Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6464–6468, July 1995
Biophysics

An A-DNA triplet code: Thermodynamic rules for predicting A- and B-DNA

(DNA structure/structure prediction/crystallography/circular dichroism/spectroscopy)

BETH BASHAM, GARY P. SCHROTH, AND P. SHING HO*

Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, OR 97331

Communicated by Kensal E. van Holde, Oregon State University, Corvallis, OR, March 20, 1995

ABSTRACT **The ability to predict macromolecular conformations from sequence and thermodynamic principles has long been coveted but generally has not been achieved. We show that differences in the hydration of DNA surfaces can be used to distinguish between sequences that form A- and B-DNA. From this, a "triplet code" of A-DNA propensities was derived as energetic rules for predicting A-DNA formation. This code correctly predicted >90% of A- and B-DNA sequences in crystals and correlates with A-DNA formation in solution. Thus, with our previous studies on Z-DNA, we now have a single method to predict the relative stability of sequences in the three standard DNA duplex conformations.**

A long held precept in biochemistry is that the conformation of a biomolecule is defined by, and thus can be predicted from, its sequence. This is the "protein folding problem" for polypeptides (1). An apparently simpler, although no less trivial problem lies in predicting the structure of DNA, a highly polymorphic molecule. Despite the large volume of data on DNA structures in solution, in fibers, and in single crystals, there are currently no general rules to accurately predict the ability of a sequence to adopt one of the standard duplex conformations of A-, B-, and left-handed Z-DNA. This is the analogous "DNA-folding problem." We present here a set of thermodynamic rules, based on hydration of DNA surfaces, to distinguish between A- and B-DNA-forming sequences in crystals and in solution.

Soon after Watson and Crick described the structure of B-DNA (2), the alternative A-DNA conformation was reported by Franklin and Gosling (3). A-DNA is a shorter, broader helix as compared with B-DNA and is characterized by (*i*) base pairs that are highly inclined and displaced away from the helix axis and (*ii*) a helical repeat of 11 bp per turn as compared with 10–10.5 bp per turn for B-DNA (4). Although the conformation of DNA in fibers apparently depends on its water content and sequence (5), this relationship has not generally been exploited to predict sequences that form A-DNA.

Predicting DNA secondary structure has traditionally relied on simple sequence rules. For example, Z-DNA is often assigned to alternating pyrimidine-purine sequences that are d(C+G) rich (6–8), while nonalternating d(G+C)-rich sequences that contain d[(CC)·(GG)] steps favor A-DNA (8). These rules, although successful in predicting simple A-DNA (9) and Z-DNA (7) sequences, are not general for the large number of oligonucleotides that have been crystallized as A- and B-DNA. For example, double-stranded d(CCAGGCCTGG) and d(ACCGGCCGGT) have the same base compositions and number of d[(CC)·(GG)] steps but crystallize as B- and A-DNA, respectively (10, 11). Furthermore, A-DNA has been crystallized in sequences that are 50% d(A+T), and at least one B-DNA sequence contains only d(C·G) base pairs (12).

We had previously shown that the sequence-dependent stability of Z-DNA is related to the free energy required to hydrate exposed DNA surfaces (the solvent free energies, or SFEs). Z-DNA was found to be more hydrophobic than B-DNA (13). This difference becomes greater for sequences that are less stable as Z-DNA (14) and is consistent with Z-DNA being stabilized by dehydrating conditions (6, 7).

A-DNA is also dehydrated compared with B-DNA (15). The position of waters around a DNA structure is correlated with its conformation (16). Therefore, A-DNA stability should also be dependent on the hydration of the DNA surface (17). Here, we use SFEs to derive a set of thermodynamic rules to predict sequences that form A- or B-DNA. First, the differences in SFEs of A- and B-DNA models ($\Delta SFE_{A\text{-}B}$) were determined for sequences that have been crystallized as A-DNA and B-DNA. From this, we derive a set of thermodynamic rules to describe the context-dependent A-DNA propensities for individual base pairs. These rules accurately predict the conformations of oligonucleotides that have been crystallized as A- and B-DNA and predict the behaviors of well-defined oligonucleotide sequences in solution.

MATERIALS AND METHODS

SFE Calculations. For this work, we assembled a data set of A- and B-DNA crystal structures (those solved to 2.7 Å or better) that contain only standard Watson–Crick base pairs, that have standard bases and phosphoribose backbones, and that did not contain any drugs or other known ligands. This includes 17 unique B-DNA sequences and 17 A-DNA sequences (Table 1), as listed in the Nucleic Acids Data Base (18) or the Brookhaven Protein Data Base (19). A- and B-DNA models were constructed for each sequence using the program HyperChem (AutoDesk, Sausalito, CA). Although the "best" model for each sequence may be its crystal structure, no sequence in this data set has been crystallized in both conformations. To treat all sequences equally and consistently, the models were generated using standard helical parameters (20).

The SFE of each DNA model was calculated as described (13). In short, the solvent-accessible surface(s) (SAS) of the DNA was determined by rolling a probe (1.45 Å in radius) over its structure (21). SFEs were calculated by applying an atomic solvation parameter (ASP) to the SAS (13, 14) ($SFE = \Sigma\, SAS_i \times ASP_i$ for each atom type i). The terminal base pairs were excluded from each SFE calculation because of our inability to accurately model the hydration of the ends (13).

We had previously shown (13) and confirm in this study that the SFEs of B-DNA crystal structures are variable at the base-pair level. When averaged over the entire sequence, however, the SFEs of the crystal and model structures are

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SFE, solvent free energy; APE, A-DNA propensity energy; TFE, trifluoroethanol.

*To whom reprint requests should be addressed.

Table 1. SFEs of A- and B-DNA sequences

Sequence	Conformation	SFE_A	SFE_B	ΔSFE_{A-B}
d(CCAGGCCTGG)*	B	−4.89	−5.47	0.58
d(CCAACGTTGG)*	B	−4.66	−5.03	0.37
d(CGATCGATCG)	B	−4.44	−4.99	0.55
d(CGATTAATCG)	B	−3.83	−4.55	0.72
d(CGATATATCG)	B	−3.72	−4.56	0.84
d(CCGGCGCCGG)	B	−5.24	−5.94	0.70
d(CATGGCCATG)	B	−4.52	−4.89	0.37
d(CCAAGCTTGG)*	B	−4.69	−5.08	0.39
d(CCATTAATGG)	B	−3.84	−4.50	0.67
d(CGTGAATTCACG)	B	−3.99	−4.87	0.88
d(CGTAGATCTACG)	B	−4.02	−4.83	0.81
d(CGCGAATTCGCG)	B	−4.26	−5.43	1.17
d(CGCATATATGCG)	B	−4.03	−5.01	0.98
d(CGCAAATTTGCG)	B	−4.17	−4.79	0.62
d(CGCAAAAAAGCG)*	B	−4.04	−4.90	0.87
d(CGCGAAAAAACG)*	B	−4.12	−4.93	0.81
d(CGCAAAAATGCG)	B	−4.17	−5.14	0.97
B-DNA average (SD)				0.72 (0.23)
d(GCCGGC)*	A	−6.32	−5.91	−0.41
d(GGGGCCCC)	A	−5.80	−5.97	0.17
d(GGGATCCC)	A	−4.99	−5.29	0.30
d(GCCCGGGC)	A	−5.97	−5.85	−0.12
d(CCCCGGGG)	A	−5.93	−5.84	−0.09
d(GTACGTAC)	A	−4.52	−4.57	0.05
d(CTCTAGAG)	A	−4.62	−4.58	−0.04
d(GTGTACAC)	A	−4.32	−4.44	0.12
d(GGGCGCCC)	A	−5.79	−5.93	0.14
d(GGGTACCC)	A	−4.90	−5.29	0.39
d(GTCTAGAC)	A	−4.56	−4.60	0.04
d(ACCGGCCGGT)	A	−5.29	−6.11	0.82
d(GCGGGCCCGC)	A	−5.25	−5.96	0.71
d(CCCGGCCGGG)	A	−5.27	−5.89	0.63
d(CCCCCGCGGGGG)	A	−4.90	−5.29	0.39
d(CCGTACGTACGG)	A	−4.18	−5.11	0.93
d(GCGTACGTACGC)	A	−4.18	−5.13	0.95
A-DNA average (SD)				0.19 (0.33)

The SFE (kcal/mol per bp) of B-DNA (SFE_B) and A-DNA (SFE_A) models built from standard helical parameters are compared. The difference (ΔSFE_{A-B}) reflects the relative stability of A- and B-DNA sequences.

*These sequences were not used in the derivation of APEs in Table 2.

nearly identical ($\Delta SFE = 0.02 \pm 0.29$ kcal/mol per bp), indicating that the hydration of B-DNA in the crystal is accurately represented by the SFE of the model structures. The crystal and model structures of A-DNA hexanucleotides are very similar (22), as are their respective SFEs. Not surprisingly, however, the SFEs of A-DNA octanucleotides, which are greatly distorted by crystal lattice effects (23), are very different. This was further impetus for using the idealized models to calculate SFEs, since it is difficult to accurately account for crystal-packing effects and to generate A-DNA models equivalent to a crystal structure for sequences that have only been crystallized as B-DNA.

Solution Studies. Oligonucleotides for solution studies were synthesized on an Applied Biosystems DNA synthesizer in the Center for Gene Research and Biotechnology at Oregon State University. These were passed over a Sephadex G-25 column and subsequently annealed in 3.33 mM Tris (pH 8.1) buffer containing 0.03 mM EDTA and 50 mM NaCl. The sequences were diluted to 0.4 absorbance unit in a 10 mM Tris (pH 8.1) buffer containing 0.02 mM EDTA and were titrated with trifluoroethanol (TFE, from Sigma). The solubility of the sequences defined the maximum concentration of TFE added in the titrations. Circular dichroism (CD) spectra of each sample were recorded on a JASCO J-720 spectrometer.

RESULTS

A- and B-DNA Data Set. In this study, we use a data base of A- and B-DNA crystal structures to define the sequence determinants for A-DNA formation. Is the ability to crystallize a DNA sequence in a particular conformation a good indicator of its stability in that conformation? It has been suggested that the DNA conformation in a crystal is strongly influenced by the crystallization solutions and by crystal lattice forces (23). To address this first point, we compared the concentration of salts and alcohol precipitants reported for crystallizing A- and B-DNA (12). We had previously observed that the cation strength ($CS = \Sigma[Z_i^2 \times C_i]$, where Z_i is the charge and C_i the concentration of each cation type i) to crystallize hexanucleotides as Z-DNA is related to the stability of these sequences as Z-DNA (14, 24). For A- and B-DNA, this relationship does not hold. On average, $CS = 0.2 \pm 0.3$ M per mmol of phosphate for A-DNA sequences and 0.2 ± 0.2 M per mmol of phosphate for B-DNA. (Crystallization conditions were taken from ref. 12.) Only cobalt hexaamine, which induces A-DNA in solution (25), directly affects the conformation of the DNA in the crystal. In addition, neither the temperature nor precipitant concentrations were correlated with the crystallization of either conformation. Although most oligonucleotides require 2-methyl-2,4-dimethylpentanediol (MPD) or some other alcohol precipitant for crystallization, several A-DNA crystals have been obtained in the absence of any alcohol (23). Whether a sequence crystallizes as A- or B-DNA is therefore not generally defined by the crystallization conditions, with the caveat that these solutions are normally more hydrophobic than standard aqueous buffers.

Some have also suggested that the conformations of DNA in crystals are defined primarily by lattice-packing forces and that sequence length is an important determinant of structure. However, Dickerson *et al.* (26) have argued against this "tyranny of the lattice." For instance, decamer and dodecamer sequences have been crystallized as both A- and B-DNA. In addition, hexanucleotides have been crystallized as B-DNA (27) and Z-DNA (6) and most recently as A-DNA (22), indicating that hexamers "fit" into the crystal lattices of all three DNA conformations. Therefore, the conformation in the crystal is not strictly defined by the DNA length. Here, we use idealized models for A- and B-DNA to avoid any potential influence of crystal packing effects on our results.

Distinguishing A- and B-DNA by SFEs. We would expect that A-forming sequences will be more hydrophilic as A-DNA than as B-DNA, and vice versa. Thus, the difference in SFE for A- versus B-DNA (ΔSFE_{A-B}) of an A-forming sequence should be lower than that of a B-forming sequence. This is indeed the case. The average ΔSFE_{A-B} for A-DNA sequences is 0.53 kcal/mol per bp lower than that of the B-DNA sequences (Table 1). The sequences fall into two distinct populations, with a 99.8% confidence limit as defined by a standard t-test (Fig. 1). Thus, ΔSFE_{A-B} can discriminate between sequences that crystallize as A- or B-DNA.

A $\Delta SFE_{A-B} = 0.50$ kcal/mol per bp [$\approx$1 standard deviation (SD) from the mean of each population] was used as the criterion to distinguish A- from B-DNA sequences. This being a positive value suggests that most of the sequences, including those crystallized as A-DNA, are primarily B-DNA in aqueous solution. The ΔSFE_{A-B} values presented here apply primarily to the more hydrophobic crystallization solutions. We would expect, however, that sequences with $\Delta SFE_{A-B} \leq 0.5$ kcal/mol per bp should have some A-DNA characteristics, even though the oligonucleotides would be predominantly B-DNA in aqueous solution. From this, we derived an A-DNA propensity energy (APE) = $\Delta SFE_{A-B} - 0.50$ kcal/mol per bp, as the thermodynamic propensity of a sequence to adopt the A-DNA conformation. The uncertainty in discriminating between A- and B-DNA according to APEs is 0.03 kcal/mol per bp (the

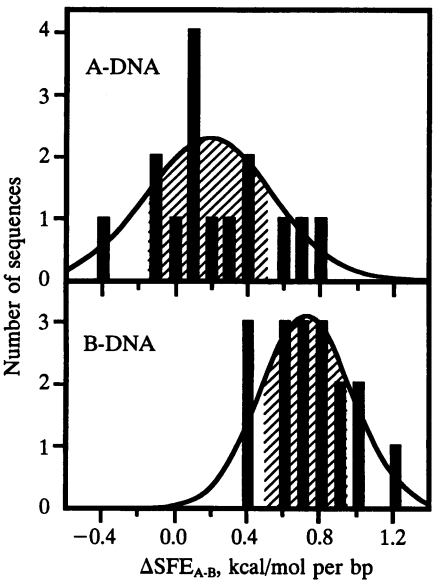

FIG. 1. Distributions of ΔSFE_{A-B} for A-DNA (*Upper*) and B-DNA (*Lower*) sequences. Values for ΔSFE_{A-B} (see Table 1) were rounded to the nearest 0.1 kcal/mol per bp. The curves represent the standard Gaussian distributions calculated from the means and SD of the ΔSFE_{A-B} for each population (Table 1). The shaded area represents 1 SD from the mean of each population.

overlap at 1 SD from the mean of each population). Thus, an APE ≤ -0.02 kcal/mol per bp should favor A-DNA, while an APE $\geq +0.02$ favors B-DNA.

Of the 17 B-DNA sequences, 14 have APEs > 0 (predicted to form B-DNA), while 3 have APEs < 0 (predicted to be A-DNA). Of the 17 A-DNA sequences, 12 have negative APEs, while 5 have positive APEs. The latter five sequences include the three decamers and two of the three dodecamers. The two dodecanucleotides were crystallized with cobalt hexaamine, which stabilizes A-DNA in solution (25), and, therefore, were induced to adopt the A-DNA conformation. As such, these sequences were excluded from the remainder of the study. No other sequence in the data set had been crystallized with cobalt hexaamine. The APEs therefore properly assigned 12 of 15 (80%) of the sequences crystallized as A-DNA.

Triplet Code To Predict A-DNA Formation. To be useful as a predictive tool, the APEs were redefined at the base-pair level. The solvation of a base pair in A- or B-DNA must be considered in the context of all possible neighboring 5′ and 3′ base pairs—that is, as a trinucleotide. A sequence is a linear combination of trinucleotides, and its hydration free energy is the average hydration across these triplets. For example, double-stranded d(GGGGCCCC) is composed of four d[(GGG)·(CCC)] and two d[(GGC)·(GCC)] triplets, and its APE (-0.35 kcal/mol per bp) is simply the weighted average of the APEs of these trinucleotides. Using the method of singular-value decomposition, we can thus determine the contribution of each trinucleotide to the average APEs of the sequences in our data set.

There are potentially 32 unique trinucleotide combinations in duplex DNA. The APEs of 25 unique trinucleotides were determined from our data set to derive a "triplet code" for A-DNA stability (Table 2). This was limited by the size of our data set. In this triplet code, negative APEs indicate that the central nucleotide favors A-DNA, while positive values favor B-DNA. Of the 25 unique triplets in Table 2, 7 are strongly A-DNA-forming (APE ≤ -0.5 kcal/mol per bp), 8 are strongly B-DNA-forming (APE ≥ 0.5 kcal/mol per bp), and 10 do not appear to strongly favor either form. Not surprisingly, d[(CCC)·(GGG)] is strongly A-DNA-forming, while triplets

Table 2. APEs of the A-DNA triplet code

	N_i				
N_{i-1}	C	G	A	T	N_{i+1}
C	−0.59 (13)	0.71 (11)	−1.76 (2)	−1.97 (1)	C
	−0.29 (11)	0.29 (11)	ND	ND	G
	−0.04 (2)	−0.33 (5)	−2.12 (2)	0.15 (3)	A
	ND	2.49 (7)	−0.10 (4)	ND	T
G	−0.49 (8)	0.69 (1)	0.69 (1)	0.69 (1)	C
	0.71 (11)	−0.59 (13)	−1.97 (1)	−1.76 (2)	G
	−0.74 (3)	−0.48 (1)	1.56 (2)	−1.74 (9)	A
	ND	1.52 (2)	0.41 (6)	ND	T
A	1.52 (2)	ND	ND	0.41 (6)	C
	2.49 (7)	ND	ND	−0.10 (5)	G
	ND	0.11 (3)	2.06 (3)	0.57 (5)	A
	ND	ND	0.58 (6)	0.58 (5)	T
T	−0.48 (1)	−0.74 (3)	−1.74 (8)	1.56 (2)	C
	−0.33 (5)	0.04 (2)	0.15 (3)	−2.12 (1)	G
	−1.06 (1)	−1.06 (1)	0.10 (2)	0.10 (2)	A
	0.11 (3)	ND	0.57 (5)	2.06 (2)	T

The A-DNA triplet code of APEs (in kcal/mol per bp) for trinucleotides consisting of a central base pair (N_i) and the 5′ flanking (N_{i-1}) and 3′ flanking (N_{i+1}) base pairs. The number of times each unique triplet is represented in the data set is shown in parentheses. An APE ≤ -0.02 kcal/mol per bp favors A-DNA, while an APE ≥ 0.02 kcal/mol per bp favors B-DNA. Triplets labeled not determined (ND) were not represented in the current data set.

that contain only d(T·A) base pairs are all very strong B-DNA-formers. All strong A-DNA triplets have at least one d(C·G) base pair, which fits the general rule that d(C·G) favors A-DNA. There were some notable exceptions, however. The alternating triplet d[(CGC)·(GCG)] is a strong B-DNA former. Also, while d[(GCC)·(GGC)] is strongly A-DNA-forming, d[(CCG)·(CGG)] is B-DNA-forming. The triplet d[(GTA)·(TAC)] is predicted to be a strong A-DNA former (APE = -1.74 kcal/mol per bp), even though it is centered around a d(T+A) base pair and is >50% d(T+A) in composition. We should note that additional crystal structures will help to improve the reliability of the APEs and to fill the gaps remaining in Table 2.

APE Predictions for A- and B-DNA in Crystals. How well does this triplet code predict the conformations of the sequences in the data set? Of the 26 sequences used to derive the APEs, 24 were predicted correctly as either A- or B-DNA and 2 were predicted incorrectly (Table 3). For 6 sequences that were not used to derive the triplet code, the conformations of 5 were predicted correctly and unambiguously, while 1 was incorrectly predicted (Table 4). Thus, for the 32 sequences that have been crystallized to date and for which we can derive APE values from Table 2, the APEs predict >90% (29 of 32) correctly and unambiguously and 9% (3 of 32) incorrectly. We suspect that the conformations of these three latter cases are strongly influenced by the crystal lattice.

Conformations of Oligonucleotides in Solution. To determine whether the APE triplet code is useful for predicting A-DNA in solution, we monitored the CD spectra of four dodecanucleotides titrated with TFE, which induces A-DNA in solution (28). Two of the sequences are predicted by the APEs to strongly favor A-DNA and two to be stable as B-DNA (Table 5).

The CD spectrum of A-DNA is characterized by, among other things, a strong positive band at 270 nm, a weaker negative band at 240 nm, and a strong negative band at about 210 nm (29). B-DNA spectra are distinguishable from A-DNA spectra in that the 270-nm band is less intense and the 240-nm band is more intensely negative (30). An increase in the intensity of the bands at 270 nm and 210 nm is indicative of a B- to A-DNA transition.

Table 3. Conformations of A- and B-DNA sequences as predicted from the APEs in Table 2

Sequence	APE	ΔSFE_{A-B}	Res.	Predicted conformation
A-DNA				
d(GGGGCCCC)	−0.56	−0.35	−0.21	A
d(GGGATCCC)	−0.22	−0.22	0.00	A
d(GCCCGGGC)	−0.26	−0.64	0.38	A
d(CCCCGGGG)	−0.30	−0.61	0.31	A
d(GTACGTAC)	−0.33	−0.47	0.14	A
d(CTCTAGAG)	−0.56	−0.56	0.00	A
d(GGGCGCCC)	−0.12	−0.38	0.26	A
d(GGGTACCC)	−0.27	−0.13	−0.14	A
d(GTCTAGAC)	−0.48	−0.48	0.00	A
d(GTGCGCAC)	−0.60	−0.60	0.00	A
d(ACCGGCCGGT)	0.40	0.30	0.10	B
d(GCGGGCCCGC)	−0.02	0.19	−0.21	A
d(CCCGGCCGGG)	−0.13	0.11	−0.24	A
d(CCCCCGCGGGGG)	−0.15	−0.13	−0.02	A
B-DNA				
d(CGATCGATCG)	0.04	0.03	0.01	B
d(CGATTAATCG)	0.19	0.20	−0.01	B
d(CGATATATCG)	0.31	0.32	−0.01	B
d(CCGGCGCCGG)	0.20	0.18	0.02	B
d(CATGGCCATG)	−0.16	−0.15	−0.01	A
d(CCATTAATGG)	0.16	0.15	0.01	B
d(CGCATATATGCG)	0.47	0.46	0.01	B
d(CGCAAAAATGCG)	0.45	0.45	0.00	B
d(CGTGAATTCACG)	0.36	0.36	0.00	B
d(CGCAAATTTGCG)	0.10	0.10	0.00	B
d(CGTAGATCTACG)	0.29	0.29	0.00	B
d(CGCGAATTCGCG)	0.65	0.65	0.00	B

The comparable ΔSFE_{A-B} values (with 0.5 kcal/mol per bp subtracted from the values in Table 1) and the residual difference (Res.) between the average APE (kcal/mol per bp) and ΔSFE_{A-B} (kcal/mol per bp) are listed.

A typical TFE-induced transition from B-DNA to A-DNA is observed for double-stranded d(GGCGGCGGCGGC) (Fig. 2*A*). This sequence (APE = 0.10 kcal/mol per bp) shows a typical B-DNA CD spectrum at low TFE (<60%). Upon titration to 71% TFE, the spectrum shifts to that of A-DNA, as evidenced by the increase in intensity at 210 nm and 270 nm. The DNA precipitates at even higher TFE concentrations.

The double-stranded sequence d(GCGCGCGCGCGC) (APE = 0.71 kcal/mol per bp) also shows a characteristic B-DNA spectrum at TFE concentrations < 68%. In 75% TFE the CD spectrum is characteristic of Z-DNA, with positive bands at 220 nm and 270 nm and a strong negative band at ≈190 nm (Fig. 2*B*) (31). The previously calculated ΔSFE of 0.15 kcal/mol per bp for Z-DNA versus B-DNA (14) suggests that the order of conformation stability for alternating d(G+C)-rich sequences is B > Z > A. Therefore, the sequence

Table 4. Conformations predicted and observed in A-DNA crystals

Sequence	APE	Conformation	
		Crystal	Predicted
d(GGATGGGAG)·d(CTCCCATCC)	−0.45	A	A
d(GGCCGGCC)	−0.23	A	A
d(GGTATACC)	0.12	A	B
d(GGGCGCCC)	−0.12	A	A
d(ATGCGCAT)	−0.04	A	A
d(GCCGGC)	−0.10	A	A

The conformations of six sequences that were not included in the data set for deriving the APEs (kcal/mol per bp) are predicted by using the values in Table 2.

Table 5. Conformations of dodecanucleotides in aqueous solution (0% TFE) and in high concentrations of TFE (concentrations shown in parentheses)

Sequence	APE	DNA conformation		
		Predicted	0% TFE	High TFE
d(GGCGGCGGCGGC)	0.10	B	B	A (71%)
d(GCGCGCGCGCGC)	0.71	B	B	Z (75%)
d(CCCCCGCGGGGG)	−0.15	A	A-like	A (68%)
d(CCCCGTACGGGG)	−0.03	A	A-like	A (75%)

Conformations were determined by CD spectroscopy (Fig. 2) and predicted from APE values (kcal/mol per bp) in Table 2.

would be expected to form Z-DNA rather than A-DNA under dehydrating conditions.

The sequence d(CCCCCGCGGGGG) (APE = −0.15 kcal/mol per bp) has been crystallized as A-DNA (32). The CD spectrum of this sequence in 0% TFE shows an intense positive band at 260 nm (Fig. 2*C*), which is more characteristic of A-DNA than of B-DNA. The addition of TFE does not dramatically change the overall shape of the spectrum, which clearly is that of A-DNA at 68% TFE. Continued titration of this sequence to 83% TFE resulted in a transition to another unidentifiable conformation. Is this sequence predominantly A-DNA even in aqueous solution? This sequence is a self-complementary oligonucleotide analog of poly(dG)·poly(dC). Poly(dG)·poly(dC) shows only an A-DNA fiber diffraction pattern (33). More recent CD and NMR studies have established that it is A-DNA in aqueous solution (34), and this has been confirmed with chiral probes (35). In addition, linear dichroism studies in neutral aqueous buffer show that the bases of poly(dG)·poly(dC) are more dramatically inclined than observed for B-DNA, and this is characteristic of A-DNA (36). The CD of double-stranded d(CCCCCGCGGGGG) in buffer is similar to that of poly(dG)·poly(dC) in buffer (37). Since the polymer is A-DNA in buffer, we conclude that our sequence is in the A-DNA form as well.

The double-stranded sequence d(CCCCGTACGGGG), a somewhat weaker A-DNA-forming sequence (APE = −0.03 kcal/mol per bp), has a CD spectrum (Fig. 2*D*) in 0% TFE that is nearly identical to that of d(CCCCCGCGGGGG) in 0% TFE. The titration to A-DNA is analogous to that of d(C-

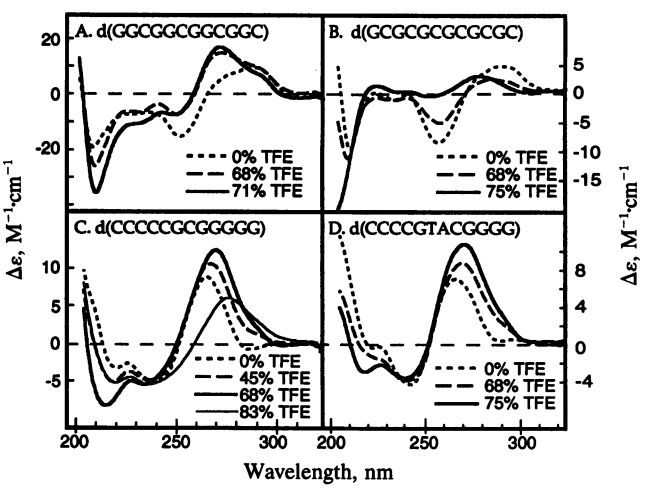

FIG. 2. CD spectra of DNA dodecanucleotides titrated with TFE. All spectra were recorded in 10 mM Tris, pH 8.1/0.02 mM EDTA/4 mM NaCl containing TFE at concentrations labeled with each spectrum. DNA concentrations were adjusted to an absorbance of 0.4 at 260 nm. Spectra are shown for double-stranded sequences d(GGCGGCGGCGGC) (*A*), d(GCGCGCGCGCGC) (*B*), d(CCCCCGCGGGGG) (*C*), and d(CCCCGTACGGGG) (*D*). Δε, change of molar extinction coefficient.

CCCCGCGGGGG). Thus, even though the APE for this sequence is only slightly negative, its behavior in solution is very similar to that of the standard A-DNA-forming sequence d(CCCCCGCGGGGG). These results show that oligonucleotides can form A-DNA in the absence of organic precipitants, including 2-methyl-2,4-dimethylpentanediol (23).

DISCUSSION

In this study, we show that sequences crystallized as A-DNA can be distinguished from those that crystallize as B-DNA by comparing the free energy required to hydrate the exposed DNA surfaces (as measured by APEs). From this, we derived an A-DNA triplet code that is better than 90% accurate in predicting sequences that crystallize as A- and B-DNA. Finally, we observed that sequences with positive APEs have typical B-DNA CD spectra in aqueous solution, while sequences with negative APEs showed characteristic A-DNA spectra, even in the absence of TFE. Therefore, the APEs provide an accurate method to predict the crystal and solution conformations of short DNA sequences.

Can the APEs be useful for predicting A-DNA formation in a cell? Recently, A-DNA has been suggested to be involved in promoter recognition by *E. coli* RNA polymerase (38), in proper phasing of CRP-binding sites (39), and in binding small acid-soluble proteins in dormant spores of *Bacillus* (40). A broader role may be found for A-DNA, but this form is difficult to detect within longer stretches of B-DNA (9). Thus, predictions based on thermodynamic rules may be highly useful, as we have seen with Z-DNA (41). To apply the APE values derived here to predicting A-DNA in genomic sequences, we must include the free energy for forming junctions between A- and B-DNA (1.4 to 2 kcal/mol) (9).

We had previously shown that SFE calculations can account for the sequence-dependent stability of left-handed Z-DNA. Here, this same method is shown to accurately predict the relative stabilities of the two right-handed forms of DNA. Thus, we can now begin to consider a more complex three-state equilibrium between A-, B-, and Z-DNA.

We thank Dr. C. Robert for his input and Prof. W. C. Johnson for help in recording and interpreting CD spectra. This work was supported by grants from the American Cancer Society (NP-740C) and the National Science Foundation (MCB 9304467). G.P.S. has been supported by a postdoctoral fellowship from the American Cancer Society (PF-3749).

1. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
2. Watson, J. D. & Crick, F. H. C. (1953) *Nature (London)* **171,** 737.
3. Franklin, R. E. & Gosling, R. G. (1953) *Nature (London)* **172,** 156–157.
4. Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, ed. Cantor, C. R. (Springer, New York), pp. 251–282.
5. Franklin, R. E. & Gosling, R. G. (1953) *Nature (London)* **171,** 740–741.
6. Rich, A., Norheim, A. & Wang, A. H. J. (1984) *Annu. Rev. Biochem.* **16,** 791–846.
7. Jovin, T. M., Soumpasis, D. M. & McIntosh, L. P. (1987) *Annu. Rev. Phys. Chem.* **38,** 521–560.
8. Peticolas, W. L., Wang, Y. & Thomas, G. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2579–2583.
9. Ivanov, V. I. & Krylov, D. Y. (1992) *Methods Enzymol.* **211,** 111–127.
10. Frederick, C. A., Quigley, G. J., Teng, M. K., Coll, M., Van Der Marel, G. A., Van Boom, J. H., Rich, A. & Wang, A. H. J. (1989) *Eur. J. Biochem.* **181,** 295–307.
11. Heinemann, U. & Alings, C. (1989) *J. Mol. Biol.* **210,** 369–381.
12. Dickerson, R. E. (1992) *Methods Enzymol.* **211,** 67–111.
13. Kagawa, T. F., Stoddard, D., Zhou, G. & Ho, P. S. (1989) *Biochemistry* **28,** 6642–6651.
14. Kagawa, T. F., Howell, M. L., Tseng, K. & Ho, P. S. (1993) *Nucleic Acids Res.* **21,** 5978–5986.
15. Dickerson, R. E. (1990) in *Structures and Methods: Vol. 3, DNA and RNA*, eds. Sarma, R. H. & Sarma, M. H. (Adenine Press, Albany, NY), pp. 1–37.
16. Schneider, B. D., Cohen, D. & Berman, H. M. (1992) *Biopolymers* **32,** 725–750.
17. Alden, C. J. & Kim, S. H. (1979) *J. Mol. Biol.* **132,** 411–434.
18. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S., Srinivasan, A. R. & Schneider, B. (1992) *Biophys. J.* **63,** 751–759.
19. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
20. Arnott, S., Smith, P. J. C. & Chandrasekaran, D. L. (1976) in *Handbook of Biochemistry and Molecular Biology*, ed. Fasman, G. D. (CRC, Cleveland), pp. 411–422.
21. Connolly, M. L. (1983) *Science* **221,** 709–713.
22. Mooers, B. H., Schroth, G. P., Baxter, W. W. & Ho, P. S. (1995) *J. Mol. Biol.* **249,** 772–784.
23. Timsit, Y. & Moras, D. (1992) *Methods Enzymol.* **221,** 409–449.
24. Ho, P. S., Kagawa, T. F., Tseng, K., Schroth, G. P. & Zhou, G. (1991) *Science* **254,** 1003–1006.
25. Xu, Q., Shoemaker, R. K. & Braunlin, W. H. (1993) *Biophys. J.* **65,** 1039–1049.
26. Dickerson, R. E., Goodsell, D. S. & Neidle, S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3579–3583.
27. Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F. & Kennard, O. (1986) *J. Mol. Biol.* **192,** 891–905.
28. Sprecher, C. A., Baase, W. A. & Johnson, W. C. J. (1979) *Biopolymers* **18,** 1009–1019.
29. Ivanov, V. I., Minchenkova, L. E., Schyolkina, A. K. & Poletayev, A. I. (1973) *Biopolymers* **12,** 89–110.
30. Brahms, J. & Mommaerts, W. F. H. M. (1964) *J. Mol. Biol.* **10,** 73–88.
31. Riazance, J. H., Johnson, W. C. J., McIntosh, L. P. & Jovin, T. M. (1987) *Nucleic Acids Res.* **15,** 7627–7635.
32. Verdaguer, N., Aymami, J., Fernandez-Forner, D., Fita, I., Coll, M., Huynh-Dinh, T., Igolen, J. & Subirana, J. A. (1991) *J. Mol. Biol.* **221,** 623–635.
33. Arnott, S., Chandrasekaran, R. & Selsing, E. (1975) in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions*, eds. Sundaralingam, M. & Rao, S. T. (University Park Press, Baltimore), pp. 577–596.
34. Sarma, M. H., Gupta, G. & Sarma, R. H. (1986) *Biochemistry* **25,** 3659–3665.
35. Mei, H. Y. & Barton, J. K. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1339–1343.
36. Kang, H. & Johnson, W. C. J. (1994) *Biochemistry* **33,** 8330–8338.
37. Gray, D. M. & Bollum, F. J. (1974) *Biopolymers* **13,** 2087–2102.
38. Warne, S. E. & DeHaseth, D. L. (1993) *Biochemistry* **32,** 6134–6140.
39. Barber, A. M., Zhurkin, V. B. & Adhya, S. (1993) *Gene* **130,** 1–8.
40. Setlow, P. (1992) *Mol. Microbiol.* **6,** 563–567.
41. Schroth, G. P., Chou, P. J. & Ho, P. S. (1992) *J. Biol. Chem.* **267,** 11846–11855.