Supporting Information for:

A Simple Molecular Model for Thermophilic Adaptation of Functional Nucleic Acids

Joshua M. Blose[*], Scott K. Silverman[†], and Philip C. Bevilacqua[*‡]

[*]Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802. [†]Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801, USA. [‡]Author to whom correspondence should be addressed; Tel: (814) 863–3812; Fax: (814) 863–8403; e-mail: pcb@chem.psu.edu

Contents:

1) Mass Spectrometry. Data were obtained on a Waters Micromass LCT Premier TOF mass spectrometer equipped with a Waters Alliance 2695 Separations Module and utilizing electrospray ionization (The Huck Institute of Life Sciences Proteomics and Mass Spectrometry Core Facility, The Pennsylvania State University).

	Sequence	Calculated MW(amu)	Measured MW (amu)
triplexes	GGGG	11862.7	11864
	AGGG	11861.7	11863
	AAGG	11860.7	11862
	AAAG	11859.7	11862
	AAAA	11858.7	11861
	G***G***GG	11877.7	11879
	GGGG	11892.7	11894
duplexes	GGGG	8257.4	8258
	AAAA	8253.4	8254

2) Native Gel Electrophoresis. In an effort to show that the DNA sequences in the triplex series form the same native structure, we carried out native gel electrophoresis experiments. In order to facilitate native folding of the triplexes, which contain several C^+•GC triples, we ran the gels in the absence of a chemical denaturant, at the low pH of 5.5, as well as the low temperature of 20 ^oC. The buffer in the electrophoresis apparatus reservoirs and the gel itself was 10 mM MES (pH 5.5). The gel was 15 % acrylamide (29:1 crosslinking), electrophoresis was for 3 h at 500 V, and the buffers were re-circulated approximately every 20 minutes.

The gel contained three sets of sequences: 1) GGGG and AAAA duplexes 2) All triplexes studied, and 3) Control triplex, which is the same as GGGG triplex but has its 3'-terminal extension changed such that it cannot form a triplex.

Control triplex: 5'-AGAGAGAGGGGG*TTTTT*CCCCTCTCTCT*TTTTT*CCTTCCC

As seen in Figure S1, the electrophoretic mobilities of the major band in the triplex lanes were identical to one another, and the mobilities of the major band in the duplex lanes were identical to each other. Moreover, the triplexes migrated slower than the duplexes, as expected given the greater bulk of the triplexes, while the control triplex migrated slower than the authentic triplexes as expected given its more open conformation. Had the DNAs in the triplex series not formed base triples, they would have been expected to co-migrate with the control triplex. These observations support the conclusion that the DNAs in the triplex series form essentially equivalent native structures.

3) Figures S1-S5

Figure S1. Native gel electrophoresis . Lanes 1-7: GGGG, AGGG, AAGG, AAAG, AAAA, G***G***GG, and ***GG***GG triplexes; Lane 8: control triplex; Lanes 9-10: GGGG and AAAA duplexes. The fold of each species is depicted on the right-hand side of the figure. The data are interpreted in Section 2 above. The depiction of the three states are adapted from Figure 1 in the text.

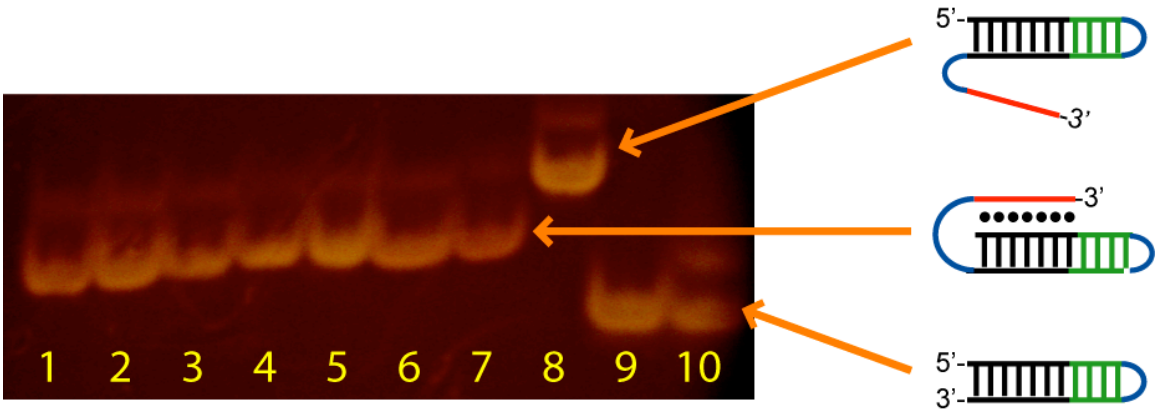


Figure S2. UV melting curves of triplexes. (A) UV melting curve of triplex GGGG at various pH values, corresponding to Figure 2A. (B) UV melting curves of triplexes at pH 7.0, corresponding to Figure 2B. (C) UV melting curves of triplexes at pH 5.5, corresponding to Figure 2C.

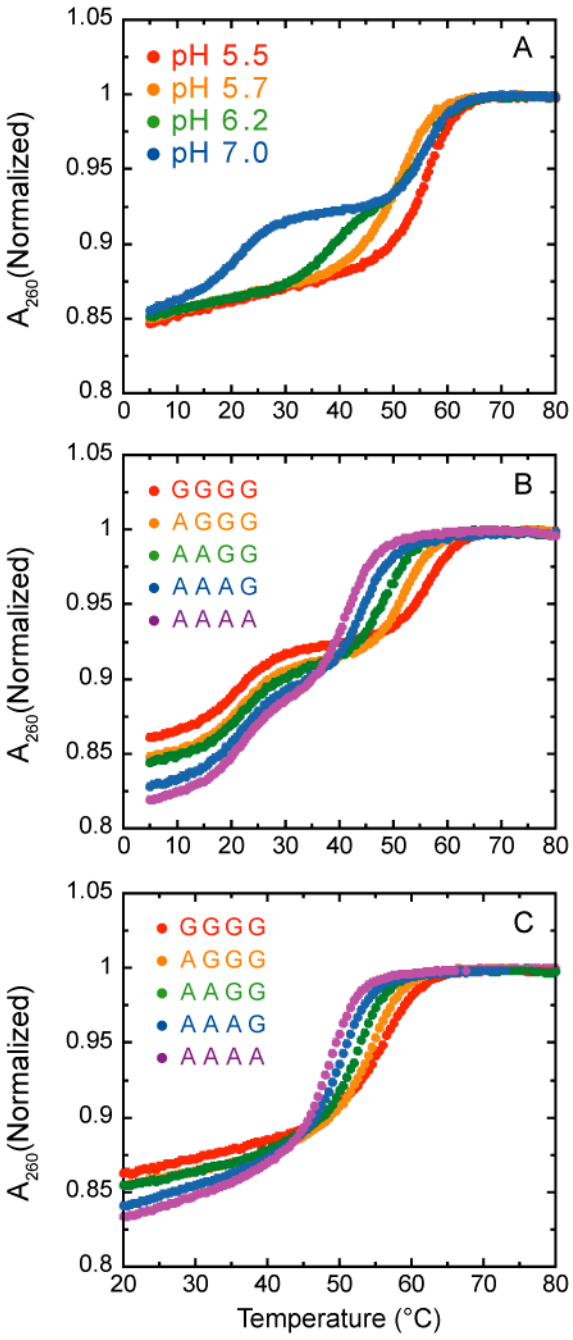


Figure S3. UV melting curves of control duplexes. (A) UV melting curves and (B) first derivative absorbance curves of control duplexes at pH 5.5 and 7.0. Both duplexes melt in a single transition (T_{M12}), which is assigned to the unfolding of the secondary structure.

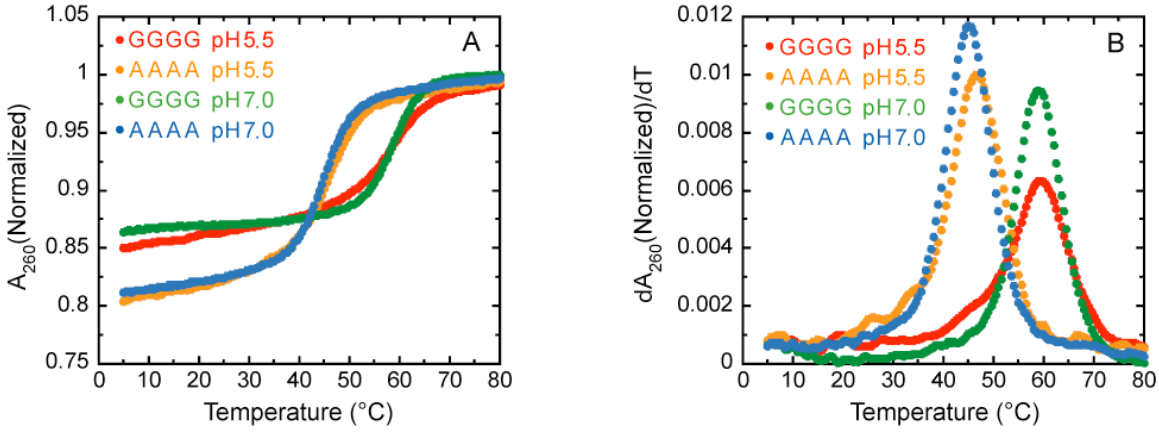


Figure S4. Melting curves of non-Watson-Crick triplexes where underlined Gs indicate a GT wobble pair. (A) UV melting curves and (B) first-derivative absorbance curves of triplex G***G***GG and triplex ***GG***GG at pH 5.5 and 7.0. As in the case of the Watson-Crick triplexes, triplex G***G***GG has two apparent transitions at pH 7.0, and each transition is assigned as for the Watson-Crick triplexes with the lower-temperature transition corresponding to unfolding of the triplex strand (T_{M23}) and the higher-temperature transition corresponding to the unfolding of the secondary structure (T_{M12}). For triplex ***GG***GG, it appears that this triplex unfolds in monophasic fashion in the UV melting profiles (A, blue curve) but derivative analysis reveals a second transition (B, blue curve). Thus, for triplex ***GG***GG the structural transitions overlap and have some cooperativity. At pH 5.5, however, both triplexes unfold in a highly cooperatively fashion with a T_{M13} that depends on base pair identity in the tunable region.

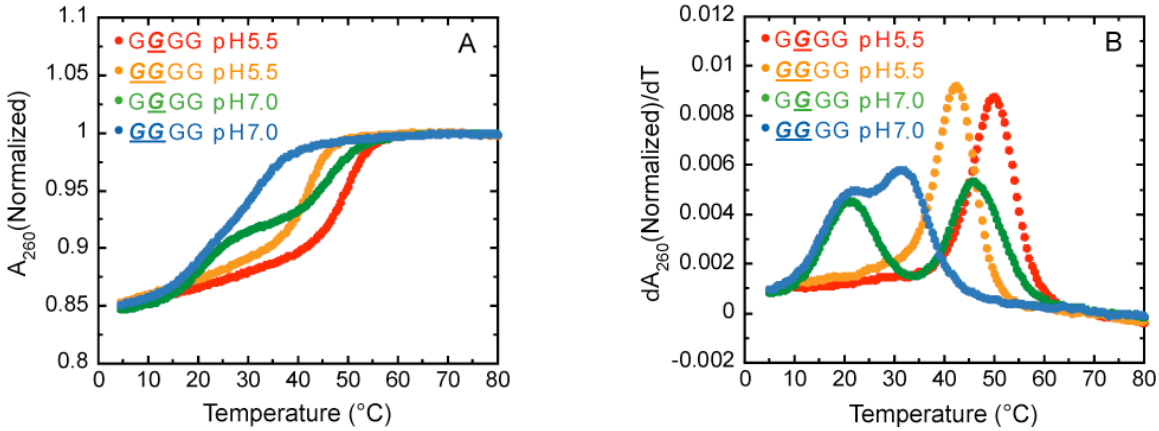
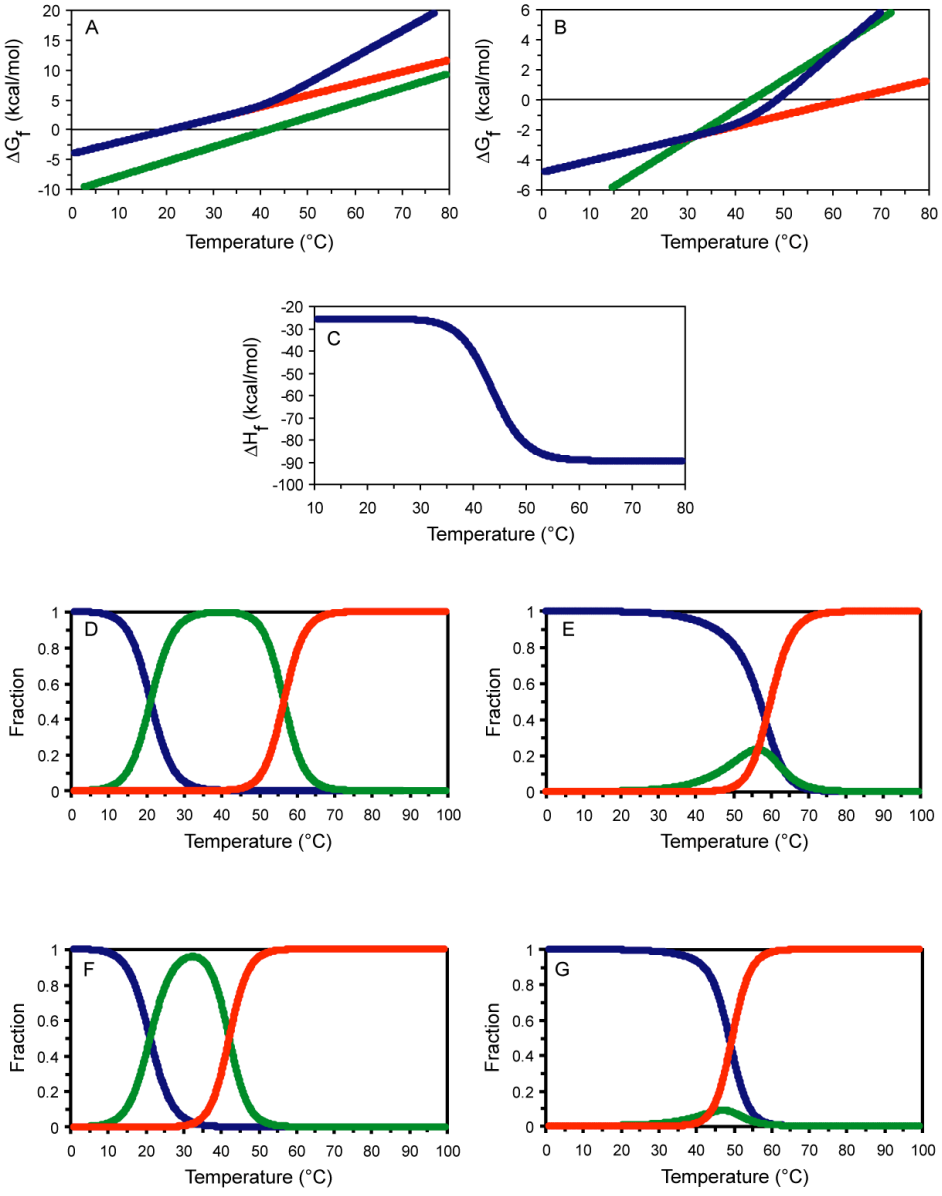


Figure S5. Thermodynamic simulations for triplexes AAAA and GGGG**.** (A) Piecewise linear analysis of AAAA at pH 7.0. The blue trace is of ΔG_f for AAAA, as in Figure 6A. The red plot is of ΔG_{23}, while the green plot is of ΔG_{12}. At low temperatures, $\Delta G_f = \Delta G_{23}$, while at high temperatures $\Delta G_f = \Delta G_{12} + \Delta G_{23}$. (B) Same as panel A but at pH 5.5, and blue trace corresponding to Figure 6B. (C) Simulation of ΔH_f versus temperature for AAAA at pH 5.5. At low temperatures, $\Delta H_f = \Delta H_{23}$ (at pH 5.5), while at high temperatures $\Delta H_f = \Delta H_{12} + \Delta H_{23}$. (D) Simulation of fractional population versus temperature for GGGG at pH 7.0. The blue, green, and red traces correspond to triplex, duplex, and unfolded states, respectively. (E) Same as panel D, but at pH 5.5. (F) Simulation of the fractional population versus temperature for AAAA at pH 7.0. The blue, green, and red traces correspond to triplex, duplex, and unfolded states, respectively. (G) Same as panel F, but at pH 5.5.



4) Tables S1 and S2

Table S1. Complete Thermodynamic Parameters for Triplex Formation at pH 7.0

Sequence	ΔH_{23} (kcal/mol)	ΔS_{23} (eu)	ΔG_{23} (kcal/mol)	$\Delta\Delta G_{23}$ (kcal/mol)	T_{M23} (TMT) (°C)	ΔT_{M23} (°C)	ΔH_{12} (kcal/mol)	ΔS_{12} (eu)	ΔG_{12} (kcal/mol)	$\Delta\Delta G_{12}$ (kcal/mol)	T_{M12} (°C)	ΔT_{M12} (°C)
Watson-Crick												
GGGG	-56.7 ± 1.0	-192.2 ± 3.3	5.43 ± 0.14	-	21.8 ± 0.5	-	-78.7 ± 1.8	-238.7 ± 5.4	-1.56 ± 0.09	-	56.5 ± 0.3	-
AGGG	-56.2 ± 1.3	-190.4 ± 4.6	5.31 ± 0.16	-0.11 ± 0.21	22.1 ± 0.2	0.3 ± 0.5	-78.4 ± 0.8	-240.9 ± 2.4	-0.60 ± 0.06	0.95 ± 0.11	52.5 ± 0.2	-4.0 ± 0.4
AAGG	-56.2 ± 1.1	-190.4 ± 3.7	5.32 ± 0.10	-0.10 ± 0.17	22.0 ± 0.4	0.3 ± 0.6	-76.8 ± 1.4	-238.3 ± 4.4	0.22 ± 0.05	1.78 ± 0.10	49.1 ± 0.2	-7.4 ± 0.4
AAAG	-56.3 ± 1.0	-190.9 ± 3.5	5.43 ± 0.13	0.01 ± 0.19	21.6 ± 0.5	-0.2 ± 0.7	-76.4 ± 1.7	-239.9 ± 5.2	1.17 ± 0.08	2.73 ± 0.12	45.1 ± 0.4	-11.4 ± 0.5
AAAA	-57.5 ± 0.9	-195.3 ± 2.9	5.64 ± 0.07	0.22 ± 0.16	21.1 ± 0.3	-0.7 ± 0.6	-77.3 ± 2.5	-245.2 ± 8.0	1.94 ± 0.11	3.49 ± 0.14	42.1 ± 0.3	-14.4 ± 0.4
Non Watson-Crick												
G<u>G</u>GG	-57.4 ± 1.4	-195.2 ± 4.7	5.65 ± 0.19	0.23 ± 0.24	21.0 ± 0.6	-0.7 ± 0.8	-65.4 ± 1.0	-204.1 ± 3.0	0.58 ± 0.09	2.13 ± 0.13	47.2 ± 0.5	-9.3 ± 0.6
<u>GG</u>GG	*Not a Clean Transition						*Not a Clean Transition					
Duplexes												
GGGG							-74.6 ± 1.6	-224.6 ± 4.6	-2.03 ± 0.11	-0.48 ± 0.14	59.0 ± 0.3	2.5 ± 0.4
AAAA[a]							-70.4 ± 0.7	-221.3 ± 2.2	1.11 ± 0.05	-0.83 ± 0.12	45.0 ± 0.3	2.9 ± 0.4

All melts were performed at 10 mM Na^+ as described in Materials and Methods. Values for ΔG are provided at 50 °C since this is closer to the T_{M12}. [a]The reference state for duplex AAAA is triplex AAAA.

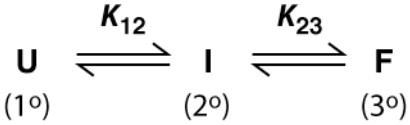
Table S2. Complete Thermodynamic Parameters for Triplex Formation at pH 5.5

Sequence	ΔH_{obs}[a] (kcal/mol)	ΔS_{obs} (eu)	ΔG_{obs} (kcal/mol)	$\Delta\Delta G_{obs}$ (kcal/mol)	T_{Mobs}(TMT) (°C)	ΔT_{Mobs} (°C)
Watson-Crick						
GGGG	-71.0 ± 1.4	-215.2 ± 4.4	-1.44 ± 0.06	-	56.7 ± 0.3	-
AGGG	-79.0 ± 3.5	-240.8 ± 10.8	-1.20 ± 0.09	0.24 ± 0.11	55.0 ± 0.5	-1.7 ± 0.5
AAGG	-85.6 ± 2.6	-262.3 ± 7.8	-0.80 ± 0.08	0.64 ± 0.10	53.0 ± 0.3	-3.6 ± 0.4
AAAG	-87.6 ± 0.6	-270.5 ± 2.1	-0.19 ± 0.09	1.25 ± 0.11	50.7 ± 0.3	-6.0 ± 0.5
AAAA	-89.3 ± 2.8	-277.1 ± 8.7	0.26 ± 0.08	1.70 ± 0.10	49.1 ± 0.3	-7.6 ± 0.4
Non Watson-Crick						
G***G***GG	-81.8 ± 1.7	-252.8 ± 5.1	-0.11 ± 0.11	1.32 ± 0.13	50.4 ± 0.4	-6.2 ± 0.5
GGGG	-87.0 ± 1.7	-275.4 ± 5.1	2.03 ± 0.05	3.46 ± 0.08	42.6 ± 0.3	-14.0 ± 0.4
Duplexes[b]						
GGGG	-55.2 ± 0.9	-166.1 ± 2.7	-1.50 ± 0.01	-0.07 ± 0.07	59.0 ± 0.1	2.4 ± 0.3
AAAA[c]	-56.8 ± 2.1	-178.1 ± 6.4	0.78 ± 0.04	0.52 ± 0.09	46.5 ± 0.3	-2.5 ± 0.4

All melts were performed at 10 mM Na^+ as described in Materials and Methods. Values for ΔG are provided at 50 °C since this is closer to T_{M13}. [a]The provided thermodynamic values are observed values ('obs') determined from fits to a two-state model, extrapolated to 50 °C since this is closer to the observed T_M. Because the observed values are for loss of tertiary structure, they are used as an operational definition of ΔG_f in the text and in Figure 4. ΔH_{obs}, ΔS_{obs}, ΔG_{obs} and $T_{M\,obs}$ values are approximately equal to ΔH_{13}, ΔS_{13}, ΔG_{13} and $T_{M\,13}$ under the cooperative conditions of pH 5.5, as described in the text. [b]For duplexes the values provided are for the 1^o to 2^o structure transition. [c]The reference state for duplex AAAA is triplex AAAA.

5) Procedure for Simulations

Background. Simulations were performed to gain insight into the thermodynamic behavior under cooperative and non-cooperative folding conditions. This section describes the mathematical definition of functional stability, the model used for the simulations, the equations derived from it, the method of the simulation, and the input thermodynamic parameters. All simulations focus either on AAAA or GGGG because these triplexes test the limits of secondary structure strength and illustrate the trends most clearly; other sequences showed intermediate behavior (not shown). We performed simulations of the three-state folding model in Figure 1 of the main text, using the measured thermodynamic parameters presented in Tables S1 and S2. Here, we denote the triplex simply as 'F' (for functional), the secondary structure as 'I' (for intermediate), and the random coil unfolded state as 'U'. This leads to Scheme S1, in which K_{12}=[I]/[U] is the intrinsic equilibrium constant for secondary structure formation from random coil, and K_{23}=[F]/[I] is the intrinsic equilibrium constant for tertiary structure formation from secondary structure. The parameters ΔG_{12}, ΔH_{12}, ΔS_{12}, and T_{M12} are associated with the first step, and ΔG_{23}, ΔH_{23}, and ΔS_{23} and T_{M23} are associated with the second step.



Scheme S1

Equations for the simulations were derived starting from the definition of *functional stability* advanced by Sosnick, Pan and co-workers (1). They define functional stability as the free energy difference between the functional state and the penultimately stable, non-functional state. The beauty of this definition is that it defines the population of the functional state relative to the next most stable state, whatever it may be. This definition is relevant to biology because only the functional state can give rise to biological function. Intriguingly, in a non-two-state system such as triplex unfolding, the identity of the penultimately stable state (reference state) can change with temperature. In our study, the triplex state is used to mimic the functional state of a nucleic acid, which typically needs tertiary structure in order to function.

We transformed the Sosnick and Pan definition of functional stability into a mathematical equation. We define a functional stability constant (K_f) as the concentration of the functional state, [F], relative to the sum of the concentration of all other states, which are [I] and [U] in the case of the triplex.[1]

$$K_f = \frac{[F]}{[I]+[U]} \quad (1)$$

[1] Note that this equation can also be arrived at from f_F=[F]/([F]+[I]+[U]) and dividing the numerator and denominator by a 'reference state' of ([I] + [U]) to give a standard relationship that is similar to the relationship between *f* and *K* for a two-state system: $f_F=K_f/(K_f+1)$.

Choosing the unfolded state as the reference state gives

$$K_f = \frac{[F]/[U]}{[I]/[U]+1} \tag{2}$$

Substituting, we obtain an expression for K_f in terms of the intrinsic constants for secondary and tertiary structure formation.

$$K_f = \frac{K_{12}K_{23}}{K_{12}+1} \tag{3}$$

This equation has two limits. When $K_{12}<<1$ (*i.e.* where folding is cooperative, which occurs at pH 5.5 for the triplex), $K_f = K_{12}K_{23}$, and functional stability constant is the same as the overall equilibrium constant between the random coil and functional state. When $K_{12}>>1$, $K_f = K_{23}$, and the functional stability is the tertiary stability.

The dependencies of K_{12} and K_{23} on temperature are given by the standard van't Hoff relationships, in which T_M and ΔH for a given step were determined directly (at pH 7.0) or indirectly (at pH 5.5) from UV melting experiments, Tables S1 and S2 respectively.

$$K_{12} = \frac{[I]}{[U]} = \exp\left[\frac{\Delta H_{12}}{R}\left(\frac{1}{T_{M12}} - \frac{1}{T}\right)\right] \tag{4}$$

$$K_{23} = \frac{[F]}{[I]} = \exp\left[\frac{\Delta H_{23}}{R}\left(\frac{1}{T_{M23}} - \frac{1}{T}\right)\right] \tag{5}$$

where R is the gas constant, T is temperature in kelvins, and T_M is the melting temperature in kelvins. The functional free energy (ΔG_f) and functional enthalpy (ΔH_f) were calculated using the standard thermodynamic relationships.

$$\Delta G_f = -RT\ln K_f \tag{6}$$

$$\Delta H_f = -R\frac{\partial \ln K_f}{\partial 1/T} \tag{7}$$

Similar to K_f, when $K_{12}<<1$ ΔG_f has a limit of $\Delta G_{12}+\Delta G_{23} = \Delta G_{13}$, and when $K_{12}>>1$ ΔG_f has a limit of ΔG_{23}. Similar limits apply to ΔH_f.

For the pH 7.0 simulations, the inputs were ΔH_{12}, ΔH_{23}, T_{M12}, and T_{M23} from the pH 7.0 melts of AAAA or GGGG. In order to perform the simulations, K_{12} and K_{23} were first calculated from the enthalpy and T_M inputs according to eqs 4 and 5; this was done at temperatures ranging from 273 to 373 K, with a point every 0.1 K. Next, K_f was calculated from K_{12} and K_{23} according to eq 3. Then, ΔG_f was calculated from the K_f

according to eq 6, and ΔH_f was calculated from $\ln K_f$ according to eq 7; in the latter case, the derivative was taken numerically rather than analytically. All plots were made using Excel (Microsoft).

For the pH 5.5 simulations, the inputs were ΔH_{12}, ΔH_{obs}, T_{M12}, and T_{Mobs}, where 'obs' means the observed parameter for tertiary structure melting, which does not correspond to either of the intrinsic parameters. The parameters of ΔH_{12} and T_{M12} are for secondary structure formation at pH 5.5; for these parameters we used ΔH_{12} and T_{M12} from melts of the core duplexes at pH 5.5 (Table S2),[2] with small corrections of $\Delta\Delta H_{12}$ and ΔT_{M12} from pH 7.0 melts, since comparison of the transition for core duplexes and the second transition of the triplexes at pH 7.0 revealed small offsets (Table S1). Because folding is cooperative at pH 5.5,

$$\Delta H_{obs} = \Delta H_{12} + \Delta H_{23} \quad \text{(pH 5.5 only)} \tag{8}$$

which allows ΔH_{23} to be calculated by subtraction. Likewise, it can be shown that

$$T_{Mobs} = (\Delta H_{12} + \Delta H_{23})/(\Delta S_{12} + \Delta S_{23}) \quad \text{(pH 5.5 only)} \tag{9}$$

which allows ΔS_{23} to be calculated, as well as T_{M23} using $T_{M23} = \Delta H_{23}/\Delta S_{23}$. Eq 9 is in agreement with the observation by Laing and Draper that the T_Ms observed in a melt of a system with a set of coupled transitions do not have to be the T_Ms of the individual transitions (2). Once ΔH_{12}, ΔH_{23}, T_{M12}, and T_{M23} at pH 5.5 were calculated using these methods, remaining functional thermodynamic parameters were simulated in the same fashion as at pH 7.0. For the low pH simulations, ΔH_{23} and T_{M23} from the AAAA triplex were used for the GGGG triplex simulations since these parameters were shown to be identical within experimental error at higher pH (Table S1) and could not be determined directly at lower pH (Table S2). For consistency, the same was done for the higher pH simulations.

See the main text for a discussion of the simulations.

5) References

(1) Fang, X. W., Golden, B. L., Littrell, K., Shelton, V., Thiyagarajan, P., Pan, T., and Sosnick, T. R. (2001) The thermodynamic origin of the stability of a thermophilic ribozyme. *Proc. Natl. Acad. Sci. U S A 98*, 4355-4360.

(2) Laing, L. G., Gluick, T. C., and Draper, D. E. (1994) Stabilization of RNA structure by Mg ions. Specific and non-specific effects. *J. Mol. Biol. 237*, 577-587.

[2] It can be noted that the thermodynamic parameters for duplex melting are mildly pH dependent (Tables S1 and S2). This may be because of protonation events in the unfolded state (2).