Volume 10 Number 3 1982 Nucleic Acids Research

The ten helical twist angles of B-DNA

W.Kabsch, C.Sander and E.N.Trifonov[†]

Abteilung Biophysik, Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 6900 Heidelberg 1, FRG

Received 27 October 1981; Revised and Accepted 21 December 1981

ABSTRACT

On the assumption that the twist angles between adjacent base-pairs in the DNA molecule are additive a linear system of 40 equations was derived from experimental measurements of the total twist angles for different pieces of DNA of known sequences. This system of equations is found to be statistically consistent providing a solution for all ten possible twist angles of B-DNA by a least squares fitting procedure. Four of the calculated twist angles were not known before (τ_{AC}, τ_{AG}, τ_{CA}, τ_{TA}). The other six twist angles calculated are very close to the experimentally measured ones (τ_{AA}, τ_{AT}, τ_{CC}, τ_{CG}, τ_{GA}, τ_{GC}). The data used were obtained by the electrophoretic band-shift method (1-3), crystallography (4) and nuclease digestion of DNA adsorbed to mica or Ca-phosphate surface (5,6). The validity of the principle of additivity of the twist angles implies that the angle between any particular two base-pairs is a function of only these base-pairs, idependent of nearest neighbours.

INTRODUCTION

Recent experimental measurements of the twist angles between A•T base-pairs in poly(dA)•poly(dT) (2,3,5) and G•C base-pairs in poly(dG)•poly(dC) (2) strongly indicate that the twist angles for different combinations of base-pairs in the DNA molecule are not identical. Another indication is provided by the structure of the selfcomplementary dodecanucleotide 5'-CGCGAATTCGCG-3' (4).

A DNA molecule can be considered as a linear array of 10 possible stereochemically different combinations of adjacent base-pairs. The total twist angle can be measured for any particular piece of DNA by one of two experimental approaches available. The total twist angles for 33 short fragments of DNA with different nucleotide sequences were estimated recently (1-3) by the electrophoretic band-shift method developed by J. Wang (1). The experiments were designed specifically to measure the average pitch (and therefore the twist angle) of natural DNA and of the polynucleotides poly(dA)•poly(dT) (2,3), poly(dG)•poly(dC) (2) and poly(dAT)•poly(dAT) (3). Similarly, nuclease

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

0305-1048/82/1003-1097$2.00/0

digestion of DNA adsorbed on mica or Ca-phosphate surfaces (7) was used to measure the pitch of natural DNA and of the polynucleotides poly(dA)·poly(dT) (5), poly(dAT)·poly(dAT) (5) and poly(dGC)·poly(dGC) (6). The values obtained by this method agree within experimental error with the electrophoretic measurements. From these experiments, the twist angles for 6 different combinations of base-pairs have been estimated so far * : τ_{AA}(2-5), τ_{AT}(4), τ_{CC}(2), τ_{CG}, τ_{GA} and τ_{GC}(4). However, assuming that the twist angles are additive, the data (1-6) contain enough information to determine all 10 twist angles. For each of the experiments, an equation can be written (40 in all) relating the 10 twist angles as unknowns to the measured total twist. Solving this overdetermined set of equations by a least squares procedure, one obtains a consistent set of values for all 10 angles, including the 4 previously unknown.

METHOD OF CALCULATION

The linear system of 40 equations with 10 unknowns as shown in Table I was solved by minimizing the sum of squared residuals. The errors of the 10 angles thus obtained were estimated by normally distributed variation of the right-hand sides τ_e according to the standard deviations σ_e .

To evaluate the consistency of the system the residuals of each equation were calculated in units of their standard deviations (see column $(\tau_e - \tau_c)/\sigma_e$ of Table I).

THE EQUATIONS

The quantities which enter the equations (Table I) are: the dinucleotide compositions of the DNA fragments (columns AA to TA); the average twist angle measured τ_e; and the experimental standard deviation σ_e of τ_e . Eqns. 1-32 are derived from the electrophoretic band-shift method (1-3), eqns. 33-37 from crystallography (4) and eqns. 38-40 from nuclease digestion of DNA adsorbed on mica, Ca-phosphate or Mg-phosphate surfaces (5,6). Fig. 1 illustrates how the coefficients of eqn. 30 are derived from the original data.

* The dinucleotides AA (5'-AA-3' on one strand) and TT (on the opposing strand) together form two stacked A·T base pairs, so that $\tau_{AA} = \tau_{TT}$. Similarly, $\tau_{AC} = \tau_{GT}$, $\tau_{AG} = \tau_{CT}$, $\tau_{CA} = \tau_{TG}$, $\tau_{CC} = \tau_{GG}$, and $\tau_{GA} = \tau_{TC}$. Therefore, the ten combinations given in Tables I and II exhaust all different possibilities.

Table I. System of equations for determination of individual twist angles

Eq.No	AA	AC	AG	AT	CA	CC	CG	GA	GC	TA	τ_e	σ_e	τ_c	$\tau_e-\tau_c$	$\frac{\tau_e-\tau_c}{\sigma_e}$	Components of σ_e: σ_1	σ_2	σ_3	Plasmids compared	Ref
1	7	8	5	6	7	9	2	5	2	6	33.96	0.15	33.93	0.03	0.2	0.06	0.10	0.09	pTR161/pTR190	1
2	6	6	5	6	6	8	2	5	2	6	33.64	0.15	33.89	-0.25	-1.7	0.06	0.09	0.10	pTR161/pTR182	1
3	4	3	2	5	3	6	2	4	1	4	33.64	0.18	33.96	-0.32	-1.8	0.06	0.09	0.15	pTR161/pTR188	1
4	3	2	2	3	1	5	2	3	1	3	33.64	0.24	33.85	-0.21	-0.9	0.06	0.09	0.21	pTR161/pTR183	1
5	3	2	2	3	1	5	1	3	1	3	33.96	0.25	34.02	-0.06	-0.2	0.06	0.10	0.22	pTR161/pTR199	1
6	3	4	3	3	5	2	0	2	1	3	33.96	0.20	33.93	0.03	0.2	0.06	0.10	0.16	pTR183/pTR182	1
7	0	0	0	0	0	0	1	0	0	0	35.29	4.20	29.93	5.36	1.3	0.07	0.10	4.20	pTR199/pTR183	1
8	1	2	0	0	1	1	0	0	0	0	36.00	0.85	34.33	1.67	2.0	0.07	0.11	0.84	pTR182/pTR190	1
9*	1	1	0	2	2	1	1	1	0	1	36.00	0.44	33.84	2.16	4.9	0.07	0.11	0.42	pTR199/pTR188	1
10*	16	0	0	1	1	0	0	1	0	0	36.36	0.13	35.43	0.93	7.1	0.07	0.11		pLP119/pBR322	2
11	34	0	0	2	2	0	0	0	1	1	35.64	0.13	35.49	0.15	1.2	0.07	0.11		pLP140/pBR322	2
12	41	0	0	1	1	0	0	1	0	0	35.64	0.13	35.54	0.10	0.8	0.07	0.11		pLP144/pBR322	2
13	0	1	1	0	2	14	0	0	1	0	33.96	0.12	33.79	0.17	1.5	0.06	0.10		pLP219/pBR322	2
14	0	1	1	0	2	17	0	0	1	0	33.64	0.11	33.77	-0.13	-1.2	0.06	0.09		pLP222/pBR322	2
15	0	0	0	0	0	23	1	0	1	0	33.96	0.12	33.76	0.20	1.7	0.06	0.10		pLP225/pBR322	2
16	0	0	0	0	0	32	1	0	1	0	33.64	0.11	33.73	-0.09	-0.8	0.06	0.09		pLP234/pBR322	2
17	7	0	0	-1	-1	0	0	1	-1	-1	36.00	0.13	36.04	-0.04	-0.3	0.07	0.11		pLP140/pLP144	2
18	18	0	0	1	1	0	0	-1	1	1	35.29	0.12	35.54	-0.25	-2.0	0.07	0.10		pLP119/pLP140	2
19	25	0	0	0	0	0	0	0	0	0	35.64	0.13	35.62	0.02	0.2	0.07	0.11		pLP119/pLP144	2
20*	0	-1	-1	0	-2	9	1	0	0	0	32.73	0.11	33.68	-0.95	-8.8	0.06	0.09		pLP219/pLP225	2
21	0	-1	-1	0	-2	15	1	0	0	0	33.64	0.11	33.67	-0.03	-0.2	0.06	0.09		pLP222/pLP234	2
22	0	-1	-1	0	-2	18	1	0	0	0	33.64	0.11	33.66	-0.02	-0.2	0.06	0.09		pLP219/pLP234	2
23*	0	0	0	0	0	12	0	0	0	0	32.73	0.08	33.65	-0.92	-8.5	0.06	0.09		pLP219/pLP222 & pLP225/pLP234	2
24	43	37	40	28	62	46	24	47	33	12	34.29	0.09	34.10	0.19	2.2	0.07	0.05	0.01	p372/pBR322	3
25	11	8	7	3	12	5	1	7	3	2	34.12	0.12	34.16	-0.04	-0.3	0.06	0.05	0.09	p60/pBR322	3
26	9	15	11	2	16	12	12	12	12	2	33.96	0.09	34.08	-0.12	-1.3	0.06	0.05	0.05	p104/pBR322	3
27	20	0	0	1	2	0	0	0	1	0	35.47	0.09	35.56	-0.09	-1.0	0.07	0.05		pAA24/pBR322	3
28	85	0	-1	0	-1	0	0	0	-1	0	35.64	0.09	35.67	-0.03	-0.4	0.07	0.05		pAA82/pBR322	3
29	0	0	2	20	0	0	0	0	1	21	33.64	0.08	33.66	-0.02	-0.2	0.06	0.05		pAT44/pBR322	3
30	-1	-2	-1	19	-1	-1	-1	0	-1	18	33.64	0.08	33.69	-0.05	-0.7	0.06	0.05		pAT29/pBR322	3
31	1	0	2	34	0	0	0	0	1	35	33.80	0.08	33.72	0.08	1.1	0.06	0.05		pAT73/pBR322	3
32	65	0	-1	-1	-3	0	0	0	-2	0	35.82	0.09	35.72	0.10	1.2	0.07	0.05		pAA82/pAA24	3
33	0	0	0	0	0	0	0	0	1	0	42.15	1.70	40.18	1.97	1.2	1.70				4
34	0	0	0	0	0	0	1	0	0	0	32.90	1.70	29.93	2.97	1.7	1.70				4
35	0	0	0	0	0	0	0	1	0	0	39.40	1.70	36.83	2.57	1.5	1.70				4
36	1	0	0	0	0	0	0	0	0	0	36.75	1.70	35.62	1.13	0.7	1.70				4
37	0	0	0	1	0	0	0	0	0	0	32.20	1.70	32.14	0.06	0.0	1.70				4
38	1	0	0	0	0	0	0	0	0	0	36.00	0.36	35.62	0.38	1.1	0.36				5
39	0	0	0	1	0	0	0	0	0	1	34.29	0.34	33.74	0.55	1.6	0.34				5
40	0	0	0	0	0	0	1	0	1	0	34.29	1.00	35.06	-0.77	-0.8	1.00				6

```
                 12 bases deleted
       .... TAAGCT|TTAATGCGGTAG|TTTATCA....              pBR322
  .... TAAGCT|TATA...........TATAT|TTTATCA ....           pAT29
                 41 bases inserted
```

Lost twist angles: $3\tau_{AA(TT)}$, $2\tau_{AC(GT)}$, τ_{AG}, τ_{AT}, $\tau_{CA(TG)}$, $\tau_{CC(GG)}$, τ_{CG}, τ_{GC}, $2\tau_{TA}$

New angles : $2\tau_{AA(TT)}$, $20\tau_{AT}$, $20\tau_{TA}$

Difference : $-\tau_{AA}-2\tau_{AC}-\tau_{AG}+19\tau_{AT}-\tau_{CA}-\tau_{CC}-\tau_{CG}-\tau_{GC}+18\tau_{TA}$

Fig. 1. Derivation of equation No. 30.

The other equations are obtained in a similar way.

The quality of the solution is sensitive to the values of the errors σ_e as these enter as weights ($1/\sigma_e^2$) in the least squares sum. A careful evaluation of σ_e is therefore necessary.

For the electrophoretic band-shift method the experimental error is quoted as 1% (1-3). The actual instrumental error, however, is only about 0.2% (±0.02 base for the pitch of DNA) (3). The overestimate, ±0.1 bases (1-3), was meant to include some uncertainties of the method which might occasionally appear in the measurements (3). We re-evaluate the errors by estimating and then compounding three components (Table I): an instrumental error σ_1 (0.02 bases for the pitch); a rounding error σ_2 (0.03 bases for the pitch for the data from ref. 1,2 and 0.015 bases for ref. 3); and an end-effect error σ_3 .

The end-effect error in eqns. 1-9, 24-26 reflects the fact that in these cases the boundary dinucleotides of the insertions are uncertain. As the twist angles are typically 30-40° per dinucleotide, the root-mean-square end-effect error is taken as 3° when one end is uncertain, $3^o\cdot\sqrt{2} = 4.2^o$ (eqns. 6-9) for two and $3^o\cdot\sqrt{3} = 5.2^o$ for three uncertain ends (eqns. 1-5, 24-26). Combining the errors as $\sigma_e^2 = \sigma_1^2+\sigma_2^2+\sigma_3^2$, the resultant total experimental error σ_e is calculated to vary between 0.08^o and 0.13^o (when there are no end-effects)which is significantly less than the originally quoted value (0.36^o). These sharper values of σ_e impose more severe constraints on the consistency of the equations. The consistency obtained in spite of these constraints (see next section) justifies the revision of the errors. In addition, a systematic correction of -2% was applied to the experimental values of the average twist angles from ref. 1, as suggested

by the authors in a later article (2).

For the crystallographic data (4), the experimental error is taken uniformly equal to 1.7^{o} which is the root mean square difference of the twist angles of identical dinucleotides positioned symmetrically in the self-complementary dodecanucleotide CGCGAATTCGCG. The twist angles for CG·CG pairs at the end of the double stranded DNA fragment are considerably larger than those of the interior CG·CG pairs (about 38° and 33° respectively). Since the complementary dinucleotides near the ends might be distorted we use the twist angles of the interior CG·CG pairs in the calculations.

RESULTS

Table II compares the estimates of the individual twist angles in DNA, experimental (2-5) and calculated (this work). The errors shown are calculated as described in Methods.

To start with, we note that the absolute differences between experimental and calculated angles ($\tau_e-\tau_c$) for the electrophoretic experiments, when no end-effects are involved, actually turn out to be significantly smaller than the error (0.36^{o}) ascribed to the measurements in ref. 2 and 3, consistent with our less conservative re-evaluation of the experimental errors.

To find out whether the system of equations is consistent we compared the experimentally measured angles τ_e with the average twist angles τ_c calculated by substitution of the solution into the equations (see Table I). The difference between calculated and experimentally measured average angles, expressed in units of the standard deviations of the measurements, is shown in column $(\tau_e-\tau_c)/\sigma_e$. If the system of equations is consistent these numbers should be normally distributed with mean zero and unit variance. As one can see, this is essentially the case.

Four equations (no. 9,10,20 and 23), however, appear to be inconsistent with the rest. Their corresponding values of τ_c deviate from the τ_e values by 5 to 9 standard deviations. We believe that the results of these four experiments might be erroneous. Indeed, experiments No. 20 and 23, designed to measure the twist angle of poly(dG)·poly(dC) (2), are contradictory to other measurements of the angle presented in the same paper (equations No. 13-16 and 21,22). Similarly, the angle τ_{AA} estimated from the experiment No. 10 (2) is about 0.8^{o} bigger than other measurements of the same series (eqs. 11,12,18 and 19). The four inconsistent equations were not used in the final calculation.

The remaining 36 equations are statistically consistent with our assumption that the twist angles of DNA are additive. As an additional test we excluded from the system all equations which correspond to measurements using synthetic polynucleotides. A priori one could suspect that the twist angles involved in these molecules (τ_{AA}, τ_{AT}, τ_{CC}, τ_{CG}, τ_{GC} and τ_{TA}) might be systematically different from the corresponding angles in the natural DNA molecules. However, the remaining 14 equations (No. 1-6, 24-26 and 33-37) are still sufficient to calculate all ten twist angles and give values practically identical to those obtained from the complete system of 36 equations.

Finally, two small changes seem to be necessary. The numbers shown in column $(\tau_e-\tau_c)/\sigma_e$ (Table I) show that the calculated average angles τ_c for equations No. 1-5 are systematically bigger and those of No. 33-37 systematically smaller than the experimental values. We believe that the systematic deviation in the first case is due to the fact that all fragments inserted in plasmids pTR182, 183, 188, 190 and 199 have the same nucleotide sequences around their 5'-ends (1). The second case (eqs. No. 33-37) corresponds to crystallographic measurements (4) where the DNA fragment crystallized might suffer some positive twist deformation distributed along the molecule. The angle τ_{AA} (eq. No. 36), for example, is about 1° bigger than the value obtained from electrophoretic and nuclease digestion measurements (2,35). The extent of this twist deformation can be estimated as mean value of $(\tau_e-\tau_c)$ for the equations No. 33-37 (1.74^{o}, see Table I). This value has to be subtracted from the crystallographic estimates.

Elimination of these two systematic errors results in slightly changed solution presented in the last column of Table II. This solution we consider the final result of our calculations.

DISCUSSION

There are two main results:

1) The fact that the system of equations is consistent, indicates that the twist angles between any two adjacent base-pairs are to a good approximation independent of the particular sequence they are included in.

2) Four of the twist angles presented in Table II (τ_{AC}, τ_{AG}, τ_{CA} and τ_{TA}) were not known before. They constitute very specific predictions for future experiments. These angles can be measured, for example, by the electrophoretic band-shift method using the shortest possible insertions containing the dinucleotides AG(CT), AC(GT), CA(TG) and TA.

Table II. Twist angles of B-DNA (°)

	Angles known before	Ref.	Angles calculated	Final solution (corrected)
τ_{AA}	35.64 ± 0.36	2	35.62 ± 0.07	35.62 ± 0.06
	35.64 ± 0.36	3		
	36.75 ± 1.7	4		
	36.00 ± 0.36	5		
τ_{AC}	-		34.0 ± 1.4	34.4 ± 1.3
τ_{AG}	-		27.9 ± 1.6	27.7 ± 1.5
τ_{AT}	32.2 ± 1.7	4	32.1 ± 1.1	31.5 ± 1.1
τ_{CA}	-		34.4 ± 1.0	34.5 ± 0.9
τ_{CC}	33.64 ± 0.34	2	33.65 ± 0.07	33.67 ± 0.07
τ_{CG}	32.9 ± 1.7	4	29.9 ± 1.1	29.8 ± 1.1
τ_{GA}	39.4 ± 1.7	4	36.8 ± 0.9	36.9 ± 0.9
τ_{GC}	42.2 ± 1.7	4	40.2 ± 1.2	40.0 ± 1.2
τ_{TA}	-		35.3 ± 1.1	36.0 ± 1.1

We realize that sequence-dependent variation in the pattern of the twist angles may be an important structural feature involved in the specific recognition of DNA by biological macromolecules. Work is in progress to analyze patterns of twist angles in natural DNA molecules and their possible relation to biological function.

One particularly interesting feature can be seen in the final set of angles presented in Table II. The twist angles for dinucleotides ending in C or A tend to be larger than average, those beginning with C or A smaller than average (the reverse in true for T and G). Expressed differently, the sum $\sum_X(\tau_{XA}-\tau_{av}) = 7.0^{\circ}$ is approximately equal with opposite sign to $\sum_X(\tau_{AX}-\tau_{av}) = -6.8^{\circ}$ (for G the corresponding sums are -10.4° and $+8.9^{\circ}$; similarly for T and C). This means that a larger than average angle tends to be compensated by a smaller than average angle in the immediately following dinucleotide. This trend prevents accumulation of over- or undertwisting along a DNA molecule. As a consequence, the structure of the B-DNA backbone typically shows a gentle zig-zag of plus/minus a few degrees. It is interesting to speculate about the biological implications of this effect.

ACKNOWLEDGEMENT

E.N.T. thanks the European Molecular Biology Organization for a short

term fellowship.

† permanent address: Polymer Research Department, The Weizmann Institute of Science, Rehovot 76100, Israel.

REFERENCES

1. Wang, J.C. (1979) Proc.Natl.Acad.Sci. USA 76, 200-203.
2. Peck, L.J. and Wang, J.C. (1981) Nature 292, 375-378.
3. Strauss, F., Gaillard, C. and Prunell, A. (1981) Eur.J.Biochem. 118, 215-222.
4. Dickerson, R.E. and Drew, H.R. (1981) J.Mol.Biol. 149, 761-786.
5. Rhodes, D. and Klug, A. (1981) Nature 292, 378-380.
6. Behe, M., Zimmerman, S. and Felsenfeld, G. (1981) Nature 293, 233-235.
7. Liu, L.F. and Wang, J.C. (1978) Cell 15, 979-984.