Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 4320–4325, April 1996
Biochemistry

Prediction of the stability of DNA triplexes

(van't Hoff enthalpy/free energy/cytosine repulsion/nearest neighbor)

RICHARD W. ROBERTS* AND DONALD M. CROTHERS†

*Department of Molecular Biology, Massachusetts General Hosipital, Boston, MA 02114; and †Department of Chemistry, Yale University, New Haven, CT 06511

Contributed by Donald M. Crothers, January 18, 1996

ABSTRACT **We present rules that allow one to predict the stability of DNA pyrimidine·purine·pyrimidine (Y·R·Y) triple helices on the basis of the sequence. The rules were derived from van't Hoff analysis of 23 oligonucleotide triplexes tested at a variety of pH values. To predict the enthalpy of triplex formation ($\Delta H°$), a simple nearest-neighbor model was found to be sufficient. However, to accurately predict the free energy of the triplex ($\Delta G°$), a combination model consisting of five parameters was needed. These parameters were (*i*) the $\Delta G°$ for helix initiation, (*ii*) the $\Delta G°$ for adding a T–A·T triple, (*iii*) the $\Delta G°$ for adding a C^+–G·C triple, (*iv*) the penalty for adjacent C bases, and (*v*) the pH dependence of the C^+–G·C triple's stability. The fitted parameters are highly consistent with thermodynamic data from the basis set, generally predicting both $\Delta H°$ and $\Delta G°$ to within the experimental error. Examination of the parameters points out several interesting features. The combination model predicts that C^+–G·C triples are much more stabilizing than T–A·T triples below pH 7.0 and that the stability of the former increases ≈1 kcal/mol per pH unit as the pH is decreased. Surprisingly though, the most stable sequence is predicted to be a CT repeat, as adjacent C bases partially cancel the stability of one another. The parameters successfully predict t_m values from other laboratories, with some interesting exceptions.**

Prediction of macromolecular physical properties using only the sequence is one of the primary goals of biophysical chemistry. The utility of such models is in the ability to predict the stability of RNA or DNA helices based on the sequence alone (1, 2), the flexibility of the DNA duplex (3), local helical geometry (4, 5), and the propensity of a particular sequence to undergo a structural transition between different helical forms (6). A predictive model for triple-helix stability could have broad application as well. Triple-helix formation can be used to recognize DNA duplexes highly specifically (for reviews, see refs. 7–10) and has potential for antisense and therapeutic applications. In this paper, we present a model for predicting DNA triplex stability using only the sequence.

In developing a predictive model, two features are important: (*i*) the appropriateness of the parameters used to construct the model and (*ii*) the distribution of the sequences used as the basis set in its parameterization. For the prediction of the enthalpy of triplex formation ($\Delta H°$), a nearest neighbor model was used, whereas a combination model (one containing a mixture of mono- and dinucleotide parameters) was found to be best for prediction of the free energy of the triplex ($\Delta G°$). A wide variety of sequences was used as the basis set for determination of the model parameters. The basis set does contain more GC than AT rich sequences. However, the accuracy of predictions in both $\Delta G°$ and $\Delta H°$ for all the sequences tested indicates that this is not a serious problem. The models presented here provide the first framework for predicting triplex stability in a broad variety of sequences and conditions.

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

MATERIALS AND METHODS

Oligonucleotides. All DNA used was synthesized as described (11). Deprotected samples were purified by denaturing PAGE, either soaked or electro-eluted from the gel using an Elutrap (Schleicher & Schuell) and desalted with a NAP 25 Sephadex column (Pharmacia).

Melting Experiments. Melting curves were taken with a Cary 1 spectrophotometer. Temperature was controlled with Peltier cell block and measured with a thermistor inserted into an adjacent temperature reference cuvette. All samples were equilibrated by cooling from 85°C to room temperature in a buffer consisting of either 100 mM [Na^+] acetate/acetic acid, 1 mM EDTA (experiments at pH 4.75–6.0) or 100 mM [Na^+] cacodylate/cacodylic acid, 1 mM EDTA (experiments at pH 6.25–7.0). The Na^+ concentration was held constant at 100 mM because it is known to affect the triplex t_m (10, 12) Melting experiments were performed by heating from low temperature to high temperature at ≈0.5°C per min and data were taken every 0.5°C. t_m values were reproducible ±0.5°C.

Determination of Triplex Stability. The stability of the 23 intermolecular triplexes was determined by van't Hoff analysis of absorbance melting curves over a pH range from 4.75 to 7.0. Between pH 4.75 and 5.5, curves were analyzed as described (11, 13). Briefly, the melting data were treated with a statistical mechanical approach where the equilibrium is broken up into discrete states (11, 14) with melting given by

$$\begin{array}{lccccc} & \text{Triplex} & \overset{K_3}{\rightleftharpoons} & \text{Hairpin} + \text{Third strand} & \overset{K_2}{\rightleftharpoons} & \text{Coil} \\ \text{State:} & (3) & & (2) & & (1) \end{array}$$

$$+ \text{ Third Strand} \qquad [1]$$

and the fraction of triplex, Θ_3, and duplex, Θ_2, given by

$$\Theta_3(t) = \frac{K_2K_3(C_t/2)(1 - \Theta_3(t))}{[K_2K_3(C_t/2)(1 - \Theta_3(t)) + K_2 + 1]} \qquad [2]$$

$\Theta_2(t)$ is given by adding K_2 to the numerator of Eq. **2**. K_2 and K_3 are the formation constants for the hairpin and triplex, respectively. C_t is the total concentration of oligonucleotide and $C_t/2$ is the concentration of the third strand. To determine the values of $\Delta H°_{\text{van't Hoff}}$ and t_m, experimental derivative melting curves (dAbs./dt vs. t) were compared with calculated curves and the parameters ($\Delta H°_2$, $\Delta H°_3$, t_{m2}, and t_{m3}) were varied to give the best fit. Trial derivative melting curves were generated by solving Eq. **2** and the equation for $\Theta_2(t)$ numerically over the temperature range of the melting curve. $\Delta G°$ and the standard entropy change $\Delta S°$ were then determined for each of the triplexes and duplexes studied using the $\Delta H°_{\text{van't Hoff}}$ and t_m data. The uncertainties in $\Delta H°_2$ and $\Delta H°_3$ are approximately ±5 kcal/mol producing an uncertainty in $\Delta G°$ of ±0.5 kcal/mol.

Above pH 5.5, van't Hoff analysis of the curves becomes difficult due to hysteresis caused by slow formation kinetics. t_m values for transitions above this pH were estimated using $\Delta H°$ values determined at a lower pH to generate a theoretical curve. This was aligned with the upper half of the experimental curve (where hysteresis artifacts should be minimized) to determine t_m. [A complete table of the thermodynamic parameters ($\Delta G°$, $\Delta H°$, $\Delta S°$, andt_m) determined using this analysis is available from the authors upon request.] As with previous work, all analysis assumes a zero or negligible value of ΔCp.

Determination of Predictive Rules. The $\Delta H°$ and $\Delta G°$ values determined from van't Hoff analysis at pH 5.0 were used as the basis set to calculate the predictive rules. Using the sequence information and energy values, a set of 23 simultaneous equations was generated. For $\Delta H°$ these equations conformed to a nearest neighbor model as shown in Eq. **3** in Table 1. TT, TC/CT, and CC represent the number of TT, CT or TC, and CC dinucleotides, respectively. Note, no attempt is made to distinguish between the TC and CT dinucleotides. This is because neither of these dinucleotides can outnumber the other by more than ±1 in a given sequence. For prediction of $\Delta G°$, both a dinucleotide and combination model were tested, the latter being used for the final set of rules. In the combination model (see Table 1, Eq. **2a**) T, C, and CC equal the number of T bases, C bases, and CC dinucleotides contained in the third strand. The dinucleotide model tested had the same form as Eq. **1** in Table 1 with an extra parameter added for helix nucleation.

To find solutions for each model, the simultaneous equations were cast as either a 23 × 3 ($\Delta H°$) or 23 × 4 ($\Delta G°$) matrix and a 23 × 1 column vector. To obtain a best fit, we minimized the sum of the residuals squared using the multilinear regression protocol described by Bevington (17). The calculations were done using the MATHMATICA software package (Wolfram Research, Champaign, IL). After the rules at pH 5.0 were calculated, a term (α), which represents the free energy decrease per pH unit per cytosine, was included to incorporate the pH dependence of the triplex. α was determined by fitting the pH dependence of 13 different sequences plotted vs. the number of CC dinucleotides present in the strand. Inclusion of this parameter (Table 1, Eq. **2b**) gives the general model for prediction of $\Delta G°$ at all pH values.

RESULTS AND DISCUSSION

Sequences. The sequences used in this study are shown in Fig. 1. There are four categories of sequences corresponding to the four different duplex sites tested. The largest category is that corresponding to the WC28/28WC + PY12 family (Fig. 1*A*). The two hairpin duplexes differ only in the position of the loop that connects the PY and PU strands. The 10 different third strands correspond to either 5′ or 3′ deletions of the full-length sequence (PY12) and can be combined with the hairpins to make 20 different triplexes. The variable length of the members was constructed to dissect the free energy penalty for removal of a single T–A·T or C^+–G·C triple. The two different hairpin loops were designed to address the effect of the loop and flanking sequences on the on the stability of the triplex.

The other three groups (Fig. 1 *B–D*) each consist of a single third strand and a hairpin duplex. These sequences were designed to contain different base composition and dinucleotide makeup than the WC28 + PY12 construct. The TC28 + TC12 triplex (Fig. 1*B*) is predominantly a redundant TC repeat with a single CC dinucleotide inserted to prohibit heterogeneity of third strand binding. The SWAP28 + 12 triplex (Fig. 1*C*) was derived from the WC28 + PY12 construct by swapping every T–A·T triple with a C^+–G·C and vice versa. The ATSWAP28 + 12 triplex contains one fewer C^+–G·C triple than the SWAP triplex.

Third Strand Deletions. To predict $\Delta G°$ for the triplex, the stabilities of several 5′ and 3′ deletions of the third strand were determined. Since the free energy change for deletion of a single base should be equal and opposite to the free energy change for adding it, examination of a number of T and C deletions could yield an accurate value for the stabilization of a T–A·T or a C^+–G·C triple.

Table 1. Equations for stability prediction (all values in kcal/mol)

Enthalpy ($\Delta H°$)

$$\Delta H° = -4.9(CC) - 8.9(TC + CT) - 7.4(TT) \quad [1]$$

Standard free energy ($\Delta G°$) at 37°C

$$\Delta G°_{37,inter} = -3.00(C) - 0.65(T) + 1.65(CC) + 6.0 \quad (pH=5.0) \quad [2a]$$

$$+(C)(pH - 5.0)(1.26 - 0.08(CC)) \quad (other\ pH\ values) \quad [2b]$$

$$\Delta G°_{37,intra} = \Delta G°_{37,inter} + \Delta G°_{loop} \quad [3]$$

Melting temperature of inter- and intramolecular triplexes

$$t_{m,inter} = \frac{310 \cdot \Delta H°}{\Delta H° - \Delta G°_{37,inter} - 310 \cdot R\ln\left(\frac{4}{C_t}\right)} \quad [4]$$

$$t_{m,intra} = \frac{310 \cdot \Delta H°}{\Delta H° - \Delta G°_{37,intra}} \quad [5]$$

K_{eq} at temperatures (t) other than 37°C

$$K_{eq} = \exp\left[\left(\frac{1}{R}\right)\left(\frac{\Delta H° - \Delta G°_{37}}{310} - \frac{\Delta H°}{t}\right)\right] \quad [6]$$

CC, no. of CC dinucleotides; TC + CT, no. of TC + CT dinucleotides; TT, no. of TT dinucleotides; C, no. of C bases; T, no. of T bases; 6.0 kcal/mol, $\Delta G°_{nucleation}$; $t_{m,inter}$, intermolecular melting temperature; $t_{m,intra}$, intramolecular melting temperature; ΔG_{loop}, see refs. 15 and 16; C_t, duplex + third strand concentration.

A

	Name
5'-GGAGAGGAGGGA T T 3'-CCTCTCCTCCCT T T	WC28
T TGGAGAGGAGGGA-3' T TCCTCTCCTCCCT-5'	28WC
5'-CCTCTCCTCCCT-3'	PY12
5' Deletions	
5'-CTCTCCTCCCT-3'	5'PY11
5'-TCTCCTCCCT-3'	5'PY10
5'-CTCCTCCCT-3'	5'PY9
5'-TCCTCCCT-3'	5'PY8
3' Deletions	
5'-CCTCTCCTCCC-3'	3'PY11
5'-CCTCTCCTCC-3'	3'PY10
5'-CCTCTCCTC-3'	3'PY9
5'-CCTCTCCT-3'	3'PY8
3' and 5' Deletion	
5'-CTCTCCTCCC-3'	5'3'PY10

B

5'-AGAGAGAGGAGA T T
3'-TCTCTCTCCTCT T T
TC28

5'-TCTCCTCTCTCT-3'
TC12

C

5'-AAGAGAAGAAAG T T
3'-TTCTCTTCTTTC T T
SWAP28

5'-TTCTCTTCTTTC-3'
SWAP12

D

5'-AAGAAAAGAAAG T T
3'-TTCTTTTCTTTC T T
ATSWAP28

5'-TTCTTTTCTTTC-3'
ATSWAP12

FIG. 1. Sequences used. The hairpin duplex sites form triplexes when combined with the pyrimidine strands in their respective groups (*A–D*). (*A*) WC28/28WC + PY12 constructs. The set consists of two hairpin duplexes WC28 and 28WC that differ only in the end of the helix in which the hairpin lies. The full-length third strand is PY12 and all the other pyrimidine strands represent 5′ or 3′ deletions of this sequence. (*B*) TC28 + TC12 construct. (*C*) SWAP12 + SWAP28 construct. (*D*) ATSWAP12 + ATSWAP28 construct.

Fig. 2 demonstrates that this simple strategy produces a confusing picture of triplex stability. In particular, the stability

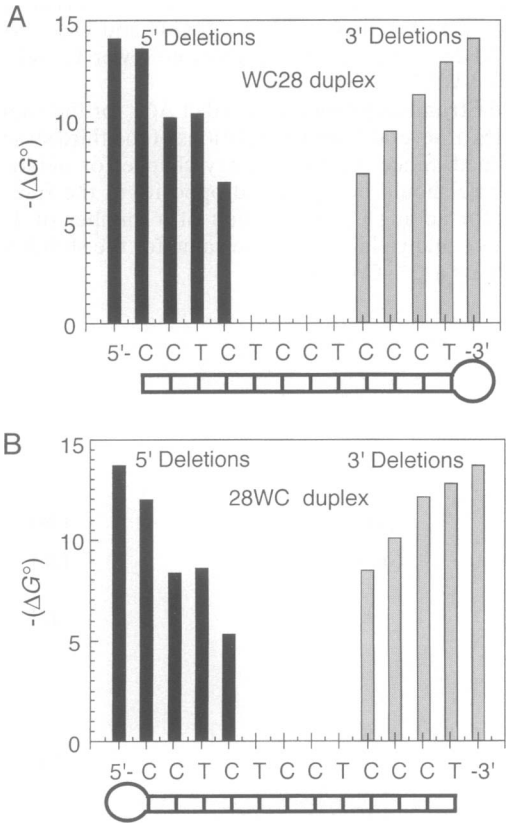

FIG. 2. Stability of 3′ and 5′ third strand deletions in the WC28/28WC + PY12 construct. The length of the bars indicates the stability (plotted as $-\Delta G°$, kcal/mol) of the triplex and is plotted over the position of the last base deleted from the full-length sequence, PY12. Bars over the 5′ and 3′ labels indicate the stability of the full-length third strand. (*A*) Stability of 3′ and 5′ third strand deletions when combined with WC28. Here, the loop lies at the 3′ end of the third strand. (*B*) Stability of 3′ and 5′ third strand deletions when combined with 28WC. Here, the duplex hairpin loop lies at the 5′ end of the third strand. In both plots, the triplex is more stable when it is adjacent to the loop as compared with the end of the helix.

of C^+–G·C triples depends strongly on where each lies in the sequence, analogous to observations for the G–T·A triple (18). For example, removal of the 5′ C from the third strand (Fig. 2*A*) produces little change in the $\Delta G°$ for the triplex. However, removal of the next C from the 5′ end results in a large drop in the stability. Third strand T bases behave more systematically, with removal being accompanied by a small destabilization in third strand binding.

Fig. 2 also demonstrates that triplex stability depends somewhat on flanking structure provided by the hairpin loop at the end of the duplex. Both the 3′ and 5′ deletions reveal that the stability of the triplex is enhanced when the end of the third strand abuts the loop. This observation runs counter to the prediction one would make based on the electrostatic contribution from the extra bases in the loop (i.e., the extra charge should destabilize the triplex).

The context dependence of C^+–G·C triples indicated that another rule was needed to describe triplex stability. Examination of all the third strand deletions revealed that addition of a cytosine adjacent to a thymine provided a much larger stabilization than addition of one next to another cytosine. Thus, adjacent C bases on the third strand could be thought of as penalized (in a free energy sense), perhaps due to the proximity of their positive charges. To quantitate this effect, we included a variable parameter in the $\Delta G°$ model multiplied by the number of CC dinucleotides present in the third strand.

Phase Diagram Crossover. Examination of the phase diagrams for the other triplexes studied reinforced the notion of repulsion by adjacent cytosines (Fig. 3). The phase diagrams of three triplexes in Fig. 3 (TC28 + TC12, SWAP28 + 12, and ATSWAP28 + 12) converge around pH 7.0. The relative stability of the three (TC28 + TC12 > SWAP28 + 12 > ATSWAP28 + 12) at pH values below 7.0 can be attributed simply to the higher stability of C^+–G·C triples compared with T–A·T triples under these conditions. However, the phase diagram for WC28 + PY12 fails to fall into this pattern, crossing the lines from the other triplexes well below pH 7.0.

One explanation for the crossover is that the WC28 + PY12 curve is just downwardly displaced from the other curves because of repulsion from adjacent cytosines. This rationale implies the other sequences show the same intercept because they are nearly devoid of CC dinucleotides (TC28 + TC12 contains one). The crossover is probably not due to anomalies in the pH dependence for the sequences. This is because the slope of each sequence is roughly proportional to the number of cytosines in the third strand and is linear over the range of the experiments.

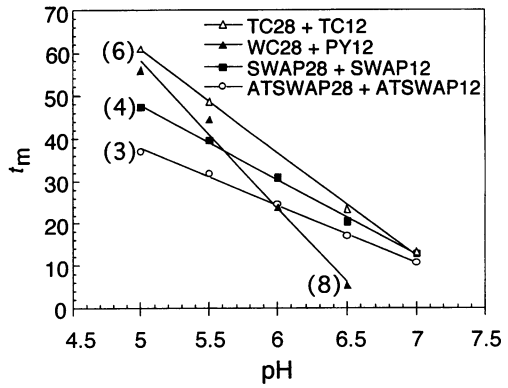

FIG. 3. Crossover in the pH dependence of triplex stability. The figure shows the pH phase diagram for the four triplex groups studied. The key is shown in the upper right and the number of cytosines in the third strand is indicated in parentheses next to each line. Three of the sequences (TC28 + TC12, SWAP28 + SWAP12, and ATSWAP28 + SWAP12) show a convergence near pH 7.0, whereas the WC28 + PY12 stability crosses the other curves well below this value.

pH Dependence. The simplest approach to predict the pH dependence of the triplex was to search for a single parameter α that would indicate the $\Delta G°$ increase per cytosine when the pH was raised one unit. In the simplest case, α would be the same for all sequences, and equal to the $\Delta G°$ for protonation of a base n pH units above its negative logarithm of association constant, −2.303 nRT (1.36 kcal/mol·unit). However, we observed that α can vary significantly from one sequence to another. The general trend shows that third strands that contain several CC dinucleotides or are C-rich have small α values ($\alpha \approx 1.0$) whereas those rich in T approach the theoretical limit ($\alpha \approx 1.3$). The observation that α is often significantly less than 1.36 revealed that the number of protons involved in triplex formation is somewhat less than the number of cytosines in the third strand. This could result from either incomplete protonation of the triplex (e.g., caused by repulsion from adjacent cytosines) or incomplete deprotonation of the third strand [e.g., from C^+–C self structure (11)] when the triplex melted. To formulate this behavior as a rule, we fit a plot of α vs. the number of CC dinucleotides in the third strand for 13 triplexes to a line and derived Eq. **2B** in Table 1.

Predictive Rules. Addition of the CC repulsion and pH dependence parameters provided the final pieces needed to develop energy models for calculation of triplex $\Delta H°$ and $\Delta G°$. Using these, the t_m and K_{eq} can be calculated for each case.

$\Delta H°$ Prediction. The most reasonable model for $\Delta H°$ prediction was a nearest-neighbor model with three parameters: (*i*) the number of CC steps, (*ii*) the total number of TC or CT steps, and (*iii*) the number of TT steps (see Eq. **1** in Table 1). This is because the enthalpy primarily reflects stacking interactions between adjacent bases. The fitting procedure confirmed this notion, producing predictions quite close to the experimentally determined numbers in all but one case (28WC + 5′PY10) and an rms residual of ≈4 kcal/mol, roughly the reproducibility of the experiment (Table 2). No terms were included in the model for temperature, pH, or buffer contributions to $\Delta H°$. The first two of these are unlikely to have an effect, as the change in heat capacity $\Delta C_p \approx 0$, and any pH dependence would likely be relevant over a very narrow range (±1 pH unit of the negative logarithm of association constant of cytosine, pH 3.5–5.5). When buffers with large enthalpies of protonation are used, our $\Delta H°$ predictions should be amended to include ionization of the buffer.

$\Delta G°$ Prediction. Two models were tested for prediction of the $\Delta G°$ for the triplex, a combination model and a nearest neighbor model (see *Materials and Methods*). The two models were evaluated by using each to predict the $\Delta G°$ of the 23 basis sequences at both 25°C and 37°C. The combination model predicted triplex stability better than the dinucleotide model at both temperatures and worked best at 37°C with an rms residual of 0.61 kcal/mol. The final model is presented in Eqs. **2a** and **2b** in Table 1, with intramolecular predictions in Eq. **3**. The model derived from $\Delta G°_{37}$ is likely to have a smaller rms deviation than the $\Delta G°_{25}$ data because it is closer to the mean t_m of the basis set, 318.9 K, making it less susceptible to errors in the $\Delta H°$ prediction. The model (Table 2) appears to provide quite good prediction for all the sequences tested, neither overnor underestimating any type of sequence in the basis set.

Physical Interpretation of Parameters. The size and sign of the parameters reveals several properties of triplex formation. First, formation of a C^+–G·C triple at pH 5.0 is about as favorable ($\Delta G° = -3.0$ kcal) as formation of G·C pair in RNA or DNA whereas formation of a T–A·T triple ($\Delta G° = -0.65$ kcal) is much less stable, even less so than an AU or AT pair (1, 2). However, several factors act to decrease the stability of C rich sequences. First, adjacent C bases are disfavored by 1.65 kcal/mol. In addition, the stability of the C^+–G·C triple decreases 1–1.3 kcal per pH unit as the pH is increased above 5.0. This closes the gap between the two triples resulting in their energetic equivalence around pH 7.0, just as the curves in Fig. 3 imply. This crossover in stability, combined with the repulsion from adjacent cytosines, produces the result that a CT repeat will be the most stable triplex below pH 7.0. Above pH 7.0 a pure T–A·T triplex would be predicted to be the most

Table 2. Comparison of measured and predicted $\Delta H°$ and $\Delta G°$ (≈37°C) for the triplexes examined at pH 5.0

Triplex	$\Delta H°$		$\Delta G°$	
	Measured	Predicted	Measured	Predicted
WC28 + PY12	−85	−82.1	−14.1	−13.9
WC28 + 5′PY11	−80	−77.2	−13.6	−12.6
WC28 + 3′PY11	−80	−73.2	−12.9	−13.3
WC28 + 5′PY10	−67.5	−68.3	−10.2	−9.6
WC28 + 3′PY10	−70	−68.3	−11.3	−11.9
WC28 + 3′5′PY10	−70	−68.3	−12.9	−11.9
WC28 + 5′PY9	−65	−59.4	−10.4	−8.9
WC28 + 3′PY9	−57.5	−63.4	−9.5	−10.6
WC28 + 5′PY8	−47.5	−50.4	−7.0	−6.0
WC28 + 3′PY8	−50	−54.5	−7.4	−7.6
28WC + PY12	−85	−82.1	−13.7	−13.9
28WC + 5′PY11	−72.5	−77.2	−12.0	−12.6
28WC + 3′PY11	−75	−73.2	−12.8	−13.3
28WC + 5′PY10	−52.5	−68.3	−8.3	−9.6
28WC + 3′PY10	−75	−68.3	−12.1	−11.9
28WC + 3′5′PY10	−72.5	−68.3	−11.7	−11.9
28WC + 5′PY9	−57.5	−59.4	−8.6	−8.9
28WC + 3′PY9	−60	−63.4	−10.1	−10.6
28WC + 5′PY8	−45	−50.4	−5.3	−6.0
28WC + 3′PY8	−50	−54.5	−8.5	−7.6
TC28 + TC12	−100	−94.2	−15.1	−14.2
SWAP28 + 12	−90	−91.9	−11.2	−11.2
ATSWAP28 + 12	−90	−88.8	−8.4	−8.8
rms Residual*	4 kcal/mol		0.61 kcal/mol	

All values are expressed as kcal/mol.
*The residual equals the absolute value of measured − predicted numbers.

stable. $\Delta G°$ for nucleation of the triplex (6.0 kcal) is about the same as that found for duplex DNA (2) and almost twice as large as that of duplex RNA (1).

The notion of C^+–G·C triples having higher stability than T–A·T triples is contrary to some initial observations made from NMR spectroscopy (19) but in line with recent thermodynamic investigations (20). NMR data revealed a triplex consisting of three separate strands preferred to have an overhanging C rather than an overhanging T. The discrepancy between this observation and our data could be due to chemical exchange of the terminal C^+–G·C imino proton, making it invisible, or to stabilization of the C overhang by CC pairing to cytosines at the end of other triplexes.

The enthalpy of triplex formation has been a matter of some contention. In some cases, the calorimetric enthalpy per base triple is much smaller than the van't Hoff enthalpy measured on the same triplexes (12, 21, 22). Our van't Hoff $\Delta H°$ parameters (Table 1, Eq. **1**) are larger than these $\Delta H°_{cal}$ data but consistent with other examples (23). In addition, our results are consistent with other work, including equilibrium competition (11), kinetic measurements (13, 24) and other van't Hoff analyses (12, 21–23, 25, 26). Smaller values of $\Delta H°$ would imply that the triplex loses very few of its protons upon melting [≈35% (12)], a relatively unexpected result.

t_m Prediction. The final test of our $\Delta H°$ and $\Delta G°$ models' utility is to use them to predict both the t_m of our data and the data of others. In Fig. 4, the measured or reported value of t_m is plotted vs. the value calculated from our energy models. Fig. 4*A* demonstrates that the models predict our data quite well, neither systematically under nor over estimating the value of t_m over more than 50°C, a wide variety of sequences, and pH values. The model seems to have little sequence bias and good

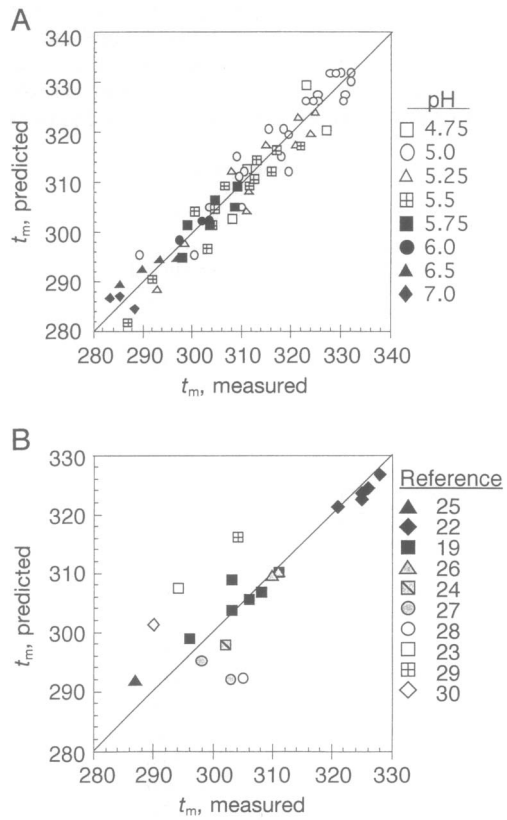

FIG. 4. Comparison of measured vs. predicted melting temperatures for triplexes in this work (*A*) and those from the literature (*B*). The pH of the experiment or the reference used is indicated at the right. In both cases, the measured/reported t_m is plotted vs. the predicted t_m derived from the model in Table 1. The construction line has a slope of 1 and passes through the origin.

precision with the rms error of 3.4°C in t_m. This corresponds to ≈1.5 kcal/mol in $\Delta G°$ for the sequences used, which we feel is quite good given the limited number of parameters in the model and the extrapolations involved.

The model's stability prediction of data from other laboratories is also quite good (Fig. 4*B*), with the exception of affinity cleavage data, even though the ionic conditions are not uniform. In many thermal melting and calorimetric experiments, we are able to predict the melting temperature (21, 23, 26–28) within 6°C, only slightly worse than the members of our data set (Fig. 4*B*). Once again, the model shows no apparent bias in GC content or pH. There are five cases where our prediction is off by ≈11–13°C (24, 29–32). In two cases (29, 30), our underestimate of t_m is likely due to the high spermine concentration used (0.5–1.0 mM). Our predictions are accurate for oligos between 8 and 16 nt long, but overestimate the stability of a 22-mer sequence (24, 31) perhaps indicating a nonadditive length dependence for long triplexes.

Our predictions only coincide with affinity cleavage when it is done under conditions we would predict to be close to t_m. The difference in $\Delta G°$ predicted from these experiments generally underestimates that which we derive for strand composition (33, 34) pH dependence (35), and length dependence (36). We note however, that the range and ranking observed mirrors the association rate constant data we have measured on similar systems (unpublished results).

R.W.R. would like to thank Maja Mataric and Michael Bolotski for help with the calculations.

1. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9373–9377.
2. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3746–3750.
3. Gartenberg, M. R. & Crothers, D. M. (1988) *Nature (London)* **333,** 824–829.
4. Trifonov, E. N. & Sussman, J. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3816–3820.
5. Tung, C.-S. & Harvey, S. C. (1986) *J. Biol. Chem.* **8,** 3700–3709.
6. Peticolas, W. L., Yang, Y. & Thomas, G. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2579–2583.
7. Cheng, Y. K. & Pettitt, M. B. (1992) *Prog. Biophys. Mol. Biol.* **58,** 225–257.
8. Nguyen, T. T. & Helene, C. (1993) *Angew. Chem. Int. Ed. Engl.* **105,** 666–690.
9. Maher, L. J. I., Wold, B. & Dervan, P. B. (1991) *Antisense Res. Dev.* **1,** 227–281.
10. Plum, G. E., Pilch, D. S., Singleton, S. F. & Breslauer, K. J. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 319–350.
11. Roberts, R. W. & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9397–9401.
12. Plum, G. E. & Breslauer, K. J. (1995) *J. Mol. Biol.* **248,** 679–695.
13. Roberts, R. W. (1993) Ph.D. thesis (Yale Univ., New Haven, CT).
14. Eisenberg, D. & Crothers, D. M. (1979) *Physical Chemistry with Applications to the Life Sciences* (Benjamin/Cummins, Menlo Park, CA).
15. Roberts, R. W. & Crothers, D. M. (1996) *J. Mol. Biol.*, in press.
16. Booher, M. A., Wang, S. & Kool, E. T. (1994) *Biochemistry* **33,** 4645–4651.
17. Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences* (McGraw–Hill, New York).
18. Kiessling, L. L., Griffin, L. C. & Dervan, P. B. (1992) *Biochemistry* **31,** 2829–2834.
19. Rajagopal, P. & Feigon, J. (1989) *Nature (London)* **339,** 637–640.
20. Wilson, W. D., Hopkins, H. P., Mizan, S., Hamilton, D. D. & Zon, G. (1994) *J. Am. Chem. Soc.* **116,** 3607–3608.
21. Plum, G. E., Park, Y.-W., Singleton, S. F., Dervan, P. B. & Beslauer, K. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9436–9440.
22. Volker, J., Botes, D. P., Lindsey, G. G. & Klump, H. H. (1993) *J. Mol. Biol.* **230,** 1278–1290.
23. Manzini, G., Xodo, L. E., Gasparotto, D., Quadrifoglio, F., van der Marel, G. A. & van Boom, J. H. (1990) *J. Mol. Biol.* **213,** 833–843.

24. Rougee, M., Faucon, B., Mergny, J. L., Barcelo, F., Giovannangeli, C., Garestier, T. & Helene, C. (1992) *Biochemistry* **31,** 9269–9278.
25. Xodo, L. E., Manzini, G. & Quadrifoglio, F. (1990) *Nucleic Acids Res.* **18,** 3557–3564.
26. Jin, R., Chapman, W. H. J., Srinivasan, A. R., Olson, W. K., Breslow, R. & Breslauer, K. J. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10568–10572.
27. Sun, J.-S., Francois, J.-C., Montenay-Garestier, T., Saison-Behmoaras, T., Roig, V., Thuong, N. T. & Helene, C. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9198–9202.
28. Mergny, J.-L., Sun, J.-S., Rougée, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J. & Helene, C. (1991) *Biochemistry* **30,** 9791–9798.
29. Sun, J. S., Giovannangeli, C., Francois, J. C., Kurfurst, R., Montenay-Garestier, T., Asseline, U., Saison-Behmoaras, T., Thuong, N. T. & Helene, C. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6023–6027.
30. Escude, C., Francois, J.-C., Sun, J.-S., Ott, G., Sprinzl, M., Garestier, T. & Helene, C. (1993) *Nucleic Acids Res.* **21,** 5547–5553.
31. Mergny, J.-L., Collier, D., Rougee, M., Montenay-Garestier, T. & Helene, C. (1991) *Nucleic Acids Res.* **19,** 1521–1526.
32. Mergny, J. L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougee, M., Montenay-Garestier, T., Bisangi, E. & Helene, C. (1992) *Science* **256,** 1681–1684.
33. Han, H. & Dervan, P. B. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 3806–3810.
34. Roberts, R. W. & Crothers, D. M. (1992) *Science* **258,** 1463–1466.
35. Singleton, S. F. & Dervan, P. B. (1992) *Biochemistry* **31,** 10995–11003.
36. Singleton, S. F. & Dervan, P. B. (1992) *J. Am. Chem. Soc.* **114,** 6957–6965.