VOLUME 123, NUMBER 7
FEBRUARY 21, 2001

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Stability and Structure of RNA Duplexes Containing Isoguanosine and Isocytidine

Xiaoying Chen, Ryszard Kierzek,[§] and Douglas H. Turner*

Contribution from the Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, and Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-714 Poznan, Noskowskiego 12/14, Poland

Received July 17, 2000

Abstract: Isoguanosine (iG) and isocytidine (iC) differ from guanosine (G) and cytidine (C), respectively, in that the amino and carbonyl groups are transposed. The thermodynamic properties of a set of iG, iC containing RNA duplexes have been measured by UV optical melting. It is found that iG−iC replacements usually stabilize duplexes, and the stabilization per iG−iC pair is sequence-dependent. The sequence dependence can be fit to a nearest-neighbor model in which the stabilities of iG−iC pairs depend on the adjacent iG−iC or G−C pairs. For 5′-CG-3′/3′-GC-5′ and 5′-GG-3′/3′-CC-5′ nearest neighbors, the free energy differences upon iG−iC replacement are smaller than 0.2 kcal/mol at 37 °C, regardless of the number of replacements. For 5′-GC-3′/3′-CG-5′, however, each iG−iC replacement adds 0.6 kcal/mol stabilizing free energy at 37 °C. Stacking propensities of iG and iC as unpaired nucleotides at the end of a duplex are similar to those of G and C. An NMR structure is reported for r(CiGCGiCG)$_2$ and found to belong to the A-form family. The structure has substantial deviations from standard A-form but is similar to published NMR and/or crystal structures for r(CGCGCG)$_2$ and 2′-*O*-methyl (CGCGCG)$_2$. These results provide benchmarks for theoretical calculations aimed at understanding the fundamental physical basis for the thermodynamic stabilities of nucleic acid duplexes.

Introduction

The structural features of nucleic acids are becoming increasingly important as the roles of nucleic acids in cellular pathways are being discovered,[1−4] as RNA and DNA become targets for therapeutics,[3,5−7] and as nucleic acids become components for design of self-assembling nanostructures.[8−11] The interactions determining nucleic acid structures, however, are not completely

* To whom correspondence should be addressed.
§ Polish Academy of Sciences.

(1) Watson, J. D.; Hopkins, N. H.; Roberts, J. W.; Steitz, J. A.; Weiner, A. M. *Molecular Biology of the Gene*; Benjamin Cummings Inc.; Menlo Park, CA, 1987.
(2) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA World*, 2nd ed.; Cold Spring Harbor Press: Cold Spring Harbor, NY, 1999.
(3) Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. *Science* **1996**, 274, 1367−1371.
(4) Cech, T. R. *Annu. Rev. Biochem.* **1990**, 59, 543−568.
(5) Crooke, S. T. *Handbook of Experimental Pharmacology*; Antisense Research and Application, Vol. 131; Springer, New York, 1998.
(6) Testa, S. M.; Gryaznov, S. M.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 2734−2739.
(7) Dervan P. B.; Burli, R. W. *Curr. Opin. Chem. Biol.* **1999**, 3; 688−693.
(8) Letsinger, R. L. *Nucleosides Nucleotides* **1998**, *17*, 1861−1869.
(9) Seeman, N. C. *Acc. Chem. Res.* **1997**, *30*, 357−363.
(10) Seeman, N. C.; Qi, J.; Li, X.; Tang, X.; Leontis, N. B.; Liu, B.; Zhang, Y.; Du, S. M.; Chen, J. *The control of DNA structure: From topological modules to geometrical modules, Modular Chemistry*; Klumer, 1997; pp 95−104.
(11) Benner, S. A.; Battersby, T. R.; Eschghaller, B.; Hutter, D.; Kodra, J. T.; Lutz, S.; Arslan, T.; Baschlin, D. K.; Blattler, M.; Egli, M.; Hammer, C.; Held, H. A.; Horlacher, J.; Huang, Z.; Hyrup, B.; Jenny, T. F.; Jurczyk, S. C.; Konig, M.; van Krosigk, U.; Lutz, M. J.; Macpherson, L. J.; Moroney, S. E.; Muller, E.; Nambiar, K. P.; Piccirilli, J. A.; Switzer, C. Y.; Vogel, J. J.; Richert, C.; Roughton, A. L.; Schmidt, J.; Schweider, K. C.; Stackhouse, J. *Pure Appl. Chem.* **1998**, 70, 263−266.

10.1021/ja002623i CCC: $20.00
Published on Web 01/27/2001

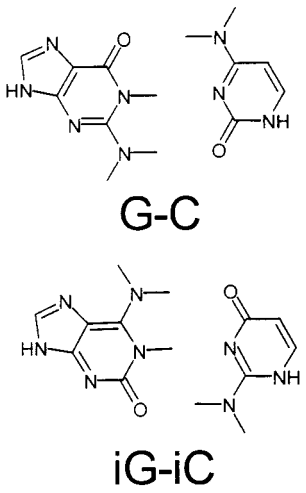

Figure 1. Schematic representation of iG−iC and G−C pairs.

understood.[12−20] A full elucidation of the sequence dependence of such interactions is prohibited by the lack of variety among naturally occurring bases. One way around this limitation is to use unnatural bases while as much as possible maintaining the RNA or DNA structure. Examples include the $\kappa-\pi$ pair,[21] Z−F pair,[22,23] the inosine−cytosine pair,[24,25] the 2,6-diaminopurine−thymine pair,[26] and the isoguanine−isocytosine (iG−iC) pair.[27−29] Here we focus on the iG−iC pair in RNA duplexes.

In the iG−iC pair, the amino and carbonyl groups in the regular nucleic acid bases G and C are transposed (Figure 1). Isoguanine and its derivatives have been isolated from natural sources and studied for a long time,[30−35] while iC is not known to occur naturally. Advances in chemical and enzymatic synthesis of oligonucleotides, however, have allowed the iG−iC pair to be used to study nucleic acid interactions,[29a,36] and to expand the codon−anticodon code.[37] Because of the rigidity of RNA duplexes and similarity between iG−iC and G−C pairs, experimental comparisons between iG−iC and G−C pairs in RNA provide a relatively simple benchmark for testing computations and therefore furthering our understanding of the interactions that shape nucleic acids.

Results

Thermodynamics of iG−iC Pairs. Thermodynamic parameters for regular RNA duplexes and those with iG−iC pairs replacing G−C pairs were measured by UV melting. Typical van't Hoff plots are shown in Supporting Information, and the results are listed in Table 1. With the exceptions of r(iCCGiG)$_2$ and r(iCCAUGiG)$_2$, the duplexes with iG−iC pairs are more stable than those with only G−C pairs. For example, $\Delta G°_{37}$ values for formation of r(iGiCiGiC)$_2$ and r(GCGC)$_2$ are −6.97 and −4.61 kcal/mol, respectively, and the T_m's at 10^{-4} M strand concentration are 45.2 and 26.6 °C, respectively. On average, the stability enhancement is 0.44 kcal/mol of iG−iC pair, while the largest change per iG−iC substitution is 0.68 kcal/mol for r(GGiCiGCC)$_2$.

The data in Table 1 for duplexes with iG−iC pairs can be fit to various nearest-neighbor models[38−41] that are able to predict duplex stability from sequence. Table 2 presents parameters for the individual nearest-neighbor-hydrogen bonding (INN-HB) model.[40] The INN-HB model includes a term for terminal pairs that are not G−C.[40] This term effectively accounts for base composition since two duplexes can have the same nearest neighbors, but different base compositions if the terminal base pairs are different types. Thus, the model assumes that the sequence dependence of duplex stability depends only on the number of each type of base pair and the interactions of adjacent base pairs. The data were fit in two ways by linear regression:[42] (1) Only parameters for nearest neighbors with at least one iG−iC pair were allowed to vary, while other parameters were taken from Xia et al.[40] (2) The data for duplexes with iG−iC pairs were combined with previous data for duplexes with only G−C pairs, and the combined data set was fit. Parameters from methods (1) and (2) are listed, respectively, without and with parentheses in Table 2. The two ways of analyzing the data give similar results, suggesting iG−iC pairs generally fit into the RNA backbone.

(12) (a) Turner, D. H. In *Nucleic Acids: Structures, Properties, and Functions*; Bloomfield, V. A., Crothers, D. M., Tinoco, I., Jr., Eds.; University Science Books; Sausalito, CA, 2000; Chapter 8, pp 259−334. (b) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167−192.

(13) Jorgensen W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008−2010.

(14) Newcomb L. F.; Gellman, S. H. *J. Am. Chem. Soc.* **1994**, *116*, 4993−4994.

(15) Friedman, R. A.; Honig, B. *Biophys. J.* **1995**, *69*, 1528−1535.

(16) Gellman, S. H.; Haque, T. S.; Newcomb, L. F. *Biophys. J.* **1996**, *71*, 3523−3525.

(17) Friedman, R. A.; Honig, B. *Biophys. J.* **1996**, *71*, 3525−3526.

(18) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134−149.

(19) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Scheils, C. J.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 2213−2222.

(20) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Intl. Ed.* **2000**, *39*, 991−1009.

(21) Piccirilli, J. A.; Krauch, T.; Moroney, S.; Benner, S. *Nature* **1991**, *343*, 33−37.

(22) Morales, J. C.; Kool, E. T. *Nat. Struct. Bio.* **1998**, *5*, 950−954.

(23) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Bio.* **1998**, *5*, 954−959.

(24) Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I., Jr. *Nucleic Acids Res.* **1985**, *13*, 8927−8938.

(25) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 3783−3785.

(26) Gaffney, B. L.; Marky, L. A.; Jones, R. A. *Nucleic Acids Res.* **1982**, *10*, 4351−4361.

(27) Abbreviations used in this paper: iG, isoguanosine; iC, isocytidine; G, guanosine; C, cytidine; eu, entropy units (cal K^{-1} mol^{-1}); T_M, melting temperature in kelvin; T_m, melting temperature in °C; C_T, total strand concentration; INN-HB, individual nearest-neighbor-hydrogen bonding; FID, free induction decay; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; DQFCOSY, double quantum filtered correlation spectroscopy; HETCOR, heteronuclear correlation spectroscopy; UV, Ultraviolet; rmsd, root mean squared deviation.

(28) Rich, A. In *Horizons in Biochemistry*; Kasha, M., Pullmann, B., Eds.; Academic Press: New York, 1962; pp 103−126.

(29) (a) Roberts, C.; Bandaru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640−4649. (b) Jurczyk, S. C.; Kim, Y. H.; Nachman, R. J.; Pavelka, L.; Mosher, H. S. *J. Nat. Prod.* **1981**, *44*, 206−214.

(30) Cherbuliez, E.; Bernhard, K. *Helv. Chim. Acta* **1932**, *15*, 464−471.

(31) Fuhrman, F. A.; Fuhrman, G. J.; Nachman, R. J.; Mosher, H. S. *Science* **1981**, *212*, 557−558.

(32) Pettit, G. R.; Ode, R. H.; Coomes, M. R.; Ode, S. L. *Lloydia* **1976**, *39*, 363.

(33) Quinn, R. J.; Gregson, R. P.; Cook, A. F.; Bartlett, R. T. *Tetrahedron Lett.* **1980**, *21*, 567−568.

(34) Kim, Y. H.; Nachman, R. J.; Pavelka, L.; Mosher, H. S. *J. Nat. Prod.* **1981**, *44*, 206−214.

(35) Bergmann, W.; Feeney, R. J. *J. Am. Chem. Soc.* **1950**, *72*, 2809−1810.

(36) Switzer, C. Y.; Moroney, S. E.; Benner, S. A. *Biochemistry* **1993**, *32*, 10489−10496.

(37) Bain, J. D.; Switzer, C. Y.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537−539.

(38) (a) Borer, P. N.; Dengler, B.; Tinoco, I., Jr.; Uhlenbeck, O. C. *J. Mol. Biol.* **1974**, *86*, 843−853. (b) Gralla, J.; Crothers, D. M. *J. Mol. Biol.* **1973**, *73*, 497−511.

(39) Gray, D. M. *Biopolymers* **1997**, *42*, 783−793.

(40) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner D. H. *Biochemistry* **1998**, *37*, 14719−14735.

(41) Goldstein, R.; Benight, A. *Biopolymers* **1992**, *32*, 1679−1693.

(42) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: New York, 1992.

Table 1. Thermodynamics of Formation of RNA Duplexes in 1 M NaCl[a]

	$1/T_m$ vs $\ln(C_T)$ parameters			
sequence (5′→3′)	$-\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	T_m[b] (°C)
Duplexes with iG−iC Pairs				
iCGCiG	4.39 ± 0.09	35.92 ± 1.79	101.7 ± 6.0	26.2
CiGiCG	4.81 ± 0.05	37.82 ± 1.61	106.4 ± 5.3	30.1
iCiGiCiG	5.64 ± 0.02	43.31 ± 1.47	121.5 ± 4.8	36.7
iGCGiC	5.70 ± 0.03	36.84 ± 1.47	100.4 ± 4.8	37.2
GiCiGC	5.38 ± 0.09	40.39 ± 2.47	112.9 ± 8.2	34.7
iGiCiGiC	6.97 ± 0.14	50.59 ± 5.35	140.6 ± 17.1	45.2
iCCGiG	4.47 ± 0.06	38.71 ± 1.83	110.4 ± 6.1	27.6
CiCiGG	5.13 ± 0.04	37.99 ± 1.64	105.9 ± 5.4	32.6
iCiCiGiG	4.83 ± 0.03	39.83 ± 1.09	112.9 ± 3.6	30.5
iGGCiC	5.68 ± 0.05	35.38 ± 2.47	108.7 ± 8.1	37.0
GiGiCC	6.26 ± 0.09	41.55 ± 3.53	113.8 ± 11.4	41.4
iGiGiCiC	7.17 ± 0.25	45.48 ± 5.48	123.5 ± 17.3	47.5
iCCAUGiG	7.20 ± 0.01	59.92 ± 1.28	170.0 ± 4.1	45.1
CiGCGiCG	10.27 ± 0.09	60.36 ± 1.20	161.5 ± 3.6	62.5
iCGGCCiG	10.52 ± 0.08	59.84 ± 1.14	159.0 ± 3.4	64.3
iGGCGCiC	12.48 ± 0.39	77.39 ± 4.58	209.3 ± 13.5	66.9
GGiCiGCC	12.69 ± 0.30	75.07 ± 3.65	201.1 ± 10.8	68.9
GCiGiCGC	12.56 ± 0.15	74.98 ± 1.68	201.3 ± 4.9	68.3
iGCCGGiC	13.19 ± 0.53	73.74 ± 5.54	195.2 ± 16.2	72.2
iGCCAUGGiC	16.47 ± 0.62	97.40 ± 5.95	261.0 ± 17.2	75.6
Duplexes with only Watson−Crick Pairs				
CGCG[c]	3.66	33.31	95.6	19.3
GCGC[c]	4.61	30.48	83.4	26.6
CCGG[c]	4.55	34.21	95.6	27.2
GGCC[c]	5.37	35.79	98.1	34.3
CCAUGG[c]	7.30	56.93	159.9	46.4
CGCGCG[d]	9.45 ± 0.05	58.40 ± 0.84	157.8 ± 2.5	58.5
GGCGCCp[c]	11.33	67.78	182.0	65.2
GCGCGCp[c]	10.62	65.98	178.5	62.1
CGGCCG[d]	10.19 ± 0.10	59.71 ± 1.39	159.7 ± 4.2	62.4
GCCGGC[d]	11.91 ± 0.09	70.74 ± 1.08	189.7 ± 3.2	67.0
GCCAUGGC	15.06 ± 0.18	93.91 ± 1.97	254.2 ± 5.8	71.4

[a] Thermodynamic parameters derived from fitting of melting curves are available in Supporting Information. The ΔH°'s derived from T_M^{-1} vs $\ln(C_T)$ plots and from curve fitting agree within 15%, consistent with the two-state approximation. [b] Calculated for an oligonucleotide strand concentration of 10^{-4} M. [c] Data from ref 40. [d] Values measured for these sequences are similar to those measured previously for the same sequences with 3′-terminal phosphates.[40]

As mentioned above, the iG−iC pair is usually more stable than the G−C pair. This trend is reflected in the nearest-neighbor parameters of Table 2. Except for 5′GiG3′/3′CiC5′, free energy increments for nearest neighbors with at least one iG−iC pair are essentially equivalent or more favorable than those for nearest neighbors with two G−C pairs. Note that isomeric substitutions make a substantially larger difference in the context 5′GC3′/3′CG5′ than in 5′CG3′/3′GC5′ or 5′GG3′/3′CC5′. That is, each iG−iC substitution in 5′GC3′/3′CG5′ enhances duplex stability by 0.6 kcal/mol, whereas single and double substitutions in 5′CG/3′GC5′ and 5′GG3′/3′CC5′ change stability by ≤0.2 kcal/mol (Table 2). Thermodynamics for the DNA-RNA hybrids d(5′AG_3AGiGGA_3G3′)/r(3′UC_3UCiCCU_3C5′) and d(5′CT_3iGT_3G3′)/r(3′GA_3iCA_3C5′) have been measured and compared to results with the iG−iC pair replaced by a G−C pair.[29a] The results for the first sequence can be compared with expectations for RNA−RNA duplexes since the iG−iC pair is flanked by G−C pairs. The substitution makes little difference in stability of the DNA−RNA hybrid,[29a] which is consistent with the expected effect in an RNA−RNA duplex (Table 2).

Stacking of bases is one interaction that contributes to duplex stability. An experimental measure of stacking propensity is provided by adding an unpaired nucleotide to each end of a duplex and then determining the increase in stability.[12,43] Table 3 lists thermodynamic parameters measured for such duplexes with "dangling ends". Table 4 lists the nearest-neighbor parameters for the unpaired nucleotides. Also shown in paren-

Table 2. Nearest-Neighbor Parameters for RNA Duplexes Containing iG−iC and G−C Pairs in 1 M NaCl[a]

	ΔG°_{37} (kcal/mol)	ΔH° (kcal/mol)	ΔS° (eu)
5′CG3′	−2.36[b]	−10.64[b]	−26.7[b]
3′GC5′	(−2.44 ± 0.10)	(−11.83 ± 1.82)	(−30.53 ± 5.54)
5′iCG3′	−2.46 ± 0.08	−10.80 ± 1.12	−27.01 ± 3.13
3′iGC5′	(−2.60 ± 0.10)	(−13.07 ± 1.96)	(−33.86 ± 5.96)
5′iCiG3′	−2.45 ± 0.17	−12.69 ± 2.23	−33.28 ± 6.29
3′iGiC5′	(−2.67 ± 0.14)	(−15.35 ± 2.76)	(−40.82 ± 8.49)
5′GG3′	−3.26[b]	−13.39[b]	−32.7[b]
3′CC5′	(−3.40 ± 0.08)	(−15.72 ± 1.56)	(−39.20 ± 4.72)
5′iGG3′	−3.46 ± 0.11	−14.94 ± 1.44	−36.80 ± 4.04
3′iCC5′	(−3.52 ± 0.12)	(−16.36 ± 2.27)	(−41.44 ± 6.96)
5′GiG3′	−3.07 ± 0.11	−14.67 ± 1.48	−37.34 ± 4.20
3′CiC5′	(−3.18 ± 0.12)	(−17.27 ± 2.41)	(−44.87 ± 7.43)
5′iGiG3′	−3.30 ± 0.17	−14.01 ± 2.26	−34.58 ± 6.35
3′iCiC5′	(−3.38 ± 0.13)	(−15.80 ± 2.55)	(−39.85 ± 7.85)
5′GC3′	−3.42[b]	−14.88[b]	−36.9[b]
3′CG5′	(−3.47 ± 0.10)	(−16.05 ± 1.87)	(−39.95 ± 5.71)
5′iGC3′	−4.00 ± 0.09	−16.90 ± 1.19	−41.32 ± 3.33
3′iCG5′	(−3.89 ± 0.10)	(−17.55 ± 1.95)	(−43.94 ± 5.97)
5′iGiC3′	−4.61 ± 0.17	−19.98 ± 2.31	−49.36 ± 6.53
3′iCiG5′	(−4.56 ± 0.13)	(−20.78 ± 2.72)	(−52.26 ± 8.40)
per terminal iG−iC[c]	−0.19 ± 0.07	−1.00 ± 0.98	−2.54 ± 2.77
	(−0.18 ± 0.06)	(−1.65 ± 1.17)	(−4.79 ± 3.61)
init	4.09[b]	3.61[b]	−1.5[b]
	(4.38 ± 0.24)	(10.34 ± 4.74)	(18.98 ± 14.5)
Q	0.009(0.19)	0.26(0.99)	0.19(0.99)
rmsd	0.53(0.44)	5.25(4.69)	15.71(14.26)
ave % of deviation	6.5(5.4)	9.8(8.8)	10.8(9.8)

[a] Values not in parentheses are from a linear regression on duplexes containing iG−iC pairs, but with parameters for nearest neighbors without iG−iC pairs fitted at values determined by Xia et al.[40] The value of *Q*, a statistical measure of randomness in the difference between measured and predicted stabilities is 0.009 for this fit of ΔG°_{37}. The average deviations between predicted and measured thermodynamic parameters are 6.5, 9.8, and 10.8% for ΔG°_{37}, ΔH° and ΔS°, respectively (see Supporting Information for comparisons of predicted and measured values). Values in parentheses are from a regression including not only duplexes with iG−iC pairs, but also those RNA duplexes containing only G−C pairs and listed in the database of ref 40. The *Q* value for this fit of ΔG°_{37} is 0.19. For both ways, the parameters of Xia et al.[40] were used to subtract the contributions from nearest neighbors involving A−U pairs in $(iCCAUGiG)_2$ and $(iGCCAUGGiC)_2$ before regression. The two data sets were also fit in the absence of the terminal iG−iC term. These fittings gave *Q* values of 0.0016 and 0.0032, respectively, with fixed and floating values of ΔG°_{37} parameters not involving iG−iC. Thus the terminal iG−iC term is statistically significant for ΔG°_{37}. [b] The values are from the individual nearest-neighbor-hydrogen bonding (INN-HB) model applied to duplexes containing only Watson−Crick pairs.[40] [c] This is a term for each terminal iG−iC pair.

Table 3. Thermodynamics of Formation for RNA Duplexes with Dangling iG or iC Nucleotides in 1 M NaCl[a]

	$1/T_M$ vs $\ln(C_T)$ parameters			
sequence (5′→3′)	$-\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	T_m[b] (°C)
iGCCGG	5.72 ± 0.04	32.71 ± 1.58	87.0 ± 5.1	37.4
GGCCiG	8.66 ± 0.09	53.98 ± 1.63	146.1 ± 5.0	55.2
CCGGiC	5.68 ± 0.02	37.74 ± 0.96	103.4 ± 3.1	37.0
GCGCiC	6.18 ± 0.03	44.11 ± 1.87	122.3 ± 6.0	40.6
CCAUGGiC	8.74 ± 0.05	66.63 ± 1.27	186.7 ± 3.9	51.9

[a] Thermodynamic parameters derived from fitting of melting curves are available in Supporting Information. The ΔH°'s derived from T_M^{-1} vs $\ln(C_T)$ plots and from curve fitting agree within 15%, consistent with the two-state approximation, except for iGCCGG where the difference is 18%. [b] Calculated for an oligomer concentration of 10^{-4} M.

Table 4. Terminal Nearest-Neighbor Parameters in kcal/mol at 37 °C for iG or iC Dangling Ends, and in Parenthesis for G or C Dangling Ends

sequence (5′→3′)	nearest neighbor	$\Delta G_{37}°$[a]
iGCCGG	5′iGC3′	−0.6 (−0.2)
GGCCiG	5′CiG3′	−1.65 (−1.7)
CCGGiC	5′GiC3′	−0.6 (−0.4)
GCGCiC	5′CiC3′	−0.8 (−0.8)
CCAUGGiC	5′GiC3′	−0.7 (−0.4)

[a] Nearest-neighbor parameter was obtained as shown in the following example: $\Delta G_{37}°$ (5′iGC3′/3′G5′) = [$\Delta G_{37}°$(iGCCGG) − $\Delta G_{37}°$(CCGG)]/2. Values in parentheses are for G or C dangling ends.[12] Unpaired nucleotides are underlined.

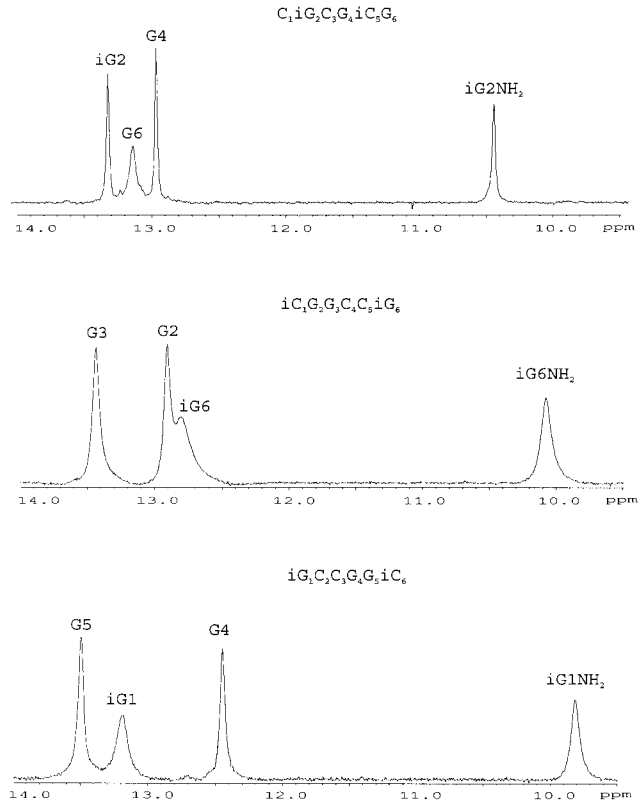

Figure 2. 1D imino proton spectra for r(CiGCGiCG)$_2$, r(iCGGCCiG)$_2$, and r(iGCCGGiC)$_2$ at 0 °C. Note the resonances at ca. 10 ppm.

theses in Table 4 are parameters previously reported for terminal unpaired G and C nucleotides.[12] Evidently, transposing amino and carbonyl groups to form the isomeric bases only modestly affects stacking.

Imino Proton NMR Spectra. Exchangeable proton NMR spectra for r(CiGCGiCG)$_2$, r(iCGGCCiG)$_2$, and r(iGCCGGiC)$_2$ are shown in Figure 2. The resonances were assigned by standard procedures.[44] All of the spectra have the number of resonances and NOEs expected, indicating that iG forms a base pair with iC.

NMR and Structural Modeling of r(CiGCGiCG)$_2$. To determine if the thermodynamics are affected by structural changes and to thereby provide a foundation for theoretical investigations of the molecular basis for the stability of iG−iC pairs, a structural model for r(CiGCGiCG)$_2$ (Scheme 1) was determined by NMR and restrained simulated annealing.

A new pulse sequence as shown in Figure 3 was used to obtain the 2D NOESY spectrum for r(CiGCGiCG)$_2$ in H_2O

Scheme 1

$C_1iG_2C_3G_4iC_5G_6$

$G_6iC_5G_4C_3iG_2C_1$

shown in Figure 4. In this pulse sequence, the H_2O resonance is attenuated by combinations of soft 90° pulses and pulsed field gradients during the mixing time period (Figure 3). For optimal attenuation of the H_2O signal, composite pulses are used at each end of the mixing time period. If necessary, the water flip-back pulse can also be used just before the final read pulse. The advantage of this pulse scheme is the equal excitation across the whole spectrum. Unlike the WATERGATE technology,[45] there is no delay after the read pulse, therefore eliminating the relaxation loss. This pulse scheme is also easy to implement since it does not need to match the pulses as in WATERGATE.

The resonances in Figure 4 were assigned by standard procedures.[44] A particularly interesting proton resonance is seen at 10.4 ppm, which straddles the regions expected for imino (10−14 ppm) and amino (6.5−8.0) protons in a regular RNA duplex. This is also seen in spectra for r(iCGGCCiG)$_2$ and r(iGCCGGiC)$_2$ (Figure 2). For r(CiGCGiCG)$_2$, a strong cross-peak was seen between the resonance at 10.4 ppm and a resonance at 6.7 ppm, suggesting the cross-peak is due to exchange between protons from the same amino group (Figure 4). The two resonances were assigned to the iG−iC pair by elimination because the amino protons of C were already assigned in a standard manner and the amino protons of G in a Watson−Crick base pair are usually missing because they are in an intermediate exchange regime. The resonances are assigned to iG because two cross-peaks are seen between C1 and C3 amino protons and the resonance at 10.4 ppm. If this resonance at 10.4 ppm is from the amino proton of iC, the expected shortest distances to the amino protons of C1 and C3 are too far (5.8 and 6.6 Å for C3 and C1, respectively, in an A-form helix) to have an NOE cross-peak. On the other hand, the distances between the amino protons of iG2 and C1, C3 are 2.9 and 3.4 Å, and 3.1 and 4.4 Å, respectively. Therefore, the resonances at 10.4 and 6.7 ppm are assigned to the amino protons of iG. The resonances of the amino protons of iC are apparently missing. Amino proton resonances with similar characteristics have been observed previously for deoxy-iG, deoxy-iC in a parallel DNA duplex, 5′-d(TiGiCAiCiGiGAiCT)-3′/5′-d(ACGTGCCTGA)-3′.[46]

Seela et al.[47] demonstrated that iG exists dominantly in a 1H/6-amino/2-oxo tautomer under neutral aqueous conditions. The existence of the iG2H1 resonance, the fact that iG2H1 disappears almost at the same temperature as G4H1, and the NOE cross-peak between iG2H1 and G4H1 protons show that this is also the tautomer in the duplex, where here and henceforth the underlined number designates the opposite strand. The similar thermal stabilities of iG−iC and G−C pairs also suggest that there are three hydrogen bonds in each iG−iC pair.

Phosphorus chemical shift is a sensitive indicator of RNA backbone conformation.[48,49] All phosphorus resonances for r(CiGCGiCG)$_2$ are within 1 ppm of each other, suggesting iG−iC fits well with the G−C pairs of this RNA duplex. The NOE

(43) Petersheim, M.; Turner, D. H. *Biochemistry* **1983**, *22*, 256−263.
(44) (a)Varani, G.; Tinoco, I., Jr. *Q. Rev. Biophys.* **1991**, *24*, 479−532. (b) Chen, X.; McDowell; J. A.; Kierzek, R.; Krugh; T. R.; Turner, D. H. *Biochemistry* **2000**, *39*, 8970−8982.
(45) Lippens, G.; Dhalluin, C.; Wieruszeski, J. M. *J. Biomol. NMR* **1995**, *5*, 327−331.
(46) Yang, X. L.; Sugimoto, H.; Ikeda, S.; Saito, I.; Wang, A. H. J. *Biophys. J.* **1998**, *75*, 1163−1171.
(47) Seela, F.; Wei, C.; Kazimierczuk, Z. *Helv. Chim. Acta* **1995**, *78*, 1843−1854.
(48) Gorenstein, D. G. *Chem. Rev.* **1994**, *94*, 1315−1338.
(49) Rife, J. P.; Stallings, S. C.; Correll, C. C.; Dallas, A.; Steitz, T. A.; Moore, P. B. *Biophys. J.* Part 1 **1999**, *76*, 65−75.

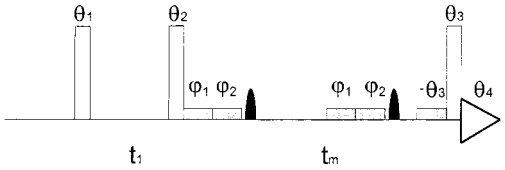

Figure 3. The water NOESY pulse scheme. Shaded rectangles represent soft water 90° pulses. Dark bells represent pulsed field gradients. Composite 90° pulse and pulsed field gradient are used in combination to selectively attenuate water signal. Water flipback pulse ($-\theta_3$) is optional. Phases (minimal): $\theta_1 = x$, $\theta_2 = (x, -x)$, $\theta_3 = (x, y, -x, -y)$, $\theta_4 = \theta_1 + \theta_2 + \theta_3$, $\varphi_1 = x$, $\varphi_2 = y$. The pulse widths are 5000 and 6 μs for the soft and hard 90° pulses, respectively. The length of the pulsed field gradient is 2 ms. Gradient recovery time is 1 ms.

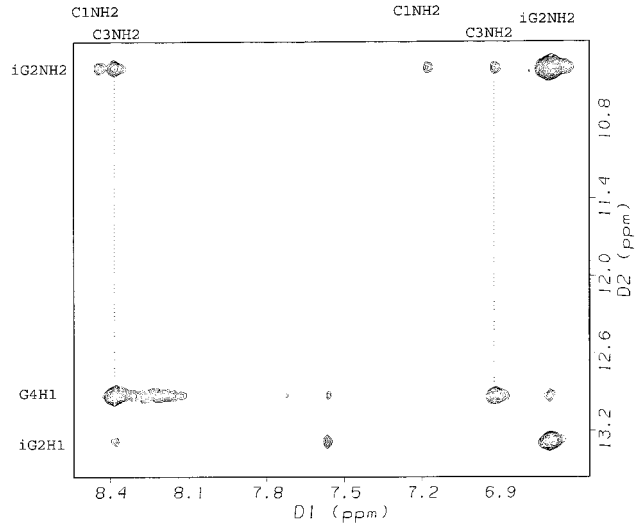

Figure 4. Portion of 400 ms NOESY spectrum of r(CiGCGiCG)$_2$ in H_2O at 0 °C. The D2-dimensional assignments are shown to the left. Some important cross-peaks are labeled.

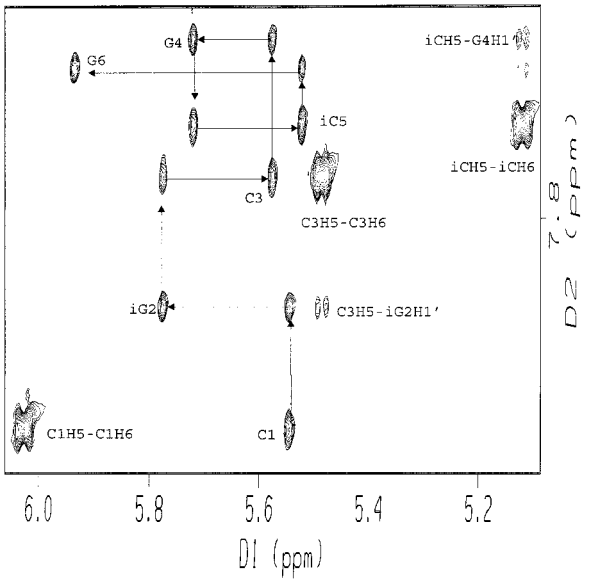

Figure 5. Portion of a 400 ms NOESY spectrum of r(CiGCGiCG)$_2$ at 30 °C in D_2O showing the H8/H6 to H1′/H5 region and the connectivities.

connectivities of (*n*)H1′/H5 to (*n*+1)H8/6 and (*n*)H2′ to (*n*+1)H8/6 in the 2D NOESY spectra in D_2O (Figure 5 and Supporting Information) demonstrate a general A-form geometry for r(CiGCGiCG)$_2$. Small *J*-coupling constants of $J_{H1'-H2'}$ (<3 Hz) are observed for all the internal sugars in r(CiGCGiCG)$_2$. This is also consistent with A-form structure, which has predominant C3′-endo sugar puckers.[50,51]

A total of 182 distance and 76 dihedral angle restraints (Supporting Information) were used to model the structure of r(CiGCGiCG)$_2$. A stereoscopic view of the model without the terminal pairs is shown in Figure 6, and the stacking patterns are shown in Figure 7. The helical parameters, such as X-displacement, according to the Cambridge convention[52] are reported as Supporting Information. The overall geometry for r(CiGCGiCG)$_2$ is closer to A-form than to B-form. This is clear from the X-displacement values, which are on average −2.9 Å, closer to the A-form value of −4.1 Å than to the B-form value of +0.8 Å.[53,54]

It is interesting to note that although the geometry of r(CiGCGiCG)$_2$ belongs to the A-form family, it has substantial deviations. In addition to the variation in X-displacement, the rise values are generally small, 2.5 Å, which is lower than the standard A-form value of 2.9 Å. The average twist angle of 38° is larger than the standard A-form value of 31°. This is not likely due to modeling, however, as similar results have been observed previously. The duplex r(CGCGCG)$_2$ has been studied extensively[50,51] and an NMR structural model at low salt was determined recently.[55] A crystal structure and NMR model of its 2′-*O*-methyl derivative were also determined.[55,56] These structures are generally A-form, but also with substantial deviations. The rmsd for all the common atoms of r(CiGCGiCG)$_2$ and the crystal structure of 2′-*O*-methyl r(CGCGCG)$_2$[56] are 1.56 and 1.03 Å with and without the terminal nucleotides, respectively. Similarly, the rmsd for all of the common atoms of r(CiGCGiCG)$_2$ and the NMR model of r(CGCGCG)$_2$[55] are 1.09 and 0.90 Å, respectively. All of these comparisons suggest that substitution of iG−iC for G−C does not affect the global structure very much.

Discussion

RNA plays an important role in cellular processes, including regulation and catalysis.[1,2,4,57−59] The interactions determining RNA structures, however, are not completely understood.[12−20] One test of current understanding is to compare theoretical predictions to experimental observations of structure and energetics. Here we provide an experimental benchmark—the energetic effects of transposing the amino and carbonyl groups in G−C pairs to form iG−iC pairs. It has been hypothesized that electrostatics are important for RNA stability and structure.[60−62] Since iG−iC and G−C pairs are isosteric, they presumably differ largely in electron density distribution. Thus, the iG−iC pair is a good model to test electrostatic calculations.

(50) Westerink, H. P.; van der Marel, G. A.; van Boom, J. H.; Haasnoot, C. A. G. *Nucleic Acids Res.* **1984**, *12*, 4323−4338.
(51) Haasnoot, C. A. G.; Westerink, H. P.; van der Marel, G. A.; van Boom, J. H. *J. Biomol. Struct. Dyn.* **1984**, *2*, 345−360.
(52) Dickerson, R. E. *Nucleic Acids Res.* **1989**, *17*, 1797−1803.
(53) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(54) Allain, F. H.-T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333−353.
(55) Popenda, M.; Biala, E.; Milecki, J.; Adamiak, R. W. *Nucleic Acids Res.* **1997**, *25*, 4589−4598.
(56) Adamiak, D. A.; Milecki, J.; Popenda, M.; Adamiak, R. W.; Dauter, Z.; Rypniewski, W. R. *Nucleic Acids Res.* **1997**, *25*, 4599−4607.
(57) Lewin, B. *Genes* VI.; Oxford University Press: Oxford, 1997.
(58) Walter, P.; Blobel, G. *Nature* **1982**, *299*, 11349−11354.
(59) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416−1419.
(60) Wu, M.; Turner, D. H. *Biochemistry* **1996**, *35*, 9677−9689.
(61) McDowell, J. A.; Turner D. H. *Biochemistry* **1996**, *35*, 14077−14089.
(62) McDowell, J. A.; He, L.; Chen, X.; Turner D. H. *Biochemistry* **1997**, *36*, 8030−8038.

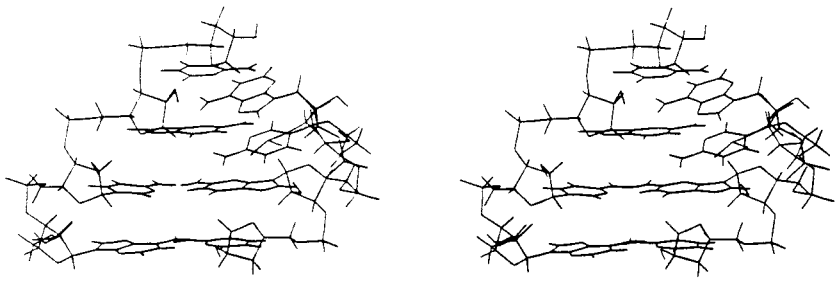

Figure 6. Stereoscopic view of the structural model for r(CiGCGiCG)$_2$ without the terminal nucleotides.

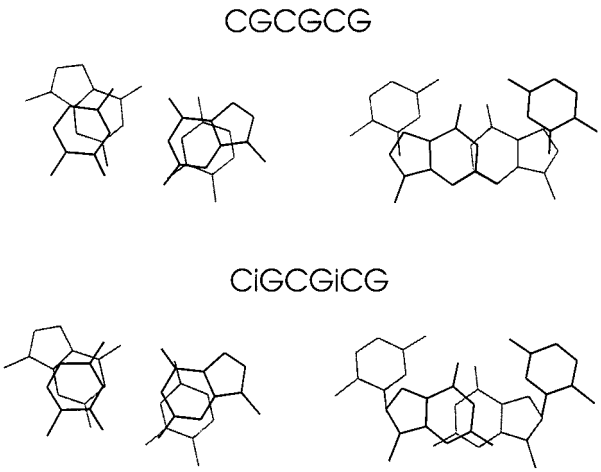

Figure 7. Stacking patterns for the 5′GC3′/3′CG5′ (top) or 5′iGC3′/3′iCG5′ (bottom), and 5′CG3′/3′GC5′ steps of r(CGCGCG)$_2$ and r(CiGCGiCG)$_2$, respectively.

An understanding of the important interactions in RNA will also provide insight into the important interactions in DNA and other polymers with similar compositions. Such understanding is important for designing self-assembling structures,[8−11] and for modeling structures from sequence,[12,38,40] NMR, and crystallographic data.[18,49,63,64]

The thermodynamics of duplex formation (Table 1) indicate that a duplex with iG−iC pairs is usually more stable than the corresponding Watson−Crick duplex but that the degree of difference in stability is sequence-dependent. This trend is also reflected in the nearest-neighbor parameters of Table 2. Except for 5′GiG3′/3′CiC5′, free energy increments for nearest neighbors with at least one iG−iC pair are essentially equivalent or more favorable than those for nearest neighbors with only G−C pairs. One possible source for this increase of stability is the strength of the hydrogen bonds in the iG−iC pair. Theoretical calculations predict that an iG−iC pair is stronger than a G−C pair.[65] The sequence dependence of the thermodynamics, however, indicates that hydrogen bonding is not the only important factor. For example, each iG−iC substitution in 5′GC3′/3′CG5′ enhances duplex stability by 0.6 kcal/mol, whereas single and double substitutions in 5′CG/3′GC5′ and 5′GG3′/3′CC5′ change stability by ≤0.2 kcal/mol (Table 2). This suggests vertical electrostatic interactions are also important. The larger sequence dependence observed with iG−iC pairs compared to that with G−C pairs significantly increases the range of free energies observed for nearest neighbors of base pairs with three hydrogen bonds. When only G−C pairs are used, the range is −2.4 to −3.4 kcal/mol, but when iG−iC pairs are included, the range is −2.4 to −4.6 kcal/mol. Reproducing these trends presents an interesting challenge for theoretical calculations. NMR and simulated annealing indicate that A-form RNA structures can be assumed in theoretical studies that attempt to reproduce the sequence dependence of stability for RNA containing iG−iC and G−C base pairs. Success in connecting theory with experiment would further our understanding of the interactions that shape RNA and related polymers.

Stacking Propensities of iG and iC Are Similar to those of G and C. Some of the differences in stability of iG−iC versus G−C nearest-neighbor parameters could come from differences in stacking propensities of G and C compared to those of iG and iC, respectively. To test for intrinsic differences in stacking propensities, thermodynamic parameters were measured for duplexes with unpaired iG or iC nucleotides as "dangling ends" (Table 3). As shown in Table 4, the resulting nearest-neighbor parameters for iG and iC dangling ends are similar to those previously measured for G and C.[12] Thus, the stacking propensities of iG and iC are close to those of G and C, respectively. Interestingly, however, the largest differences are observed for 5′-iGC-3′ and 5′-GiC-3′ sequences, where the unpaired nucleotide is underlined. This correlates with iG−iC replacement having the largest effect on the 5′-GC-3′/3′-CG-5′ nearest neighbor and is consistent with the proposed importance of vertical electrostatic interactions.

The Sign and Magnitude of the Terminal iG−iC Free Energy Term Are Consistent with an Important Contribution from Hydrogen Bonds to the Free Energies of Duplex Formation. Goldstein and Benight[41] and Gray[39] have pointed out that a simple nearest-neighbor model allows extraction of a maximum of 12 sequence-dependent parameters from data on duplexes with only two different types of base pairs. In the individual nearest-neighbor-hydrogen bonding (INN-HB) model, one of these parameters, the terminal base pair parameter, is attributed to half the difference in hydrogen bonding strength between the two types of base pairs.[40] This accounts for the fact that a duplex with two terminal A−U pairs can have the same nearest-neighbor composition as a duplex with two terminal G−C pairs, while having one more AU and one less G−C pair. In particular, results for duplexes containing only G−C and A−U pairs provided an unfavorable free energy increment of 0.45 kcal/mol at 37 °C for each terminal A−U pair; thus, the favorable free energy increment attributed to the third hydrogen bond in a G−C pair has a magnitude of 2 × 0.45 = 0.9 kcal/mol.[40] On the basis of this model, it is expected that the free energy increment for having one more iG−iC pair and one less G−C pair in a duplex with the same nearest neighbors would be smaller in magnitude and probably of opposite sign to that observed for A−U. The increment is expected to be smaller because iG−iC pairs have the same

(63) Brunger, A. *X-PLOR Version 3.1. System for X-ray Crystallography and NMR;* Yale University Press: New Haven, CT, 1992.
(64) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(65) Leach, A. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1992**, *114*, 3675−3683.

number of hydrogen bonds as G−C pairs, whereas A−U pairs have one less hydrogen bond. The increment is expected to be of opposite sign because theoretical calculations predict the hydrogen bonds in an iG−iC pair are stronger than those in a G−C pair.[65] Both expectations are met. The free energy increment per terminal iG−iC pair is a favorable −0.19 kcal/mol compared to the unfavorable 0.45 kcal/mol for each terminal A−U pair.[40] Thus the results support interpretations that hydrogen bonds contribute to stabilities of Watson−Crick pairs[25] and that this is the physical basis for the terminal base pair term in the INN-HB model for predicting duplex stabilities.[40]

There May Be Thermodynamic Reasons for Selection against iG−iC Pairs during Evolution. There are at least three explanations that have been proposed to support the belief that iG and iC were available for the early stages of evolution[28] but are not found in living organisms: (1) iG has another significant tautomeric form,[36] (2) deoxy-iC degrades fairly quickly into deoxy-U,[36] and (3) iG and iC cannot form the R-form triplex that is required for repairing genetic material.[66] Properties (1) and (2) would significantly decrease the intrinsic fidelity in replication of genetic information. There may also be thermodynamic reasons for selection against iG−iC pairs. If iG−iC replaced either A−U or G−C pairs, helix stability would increase. This would make it more difficult to unwind a helix for replication. Moreover, if iG−iC replaced G−C, then the range of free energy parameters for helix propagation in RNA at 37 °C would expand from the current range of −0.9 to −3.4 kcal/mol[40] to −0.9 to −4.6 kcal/mol. Thus, mutations might have more dramatic effects on RNA folding with an A−U/iG−iC code than with an A−U/G−C code, thereby limiting the capacity of RNA to make small evolutionary excursions.

The iG−iC Pair Can Facilitate Design of Self-Organizing Structures. Nucleic acids and similar polymers are being used to design self-organizing structure.[8−11] As these structures get larger, the possibilities for alternative foldings increase,[67] thus complicating design. Due to kinetic trapping, alternative foldings can also impose strict requirements on conditions for self-assembly.[68] Such effects can be observed in sequences containing as few as 16 nucleotides.[69] Expanding the nucleic acid alphabet can reduce the number of possible folds and thus simplify design of a sequence to produce a unique fold. The results presented here show that the iG−iC pair will be useful for such applications because it fits well into the same helix as G−C and A−U pairs, and its contributions to the energetics of folding are well approximated by a nearest-neighbor model.

Experimental Section

Synthesis of Phosphoramidites of iG and iC. Protected isoguanosine (iG) was synthesized as described previously.[44b] The substrate for isocytidine synthesis was uridine which was first converted to 2,5′-anhydrouridine and then treated with ammonia-saturated methanol to obtain isocytidine.[70] The 2-amino group of the isocytidine was protected with dimethylchloromethyleneammonium chloride[47,71] followed by protection of the 5′-hydroxyl with dimethoxytrityl chloride. At this stage, the overall yield was 46%. After treatment of the 5′-*O*-dimethoxytrityl-2-*N*-dimethylaminoethylideneamino-isocytidine with *tert*-butyldimethylsilyl chloride in the presence of imidazole, the 2′-silylated isomer of the protected isocytidine was obtained in 55% yield. This last derivative was transformed into the 3′-phosphoramidite with 88% yield.[72]

Base Composition Analysis. Oligonucleotide base composition was confirmed for r(iCGGCCiG)$_2$ and r(iGCCGGiC)$_2$ by enzymatic digestion of 0.4 A_{260} unit of sample. Stock $MgSO_4$ and $ZnCl_2$ solutions were added so that the final concentration of each was 1 mM. After addition of 10 units of nuclease P1 (Boehringer Mannheim) to 0.3 mL of sample, the mixture was incubated at 65 °C for 24 h. After this, 40 units of calf intestine phosphatase (Boehringer Mannheim) was added and the reaction mixture was incubated at 37 °C for 12 h. To check for completion, 1 unit of nuclease P1 and 4 units of calf intestine phosphatase were added, and the total reaction mixture was incubated at 37 °C for another 6 h. No further increase of absorbance was observed. The digestion mixture was injected into a C-18 reverse-phase column (Supercosil) using 50% acetonitrile as mobile phase. Both oligonucleotides had the compositions expected.

Determination of Extinction Coefficients for iG, iC Containing Duplexes. The extinction coefficient was measured by a microdetermination method of phosphorus.[73] Reagent C was prepared freshly daily from 3 volumes of 2 N sulfuric acid, 1 volume of 2.5% (w/w) sodium molybdate, and 1 volume of 10% (w/w) ascorbic acid. The 10% (w/w) ascorbic acid solution can be kept for several weeks in the dark. Distilled water was mixed with a completely digested reaction mixture of the type used for analysis of base composition to give a final volume of 1.5 mL. Then 1.5 mL of reagent C was added to the solution in a 15 mL test tube followed by incubation at 37 °C for 1.5−2 h. After cooling to room temperature, absorbance at 830 nm was measured. Each 0.1 μmol of phosphorus gives an absorbance at 830 nm of 0.8 for a 3 mL solution in a 1 cm cell. The measured extinction coefficients at 280 nm for iCGGCCiG and iCCAUGiG are 33 200 and 31 500 M^{-1} cm^{-1}, respectively. The calculated values according to the nearest-neighbor model[74,75] are 33 340 and 30 620 M^{-1} cm^{-1}, respectively, assuming that iG and iC are equal to G and C, respectively. This is within experimental error of the measured extinction coefficients. Therefore, the extinction coefficients for iG, iC containing duplexes studied here were calculated by assuming iG and iC are equal to G and C, respectively.

Thermodynamic Measurements. Optical melting curves at 280 nm were measured on a Gilford 250 spectrometer with a heating rate of 1 °C min^{-1}. The buffer system was 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA at pH 7.0. Thermodynamic parameters for duplex formation were calculated by two methods: (1) by fitting the shape of each melting curve to a two state model[43] and (2) by plotting the reciprocal of the melting temperature, T_M^{-1}, versus $\ln(C_T)$,[38a] where T_M is the melting temperature in kelvins and C_T is the total strand concentration.

$$1/T_M = (R/\Delta H^\circ)\ln(C_T) + \Delta S^\circ/\Delta H^\circ \quad (1)$$

Note that the free energy analysis is relatively insensitive to extinction coefficient errors. A 10% difference in the extinction coefficient results in only a difference of $RT\ln(1.1)$ in ΔG°, which is equal to about 0.06 kcal/mol at 37 °C.

Extraction of Nearest-Neighbor Parameters. The thermodynamic properties for formation of an RNA duplex can be approximated by a nearest-neighbor model.[38,40] It is assumed in the nearest-neighbor model that the contribution of a given base pair to the thermodynamic properties of an RNA duplex depends on the identity of two directly adjacent neighbors and has a linear dependence on the occurrence of these nearest neighbors. For example, in the individual nearest-neighbor-

(66) Cox, M. M. *Mutation Res.* **1997**, *384*, 15−22.
(67) Zuker, M.; Sankoff, D. *Bull. Math. Biol.* **1984**, *46*, 591−621.
(68) Stage, T. K.; Hertel, K. J.; Uhlenbeck, O. K. *RNA* **1995**, *1*, 95−101.
(69) Schroeder, S. J.; Turner, D. H. *Biochemistry* **2000**, *39*, 9257−9274.
(70) Strobel, S. A.; Cech, T. R.; Usman, N.; Beigelman, L. *Biochemistry* **1994**, *33*, 13824−13835.
(71) Zemlicka, J.; Sorm, F. *Collect. Czech. Chem. Commun.* **1965**, *30*, 1880−1889.
(72) Wincott, F.; DiRenzo, A.; Shaffer, C.; Grimm, S.; Tracz, D.; Workman, C.; Sweedler, D.; Gonzalez, C.; Scaringe, S.; Usman, N. *Nucleic Acids Res.* **1995**, 23, 2677−2684.
(73) Chen, P. S.; Toribara, T. Y., Jr.; Warner, H. *Anal. Chem.* **1956**, *28*, 1756−1758.
(74) Borer, P. N. In *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1975; Vol. 1, p 597.
(75) Richards, E. G. In *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed.; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1975; Vol. 1, p 579.

hydrogen bonding (INN-HB) model when only G−C and iG−iC pairs are considered, the free energy for RNA duplex formation can be approximated by:[40]

$$\Delta G^\circ = \Delta G^\circ_{\text{init}} + \Delta G_{\text{sym}} + m\Delta G^\circ_{\text{term-iG-iC}} + \sum_j n_j \Delta G^\circ_j(\text{NN}) \quad (2)$$

Here $\Delta G^\circ_{\text{init}}$ is the initiation free energy, ΔG_{sym} at 37 °C is 0.43 and 0 kcal/mol for self- and nonself-complementary sequences, respectively, $\Delta G^\circ_{\text{term-iG-iC}}$ is the free energy term for each terminal pair that is an iG−iC pair, m is the number of terminal iG−iC pairs, n_j is the number of jth nearest neighbors, and ΔG°_j(NN) is the free energy for the jth nearest neighbor.

The linear regression was done with the program MATHEMATICA version 3.0, assuming standard errors of 4, 12, and 13.5% for ΔG°_{37}, ΔH°, and ΔS°, respectively. The goodness-of-fit is revealed by the Q values. A value larger than 0.001 is generally considered acceptable.[42]

NMR Spectroscopy. The NMR buffer system was 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA at pH 7.0. Sample preparation, data collection, and processing were the same as those described previously.[44b] Proton chemical shifts were referenced to the internal EDTA. The one-dimensional exchangeable proton spectra were acquired with a binomial 1:3:3:1 pulse sequence[76] for H_2O peak suppression with a 12 000 Hz sweep width. Pulse delays were calculated to achieve a signal-to-noise maximum at 12 ppm, which is in the imino proton resonance region. One-dimensional NOE experiments were performed by irradiation for 2 s at low power.

Structural Modeling. Distance and dihedral angle restraints were generated as described previously.[44b] Hydrogen bonds for iG−iC pairs were restrained as: 2.81−3.01 for iGO2 to iCN2, 2.85−3.05 for iGN1 to iCN3 and 2.76−2.96 for iGN6 to iCO4.[44b,77,78] A total of 182 distance and 76 dihedral restraints were used, and 40 structures were generated and converged very well. Structural modeling was accomplished by restrained molecular dynamics and energy minimization with the force field AMBER95.[64] The atom potential types for iG and iC were assigned following the potential type definitions in Cornell et al.[64] To accommodate the unusual bases iG and iC, some new molecular mechanical parameters were added. The values for these parameters were obtained by comparing with similar, known parameters (templates) in the original AMBER95 force field.[64] A table of all of the additional parameters is supplied as Supporting Information. The partial charges for atoms in bases of iG and iC nucleotides were obtained from G and C, respectively, by flipping the charges of the amino and carbonyl groups as well as the carbons to which they attach in the six-membered rings. Modeling was also carried out without partial charges for the bases of iG and iC nucleotides. The structural model was the same with and without partial charges for the bases of iG and iC, suggesting the model was determined by factors other than the partial charges. Coordinates for the final model are available from the Protein Data Bank.[79]

Acknowledgment. This work was supported by NIH Grants GM 22939 (D.H.T.) and TW1068 (D.H.T. and R. K.). We thank Dr. Tianbing Xia for help with MATHEMATICA and Professors Jiali Gao and Thomas R. Krugh for stimulating discussions.

Supporting Information Available: One table of thermodynamic parameters measured from $1/T_M$ versus $\ln(C_T)$ plots and predicted with the individual nearest-neighbor-hydrogen bonding (INN-HB) model, one table of chemical shifts for r(CiGCGiCG)$_2$, one table of distance restraints, one table of dihedral restraints, one table of molecular dynamics parameters, two tables of helical parameters for r(CiGCGiCG)$_2$, two tables containing thermodynamic parameters derived from fitting shapes of melting curves, eight $1/T_M$ versus $\ln(C_T)$ plots for RNA duplexes, one plot of the drawing of G−C and iG−iC pairs with the atom potential types, and one plot of 2D D_2O NOESY for r(CiGCGiCG)$_2$ showing the (n)H2′ to (n+1)H8/6 connectivity (PDF). This material is available free of charge via the Internet at http:://pubs.acs.org.

JA002623I

(76) Hore, P. J. *J. Magn. Reson.* **1983**, *55*, 283−300.
(77) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 98−980.
(78) Dallas, A.; Moore, P. B. *Structure* **1997**, *5*, 1639−1653.
(79) Coordinates for the NMR structure of r(CiGCGiCG)$_2$ have been deposited with the Protein Data Bank, accession numbers 1G3A and RCSB012179.