Sequence-Dependent Dynamics of Duplex DNA: The Applicability of a Dinucleotide Model

T. M. Okonogi,* S. C. Alley,*† A. W. Reese,* P. B. Hopkins,* and B. H. Robinson*
*Department of Chemistry, University of Washington, Seattle, Washington 98195-1700 USA; and
†Pathogenesis, Seattle, Washington 98119 USA

ABSTRACT The short-time (submicrosecond) bending dynamics of duplex DNA were measured to determine the effect of sequence on dynamics. All measurements were obtained from a single site on duplex DNA, using a single, site-specific modified base containing a rigidly tethered, electron paramagnetic resonance active spin probe. The observed dynamics are interpreted in terms of single-step sequence-dependent bending force constants, determined from the mean squared amplitude of bending relative to the end-to-end vector using the modified weakly bending rod model. The bending dynamics at a single site are a function of the sequence of the nucleotides constituting the duplex DNA. We developed and examined several dinucleotide-based models for flexibility. The models indicate that the dominant feature of the dynamics is best explained in terms of purine- and pyrimidine-type steps, although distinction is made among all 10 unique steps: It was found that purine-purine steps (which are the same as pyrimidine-pyrimidine steps) were near average in flexibility, but the pyrimidine-purine steps (5′ to 3′) were nearly twice as flexible, whereas purine-pyrimidine steps were more than half as flexible as average DNA. Therefore, the range of stepwise flexibility is approximately fourfold and is characterized by both the type of base pair step (pyrimidine/purine combination) and the identity of the bases within the pair (G, A, T, or C). All of the four models considered here underscore the complexity of the dependence of dynamics on DNA sequence with certain sequences not satisfactorily explainable in terms of any dinucleotide model. These findings provide a quantitative basis for interpreting the dynamics and kinetics of DNA-sequence-dependent biological processes, including protein recognition and chromatin packaging.

INTRODUCTION

Flexibility is an intrinsic property of duplex DNA and plays an important role in its functionality, particularly in the areas of ion and drug binding (Takahara et al., 1995), architectural transcription protein interactions (Grove et al., 1996; Nardulli et al., 1996; Wolffe, 1994), chromatin packaging (Hagerman, 1988; Patikoglou and Burley, 1997; Richmond et al., 1984), strand exchange (Thompson et al., 1976), deletion formation (Chedin et al., 1994), and the number of processes occurring during replication, repair, and recombination known as dynamic mutations (Pearson and Sinden, 1998a). Past efforts to understand the sequence-dependent nature of flexibility have been manifested in the form of circular DNA formation kinetic studies (Dlakic and Harrington, 1995; Harrington and Winicov, 1994). A number of specific sequences have been suggested as more flexible; in particular, the dinucleotides CA and TA are candidates for regions of increased flexibility based on the results of gel mobility assays (Harrington and Winicov, 1994). Harrington (Harrington and Winicov, 1994) suggested that sequence-dependent flexibility may account for unexplained differences in the gel mobilities between G_3C_3-motifs and A_5-tracts, both thought to contain bends equal in magni-tude, although not in direction (Brukner et al., 1993). Hogan and Austin examined bacteriophage 434 repressor binding affinity to DNA sequences (Hogan and Austin, 1987; Hogan et al., 1989). They concluded that protein binding diminished when the central 4-bp ATAT within the 14-bp long operator were replaced with other sequences. The repressed binding affinities at modified sites were due to the replacement basepairs' inability to twist or flex like the AT basepairs they replaced. An energetic bias towards nucleosome formation in the form of conformationally correct preexisting bends and bendability dictates translational and rotational wrapping about the nucleosome (Anderson and Widom, 2000; Anselmi et al., 1999; Filesi et al., 1999; Sivolob and Khrapunov, 1995; Thastrom et al., 1999) and is reflected in the following recent nucleosome-positioning experiments. Among the natural and nonnatural nucleosome-positioning sequences tested in nucleosome formation using SELEX methods, natural sequences with the highest formation affinities contained TATA runs and neurodegenerative disease associated CAG repeats (Thastrom et al., 1999). In contrast, nonnatural sequences with the highest formation affinities contained ~10-bp periodically repeated TA and AA steps (Anderson and Widom, 2000; Thastrom et al., 1999; Widlund et al., 1999). TA steps, AA steps, TATA sequences, and CAG repeats have all been reported as either highly deformable, flexible, or occurring as a bend or distorted DNA structure. For example, AA and AT patches commonly adopt narrowed minor grooves that bend towards the protein surface where phosphate-residue contacts are made in the catabolite activator protein (CAP) and the nucleosome (Luger et al.,1997; Schultz et al., 1991). The

Submitted September 4, 2001; and accepted for publication May 23, 2002.
Address reprint requests to Dr. Bruce H. Robinson, University of Washington, Department of Chemistry, Box 351700, Seattle, WA 98195. Tel.: 206-543-1773; Fax: 206-685-8665; E-mail: robinson@chem.washington.edu.

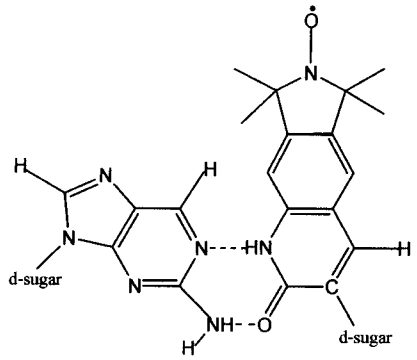
0006-3495/02/12/3446/14 $2.00

TATA binding box is bent by ~100° upon binding with the TATA binding protein (TBP) (Parkhurst et al., 1999). To date, sequence-dependent flexibility is the only physical characteristic to correlate with human hereditary neurodegenerative diseases caused by $(CTG)_n$ $(CAG)_n$, $(CGG)_n(CCG)_n$, and $(GAA)_n$ $(TTC)_n$ repeat expansion (Bacolla et al., 1997; Chastain and Sinden, 1998; Pearson and Sinden, 1998a, 1998b; Sinden, 1999). Several experimental techniques exist for measuring the dynamic flexibility of DNA (Fujimoto and Schurr, 1990; McAteer and Kennedy, 2000; Okonogi et al., 1999), and certain sequences, for example TA, CA, and CG steps, have been implicated as being more conformationally mobile (McAteer et al., 1995; McAteer and Kennedy, 2000) or more flexible than average (Okonogi et al., 2000).

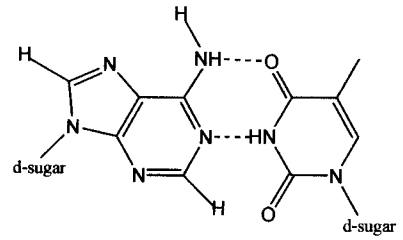
There are two distinct mechanisms by which the sequence may alter the dynamics of DNA. The first mechanism involves a global change in the set of secondary structures induced by a specific sequence. For example, Schurr and co-workers inserted a 16-bp $(CG)_8$ sequence into a linear 1.1-kb duplex DNA sequence and found that the relatively small change in sequence induced a large change in secondary structure throughout the 1.1-kb molecule (Kim et al., 1993). The second mechanism by which sequence may alter dynamics is local differential sequence-dependent flexibility, which may affect the dynamics of other bases through the internal collective modes. An example of a local change in flexibility affecting dynamics or kinetics elsewhere in the DNA is the process by which inserts are deleted between directing 18-bp repeats flanking the deletion segment (Chedin et al., 1994).

Generally, if a local effect is communicated via the DNA to other parts, then the communication of that effect should fall away as the distance from it increases. Fall-off of an effect with distance is a characteristic of both structural alterations (allosterism) and differential internal flexibility. One example of an allosteric effect falling-off is that of a $(CG)_4$ insert increasing the rate of cleavage by restriction enzymes at specific sites. As an example, an enhancement in the cutting rate was observed when an insert was placed up to 50 bp away from the cutting site in linear DNA (Aloyo et al., 1993; Ramsauer et al., 1997). This illustrates how long-range structural correlations can propagate over a finite domain and die off with distance. Dynamic effects have similarly been shown to behave as though local and have a characteristic decrease in effect as the distance between the insert and the observer is increased (Okonogi et al., 2000; Schurr and Fujimoto, 1999).

In the present study we examined the dynamics of 40 duplex B-form DNAs, differing only in sequence at positions 12 through 50. The observer or reporter group is a 2-aminopurine-quinolone base containing a rigidly attached nitroxide moiety, basepaired as 2AP-Q (Fig. 1). Because the nitroxide in the 2AP-Q basepair is rigidly attached, motion of this probe reflects the dynamics of the DNA. The orientation of the nitroxide probe in 2AP-Q makes this probe particularly sensitive to bending dynamics and thus provides an excellent method to quantitatively determine the internal flexibility of DNA. The mean square amplitude of the dynamics at the probe, $\langle\beta_6^2(\infty)\rangle_{exp}$, was measured using methods previously reported (Hustedt et al., 1993, 1995; Okonogi et al., 1998, 1999, 2000; Robinson and Drobny, 1995a, 1995b). The mean square amplitude of motion of basepair 6, $\langle\beta_6^2(\infty)\rangle_{exp}$, is related to the order parameter, S_6. From electron paramagnetic resonance, (EPR) experiments at x-band, the spectral distance between the two outmost features of the spectra, $\langle A_{zz}\rangle$, defines the order parameter:



2AP-Q



A-T

FIGURE 1 2AP-Q spin-probe basepair and the naturally occurring adenosine-thymidine (A-T) basepair shown for comparison.

$$S_6 = \frac{\langle A_{zz}\rangle - \bar{a}}{A_{zz} - \bar{a}} = \frac{3e^{-2\langle\beta_6^2\rangle} + 1}{4} \quad (1)$$

Here A_{zz} is the spectral distance in the absence of any motion and $\bar{a}$ is a motion-independent constant. Both A_{zz} and $\bar{a}$ are known quantities. The values of $\langle\beta_6^2(\infty)\rangle_{exp}$ are determined by spectral simulation. The data are analyzed in terms of the modified-weakly- bending-rod (MWBR) model (Okonogi et al., 2000). This model allows that 1) differential motion may be due to a local change in flexibility of the DNA molecule and 2) different sequences of DNA have different propensities for bending.

TABLE 1 Sequences used to determine the force constant ratios

Number, name	Sequence
1. NT	CCTCGQATCGTGCTCCTCATCTTCGTGTCCGGAAAGGTGGTGCTCACTG
2. AT4	CCTCGQATCGTATATATATTCTTCGTGTCCGGAAAGGTGGTGCTCACTG
3. AT10	CCTCGQATCGTATATATATATATATATATATGAAAGGTGGTGCTCACTG
4. AT15	CCTCGQATCGTATATATATATATATATATATATATATGCTCACTG
5. AT7A	CCTCGQATCGTATATATATATATATAGTCCGGAAAGGTGGTGCTCACTG
6. AT7A_s5	CCTCGQATCGTGCTCCATATATATATATATAGAAAGGTGGTGCTCACTG
7. AT7A_s12	CCTCGQATCGTGCTCCTCATCTTATATATATATATATAGGTGCTCACTG
8. AT7A_s24	CCTCGQATCGTGCTCCTCATCTTCGTGTCCGGAAAATATATATATATAT
9. AA7A or AAA5	CCTCGQATCGTAAAAAAAAAAAAAAGTCCGGAAAGGTGGTGCTCACTG
10. AA10	CCTCGQATCGTAAAAAAAAAAAAAAAAAAAAGAAAGGTGGTGCTCACTG
11. AA7A_s5	CCTCGQATCGTGCTCCAAAAAAAAAAAAAAGAAAGGTGGTGCTCACTG
12. CG7C	CCTCGQATCGTCGCGCGCGCGCGCGCGTCCGGAAAGGTGGTGCTCACTG
13. CG10	CCTCGQATCGTCGCGCGCGCGCGCGCGCGCGGAAAGGTGGTGCTCACTG
14. CG7C_s5	CCTCGQATCGTGCTCCCGCGCGCGCGCGCGCGAAAGGTGGTGCTCACTG
15. CC7C or CCC	CCTCGQATCGTCCCCCCCCCCCCCCCGTCCGGAAAGGTGGTGCTCACTG
16. CC10	CCTCGQATCGTCCCCCCCCCCCCCCCCCCCCGAAAGGTGGTGCTCACTG
17. CC7C_s5	CCTCGQATCGTGCTCCCCCCCCCCCCCCCCCGAAAGGTGGTGCTCACTG
18. AC7A	CCTCGQATCGTACACACACACACACAGTCCGGAAAGGTGGTGCTCACTG
19. AG7A	CCTCGQATCGTAGAGAGAGAGAGAGAGTCCGGAAAGGTGGTGCTCACTG
20. AAT5	CCTCGQATCGTAATAATAATAATAATGTCCGGAAAGGTGGTGCTCACTG
21. AAC5	CCTCGQATCGTAACAACAACAACAACGTCCGGAAAGGTGGTGCTCACTG
22. AAC5_s5	CCTCGQATCGTGCTCCAACAACAACAACAACGAAAGGTGGTGCTCACTG
23. AAG5	CCTCGQATCGTAAGAAGAAGAAGAAGGTCCGGAAAGGTGGTGCTCACTG
24. AGG5	CCTCGQATCGTAGGAGGAGGAGGAGGGTCCGGAAAGGTGGTGCTCACTG
25. AGG5_s5	CCTCGQATCGTGCTCCAGGAGGAGGAGGAGGGAAAGGTGGTGCTCACTG
26. ACG5	CCTCGQATCGTACGACGACGACGACGGTCCGGAAAGGTGGTGCTCACTG
27. ACG5_s5	CCTCGQATCGTGCTCCACGACGACGACGACGGAAAGGTGGTGCTCACTG
28. ACT5	CCTCGQATCGTACTACTACTACTACTGTCCGGAAAGGTGGTGCTCACTG
29. ATC5	CCTCGQATCGTATCATCATCATCATCGTCCGGAAAGGTGGTGCTCACTG
30. CAG5	CCTCGQATCGTCAGCAGCAGCAGCAGGTCCGGAAAGGTGGTGCTCACTG
31. CCG5	CCTCGQATCGTCCGCCGCCGCCGCCGGTCCGGAAAGGTGGTGCTCACTG
32. CCG5_s5	CCTCGQATCGTGCTCCCCGCCGCCGCCGCCGGAAAGGTGGTGCTCACTG
33. 1/2CAP:TGTGACAT	CCTCGQATCGTTGTGACATTCTTCGTGTCCGGAAAGGTGGTGCTCACTG
34. TATA:TATATAAA	CCTCGQATCGTTATATAAATCTTCGTGTCCGGAAAGGTGGTGCTCACTG
35. G_3C_3motif	CCTCGQATCGTGGGCCCCATCTTCGTGTCCGGAAAGGTGGTGCTCACTG
36. G_3C_3motif_s1	CCTCGQATCGTGGGGCCCATCTTCGTGTCCGGAAAGGTGGTGCTCACTG
37. G_3C_3motif_s8	CCTCGQATCGTGCTCCTCAGGGCCCTGTCCGGAAAGGTGGTGCTCACTG
38. G_3C_3motif_s0_s10	CCTCGQATCGTGGGCCCCATCGGGCCCTCCGGAAAGGTGGTGCTCACTG
39. G_3C_3motif/A_5tract	CCTCGQATCGTGGGCCCAAAAAGGGCCCAAAAAAAGGTGGTGCTCACTG
40. A_5tract/G_3C_3motif	CCTCGQATCGTAAAAAGGGCCCAAAAAGGGCCCAAGGTGGTGCTCACTG

T replaces Q at position 6 in calculations.

MATERIALS AND METHODS

A set of 50-bp-long duplex DNAs (Table 1) was constructed. The spin probe base, Q (Fig. 1), is always at position 6 in the initial 11-bp sequence, and the test sequence never begins before position 12. The control or reference sequence (NT) is basepairs 1087 to 1136 from *Drosophila melanogaster* TATA-box binding protein TFIID gene (Table 1). Dynamics seen by the probe comes from several sources: 1) the overall tumbling of the DNA, 2) the internal collective modes of motion, and 3) the motion of the probe (including the basepair) independent of the macromolecular environment. The following experiments were designed to measure only the sequence effects on the collective modes of motion on the probe. In doing so, the following controls were used: 1) a constant 11-bp region contained the probe and 2) the overall length of the DNA remained constant throughout the experiments. These two constraints on the DNA ensure that the correlation times of the uniform modes and the length independent motions of the probe are nearly constant among the set of DNAs constructed. The internal collective modes provide the only mechanism for motions in a distal region of the DNA to affect the dynamics of the probe.

All oligomers were prepared on an ABS 392 DNA synthesizer, sized with denaturing polyacrylamide gel electrophoresis, and purified on a Sephadex column or purified by trityl-on reverse-phase high pressure liquid chromatograph (HPLC) and the trityl removed by 85% glacial acetic acid in water. The spin-labeled sequences were combined with a 10% excess of the corresponding unlabeled complementary strand. The buffer in all cases was 10 mM sodium phosphate (pH 7.0), 0.1 mM EDTA, and 100 mM NaCl. All EPR spectra were recorded digitally on an EPR 9.4-GHz (x-band) Bruker EMX spectrometer. The EPR parameters were 10 and 100 kHz of modulation frequency, 1.0 Gauss of modulation amplitude, and 1 and 2 mW of microwave (nonsaturating) power. All samples were kept in a refrigerator at 4°C between experiments and kept at 30°C while in the spectrometer.

One set of constructs contains the six distinct dinucleotide repeats $(AT)_n$, $(CG)_n$, $(AC)_n$, $(TC)_n$, $(GG)_n$, and $(AA)_n$ (see Table 5). The largest set of dinucleotides produced were the $(AT)_n$ inserts. Initially, the $(AT)_n$ sequences were analyzed independently (Okonogi et al., 2000) for the following reasons: 1) $(AT)_n$ sequences lack retarded polyacrylamide gel

TABLE 2 Names of the controls and all di- and tri-nucleotide sequences made with the probe, Q, at position 6, the length of the test sequence *L*, the start, N_1, and stop, N_2, positions of the test sequence, and the corresponding $\langle\beta_6^2(\infty)\rangle_{exp}$ values

Number name	*L*	N_1	N_2	$\langle\beta_6^2(\infty)\rangle_{exp} \pm \sigma\langle\beta_6^2(\infty\rangle_{exp}$ ($\langle\beta_6^2(\infty)\rangle_{mod}$) rad^2
1. NT	0	—	—	0.087 ± 0.002
2. AT4	8	12	19	0.090 ± 0.001
3. AT10	20	12	31	0.093 ± 0.001
4. AT15	30	12	41	0.097 ± 0.003
5. AT7A	15	12	26	0.093 ± 0.001
6. AT7A s5	15	17	31	0.091 ± 0.001
7. AT7A s12	15	24	38	0.088 ± 0.001
8. AT7A s24	15	36	50	0.087 ± 0.002
9. AA7A or AAA5	15	12	26	0.083 ± 0.001
10. AA10	20	12	31	0.089 ± 0.001(0.084)
11. AA7A s5	15	17	31	0.086 ± 0.001(0.084)
12. CG7C	15	12	26	0.085 ± 0.002
13. CG10	20	12	31	0.086 ± 0.001
14. CG7C s5	15	17	31	0.086 ± 0.001
15. CC7C or CCC5	15	12	26	0.086 ± 0.002
16. CC10	20	12	31	0.084 ± 0.001
17. CC7C s5	15	17	31	0.087 ± 0.001
18. AC7A	15	12	26	0.088 ± 0.002
19. AG7A	15	12	26	0.089 ± 0.002
20. AAT5	15	12	26	0.089 ± 0.004
21. AAC5	15	12	26	0.088 ± 0.001
22. AAC5 s5	15	17	31	0.087 ± 0.001
23. AAG5	15	12	26	0.087 ± 0.002
24. AGG5	15	12	26	0.089 ± 0.001
25. AGG5 s5	15	17	31	0.087 ± 0.001
26. ACG5	15	12	26	0.089 ± 0.001
27. ACG5 s5	15	17	31	0.085 ± 0.001
28. ACT5	15	12	26	0.088 ± 0.001
29. ATC5	15	12	26	0.087 ± 0.002
30. CAG5	15	12	26	0.089 ± 0.002
31. CCG5	15	12	26	0.087 ± 0.001
32. CCG5 s5	15	17	31	0.082 ± 0.002
33. 1/2CAP: TGTGACAT	8	12	19	0.089 ± 0.002
34. TATA: TATATAAA	8	12	19	0.093 ± 0.002(0.088)
35. G_3C_3-motif	6	12	17	0.084 ± 0.001(0.083)
36. G_3C_3-motif_s1	6	13	18	0.081 ± 0.001(0.077)
37. G_3C_3-motif_s8	6	20	25	0.076 ± 0.001(0.070)
38. G_3C_3-motif_s0-s10	6	12	17	0.088 ± 0.001(0.083)
G_3C_3-motif_s10	6	22	27	
39. G_3C_3-motif/A_5-tract				0.089 ± 0.001(0.086)
G_3C_3	6	12	17	
A_5-tract	5	18	22	
G_3C_3	6	23	28	
A_5-tract	5	29	33	
40. A_5-tract/G_3C_3-motif				0.087 ± 0.001(0.072)
A_5-tract	5	12	16	
G_3C_3	6	17	22	
A_5-tract	5	23	27	
G_3C_3	6	28	33	

All sequences are in total 50 basepairs in length. $\langle\beta_6^2(\infty)\rangle_{mod}$ in parenthesis is the value after bend analysis for sequences where the analysis showed a change that was significant based on the standard error.

electrophoretic mobilities, indicating that these sequences do not contain permanent bends; 2) $(AT)_n$ sequences are thought to be very flexible (Harrington and Winicov, 1994); and 3) $(AT)_n$ sequences have not been implicated in inducing allosteric transitions. Sequences (Table 2) containing the trinucleotide combinations $(ACG)_n$, $(AAG)_n$, $(AAT)_n$, $(CAG)_n$, $(AAC)_n$, $(ACT)_n$, $(AAA)_n$, $(AGG)_n$, $(CCC)_n$, $(CCG)_n$, and $(ATC)_n$ as well as G_3C_3-motifs, A-tracts, the one-half CAP (TGTGACAT) and TATA-box (TATATAAA) were also constructed. In principle, the trinucleotide sequences provide the additional discrimination needed to distinguish between the two distinct types of steps in the set of dinucleotide repeat sequences. For example, one might be able to distinguish between an ApT step and a TpA step. The names of the test sequences describing the insert composition, the test sequence insert lengths L, test sequence base starting positions N_1, and stopping positions N_2, are listed in Table 2 accompanied by their experimental mean square amplitude of flexure value, $\langle(\beta_6^2(\infty)\rangle$, as reported by the probe at position 6. Sequences containing A_5-tracts, G_3C_3-

TABLE 3a List of sequences removed in order from 1 to 10 in the least squares minimization of the models using 2 force constants (f.c.), 3 f.c., 6 f.c., and 10 f.c. ($\nu = 10$ and $\nu = 8$) for the $\langle\beta_6^2(\infty)\rangle$ values with A-tract modified values

	2 f.c. model	3 f.c. model	6 f.c. model	10 f.c. model $\nu = 10$	10 f.c. model $\nu = 8$
1	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8
2	AA10	G_3C_3_s1	G_3C_3_s1	G_3C_3_s1	G_3C_3_s1
3	G_3C_3_s1	TATA	TATA	TATA	TATA
4	AA7A or	C_3G_3	G_3C_3/A_5-tract	G_3C_3	G_3C_3
5	AA7A_s5	CG10	ACG5_s5	G_3C_3/A_5-tract	ACG5_s5
6	TATA	ACG5_s5	G_3C_3	ACG5_s5	$G_3C_3/A_5/A_5$-tract
7	G_3C_3	AGG5	ATC5	CCG5_s5	ATC5
8	ACG5_s5	CG7C_s5	AT7A_s12	ATC5	CCG5_s5
9	AAC5_s5	G_3C_3/A_5-tract	ACT5	CC10	CC10
10	A_5-tract/G_3C_3	CCG5_s5	A_5-tract/G_3C_3	AT7A_s12	AT7A_s12

G_3C_3-motif is abbreviated to G_3C_3.

TABLE 3b Inverse force constant ratios corresponding to Table 3a for $N = 31$ with sequences 1 through 9 removed

Nucleotide pair	2 f.c. model	3 f.c. model	6 f.c. model	10 f.c. model $\nu = 10$	10 f.c. model $\nu = 8$
AT	1.12 ± 0.01	1.1 ± 0.1	1.15 ± 0.01	0.9 ± .3	0.80 ± 0.09
AA = TT	1.12 ± 0.01	0.90 ± 0.01	0.87 ± 0.01	0.86 ± 0.01	0.85 ± 0.01
GT = AC	1.03 ± 0.01	1.1 ± 0.1	1.05 ± 0.01	0.45 ± 0.09	0.39 ± 0.05
CT = AG	1.03 ± 0.01	0.90 ± 0.01	1.05 ± 0.01	1.3 ± 0.1	1.20 ± 0.04
TA	1.12 ± 0.01	1.2 ± 0.1	1.15 ± 0.01	1.4 ± .2	1.48 ± 0.08
GA = TC	1.03 ± 0.01	0.90 ± 0.01	1.05 ± 0.01	0.69 ± 0.09	0.74 ± 0.04
CA = TG	1.03 ± 0.01	1.2 ± 0.1	1.05 ± 0.01	1.66 ± 0.08	1.71 ± 0.04
GC	0.94 ± 0.01	1.1 ± 0.1	0.94 ± 0.01	0.00 ± 0.00	0.00 FIXED
CC = GG	0.94 ± 0.01	0.90 ± 0.01	0.93 ± 0.01	1.01 ± 0.01	1.00 ± 0.01
CG	0.94 ± 0.01	1.2 ± 0.1	0.94 ± 0.01	1.85 ± 0.01	1.85 FIXED

motifs, or the TATA sequence have two $\langle\beta_6^2(\infty)\rangle$ values listed and correspond to the experimentally measured value and the modified value based on a static bend analysis when significantly different, respectively.

Data analysis

We use the MWBR theory to analyze for the dynamics observed at position 6 (Okonogi et al., 2000). The theory assumes a force constant κ may be used to characterize the bending potential between basepairs. The potential is assumed to be harmonic and isotropic. The force constant may depend on many bases in the sequence around the basepair. The most general form for the potential, in terms of two bending angles, η and θ, in the x and y direction that describe the relative orientation of 1 bp abstracted to a single, solid structure with cylindrical symmetry, relative to its neighbor is:

$$U = \tfrac{1}{2}\,\vec{\eta}\bar{\kappa}A(r)\vec{\eta} + \tfrac{1}{2}\,\vec{\Theta}^{\,+}\,\bar{\kappa}A(r)\Theta \qquad (2)$$

Where the matrix A depends on the force constant, $\{\kappa\}$:

$$A(r) = \frac{1}{\bar{\kappa}}\begin{bmatrix} \kappa_1 & -\kappa_1 & 0 & 0 \\ -\kappa_1 & \kappa_1+\kappa_2 & -\kappa_2 & 0 \\ 0 & -\kappa_2 & \kappa_2+\kappa_3 & -\kappa_3 \\ 0 & 0 & -\kappa_3 & \kappa_3 \end{bmatrix} = \begin{bmatrix} r_1 & -r_1 & 0 & 0 \\ -r_1 & r_1+r_2 & -r_2 & 0 \\ 0 & -r_2 & r_2+r_3 & -r_3 \\ 0 & 0 & -r_3 & r_3 \end{bmatrix} \qquad (3)$$

A force constant ratio, r_i, is related to a force constant, κ_i, by $r_i = \kappa_i/\bar{\kappa}$, in which $\bar{\kappa}$ is the force constant for average DNA. It is more convenient to compare force constants as ratios rather than absolute values. The mean square dynamics at equilibrium is:

$$\langle\vec{\beta}\vec{\beta}^{+}\rangle = 2\langle\vec{\eta}\vec{\eta}^{+}\rangle + \beta_0^2 = 2\left(\frac{k_BT}{\bar{\kappa}}\right)(A^{-1}) + \beta_0^2 \qquad (4)$$

in which A^{-1} is the pseudo-inverse of A and $\langle\beta_6^2(\infty)\rangle = \langle\,\vec{\beta}\vec{\beta}^{+}\rangle_{6,6}$. k_BT is Boltzmann's constant times the absolute temperature. Therefore, the mean square amplitude of motion at position 6 is a well-defined function of the set of force constants.

We consider different possible descriptions of the set of possible force constants. While each force constant can depend on many basepairs in the region around the basepair of interest, we restrict ourselves to admitting only nearest neighbor sequence dependencies to describe the force constants. The most general description of nearest neighbor force constants admits 16 possible combinations but only 10 are distinct due to symmetry. The notation $\kappa(X)$ represents the force constant for the dinucleotide step, X. When $\kappa(X) = \kappa(Y)$, the shortened notation $\kappa(X = Y)$ is used. The 10 distinct force constants are: κ(AT), κ(AA = TT), κ(GT = AC), κ(AG = CT), κ(TA), κ(GA = TC), κ(CA = TG), κ(GC), κ(CC = GG), and κ(CG) (Tables 3b and 4b). All steps are taken in the 5′ to 3′ direction: an AT or ApT step is 5′-ApT-3′.

Previously, we found that the mean force constant ratio for the average ApT and TpA dinucleotide step, $\bar{r}$ (AT = TA), was independent of $k_BT/\bar{\kappa}$ over a wide range of possible values of $\bar{\kappa}$ (i.e., $0.00227 \le k_BT/\bar{\kappa} \le 0.00283$). Therefore, the previously determined value of $k_BT/\bar{\kappa} = 0.00227$ (rad^2) was used in all of the following analyses. The length independent $\langle\beta_0^2$

TABLE 4a Same as Table 3a except with A-tract, TATA, and G_3C_3-motif modified $\langle\beta_6^2(\infty)\rangle$ values

	2 f.c. model	3 f.c. model	6 f.c. model	10 f.c. model $\nu = 10$	10 f.c. model $\nu = 8$
1	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8	G_3C_3_s8
2	A_5-tract/G_3C_3	A_5-tract/G_3C_3	A_5-tract/G_3C_3	A_5-tract_G_3C_3	A_5-tract/G_3C_3
3	G_3C_3_s1	G_3C_3_s1	G_3C_3_s1	G_3C_3_s1	G_3C_3_s1
4	AA10	AGG5	G_3C_3	G_3C_3	G_3C_3
5	AA7A	CG7C	ACG5_s5	ACG5_s5	ACG5_s5
6	AA7A_s5	CG10	ATC5	ATC5	ATC5
7	G_3C_3	ACG5_s5	AT7A_s12	CC7C_s5	CC7C_s5
8	AGG5	G_3C_3	ACT5	CCG5_s5	CCG5_s5
9	ACG5_s5	CG7C_s5	CCG5_s5	AT7A_s12	AT7A_s12
10	TATA	CCG5_s5	CC10	G_3C_3_s0_s10	G_3C_3_s0_s10

TABLE 4b Inverse force constant ratios corresponding to Table 4a for $N = 31$ with sequences 1 through 9 removed

Nucleotide pair	2 f.c. model	3 f.c. model	6 f.c. model	10 f.c. model $\nu = 10$	10 f.c. model $\nu = 8$
AT	1.11 ± 0.01	1.30 ± 0.09	1.15 ± 0.01	0.9 ± 0.2	0.77 ± 0.08
AA = TT	1.11 ± 0.01	0.91 ± 0.01	0.86 ± 0.01	0.84 ± 0.01	0.84 ± 0.01
GT = AC	1.02 ± 0.01	1.30 ± 0.09	1.05 ± 0.01	0.42 ± 0.09	0.37 ± 0.04
CT = AG	1.02 ± 0.01	0.91 ± 0.01	1.04 ± 0.01	1.2 ± 0.1	1.20 ± 0.04
TA	1.11 ± 0.01	0.93 ± 0.09	1.15 ± 0.01	1.4 ± 0.2	1.51 ± 0.08
GA = TC	1.02 ± 0.01	0.91 ± 0.01	1.04 ± 0.01	0.77 ± 0.09	0.80 ± 0.04
CA = TG	1.02 ± 0.01	0.93 ± 0.09	1.05 ± 0.01	1.66 ± 0.08	1.71 ± 0.04
GC	0.94 ± 0.01	1.30 ± 0.09	0.94 ± 0.01	0.00 ± 0.00	0.00 FIXED
CC = GG	0.94 ± 0.01	0.91 ± 0.01	0.92 ± 0.01	0.92 ± 0.02	0.91 ± 0.01
CG	0.94 ± 0.01	0.93 ± 0.09	0.94 ± 0.01	1.85 ± 0.01	1.85 FIXED

value, previously determined, (Okonogi et al., 1999) was also used throughout.

RESULTS

Correcting for structural anomalies

The $\langle\beta_6^2(\infty)\rangle$ for different A-tracts (AA7A, AA7A_s5, AA10 of Table 1) are examined as a function of the distance of the A-tract from the probe, shown in Fig. 2 (open circles). It is clear that the initial data do not move monotonically toward the NT sequence as the distance from the A-tract to the spin probe is increased. Here, the NT control sequence has all $\bar{r}$ (NT) = 1 between basepairs and is neither flexible nor rigid; it is average DNA where all three AA7A, AA7A_s5, and AA10 converge when their test insert sequences are moved off the end of their respective DNA sequences. Later, NT is treated like any other sequence in Table 1 when determining the force constant ratios between basepairs. It has been established that A_5-tracts form curved helical DNA, with a curvature of 4°/basepair titled into the minor groove. One may correct for the effect of this tilting on the uniform motional modes using the general theory of Garcia de la Torre for the motion of rigid objects of arbitrary shape (Garcia Bernal and Garcia de la Torre, 1980). The theory predicts that, for a DNA bent due to A-tracts, the uniform modes may change very slightly, but the tilt angle between the spin probe and the principal axis of diffusion may

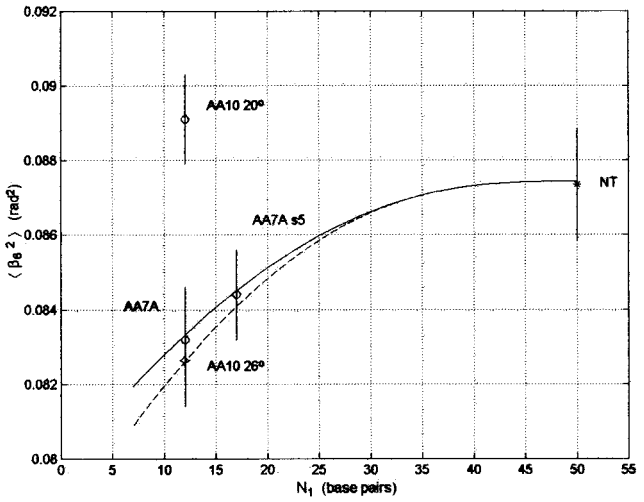


FIGURE 2 $\langle\beta_6^2\infty\rangle$ (rad^2) values for the NT (control) (∗), AA7A (○), AA10 (○), and AA7A_s5 (○) sequences with the tilt angle of the probe assumed to be 20° for all, plotted against test sequence start positions N_1 (basepairs). It was only necessary to modify the AA10 $\langle\beta_6^2\infty\rangle$ value (◇) when it was found to have a different tilt angle to the probe, that of 26°. Curves overlaying the data points are the theoretical predictions for $\langle\beta_6^2\infty\rangle$ based on test sequence insert start position N_1, length of the insert, r^{-1} (AA) = 0.86 for AA pairs and $r^{-1} = 1$ for all other pairs that are not AA (Eq. 4). Solid line is for an AA7A sequence within a 50-mer, and dashed line is for an AA10 sequence within a 50-mer.

change significantly. A significantly altered probe tilt angle to the principal axis will change the apparent dynamics measured at the probe. We can correct for this by creating a representation of the bent duplex DNA and computing new diffusion coefficients and orientation of the principle axes of diffusion, to give a new tilt angle. The new diffusion coefficients and tilt angle are then used in the line-shape simulation, which gives the value of $\langle\beta_6^2(\infty)\rangle$. Of the new tilt angles calculated for the three sequences, only the AA10 sequence probe angle deviated significantly (more than 1°) from the unperturbed probe angle of 20° with a new tilt angle of 26°. Imagine the helix aligned along the *z* axis in the *z-x* plane. The spin probe is tilted ~20° off the *z* axis (the helix) into the *z-y* plane. The bend in AA7A sequence realigns the helix along the *x* axis in the *z-x* plane. In this way, the tilt angle to the probe is preserved. The bend in the AA7A_s5 sample tilts the helix in the opposite direction onto the $-x$ axis in the *z-x* plane, and again the original tilt angle is preserved. The bends in the AA10 sample, however, take the helix out of the *z-x* plane and bend it into the *z*-($-y$) plane, therefore increasing the tilt angle to the probe. The corrected value for the AA10 sequence is plotted in Fig. 2 (open diamond) with the original values for AA10, AA7A, and AA7A_s5 (open circles). One can see that the new value is a monotonic function of the distance, which we take as confirmation that the corrections were reasonable.

Other sequences are known to bend DNA. G_3C_3 (GGGCCC) is thought to bend DNA by about the same amount as an A_5-tract but in the opposite direction (Brukner et al., 1993). Sequences containing this motif were corrected in a similar fashion to the A-tract containing sequences. The TATATAAA sequence was also analyzed for a bend due to its similarity to an A_5-tract. The four forms of analyses (discussed immediately below) were done with both the corrected and uncorrected tilt angles and diffusion tensors.

Analysis of sequence-dependent flexibility

Four forms of analysis were performed on the entire data set consisting of all 40 sequences and their $\langle\beta_6^2(\infty)\rangle$ values as given in Tables 1 and 2. The criterion for goodness of fit for all models is the reduced χ^2, χ_R^2, statistic, which is also used to evaluate the significance of the fit.

$$\chi_R^2 = \frac{1}{d}\sum_{i=1}^{N}\delta_i^2 \text{ in which } \delta_i = \left\{\frac{\langle\beta_6^2(\infty)\rangle_{i-exp} - \langle\beta_6^2(\infty)\rangle_{i-calc}}{\sigma_i}\right\} \tag{5}$$

The degrees of freedom is $d = N - v$, in which N is the number of DNA sequences used in the fit, and v is the number of force constants allowed to float in the model. σ_i is the experimentally determined standard error of $\langle\beta_6^2(\infty)\rangle_{i-exp}$. The standard errors were determined from replicate spectral data sets and verified in some instances with multiple sample preparations. The relative error of the fit of the model to the *i*th datum is δ_i.

The four models are listed in order of the number of parameters allowed in each model. There are 16 dinucleotide pairs possible. By symmetry, however, only 10 are unique. For example, an 5′ApA3′ step is identical to a TpT step, therefore the force constant for AA and TT steps would be the same, $\kappa(AA) = \kappa(TT)$ or $\kappa(AA = TT)$. We chose to examine different dinucleotide models to determine whether the analyses are mutually consistent. It may be that a particular data analysis may not be robust enough to handle many adjustable parameters. There may be unknown symmetries that make the parameters of a multiparameter model linearly dependent. Therefore, we are considering four different models, which we present in order of increasing numbers of parameters.

The first model is based on the premise that there are two distinct types of steps. Steps containing only A and T in any form are all the same and have the same force constant, $\kappa(AT = TA = AA = TT)$, and steps containing only G and C in any combination are also all the same and have the same force constant, $\kappa(GC = CG = GG = CC)$. Those steps containing a combination of A or T with C or G have a force constant $\kappa(TG = CA = AC = GT = TC = GA = AG = CT)$, which is the average of the two distinct force constants; i.e., $\kappa^{-1}(TG) = \frac{1}{2}(\kappa^{-1}(AT) + \kappa^{-1}(GC))$. This model is similar to a model used by Hogan and Austin, in which they speculate that GpG and GpC steps are fourfold stiffer than ApA or ApT steps with all remaining basepairs ranking between the two. Their results are based on a Young's Modulus calculation using only poly(dA)-poly(dT), poly(dAdT)-poly(dAdT), poly(dG)-poly(dC), poly(dG-dC)-poly(dG-dC), poly(dAdC)-poly(dGdT), and random type sequences (Hogan et al., 1987).

The second model is a three-parameter model, which assumes that sequences depend only on the purine and pyrimidine nature of the steps. There are then three types of dinucleotide steps: 1) a pyrimidine-purine step with a force constant $\kappa(TA = TG = CG)$, 2) a purine-purine step (which, by symmetry, is the same as a pyrimidine-pyrimidine step) with a force constant $\kappa(TC = CT = AA = GG)$, and 3) a purine-pyrimidine step, where the force constant is $\kappa(AT = GT = GC)$.

The third model allows for six independent force constants on the assumption that the 5′-3′ order of the basepairs in the step does not matter and that the force constant depends only on the type of basepairs in the step. This model may be contrasted with the three-parameter model, which assumes that the order of the base pairs is important. Of the 10 distinct force constant ratios, then, an ApT step is considered the same as a TpA step, and a CpG is the same as GpC, and finally TpG is the same as GpT. Therefore, the six independent force constants are $\kappa(AT = TA)$, $\kappa(AG = CT = TC = GA)$, $\kappa(AC = GT = TG = CA)$, $\kappa(GC = CG)$, $\kappa(CC = GG)$, and $\kappa(AA = TT)$. As an example, the

TABLE 5 Six inverse force constant ratios compared with the averages of the 10 inverse force constant ratios as defined by the formulas below when $N = 31$, χ^2_R probability $\geq 93\%$ for both 10-parameter models ($v = 10$ and $v = 8$), and $\langle\beta^2_6(\infty)\rangle$ values include both A-tract, TATA, and G_3C_3-motif modified values

Name	Averaging formula	$\bar{r}^{-1}$, Average of ratios^{-1} for $v = 10$ model	$\bar{r}^{-1}$, Average of ratios^{-1} for $v = 8$ model	r^{-1}, Six ratios^{-1}, best-fit
AT, TA	$\bar{r}^{-1} = 0.52\, r^{-1}_{(AT)} + 0.48\, r^{-1}_{(TA)}$	$1.2 \pm 0.$	1.14 ± 0.06	1.15 ± 0.01
CG, GC	$\bar{r}^{-1} = 0.52\, r^{-1}_{(CG)} + 0.48\, r^{-1}_{(GC)}$	0.96 ± 0.01	0.96 Fixed	0.94 ± 0.01
AC = GT, TG = CA	$\bar{r}^{-1} = 0.52\, r^{-1}_{(AC = GT)} + 0.48\, r^{-1}_{(TG = CA)}$	1.02 ± 0.06	1.01 ± 0.03	1.05 ± 0.01
TC = GA, AG = CT	$\bar{r}^{-1} = 0.52\, r^{-1}_{(TC = GA)} + 0.48\, r^{-1}_{(AG = CT)}$	0.98 ± 0.07	0.99 ± 0.03	1.04 ± 0.01
GG = CC	No averaging	0.92 ± 0.01	0.91 ± 0.01	0.93 ± 0.01
AA = TT	No averaging	0.84 ± 0.01	0.84 ± 0.01	0.86 ± 0.01

κ(AT = TA) for long runs of ApT repeating steps is some average of ApT and TpA steps. We previously reported κ(AT = TA) (Okonogi et al., 2000), which was found to be ~20% more flexible than average DNA, where κ^{-1} (AT = TA) = $0.52\kappa^{-1}$ (AT) + $0.48\kappa^{-1}$ (TA). The relative weighting of the two types of steps arises because ApT is the first dinucleotide in the test sequence, is therefore closest to the probe, and subsequently has a larger effect on the amplitudes seen at the probe than the other basepairs downstream.

The fourth model allows for 10 independent force constants. The 10 symmetry-constrained, independent force constants are: κ(AT), κ(TA), κ(GC), κ(CG), κ(AG = CT), κ(GA = TC), κ(AC = GT), κ(CA = TG), κ(AA = TT), and κ(GG = CC). This model is the most general dinucleotide model possible and in a sense subsumes the other three. For example, we can now distinguish between κ(AT) and κ(TA) and they are distinct. The parameters in this fourth model are allowed to minimize in two ways. Initially, all 10 force constants are allowed to minimize to their best fit values. Unfortunately, there is minimal sensitivity of the model to those force constants that are quite large. Therefore, in the second minimization, κ(GC) (which is very large) and its partner κ(CG) are held fixed at the initial best fit values when $v = 10$ and the other eight force constants are allowed to minimize to their best-fit values. In this way, the 10 force constants change only slightly (Table 5), and the errors associated with each are statistically meaningful.

The data were analyzed in all four models assuming the probe-containing Q-2AP basepair could be replaced by either a CG base pair or (in a separate analysis) a TA base pair; the results were the same for either assumption.

The χ^2_R value is computed for each model based on the differences between the experimental and calculated $\langle\beta^2_6(\infty)\rangle$ values, the standard errors σ_i, and the degrees of freedom d. The sequence with the largest relative error, δ_i, is removed from the data set, and the set of force constant ratios is recomputed with one fewer sequence. In this way χ^2_R is computed for each model as a function of N. Each datum contributing the largest relative error is identified at each step for each model (see Tables 3a and 4a). The values of χ^2_R are plotted as a function of N, the number of sequences used in the fit, for each of the four models (Figs. 3 *a* and 4 *a*). The reciprocal force constant ratios are listed in Tables 3b and 4b for all the models when $N = 31$. At $N = 31$, nine sequences have been eliminated and the χ^2_R probability $\geq 93\%$ for the 6 and 10 force constant ratios models.

A satisfactory $\chi^2_R \leq 1.0$ value was obtained for the 6 and 10 force constant ratio models using only A-tract modified $\langle\beta^2_6(\infty)\rangle$ values with a χ^2_R probability of 46% (6 ratios), 65% (10 ratios), and 74% (10 ratios with κ(GC) and κ(CG) held fixed) after a total of eight sequences were removed from the data set (32 remaining). Fig. 3 shows the values of χ^2_R as a function of the number of data sets used in the analysis for the four different models, where $v = 2$, 3, 6, 8, and 10. Fig. 4 shows the values of χ^2_R as a function of the number of data sets used in the analysis for the four different models, where $v = 2,3,6,8$, and 10 for the case in which both the A-tract, TATA, and G_3C_3-motif containing sequences $\langle\beta^2_6(\infty)\rangle$ values were modified. A satisfactory $\chi^2_R \leq 1.0$ value was obtained for the 6 and 10 force constant ratio models using

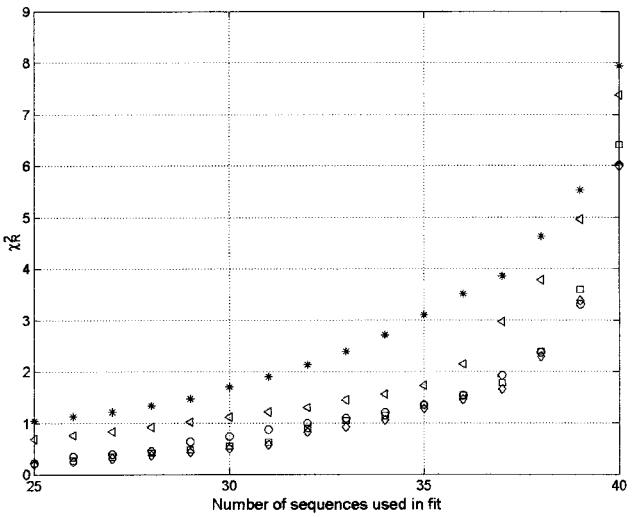


FIGURE 3 χ^2_R values plotted against the number of sequences used in each fit for the two-parameter model (◁), the three-parameter model (*), the six-parameter model (○), the 10-parameter model, $v = 10$ (□), and the 10-parameter model, $v = 8$ (◇) using $\langle\beta^2_6\infty\rangle$ values using A-tract modified values.

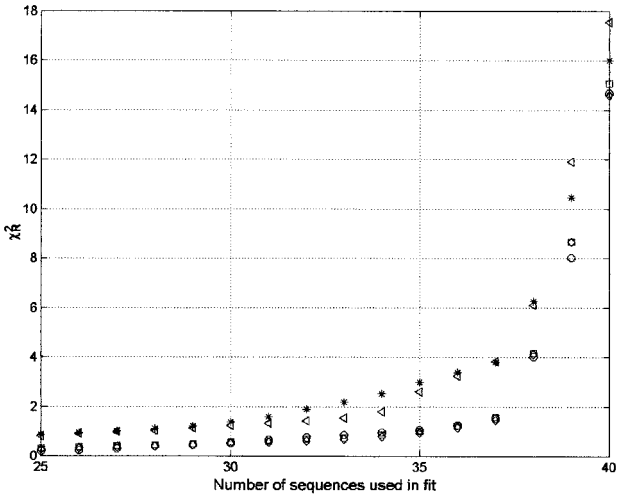


FIGURE 4 Same as Fig. 3 using A-tract, TATA, and G_3C_3-motif modified values.

the A-tract, TATA, and G_3C_3-motif modified values with a χ^2_R probability of 35% (6 ratios), 46% (10 ratios), and 65% (10 ratios with κ(GC) and κ(CG) held fixed) after a total of five sequences were removed from the data set (35 remaining). Fig. 4 shows the values of χ^2_R as a function of the number of data sets used in the analysis for the four different models, where $\nu = 2$, 3, 6, 8, and 10, respectively. When $\chi^2_R \leq 1.0$, the significance of the fit is high. The agreement with the data was such that all models described above eventually obtained a high significance. The models using 6 and 10 force constants obtained significance after discarding eight sequences (for A-tract modified $\langle\beta^2_6(\infty)\rangle$ values only) and after discarding five sequences (for A-tract, TATA, and G_3C_3-motif modified $\langle\beta^2_6(\infty)\rangle$ values). The models using two and three parameters required discarding more sequences to obtain significance (Taylor, 1982, Appendix D).

In Fig. 5 (*a-c*), we show the inverse force constant ratios for the 6 force constant model and the 10 force constant model as a function of the number of data points retained in the analysis when $0.45 \leq \chi^2_R \leq 0.95$ for $\nu = 6$ (Fig. 5 *a*) and when $0.3 \leq \chi^2_R \leq 1.0$ for $\nu = 10$ (Fig. 5 *b*), and $\nu = 8$ (Fig. 5 *c*) floated parameters (for A-tract, TATA, and G_3C_3-motif modified $\langle\beta^2_6(\infty)\rangle$ values). The ratios of the force constants are plotted as reciprocal ratios because the smaller ratios reflect the more flexible sequences; reciprocals show the differences for the more sensitive sequences more clearly. Moreover, large ratios correspond to steps that are very stiff, and a ratio greater than two is essentially infinite. The overall effect of the force constant on dynamics is through the reciprocals of the force constants. The important feature to notice is that the ratios, essentially, do not change within errors over the range shown. This seems to be true for all four models. This suggests that these ratios are close to consensus values. One test of consistency is that the ratio values in the six-parameter model are indeed very close to the averages of the respective values of the 10-parameter model, as illustrated in Table 5. Using the set of $\langle\beta^2_6(\infty)\rangle$ values modified for A-tract, TATA, and G_3C_3-motif sequences for $\nu = 10$ and $\nu = 8$ parameters, the average dinucleotide values are compared with the set of six reciprocal force constant ratios. The averages are taken as reciprocals and illustrates how the reciprocals of the ratios enter into the observed dynamics. Table 5 compares the results of the 10-parameter models ($\nu = 10$ and $\nu = 8$) and the six-parameter model when $N = 31$. Notice that all average values, except for the $r^{-1}(TC = GA, AG = CT)$ average, are the same as the six-parameter model within error.

The results from the list of the 10 force constants in the 10 force constant models used are listed in Table 6 and are indeed quite interesting. The first three entries are pyrimidine-purine type pairs that have inverse ratios greater than 1 with a mean value of $\bar{r}^{-1} = 1.7 \pm 0.2$. The next four entries are purine-purine type pairs with a mean value $\bar{r}^{-1} = 0.9 \pm 0.2$. The final three entries are purine-pyrimidine type pairs with values less than 1 with a mean of $\sim 0.4 \pm 0.4$. The most flexible sequences are those that begin with pyrimidines and end with purines. The least flexible are those that begin with purines and end with pyrimidines. Consider the sequence 5′-XpY. The order of flexibility decreases as X is changed from C to T to A to G, for any Y. Moreover, the order of flexibility increases as Y goes in the same order, from C to T to A to G, for any X. The difference between C and T is very small, as is the difference between A and G. The three-parameter model tests whether the ratios of pyrimidine-purine type steps are close enough to pool, and the answer is that they are not. The three-parameter model did not reduce to a better χ^2_R value than the 10-parameter model for any number of data sets tested (Figs. 3 and 4). This illustrates that the 10-parameter model does have sensitivity to all 10 parameters and that while basepair-type steps can predict the range of flexibility, those are not the only criteria governing flexibility.

DISCUSSION

For all models, the dinucleotide parameterization accounts for the data for the majority of sequences and shows a strong discrimination among the three classes of basepair steps. The four models give consistent results in terms of most and least flexible sequences, and the values from the models with the smaller number of parameters do reflect the averages of the values from the models with more parameters. The 10-parameter model gives the best χ^2_R value for the fewest sequences removed. The six-parameter model fairs quite well in comparison to the 10 because they are the average values, which may make the six-parameter model useful when ranking sequence flexibility. For example, al-

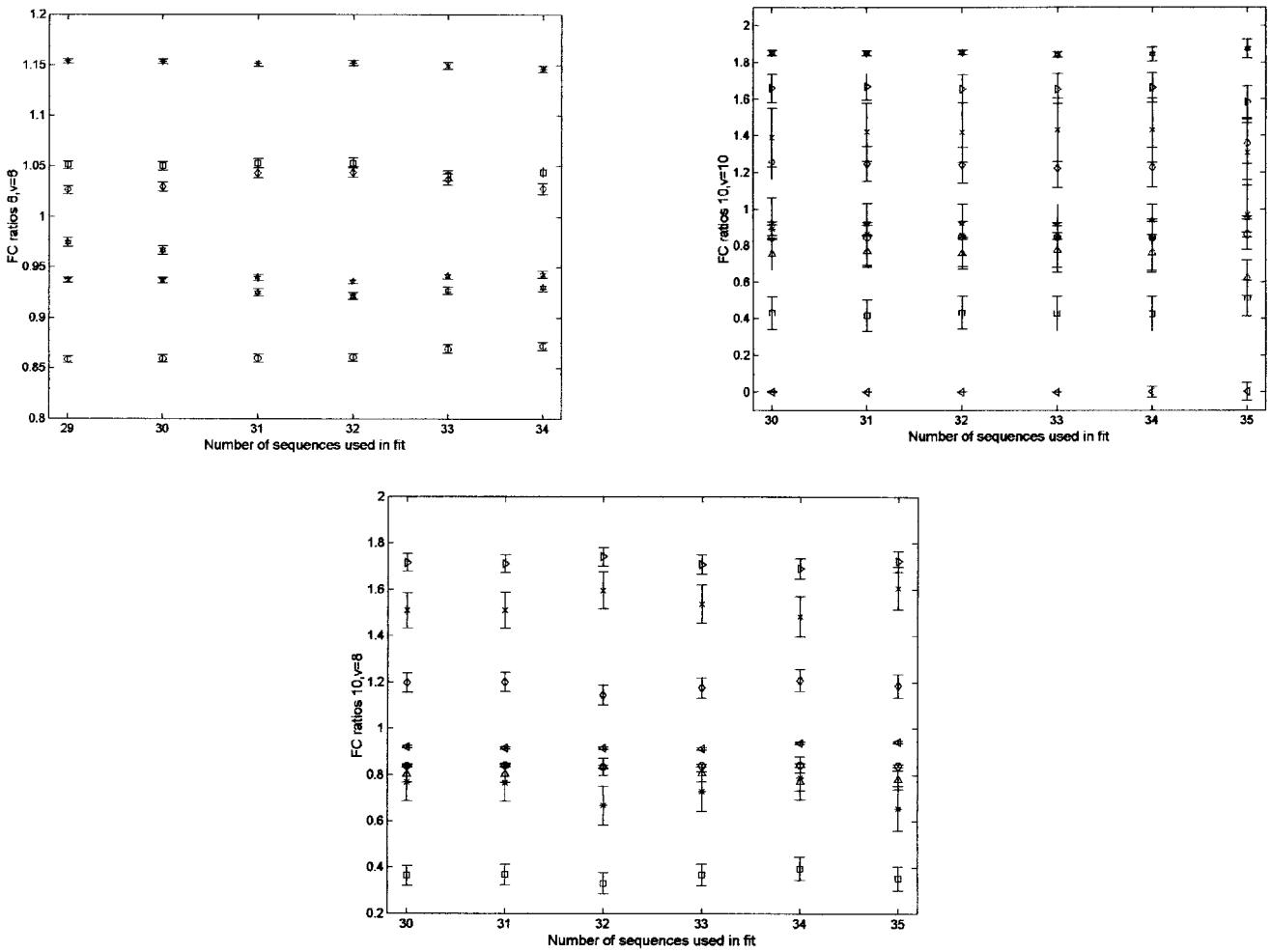


FIGURE 5 Inverse force constant ratios and errors plotted against the number of sequences used in fit for A_5-tract, TATA, and G_3C_3-motif modified $\langle\beta_6^2\infty\rangle$ values when $0.3 \leq \chi_R^2 \leq 1.0$ for (*a*) six-parameter model, $v = 6$, r^{-1} (AT = TA) (∗), r^{-1} (AA = TT) (○), r^{-1} (GT = AC = TG = CA) (□), r^{-1} (CT = GA = TC = AG) (◇), r^{-1} (CC = GG) (☆), r^{-1} (CG = GC) (◯); (*b*) 10-parameter model, $v = 10$, r^{-1} (AT) (∗), r^{-1} (AA = TT) (○), r^{-1} (GT = AC) (□), r^{-1} (CTA = G) (◇), r^{-1} (TA) (×), r^{-1} (GA = TC) (△), r^{-1} (CA = TG) (▷), r^{-1} (GC) (◁), r^{-1} (CC = GG) (★), r^{-1} (CG) (◯); and (*c*) 10-parameter model, $v = 8$, r^{-1} (AT) (∗), r^{-1} (AA = TT) (○), r^{-1} (GT = AC) (□), r^{-1} (CT = AG) (◇), r^{-1} (TA) (×), r^{-1} (GA = TC) (△), r^{-1} (CA = TG) (▷), r^{-1} (CC = GG) (◁), and r^{-1} (CG) and r^{-1} (GC) are fixed. The table below corresponds to the number of sequences used in the fit, degree of freedom, χ_R^2 and percent probability associated with χ_R^2 for the ratios listed in Tables 4b, 5, and 6.

Fig.	Number of sequences	Degrees of freedom	χ_R^2 value	% Probability
5 *a*	31	25	0.67	86
5 *b*	31	21	0.60	93
5 *c*	31	23	0.56	96

though CpG is the most flexible of all dinucleotides in this study, its partner GpC is the most inflexible, and so their average flexibility is closer to 1 than the average for the combined ApT and TpA steps. Therefore, it is not surprising that sequences containing ApT and TpA steps are believed to be the most flexible of all dinucleotides (Harrington and Winicov, 1994).

All four models agree in large measure on the set of the first eight sequences that are to be discarded. These sequences invariably include a few trinucleotide sequences, an AT sequence, and the G_3C_3-motif containing sequences (thought to contain static bends (Brukner et al., 1993)) and a duplex structure resembling A-form DNA (Jayasena and Behe, 1989)). We corrected the mean square amplitudes for major groove bending predicted for G_3C_3-motifs (Brukner et al., 1993); we did not take into account the possibility of A-form behavior when analyzing sequences containing the motif. The corrections we did make could not account for

TABLE 6 Ten inverse force constant ratios $r_i^{-1} = (\kappa_i/\bar{\kappa})^{-1}$ when $N = 31$ and χ^2_R probability $\geq 92\%$

Type	Dinucleotide Name	A_5-tract, TATA, and G_3C_3-motif modified $\langle\beta_6^2(\infty)\rangle$ $(r)^{-1}$, $\nu = 10$	$(r)^{-1}$, $\nu = 8$
pyr-pur	CG	1.85 ± 0.01	1.85 Fixed
pyr-pur	CA = TG	1.67 ± 0.08	1.71 ± 0.04
pyr-pur	TA	1.4 ± 0.2	1.51 ± 0.08
pyr-pyr	CT = AG	1.2 ± 0.1	1.20 ± 0.04
pyr-pyr	CC = GG	0.92 ± 0.01	0.91 ± 0.01
pyr-pyr	TT = AA	0.84 ± 0.01	0.84 ± 0.01
pyr-pyr	TC = GA	0.77 ± 0.09	0.80 ± 0.04
pur-pyr	AT	0.9 ± 0.2	0.77 ± 0.08
pur-pyr	AC = GT	0.42 ± 0.09	0.37 ± 0.04
pur-pyr	GC	0.00 ± 0.00	0.00 Fixed

$\langle\beta_6^2(\infty)\rangle$ values include A-tract, TATA, and G_3C_3-motif modified values.

the anomalous G_3C_3-motifs and they were eliminated. Two sequences (39 and 40 in Table 2) are permutations of one another and phased by 5 bp. Sequence 40 of Table 2 was adequately explained, but 39 was not when no correction for the G_3C_3-motifs bending was applied. After the correction for bending for the G_3C_3-motifs, sequence 39 of Table 2 was adequately explained but 40 was not (see Table 7.) Table 7 lists all 40 sequences and their associated δ_i values using the set of 10 best-fit force constant ratios from Table 6 and the values of $\langle\beta_i^2\rangle_{exp}$ corrected for bends in A-tract and G_3C_3 motifs. The nine sequences with the largest δ_i values were removed sequentially until the final fit was obtained as previously discussed. Positive δ_i values ($\langle\beta_i^2\rangle_{exp} - \langle\beta_i^2\rangle_{calc} > 0$) correspond to sequences that are more flexible than predicted by the model, and negative δ_i values correspond to sequences that are stiffer than predicted. Of the trinucleotide sequences, only CCG5_s5, CC7C_s5, ATC5, and ACG5_s5 were removed, and all but CC7C_s5 are stiffer than consensus. Each of the eliminated trinucleotide sequences has a partner, shifted by 5 bp; that was determined to be acceptable. We suggest that there may be either some permanent bends in these structures or additional structures that interconvert on a timescale long compared with the timescale of the EPR experiments. Structures of this sort we term slowly relaxing structures (Schurr et al., 1997b). In summary, there are only a few outlying triplet repeats that cannot be explained by the model, indicating that what determines duplex DNA's flexibility and structure, in some cases may be more than the simple sum of the dinucleotide parts. It is noteworthy that the only dinucleotide sequence considered to be anomalous is AT7A_s12 (see Table 4a and 7).

Lundin et al. (1994) investigated changes in MIG1 zinc finger binding by mutating MIG1's GC box binding site and the flanking AT box found naturally in all MIG1 sites. Out of dozens of mutations to the GC box, all but three halted binding while no single mutation within the AT box had the same effect. Any one G or C mutation in the AT box only repressed binding, whereas any combination of all AT steps guaranteed high affinity binding. Two or more G/C muta-

TABLE 7 δ_i for all 40 sequences using the final set of 10 force constant ratios ($\kappa_i\bar{\kappa}$) from Table 6 for $\nu = 8$

	Sequence	δ_i ($N = 31$)
1	CG10	−0.003
2	ACG5	−0.015
3	AAG5	−0.113
4	AG7A	0.128
5	AT7A_s5	0.142
6	AT4	−0.256
7	AT10	−0.258
8	AAT5	−0.260
9	NT	−0.300
10	CG7C_s5	0.370
11	AT7A_s24	−0.405
12	AT7A	0.431
13	G_3C_3-motif/A_5tract	0.465
14	CAG5	0.506
15	AAC_s5	−0.509
16	ACT5	−0.517
17	AA7A (AAA5)	−0.561
18	AGG5_s5	−0.632
19	AA10	0.643
20	TATA	−0.671
21	CC10	−0.674
22	CC7C (CCC5)	0.700
23	AA7A_s5 (AAA5_s5)	−0.744
24	CCG5	0.761
25	AC7A	−0.835
26	CG7C	−0.904
27	1/2CAP	0.923
28	AT15	0.966
29	AAC5	1.033
30	AGG5	1.186
31	G_3C_3-motif_s0_s8	−1.235
32	AT7A_s12	−1.510
33	CCG5_s5	−1.554
34	CC7C_s5 (CCC5_s5)	2.059
35	ATC5	−2.634
36	ACG5_s5	−3.427
37	G_3C_3-motif	−3.616
38	G_3C_3-motif_s1	−10.357
39	A_5tract/G_3C_3-motif	−12.671
40	G_3C_3-motif_s8	−16.913

Sequences are thrown out in fit based on δ_i.

tions in the AT box halted binding altogether. The investigators concluded that the flexibility of the AT-rich flanking sequence guaranteed the bending essential for GC box binding. Gel shift assays confirmed that MIG1 does indeed bend the DNA at the AT box upon binding. They concluded that "MIG1 recognizes the AT box directly, but in a way that requires bendable DNA rather than a unique sequence motif." It would seem that here again increased flexibility is an average of ApT and TpA steps.

The timescale of the dynamics of the EPR experiments are submicrosecond, which in large part then only pertain to short time processes. However, it is interesting to compare with other dynamics studies carried out on longer timescales. Kennedy and co-workers (McAteer et al., 1995; McAteer and Kennedy, 2000) studied the line widths of the NMR signals of hydrogens on adenines in short sequences of DNA; they observed that the proton NMR spectra for hydrogen on the A in TpA steps were very mobile. This type of step was contrasted with the ApT step, also contained in the sequences, which showed no enhanced mobility. The NMR experiments also revealed context effects on the dynamics, which requires developing models with sequences greater than dinucleotide steps. The dynamics of these NMR events are on the order of micro- to milliseconds and are most likely due to the existence of slowly relaxing structures and the interconversion among B-form structures of DNA. The submicroscond dynamics, which are observed in the EPR studies, are only part of the dynamics seen on these longer timescales. However, the trend we observe, which is that TpA steps are flexible while ApT steps are inflexible, correlates quite well with the dynamics inferred from the NMR experiments.

Similarly, Drobny and co-workers (Geahigan et al., 2000; Hatcher et al., 1998) when studying the dodecamer d-($C_1GC_3GAATTC_9GC_{11}G$) via solid-state deuterium NMR, reported large amplitude motions in the furanose ring and backbone for only the internal CpG steps (at the 3 and 9 positions only). The CpG steps at positions 1 and 11 did not show any large amplitude motions, but it is not clear whether end effects play a role in the dynamics. All other steps showed very small amplitude furanose ring and backbone dynamics. The dynamics observed by NMR are on the micro- to millisecond timescale. Nonetheless, it is interesting to notice that the CpG is the only one to show large amplitude internal dynamics. This correlates well with the general findings of this work, that pyrimidine-purine steps are the most flexible and CpG is the only pyrimidine-purine step in this sequence (Table 6).

Circularization rate studies performed by Bacolla et al. (1997) are on a much longer timescale (minutes to hours) than our EPR studies. Bacolla et al. (1997) studied the flexibility of the two triplet repeat sequences (TRS), $(CTG)_n$ $(CAG)_n$, and $(CGG)_n(CCG)_n$ where *n* varied from 36 to 80 and from 24 to 73, for the two TRSs, respectively (Bacolla et al., 1997). Bacolla et al. (1997) concluded that the more rapid rate of closure, or circularization, of these sequences could be attributed to a 40% increase in flexibility of these TRSs relative to "normal" B-form DNA. From enzymatic digest experiments, Bacolla et al. (1997) concluded that $(CGG)_n$ $(CCG)_n$ remained within the family of B-form structures. They further note that the increased flexibility of these sequences is the only physical characteristic to correlate with the diseases these sequences induce. Pearson and Sinden (1996) review Bacolla et al.'s results and place them in the context of altered structures such as slipped-strands and hairpins and describe the possible multiple morphologies of structures accessible to TRS. The results presented here when compared with those of Bacolla et al. may enhance the discussion on the physical characteristics of TRSs. As we have noted CG steps are the most flexible and GC steps are the least flexible, being approximately fourfold stiffer than CG steps. As an averaged structure, based on our estimates, the $(CGG)_n(CCG)_n$ sequences would only be ~10% more flexible than "normal" DNA. This averaged effect does not rule out interesting local effects on dynamics, but local regions of differential flexibilities based on submicrosecond dynamics are unlikely to explain the results of Bacolla et al. (1997).

The dynamics of DNA occur on multiple time scales: subnanosecond dynamics characterize DNA interactions on the individual base pair level, whereas submicrosecond interactions involve dynamics of the collective modes (Okonogi *et al.*, 1999), which are the quantities reported in these studies. As previously published (Okonogi et al., 1999, 2002; Schurr et al., 1997a), the timescale of our measurements leaves open the possibility that additional dynamics are occurring in the millisecond (and longer) timescale. Our suggestion is that the submicrosecond dynamics, observed in our experiments, are sensitive to excursions of basepairs and collective modes of basepairs about local minima. There may be many local minima (or local structural basins) that are nearly energetically equivalent and that interconvert on longer time scales. This is another way of saying that DNA is polymorphic. Motion not seen on the submicrosecond (and probably on the submillisecond) time scales is motion among the slowly relaxing structures. The principle of polymorphic DNA applied to the TRSs suggests that there might be a set of bent structures within the family of B-forms whose set is not explored in the submicrosecond dynamics. It may well be that on the short timescale TRS sequences are stiff. Moreover, a stiff sequence implies that there is little entropic stability in that lowest energy form, which would suggest that conversion to other forms would be relatively favorable entropically. If this were the case, and there were additional forms easily accessed by TRSs (such as $(CGG)_n(CCG)_n$ sequences), these structures might appear to be more flexible in a circularization rate study such as that carried out by Bacolla et al. (1997).

Pearson and Sinden point out that "$(CGG)_n(CCG)_n$ repeats assemble into nucleosomes with the least efficiency of any known sequence" (Pearson and Sinden 1998b). Pearson and Sinden also state that although $(CTG)_n(CAG)_n$ and $(CGG)_nCCG)_n$ "triple repeats are flexible, they almost certainly have very different biological properties in terms of the chromatin organization upon expansion in the disease state." They conclude by noting, "any direct relationship between DNA structure and loci-specific expansion leading to human disease remains an enigma." Whereas there is some controversy concerning nucleosomal assembly of the TRS, it seems that unmethylated triplet repeats show poor nucleosome assembly but do assemble (Coffee et al., 1999). At first it seems difficult to reconcile poor nucleosomal assembly with the observation that this TRS appears to be 40% more flexible, because more flexible DNAs generally assemble easily into nucleosomes. In light of all of these results one might consider that $(CGG)_n(CCG)_n$ might be difficult to assemble on nucleosomes because the individual members of the set of slowly relaxing structures are stiff and that supramillisecond time-dependent interconversion among the set of relaxing structures is the basis for the enhance flexibility observed in the circularization rate study.

CONCLUSION

The dynamic bending force constants for the 10 possible unique dinucleotides have been measured and rank ordered according to their increase or decrease in flexibility relative to the average bending force constant for random DNA. The submicrosecond, sequence-dependent dynamic bending force constants, directly proportional to the dynamic persistence lengths (Okonogi et al., 1999), based on a number of dinculeotide models, are presented here for the first time. The novel finding is that pyrimidine-purine type steps are the most flexible, purine-purine steps are about average, and purine-pyrimidine steps are the most inflexible. The range in flexibility is approximately fourfold. Whereas the dinucleotide model is valuable in understanding or predicting increased flexibility among some sequences, it cannot be used universally: certain combinations of dinucleotides may possess dynamics characteristics that are more than simply additive due to structural changes or anomalies occurring on a wide range of timescales. These new results provide a set of rank ordering of sequence-dependent flexibility on the submicrosecond timescale and may provide basic information for understanding the dynamics within the DNA regulation and repair processes.

This work was supported in part by grants from the National Institute of Health (GM 32681, GM 55963, and GM 65944) and NIEHS Environmental Sciences Center Grant P30 ESO7033. We are grateful to J.M. Schurr for reading this manuscript and many helpful suggestions.

REFERENCES

Aloyo, M. C., D. Campbell, N. J. Combates, J. Gonzales, Y. Kwok, R. D. Sheardy, and S. A. Winkle. 1993. Restriction enzymes have altered cleavage rates at sites near certain sequences. *Biophys. J.* 64:A280.

Anderson, I. D., and J. Widom. 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.* 296:979–987.

Anselmi, C., G. Bocchinfuso, P. De Santis, M. Savino, and A. Scipioni. 1999. Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability. *J. Mol. Biol.* 286:1293–1301.

Bacolla, A., R. Gellibolian, M. Shimizu, S. Amirhaeri, S. Kang, K. Ohshima, J. E. Larson, S. C. Harvey, B. D. Sollar, and R. D. Wells. 1997. Flexible DNA: genetically unstable CTG-CAG and CGG-CCG from human hereditary neuromuscular disease genes. *J. Biol. Chem.* 272: 16783–16792.

Brukner, I., M. Dlakic, A. Savic, S. Susic, S. Pongor, and D. Suck. 1993. Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucleic Acids Res.* 21:1025–1029.

Chastain, P. D., and R. R. Sinden. 1998. CTG repeats associated with human genetic disease are inherently flexible. *J. Mol. Biol.* 275: 405–411.

Chedin, F., E. Dervyn, R. Dervyn, S. D. Ehrlich, and P. Noirot. 1994. Frequency of deletion formation decreases exponentially with distance between short direct repeats. *Mol. Microbiol.* 12:561–569.

Coffee, B., F. Zhang, S. T. Warren, and D. Reines. 1999. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat. Genet.* 22:98–101.

Dlakic, M., and R. E. Harrington. 1995. Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J. Biol. Chem.* 270:29945–29952.

Filesi, I., S. Cacchione, P. De Santis, L. Rossetti, and M. Savino. 1999. The main role of the sequence-dependent DNA elasticity in determining the free energy of nucleosome formation on telomeric DNAs. *Biophys. Chem.* 83:223–237.

Fujimoto, B. S., and J. M. Schurr. 1990. Dependence of the torsional rigidity of DNA on base composition. *Nature.* 344:175–177.

Garcia Bernal, J. M., and J. Garcia de la Torre. 1980. Transport properties and hydrodynamic centers of rigid macromolecules with arbitrary shapes. *Biopolymers.* 19:751–766.

Geahigan, K. B., G. A. Meints, M. E. Hatcher, J. Orban, and G. P. Drobny. 2000. The dynamic impact of CpG methylation in DNA. *Biochemistry.* 39:4939–4946.

Grove, A., A. Galeone, L. Mayol, and E. P. Geiduschek. 1996. Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J. Mol. Biol.* 260:120–125.

Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Chem.* 17:265–286.

Harrington, R. E., and I. Winicov. 1994. New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. *Prog. Nucleic Acid Res. Mol. Biol.* 47:195– 270.

Hatcher, M. E., D. L. Mattielo, G. A. Meints, J. Orban, and G. P. Drobny. 1998. A solid-state deuterium NMR study of the localized dynamics at the C9pG10 step in the DNA Dodecamer [d(CGCGAATTCGCG)]2. *J. Am. Chem. Soc.* 120:9850–9862.

Hogan, M. E., and R. H. Austin. 1987. Importance of DNA stiffness in protein-DNA binding specificity. *Nature.* 329:263–266.

Hogan, M. E., M. W. Roberson, and R. H. Austin. 1989. DNA flexibility variation may dominate DNase I cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 86:9273–9277.

Hogan, M. E., T. F. Rooney, and R. H. Austin. 1987. Evidence for kinks in DNA folding in the nucleosome. *Nature.* 328:554–557.

Hustedt, E. J., J. J. Kirchner, A. Spaltenstein, P. B. Hopkins, and B. H. Robinson. 1995. Monitoring DNA dynamics using spin-labels with different independent mobilities. *Biochemistry.* 34:4369–4375.

Hustedt, E. J., A. Spaltenstein, J. J. Kirchner, P. B. Hopkins, and B. H. Robinson. 1993. Motions of short DNA duplexes: an analysis of DNA dynamics using an EPR-active probe. *Biochemistry.* 32:1774–1787.

Jayasena, S. D., and M. J. Behe. 1989. Nucleosome reconstitution of core-length poly(dG)-poly(dC) and poly(rG-dC)-poly(rG-dC). *Biochemistry.* 28:975–980.

Kim, U. S., B. S. Fujimoto, C. E. Furlong, J. A. Sundstrom, R. Humbert, D. C. Teller, and J. M. Schurr. 1993. Dynamics and structures of DNA: long-range effects of a 16 base-pair (CG)8 sequence on secondary structure. *Biopolymers.* 33:1725–1745.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature.* 389:251–260.

Lundin, M., J. O. Niehlin, and H. Ronne. 1994. Importance of a flanking AT-rich region in target site recognition by the GC box-binding zinc finger protein MIG1. *Mol. Cell. Biol.* 14:1979–1985.

McAteer, K., P. D. Ellis, and M. A. Kennedy. 1995. The effects of sequence context on base dynamics at TpA steps in DNA studied by NMR. *Nucleic Acids Res.* 23:3962–3966.

McAteer, K., and M. A. Kennedy. 2000. NMR evidence for base dynamics at all TpA steps in DNA. *J. Biomol. Struct. Dynam.* 17:1001–1009.

Nardulli, A. M., L. E. Romine, C. Carpo, G. L. Greene, and B. Rainish. 1996. Estrogen receptor affinity and location of consensus and imperfect estrogen response elements influence transcription activation of simplified promoters. *Mol. Endocrinol.* 10:694–704.

Okonogi, T. M., S. C. Alley, E. A. Harwood, P. B. Hopkins, and B. H. Robinson. 2002. Phosphate backbone neutralization increases duplex DNA flexibility. *Proc. Natl. Acad. Sci. U. S.A.* 99:4156–4160.

Okonogi, T. M., S. C. Alley, A. W. Reese, P. B. Hopkins, and B. H. Robinson. 2000. Sequence dependent dynamics in duplex B-DNA. *Biophys. J.* 78:2560–2571.

Okonogi, T. M., A. W. Reese, S. C. Alley, P. B. Hopkins, and B. H. Robinson. 1998. How far does bending flexibility transmit in duplex DNA. *Biophys. J.* 72:A287 (Poster).

Okonogi, T. M., A. W. Reese, S. C. Alley, P. B. Hopkins, and B. H. Robinson. 1999. Flexibility of duplex DNA on the sub-microsecond timescale. *Biophys. J.* 77:3256–3276.

Parkhurst, K. M., R. M. Richards, M. Brenowitz, and L. J. Parkhurst. 1999. Intermediate species possessing bent DNA are present along the pathway to formation of a final TBP-TATA complex. *J. Mol. Biol.* 289: 1327–1341.

Patikoglou, G., and S. K. Burley. 1997. Eukaryotic transcription factor - DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* 26:289–325.

Pearson, C. E., and R. R. Sinden. 1998a. Slipped strand DNA, dynamic mutations, and human disease. In: Genetic Instabilities and Hereditary Neurological Diseases. R. D. Wells, S. T. Warren, editors Academic Press, San Diego CA. 585–621.

Pearson, C. E., and R. R. Sinden. 1998b. Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr. Opin. Struct. Biol.* 8:321–330.

Ramsauer, V., L. Aguilar, M. Ballester, R. D. Sheardy, and S. A. Winkle. 1997. The conformation of (CG) DNA segments regulates RNA polymerase activity. *Biophys. J.* 72:A98.

Richmond, T. J., J. T. Finch, B. Rushton, D. Rhodes, and A. Klug. 1984. Structure of the nucleosome core particle at 7 A resolution. *Nature.* 311:532–537.

Robinson, B. H., and G. P. Drobny. 1995a. Site-specific dynamics in DNA: theory. *Annu. Rev. Biophys. Biomol. Struct.* 24:523–549.

Robinson, B. H., and G. P. Drobny. 1995b. Site-specific dynamics in DNA: theory and experiment. *Methods Enzymol.* 261:451–509.

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

Schurr, J. M., J. J. Delrow, B. S. Fujimoto, and A. S. Benight. 1997a. The question of long-range allosteric transitions in DNA. *Biopolymers.* 44: 283–308.

Schurr, J. M., and B. S. Fujimoto. 2000. The distribution of end-to-end distances of the weakly bending rod model. *Biopolymers.* 54:561–571.

Schurr, J. M., B. S. Fujimoto, and B. H. Robinson. 1997b. Diffusional spinning as a probe of DNA fragment's conformation [Comment]. *J. Chem. Phys.* 106:815–816.

Sinden, R. R. 1999. Human genetics '99: trinucleotide repeats biological implications of the DNA structures associated with disease-causing triplet repeats. *Am. J. Hum. Genet.* 64:346–353.

Sivolob, A. V., and S. N. Khrapunov. 1995. Translational positioning of nucleosomes on DNA: the role of sequence-dependent isotropic DNA bending stiffness. *J. Mol. Biol.* 247:918–931.

Takahara, P. M., A. C. Rosenzweig, C. A. Frederick, and S. J. Lippard. 1995. Crystal structure of double-stranded DNA containing the major adduct of the anticancer drug cisplatin. *Nature.* 377:649–652.

Taylor, J. R. 1982. An introduction to error analysis. University Science Books, Mill Valley, CA.

Thastrom, A., P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, and J. Widom. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome postioning DNA sequences. *J. Mol. Biol.* 288: 213–229.

Thompson, B. J., M. N. Camien, and R. C. Warner. 1976. Kinetics of branch migration in double-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* 73:2299–2303.

Widlund, H. R., P. N. Kuduvalli, M. Bengtsson, H. Cao, T. D. Tullius, and M. Kubista. 1999. Nucleosome structural features and intrinsic properties of the TATAAACGCC repeat sequence. *J. Biol. Chem.* 274: 31847–31852.

Wolffe, A. P. 1994. Architectural transcription factors [comment]. *Science.* 264:1100–1101.