Influence of Sequence on the Conformation of the B-DNA Helix

Juan A. Subirana and Tiago Faria
Departament d'Enginyeria Química, ETSEIB, Universitat Politècnica de Catalunya, Barcelona E-08028, Spain

ABSTRACT We have tried to ascertain whether the variability found in the conformational features of the 10 base steps in B-DNA is mainly due to the flanking sequences or to interactions with the environment. From an analysis of the twist parameter of the base-pair steps available from crystals of oligonucleotides and protein/oligonucleotide complexes, we conclude that in most cases the flanking sequences show little influence: the conformation of a DNA region results from the combination of the independent intrinsic features of each base step (average conformation and intrinsic variability), modulated by their interactions with the environment. Only in some cases (YR steps, in particular CG and CA/TG) does it appear that flanking sequences have an influence on the conformation of the central base step. The values obtained allow an approximation to the parameters expected for repetitive DNA sequences. In particular, it is found that poly[d(AG/CT)] should have a strongly alternating conformation, in agreement with recently reported oligonucleotide structures.

INTRODUCTION

The conformation of individual base steps in B-DNA has been determined by x-ray diffraction of oligonucleotide crystals. The values found in different structures vary around an average value that is different for each of the 10 base steps (Gorin et al., 1995; Subirana et al., 1995). The purpose of this paper is to find out whether the particular conformation found for each base step in a given DNA molecule is influenced by the flanking sequence or is due mainly to the interactions of the DNA molecule with its environment. In particular, we want to ascertain whether the effect of sequence can be separated from crystal lattice effects. In fact, we have shown (Subirana et al., 1995) that neighbor base pairs do have an influence on the presence of either the BII or the sugar C3′-endo conformations. For example, the cytosine sugar in the GCA sequence has a strong tendency toward the A-like C3′-endo conformation.

Some time ago, Yanagi et al. (1991) had already suggested that base step geometry cannot be correlated solely with the bases that define the step in question; the two flanking steps also must be taken into account. However, only a few of the 136 four-base steps (or tetrads) were represented in the structures available to Yanagi et al. (1991). At present, more data have accumulated, but still they are not sufficient to analyze all 136 possible tetrads. Nevertheless, enough data have been obtained on some of them, and we think it is of interest to determine if any conclusion can be drawn from such study. In particular, is the tetrad approach valid, or is there no correlation between the conformations of consecutive base steps? Is it worthwhile to study additional oligonucleotide structures to obtain suitable data for all 136 tetrads? In fact, we are very far from such a goal; about 10 times more oligonucleotide crystal structures than those available should be determined for a complete knowledge of all tetrads in the B-form.

In this paper we will assess the conformational variability of the 10 individual base pair steps in B-form DNA as a function of flanking sequences. We will use mainly the value of twist for this purpose (Dickerson et al., 1989). The availability of a significant number of oligonucleotide structures, crystallized either alone or in association with proteins, allows a statistical approach to this problem. Recently other groups have presented several statistical studies on the conformational features of the different base steps. Gorin et al. (1995) have focused their analysis on the geometrical features that determine the morphology of dimer base pair steps. They have found a very interesting twist-clash correlation for the 10 base-pair steps. However, no attempt was made to determine the influence of the surounding DNA sequence. Other studies (Suzuki and Yagi, 1995; Young et al., 1995) have been mainly directed at analyzing the contribution of the different base steps to overall bending in DNA. Our approach is different from and complementary to those studies. We will show that neighbor bases do not appear to have much influence on the conformation of individual base steps, with some clear exceptions. The results obtained may help make it possible to analyze the intrinsic conformational variability of different sequences and predict the average structure of simple repetitive DNA sequences. Our study is aimed at understanding the conformational features of B-form DNA. Major distortions induced by protein or by interactions in the crystal (in particular, end-to-end interactions) will not be considered. In any case, it should be noted that in single crystals of oligonucleotides, there are interactions between neighbor molecules that are of the same nature (hydrogen bonds, ionic interactions) as those found in DNA in solution and when associated with proteins. Therefore their fine structural details should be useful for understanding the variability of DNA structure in biological systems.

Received for publication 19 November 1996 and in final form 26 March 1997.

Address reprint requests to Dr. Juan A. Subirana, Departament d'Enginyeria Química, ETSEIB, Universitat Politècnica de Catalunya, Av. Diagonal 647, Barcelona E-08028, Spain. Tel.: 34-3-4016688; Fax: 34-3-4017150; E-mail: subirana@eq.upc.es.

0006-3495/97/07/333/06 $2.00

CHOICE OF PARAMETERS

The values of the torsion angles are not very useful for visualizing the overall oligonucleotide conformation. Some of these angles are correlated among themselves, as was shown by Fratini et al. (1982). In the present analysis we use the twist angle ω between two consecutive base pairs to represent the most basic feature of DNA structure. Furthermore, its value is rather accurate, because it does not depend much on the method of crystallographic refinement (Hahn and Heinemann, 1993). As an additional parameter, in some cases we have used rise, *Dz*, which gives an approximate measure of the local pitch of the helix. The value of both parameters has been calculated with NEWHELIX91.

The coordinates of the structures analyzed have been obtained from the Nucleic Acids Data Base (Berman et al., 1992) and are listed in Tables 1 and 2. (Biographical references, coordinates, and other details are available on the World Wide Web (http://ndbserver.rutgers.edu).) Only structures with a resolution better than 3.0 Å have been used. Terminal base steps are not included in the calculations. A few base steps are strongly distorted as a result of protein/DNA interactions and give rise to prominent bends. They have been excluded from the calculations by the use of partial helical axes, as indicated in Table 2. A change in helical axis was introduced when it was visually detected by inspection of the graphic appearance of the whole molecule. The presence of a bend was confirmed from the values of the base-pair normal cosines, as calculated with NEWHELIX91. The number of anomalous base-pair steps excluded from the calculations represents only ~4% of the total values derived from the sequences given in Table 2.

TABLE 1 Oligonucleotide structures analyzed

NDB code	Sequence
BDL001	CGCGAATTCGCG
BDL002	CGCGAATTCGCG
BDLB03	CGCGAATTCGCG
BDL006	CGCAAAAAAGCG
BDL007	CGCATATATGCG
BDLB13	CGCGAATTCGCG
BDL015-1	CGCAAAAATGCG
BDL015-2	CGCAAAAATGCG
BDJ017	CCAGGCCTGG
BDL018	ACCGGCGCCACA
BDJ019	CCAACGTTGG
BDJ025	CGATCGATCG
BDJB27	CCAGGCCTGG
BDL028	CGTGAATTCACG
BDL029	CGTGAATTCACG
BDJ031	CGATTAATCG
BDJ036	CGATATATCG
BDJ037	CGATATATCG
BDL038	CGCAAATTTGCG
BDJ039	CCGGCGCCGG
BDLB40	CGCIAATTCGCG
BDL042	CGTAGATCTACG
BDJB43	CCAACITTGG
BDJB44	CCAACITTGG
BDL047-1	CGCGAAAAAACG
BDL047-2	CGCGAAAAAACG
BDL047-3	CGCGAAAAAACG
BDJB48	CGATCGATCG
BDJ051	CATGGCCATG
BDJ052	CCAAGCTTGG
BDJ055	CCATTAATGG
BDJ060	CTCTCGAGAG
BDJ061	CCACTAGTGG
BDL070-1	CGCTCTAGAGCG
BDL070-2	CGCTCTAGAGCG
BDH071	CGCTAGCG

The coordinates have been obtained from the Nucleic Acid Database (Berman et al., 1992). In the case of BDH071, only one molecule has been analyzed, because the three molecules in the asymmetrical unit are related by pseudosymmetry. In the two sequences that contain inosine (I), it has been counted as a guanine (G). Three sequences have been crystallized with a single strand in the asymmetrical unit: BDJ017, BDJ019, and BDJB44. Identical sequences in these three duplexes, which are related by crystallographic symmetry, have not been counted.

LIMITATIONS OF THE AVAILABLE DATA

The average value and standard deviation of twist for all of the base steps studied are given in Table 3. The values reported are similar to those obtained by Gorin et al. (1995), who used a different method for calculation and a slightly different data set. It is interesting to note that most base steps (AT, AC/GT, and AG/CT) that start with A, except AA/TT, have a low average twist, ~32°.

In general, the oligonucleotide crystals have been studied to a higher resolution than the DNA/protein complexes, so that these data should be expected to be more accurate, although significant errors are certainly present, as discussed in detail by different authors (Westhof, 1987; Hahn and Heinemann, 1993; Subirana et al., 1995). On the other hand, the constraints of the lattice restrict the variability in oligonucleotide crystals (Dickerson et al., 1994) when compared with DNA/protein complexes that have been crystallized in a large variety of unit cells. There is certainly a need to study oligonucleotide structures in new crystallographic environments.

As an example of the comments made in the previous paragraph, the structure of the complex of *Pru*II endonuclease with DNA (PDE017 in Table 2) may be considered. In this case (Cheng et al., 1994) the DNA is considerably untwisted, a feature that cannot be achieved in the space groups usually found in either DNA decamers or dodecamers. In other cases, approximations are introduced during refinement of the structure of the complexes that limit the accuracy of the results. For example, in the case of the glucocorticoid receptor (PDRCO1 in Table 2), the sugar pucker was restrained to the C2′-endo conformation (Luisi et al., 1991), whereas it is known that in ~30% of cases the C3′-endo conformation is present (Subirana et al., 1995). Similar simplifications have been introduced in other structures, but in many cases the details of the refinement process of the DNA parameters have not been reported.

From a statistical point of view, a striking feature of the data shown in Table 3 is that the standard deviation for protein/DNA complexes is rather uniform in all cases, whereas it varies more in the data obtained from oligonucleotide crystals. We tentatively interpret this behavior as

TABLE 2 DNA/Protein complexes analyzed

NDB code	Sequence
PDE001	CGCGA/AT/TCGCG
PDE002	CGAGCTCG
PDE006	GATCGC
PDE009	GTTTTTGATAAGA
PDE011	CGGCGCC
PDE017	GACCAG/CTGGTC
PDR001	AGTAC/AAACTTTCTT/GTAT
PDR004	AGTAC/AAACTTTCTTGTAT
PDR006	ACACTTTTC
PDR009	GTACT/AGTTAACT/AGTAC
PDR010	ATACC/ACTGGCGGT/GATAT
PDR011	ATACAATGTATCTTGTTT
PDR012	ATAGT(A)GAGTGCTT(C)TATCAT
PDR013	AGCGTACT/AGTACGCT
PDR015	AGTACAGTTTTTCT/TGTAT
PDR016	ATACC/ACTGGCGGT/GATAT
PDRC01	CAGAACATCGATGTTCTG
PDT002	TCCTATGACTCATCCA
PDT003	CCGGAGGACAGTCCTCCGG
PDT004	TTTGCCATGTAATTACCTAA
PDT005	CATGTAATTCATTTACACGC
PDT006	GCGTGGGCGT
PDT007	GGAGATGACGTCATCTCC
PDT011	TAATAAGGATAACGTCCG
PDT016	TCAACAGCTGTTGA
PDT019	GTAT(G)CAAAT(A)AGG
PDV001	CGACC/GACGTC/GGTCG

The coordinates have been obtained from the Nucleic Acid Data Base (NDB; Berman et al., 1992). The dashes and parentheses in the sequence indicate regions in which a distortion of the helix, usually a bend, was present. Sequences between dashes and parentheses were analyzed independently as described in the text. Bases in parentheses were included in both sequences, at its right and left. Terminal bases not included in our analysis are not shown in the table.

being due to a lower accuracy of the protein/DNA data, which are subjected to random errors, whereas the different standard deviations of oligonucleotide data may reflect the true differences in intrinsic variability of each base step. These considerations indicate that in general the oligonucleotide data are more accurate. However, we have included in the tables the data obtained from protein/DNA complexes for comparative purposes.

Another limitation of the available data is that many structures in the data set are related. The values obtained from them will give some bias to the average values, particularly in those cases in which similar sequences have been studied in similar environments. In the case of most dodecamer crystals, the two halves of the molecule have identical sequences and show a different conformation. However, it is not feasible to "measure" how similar the environments of the two halves of the molecule are.

A comparison of the data obtained for oligonucleotides and for protein/DNA complexes presented in Table 3 shows that in most cases the average values are similar for the two sets of data. However, there are three exceptions in which the difference between average values is about one standard deviation. In the case of the GG/CC step, the difference is probably due to the fact that only a few data are available for oligonucleotide crystals. Because this step exhibits considerable variability, it is likely that the data obtained for oligonucleotide crystals have sampled only a limited region of the conformational space available. In the case of the GC and CG steps, most of the values obtained for oligonucleotides correspond to the second and third base steps starting from the ends of dodecamers. Because all dodecamers pack in similar lattices, this will introduce a bias to the oligonucleotide data, which again will sample only a region of the available conformational space. For these two steps the average twist obtained for protein/DNA complexes might be more representative. In fact, the average values are then more consistent with the low average twist of RY steps and the higher average twist for YR steps. On the other hand, Gorin et al. (1995), through an analysis of steric clashes, justified the oligonucleotide data with a high twist for the GC step and a low twist for the CG step. Taking all of this information into account, we must conclude that the reason for the discrepancy between the protein/DNA and oligonucleotide data for these two steps remains unclear.

THE INFLUENCE OF SEQUENCE ON TWIST

The values given in Table 3 show that individual base steps do clearly differ in their average twist, as already shown by Gorin et al. (1995) and Subirana et al. (1995). They also have a different intrinsic variability, as measured by their standard deviations, also given in Table 3. It should be noted that the base steps with an intermediate variability are those for which fewer data are known, so that their position in the table may change as more data become available.

We will now study the influence of the neighbor sequence on each base step. Table 4 is a list of the average values of twist of the central step in the 15 tetrads present in oligonucleotide crystals for which more than four examples have been studied. In Table 5 the 27 cases available from protein/DNA complexes are also presented. Because there are 136 possible tetrads, it is obvious that the tables are not at all complete; many more data are needed for a full understanding of DNA variability. Inspection of the tables indicates that only in two cases, CAAA and AAAA, are enough data available in both data sets (Tables 4 and 5). In these two cases the average values observed are practically identical.

The RY and RR steps

In all purine-pyrimidine (RY) and purine-purine steps (RR), there is little influence of neighbor sequences. The twist behavior is mainly determined by the central base steps. Each of them has a characteristic average value. On the other hand, the standard deviations of the values shown in Table 4 may change when more data become available. Most of the tetrads with RY and RR steps occur in similar regions of the oligonucleotide in the crystal, so that their variability is probably restricted as a result of similar packing interactions.

TABLE 3 Average values of twist in B-DNA base steps

	Protein/DNA		DNA		
Step	N	Twist (°)	N	Twist (°)	Intrinsic variability
RY	109	32.1 (3.8)	83	34.4 (4.1)	
GC	17	34.2 (3.8)	36	37.5 (2.7)	−
AC/GT	69	32.0 (3.5)	17	32.4 (4.8)	+
AT	23	30.6 (3.9)	30	32.0 (2.8)	−
RR/YY	172	34.6 (4.0)	129	35.7 (4.1)	
GA/TC	41	36.4 (3.2)	41	38.4 (3.1)	−
AA/TT	55	35.9 (3.1)	57	35.3 (3.2)	−
GG/CC	25	32.4 (4.8)	10	35.5 (4.6)	+
AG/CT	51	32.9 (4.0)	21	31.6 (4.3)	+
YR	95	37.8 (4.4)	67	37.1 (7.4)	
TA	38	38.6 (3.6)	14	40.3 (4.8)	+
CG	19	38.0 (3.5)	24	33.6 (5.6)	++
CA/TG	38	36.9 (5.3)	29	38.5 (8.6)	++

N = Number of base steps considered. The intrinsic variability is determined from the standard deviations of twist for oligonucleotides (−, $\sigma < 3.5$; +, $3.5 < \sigma < 5.0$; ++, $\sigma > 5.0$). The choice of limits (3.5, 5.0) is based on the fact that the less variable base pairs have σ values between 2.7 and 3.2, the intermediate base pairs lie between 4.3 and 4.8, whereas the most variable cases have σ values 5.6 and 8.6. Thus it appears that there are three categories: low variability ($\sigma = 3 \pm 0.3$), moderate variability ($\sigma = 4.5 \pm 0.3$), and high variability ($\sigma > 5.5$).

The YR steps

In the case of the pyrimidine-purine steps (YR), the tetrad sequence appears to have an influence on the average twist of the central step, although the large standard deviation observed in some cases indicates that the same sequence may vary considerably in conformation when placed in different environments.

It appears that the high variability of these steps is mainly determined by the smaller degree of stacking between consecutive base pairs. In particular, a high twist (43°-53°) is only available in YR steps. No influence of the surrounding sequence has been detected in the TA steps, whereas a large influence appears to be present in the CG steps, as shown in Tables 4 and 5. Unfortunately, only two cases with a central CG step are available, and this fact prevents further analysis. In any case, it should be noted that the large difference in the average values for this step for both sets of data (Table 3) may be influenced by the fact that different tetrad sequences have been studied in the two sets.

The CA/TG step is of particular interest, because the data shown in Tables 4 and 5 indicate a strong influence of the flanking sequence. To visualize this more clearly, the twist values of this step have been represented versus rise, as shown in Fig. 1. A very high correlation coefficient between rise and twist ($r = 0.91$) is found for the data from oligonucleotides for this step, whereas the values for DNA/protein complexes show a large scatter ($r = 0.22$, data not shown), probably because of their lower accuracy, as discussed above. Plots similar to that shown in Fig. 1 could be drawn for the other base steps. In general, they have a greater scatter and lower correlation coefficients, but a correlation between rise and twist is apparent in most cases. Fig. 1 is presented here because it clearly shows the influ-

TABLE 4 Average values of twist in B-DNA tetrads in oligonucleotide crystals

Step	N	Twist (°)
RY	83	34.4 (4.1)
CGCG	13	38.2 (2.3)
CGCA/TGCG	10	38.2 (1.5)
CGCT/AGCG	7	36.7 (2.1)
GATC	5	33.5 (1.2)
AATT	8	32.1 (1.4)
AACG/CGTT	7	31.3 (2.3)
RR/YY	129	35.7 (4.1)
CGAA/TTCG	13	39.0 (2.4)
CGAT/ATCG	14	38.3 (3.6)
AGAG/CTCT	6	36.2 (2.2)
CAAA/TTTG	5	36.2 (3.7)
GAAT/ATTC	14	36.1 (1.4)
AAAA/TTTT	17	36.0 (1.7)
TAGA/TCTA	6	30.7 (3.8)
YR	67	37.1 (7.4)
ATAT	6	40.9 (2.6)
TCGC/GCGA	13	30.9 (3.0)
CCAA/TTGG	6	45.4 (6.6)
GCAA/TTGC	5	30.0 (2.6)

N = Number of structures available. Only cases with $N > 4$ are given in the table. The average values in the central base step are given, with the standard deviations in parentheses.

TABLE 5 Average values of twist in B-DNA tetrads from protein/DNA complexes

Step	*N*	Twist (°)
RY	109	32.1 (3.8)
CGCC/GGCG	5	35.1 (4.6)
AACT/AGTT	7	34.4 (0.5)
GACG/CGTC	5	34.3 (2.5)
TACT/AGTA	12	33.4 (2.6)
CACT/AGTG	5	32.2 (1.3)
AACA/TGTT	7	32.0 (3.0)
TACA/TGTA	11	30.5 (2.6)
TATC/GATA	7	28.7 (3.1)
RR/YY	172	34.6 (4.0)
TAAC/GTTA	6	38.3 (2.0)
GAAA/TTTC	5	37.5 (1.6)
AAAA/TTTT	6	37.3 (2.7)
AAAC/GTTT	5	37.0 (2.7)
TAAT/ATTA	5	36.4 (2.4)
CAAA/TTTG	5	35.1 (3.3)
AGAA/TTGT	6	36.2 (3.6)
AAGA/TCTT	5	34.7 (2.7)
TAGT/ACTA	13	32.7 (3.1)
AGGA/TCCT	5	32.0 (4.4)
YR	95	37.8 (4.4)
ACGC/GCGT	7	40.7 (2.9)
GTAC	11	40.6 (2.6)
TTAT/ATAA	6	39.4 (1.8)
ATAC/GTAT	8	39.0 (2.6)
TCAA/TTGA	5	42.1 (3.0)
ACAT/ATGT	5	38.3 (2.6)
ACAG/CTGT	6	36.6 (2.0)
TCAT/ATGA	5	32.7 (4.9)

N = Number of structures available. Only cases with $N > 4$ are given in the table. The average values in the central base step are given, with the standard deviations in parentheses.

ence of the surrounding sequence on the values available for the CA/TG steps. Some sequences, like GCAA, prefer low twist, whereas CCAA favors high twist, as shown in Table 4. However, it should be noted that all CCAA sequences available correspond to the 5′-terminal region of decamers, whereas all GCAA sequences correspond to nucleotides 2–5 in dodecamers, so that the different behaviors of these two tetrads might also be influenced by the different packing environments of these two sequences. On the other hand, the tetrads available in protein/DNA complexes (Table 5) also show an influence of the flanking sequence on the average twist of the CA/TG step. Again, it might be argued that different protein environments may have an influence on this behavior, a question that we plan to investigate. In summary, the data presented in Fig. 1 and in Tables 4 and 5 indicate that in the CA/TG step there is an influence of the flanking sequence on its conformation.

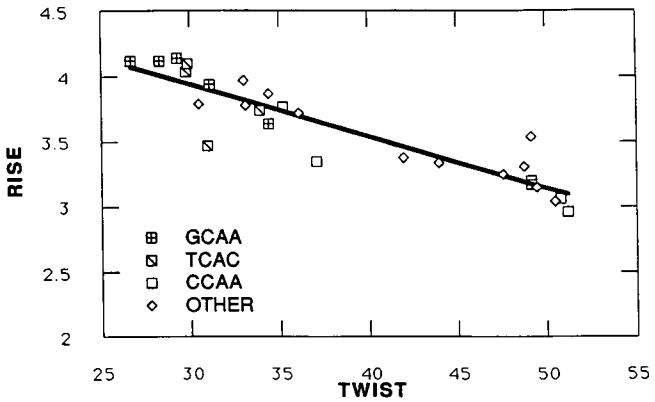

FIGURE 1 Rise/twist values for the CA/TG step in oligonucleotide crystals, showing the influence of neighbor bases.

The CTAG tetrad

This tetrad has recently been studied in detail (Urpí et al., 1996). It has not been included in Tables 4 and 5, because there are fewer than five known data points in either set. The average value of twist for the TA step is 44.0° (σ = 3.4°), including data from a decamer (Shakked et al., 1994), a complex with the *met* repressor (Somers and Phillips, 1992), one octamer, and two dodecamers (Urpí et al., 1996). These values are consistent with the high average twist value for this step found in other cases (Tables 3–5), so that it may be concluded that this is a conventional TA step, with little influence of the flanking C, G bases. However, the twist value for TA in the same sequence in the *trp* repressor/operator complex is only 31.2° (Otwinowski et al., 1988), quite different from the values mentioned above. We may conclude that the statistical analysis that we have presented gives average values and trends, but exceptions will be found upon interaction with proteins, as in this case (Otwinowski et al., 1988). An even more spectacular distortion of the TA steps is found in complexes of the TATA-binding proteins (Juo et al., 1996).

Poly [d(AG/CT)]

The sequence poly[d(AG/CT)] is particularly interesting, because two new oligonucleotide structures have recently been solved that contain such sequences: d(CTCTC-GAGAG) (Goodsell et al., 1995) and d(CGCTCTA-GAGCG) (Urpí et al., 1996). The average values for the AG/CT and GA/TC steps given in Table 3 were mainly obtained from isolated steps embedded in oligonucleotides of mixed sequence. Practically identical average values are obtained when the two alternating sequences just mentioned are omitted. From the calculated average values it would be predicted that poly[d(AG/CT)] should have an alternating structure, with high twist/low rise in the GA step and an opposite behavior in the AG step. Interestingly, this is the

dominant feature in the two recently determined oligonucleotide structures mentioned above. Leslie et al. (1980) also reported that poly[d(AG/CT)] could form fibers with a dinucleotide repeat. Thus the values given in Table 3 have a predictive value in this case. In such a conformation the low twist of the AG step appears to be stabilized by three-center hydrogen bonds in the narrow groove (Tereshko et al., 1996). This peculiar conformation may be essential for recognition by the GAGA factor (Granok et al., 1995).

DISCUSSION

The available data clearly demonstrate that no base step is rigid—all show a considerable structural variability, as measured by the standard deviation of twist (Table 3). Such an observation was already reported by Gorin et al. (1995) and Subirana et al. (1995). Similar DNA sequences may show differences in structure as a function of their interactions with neighbor molecules. However, the intrinsic variabilities of the various base steps are different, as also shown in Table 3.

With regard to the main objective of this work, our analysis shows that flanking sequences have only a small influence on the behavior of most individual base steps. Only in the YR base steps, particularly in the CG and CA/TG cases, are there indications that flanking sequences do influence their conformational parameters. The dual conformation of the CA/TG step, analyzed by Gorin et al. (1995), may depend on the flanking sequences. We conclude that the variability of individual base steps in B-DNA is due mainly to the interactions with neighbor molecules; the flanking sequence has an influence only in some cases. Furthermore, it is obvious that many more data are needed to understand the conformation of all tetrad sequences. In particular, it would be of interest to obtain more data on the YR sequences to better understand their high variability.

Finally, we should mention that the conclusions we have derived from the twist values could have also been derived by the use of other DNA conformational parameters, because many of them show a high correlation with twist (Yanagi et al., 1991). In particular, Gorin et al. (1995) have found a high correlation betwen twist and roll for the YR steps, which is similar to that presented in Fig. 1 for twist and rise. This is not surprising, because rise and roll are geometrically related and have been found to be correlated (Yanagi et al., 1991).

We are grateful to Dr. L. Malinina, Dr. J. Bravo, and Mr. X. Salas for discussions and advice, and to Dr. R. Dickerson for providing NEWHELIX91.

This work has been supported in part by DGICYT grant PB93-1067. TF was a participant in the CENET Erasmus program (94-NL-1067/13).

REFERENCES

Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751–759.

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

Cheng, X., K. Balendiran, I. Schildkraut, and J. E. Anderson. 1994. Structure of *Pvu*II endonuclease with cognate DNA. *EMBO J.* 13: 3927–3935.

Dickerson, R. E., D. S. Goodsell, and S. Neidle. 1994. ". . . the tyranny of the lattice. . . " *Proc. Natl. Acad. Sci. USA.* 91:3579–3583.

Dickerson, R. E., M. Sansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C.-S. Tung, A. H.-J. Wang, and V. B. Zhurkin. 1989. Definitions and nomenclature of nucleic acid structure parameters. *EMBO J.* 8:1–4.

Fratini, A. V., M. L. Kopka, H. R. Drew, and R. E. Dickerson. 1982. Reversible bending and helix geometry in B-DNA dodecamer: CGCGAATT^{Br}CGCG*. *J. Biol. Chem.* 257:14686–14707.

Goodsell, D. S., K. Grzeskowiak, and R. E. Dickerson. 1995. Crystal structure of C-T-C-T-C-G-A-G-A-G. Implications for the structure of the Holliday junction. *Biochemistry.* 34:1022–1029.

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

Granok, H., B. A. Leibovitch, C. D. Shaffer, and S. C. R. Elgin. 1995. Ga-ga over GAGA factor. *Curr. Biol.* 5:238–241.

Hahn, M., and U. Heinemann. 1993. DNA helix structure and refinement algorithm: comparison of models for d(CCAGGCm^{5}CTGG) derived from *NUCLSQ*, *TNT* and *X-PLOR*. *Acta Crystallogr.* D49:468–477.

Juo, Z. S., T. K. Chiu, P. M. Leiberman, I. Baikalov, A. J. Berk, and R. E. Dickerson. 1996. How proteins recognize the TATA box. *J. Mol. Biol.* 261:239–254.

Leslie, A. G. W., S. Arnott, R. Chandrasekaran, and R. L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* 143:49–72.

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

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

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

Somers, W. S., and S. E. V. Phillips. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature.* 359:387–393.

Subirana, J. A., X. Salas, L. Urpí, E. Font, and N. Verdaguer. 1995. Aportaciones de la cristalografía al estudio del ADN. In Nuevas tendencias en cristalografía. F. H. Cano, C. Foces-Foces, and M. Martínez-Ripoll, editors. Ediciones CSIC, Madrid. 223–240.

Suzuki, M., and N. Yagi. 1995. Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.* 23:2083–2091.

Tereshko, V., L. Urpí, L. Malinina, T. Hunyh-Dinh, and J. A. Subirana. 1996. Structure of the B-DNA oligomers d(CGCTAGCG) and d(CGCTCTAGAGCG). *Biochemistry.* 35:11589–11596.

Urpí, L., V. Tereshko, L. Malinina, T. Huynh-Dinh, and J. A. Subirana. 1996. Structural comparison between the d(CTAG) sequence in oligonucleotides and *trp* and *met* repressor-operator complexes. *Nature Struct. Biol.* 3:325–328.

Westhof, E. 1987. Re-refinement of the B-dodecamer d(CGCGAATTCGCG) with a comparative analysis of the solvent in it and in the Z-hexamer d(5BrCG5BrCG5BrCG). *J. Biomol. Struct. Dyn.* 5:581–600.

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

Young, M. A., G. Ravishanker, D. L. Beveridge, and H. M. Berman. 1995. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* 68:2454–2468.