the less variable parameters that were not considered by Hunter are actually significant in terms of interaction energies. Finally, it seems possible that these electrostatic energy calculations are likely to be somewhat sensitive to the choice of charge distributions. Although this sensitivity might not necessarily impinge on the usefulness of the calculations in giving a proper insight to the problem of sequence-dependent structure of DNA, it might affect the suitability of the calculations for a detailed determination of conformational geometries. In any case, a careful 'calibration' of the results of such calculations needs to be made with respect to the empirical results of one or two well-characterized steps before a detailed analysis of the conformations of all possible dinucleotide steps can be made.

9. Conformational characteristics of the sugar phosphate backbone

Hitherto, examination of the backbone conformation in the literature has mainly concentrated on detailed analyses of the large number of *internal* angles (torsion angles α, β, γ, δ, ϵ and ζ, the Glycosyl angle χ and the internal sugar ring angles ν_1, ν_2, ..., ν_5) that describe the fine detail of the backbone structure. Although some correlations between some of these various angles have been suggested (Fratini *et al.* 1983; Saenger 1984), detailed analysis of the backbone structure has had relatively little impact; and indeed, as already indicated, current thinking in the field of DNA structure has relegated the role of the backbone to that of a sort of passive 'string' that delineates, somehow, the boundaries of the dinucleotide step conformations. One reason why a complicated description of the backbone is not particularly useful in giving a good global picture, is that these angular parameters may vary internally in a largely self-compensating manner. In this paper we use a simple *two*-parameter model for describing the overall backbone conformation and relating it to the kinematics of a base pair. These parameters are: (1) the position of the phosphate group ((X, Y, Z) coordinates) with respect to the mid-step triad of the CEHS scheme†; and (2) the same-strand C1′–C1′ distance, which we shall denote by the symbol d_{C1C1}.

An examination of the phosphate positions in all entries in our database revealed that the Y-coordinate is more or less fixed at a value *ca.* 8 Å. The X and Z coordinates, on the other hand, show more variation. Both coordinates have a tendency to cluster into two regions, with the Z-coordinate in particular clustering into two distinct regions in the vicinity of −0.5 Å and 2.5 Å respectively. Moreover, a positive correlation is seen between the X-coordinates and the corresponding Z-coordinates ($X = 0.43Z - 2.71$; see figure 13*a*). Although the correlation coefficient is high ($R = 0.85$), it reflects a bimodal relationship rather than a continuous one.

The high-Z cluster and the low-Z cluster in figure 13*a* correspond to data from 'A-form' and 'B-form' oligomers respectively. Calladine & Drew (1984) suggested that there is a strong linkage between the phosphate Z-coordinate and slide, which they illustrated diagrammatically as shown in figure 13*c*. In order to give a more complete

† If we use a *single* reference frame, there would be a reversal in the signs of the Y and Z coordinates between the two phosphate groups associated with the same dinucleotide step. We have therefore decided to use *two* mid-step triad related by a 180° rotation about the X-axis for each dinucleotide step: one frame for each strand. This ensures that Y is always *positive* and that X and Z are *positive* if the phosphate lies towards the towards the major-groove side and the 3′ base in the corresponding strand respectively.

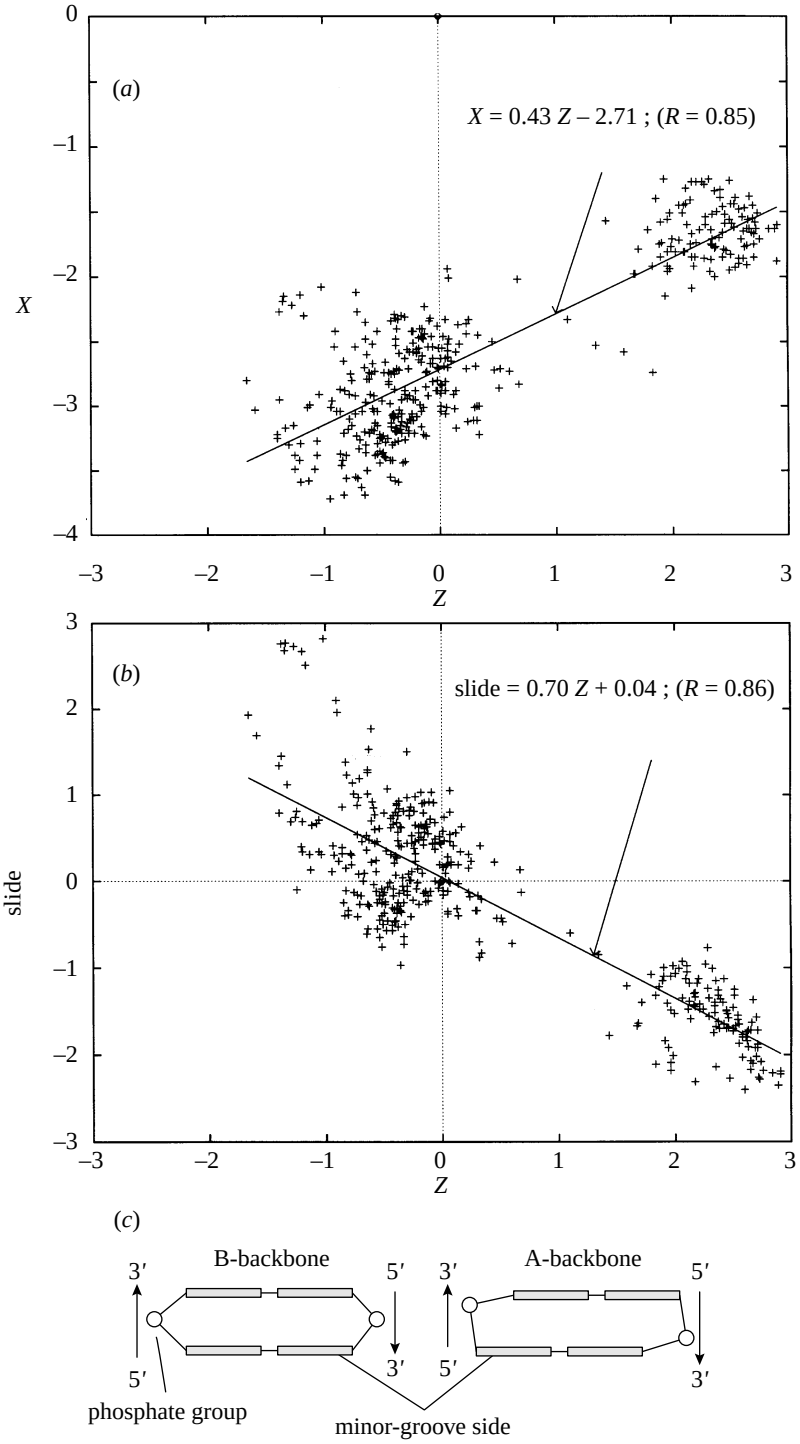

Figure 13. Correlations between (*a*) the phosphate *X* and *Z* coordinates and (*b*) the phosphate *Z* coordinate and slide. (*c*) Schematic diagram, adapted from Calladine & Drew (1984), to illustrate the correlation between the phosphate *Z* coordinate and slide.

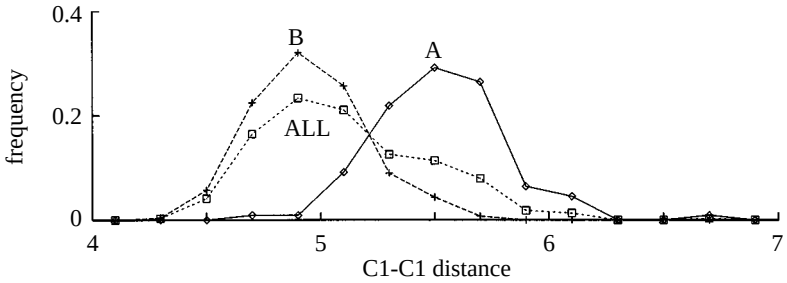

Figure 14. Frequency plots for the same-strand C1′–C1′ distance for A-backbone data, B-backbone data, and all data taken together.

quantitative examination of this linkage between slide and Z, we have plotted values of these parameters for all steps in our database, as shown in figure 13*b*. The figure shows a very clear negative bimodal correlation, with one backbone conformation preferring low slide ($\not> -1$ Å) and the other preferring higher slides. The correlation statistics are given by $D_y = -0.70Z + 0.04$; ($R = -0.86$). The bimodality of the correlation is demonstrated by the fact that when we try to correlate the low-slide data and the high-slide data separately, we find that the correlation coefficients are -0.52 and -0.46 respectively while the slopes are -0.42 and -0.31 respectively. In other words, despite the existence of relatively weak correlations between Z and slide for each backbone conformation, the overall correlation expresses mainly the separation of the data into two regions.

We may thus conclude that the backbone exists in two different conformations that are clearly described by the Z coordinate of the phosphate as measured relative to the step-reference frame. This switching of conformations is strongly slide-dependent; and this behaviour can be summarized by:

$$\text{A-backbone:}\quad [Z \approx 2 \longrightarrow 3\ \text{Å}] \Longleftrightarrow [D_y \approx -2.5 \longrightarrow -1\ \text{Å}]$$

$$\text{B-backbone:}\quad [Z \approx -1 \longrightarrow 0\ \text{Å}] \Longleftrightarrow [D_y \approx -1 \longrightarrow 3\ \text{Å}]$$

We have already noted the ambiguity of the terms ‘A-form’ and ‘B-form’ as used by many workers in the field, and we have also seen that as far as dinucleotide step-conformations are concerned there is more or less a continuum of conformations, with A-backbone steps occupying the low-slide/high-roll/low-helical twist end of this allowed conformational space and the B-backbone data occupying the rest of it. We shall therefore use from now onwards the terms ‘A-backbone’ and ‘B-backbone’ instead of the rather imprecise terms ‘A-form’ and ‘B-form’; and we shall base our definitions on the above-deduced guidelines, namely according to the value of the Z-coordinate of the phosphate group.

Figure 14 shows frequency plots of d_{C1C1} for all data taken together, and all A-backbone and B-backbone data taken separately; while in table 5 we give the averages and standard deviations of these distances. Although the standard deviations of table 5 indicate a more or less constant backbone length, the frequency diagrams show that there is a spread of about 1 Å for each particular conformation and about 1.5 Å for the two taken collectively. Moreover, there is a very clear difference between the average distances for each conformation; thus the A-backbone ($Z \approx 2.5$ Å) prefers an average $d_{\mathrm{C1C1}} \approx 5.5$ Å, while the B-backbone ($Z \approx -0.5$ Å) prefers a lower average value *ca.*

Table 5. *Averages and standard deviations of the* C1′–C1′ *distances of all steps, A-backbone steps and B-backbone steps*

class	$\overline{d_{\mathrm{C1C1}}}$ (Å)	sd(d_{C1C1}) (Å)
ALL	5.11	0.36
A-backbone	5.52	0.28
B-backbone	4.95	0.24

4.9 Å. In other words, the two backbone conformations may be distinguished on the basis of either the phosphate Z-coordinate or d_{C1C1}.

It is physically obvious that if d_{C1C1} were to vary then it would do so in a way that correlates with roll, slide and twist of the step in question. In order to see if any such correlations exist, we need to examine plots of d_{C1C1} *versus* roll, slide and twist for each step individually. It turns out, as expected, that rigid steps AA/TT, GA/TC and AT all occupy single B-conformation clusters†, while the bistable steps GG/CC, CG, GC and AC/GT all occupy two clusters. The flexible steps CA/TG and TA on the other hand are more interesting. The B-form examples of these steps show continuous variation, with d_{C1C1} linearly correlated with all three parameters roll, slide and twist. (The d_{C1C1} roll correlation of TA was very poor. Roll values were biased to large positive values and showed little correlation with d_{C1C1}.) In order to illustrate these points, we have included the d_{C1C1} roll-slide-twist plots for CA/TG in figure 15*a*. As the figure shows, the backbone of the CA/TG is almost entirely in the B-conformation: only two of the 26 examples adopt the A-backbone conformation. Excellent linear positive correlations can be seen between d_{C1C1} and both slide and twist and a negative correlation with roll for the B-backbone data. The correlation coefficients are all good and the relationships can be summarized by the following formulae in which all distances are in angströms and all angles in degrees:

$$d_{\mathrm{C1C1}} = 0.18D_y + 4.89 \qquad (R = 0.74), \tag{9.1}$$

$$d_{\mathrm{C1C1}} = 0.025\Omega + 4.24 \qquad (R = 0.88), \tag{9.2}$$

$$d_{\mathrm{C1C1}} = -0.03\rho + 5.18 \qquad (R = -0.81). \tag{9.3}$$

In order to illustrate the difference between *flexible* and *bistable* steps we have included plots of d_{C1C1} *versus* slide, twist and roll for GG/CC in figure 15*b*. A clear conformational bistability can be seen particularly in the d_{C1C1} slide plot. All A-backbone steps occupy the low-slide/high-roll/low-twist region, while the B-backbone steps occupy the high-slide/low-roll/high-twist region. Moreover the two subsets in each of the three plots are all reasonably tightly clustered. The only exception here is the B-backbone conformation set of the d_{C1C1} helical twist plot, which indicates a good positive correlation between helical twist and d_{C1C1}:

$$d_{\mathrm{C1C1}} = 0.04\Omega + 3.71 \qquad (R = 0.86).$$

It is important, however, to distinguish between this behaviour and that of CA/TG. Whereas the correlations in CA/TG involve all of roll, slide and helical twist, and as

† AT and GA/TC have a few (four) examples that adopt the A-conformation; and these have already been highlighted.

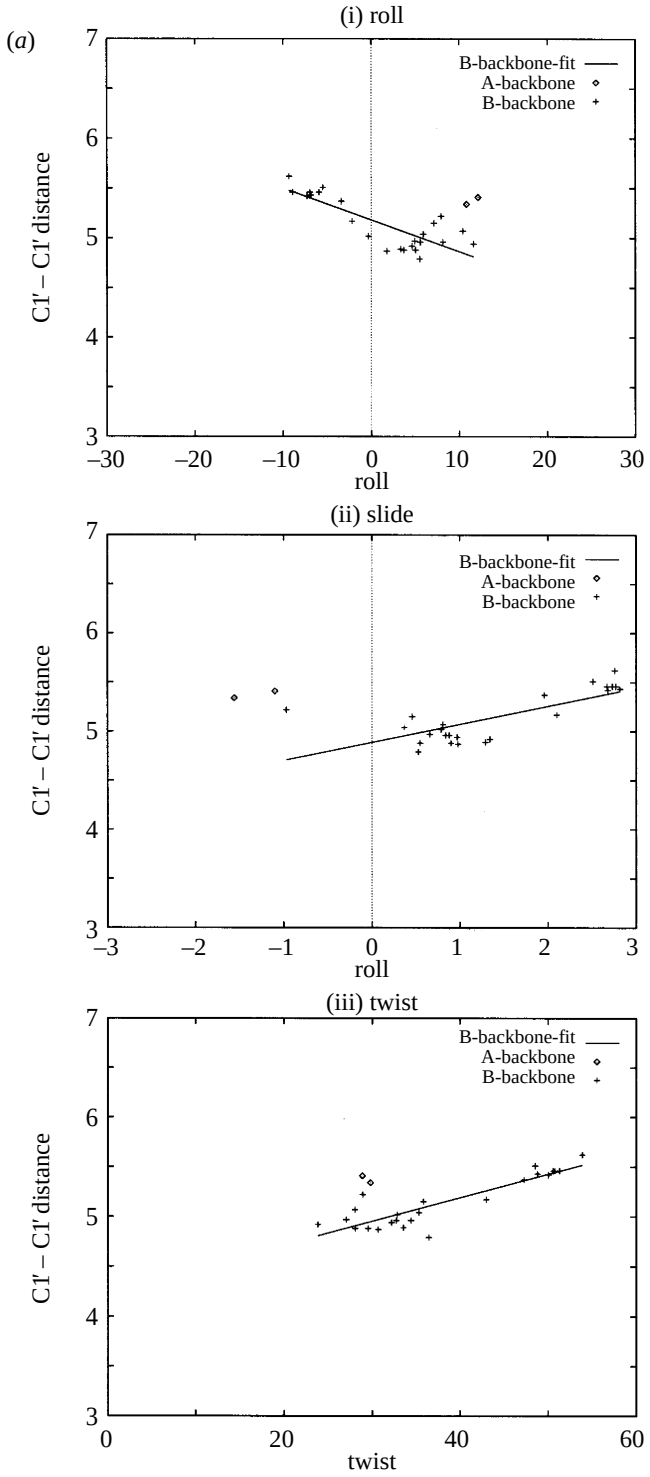

Figure 15. C1′–C1′ distance plotted against roll, slide and twist for (*a*) CA/TG. Best-fit lines for B-backbone data are included for all of the roll, slide and twist plots in the case of CA/TG and for the twist plot in the case of GG/CC.

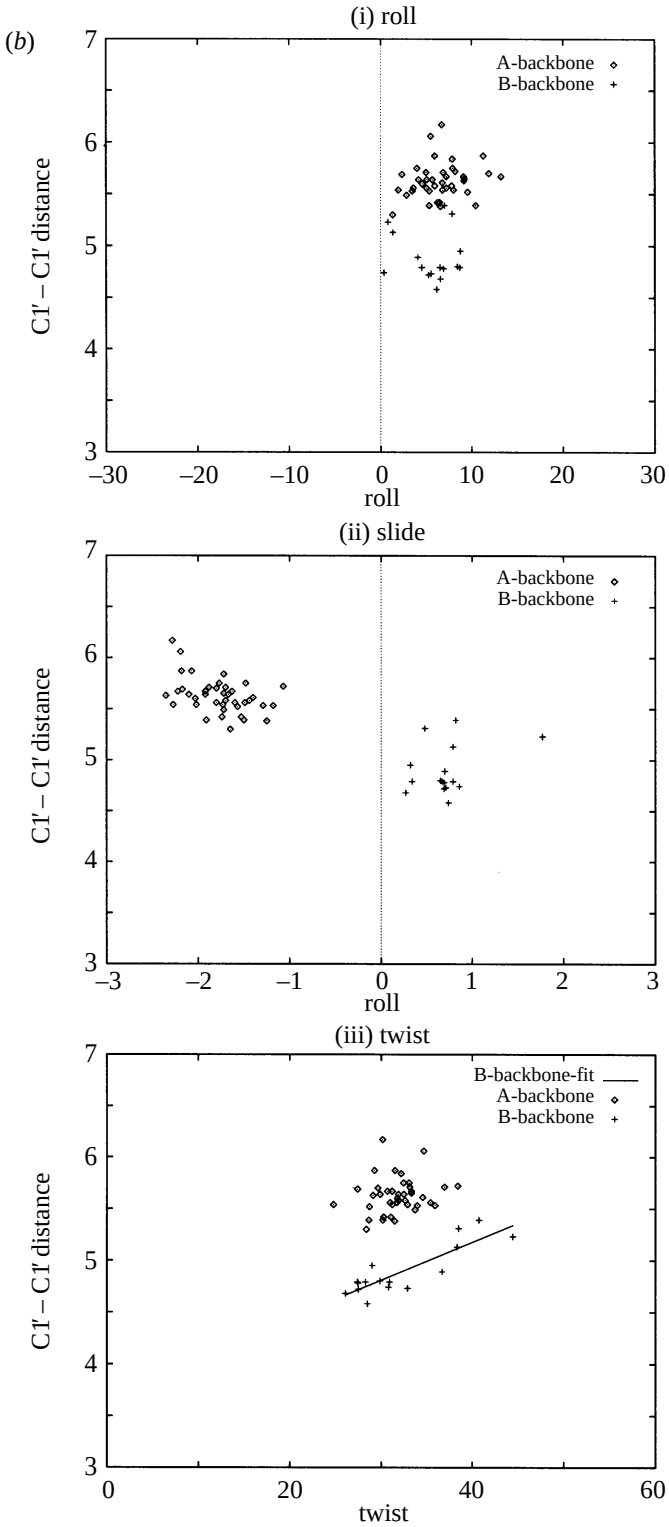

Figure 15. *Cont.* (*b*) GG/CC.

a result identify a single-degree-of-freedom mechanism, the apparent flexibility of the B-backbone data of GG/CC involves only helical twist, and it takes place at more or less constant high slide and positive roll. We shall return to this point in the § 10.

Figure 16*a* gives a schematic representation of the backbone and its relation to the step conformation. The first picture of this figure shows that there are two allowable regions in the d_{C1C1} slide space: the B-backbone region, which linearly correlates d_{C1C1} and slide; and the A-backbone region to which the backbone switches if slide becomes intolerably low for the B-backbone. The second and third pictures show the corresponding plots for d_{C1C1} *versus* twist and d_{C1C1} *versus* roll respectively. Together, the three plots describe the step's backbone-dictated conformations which constitute a single-degree-of-freedom motion with all three parameters linearly correlated: (roll and twist) and (twist and slide) positively correlated, and (roll and slide) negatively correlated. The two backbone conformations are illustrated by the stereoscopic pictures of figure 16*b*.

The complete absence of any example in the database where the two backbone conformations co-exist within the same oligomer indicates that there is some persistence in the backbone-conformation; and indeed this might play a part in determining the conformations of some dinucleotide steps. The few A-backbone examples of AT and GA/TC best illustrate this point. We have seen that AT and GA/TC on the whole prefer a rigid, near zero-slide conformation. The few (four) low-slide examples might suggest that AT, and GA/TC are capable of adopting low-slide (A-backbone) conformations; but they are unlikely to be pointing to a clear bistability as seen in the case of the homogeneous G|C steps. It is more likely that the persistence of the backbone conformation throughout any given oligomer is responsible for imposing low-slide conformations on these steps. In other words, it might be preferable for the AT and GA/TC steps to break loose from their preferred near-zero slide conformation and adopt instead lower slide conformations if they appear in the middle of an oligomer whose backbone is in the A-form, than it is for the backbone to switch conformations so as to accommodate the conformational preference of these steps.

10. A hypothesis for the sequence-dependent structure of DNA

We shall now present a simple first-order argument that attempts to account qualitatively for the conformational characteristics of the dinucleotide steps that we have described empirically above. Our argument will be based on the backbone effect, propeller-twisting and Hunter's special electrostatics of the G–C base pair; and we shall demonstrate that these three effects are, to first-order, responsible for the observed conformations of most of the dinucleotide steps examined in this work.

Base stacking interactions may be divided into *mechanical* and *chemical* effects. The mechanical (or stereochemical) effects are mainly steric clashes due to the deformation of base pairs, and particularly propeller. Propeller leads to definite mechanical locking at either the major- or minor-groove side. These locks might take the form of steric clashes of the type proposed by Calladine (1982), or they might take the form of attractive interactions such as bifurcated Hydrogen bonds postulated by Nelson *et al.* (1987). The chemical or electrostatic effects are mainly stacking preferences due to the electrostatic charge distributions of the constituent base pairs (figure 12*b*). The chemical aspect of our argument will be based on some elements of Hunter's π–π chemical theory, particularly the special electrostatics of G–C base pairs.

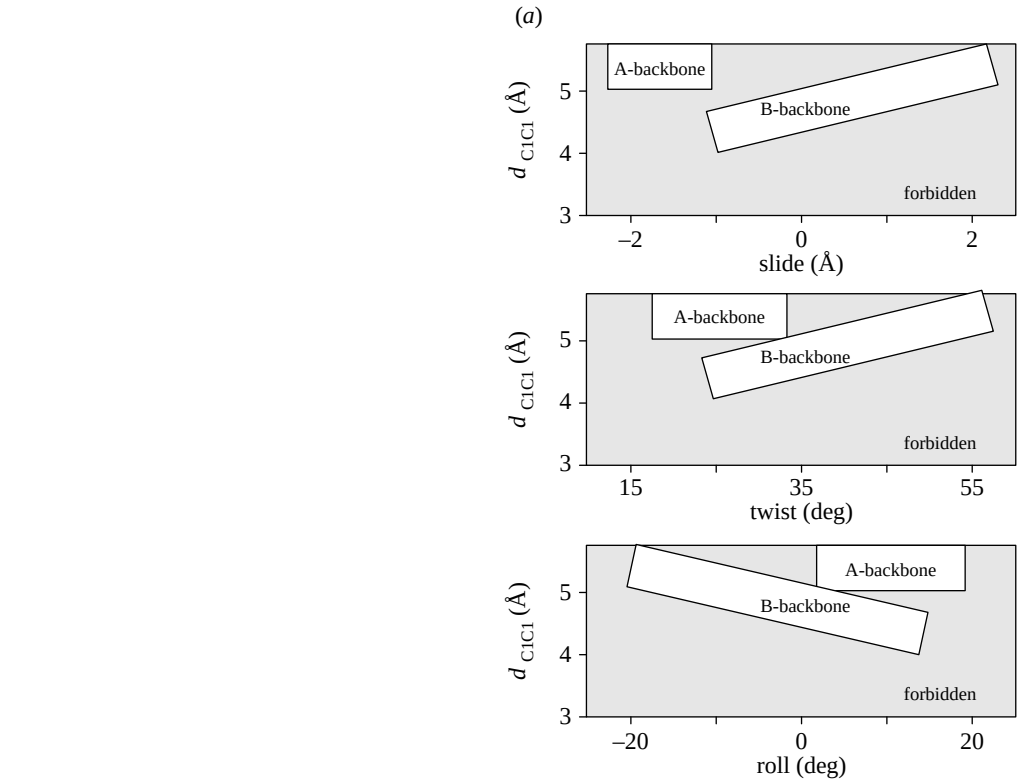

Figure 16. (*a*) Summary of the backbone-dictated conformations. The figure gives a schematic representation of the variation of the same strand C1′–C1′ distance with (1) slide, (2) twist and (3) roll; each plot shows the range allowed by the B-backbone and A-backbone. (*b*) Stereoscopic pictures illustrating the two main backbone conformations. (i) A typical A-backbone step: GG/CC from d(GGGGCCCC) (McCall *et al.* 1985) and (ii) a typical B-backbone step: AA/TT step from d(CGCGAAAAAAGCG) (Nelson *et al.* 1987).

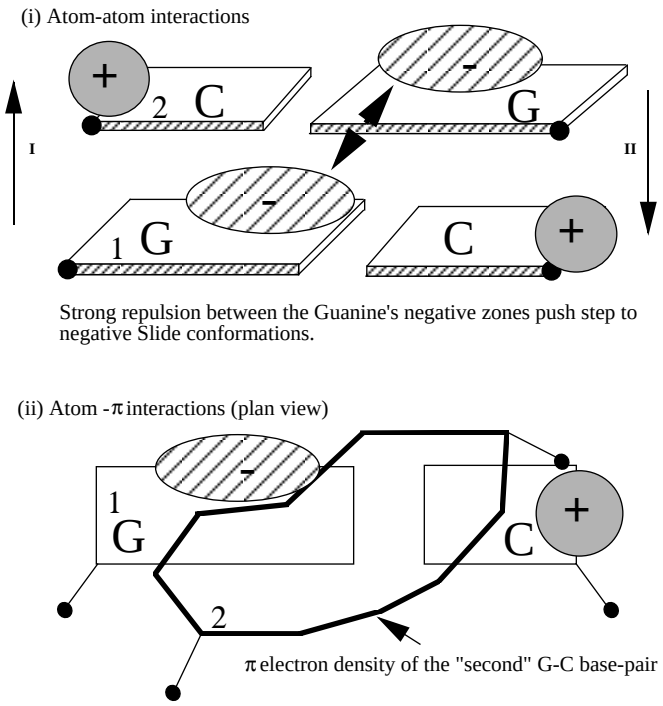

Figure 17. Atom–atom interactions and atom–π interactions in the GC step.

In the absence of base-stacking interactions, whether mechanical or chemical, one would expect the backbone to guide the step's conformations; and this is in fact what we observe in the case of CA/TG and, to a lesser extent, TA, which together span the total range of conformations that is permitted by the backbone. In other words, there can be no special base-stacking interactions in these two steps. This is confirmed by the fact that the levels of propeller in both of these steps is low; and neither of the steps is of the homogeneous G|C kind for which the special electrostatics play a very significant role. AA/TT, AT and GA/TC, on the other hand, all have high propeller; which would inhibit the freedom seen in the CA/TG step by giving rise to a cross-strand major-groove clash (or minor-groove clash at sufficiently low slide/twist). We therefore expect these steps to span a narrow region of the roll/slide/twist conformational space; and hence they constitute the rigid subset of the dinucleotide steps.

The homogeneous G|C steps all have low propeller, and hence the mechanical locking seen in AA/TT, AT and GA/TC is not relevant to them. The special electrostatics of the G–C base pair suggested by Hunter must therefore be largely responsible for the observed bistabilities of these steps. Indeed, as we have seen earlier, Hunter predicts a conformational bistability in both GG/CC and CG that correlates well with our empirical data. GC, however, was predicted by Hunter to have a single low-slide conformation rather than a bimodal one. We shall argue, on the basis of Hunter's main effects, that if we allow some front-to-back displacement or shift, then GC could show some degree of conformational bistability. Figure 17, which depicts the atom–atom interactions for GC, shows that these clearly favour low-slide confor-

mations; and indeed so do the $\pi\sigma$–atom interactions. Now at positive slide, the two large negative partial atomic charges of the guanines straddle each other; and this is clearly an unfavourable conformation. However, it can be seen that this atom–atom repulsion at positive slide may be relieved by a small ± *shift*. The atom–$\pi\sigma$ effects will probably not be influenced much by shift, since any shift that relieves the repulsion between the guanine's partial atomic charge of base pair (1) and the π electron density of base pair (2) would probably introduce a compensatory repulsion between the guanine's partial atomic charge of base pair (2) and the π electron density of base pair (1). In support of our argument, recall (§ 7) that we have already demonstrated that high-slide GC steps have a tendency to adopt unusually high shift conformations. The average (and standard deviation) of shift in the high-slide G–C steps are 0.75(0.32) Å, compared to the global values of 0.21(0.4) Å.

It is now clear that the bistability of all homogeneous G|C steps can be accounted for in terms of Hunter's base–base interactions. Whether or not the backbone-conformation is also involved in determining this bistability is not clear. Thus, on the one hand, it might be plausible to suggest that since all low-slide examples come from oligomers whose backbones are in the A-conformation and all high-slide conformations come from oligomers whose backbones are in the B-conformation, then the backbone must be responsible for the observed conformation. However, it is equally plausible, if not more so, to suggest that since A-backbone oligomers are dominantly made up of homogeneous G|C steps, then it is the tendency of these steps to adopt the strongly negative slide that drives the backbone into the low-slide-compatible A-conformation. What is certain about the role of the backbone is that each backbone conformation is associated with a certain range in the roll/slide/twist conformational space, and that each backbone conformation is persistent (as discussed in § 9).

Now within each subset (low-slide or high-slide) of the homogeneous G|C steps there is no restraint apart from that provided by the backbone. The flexibility of the backbone therefore implies that considerable scatter or otherwise may be seen within each subset in the d_{C1C1} roll and d_{C1C1}/twist plots. For example, while both GG/CC subsets are particularly rigid with respect to roll, the high-slide subset is particularly flexible with respect to helical twist; and the CG clusters are in general flexible with respect to roll and, to a lesser extent, twist. The important point to note is that the roll/slide/twist correlation that is striking in the case of CA/TG is not seen in the bistable steps such as GG/CC. While helical twist varies in GG/CC, slide cannot vary as freely in accordance with the backbone constraint (figure 15*b*). The reason for this is that zero-slide conformation in GG/CC is blocked by the strong electrostatic interactions.

So far we have accounted for the empirically observed conformational characteristics of most of the dinucleotide steps in terms of a simple first-order argument. In fact the only step that we cannot deal with easily in this scheme is AC/GT. Backbone switching can be invoked to account for the rigid low-slide conformation, but the lack of pattern in the B-backbone subset makes it difficult to formulate a hypothesis that accounts for the conformation of AC/GT. It is interesting that this step, in contrast to the homogeneous G|C steps, is characterized by propeller values of both its A-backbone and B-backbone subsets that are somewhere between those of the rigid B-backbone steps (= AA/TT, AT and GA/TC) and the flexible B-backbone steps (= CA/TG and TA): average propeller (and standard deviation) for the B-backbone and A-backbone subsets of AC/GT are $-14.8°(3.6°)$ and $-12.4°(2.9°)$ respectively. Although the average propeller of A-backbone AC/GT steps is generally low, its

marginally higher value (cf. GG/CC, CG and GC) is compatible with the less negative slide seen in the low-slide AC/GT step when compared to the GG/CC step, for example. The slide of AC/GT is negative but not too negative to limit propeller to the values (*ca.* $-10°$) seen in typical A-backbone steps, namely GC, CG and GG/CC.

11. On the propeller/flexibility linkage

We have demonstrated elsewhere (El Hassan & Calladine 1996) that there is an excellent correlation between propeller and flexibility (as measured by the standard deviation of slide: sd(D_y)) with low-propeller accompanying high standard deviations of slide and vice versa. We have also argued in that paper that high-propeller is probably responsible for locking the dinucleotide step 'mechanically' into a near-zero-slide conformation by giving rise to minor or major groove cross-strand clashes if some slide freedom were to be exercised. Furthermore, we have argued that propeller is *not* a free variable that may or may not be deployed depending on the geometry of the particular step in question. Rather, it is strongly influenced by some specific sequence-related effects and it therefore positively dictates the conformations adopted by the various steps: low-propeller allows the relevant steps significant freedom and high-propeller limits this freedom in other steps. Sequence-specific effects on propeller that are consistent with our data include Hunter's thymine methyl clash for AA/TT, AT and AC/GT; Nelson's bifurcated H-bonding for AA/TT (Nelson *et al.* 1987); and, to a lesser extent, the (possible) bifurcated H-bonds at the major-groove side of A-backbone AC/GT steps (Jain *et al.* 1989). Another factor that is also consistent with the empirical findings of our work is the inherent deformability of the A–T base pair with only two Watson–Crick H-bonds as compared to the G–C base pair with three.

The propeller-flexibility correlation (El Hassan & Calladine 1996), which we reproduce by line (1) of figure 18, is given by

$$\mathrm{sd}(D_y) = 0.09\overline{\omega} + 2.01 \qquad (R = 0.95). \tag{11.1}$$

This correlation (11.1) involves *all* steps, including A-backbone steps, B-backbone steps and all homogeneous G|C steps. However, we have just argued that the 'mechanical locking' idea is relevant for B-backbone steps and that it restricts such steps to a narrow region of the range of conformations allowed by the B-backbone. It turns out that the propeller/flexibility correlation for all B-backbone steps (i.e. excluding all A-backbone steps, and all homogeneous G|C steps (both A-backbone and B-backbone)) is given by

$$\mathrm{sd}(D_y) = 0.09\overline{\omega} + 1.84 \qquad (R = 0.92). \tag{11.2}$$

The similarity between the statistics of these two correlations ((11.1) and (11.2)), which can also be seen in figure 18, is striking: there is practically no difference between the two lines. In other words, the homogeneous G|C steps satisfy the same propeller/flexibility relationship as other steps, but for different reasons. The flexibility – or more accurately the bistability – in these steps is due to Hunter's electrostatic interactions, while their low-propeller is consistent with (*a*) the intrinsic predisposition of GC base pairs to low-propeller and (*b*) the need for homogeneous G|C steps to exercise their electrostatically driven bistability. Although the overall correlation statistics (slope, intercept and correlation-coefficient) do not seem to be affected, an interesting change takes place upon elimination of A-backbone data: the correlation begins to assume more 'bistable' nature. In other words, the data tend

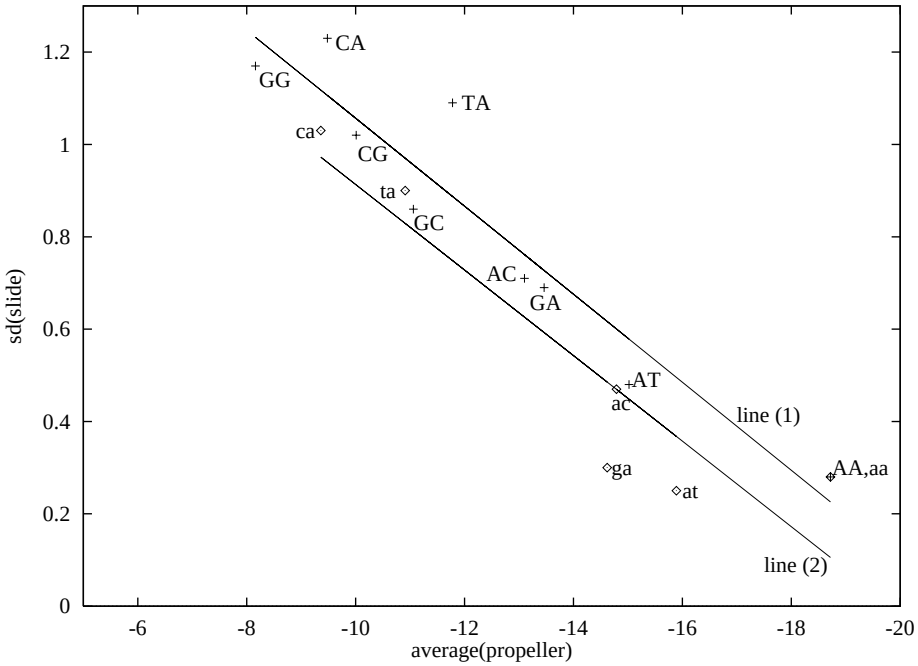

Figure 18. Flexibility, as measured by the standard deviation of slide, *vs* average propeller. 'Upper case' steps (+) and the corresponding best-fit (line (1)) are for all the data taken collectively; while 'lower case' steps (◊) and the corresponding best-fit (line (2)) are for B-backbone data only, excluding all homogeneous G|C steps.

to cluster into two broad regions: high-propeller/low-standard deviation (slide) and low-propeller/high-standard deviation (slide). What happens is that by eliminating the few A-backbone examples of GA/TC, AT and, to a lesser extent, AC/GT, the points corresponding to these steps in the propeller/flexibility plot of figure 18 move closer to the AA/TT point, while the TA and CA/TG points remain more or less unchanged. This tendency for rigid (B-backbone) steps to be associated with high-propeller, and flexible (B-backbone) steps with low-propeller, is clearly consistent with the mechanical locking idea to which we have hitherto appealed in order to account for the rigidity of steps such as AA/TT and AT.

12. Overall classification map

We are now in a position to present an overall classification map that summarizes the discussion that we have presented so far. This scheme is summarized by the roll/twist/slide diagram shown in figure 19. The plot shows the overall conformational space available for all dinucleotide steps, and how different parts of it are mobilized by the nine dinucleotide steps that we have examined. In essence, the figure represents the *intersections* of the regions in the roll/slide/twist conformational space that corresponds to the following three effects. First the existence of a narrow channel along the roll/twist (negative slope) and twist/slide (positive slope) diagrams which is permitted by the B-backbone. Second, the chemical effects such as the special electrostatics of the G–C base pair favour one of two discrete conformational zones in the roll/slide/twist space (Hunter 1993). Third, the mechanical effects, namely the

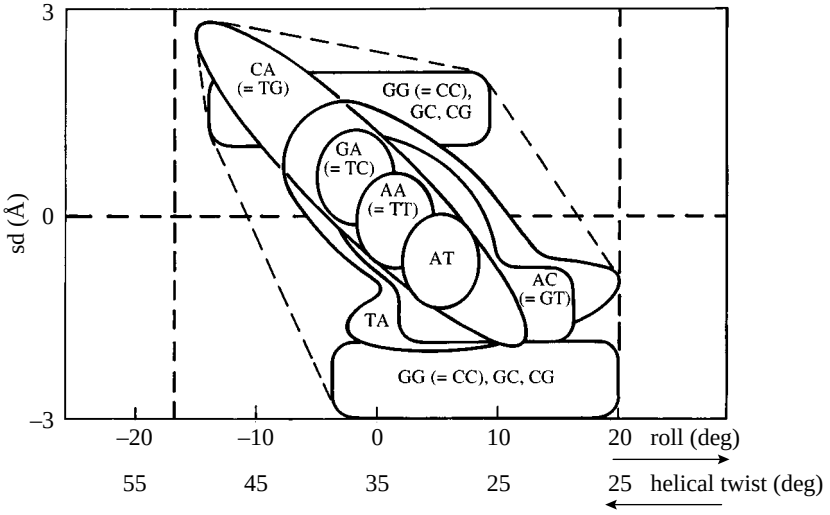

Figure 19. A complete kinematic classification map of all dinucleotide steps except AG/CT.

propeller locking as seen in the rigid steps, e.g. AA/TT, favour a very small region in the roll/slide/twist conformational space near the zero-slide, zero-roll and 36°-twist region.

The CA/TG step occupies a long, thin channel with roll, slide and twist all linearly correlated. TA is similar to CA/TG except that it has a broad low-slide region in addition to the long thin channel; this is clearly indicative of some degree of bistability. The AA/TT step occupies a narrow central region centred about 36° helical twist while the GA/TC and the AT steps occupy similar clusters that are at slightly higher and lower helical twists respectively. The homogeneous G|C steps are shown to occupy two distinct regions at different slide levels, but with very variable roll and twist values. Low-slide AC/GT occupies a region adjacent to the homogeneous G|C low-slide region. The high-slide AC/GT is represented by a relatively wide zone centred at zero-slide, zero-roll and 36°-twist. Note that the relatively smaller size of the region of the low-slide AC/GT and its offset from that of the homogeneous G|C steps, is consistent with the bifurcated H-bonds postulated by Jain *et al.* (1989).

13. General discussion and conclusion

Despite the widespread agreement that DNA molecules possess conformational flexibility to varying extents (Drew *et al.* 1990; Dickerson 1992; Dickerson *et al.* 1994, 1996), and that this flexibility is important for the many functions of the molecule, little progress has been made in understanding the basis of such conformational flexibility, i.e. its relation to the constituent dinucleotide steps. Thus, there is no consensus as to whether one should take a dinucleotide, a trinucleotide, or even a tetranucleotide as the fundamental unit for local structural description of DNA. In this paper we have argued for the dinucleotide step model of structural description; and on this basis we have presented an extensive empirical study of the conformational characteristics of dinucleotide steps in DNA. Our classification has been derived from an extensive database of atomic coordinates of some 60 solved oligomer structures containing 400 dinucleotide steps. We have described the results of this

classification in detail. The map of figure 19 shows the conformational characteristics of all but one step type, AG/CT (due to its extreme under-representation). We have compared our empirical results with the theoretical findings of Hunter (1993); and we have also attempted to account for the observed preferences in terms of a hypothesis based on (i) some of the theoretical results of Hunter (1993), particularly in relation to the electrostatically driven homogeneous G|C steps; (ii) some simple mechanical ideas relating to the role of propeller in determining the rigidity of other steps; and (iii) some simple features of the conformational space permitted by the backbone constraint.

The main features that distinguish our work from that of others (Yanagi *et al.* 1991; Dickerson *et al.* 1994; Gorin *et al.* 1995) lie in our underlying philosophy and general approach to the problem. First, we believe that in order to see a complete and unbiased picture, we need to observe the behaviour of the complete set of available data without any preconceived ideas or convictions regarding the suitability or otherwise of certain subsets within the database. Thus, unlike most workers in the field, we have not, in particular, excluded low-slide, A-backbone DNA. We have also retained dodecamer end-sequences d(CGCG) despite the fact that these regions appear to be subjected to significantly more direct crystal-packing effects than others. Once again, this move stems from our underlying philosophy of taking an unbiased and global view of the database, and regarding the crystalline environment as a test-bed into which the various oligomer sequences are fitted. The fact that certain portions such as dodecamer-end d(CGCG) sequences are subjected to more direct crystal-packing effects than others might warrant a closer examination of these regions; but it does not justify, in our view, their complete elimination from the database. Indeed, as we have shown, the conformational characteristics of these regions enable us to understand, in terms of Hunter's electrostatics of the G–C base pairs, the adoption of positive slide conformations by GC steps.

In this paper we have dealt with a number of issues regarding DNA structure. However, for the sake of continuity, we have deliberately omitted certain aspects of DNA structure; most notably the effect of flanking bases on the conformations of dinucleotide steps. The obvious difficulty in trying to classify structurally longer DNA units (tri- or tetra-nucleotides) is the present lack of data needed to cover all possible trimers and/or tetramers. However, it is noteworthy that, contrary to what has been asserted by many in the field (see, for example, Yanagi *et al.* 1991), we have managed to make significant progress on the basis of a simple dinucleotide model. The key point here is that trinucleotides or longer units of DNA structure description are only needed if the dinucleotide step in question is known to adopt more than one conformation. This immediately excludes all 'rigid' steps, leaving only the flexible and bistable ones. Even then, it might not necessarily be the case that the adopted conformation of a flexible step is dictated only by the flanking bases. Indeed it seems reasonable to suppose that the conformations adopted by flexible steps are dictated by the overall global constraint on the DNA molecule, rather than the immediately adjacent bases. Thus if the oligomer is sharply bent, then this sets up the conformation of the flexible/bistable step as one having the appropriate roll and hence slide and twist values (recall that roll, slide and twist are correlated). A good example of a step that appears to have its conformation influenced by its immediate near-neighbours is GC. As we have already pointed out in § 7, the GC step preceded by GG/CC appears to prefer positive roll in almost all of the available examples. Indeed this tendency has been noted by several previous workers (Calladine & Drew

1986; Brukner *et al.* 1993; Travers 1995). Goodsell *et al.* (1993) have concluded that the GGCC tetramer prefers to bend in a direction that compresses the major groove. Based on the crystal structure of d(CATGGCCATG), they have proposed a more detailed mechanism whereby the bending that compresses the major groove side of this tetramer comes about from high propeller in the two base pairs of the central GC step and more or less zero propeller in the flanking G–C base pairs. This proposed mechanism, at least in its schematic form (see fig. 1 of Goodsell *et al.* 1993), appears to suggest that the central GC step has more or less zero roll and that it therefore plays little part in the observed bending; which comes about instead from the positive roll angles in the flanking GG/CC steps. However, an examination of the roll values of the GG/CC, GC and GG/CC steps in d(CATGGCCATG) (Goodsell *et al.* 1993) reveals that they are 6.5°, 9.5° and 8.4° respectively. In other words, the highest roll value in this particular examples does in fact occur in the central GC step.

Although the present paper has dealt with the problem of the conformation and structure of naked DNA, it is hoped that the findings of this work may be of value in understanding the various structure-related functions of DNA such as the recognition of dimeric proteins. DNA bending and twisting by proteins is a well known phenomenon (Travers 1991, 1993, 1995), and the significance of the deformation of the DNA at certain preferred locations has been recognized since the mid-1980s. Thus Drew & Travers (1985) used statistical sequencing to deduce that the helical repeat of DNA reduces on winding around chicken nucleosome cores. Satchwell *et al.* (1986) deduced that CA/TG steps are capable of lying around the nucleosome in two opposite orientations, with their minor grooves pointing either inwards or outwards from the center of the circular path. More recently, the solution of a number of dimeric DNA/protein complexes has revealed that very often a flexible step, CA/TG or TA, tends to be present either in the directly contacted sites or in the region that bridges the two contacted sites (Aggarwal *et al.* 1988; Mondragon & Harrison 1991; Luisi *et al.* 1991; Schultz *et al.* 1991; Y. Kim *et al.* 1993; J. L. Kim *et al.* 1993; Marmorstein *et al.* 1992; Beamer & Pabo 1992; Schwabe *et al.* 1994; Rodgers & Harrison 1993; Shimon & Harrison 1995). We shall describe the analysis of oligomeric protein-bound DNA elsewhere in some detail; and we shall argue that many of the statements made here about naked dinucleotide conformations, apply equally well to protein-bound DNA.

We thank Dr Chris Hunter, Dr Daniella Rhodes, Dr Andrew Travers and Dr Horace Drew for help and comments. M.A.E. acknowledges financial support (Ph.D. studentship (1991–1994)) by the Cambridge Commonwealth Trust and Jesus College (Cambridge); M.A.E. is grateful for a College Research Fellowship at Peterhouse, Cambridge (1994–1997).

References

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

Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. & Dickerson, R. E. 1993 The crystal structure of the trigonal decamer CGATCG6meATCG: a B-DNA helix with 10.6 base pairs per turn. *J. Mol. Biol.* **231**, 768–784. bdjb48.

Beamer, L. J. & Pabo, C. O. 1992 Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* **227**, 177–196.

Biburger, M., Niederweis, M. & Hillen, W. 1994 Oligo[d(C).(G)] runs exhibit a helical repeat of 11.1 bp in solution and cause slight DNA curvature when properly phased. *Nucleic Acids Res.* **22**, 1562–1566.

Bingman, C. A., Zon, G. & Sundaralingam, M. 1992*a* Crystal and molecular structure of the A-DNA dodecamer d(CCGTACGTACGG): choice of fragment helical axis. *J. Mol. Biol.* **227**, 738–756. adl045.

Bingman, C. A., Jain, S., Zon, G. & Sundaralingam, M. 1992*b* Crystal and molecular structure of the alternating dodecamer d(GCGTACGTACGC) in the A-DNA form: comparison with the isomorphous non-alternating dodecamer d(CCGTACGTACGG). *Nucleic Acids Res.* **20**, 6637–6647. adl046.

Brown, T., Leonard, G. A., Booth, E. D. & Chambers, J. 1989 Crystal structure and stability of a DNA duplex containing A(anti).G(syn) base pairs. *J. Mol. Biol.* **207**, 455–457. DODBAD.

Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. 1986 Molecular structure of the G.A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natn. Acad. Sci. USA* **83**, 2402–2406. DODBAH.

Brukner, I., Diakic, M., Savic, A., Susic, S., Pongor, S. & Suck, D. 1993 Evidence for opposite groove-directed curvature of GGGCCC and AAAAAA tracts. *Nuc. Acids. Res.* **21**, 1025–1029.

Calladine, C. R. 1982 Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* **161**, 343–352.

Calladine, C. R. & Drew, H. R. 1984 A base-centred explanation of the B-to-A transition in DNA. *J. Mol. Biol.* **178**, 773–782.

Calladine, C. R. & Drew, H. R. 1986 Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.* **192**, 907–918.

Calladine, C. R. & Drew, H. R. 1992 *Understanding DNA; the molecule and how it works.* Academic Press.

Cervi, A., Langlois D'Estaintot, B. & Hunter, W. N. 1992 Crystal and molecular structure of d(GTCTAGAC). *Acta Crystallogr.* B **48**, 714–719. adh041.

Conner, B. N., Yoon, C., Dickerson, J. L. & Dickerson, R. E. 1984 Helix geometry and hydration in an A-DNA tetramer: C-C-G-G. *J. Mol. Biol.* **174**, 663–695. addb01.

Corfield, P. W. R., Hunter, W. N., Brown, T., Robinson, P. & Kennard, O. 1987 Inosine-Adenine base pairs in a B-DNA duplex. *Nucleic Acids Res.* **15**, 7935–7949. bdlb10.

Cruse, W. B. T., Aymami, J., Kennard, O., Brown, T., Jack, A. G. C. & Leonard, G. A. 1989 Refined crystal structure of an octanucleotide duplex with I.T mismatched base pairs. *Nucleic Acids Res.* **17**, 55–72. adhb17.

Dickerson, R. E. 1992 DNA structure from A to Z. *Meth. Enzymology* **211**, 67–111.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., von Kitzing, E., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Sklenar, H., Soumpasis, D. M., Tung, C.-S., Wang, A. H.-J. & Zhurkin, V. B. 1988 Definition and nomenclature of nucleic acid structure parameters, 'The Cambridge Accord'. *EMBO. J.* **8**, 1–4.

Dickerson, R. E., Grzeskowiak, K., Grzeskowiak, M., Kopka, M. L., Larsen, T., Lipanov, A., Privé, G. G., Quintana, J., Schultz, P., Yanagi, K., Yuan, H. & Hyo-Chun Yoon 1991 Polymorphism, packing, resolution, and reliability in single-crystal DNA oligomer analysis. *Nucleosides Nucleotides* **10**(1–3), 3–24.

Dickerson, R. E., Goodsell, D. S. & Neidle, S. 1994 '... The tyranny of the lattice...'. *Proc. Natn. Acad. Sci. USA* **91**, 3579–3583.

Dickerson, R. E., Goodsell, D. & Kopka, M. L. 1996 MPD and DNA bending in crystals and in solution. *J. Mol. Biol.* **256**, 108–125.

diGabrielle, A. D., Sanderson, M. R. & Steitz, T. A. 1989 Crystal lattice packing is important in determining the bend of a DNA dodecamer containing an adenine tract. *Proc. Natn. Acad. Sci. USA* **86**, 1816–1820. bdl015/1/2.

diGabrielle, A. D. & Steitz, T. A. 1993 A DNA dodecamer containing an adenine tract crystallises in a unique lattice and exhibits a new bend. *J. Mol. Biol.* **231**, 1024–1039. bdl047/1/2/3.

Drew, H. R. & Travers, A. A. 1985 DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186**, 773–790.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. 1981 Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natn. Acad. Sci. USA* **78**, 2179–2183. DODBAJ.

Drew, H. R., McCall, M. J. & Calladine, C. R. 1990 New approaches to DNA in the crystal and in solution. In *DNA topology and its biological effects* (ed. N. R. Cozzarelli & J. C. Wang), pp. 1–56. Cold Spring Harbor Laboratory Press.

Edwards, K. J., Brown, D. G., Spink N. & Neidle, S. 1992 Molecular structure of the B-DNA dodecamer d(C-G-C-A-A-A-T-T-T-G-C-G)2; an examination of propeller twist and minor-groove water structure at 2.2 Å resolution. *J. Mol. Biol.* **226**, 1161–1173. bdl038.

Egli, M., Usman, N., Zhang, S. & Rich, A. 1992 Crystal structure of an okazaki fragment at 2 Å resolution. *Proc. Natn. Acad. Sci. USA* **89**, 534–538. ahj040.

Egli, M., Usman, N. & Rich, A. 1993 Conformational influence of the ribose 2′-hydroxyl group: crystal structures of DNA–RNA chimeric duplexes. *Biochem.* **32**, 3221–3237. ahj043/44.

Eisenstein, M., Frolow, F., Shakked, Z. & Rabinovich, D. 1990 The structure and hydration of the A-DNA fragment d(GGGTACCC) at room temperature and low temperature. *Nucleic Acids Res.* **18**, 3185–3194. OCTAAR.

El Hassan, M. A. & Calladine, C. R. 1995 The Assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.* **251**, 648–664.

El Hassan, M. A. & Calladine, C. R. 1996 Propeller-twisting of base pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.* **259**, 95–103.

Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. 1982 Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 14686–14707. bdlb04.

Frederick, C. A., Quigley, G. J., Teng, M. K., Coll, M., van der Marel, G. A., van Boom, J. H., Rich, A. & Wang, A. H.-J. 1989 Molecular structure of an A-DNA decamer d(A-C-C-G-G-C-C-G-G-T). *Eur. Jl Biochem.* **181**, 295–307. adj022.

Frederick, C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J. & Rich, A. 1988 Methylation of the Ecori recognition site does not alter DNA conformation. The crystal structure of d(CGCGAm6ATTCGCG) at 2.0 Å resolution. *J. Biol. Chem.* **263**, 17872–17879. bdlb13.

Goodsell, D. S, Kopka, M. K., Gascio, D. & Dickerson, R. E. 1993 Crystal structure of CATG-GCCATG and its implication for A-tract bending models. *Proc. Natn. Acad. Sci. USA* **90**, 2930–2934. bdj051.

Gorin, A. A., Zhurkin, V. B. & Olson, W. K. 1995 B-DNA twisting correlated with base pair morphology. *J. Mol. Biol.* **247**, 34–48.

Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. 1991 The structure of B-helical CGATCGATCG and comparison with CCAACGTTGG. The effect of base pair reversals. *J. Biol. Chem.* **266**, 8861–8883. bdj025.

Hahn, M. & Heinemann, U. 1993 DNA helix structure and refinement algorithm: comparison of models for d(CCAGGCM==5==CTGG) derived from NUCLSQ, TNT, and X-PLOR. *Acta Crystallogr.* D **49**, 468–477. bdjb49.

Haran, T. E., Shakked, Z., Wang, A. H.-J. & Rich, A. 1987 The crystal structure of d(CCCCGGGG): a new A-Form variant with an extended backbone conformation. *J. Biomol. Struct. Dyn.* **5**, 199–217. OCTAAA.

Heinemann, U., Lauble, H., Frank, R. & Bloecker, H. 1987 Crystal structure analysis of an A-DNA fragment at 1.8 Å resolution. d(GCCCGGGC). *Nucleic Acids Res.* **15**, 9531–9550. adh008.

Heinemann, U. & Alings, C. 1989 Crystallographic study of one turn of G/C-Rich B-DNA. *J. Mol. Biol.* **210**, 369–381. bdj017.

Heinemann, U., Rudolph, L. N., Alings, C., Morr, M., Heikens, W., Frank, R. & Bloecker, H. 1991 Effect of a single 3′-methylene Phosphonate linkage on the conformation of an A-DNA octamer double helix. *Nucleic Acids Res.* **19**, 427–433. adhp36.

Heinemann, U., Alings, C. & Bansal, M. 1992 Double helix conformation groove dimensions and ligand-binding potential of a G/C-stretch in B-DNA. *EMBO J.* **11**, 1931–1939. bdj039.

Heinemann, U. & Hahn, M. 1992 CCAGGCm5CTGG; helical fine structure, hydration, and comparison with CCAGGCCTGG. *J. Biol. Chem.* **267**, 7332–7341.bdjb27.

Hunter, C. A. 1993 Sequence dependent DNA structure; the role of base stacking interactions. *J. Mol. Biol.* **230**, 1025–1054.

Hunter, W. N., Kneale, G., Brown, T., Rabinovich, D. & Kennard, O. 1986 Refined crystal structure of an octanucleotide duplex with G.T mismatched base pairs. *J. Mol. Biol.* **190**, 605–618. OCTAAJ.

Hunter, W. N., Brown, T. & Kennard, O. 1987*a* Structural features and hydration of a dodecamer duplex containing two C.A mispairs. *Nucleic Acids Res.* **15**, 6589–6605. DODBAC.

Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D. & Kennard, O. 1987*b* The structure of guanosine-thymidine mismatches in B-DNA at 2.5 Å resolution. *J. Biol. Chem.* **262**, 9962–9970. bdl009.

Hunter, W. N., Langlois D'Estaintot, B. & Kennard, O. 1989 Structural variation in d(CTCTAGAG). Implications for protein-DNA interactions. *Biochem.* **28**, 2444–2451. adh020.

Jain, S., Zon, G. & Sundaralingam, M. 1989 Base-only binding spermine in the deep groove of the A-DNA octamer d(GTGTACAC). *Biochem.* **28**, 2360–2364. adh014.

Kennard, O., Cruse, W. B. T., Nachman, J., Prange, T., Shakked, Z. & Rabinovich, D. 1986 Ordered water structure in an A-DNA octamer at 1.7 Å resolution. *J. Biomol. Struct. Dyn.* **3**, 623–647. adhb11.

Kim, J. L., Nikolov, D. B. & Burley, S. K. 1993 Co-crystal structure of TBP recognising the minor groove of a TATA element. *Nature* **365**, 520–527.

Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. 1993 Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z. & Steitz, T. A. 1979 A hypothesis on a specific sequence-dependent conformation of DNA and its relation to the binding of the *lac*-repressor protein. *J. Mol. Biol.* **131**, 669–680.

Kneale, G., Brown, T., Kennard, O. & Rabinovich, D. 1985 G.T base pairs in a DNA helix. The crystal structure of d(GGGGTCCC). *J. Mol. Biol.* **186**, 805–814. OCTAAK.

Larsen, T. A., Kopka, M. L. & Dickerson, R. E. 1991 Crystal structure analysis of the B-DNA dodecamer CGTGAATTCACG. *Biochem.* **30**, 4443–4449.bdl029.

Lauble, H., Frank, R., Bloecker, H. & Heinemann, U. 1988 Three-dimensional structure of d(GGGATCCC) in the crystalline state. *Nucleic Acids Res.* **16**, 7799–7816. OCTAAG.

Leonard, G.A. & Hunter, W. N. 1993 Crystal and molecular structure of d(CGTAGATCTACG) at 2.25 Å resolution. *J. Mol. Biol.* **234**, 198–208. bdl042.

Leonard, G. A., Thomson, J., Watson, W. P. & Brown, T. 1990 High-resolution structure of a mutagenic lesion in DNA. *Proc. Natn. Acad. Sci. USA* **87**, 9573–9576. bdlb26.

Leonard, G. A., Booth, E. D., Hunter, W. N. & Brown, T. 1992 The conformational variability of an adenosine–inosine base pair in a synthetic DNA dodecamer. *Nucleic Acids Res.* **20**, 4753–4759. bdlb41.

Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. 1993 The structure of the B-DNA decamer CCAACITTGG in two different space groups: conformational flexibility of B-DNA. *Biochem.* **32**, 1373–1389. bdjb43/44.

Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto K. R. & Sigler, P. B. 1991 Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505.

Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. 1992 DNA recognition by GAL4: structure of a protein–DNA complex. *Nature* **356**, 408–414.

McCall, M. J., Brown, T. & Kennard, O. 1985 The crystal structure of d(GGGGCCCC). A model for poly(DG).poly(DC). *J. Mol. Biol.* **183**, 385–396. OCTAAI.

Mondragon, A. & Harrison, S. C. 1991 The phage 434CRO/OR1 complex at 2.5 Å resolution. *J. Mol. Biol.* **219**, 321–334.

Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. 1987 The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **330**, 221–226. DODBAA.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L. & Dickerson, R. E. 1987 Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science* **238**, 498–504. bdj008.

Privé, G. G., Yanagi, K. & Dickerson, R. E. 1991 Structure of B-DNA decamer CCAACGTTGG and comparison with the isomorphous decamers decamers CCAAGATTGG and CCAGGCCTGG. *J. Mol. Biol.* **217**, 177–199. bdj019.

Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. 1992 The structure of a B-DNA decamer with a central T-A step: CGATTAATCG. *J. Mol. Biol.* **225**, 379–395. bdj031.

Rodgers, D. W. & Harrison, S. C. 1993 The complex between phage 434 repression DNA-binding domain and the operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure* **1**, 227–240.

Saenger, W. 1984 *Principles of nucleic acid structure.* Springer.

Satchwell, S. C., Drew, H. R. & Travers, A. A. 1986 Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675.

Schultz, S. C., Shields, G. C. & Steitz, T. A. 1991 Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science* **253**, 1001–1007.

Schwabe, J. W. R., Chapman, L., Finch, J. T. & Rhodes, D. 1994 The crystal structure of the oestrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* **75**, 567–578.

Shakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. 1989 The conformation of the DNA double helix in the crystal is dependent on its environment. *Nature.* **342**, 456–460. OCTAAH.

Shimon, L. J. W. & Harrison, S. C. 1993 The phage 434 OR2/R1-69 complex at 2.5 Å resolution. *J. Mol. Biol.* **232**, 826–838.

Skelly, S. K., Edwards, K. J., Jenkins, T. C. & Neidle, S. 1993 Crystal structure of an oligonucleotide duplex containing G.G base pairs: the influence of mispairing on DNA backbone conformation. *Proc. Natn. Acad. Sci. USA* **90**, 804–808. bdl046.

Takusagawa, F. 1990 The crystal structure of d(GTACGTAC) at 2.25 Å resolution. Are the A-DNA's always unwound approximately 10° at the C-G steps? *J. Biomol. Struct. Dyn.* **7**, 795–809. adh024.

Thota, N., Li, X. H., Bingman, C. & Sundaralingam, M. 1993 High resolution refinement of the hexagonal A-DNA octamer d(GTGTACAC) at 1.4 Å resolution. *Acta Crystallogr.* D **49**, 282–291. adh038.

Travers, A. A. 1991 DNA bending and kinking-sequence dependence and function. *Current Opinion Struct. Biol.* **1**, 114–122.

Travers, A. A. 1993 *DNA-protein interactions.* London: Chapman and Hall.

Travers, A. A. 1995 DNA bending by sequence and proteins. In *DNA–protein; structural interaction* (ed. D. M. J. Lilley), pp 49–75. Oxford: IRL Press.

Wang, A. H.-J., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A. A. & Rich, A. 1982 Molecular structure of r(GCG)d(TATACGC): a DNA–RNA hybrid helix joined to double helical DNA. *Nature* **299**, 601–604. ahj015.

Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P. F., Li, B. F. & Tickle, I. J. 1990 Crystal structure and sequence-dependent conformation of the A.G mispaired oligonucleotide d(CGCAAGCTGGCG). *Proc. Natn. Acad. Sci. USA* **87**, 6693–6697. bdl022.

Xuan, J. C. & Weber, I. T. 1992 Crystal structure of a B-DNA dodecamer containing inosine, d(CGCIAATTCGCG), at 2.4 Å resolution and its comparison with other B-DNA dodecamers. *Nucleic Acids Res.* **20**, 5457–5464. bdlb40.

Yanagi, K., Privé, G. G. & Dickerson, R. E. 1991 Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201–214.

Yoon, C., Privé, G. G., Goodsell, D. G. & Dickerson, R. E. 1988 Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natn. Acad. Sci. USA* **85**, 6332–6336. DODBAF.

Yuan, H., Quintana, J. & Dickerson, R. E. 1992 Alternative structures for alternating poly(dA-dT) tracts: the structure of the B-DNA decamer CGATATATCG. *Biochem.* **31**, 8009–8021. bdj036.

Received 12 September 1995; accepted 2 November 1995