Stability of DNA Duplexes with Watson−Crick Base Pairs: A Predicted Model

M. Sundaralingam[‡] and P. K. Ponnuswamy*[,§]

Macromolecular Center, Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, and Madurai Kamaraj University, Madurai 625 021, Tamil Nadu, India

Received August 26, 2004; Revised Manuscript Received September 1, 2004

ABSTRACT: The conformational stability (difference between the free energies of the folded and unfolded states, $\Delta G°$) of a DNA duplex is considered as a function of component energy terms, hydrophobic, base stacking, hydrogen bonding, van der Waals, and electrostatic, and a trinucleotide-level helix stiffness parameter measured in terms of its Young's modulus. Hydrophobic and base stacking energy components were determined with the use of the crystal structure data of 30 DNA duplexes judicially selected within a resolution of 1.5 Å, and hydrogen bonding, van der Waals and electrostatic terms were determined through an extensive review of experimental and theoretical studies. The stiffness indices for the trinucleotides were the ones realized by M. M. Gromiha [(2000) *J. Biol. Phys*. *26*, 43−50] using the crystal structure data of 70 DNA duplexes. The unfolded state was treated in the classical way to determine its stability. Thermodynamically determined $\Delta G°$ values for 111 DNA duplexes, with the number of base pairs ranging from 4 to 16, were selected in two sets, and the regression equation formed with one set was used to predict the stabilities of the other set, taking the energy components and the stiffness parameter to be independent variables. The computed energy terms indicate that the base stacking and hydrogen bonding forces are the dominant and the hydrophobic and electrostatic forces the weak partners in imparting stability to the duplexes. This model predicts $\Delta G°$ values for DNA duplexes examined with a level of accuracy similar to that used for predictions made by the widely used nearest-neighbor models. The uniqueness of this model is that it combines the crystal and thermodynamic data for interpretation of conformational stability.

A number of knowledge-based secondary structure prediction algorithms have been developed with the use of crystal structure data of globular proteins (*1*−*7*). These algorithms, apart from delineating pathways from primary structure to secondary structure, have provided valuable information about the differential role of stability factors from nonbonded atoms as well as the influence of surrounding solvent medium on the conformational preferences of protein chains. Because of the inherent limited number of basic building block units in nucleic acids compared to proteins (5 nucleotides vs 20 amino acids), and the associated limit of external input parameters for the building block units, developing successful secondary structure prediction algorithms remains a difficult task for workers in the field, and hence, only a few attempts have so far been reported (*8*−*12*). An encouraging development of the recent past is the rate of reporting of thermodynamic parameters for a variety of oligonucleotides (*13*). A judicial combination of these parameters with the wealth of crystal structure data (*14*) for related oligonucleotides may form a strong database for use in newer theoretical techniques targeting secondary and higher-order structural elements and their associations in nucleic acid systems, especially in RNA molecular families.

Preludes to algorithms for localized secondary structural elements are algorithms for predicting the global stability of oligo/polynucleotides, defined as the free energy difference ($\Delta G°$) between their folded native state and their unfolded structure-less state, which ignores any sequence dependence for the structure. The widely used models for predicting the thermodynamic stability of oligonucleotides from their sequence is the nearest-neighbor model originally developed by Zimm (*15*) and the ones then evolved from it (*16*−*24*). The basic input in these models is the stacking characteristics of neighboring base pairs in the sequence, which are extracted from stability comparisons of oligonucleotides differing in appropriate base sequences. These models incorporate parameters, such as duplex initiation, symmetry correction, end effect, etc., in addition to stacking and hydrogen bonding abilities. Analysis of thermodynamic results (*22*) indicates two distinctive observations. First, duplexes with the same nearest neighbors but different base compositions and therefore different ends consistently exhibit different stabilities (see Table 3 of ref *22*). Second, pairs of groups of sequences with identical nearest neighbors and identical ends do not have identical thermodynamic properties, indicating the presence of non-nearest-neighbor effects (in Table 2 of ref *22*, duplexes AACUAGUU and ACUUAAGU differ in their $\Delta G°$ values by 1 kcal/mol; see also the duplex families which have $\Delta G°$ values differing by more than 0.5 kcal/mol). Alternative models so far proposed to incorporate additional parameters to take care of the two distinctive experimental observations stated above have

* To whom correspondence should be addressed. Phone: 91-452-2459166 or 91-452-2458220. Fax: 91-452-2458449.

‡ The Ohio State University.

§ Madurai Kamaraj University.

10.1021/bi048158+ CCC: $27.50
Published on Web 12/03/2004

improved the predictions to a certain extent. Also, as the thermodynamic experiments are carried out under different experimental conditions, and as there are more than a dozen nearest-neighbor-type theoretical models which widely differ in their predictive abilities (*21*, *23*, *24*), much work must be done to obtain results that could be interpreted in terms of fundamental interactions that are responsible for the molecular stabilities. This also means that it is advisable to look for predictive models that incorporate parameters outside the thermodynamic measurements.

One approach is to develop models for describing the stability that will include, in addition to thermodynamic parameters, knowledge-based inputs derived from the vast crystal structure data for oligonucleotides available today. For example, the hydrophobic factor realized via water accessibility computations enriched the protein models in an exciting way (*25*−*28*). Second, the base stacking parameter set could be determined through atom−atom interactions with the use of the crystal data. Third, the classical secondary force factors, van der Waals, electrostatic, charge−charge, charge−dipole, etc., can be incorporated into the model.

In the early 1990s, P. K. Ponnuswamy made attempts to enumerate the stability factors, hydrogen bonding, base stacking, van der Waals, electrostatic, and hydrophobic interactions, from the knowledge of the crystal structures of DNA−RNA duplexes and tRNA molecules, and translated them into contributions of the free energy to the stability of nucleic acid systems (*29*). Using the thermodynamic experimental (ΔG°) values and computed component free energy terms for a set of duplexes, multiple regression equations were set up to predict duplex stabilities. There was very good agreement between the theoretical predictions and experimental observations for the considered duplexes with 4−16 base pairs and tRNA molecules. These authors found no relationship between ΔG° and N, the number of nucleotides in the molecule. They also pointed out that due to the basic difference between the numbers and kinds of building units and the chain architectures, the unfolded and folded states of nucleic acids substantially differed from those of proteins in their physical ability to associate with water. This has resulted in a dominant role for the hydrophobic factor in protein stability and a smaller role in nucleic acid stability.

These findings were derived by using the then available smaller set of crystal structures of 15 duplexes and two tRNA molecules. The phenomenal growth of thermodynamic (*13*) and crystal structure (*14*) data on DNA−RNA duplexes during the past 10 years has prompted us to improve our model for investigating nucleic acid stability. Using the enriched crystal data, we improved the two important stability factors, namely, hydrophobic and base stacking. We included three additional factors, namely, duplex rigidity (which includes the non-neighbor effect on stability), duplex initiation, and symmetry correction. This improved model is very simple, and it predicts duplex stability satisfactorily. Also, it provides a physical basis for the stability in terms of classical and derived forces that cooperatively help the duplex structures to be strong enough to maintain a stable structure and weak enough to have flexibility to perform a function. In this article, we develop the model and describe the results for DNA duplexes, and in the following article, we deal with the model for RNA duplexes.

METHODS

The main feature of this model is the computation of free energies G_f° and G_u° of the folded native and unfolded structure-less states, respectively, of the duplex molecules. Once these quantities are evaluated in terms of various stability factors, the conformational stability of a duplex molecule could be taken as the difference between these free energies:

$$\Delta G^\circ = G_f^\circ - G_u^\circ \quad (1)$$

Free Energy of the Folded State. The free energy of the folded state (G_f°) is taken to be the sum of various classical/empirical stability factors:

$$G_f^\circ = G_{hy}^\circ + G_{bs}^\circ + G_{sf}^\circ + G_{hb}^\circ + G_{vw}^\circ + G_{el}^\circ \quad (2)$$

The partial energy terms on the right-hand side of the equation are the contributions from hydrophobic (G_{hy}°), base stacking (G_{bs}°), stiffness (G_{sf}°), hydrogen bonding (G_{hb}°), van der Waals (G_{vw}°), and electrostatic (G_{el}°) interactions within the molecules. Below, we briefly describe the methods for computing each of these terms.

G_{hy}°. The differential effect of the solvent on the folded state when compared to the unfolded state is represented by the term "hydrophobic". This part of free energy is computed using the expression

$$G_{hy}^\circ = \Delta\sigma_i[A_i(\text{folded}) - A_i(\text{unfolded})] \quad (3)$$

where $\Delta\sigma_i$ values are the atomic solvation parameters (ASPs) which help to quantify the hydrophobic free energy terms as formulated by Eisenberg and McLachlan (*30*) and A_i(folded) and A_i(unfolded) represent the water accessible surface areas (ASAs) of corresponding atoms in the folded and unfolded states, respectively, of molecule i. We used the same set of $\Delta\sigma$ values for the atom species [nonpolar carbon, C (2.37 cal mol^{-1} Å^{-2}), polar neutral, N/O (−2.81 cal mol^{-1} Å^{-2}), polar charged, O (−17.05 cal mol^{-1} Å^{-2}), and phosphate, P (0.0 cal mol^{-1} Å^{-2})] delineated by us in our earlier work (*29*).

The ASA of any molecule is simply the sum of ASAs of constituent atoms in the molecule. We used ACCESS, developed by Richmond and Richards (*31*), to compute the ASAs by using the atomic coordinates of molecules in the crystal state. While it is straightforward to compute A_i(folded), it is difficult to compute A_i(unfolded), as there is no crystal geometry for the unfolded state. We assumed that the unfolded state is the fully extended state having the maximum water accessible surface area for each residue. As in practice, we considered a set of triresidue segments (Q-X-P) in extended conformations and computed the ASA for X as the average of ASAs of all X residues in the considered segments. In our earlier work (*29*), we considered 226 Q-X-P-type trinucleotide segments found in the nucleic acid data bank and computed their ASAs. The highest possible values for each of the atoms in the common nucleotides [X = adenine (A), thymine (T), guanine (G), cytosine (C), or uracil (U)] were taken to be the accessibilities of the atoms in the respective nucleotide when it is in the extended reference conformation. The sum of the highest ASA values of the constituent atoms was taken to be the highest ASA of the

respective X residue. Since the highest ASAs for the five common nucleotides are thus known, the ASA for any molecule in the extended state can be computed as the sum of the ASAs of the constituent nucleotides in it.

Once $\Delta\sigma_i$, A_i(folded), and A_i(unfolded) are known for a duplex, we can compute $G_{hy}°$ from eq 3.

$G_{bs}°$. The base stacking energy term was computed using the model proposed by Olson (*32*). This model uses the expression

$$G_{bs}° = -166(q_i^2\alpha_j + q_j^2\alpha_i)/Dr_{ij}^4 \quad (4)$$

where q_i and q_j are the partial charges of, α_i and α_j are the polarizabilities of, and r_{ij} values are the distances between atoms *i* and *j* from the stacking bases, respectively. *D* is the dielectric constant, which was taken to be 4*r*, a distance-dependent value (*32*). Since the van der Waals and electrostatic energy terms are computed separately, the charge-induced dipole interaction alone is considered here as the base stacking energy. The partial charges were taken from the work of Cieplak *et al.* (*33*) and the polarizabilities from the work of Halgren (*34*). We selected the same 30 DNA duplex crystal structures (which were used in computing the $G_{hy}°$ term above) for computing the base stacking energy term. $G_{bs}°$ was computed for each sequentially neighboring doublet of bases in the individual and paired strands of the 30 DNA duplex structures. In a double helix, considering two nearby base pairs involving four bases, bases 1 and 2 in one strand and the complementing bases 3 and 4 in the other strand, the stacking interactions in bases 1 and 2 and in bases 3 and 4 pertain to intrastrand base stacking, and stacking interactions in bases 1 and 3 and in bases 2 and 4 pertain to interstrand base stacking. Accordingly, in a set of double helices, we obtain (on averaging) a 5×5 matrix of stacking parameters for the interactions in the two independent (intra) strands, and another 5×5 matrix of parameters for the interactions between the two (inter) strands. The parametric entries of these two matrices were then used to compute the term $G_{bs}°$ for any duplex.

$G_{sf}°$. Gromiha (*35*) has recently computed the Young's modulus for the 32 possible trinucleotide units in DNA molecules using the crystal data of 70 DNA duplexes and assuming the elastic rod model for DNA. This parameter set, called the structure-based stiffness scale, being sequence dependent, is reported to highly correlate with protein−DNA binding data for the molecular assemblies, 434-repressor, and Cro-repressor. Since this scale helps to characterize the behavior of individual trinucleotides we included it in our model to incorporate the sequence effect on the free energy of folding of duplex molecules. In essence, this scale reflects the relative rigidities of the 32 trinucleotides with respect to each other, and hence reflects the relative strengths of operating forces in the trinucleotides. Inclusion of this parameter in our model simply means the normalization of forces that impart stability to the trinucleotide units in the duplexes.

$G_{hb}°$, $G_{vw}°$, and $G_{el}°$. In our earlier studies (*29*), we computed these terms on the basis of an extensive survey of experimental and theoretical results on the strengths of these three forces reported in the literature. We indicated that the $G_{hb}°$ term contributes approximately 1.0 kcal/mol per hydrogen bond and the $G_{vw}°$ and $G_{el}°$ terms contribute 1.0 and 0.2 kcal/mol per nucleotide, respectively. Accordingly, the expressions used for computing these terms were

$$G_{hb}° = 1.0N_{hb} \quad (5)$$

$$G_{vw}° = 1.0(N - 1) \quad (6)$$

$$G_{el}° = 1.0(N - 1) \quad (7)$$

where N_{hb} is the number of hydrogen bonds (two per A·T pair and three per G·C pair) and *N* is the number of nucleotides in the duplex. We also used these expressions in the work presented here.

Free Energy of the Unfolded State ($G_u°$). The ideal unfolded state for a polynucleotide chain is the random coil in which the dihedral angles about all the bonds in the backbone and bases are independent so that the free energy is mainly from chain configurational entropy ($G_{en}°$), the energetics of interactions between different parts of the chain being balanced by interactions from the solvent (*36*). However, this situation is highly impossible to achieve, as there may be certain interactions between parts of the chain which are energetically more favored than the chain solvent interactions. This means that there is a nonentropic ($G_{ne}°$) energy contribution that affects the degree of disorder in the random coil state. Two possible classes of interactions in the unfolded state are the weak hydrogen bonds between the chain and the solvent and steric interactions among parts of the chain, which may occur for fractions of the totally random configurations. We presume the steric (including base stacking) interactions are of negligible magnitude and the hydrogen bond interactions are similar to those between molecules of liquid water. As an approximation, we omit the steric interactions and assume that the hydrogen bonds present in the fractional random configurations similar to those of liquid water could be taken to be half the number found in the folded state (*37*, *38*). Accordingly, the free energy term for the unfolded state is taken to be the sum of the entropic $G_{en}°$ and nonentropic $G_{ne}°$ terms

$$G_u° = G_{en}° + G_{ne}° \quad (8)$$

and the expressions for computing these terms are

$$G_{en}° = -T\Delta S \quad (9)$$

$$G_{ne}° = {}^1/_2 G_{hb}° \quad (10)$$

ΔS could be computed with the number of points of flexibility in the molecule and the number of orientations of equal energy at each such point as suggested by Kauzmann (*36*).

Duplex Formation and Symmetry Considerations. When the duplex is nucleated joining the two independent strands, the process encounters an unfavorable entropic situation, and this is represented by the term $G_{di}°$. This duplex association process also differs depending upon whether the two strands are self-complementary. The energetics of this symmetry feature is represented by the term $G_{sm}°$. Different comparable values have been reported for $G_{di}°$ by various authors (*19*, *20*). We assumed a value of 3.4 kcal/mol which was followed by Sugimoto *et al*. (*19*). As widely used (*19*, *20*), we adopted a value of 0.4 kcal/mol for $G_{sm}°$ if the duplex was non-self-

complementary and 0.0 if it was self-complementary. Both of these parameters were included as unfavorable energy terms in the regression equations of the individual duplexes.

Regression Analysis. Taking the component energy terms and the experimental $\Delta G°$ (eqs 1 and 2) for each of the molecules, one could set up multiple linear regression equations of the type

$$(X_1G_{hy}°)_i + (X_2G_{bs}°)_i + (X_3G_{sf}°)_i + (X_4G_{hb}°)_i + (X_5G_{el}°)_i + (X_6G_{vw}°)_i - G_{us}° = \Delta G° \quad (11)$$

where

$$G_{us}° = G_{di}° + G_{sm}° + G_u° \quad (12)$$

where $i = 1-n$, n being the number of molecules, $\Delta G°$ values being the experimentally known variables, and X being the coefficients. Since the unfolded state is being considered by a single energy term, $G_u°$, the relevant coefficient is taken to be 1; $G_{di}°$ and $G_{sm}°$ are also constant values for a duplex, and their coefficients are also taken to be 1. Treating all the energy terms as independent variables, we could set up a predictive regression equation

$$\Delta G° = A_0 + A_1G_{hy}° + A_2G_{bs}° + A_3G_{sf}° + A_4G_{hb}° + A_5G_{el}° + A_6G_{vw}° - G_{us}° \quad (13)$$

where A values are regression coefficients.

As per our model, each of the four energy terms ($G_{hy}°$, $G_{hb}°$, $G_{el}°$, and $G_{vw}°$) is approximated to be linearly dependent on $N - 1$, where N is the number of nucleotides in a molecule, and treating them as independent variables in the regression technique may not provide unique information about the characteristic behavior of each of these energy components. Hence, we considered the sum of these four energy terms

$$G_{lc}° = G_{hy}° + G_{hb}° + G_{el}° + G_{vw}° \quad (14)$$

as the common single term in the regression equation. Accordingly

$$\Delta G° = A_0 + A_1G_{lc}° + A_2G_{bs}° + A_3G_{sf}° - G_{us}° \quad (15)$$

This equation was considered in the predictive attempts presented here.

RESULTS AND DISCUSSION

Hydrophobic Energy of Duplexes. In our previous work (*29*) when we analyzed the computed $G_{hy}°$ for a set of oligonucleotides for which the crystal structures were then known, we found a linear relationship between $G_{hy}°$ and $N - 1$, where N is the number of nucleotides in a molecule. In the work presented here, we considered 30 DNA oligonucleotides (which had no drug association bulge, mismatch, etc.) with a resolution of ≤1.5 Å. This set of duplexes was selected from the Nucleic Acid Data Bank (*14*) after duplexes with similar lengths and similar experimental $\Delta G°$ values had been eliminated. The duplex structures thus selected (with NAD entries) are as follows: AD003 for AGGGGCCCCT (*39*), AD004 for AGGGGCCCCT (*39*), AD007 for GCGTATACGC (*40*), ADH039 for GTGTACAC (*41*), BD0006 for GGCCAATTGG (*42*), BD0012 for CGCGAATTCGCG (*43*), BD0013 for CGCGAATTCGCG (*43*), BD0014 for CGCGAATTCGCG (*44*), BD0018 for GCGAATTCGCG (*45*), BD0024 for CCGAATGAGG (*46*), BD0030 for CGCGAATTCGCG (*40*), BD0035 for CCGCGCTGG (*47*), BD0037 for GCGAATTCG (*48*), BD0038 for CGCGAATTCGCG (*49*), BD0041 for CGCGAATTCGCG (*50*), BDJ008 for CCAAGATTGG (*51*), BDJ025 for CGATCGATCG (*52*), BDJ031 for CGATTAATCG (*53*), UD0005 for GUTGCAAC (*54*), UDB004 for CGCG (*55*), UDB005 for CGCG (*56*), UDJ032 for AGGCATGCCT (*57*), ZD0003 for CGCGCG (*58*), ZDD023 for CGCG (*59*), ZDF002 for CGCGCG (*60*), ZDF013 for CGCGTG (*61*), ZDF028 for CGCGCG (*62*), ZDF053 for CGCGCG (*63*), ZDF060 for TGCGCA (*64*), and ZDFB05 for CGCGCG (*65*).

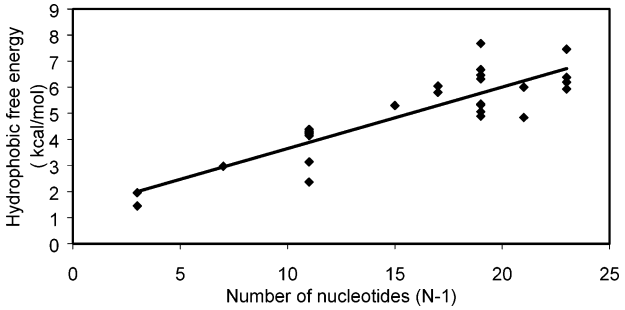

FIGURE 1: Relationship between hydrophobic free energy and the number of nucleotides in DNA duplexes.

Table 1: Intra- and Interstrand Base Stacking Energies (kilocalories per mole)[a]

		A	T	C	G
intrastrand	A	0.82 (26)	0.55 (34)	0.67 (10)	0.48 (10)
	T	0.45 (8)	0.32 (22)	0.45 (22)	0.31 (17)
	C	0.45 (15)	0.25 (11)	0.8 (19)	0.85 (118)
	G	0.71 (28)	0.58 (7)	0.88 (75)	0.69 (27)
interstrand	A	0.07 (3)	0.24 (29)	0.14 (3)	0.17 (4)
	T	0.22 (32)	0.06 (2)	0.02 (2)	0.14 (2)
	C	0.15 (3)	0.04 (3)	0.09 (6)	0.57 (55)
	G	0.18 (4)	0.15 (2)	0.41 (87)	0.15 (13)

[a] See the Methods.

Table 2: Structure-Based DNA Stiffness (Young's Modulus[a]) Parameters for Trinucleotides

trinucleotide	Young's modulus	trinucleotide	Young's modulus	trinucleotide	Young's modulus
AAA/TTT	4.80	AAC/GTT	3.90	AAG/CTT	1.91
AAT/ATT	2.96	ACA/TGT	4.70	ACC/GGT	1.57
ACG/CGT	7.09	ACT/AGT	3.63	AGA/TCT	4.03
AGC/GCT	4.58	AGG/CCT	4.34	ATA/TAT	2.36
ATC/GAT	1.83	ATG/CAT	3.19	CAA/TTG	2.53
CAC/GTG	3.36	CAG/CTG	2.40	CCA/TGG	3.25
CCC/GGG	6.07	CCG/CGG	2.40	CGA/TCG	2.82
CGC/GCG	3.33	CTA/TAG	4.75	CTC/GAG	4.03
GAA/TTC	2.70	GAC/GTC	7.83	GCA/TGC	3.75
GCC/GGC	3.16	GGA/TCC	3.69	GTA/TAC	2.19
TAA/TTA	2.72	TCA/TGA	2.97		

[a] In units of $\times 10^8$ newtons per square meter.

The $G_{hy}°$ values computed using eq 3 were used to determine the relationship between $G_{hy}°$ and $N - 1$. The data plotted in Figure 1 show a good linear relationship

$$G_{hy}° = 1.608 + 0.214(N - 1) \quad (16)$$

with a coefficient of determination (R^2) of 0.74 and a standard

Table 3: Energy Components and Experimental and Predicted Conformational Stabilities of DNA Duplexes[a]

Set 1 (from ref *20*)

	duplex	*N*	ΔG_{hy}°	ΔG_{hb}°	ΔG_{el}°	ΔG_{vw}°	ΔG_{bs}°	ΔG_{sf}°	G_u°	ΔG_e°	ΔG_p°	ΔG_{sl}°
1	CCGG	8	3.1	12	1.4	7	6.6	6.2	30.0	3.5	4.2	3.4
2	CGCG	8	3.1	12	1.4	7	7.1	7.8	30.0	4.0	4.3	4.2
3	GCGC	8	3.1	12	1.4	7	7.2	7.8	30.0	4.4	4.3	4.4
4	CCGCGG	12	4.0	18	2.2	11	11.1	10.5	45.0	8.0	8.0	7.8
5	CGATCG	12	4.0	16	2.2	11	9.2	9.1	44.0	5.3	5.7	5.6
6	CGCGCG	12	4.0	18	2.2	11	11.5	11.7	45.0	8.3	8.1	8.6
7	CGGCCG	12	4.0	18	2.2	11	11.1	10.3	45.0	8.3	8.0	7.8
8	CGTACG	12	4.0	16	2.2	11	9.2	15.1	44.0	5.4	5.8	5.4
9	GACGTC	12	4.0	16	2.2	11	8.9	22.5	44.0	5.6	6.0	5.7
10	GCATGC	12	4.0	16	2.2	11	8.5	12.1	44.0	5.6	5.7	5.8
11	GCCGGC	12	4.0	18	2.2	11	11.1	10.3	45.0	8.5	8.0	8.0
12	GCGAGC	12	4.0	17	2.2	11	9.7	13.2	44.5	7.7	6.9	7.5
13	GCGCGC	12	4.0	18	2.2	11	11.6	11.7	45.0	9.1	8.1	8.8
14	GCTAGC	12	4.0	16	2.2	11	8.3	15.2	44.0	5.3	5.8	5.3
15	GGATCC	12	4.0	16	2.2	11	8.8	10.2	44.0	5.0	5.6	5.0
16	GGCGCC	12	4.0	18	2.2	11	11.1	11.5	45.0	7.9	8.0	8.0
17	GGGACC	12	4.0	17	2.2	11	9.5	15.4	44.5	6.5	6.9	6.4
18	GTGAAC	12	4.0	15	2.2	11	7.6	11.5	43.5	5.1	4.5	4.9
19	CAAAAAG	14	4.4	16	2.6	13	8.2	14.1	50.0	4.8	4.2	4.8
20	CAAAAAAG	16	4.8	18	3.0	15	9.5	17.2	57.0	5.7	5.0	5.8
21	CAAGCTTG	16	4.8	20	3.0	15	9.8	13.8	58.0	7.0	7.0	7.2
22	CATCGATG	16	4.8	20	3.0	15	10.6	13.0	58.0	7.6	7.1	7.0
23	CGATATCG	16	4.8	20	3.0	15	11.6	12.1	58.0	6.9	7.1	6.9
24	CGTCGACG	16	4.8	22	3.0	15	13.3	25.9	59.0	9.8	9.9	9.8
25	GAAGCTTC	16	4.8	20	3.0	15	10.6	14.0	58.0	7.0	7.1	7.3
26	GATCGATCG	18	5.3	22	3.4	17	9.3	13.9	65.0	7.6	7.6	7.2
27	GATGCATC	16	4.8	20	3.0	15	10.6	15.1	58.0	7.3	7.1	7.2
28	GGAATTCC	16	4.8	20	3.0	15	11.5	16.2	58.0	6.8	7.2	7.0
29	GGACGTCC	16	4.8	22	3.0	15	12.9	26.5	59.0	9.0	9.8	9.2
30	GGAGCTCC	16	4.8	22	3.0	15	11.9	19.3	59.0	8.7	9.5	8.8
31	GGTATACC	16	4.8	20	3.0	15	11.2	12.2	58.0	5.5	7.0	6.1
32	GTACGTAC	16	4.8	20	3.0	15	10.3	17.1	58.0	7.1	7.2	6.8
33	GTAGCTAC	16	4.8	20	3.0	15	10.3	18.3	58.0	7.0	7.2	6.4
34	GTTGCAAC	16	4.8	20	3.0	15	10.9	16.9	58.0	7.8	7.2	7.7
35	CAAAAAAAG	18	5.3	20	3.4	17	10.9	20.3	64.0	7.2	5.8	6.8
36	CAAACAAAG	18	5.3	21	3.4	17	11.0	18.9	64.5	7.7	6.8	7.6
37	CAAATAAAG	18	5.3	20	3.4	17	10.6	17.5	64.0	6.5	5.6	6.1
38	CAAAGAAAG	18	5.3	21	3.4	17	10.7	17.8	64.5	7.3	6.7	7.4
39	GCGAATTCGC	20	5.7	26	3.8	19	16.4	19.0	73.0	12.9	11.1	12.2
40	CCATCGCTACC	22	6.1	29	4.2	21	17.0	19.6	80.5	13.3	12.7	12.9
41	GCGAAAAGCG	20	5.7	26	3.8	19	15.9	21.7	73.0	11.9	11.1	12.6
42	CCATTGCTACC	22	6.1	28	4.2	21	15.7	20.2	80.0	12.2	11.5	11.7
43	CTGACAAGTGTC	24	6.5	30	4.6	23	15.2	30.7	87.0	12.6	12.4	12.9
44	CATATGGCCATATG	28	7.4	34	5.4	27	18.7	24.7	101.0	12.7	13.8	13.2
45	TCATGA	12	4.0	14	2.2	11	6.8	15.3	43.0	3.3	3.5	3.4
46	TGATCA	12	4.0	14	2.2	11	6.8	13.5	43.0	2.8	3.5	3.4
47	AAAAAAAA	16	4.8	16	3.0	15	9.8	26.0	56.0	4.5	3.2	3.9
48	TAGATCTA	16	4.8	18	3.0	15	9.0	22.3	57.0	5.1	5.1	4.2
49	TCTATAGA	16	4.8	18	3.0	15	9.0	22.8	57.0	4.3	5.1	4.2
50	ATGAGCTCAT	20	5.7	23	3.8	19	11.8	24.2	71.5	10.0	7.6	9.5
51	TTTTATAATAAA	24	6.5	24	4.6	23	12.2	25.9	84.0	5.5	5.6	5.8
52	CAACTTGATATTATTA	32	8.2	35	6.2	31	19.6	25.9	113.5	12.4	11.8	12.7
53	CCCGGG	12	4.0	18	2.2	11	10.6	17.2	45.0	6.9	8.2	7.0
54	CCCAGGG	14	4.4	20	2.6	13	10.6	21.7	52.0	7.9	8.9	7.8
55	CGCGAATTCGCG	24	6.5	32	4.6	23	20.8	26.0	88.0	20.6	15.0	16.4
56	GTATACCGGTATAC	28	7.4	34	5.4	27	20.2	21.8	101.0	12.3	16.8	17.4
57	CGCATGGGTACGC	26	7.0	35	5.0	25	21.1	31.1	95.5	14.4	13.7	13.0
58	CATATTGGCCAATATG	32	8.2	38	6.2	31	21.4	27.7	115.0	13.2	15.2	15.3
59	GTATAACCGGTTATAC	32	8.2	38	6.2	31	22.9	25.2	115.0	14.3	15.2	15.0
60	CGCGTACGCGTACGCG	32	8.2	44	6.2	31	29.0	35.8	118.0	29.1	22.6	24.1
										rmsd[b]	1.4	1.2
										rmsd[c]	0.7	0.4

Set 2 (from ref *19*)

	duplex	*N*	ΔG_{hy}°	ΔG_{hb}°	ΔG_{el}°	ΔG_{vw}°	ΔG_{bs}°	ΔG_{sf}°	G_u°	ΔG_e°	ΔG_p°	ΔG_{su}°
1	AGCCG	10	3.5	14	1.8	9	7.9	8.8	37.0	5.6	4.9	5.3
2	ACCGCA	12	4.0	16	2.2	11	9.5	10.9	44.0	6.7	5.7	7.0
3	ATGCGC	12	4.0	16	2.2	11	9.5	11.8	44.0	7.3	5.8	6.6
4	CGGTGC	12	4.0	17	2.2	11	9.5	10.9	44.5	7.2	6.8	7.0
5	CGTGCC	12	4.0	17	2.2	11	9.7	13.6	44.5	6.9	6.9	7.0
6	TGCGCA	12	4.0	16	2.2	11	9.1	12.3	44.0	6.9	5.8	7.0

[a] Subscripts: e, experimental; p, predicted by this work; sl, SantaLucia *et al*. (*20*) prediction; su, Sugimoto *et al*. (*19*) prediction. [b] For all 60 duplexes. [c] For 55 duplexes [excluding five duplexes (50, 55, 56, 58, and 60 in the list) which have an rmsd of >2]. All *G* values (kilocalories per mole) are negative. ΔG_p° of set 1 determined by eq 17 and ΔG_p° of set 2 determined by eq 18.

Table 3. (Continued)

	Set 2 (from ref *19*)											
	duplex	N	$\Delta G_{hy}°$	$\Delta G_{hb}°$	$\Delta G_{el}°$	$\Delta G_{vw}°$	$\Delta G_{bs}°$	$\Delta G_{sf}°$	$G_u°$	$\Delta G_e°$	$\Delta G_p°$	$\Delta G_{su}°$
7	AATACCG	14	4.4	17	2.6	13	10.0	10.7	50.5	5.9	5.3	6.0
8	AGCCGTG	14	4.4	19	2.6	13	10.5	17.5	51.5	8.5	7.7	8.5
9	AGCTTCA	14	4.4	17	2.6	13	8.7	13.8	50.5	6.1	5.3	6.3
10	GGACTTA	14	4.4	17	2.6	13	8.9	14.6	50.5	5.6	5.3	5.3
11	AAGCGTAG	16	4.8	20	3.0	15	11.0	18.4	58.0	8.0	7.3	8.3
12	AATCCAGT	16	4.8	19	3.0	15	10.3	15.4	57.5	6.8	6.0	7.0
13	ACATATGT	16	4.8	18	3.0	15	9.4	17.4	57.0	5.7	4.9	5.3
14	ACCTAGTC	16	4.8	20	3.0	15	10.5	21.7	58.0	6.5	7.3	7.1
15	ACGACCTC	16	4.8	21	3.0	15	12.0	20.7	58.5	8.7	8.4	9.0
16	AGAGAGAG	16	4.8	20	3.0	15	9.0	19.0	58.0	7.4	7.1	7.1
17	AGCGTAAG	16	4.8	20	3.0	15	10.9	16.5	58.0	7.8	7.2	8.3
18	AGTCCTGA	16	4.8	20	3.0	15	10.1	18.2	58.0	7.5	7.1	7.9
19	ATGCGCAT	16	4.8	20	3.0	15	11.8	17.0	58.0	9.0	7.3	8.8
20	CACGGCTC	16	4.8	22	3.0	15	12.4	19.5	59.0	10.0	9.5	10.0
21	CCATATGG	16	4.8	20	3.0	15	10.4	15.5	58.0	6.8	7.1	6.5
22	CGCGTATA	16	4.8	20	3.0	15	12.1	16.1	58.0	8.4	7.3	8.7
23	CGCTGTAA	16	4.8	20	3.0	15	11.1	15.8	58.0	7.9	7.2	8.5
24	CTAGTGGA	16	4.8	20	3.0	15	9.7	18.2	58.0	7.5	7.1	7.3
25	CTCACGCG	16	4.8	22	3.0	15	12.6	18.4	59.0	9.8	9.5	10.0
26	CTGAGTCC	16	4.8	21	3.0	15	10.4	20.7	58.5	8.0	8.3	7.9
27	GAATATTC	16	4.8	18	3.0	15	10.0	14.5	57.0	4.3	4.9	4.3
28	GACTAGTC	16	4.8	20	3.0	15	10.0	24.1	58.0	6.6	7.3	6.1
29	GAGTACTC	16	4.8	20	3.0	15	10.0	16.9	58.0	5.8	7.1	6.1
30	GATTAATC	16	4.8	18	3.0	15	10.0	13.6	57.0	4.3	4.9	4.3
31	GCATATGC	16	4.8	20	3.0	15	11.0	16.1	58.0	7.8	7.2	6.9
32	GCCAGTTA	16	4.8	20	3.0	15	10.9	16.2	58.0	8.2	7.2	7.8
33	GGTGCCAA	16	4.8	21	3.0	15	11.7	15.2	58.5	9.0	8.3	9.2
34	GTCGAACA	16	4.8	20	3.0	15	11.3	19.5	58.0	8.3	7.3	8.3
35	GTCTAGAC	16	4.8	20	3.0	15	9.8	24.2	58.0	6.5	7.3	6.1
36	TAGGCCTA	16	4.8	20	3.0	15	10.8	19.6	58.0	8.2	7.3	7.5
37	TATGCATA	16	4.8	18	3.0	15	9.6	15.5	57.0	5.9	4.9	5.5
38	AAAAAAAAA	18	5.3	18	3.4	17	11.2	23.8	63.0	6.2	3.8	6.2
39	ATAACTGGC	18	5.3	22	3.4	17	12.2	17.1	65.0	9.0	7.9	8.7
40	ATCTATCCG	18	5.3	22	3.4	17	12.5	16.8	65.0	8.7	7.9	8.7
41	CGCTGTTAC	18	5.3	23	3.4	17	12.9	17.5	65.5	9.9	9.0	10.0
42	GCCAGTTAA	18	5.3	22	3.4	17	12.2	16.9	65.0	8.8	7.9	9.0
43	AAAAAAAAAA	20	5.7	20	3.8	19	12.6	26.9	70.0	6.7	4.6	7.4
44	CGGCAAGCGC	20	5.7	28	3.8	19	17.1	19.8	74.0	13.3	13.3	15.6
45	TAGGTTATAA	20	5.7	22	3.8	19	12.1	19.0	71.0	7.0	6.4	7.7
46	ACGTATTATGC	22	6.1	26	4.2	21	15.3	23.7	79.0	10.4	9.5	11.2
47	ATTGGATACAAA	24	6.5	27	4.6	23	15.3	23.6	85.5	10.3	9.0	11.4
48	ACATTATTATTACA	28	7.4	30	5.4	27	17.2	25.9	99.0	11.3	9.3	12.0
49	CACAG	10	3.5	13	1.8	9	5.5	13.2	36.5	3.6	3.8	3.0
50	CAACCAACCAAC	24	6.5	30	4.6	23	17.2	24.4	87.0	14.1	12.3	14.0
51	CTTCCTTCCTTC	24	6.5	30	4.6	23	16.7	24.0	87.0	14.1	12.3	13.4
										rmsd	1.0	0.6

error of 0.76. This equation was used to compute $G_{hy}°$ for molecules considered in our predictive scheme.

Base Stacking Energy. For computing this energy term using eq 4, we used the same 30 DNA duplexes (eliminating two duplexes for which $N = 4$) used for computing the $G_{hy}°$ term above. When the computed results were averaged over the two possible 5×5 matrices (as enumerated in the Methods) of base doublets, one set for intrastrand and another for interstrand, we obtained the values listed in Table 1. The entries in parentheses are the frequencies of occurrence of the respective doublet neighbors. Since the frequencies of occurrence of doublets involving the base U are small or zero, we did not consider duplexes with U for our predictive attempts. We note from this table that the stacking character of doublets varies with the order of sequence. For example, the intrastrand stacking energies indicate that G·C, C·G, A·A, and A·C base pairs come in order of stability, while the interstrand stacking energies indicate that C·G, G·C, A·T, and T·A base pairs come in order. Such distinctive behavior is seen in each of the stacking base·base units. In general, the intrastrand stacking is stronger than the interstrand stacking in any pairing units. We used this table of sequence-dependent $G_{bs}°$ values in the predictive cases.

Stiffness Term. Table 2 lists the structure-based stiffness (Young's modulus) parameters for the 32 relevant trinucleotide sequences in DNA duplexes as determined by Gromiha (*35*). From this table, we note that a strong sequence-dependent character is exhibited by the triplets; GAC/CGT, ACG/CGT, and CCC/GGG are the stiffest and ACC/GGT, ATC/GAT, and AAG/CTT the most flexible trinucleotides.

Regression Results. We selected 111 DNA duplexes in two sets. Set 1 includes 60 duplexes taken from Table 2 of the work of SantaLucia *et al*. (*20*), and set 2 includes 51 duplexes taken from Table 2 of Sugimoto *et al*. (*19*) (Table 2 of Sugimoto *et al*. contains 65 duplexes, and of these, 14 are duplicates to duplexes listed in Table 2 of SantaLucia *et al*.; we eliminated these duplicated cases). For all these duplexes, the experimental $\Delta G°$ values were determined at 37 °C in 1 M NaCl. The various energy terms in eq 15 were then computed for both sets of molecules with the use of expressions described in the Methods. The computed values are given in Table 3. Taking the numerical values of $G_{lc}°$,

$G_{bs}°$, $G_{sf}°$, and $G_{us}°$ for each molecule in set 1 to be independent variables on the right-hand side and the experimental $\Delta G°$ of that molecule to be the known variable on the left-hand side of eq 15, we constructed 60 linear equations. Solving these equations by regression technique, we obtained a preliminary regression equation. Using the coefficients of this equation, we back predicted the $\Delta G°$ values for each of the same 60 duplexes. Five duplexes (ATGAGCTCAT, CGCGAATTCGCG, CGCATGGGTACGC, CATATTGGCCAATATG, and CGCGTACGCGTACGCG) exhibited residuals ($\Delta G_{pred}° - \Delta G_{exp}°$) greater than 2 kcal/mol. The linear equations corresponding to these five duplexes were then eliminated from the input. The 55 remaining linear equations were solved to obtain a regression equation. This predictive equation was

$$\Delta G° = -7.265 + 1.570G_{lc}° + 0.080G_{bs}° + 0.035G_{sf}° - G_{us} \quad (17)$$

The coefficient of determination (R^2) and the standard error for this equation were 1.0 and 0.75, respectively. Using the equation described above, we back predicted $\Delta G°$ values for the same 55 DNA duplexes of set 1. The back-predicted results ($\Delta G_p°$ values in set 1, Table 3) agreed with the experimental values with an rmsd of 0.7.

The back-predicted $\Delta G°$ values and the experimental values were found to be related by the linear equation

$$\Delta G_p° = 0.295 + 0.962G_e° \quad (18)$$

which had a coefficient of determination (R^2) of 0.97 and a standard deviation of 0.73. Now the rmsd of the residuals was 0.67.

Predictions about New Duplexes (Set 2). Equation 17 was then used to predict the $\Delta G°$ values for the 51 duplexes of set 2 (Table 3), which did not form part of the linear equations used to delineate this equation.

Sample Prediction. Duplex 1 in set 2 (AGCCG): $G_{hy}°$ (from eq 3) = 3.53, $G_{lc}°$ (=$G_{hy}° + G_{hb}° + G_{el}° + G_{vw}°$ = 3.53 + 14.0 + 1.8 + 9.0) = 28.33, $G_{bs}° = 7.9$, $G_{sf}° = 8.84$, $G_u° = 37.0$, $G_{di}° = 3.4$, and $G_{sm}° = 0.4$. Substituting these numerical values into eq 17, we have

$$\Delta G_p° = -7.265 + 1.57 \times 28.33 + 0.08 \times 7.9 + 0.035 \times 8.84 - (37.0 - 3.4 - 0.4) = 4.9.$$

Table 3 includes $G_p°$ values predicted as a part of this work as well as the values predicted by Sugimoto *et al.* (*19*) for the same molecules.

The computed residuals for the predicted results of this model and the two sets of predicted results of Sugimoto *et al.* (*19*) with reference to the experimental values are given in Table 3. The residuals of this work had an rmsd value of 1.0. The two duplexes, A_9 and A_{10}, had residuals of >2, viz., 2.4 and 2.8. (The experimental $\Delta G°$ values given for A_8 by SantaLucia *et al.* (*20*) and by Sugimoto *et al.* (*19*) differ by 0.6 kcal/mol. Probably, the values of these two oligo-A duplexes need scrutiny. Or else, it could be that crystal stacking conditions may differ due to solution conditions.) When these two duplexes are excluded, the rmsd value was reduced to 0.9. The predictions of Sugimoto *et al.* (*19*) had residuals of 0.6. Figure 2a is a plot of experimental versus predicted $\Delta G°$ values for the duplexes listed in Table 3, and Figure 2b is a plot of the corresponding residuals.

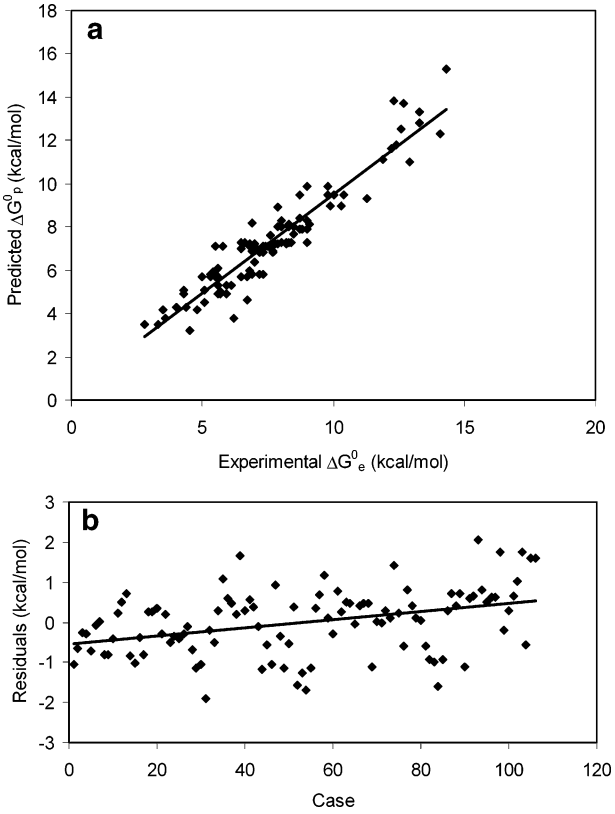

FIGURE 2: (a) Plot of predicted and experimentally observed conformational stabilities of DNA duplexes. (b) Plot of residuals for DNA duplexes.

Figure 3a shows the variation of individual energy terms with *N*, the number of nucleotides in four selected duplexes. We note from Table 3 and Figure 3a that while the hydrophobic and electrostatic terms contribute minimally and more or less equally, hydrogen bonding, van der Waals, and base stacking terms contribute significantly. Though we observe distinctive values for the base stacking term of the individual duplexes, they show a linear relationship with *N* as displayed in Figure 3b. The trend of variation of these energy terms is similar with those of all the duplexes listed in Table 3. The stiffness parameter (Figure 3c) indicates that the rigidity of the duplex varies with duplex length, the individual equally long molecules exhibiting a high order variation, reflecting the effect of sequence.

There are several points with which to assess the predictive ability level of the model presented here. We have at present restricted the size of the sample of crystal data used for computing the hydrophobic and base stacking energy parameters to 30, which had a resolution within 1.5 Å, and without any kind of modifications in the duplexes. It is likely that this small set could not present all possible base•base doublet situations that exist in the larger duplex set considered for prediction. Also, it is likely that the limitations of treating the total energy as partitioned components (*65*) and residue averages have resulted in an inadequate representation of atomic interactions that determine the conformation of the duplexes. Probably, because of these reasons, the back-predicted $\Delta G_p°$ values for the input molecules themselves resulted at a slightly higher rmsd of 0.7. In light of this inherent level of predictive accuracy in the model presented

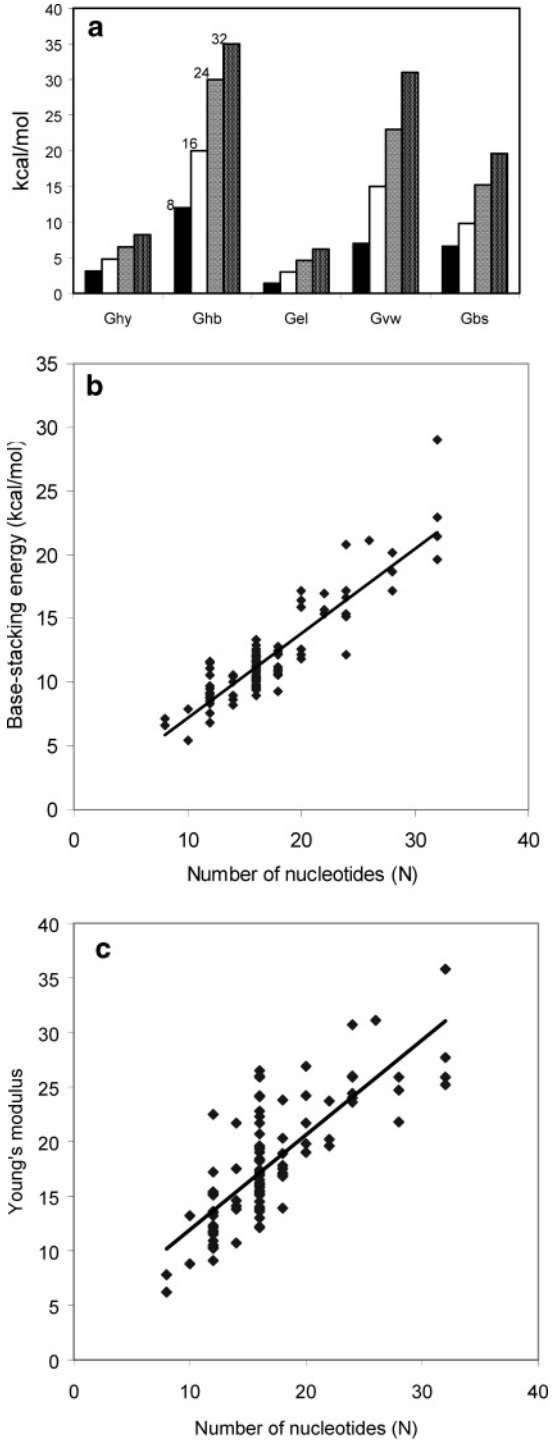

FIGURE 3: (a) Relative contributions of different energy terms to DNA duplex stability ($N = 8$, 16, 24, and 32). (b) Variation of base stacking energy with DNA duplex length. (c) Variation of Young's modulus with DNA duplex length.

here, an rmsd value of 0.9 for the duplexes, which did not form part of the input linear equations, is a very reasonable value for accuracy considerations.

CONCLUSION

This model differs from the widely used nearest-neighbor (NN) model (and the improved ones) in many aspects. Apart from hydrogen bonds, it incorporates the hydrophobic and base stacking energy contributions into the stability which are empirically determined from the crystal structure data of DNA duplexes. It incorporates stability contributions emanating from triplets in the sequence. It also includes effects of duplex initiation and self-complementarity of sequences. Because of these features, we are able to conceive the thermodynamically derived stability of the duplex systems in the form of relative competitive/additive forces in the water environment and predict the stabilities of DNA duplexes very satisfactorily. We are currently working to improve the predictive ability of the model by enlarging the crystal input in a variety of situations, including mismatches, bulges, modifications, etc.

ACKNOWLEDGMENT

We thank M. M. Gromiha, J. P. S. Kumaravel, Ki Shi, and C. Sudarsanakumar for their help in various forms while the work was carried out.

REFERENCES

1. Ramachandran, G. N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins, *Adv. Protein Chem. 23*, 283−438.
2. Fasman, G. D. (1989) in *Prediction of Protein Structure and Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 193−316, Plenum Press, New York.
3. Ponnuswamy, P. K., and Gromiha, M. M. (1993) Prediction of transmembrane helices from hydrophobic characteristics of proteins, *Int. J. Pept. Protein Res. 42*, 326−341.
4. Sternberg, M. J., Bates, B. A., Kelley, L. A., and MacCallum, R. M. (1999) Progress in protein structure prediction: assessment of CASP3, *Curr. Opin. Struct. Biol. 9* (3), 368−373.
5. Cuff, J. A., and Barton, G. J. (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction, *Proteins 34*, 508−519.
6. Mart-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct. 29*, 291−325.
7. Al-Lazikani, B., Jung, J., Xiang, Z., and Honig, B. (2001) Protein structure prediction, *Curr. Opin. Chem. Biol. 5* (1), 51−56.
8. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. (1994) Coaxial Stacking of Helices Enhances Binding of Oligoribonucleotides and Improves Predictions of RNA Folding, *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
9. Pace, C. N., Shirley, B. A., Mcnutt, M., and Gajiwala, K. (1996) Forces Contributing to the Conformational Stability of Proteins, *FASEB J. 10* (1), 75−83.
10. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure, *J. Mol. Biol. 288*, 911−940.
11. Juan, V., and Wilson, C. (1999) RNA Secondary Structure Prediction Based on Free Energy and Phylogenetic Analysis, *J. Mol. Biol. 289*, 935−947.
12. Mathews, D. H., and Turner, D. H. (2002) Dynalign: An Algorithm for Finding the Secondary Structure Common to Two RNA Sequences, *J. Mol. Biol. 317* (2), 191−203.
13. Chiu, W. L. A. K., Sze, C. N., Yip, L. N., Chan, S. K., and Au-Yeung, S. C. F. (2001) NTDB: Thermodynamic Database for Nucleic Acids, *Nucleic Acids Res. 29* (1), 230−233.
14. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R., and Schneider, B. (1992) The Nucleic Acid Database, *Biophys. J. 63*, 751−759.
15. Crothers, D. M., and Zimm, B. H. (1964) Theory of Melting Transition of Synthetic Polynucleotides: Evaluation of Stacking Free Energy, *J. Mol. Biol. 9*, 1−9.
16. Borer, P. N., Dengler, B., Tinoco, I., and Uhlenbeck, O. C. (1974) Stability of ribonucleic acid double-stranded helices, *J. Mol. Biol. 86*, 843−853.
17. Wartell, R. M., and Benight, A. S. (1985) Thermal denaturation of DNA molecules: A comparison of theory with experiment, *Phys. Rep. 126*, 67−107.

18. Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986) Predicting DNA duplex stability from the base sequence, *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746−3750.
19. Sugimoto, N., Nagano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA Duplexes, *Nucleic Acids Res. 24* (22), 4501−4505.
20. SantaLucia, J., Jr., Allawi, H. T., and Ananda Seneviratne, P. (1996) Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability, *Biochemistry 35*, 3555−3562.
21. Owczarzy, R., Vallone, P. M., Gallo, F. J., Paner, T. M., Lane, M. J., and Benight, A. S. (1998) Predicting Sequence Dependent Melting Stability of Short Duplex DNA Oligomers, *Biopolymers 44*, 217−239.
22. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs, *Biochemistry 37*, 14719−14735.
23. Gray, D. M. (1997) Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest-neighbors, *Biopolymers 42*, 783−793.
24. SantaLucia, J., Jr. (1998) A Unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465.
25. Chothia, C. (1974) Hydrophobic Bonding and Accessible Surface Area in Proteins, *Nature 248*, 338−339.
26. Manavalan, P., and Ponnuswamy, P. K. (1978) Average surrounding hydrophobicity, *Nature 275*, 73−74.
27. Ponnuswamy, P. K. (1993) Hydrophobic characteristics of folded proteins, *Prog. Biophys. Mol. Biol. 59*, 57−102.
28. Ponnuswamy, P. K., and Gromiha, M. M. (1994) On the conformational stability of folded proteins, *J. Theor. Biol. 166*, 63−74.
29. Ponnuswamy, P. K., and Gromiha, M. M. (1994) On the Conformational Stability of Oligonucleotide Duplexes and tRNA Molecules, *J. Theor. Biol. 169*, 419−432.
30. Eisenberg, D., and McLachlan, A. D. (1986) Solvation energy in protein folding and binding, *Nature 319*, 199−203.
31. Richmond, T. J., and Richards, F. M. (1978) Packing of α-helices: Geometrical constraints and contact areas, *J. Mol. Biol. 119*, 537−555.
32. Olson, W. K. (1978) Spatial configuration of ordered polynucleotide chains. V. Conformational energy estimates of helical structure, *Biopolymers 17*, 1015−1040.
33. Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA and proteins, *J. Comput. Chem. 16*, 1357−1377.
34. Halgren, T. A. (1992) Representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters, *J. Am. Chem. Soc. 114*, 7827−7843.
35. Gromiha, M. M. (2000) Structure Based Sequence Development Stiffness Scale for Trinucleotides: A Direct Method, *J. Biol. Phys. 26*, 43−50.
36. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv. Protein Chem. 14*, 1−63.
37. Stillinger, F. H. (1980) Water revisited, *Science 209*, 451−457.
38. Creighton, T. E. (1983) An empirical approach to protein conformation stability and flexibility, *Biopolymers 22*, 49−58.
39. Gao, Y. G., Robinson, H. H., and Wang, A. H.-J. (1999) High-resolution A-DNA Crystal Structures of d(AGGGGCCCCT): An A-DNA Model of poly(dG)·poly(dC), *Eur. J. Biochem. 261*, 413−420.
40. Egli, M., Tereshko, V., Teplova, M., Minasov, G., Joachimiak, A., Sanishvili, R., Weeks, C. M., Miller, R., Maier, M. A., An, H., Dan Cook, P., and Manoharan, M. (1998) X-ray Crystallographic Analysis of the Hydration of A- and B-form DNA at Atomic Resolution, *Biopolymers 48*, 234−252.
41. Thota, N., Li, X. H., Bingman, C., and Sundaralingam, M. (1993) High-resolution refinement of the hexagonal A-DNA octamer d(GTGTACAC) at 1.4 Å, *Acta Crystallogr. D49*, 282−291.
42. Vlieghe, D., Turkenburg, J. P., and Van Meervelt, L. (1999) B-DNA at atomic resolution reveals extended hydration patterns, *Acta Crystallogr. D55*, 1495−1502.
43. Tereshko, V., Minasov, G., and Egli, M. (1999) A "Hydrat-Ion Spine" in a B-DNA minor groove, *J. Am. Chem. Soc. 121*, 3590−3595.
44. Liu, J., and Subirana, J. A. (1999) Structure of d(CGCGAATTCGCG) in the presence of Ca^{2+} ions, *J. Biol. Chem. 274*, 24749−24752.
45. Minasov, G., Tereshko, V., and Egli, M. (1999) Atomic-resolution Crystal Structures of B-DNA Reveal Specific Influences of Divalent Metal Ions on Conformation and Packing, *J. Mol. Biol. 291*, 83−99.
46. Gao, Y. G., Robinson, H., Sanishvili, R., Joachimiak, K., and Wang, A. H.-J. (1999) Structure and recognition of sheared tandem G·A base pairs associated with human centromere DNA sequence at atomic resolution, *Biochemistry 38*, 16452−16460.
47. Chiu, T. K., and Dickerson, R. E. (2000) A crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium, *J. Mol. Biol. 301*, 915−945.
48. Soler-Lopez, M., Malinina, L., and Subirana, J. A. (2000) Solvent organization in an oligonucleotide crystal. The structure of $d(GCGAATTCG\$)_2$ at atomic resolution, *J. Biol. Chem. 275*, 23034−23044.
49. Minasov, G., Teplova, M., Neilson, P., Wengel, J., and Egli, M. (2000) Structural basis of cleavage by RNase H of hybrids of arabinonucleic acids and RNA, *Biochemistry 39*, 3525−3532.
50. Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D., and Williams, L. D. (2000) Cations Mediate B-DNA Conformational Heterogeneity, *J. Am. Chem. Soc. 122*, 11048−11056.
51. Prive, G. G., Yanagi, K., and Dickerson, R. E. (1991) Structure of the B-DNA Decamer C-C-A-A-C-G-T-T-G-G and Comparison with Isomorphous Decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G, *J. Mol. Biol. 217*, 177−199.
52. Grzeskowiak, K., Yanagi, K., Prive, G., and Dickerson, R. E. (1991) The Structure of B-Helical C-G-A-T-C-G-A-T-C-G and Comparison with C-C-A-A-C-G-T-T-G-G. The Effect of Base Pair Reversals, *J. Biol. Chem. 266*, 8861−8883.
53. Quintana, J. R., Grzeskowiak, K., Yanagi, K., and Dickerson, R. E. (1992) The Structure of a B-DNA Decamer with a Central T-A Step: C-G-A-T-T-A-A-T-C-G, *J. Mol. Biol. 225*, 379−395.
54. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K., and Rees, D. C. (2000) Intercalation of a DNA helix by a photoexcitable rhodium complex at 1.2 Å resolution, *Nat. Struct. Biol. 7*, 117−121.
55. Cruse, W. B. T., Egert, E., Kennard, O., Sala, G. B., Salisbury, S. A., and Viswamitra, M. A. (1983) Self-base pairing in a complementary deoxydinucleoside monophosphate duplex: crystal and molecular structure of deoxycytidylyl-(3′−5′)-deoxyguanosine, *Biochemistry 22*, 1833−1839.
56. Coll, M., Solans, X., Font-Altaba, M., and Subirana, J. A. (1987) Crystal and Molecular Structure of the Sodium Salt of the Dinucleotide Duplex d(CpG), *J. Biomol. Struct. Dyn. 4*, 797−811.
57. Nunn, C. M., and Niedle, S. (1996) The High-Resolution Crystal Structure of the DNA Decamer d(AGGCATGCCT), *J. Mol. Biol. 256*, 340−351.
58. Ohishi, H., Tomita, K.-i., Naganishi, I., Ohtsuchi, M., Hakoshima, T., and Rich, A. (2004) Crystal Structure of a Molecular Complex of a Left-handed Z-DNA Hexamer, $d(CG)_3$, with Synthetic Polyamine: Polyamine Binding in the Minor Groove at Room Temperature (to be submitted for publication).
59. Crawford, J. L., Kolpak, F. J., and Wang, A. H.-J. (1980) The Tetramer d(CpGpCpG) Crystallizes as a Left-Handed Double Helix, *Proc. Natl. Acad. Sci. U.S.A. 77*, 4016−4020.
60. Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A., and Wang, A. H.-J. (1989) The Molecular Structure of the Left-Handed Z-DNA Double Helix at 1.0 Å Atomic Resolution. Geometry, Conformation, and Ionic Interactions of d(CGCGCG), *J. Biol. Chem. 264*, 7921−7935.
61. Ho, P. S., Frederick, C. A., Quigley, G. J., van Der Marel, G. A., Van Boom, J. H., Wang, A. H.-J., and Rich, A. (1985) GT Wobble Base-Pairing in Z-DNA at 1.0 Å Atomic Resolution: The Crystal Structure of d(CGCGTG), *EMBO J. 4*, 3617−3623.
62. Kagawa, T. F., Geierstanger, B. H., Wang, A. H.-J., and Ho, P. S. (1991) Covalent Modification of Guanine Bases in Double Stranded DNA: The 1.2 Å Z-DNA Structure of d(CGCGCG) in the Presence of $CuCl_2$, *J. Biol. Chem. 266*, 20175−20184.
63. Ohishi, H., Terasoma, N., Naganishi, I., Van Der Marel, G., Van Boom, J. H., Rich, A., Wang, A. H.-J., Hakoshima, T., and Tomita,

K.-I. (1996) Interaction Between Left-Handed Z-DNA and Polyamine-3. The Crystal Structure of the $d(CG)_3$ and Thermospermine Complex, *FEBS Lett. 398*, 291−296.
64. Harper, N. A., Brannigan, J. A., Buck, M., Lewis, R. J., Moore, M. H., and Schneider, B. (1998) The Structure of $d(TGCGCA)_2$ and a Comparison to Other Z-DNA Hexamers, *Acta Crystallogr. D54*, 1273−1284.
65. Chevrier, B., Dock, A. C., Hartmann, B., Leng, M., Moras, D., Thuong, M. T., and Westhof, E. (1986) Solvation of the Left-Handed Hexamer d(5BrC-G-5BrC-G-5BrC-G) in Crystals Grown at Two Temperatures, *J. Mol. Biol. 188*, 707−719.
66. MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, *J. Phys. Chem. B 102*, 3586−3616.

BI048158+