5978–5986 Nucleic Acids Research, 1993, Vol. 21, No. 25

Effects of base substituents on the hydration of B- and Z-DNA: correlations to the B- to Z-DNA transition

Todd F.Kagawa, Meredith L.Howell, Kaihan Tseng and P.Shing Ho*
Department of Biochemistry and Biophysics, ALS 2011, Oregon State University, Corvallis, OR 97331, USA

Received August 19, 1993; Revised and Accepted November 5, 1993

ABSTRACT

We present a study of how substituent groups of naturally occurring and modified nucleotide bases affect the degree of hydration of right-handed B-DNA and left-handed Z-DNA. A comparison of poly(dG-dC) and poly(dG-dm^{5}C) titrations with the lipotropic salts of the Hofmeister series infers that the methyl stabilization of cytosines as Z-DNA is primarily a hydrophobic effect. The hydration free energies of various alternating pyrimidine-purine sequences in the two DNA conformations were calculated as solvent free energies from solvent accessible surfaces. Our analysis focused on the N2 amino group of purine bases that sits in the minor groove of the double helix. Removing this amino group from guanine to form inosine (I) destabilizes Z-DNA, while adding this group to adenines to form 2-aminoadenine (A') stabilizes Z-DNA. These predictions were tested by comparing the salt concentrations required to crystallize hexanucleotide sequences that incorporate d(CG), d(CI), d(TA) and d(TA') base pairs as Z-DNA. Combining the current results with our previous analysis of major groove substituents, we derived a thermodynamic cycle that relates the systematic addition, deletion, or substitution of each base substituent to the B- to Z-DNA transition free energy.

INTRODUCTION

The sequence-dependent behavior of DNA structure has been suggested to play a role in a number of transcriptional and replicative processes (reviewed in 1). An understanding of how DNA conformations are affected by the various substituent groups of the nucleotide bases, therefore, helps to extend our understanding of the various mechanisms available to control cellular functions. Aside from the canonical right-handed B-form, the left-handed Z-conformation is perhaps the best studied structural form of DNA (reviewed in 2). The structural and thermodynamic differences between B- and Z-DNA are fairly well understood in empirical terms. Studies show that, for the basic pyrimidine-purine dinucleotide (dn) repeat of Z-DNA, substitution of the cytosines with thymines, and guanines with adenines reduces the ability of sequences to adopt the left-handed conformation in essentially an additive manner (reviewed in 3). The d(TA) dinucleotide is therefore less stable as Z-DNA than d(CA)–d(TG), and both are less stable than the prototypical d(CG) Z-DNA dinucleotide. The relative abilities of these dinucleotides to adopt the Z-conformation have been measured as the free energy of transition for the negative supercoiled induced B- to Z-DNA transition ($\Delta G_{T(B-Z)}$). The values for $\Delta G_{T(B-Z)}$ have been determined experimentally to be 0.66 kcal/mol for d(CG) (4), 1.2 kcal/mol for the d(CA)–d(TG) (5), and 2.4 kcal/mol for d(TA) dinucleotides (6). The B- to Z-DNA transition, therefore, provides a unique system for studying the effects of base substituent groups on the thermodynamic differences between two interconvertable conformations of DNA, and allows one to test models and methods that predict the stability of macromolecular structures.

A number of theoretical methods have been employed, with varying success, to study the effects of substituent groups on the stability of the Z-conformation. Molecular mechanics calculations suggested that the Z-DNA stabilizing effect of methylating cytosines at the C5 position could be understood in enthalpic terms (7). The inability of d(TA) dinucleotides to form Z-DNA, however, could not be readily explained by these methods, primarily because solvent interactions were not included in the calculations. Free energy perturbation calculations in which d(CG) dinucleotides were gradually 'mutated' into d(TA) in an aqueous environment showed that mutation of the first base pair was detrimental to the stability of Z-DNA, but that a transformation of the second d(CG) base pair to d(TA) should potentiate formation of Z-DNA (8). These studies properly predicted that d(CA)–d(TG) would be less stable as Z-DNA compared to d(CG) dinucleotides; however, they also predicted that d(TA) dinucleotides would adopt the Z-conformation more readily than d(CA)–d(TG). This discrepancy between the calculated and the experimental results could not be readily explained.

In our own studies on B- and Z-DNA stabilities, we analyse solvent accessible surfaces (SASs) to determine the difference in solvent free energies (SFEs) for sequences as B- and Z-DNA (9). The differences in free energies between the two conformations for d(CG), d(CA)–d(TG) and d(TA) were of the

*To whom correspondence should be addressed

same order of magnitude as those determined experimentally, and followed the trend that d(CG) < d(CG)−d(TG) < d(TA) in terms of hydrophobicity as Z-DNA. The general order of Z-DNA stability for these dinucleotides could thus be explained in terms of the differences in SFE associated with the base substituents in B- *versus* Z-DNA.

Using this same method, we have also shown that the stabilizing effect of methylating cytosines at the C5 position results from the methyl group actually filling in a hydrophobic pocket at the major groove of Z-DNA (10). The increase in hydrophobic surface that would be expected from adding a methyl group to the major groove surface of Z-DNA is thus more than compensated for by the decrease in the exposed hydrophobic surface of the C5 pocket. An analysis of the C5 methyl of thymine, however, led to an entirely different conclusion (9). For d(TA) dinucleotides in Z-DNA, the C5 pocket was found to be overall more hydrophilic than that of d(CG). The C5 methyl group of thymine actually buries both hydrophilic as well as hydrophobic surfaces. We predicted, therefore, that demethylating the thymine base would actually facilitate the formation of Z-DNA. Thus, the effect of methylation at the C5 carbon of pyrimidine bases on the relative stability of Z- *versus* B-DNA is dependent on the specific base being modified.

Crystallographic studies on the self-complementary sequence d(m^{5}CGUAm^{5}CG) supported the prediction that the decreased hydrophobicity of the Z-DNA major groove surface helped to stabilize d(UA) dinucleotides in the left-handed conformation (11). We had previously shown that the ion concentrations in solutions that yield Z-DNA crystals of hexanucleotide duplexes are related to the ability of these sequence to adopt the Z-conformation in solution. This relationship thus allows us to systematically test predictions for how sequence modifications affect the ability of hexanucleotides to form Z-DNA, and, in fact, has been useful in this laboratory to predict how to crystallize a particular sequence as Z-DNA (12). The self-complimentary hexamer sequence d(m^{5}CGUAm^{5}CG) was crystallized as Z-DNA under significantly lower cation concentrations than the analogous d(TA)-containing sequence. The atomic resolution structure of this d(UA)-containing sequence showed that the absence of the methyl group at the C5 position of the uridine base allowed interactions between a well ordered hexaaquomagnesium complex with the O4 oxygens of the uridine base. The analogous position in the isomorphous d(m^{5}CGTAm^{5}CG) crystal was occupied by a very distorted and poorly defined cluster of cation complexes. Thus the methyl groups of the thymine base were shown to disrupt the solvent interactions at the major groove surface, as predicted from the SAS analyses.

The inability of d(TA) dinucleotides to adopt the Z-conformation was, from our analysis, predicted to be only partially related to the destabilizing effect of the thymine methyl group. The other major factor was observed to be the lack of an amino group at the C2 position of the adenine base, thus rendering the minor groove crevice of Z-DNA less hydrophilic than that of d(CG) dinucleotides. In the present study, we first demonstrate experimentally that the methylation of cytosine stabilizes Z- *versus* B-DNA by affecting the hydrophobicity of the two conformations. We then focus on the contribution of other substituent groups, particularly the N2 amino group in the minor groove of the purine bases, on the thermodynamic stability of Z-DNA. Using the results of these studies, we have derived a thermodynamic cycle which describes the stability of Z-DNA as the base substituents of alternating pyrimidine−purine (APP) dinucleotides are systematically substituted to evolve d(m^{5}CG) to d(TA) and back again.

MATERIALS AND METHODS

Salt titrations of poly(dG-dC) and poly(dG-dm^{5}C)

The polynucleotides poly(dG-dC) and poly(dG-dm^{5}C) were obtained from Pharmacia. Titrations were performed by adding polynucleotide to various concentrations of $MgCl_2$, LiCl, NaCl, and KCl solutions to give approximately 1 OD at 260 nm. The solutions were heated to 60°C for 10 minutes to facilitate formation of Z-DNA (13). Spectra were recorded at room temperature on an HP8453 diode array spectrophotometer. The formation of Z-DNA at each salt concentration was determined by monitoring the ratio of absorbance at 294 nm versus 260 nm. B-DNA has an absorbance ratio of 0.15−0.2 for poly(dG-dC) and 0.25−0.4 for poly(dG-dm^{5}C), while Z-DNA has a ratio of 0.35−0.5 for poly(dG-dC) and 0.35−0.7 for poly(dG-dm^{5}C), depending on the salt.

Solvent accessible surface and solvent free energy calculations

The general methods for calculating the surfaces of DNA structures that are exposed to solvent (solvent accessible surfaces, or SASs) and the free energies for solvating these surfaces (solvent free energies, or SFEs) were previously described (9). The method involves first building models for hexanucleotide sequences in either the B- or the Z-conformations. The atomic coordinates of sequences as B-DNA were generated using standard helical parameters for B-DNA. The atomic coordinates of these same sequences as Z-DNA were generated from the crystal structures of previously crystallized sequences. Models of d(TA′)-containing sequences, where A′ is an aminated adenine at the C2 position, were constructed by adding an sp^2 amino group at the C2 carbon of an adenine base using standard distances and geometries. Sequences containing d(CI) base pairs, where I is an inosine base, were constructed by removing this same amino group from the guanine of a d(CG) base pair. The d(UA) and d(UA′)-containing sequences were constructed by demethylating d(TA) and d(TA′) base pairs, respectively. In specific cases where these bases have been crystallized, we have incorporated the conformations from single crystal structures into our models. Thus, for the d(TA′)- and d(UA′)-containing sequences, we can compare the simplest model for these base pairs in Z-DNA, as generated by adding or removing substituents from the naturally occurring bases, to the actual conformations of these base pairs in the crystal structures to assess their effects on Z-DNA stability.

Table I. Atomic solvation parameters (ASP) of hydrophilic, hydrophobic and charged phosphate surfaces in nucleic acids as derived from the partition coefficients and calculated solvent-accessible surfaces of small organic molecules

Group	Surface type	ASP (kcal/mol^{-1}Å^{-2})
Base	Hydrophobic (C)	0.034
	Methyl (C)	0.043
	Hydrophilic (O/N)	−0.068*
Ribose	Hydrophobic (C)	0.043
	Hydrophilic (O)	−0.038
Phosphate	Charged (O/P)	−0.100

*This value differs slightly from that reported by Kagawa *et al.*(9).

The calculation of the solvent free energies (SFEs) of each DNA structure requires first a calculation of the solvent accessible surfaces (SASs) for each sequence in both the B- and the Z-conformatons. The SASs of the internal four base pairs of each hexanucleotide sequence were calculated using the Connally rolling ball method (14). The SFEs of these models were calculated as previously described. In short, each surface type was converted to a free energy for hydration by applying an atomic solvation parameter (ASP) that describes the energy required to transfer that surface type from an organic phase to an aqueous solvent phase (see Table I). The total free energy of hydration (ΔG_H) for each hexamer model was calculated using Equation 1 (15), where SAS_i is the solvent accessible surface for each surface type i, and ASP_i is the atomic solvation parameter for that surface.

$$\Delta G_H = \Sigma(SAS_i \times ASP_i) \quad \text{(Eq. 1)}$$

The two base pairs at either end of the hexamers were not included in the calculations to eliminate possible artifacts due to end effects. The total ΔG_H was then divided by 2 to obtain the average SFE for the dinucleotide.

The differences in the hydration free energy between Z- and B-DNA ($\Delta\Delta G_{H(Z-B)}$) for the previously studied d(m^{5}CG), d(CG), d(CA)–d(TG), and d(TA) dinucleotides are linearly related to the B- to Z-DNA transition free energies ($\Delta G_{T(B-Z)}$) by Equation 2 (9).

$$\Delta\Delta G_{H(Z-B)} = 0.71\Delta G_{T(B-Z)} - 0.464 \quad \text{(Eq. 2)}$$

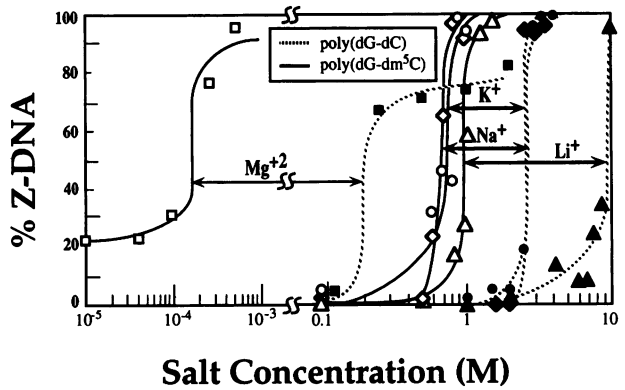

Figure 1. Titrations of poly(dG-dC) and poly(dG-dm^{5}C) with $MgCl_2$, LiCl, NaCl and KCl. The percent of Z-DNA induced by titration of polynucleotides were measured by monitoring the ratio of absorbance at 294 nm versus 260 nm as a function of the salt concentrations. Interpolated curves are drawn to facilitate analysis of the results and are not fit using any theoretical model. Titrations of poly(dG-dC) are shown as closed symbols and are traced with dotted lines, while those of its methylated analogue poly(dG-dm^{5}C) are shown as open symbols are traced using a solid line. The double-headed arrows indicate the difference between the approximate midpoint of titrations with each salt for methylated and unmethylated polynucleotides. The squares represent titrations with $MgCl_2$, diamonds with NaCl, circles with KCl and triangles with LiCl.

Crystallization solutions for Z-DNA hexanucleotides

In general, the most variable component in the crystallization of Z-DNA are the types and concentrations of the cations in the solutions. To normalize all the different types of solutions, we converted the various crystallization conditions for Z-DNA to a measure of the effective cation concentration (the cation strength) in each solution. The cation strengths (*CS*) were estimated as the sum of the concentrations of cation added to the crystallization solutions ([cation]), upon equilibration against the precipitant in the reservoir, corrected for the effective charge of each cation species (Z^2) (i.e. $CS = \Sigma Z^2$[cation], as previously described (9)).

To determine whether the calculated trnsition free energies were related to the actual driving force required for inducing a B- to Z-DNA transition, we compared $\Delta G_{T(B-Z)}$ calculated from Equation 2 to the cation strengths required to crystallize hexanucleotides as Z-DNA. The $\log_{10}$ of the *CS* values (*logCS*) were compared to the $\Delta G_{T(B-Z)}$ of each sequence. With the exception of the d(CICGCG) sequence, the crystallization conditions for the various hexanucleotides used in this study were those published in the original papers (see Table IV). The d(CICGCG) sequence was crystallized from a solution containing 2.0 mM DNA, 33 mM sodium cacodylate buffer at pH 7.0, 0.2

Table II. Solvent accessible surface areas (Å^2) of alternating pyrimidine–purine sequences (Py–Pu Seq) as B and Z-DNA*

Conf. (B/Z)	Py–Pu seq.	Nucleotide base C	N	O	N2	C5methyl	Ribose–phosphate C	O	P	Total
B	TA	43.6	55.6	32.6	NA	44.8	182.8	51.4	132.8	543.8
Z	TA	46.0	50.8	27.0	NA	46.2	188.2	42.0	133.6	533.8
B	TA'	28.8	46.8	29.8	26.0	44.8	183.8	43.8	132.8	536.6
Z	TA'	33.3	49.7	27.4	21.5	46.3	170.9	41.8	133.2	524.1
B	CG	49.4	59.1	31.0	23.6	NA	185.5	47.2	132.6	528.4
Z	CG	56.3	48.7	44.2	19.6	NA	184.0	47.4	132.1	532.4
B	CI	64.8	65.2	38.0	NA	NA	197.7	52.3	132.6	538.6
Z	CI	71.4	48.8	47.2	NA	NA	199.4	46.8	133.0	546.2
B	UA	63.2	57.6	39.2	NA	NA	190.0	51.4	133.6	534.6
Z	UA	68.6	57.0	37.8	NA	NA	194.6	42.0	133.8	534.0
B	UA'[1]	54.0	45.8	34.0	28.2	NA	197.4	47.2	124.2	530.6
Z	UA'[1]	60.0	54.4	33.6	20.6	NA	193.0	40.0	130.8	540.2
B	UA'[2]	45.6	45.8	33.4	29.2	NA	192.8	42.2	133.6	522.4
Z	UA'[2]	56.2	57.0	36.4	18.4	NA	182.8	41.8	133.0	525.6

*Surfaces were calculated for the four internal base pairs of model hexanucleotides of alternating pyrimidine–purine sequence in the table. The total for each surface type are therefore reported for four base pairs (or two dinucleotides). The corresponding values for a single dinucleotide would be half of those in the table.
[1]Calculated from models where the UA' base pairs were constructed by simply removing the methyl from the thymine and adding an amino to the adenine bases of a d(TA) base pair.
[2]Calculated from models constructed using the atomic coordinates of d(UA') dinucleotides of the Z-DNA structure d(m^{5}CGUA'm^{5}CG) (21).

M $MgCl_2$, and 10% 2-methyl-2,4-dimethylpentanediol (MPD) equilibrated against a reservoir of 50% MPD.

RESULTS

Our previous analysis of the hydrated surfaces of DNA duplexes suggested that the effect of base substituent groups on the relative stability of left-handed Z-DNA *versus* right-handed B-DNA is due to differences in the hydrophobicity of the solvent accessible surfaces (SAS) of the DNA structures. In this analysis, these base substituent effects are quantitated by converting the SAS values to solvent free energies (SFE). Here, we present experimental evidence that stabilization of Z-DNA by methylating cytosine bases is associated with differences the hydrophobicity of the DNA conformations. In addition, we have extended the surface and hydration analyses to the minor groove substituents of the DNA bases to study the role of the N2 amino substituent group of purine bases in stabilizing the minor groove crevice of Z-DNA. Our previous studies (9,11) suggested that the inability of d(TA) dinucleotides to adopt the Z-conformation was partially related to the increased hydrophobicity of the major groove surface of Z-DNA associated with the C5 methyl group of the thymine bases. In this study, we focus on the role of the amino group of the purine bases in stabilizing Z-DNA by comparing the hydration free energies of d(TA′) and d(CI) base pairs, where A′ is adenine aminated at the C2 carbon and I is inosine, in their B- and Z-conformations. For these two sequences, we address the question of how adding an amino group to an adenine base pair affects the ability of d(TA), and how removing the amino group from guanine would affect the ability of d(CG) to adopt the Z-conformation. Using the results from our current and previous studies, we have constructed a thermodynamic scheme that describes the contribution of each nucleotide base substituent group to the stability of Z- *5C) versus* B-DNA.

The Hofmeister hydrophobicity cation series and the B- to Z-DNA transition

One of the accepted measures of hydrophobicity is to monitor the influence of certain cations and anions on a process or transition. For cations, it has been shown that the hydrophobic effect follows the Hofmeister series $Mg^{2+} > Li^+ > Na^+ > K^+ > NH_4^+$. According to this series, Mg^{2+} and Li^+ would

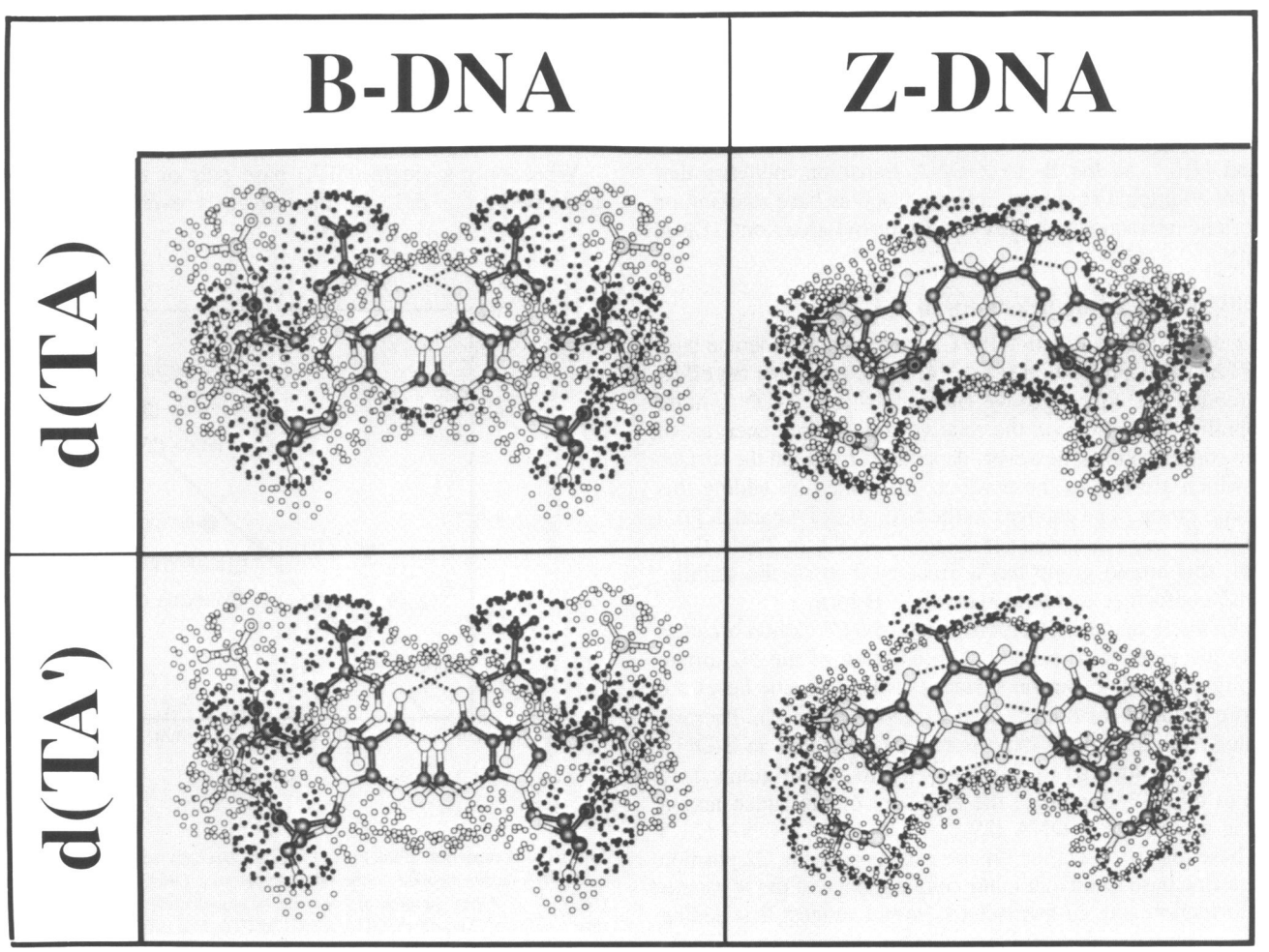

Figure 2. Solvent accessible surfaces of d(TA) and d(TA′) dinucleotides as B- and Z-DNA. The dinucleotides are viewed down the helical axis. The *anti-syn* over *syn-anti* stacking of the base pairs are shown for the dinucleotides in Z-DNA. Hydrogens have been omitted from this figure for clarity, although they were included for the surface calculations. The open dots represent hydrophilic contact points with water, while the filled dots represent hydrophobic contact points. SAS were calculated using a probe radius of 1.45 Å and a dot density of 10 dots/$Å^2$ (5 dots/$Å^2$ are shown for clarity).

show the most pronounced hydrophobic effect, while K^+ and NH_4^+ de-emphasize the effect (16). We can use this series to test our hypothesis that substituent groups of the DNA bases stabilize or destabilize Z-DNA by either decreasing or increasing the hydrophobicity of the Z- and B-conformations.

To determine whether the Z-DNA stabilizing effect of methylating cytosine bases at the C5 position is related to hydrophobicity, we monitored the B- to Z-DNA transition for poly(dG-dC) to poly(dG-dm^{5}C) as induced by cations of the Hofmeister series (Figure 1). If indeed the methyl group is primarily affecting the hyrophobicity of the DNA structures, there should be a dramatic difference in the amount of Mg^{2+} required to induce a B- to Z-DNA transition between the methylated and unmethylated polynucleotides. This difference should decrease as we proceed down the series, with NH_4^+ showing very little if any difference between the two polynucleotides. The titration curves in Figure 1 show the effects of cations on the midpoint of the salt induced transition from B- to Z-DNA. The largest difference was observed for Mg^{2+} (>1000-fold difference for the midpoints of the titrations), followed by Li^+ (> 10-fold difference). Titrations with Na^+ and K^+ showed nearly identical differences (~4-fold differences) in their midpoints. Titrations with NH_4^+, which are not included in Figure 1, were again nearly identical to Na^+ and K^+, with an approximate 5-fold difference for the unmethylated *versus* the methylated sequences. The decreasing differences going from Mg^{2+} to Li^+ to Na^+ suggest that there is a significant hydrophobicity component to the Z-DNA stabilizing effect of cytosine methylation, as we had previously suggested (10). The near identical behaviour of Na^+, K^+ and NH_4^+ on the B- to Z-DNA transition indicates that there are additional stabilizing effects, such as base stacking or electrostatic interactions, of the cytosine methyl group on Z-DNA stability.

Stability of d(TA') dinucleotides in Z-DNA

Adding an N2 amino group to the C2 carbon of the adenine base in a d(TA) base pair generates a d(TA') base pair. The net effect is to render the minor groove of both B- and Z-DNA more hydrophilic. The effect on the relative solvent free energies of the two conformations, however, depends as much on the surface types which are lost as those which are gained by adding this N2 amino group. The changes in the SAS of d(TA) and d(TA') dinucleotides are compared for B- and Z-DNA in Table II. In general, this amino group has a greater effect on the stability of the Z-conformer as compared to the B-form.

In both the B- and the Z-conformations, the C2 carbon becomes entirely inaccessible to solvent with addition of the N2 amino group; the accessible carbon surfaces of the aromatic bases are therefore reduced for both B- and Z-DNA (Table II). There is a greater reduction in the exposure of base carbons in B-DNA (>17 Å^2) as compared to Z-DNA (~12.6 Å^2) in going from d(TA) to d(TA'). Similarly, the exposure of the added amino group is greater for B-DNA (29.2 Å^2) than for Z-DNA (21.5 Å^2). These results reflect the greater exposure of the C2 position of the purine base to solvent in the minor groove of the B- *versus* Z-conformation and, in themselves, would suggest that adding the N2 amino to adenine bases should destabilize Z-DNA. However, if we include in our comparisons the additional surfaces that are buried, we see that neighboring hydrophilic atoms also become less accessible in the d(TA') dinucleotides. For B-DNA, there is a total loss of >19 Å^2 of base oxygen and nitrogen surfaces with the addition of the N2 amino. In the case of Z-DNA, the loss of hydrophilic surface is only 1.3 Å^2. In addition, the added N2 amino of the A' base is stacked directly above the ribose sugar of an adjacent base pair in Z-DNA. This greatly reduces the accessible surfaces, particularly of the C1' and C2' carbons, of the ribose (Figure 2). In total, the surface of B-DNA becomes only slightly more hydrophilic (with a net loss of 17.4 Å^2 of exposed hydrophobic surface and only a 6.8 Å^2 net increase in hydrophilic surface), while for the Z-conformer, the loss of hydrophobic surface is 30 Å^2, with a concomitant gain of 20.2 Å^2 of hydrophilic surface.

When these changes in SASs are translated into SFE values (Table III), we see that there is a 1.35 kcal/mol per dinucleotide (kcal/mol-dn) difference in the hydration free energy of Z *versus* B-DNA ($\Delta\Delta G_{H(Z-B)}$) in the case of the d(TA) dinucleotide, while for the d(TA') dinucleotide, $\Delta\Delta G_{H(Z-B)} \cong 0$. Using the relationship between $\Delta\Delta G_{H(Z-B)}$ and the B- to Z-DNA transition free energy ($\Delta G_{T(Z-B)}$) (9), this translates to 1.01 kcal/mol-dn difference in stability of Z- *versus* B-DNA for the d(TA') dinucleotides, as compared to the 2.4 kcal/mol-dn for d(TA). Thus mutation of d(TA) dinucleotides to d(TA') would have an overall effect of stabilizing the Z-conformation. The extent of this stabilization was greater than expected. The SAS of the d(TA') dinucleotide is overall slightly less hydrophilic compared to the d(CG), but this was true for both the B- and the Z-conformations. We therefore would predict that, compared to the naturally occurring APP dinucleotides of d(CG), d(C-A)−d(TG), and d(TA), the d(TA') dinucleotide would have a propensity to adopt the Z-conformation that is comparable to that of the d(CG).

When only a single d(TA) base pair of the dinucleotide is converted to a d(TA'), the effect is even more dramatic.

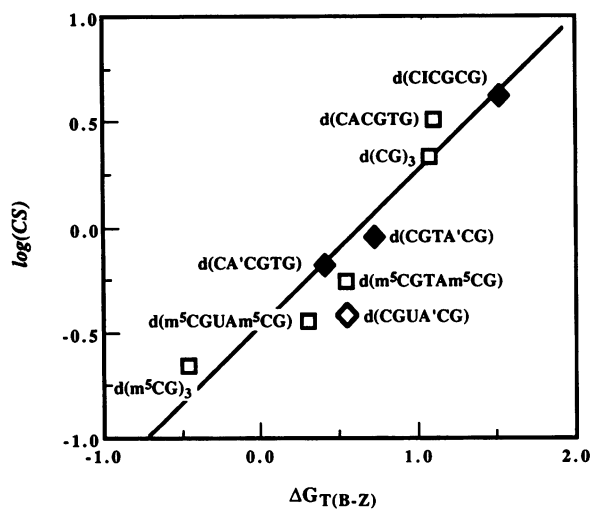

Figure 3. Comparison of the log of the cation strength (*logCS*) for crystallization of Z-DNA hexanucleotides to the calculated B- to Z-DNA transition free energies ($\Delta G_{T(B-Z)}$). Open squares are for the hexanucleotide sequences d(m^{5}CG)$_3$, d(m^{5}CGTAm^{5}CG), d(m^{5}CGUAm^{5}CG), d(CG)$_3$ and d(CACGTG) from earlier studies (see references in Table IV). The diamonds represent sequences for this study. Closed diamonds are for the sequences d(CA'CGTG), d(CGTA'CG) and d(CICGCG). The open diamond is for the sequence d(CGUA'CG). The cation strength (*CS*) is estimated from the crystallization conditions by the relationship $CS = \Sigma Z^2[\text{cation}]$ and where Z is the charge of the cation. The line represents a linear least square fit to the data, giving a relationship: $logCS = 0.74\Delta G_{T(B-Z)} - 0.47$ (R = 0.93).

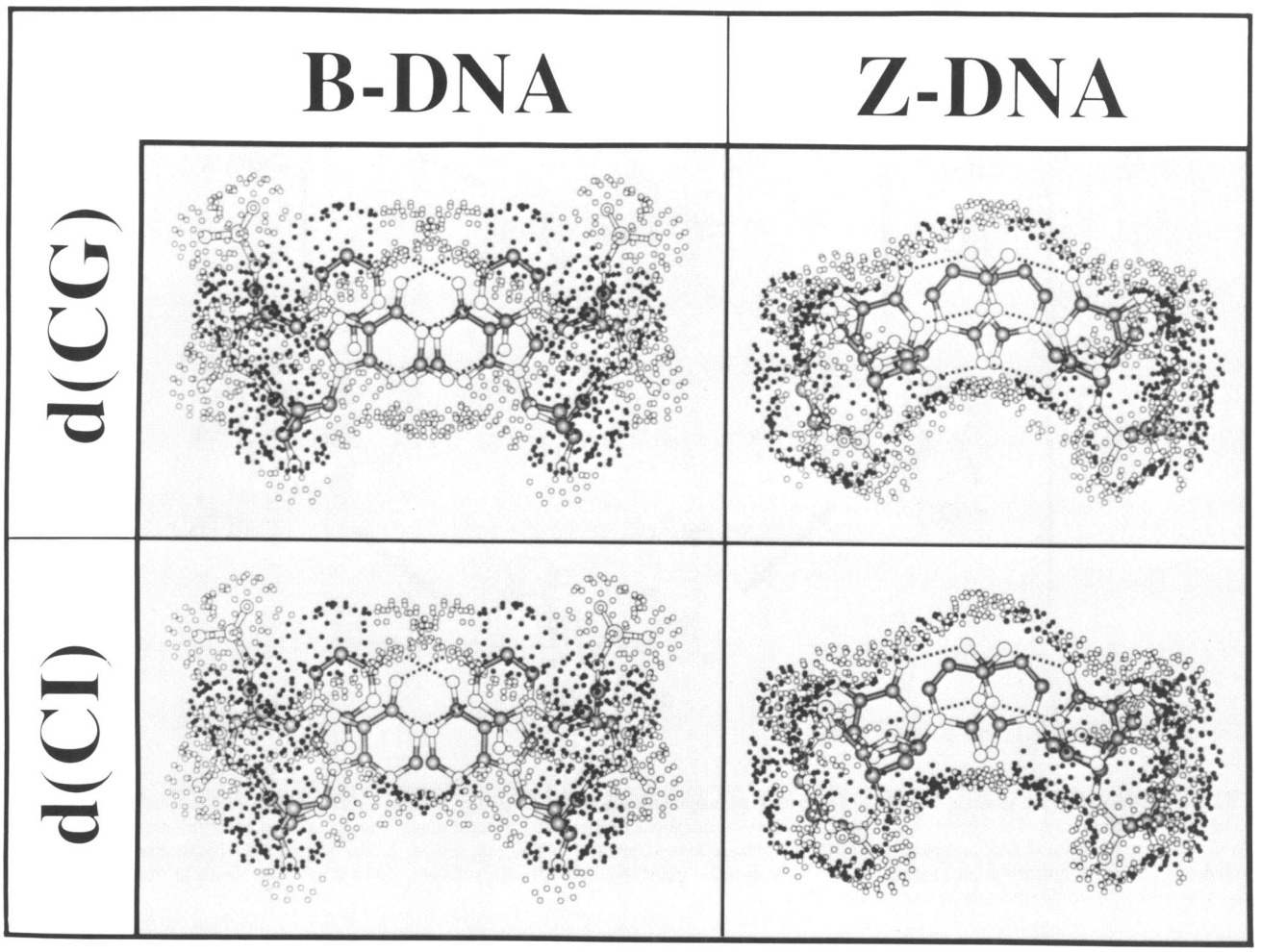

Figure 4. Solvent accessible surfaces of d(CG) and d(CI) dinucleotides as B- and Z-DNA. The dinucleotides are viewed down the helical axis. The *anti-syn* over *syn-anti* stacking of the base pairs are shown for the dinucleotides in Z-DNA. Hydrogens have been omitted from this figure for clarity, although they were included for the surface calculations. The open dots represent hydrophilic contact points with water, while the filled dots represent hydrophobic contact points. SAS were calculated using a probe radius of 1.45 Å and a dot density of 10 dots/Å^2 (5 dots/Å^2 are shown for clarity).

Comparing the areas of d(CPu)−d(PyG) dinucleotides (where Pu is the purine nucleotide G, A or A′ and Py is the pyrimidine base C or T), we calculated $\Delta G_{T(B-Z)}$ = 0.08 kcal/mol-dn for the d(CA′)−d(TG) dinucleotide, which is lower than we would expect for the average between the d(CG) and a d(TA′) dinucleotides. Thus, the effect of each substituent on the stability of a dinucleotide in Z-DNA cannot be considered as simply the sum of the $\Delta G_{T(B-Z)}$ for the two base pairs that form the dinucleotide, as has been suggested (17). A comparison of the predicted $\Delta G_{T(B-Z)}$ values in Table II suggests that the dinucleotide must be considered as the minimum unique repeat unit for Z-DNA in terms of the thermodynamic stability.

We have previously observed a correlation between the cation strengths (*CS*) required to crystallize a series of isomorphous hexamer duplex structures as Z-DNA and $\Delta G_{T(B-Z)}$ calculated for these sequences (12). Our spectroscopic studies on Z-DNA formation under solutions for crystallization showed that the conditions for crystallization of hexanucleotides as Z-DNA were also conditions that stabilize the left-handed conformation in solution (12). Thus, the Z-DNA crystallization conditions for hexamer sequences mirror, and are indicative of the ability of these sequences to adopt the Z-conformation. The incremental increase in the $\Delta G_{T(B-Z)}$ expected for introducing a d(TA′) dinucleotide into a Z-DNA forming sequence was predicted from the SAS analysis to be ~1.01 kcal/mol-dn. This suggests that

Table III. Calculated solvation free energies (SFEs) and the corresponding B- to Z-DNA transition free energies ($\Delta G_{T(B-Z)}$) of dinucleotide sequences

Dinucleotide sequence	SFE (kcal/mol per dinucleotide) $\Delta G_{H(B)}$	$\Delta G_{H(Z)}$	$\Delta\Delta G_{H(Z-B)}$	$\Delta G_{T(B-Z)}$* (kcal/mol per dinucleotide)
d(m^{5}CG)	−10.49	−11.36	−0.87	−0.45
d(CG)	−13.05	−12.76	0.29	1.07
d(CA)−d(TG)	−11.67	−11.34	0.33	1.12
d(TA)	− 9.90	− 8.53	1.35	2.45
d(UA)	−11.50	−10.64	0.86	1.81
d(CI)	−12.01	−10.52	1.50	2.64
d(TA′)	−10.80	−10.56	0.25	1.01
d(CI)−d(CG)	−12.51	−11.69	0.82	1.76
d(CA′)−d(TG)	−11.87	−12.34	−0.47	0.08
d(UA′)	−11.15	−11.87	−0.72	−0.25

*Calculated using the relationship $\Delta G_{T(B-Z)} = 1.30 \times \Delta\Delta G_{H(Z-B)} + 0.69$ kcal/mol per dinucleotide, derived from the relationship analogous to that in Kagawa *et al.* (9).

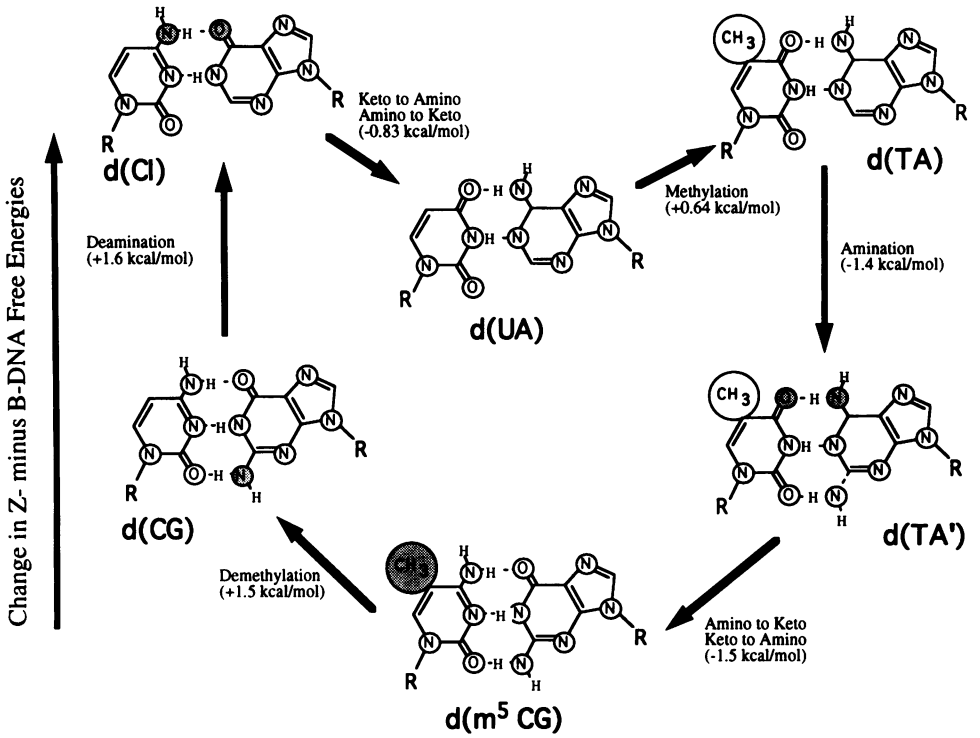

Figure 5. Thermodynamic cycle comparing the substituent effects at the major and minor grooves of the double helix on the stability of Z- versus B-DNA for alternating pyrimidine-purine dinucleotides. The effects of systematic additions or mutations of substituent groups, going from the most stable d(m^{5}CG) dinucleotide to d(CG) to d(CI) to d(UA) to d(TA) to d(TA') and back to d(m^{5}CG), on the relative stability of Z-DNA are shown. Atoms or groups in shaded spheres indicate positions that are added, removed, or mutated at each step in the cycle. The direction of the cycle was arbitrarily chosen, although each step should be treated as an equilibrium. Only one base pair of each dinucleotide is shown for clarity.

replacing a single d(CG) dinucleotide by d(TA') in a hexanucleotide sequence would not significantly affect the cation strength required to crystallize this sequence as Z-DNA. Alternatively, replacing a d(TA) dinucleotide with d(TA') would lower the salt requirement for crystallization. Furthermore, the even lower $\Delta G_{T(B-Z)}$ of 0.08 kcal/mol-dn predicted for d(C-A')–d(TG) would suggest that a sequence that contains this dinucleotide would require less salt to crystallize than an analogous d(CG)-containing sequence.

Coll *et al.* (18) obtained single crystals of the self-complimentary sequence d(CGTA'CG), which was disordered but presumed to be in the Z-conformation, and of d(CA'CGTG), which was indeed in the Z-conformation and isomorphous to other Z-DNA hexanucleotide crystals. Table IV compares the crystallization conditions published for these two hexanucleotides, and also compares the Z-DNA crystallization conditions and the calculated $\Delta G_{T(B-Z)}$ for other well studied hexamer sequences. The cation strength required to crystallize d(CA'CGTG) was indeed similar to that for d(CG)$_3$ and lower than that for d(C-ACGTG), as was predicted from the SFE calculations (Figure 3). The sequence d(CGTA'CG) also behaved as predicted, requiring less salt for crystallization than even d(CG)$_3$.

The structure of the d(CA'CGTG) sequence provides a molecular test for some aspects of this hydration model (18). Of particular interest was the observation that a continuous spine of water molecules was located in the minor groove crevice of the d(TA')-containing hexanucleotide. Similar sets of waters were observed in the crystal structures of d(CGCGCG) and d(m^{5}CGm^{5}CGm^{5}CG), suggesting that these help to stabilize Z-DNA conformation. The d(TA) base pairs of the d(m^{5}CGTAm^{5}CG) and d(CACGTG) sequences, however, disrupt this spine of water. Our results are consistent with these observations, and would have predicted this difference in the interaction of solvent molecules in the minor groove crevice of Z-DNA.

Table IV. Comparison of the calculated B- to Z-DNA transition free energies ($\Delta G_{T(B-Z)}$) to cation strength (*CS*) in solutions that yield crystals of hexanucleotide sequences as Z-DNA

Hexanucleotide sequence	$\Delta G_{T(B-Z)}$* (kcal/mol-dn)	*CS*** (M)	*log(CS)*	Reference
d(m^{5}CG)$_3$	−0.45	0.23	−0.64	(20)
d(m^{5}CGUAm^{5}CG)	0.31	0.36	−0.44	(11)
d(m^{5}CGTAm^{5}CG)	0.55	0.56	−0.25	(23)
d(CA'CGTG)	0.41	0.66	−0.18	(18)
d(CGUA'CG)	0.63	0.38	−0.42	(21)
d(CGTA'CG)***	1.05	0.80	−0.10	(18)
d(CG)$_3$	1.07	2.19	0.34	(24)
d(CACGTG)	1.09	3.24	0.51	(25)
d(CICGCG)	1.52	4.17	0.62	This work

*$\Delta G_{T(B-Z)}$ calculated as the average across the three dinucleotides of the hexamer sequence (units are in kcal/mol per dinucleotide = kcal/mol-dn).
**CS estimated at equilibrium from crystallization solutions by the equation $CS = \Sigma Z^2$[cation] (where Z is the charge of the cation).
***This structure was crystallized in a disordered crystal lattice, but presumed to be in the Z-conformation.

Stability of d(CI) dinucleotides in Z-DNA

The N2 amino groups of guanine bases render the minor grooves of both B- and Z-DNA very hydrophilic (Figure 4). As we would suspect from extrapolation of the results from the analysis of d(TA) and d(TA′), removing this amino group from the minor groove of a d(CG) base pair would greatly reduce the stability of Z-DNA. The SAS calculations show that there is a slight increase of the hydrophobic surfaces of the bases for both B- and Z-DNA due to the exposure of the C2 carbon of the purine base, but, again, it is the now increased exposure of other neighboring hydrophilic groups of the B-DNA bases that helps to destabilize the Z-form (Figure 4 and Table II).

These calculations show that the N2 amino group, when removed, would destabilize the Z-form much more than the B-conformer. The overall loss in stability is observed as a $\Delta\Delta G_{H(Z-B)}$ = 1.5 kcal/mol-dn, or a 2.64 kcal/mol-dn for $\Delta G_{T(B-Z)}$ (Table III). A comparison of the actual Z-DNA crystallization conditions for the sequence d(CICGCG) to those of other Z-DNA hexanucleotides and to the predicted cation strength show that indeed the d(CI) dinucleotide requires a dramatically higher salt for crystallization as compared to other APP hexanucleotide sequences. Thus d(CI) greatly diminishes the stability of the left-handed conformation (Table IV). The degree of destabilization resulting from deamination of guanines is comparable to the stabilizing effect of aminating adenine bases. Consequently, the d(CI) dinucleotide is predicted to have the lowest propensity to form Z-DNA of any APP dinucleotide studied so far.

DISCUSSION

The energetics of DNA folding, as defined by a purely thermodynamic approach, have yet to be fully described in a satisfatory manner. Our work has focused on the contribution of the hydrophobic effect on the ability of various APP sequences to adopt the left-handed structure of Z-DNA. In the present studies, we bring together spectroscopic and crystal growth experiments, along with SFE calculations, to show that indeed there is a strong hydrophobic effect on the stability of Z-DNA and that SFE calculations can be used in a predictive manner to estimate the $\Delta G_{T(B-Z)}$ of APP sequences.

Our results from the titration of poly(dG-dC) and poly(dG-dm^{5}C) with the cations of the lipotropic Hofmeister series show that hydrophobicity contributes to the Z-DNA stabilizing effect of cytosine methylation, as had been previously suggested (10). The results from these experiments, however, also demonstrate that there is an intrinsic ability of the methyl group to stabilize Z-DNA. This additional stabilization may arise from perturbations to the electrostatic properties of the nucleotides as suggested by Soumpasis, *et al.* (19), or by affecting the base stacking, as suggested from the original crystal structure of d(m^{5}CG)$_3$ (20). This is consistent with the relationship that we have derived between the calculated SFEs and $\Delta G_{T(B-Z)}$, which showed that solvent interactions account for approximately 70% of the sequence dependence for Z-DNA formation.

We have extended this SFE analysis to other substituent groups of the DNA bases, and predicted that the N2 amino group is critically important in defining the stability of Z- *versus* B-DNA. The resulting estimates for $\Delta G_{T(B-Z)}$ faithfully predict the crystallization conditions, and thus the driving force required for the salt induced B- to Z-DNA transition of d(CI) and d(TA′)-containing hexanucleotide sequences. The results from the current studies on the minor groove substituents and our previous work on methylation at the major groove surface (9,11) demonstrate that the approach of estimating SFEs from the solvent exposed surfaces of a DNA sequence is useful for predicting the relative abilities of various sequences to adopt the Z-DNA conformation.

We have used the results of these analyses to construct a thermodynamic cycle that relates the effects of systematically adding, removing, or inverting the positions of the various substituent groups on the stability of APP dinucleotides as Z-DNA (Figure 5). Starting with the d(m^{5}CG) dinucleotide, demethylation destabilizes Z-DNA by 1.5 kcal/mol-dn. Removing the amine in the minor groove crevice of the guanine base to form the d(CI) dinucleotide has a more dramatic effect of destabilizing Z-DNA by 1.6 kcal/mol-dn. If we now reverse the positions of the O6 keto oxygen of the guanine base and the N4 amino of the cytosine base to form a d(UA) dinucleotide, Z-DNA is stabilized by −0.83 kcal/mol-dn. Methylation of the uridine bases generates the d(TA) dinucleotide which is now 0.64 kcal/mol-dn less stable as Z-DNA than d(UA). Adding an amino group to the adenine base to give the d(TA′) dinucleotide stabilizes Z-DNA by −1.4 kcal/mol. Finally, we can again reverse the positions of the keto oxygen of the pyrimidine and the amine of the purine at the major groove surface to now complete the thermodynamic cycle, bringing us back to the original d(m^{5}CG) dinucleotide, which is −1.5 kcal/mol-dn more stable than the d(TA′) dinucleotide.

Of the base pairs in this cycle, each has been incorporated in at least one sequence that has been crystallized as Z-DNA. The crystallization conditions can in fact be predicted for these hexanucleotides of APP sequences using the simple rules in this thermodynamic cycle. In addition, the free energies for the negative supercoil induced B- to Z-DNA transition have been determined for three of the dinucleotides in this cycle, and follow the rules d(m^{5}CG) > d(CG) > d(TA) as predicted by this thermodynamic scheme.

One interesting observation from the scheme in Figure 5 is that the positions of the keto- and amino groups at the major groove surface have a greater effect on the stability of Z-DNA than expected. The positions of these two substituents specify whether C5 methylation of the pyrimidines stabilizes or destabilizes Z-DNA. When the keto oxygen is placed on the pyrimidine (U or T), the C5 methyl has a destabilizing effect, while an amino at the C4 position gives the methyl group a stabilizing influence on Z-DNA. The hydration model suggests that the placement of these two substituents defines the pocket at the C5 position of the pyrimidine in Z-DNA as either overall hydrophilic or hydrophobic. In the case of the d(CG) dinucleotide, C5 methylation buries a hydrophobic pocket and thus stabilizes the Z-form. For the d(TA) dinucleotides, demethylation of the thymine exposes a more hydrophilic pocket and thus stabilizes the Z-form.

There are a number of possible pyrimidine-purine dinucleotides that are not represented in this thermodynamic cycle. The most obvious is d(CA)−d(TG). Our previous studies show that this dinucleotide would be as stable in the Z-conformation as would d(UA). The most interesting dinucleotide that is not in this figure is d(UA′). Considering the major groove and minor groove substituent effects that that we observe here, this pair of unusual bases was predicted to be one of the most stable dinucleotides as Z-DNA. In a two state system in which the DNA duplex can adopt only the B- or the Z-conformations, alternating d(UA′)

would be expected to be even more stable in the Z-form than even d(CG), and may be comparable to d(m^5CG) (Table III). This is consistent with the low concentration of salt reported in the crystallization of the sequence d(CGUA'CG) (21) (Table IV).

Obviously, there may be alternative structures of DNA, other than the B- or the Z-forms, that would affect the ability of sequences to adopt the Z-conformation. For example, polymeric d(UA), which should be as stable in the Z-form as a d(C-A)$_n$–d(TG)$_n$ polymer, has not been observed to form Z-DNA in solution. There appears to be an alternative conformation (the yet undefined X-form) which d(UA) prefers under conditions that would normally induce other APP polymers to form Z-DNA (22). Whether there are any alternative structures for d(UA') that could effectively compete with Z-DNA is yet to be determined.

ACKNOWLEDGEMENTS

This work has been supported by grants from the American Cancer Society (NP 740) and the Office of Naval Research (N00014-91-J-1542). P.S.H. was the recipient of an American Cancer Society Junior Faculty Research Award during the period of this research. We thank Drs G.P.Schroth and C.H.Robert, B.Mooers, and other members of the group for their suggestions.

REFERENCES

1. Sinden, R. R. and Wells, R.D. (1992) *Curr. Opin. Biotechnol.*, **3**, 612–622.
2. Jovin, T. M., Soumpasis, D.M. and McIntosh, L.P. (1987) *Ann. Rev. Phys. Chem.*, **38**, 521–560.
3. Rich, A., A. Nordheim and A. H. J. Wang (1984) in *Ann. Rev. Biochem.*, (C. C. Richardson, ed.) **XVI**, Annual Reviews, Inc., pp 791–846.
4. Peck, L. J. and Wang J.C. (1983) *Proc. Natl. Acad. Sci., USA*, **80**, 6206–6210.
5. Vologodskii, A. V. and M. D. Frank-Kamenestskii (1984) *J. Biomol. Struct. Dynam.*, **1**, 1325–1333.
6. Ellison, M. J., Feigon, J., Kelleher, R. J., III, Wang, H. J., Habener, J. F. and Rich, A. (1986) *Biochemistry*, **25**, 3648–3655.
7. Kollman, P., Weiner, P., Quigley, G. and Wang, A. (1982) *Biopolymers*, **21**, 1945–1970.
8. Dang, L. X., D. A. Pearlman and P. A. Kollman (1990) *Proc. Natl. Acad. Sci, USA*, **87**, 4630–4634.
9. Kagawa, T. F., Stoddard, D., Zhou, G. and Ho, P.S. (1989) *Biochemistry*, **28**, 6642–6651.
10. Ho, P. S., Quigley, G.J., Tilton, R. and Rich, A. (1988) *J. Phys. Chem.*, **92**, 939–945.
11. Zhou, G. and P. S. Ho (1990) *Biochemistry*, **29**, 7229–7236.
12. Ho, P. S., Kagawa, T.F., Tseng, K., Schroth, G. P. and Zhou, G. (1991) *Science*, **254**, 1003–1006.
13. Behe, M. and G. Felsenfeld (1981) *Proc. Natl. Acad. Sci., U.S.A.*, **78**, 1619–1623.
14. Connally, M. L. (1983) *Science*, **221**, 709–713.
15. Eisenberg, D. and A. D. MacLachlan (1986) *Nature (London)*, **319**, 199–203.
16. Melander, W. and Horvath, C. (1977) *Arch. Biochem. Biophys.*, **183**, 200–215.
17. Anshelevich, V. V., Vologodskii, A.V. and Frank-Kamenetskii, M.D. (1988) *J. Biomol. Struct. Dynam.*, **6**, 247–259.
18. Coll, M., Wang, A. H. J., van der Marel, G. A., van Boom, J. H. and Rich, A. (1986) *J. Biomol. Struct. Dynam.*, **4**, 157–172.
19. Soumpasis, D. M., Robert-Nicoud, M. and Jovin, T.M. (1986) *FEBS Lett.*, **213**, 341–344.
20. Fujii, S., A. H.-J. Wang, G. van der Marel, J. van Boom and A. Rich (1982) *Nucleic Acids Res.*, **10**, 7879–7892.
21. Schneider, B., Ginell, S.L., Jones, R., Gaffney, B. and Berman, H.M. (1992) *Biochemistry*, **31**, 9622–9628.
22. Vorlickova, M. and Jaroslav, K. (1984) *Int. J. Biol. Macromol.*, **6**, 77–80.
23. Wang, A. H. J., Hakoshima, T., van der Marel, G., van Boom, J. H. and Rich, A. (1984 b) *Cell*, **37**, 321–332.
24. Wang, A. H. J., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich (1979) *Nature*, **282**, 680–686.
25. Coll, M., Fita, I., Lloveras, J., Subirana, J.A., Bardella, F., Huynh-Dinh, T. and Igoleu, J. (1988) *Nucleic Acids Res.*, **16**, 8695–8705.