Supporting Information

Testing the nearest neighbor model for canonical RNA base pairs: Revision of GU parameters

Jonathan L. Chen[1], Abigael L. Dishler[2], Scott D. Kennedy[3], Ilyas Yildirim[1,4], Biao Liu[1,5], Douglas H. Turner[1,6]*, and Martin J. Serra[2]

[1] Department of Chemistry, University of Rochester, Rochester, New York, 14627, USA

[2] Department of Chemistry, Allegheny College, Meadville, Pennsylvania 16335, USA.

[3] Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, USA

[4] Current address: Department of Chemistry, Northwestern University, Evanston, Illinois, 60208, USA

[5] Current address: Roswell Park Cancer Institute, Buffalo, New York, 14263.

[6] Center for RNA Biology, University of Rochester, Rochester, New York, 14627, USA

Jonathan L. Chen: jchen46@z.rochester.edu
Abigael L. Dishler: dishlea@allegheny.edu
Scott D. Kennedy: Scott_Kennedy@urmc.rochester.edu
Ilyas Yildirim: i-yildirim@northwestern.edu
Biao Liu: biao.liu@roswellpark.org
Douglas H. Turner: turner@chem.rochester.edu
Martin J. Serra: mserra@allegheny.edu

*To whom correspondence should be addressed. Phone: (585) 275-3207. Fax: (585) 276-0205. E-mail: turner@chem.rochester.edu

This work was supported by NIH Grant GM22939 (D.H.T.)

Table S1 Thermodynamic parameters for duplex formation for Watson-Crick sequences

Table S2 Experimental absolute error limits of new sequences used in this study. $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ values were obtained from $1/T_M$ vs. $\ln(C_T/a)$ fits.

Table S3 Component free energies of GU pairs.

Table S4 Component enthalpies of GU pairs.

Table S5 Comparison of measured and predicted $\Delta G°_{37}$ values for terminal GU pairs.

Table S6 The number of occurrences of each triplet, the component $\Delta G°_{37}$, and in brackets, the sum of the $\Delta G°_{37}$'s of each INN in triplet.

Table S7 Free energy differences when GC pairs are replaced with AU pairs.

Table S8 INN parameters for GU base pairs in 1 M NaCl with a separate parameter for terminal GU pairs.

Table S9 Probability density function of Student's t-distribution for each INN motif with a separate parameter for terminal GU pairs.

Figure S1 WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 100 ms mixing time at 0 °C showing the H2´-H6/H8 region.

Figure S2 WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 400 ms mixing time at 8 °C showing the H6/H8/H2/H5/H1´ region. C and U H5-H6 NOEs are labeled with crosses.

Figure S3 WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 400 ms mixing time at 8 °C showing the H2/H6/H8 region.

Figure S4 NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H2´-H6/H8 region.

Figure S5 NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H6/H8/H2/H5/H1´ region.

Figure S6 NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H2/H6/H8 region.

Table S1: Thermodynamic parameters for duplex formation for Watson-Crick sequences

	T_M^{-1} vs log C_T				average of curve fits				
Oligomer[a]	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M[b] (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_M[b] (°C)	Ref.
CAUGCG/	7.00	48.6	134.0	42.9	7.01	52.2	145.6	42.5	(*1*)
CUGGUC/	8.02	55.1	151.7	45.6	8.13	57.7	159.7	45.9	(*2*)
GAGCUG/	7.49	51.6	142.2	45.5	7.95	55.1	152.0	47.9	(*1*)
GCUGAG/	7.72	55.9	155.2	46.2	7.95	55.1	152.0	47.9	(*1*)
GUGCAG/	7.67	55.9	155.6	46.0	7.87	54.0	148.8	47.5	(*1*)
GAGUGAG/	9.71	70.5	196.0	51.6	10.03	80.1	226.1	51.2	(*3*)
UAAGGUA/	6.95	51.3	142.9	42.2	6.18	46.4	129.8	37.4	(*1*)
AGGCCGGA/	15.26	84.1	221.9	72.9	14.95	81.1	213.4	72.9	(*1*)
CGAGCUCG	13.16	88.0	241.4	65.8	12.82	83.7	228.4	65.9	(*4*)
GAGAUCUC	10.11	75.0	209.2	56.5	10.14	76.5	213.8	56.4	(*1*)
GAGCCGAC/	13.94	80.1	214.0	68.5	14.34	84.7	226.8	68.5	(*5*)
GAGUACUC	9.70	70.0	194.3	55.9	10.20	79.5	223.3	55.9	(*6*)
GAGUUGAC/	10.57	88.7	251.8	51.8	10.23	79.9	224.6	52.0	(*7*)
GCAGCUGC	13.87	88.8	241.7	68.5	13.63	86.0	233.5	68.6	(*4*)
GCCAUGGC	15.06	93.9	254.2	71.4	14.16	83.2	222.6	72.2	(*1*)
GCUGCAGC	14.50	96.5	264.5	68.2	14.34	94.8	259.5	68.2	(*4*)
GGCAUGCC	13.30	82.2	222.0	68.8	14.00	90.2	245.7	68.6	(*6*)
GCUGCGAC/	13.93	86.2	233.0	67.9	13.89	81.6	218.0	70.0	(*1*)
GUCGCAGC/	13.93	86.2	233.0	66.1	13.80	84.5	227.9	66.2	(*5*)
UCACCUGA/	10.84	71.5	195.6	56.9	11.12	75.4	207.2	57.1	(*5*)
UCCGCGCA/	14.59	81.2	214.6	73.2	14.29	76.3	199.8	74.3	(*5*)
AAGGCCGGAA/	18.16	98.3	258.5	78.6	17.00	88.0	228.8	78.9	(*5*)

[a] Listed in order of length of the oligoribonucleotide and then alphabetically for sequences of the same length. All non-self-complementary sequences have a slash and only one sequence shown. [b] Calculated at a total strand concentration of 1×10^{-4} M.

Table S2: Measured thermodynamics and error limits for new sequences used in this study. $\Delta G°_{37}$, $\Delta H°$ and $\Delta S°$ values were obtained from $1/T_M$ vs. $\ln(C_T/a)$ fits and average of curve fits.

Sequence[a]	$-\Delta G°_{37}$ (kcal/mol)[b]	$-\Delta G°_{37}$ error (kcal/mol)[b]	$-\Delta G°_{37}$ (kcal/mol)[c]	$-\Delta G°_{37}$ error (kcal/mol)[c]	$-\Delta H°$ (kcal/mol)[b]	$-\Delta H°$ error (kcal/mol)[b]	$-\Delta H°$ (kcal/mol)[c]	$-\Delta H°$ error (kcal/mol)[c]	$-\Delta S°$ (eu)[b]	$-\Delta S°$ error (eu)[b]	$-\Delta S°$ (eu)[c]	$-\Delta S°$ error (eu)[c]
GCGGGAC/	9.00	0.16	9.30	0.16	45.20	3.68	50.60	5.69	116.8	11.5	133.1	17.3
CAGGGCUC/	11.10	0.27	11.50	0.27	62.80	4.34	65.60	3.30	166.6	13.2	174.2	9.8
CUGGCUAG[d]	7.10	0.10	7.06	0.09	60.38	4.64	62.56	2.05	171.8	15.0	179.0	6.6
GACGCCAG/	10.50	0.17	11.00	0.17	63.80	3.29	74.60	6.41	171.8	10.1	205.1	19.2
GACGCGUU	9.50	0.32	9.60	0.32	62.20	5.71	62.70	4.77	169.9	17.5	171.4	14.2
GCAGCUGU	10.30	0.26	10.40	0.26	72.30	4.64	72.00	8.59	199.8	14.1	198.4	13.1
GGAGCUCU	10.50	0.21	10.60	0.21	66.57	3.32	67.60	7.75	180.9	10.1	184.0	13.2
GUAGCUAU	7.30	0.05	7.40	0.05	50.30	2.18	57.90	11.92	138.7	7.0	162.8	35.7
GUCGGGCC/	15.00	0.05	14.20	0.05	96.00	2.95	83.30	3.59	261.3	9.6	222.7	14.3
UACCGGUG	9.70	0.26	10.00	0.26	51.70	3.85	54.70	8.18	135.4	11.6	144.1	14.7
UCACGUGG	8.40	0.32	8.60	0.32	46.90	6.57	45.00	8.96	124.2	20.7	117.1	28.7
UGACGUCG	10.40	0.37	10.60	0.37	63.80	5.62	67.30	5.71	172.3	17.0	182.7	17.0
UUACGUAG	6.20	0.04	6.20	0.04	44.60	2.10	49.20	2.76	124.0	6.8	138.9	8.8
CAGAGGAGAC/	9.43	0.06	9.39	0.22	98.95	2.36	95.19	8.89	288.6	7.4	276.7	28.0
CAGCGCGUUG	12.31	0.41	11.59	0.41	77.02	5.82	66.84	5.34	208.6	17.5	178.2	16.9
CAGUCGAUUG	8.70	0.10	8.54	0.24	92.33	3.45	84.54	8.52	269.7	10.8	245.0	26.8
CCGAAUUUGG[d]	6.99	0.08	7.00	0.08	76.76	9.49	76.35	5.88	225.0	30.8	223.5	19.1
CGGAAUUUCG	7.88	0.05	7.77	0.14	90.52	2.91	84.47	5.00	266.5	9.2	247.3	15.8
CGGAUAUUCG[d]	8.78	0.14	8.49	0.18	88.20	4.00	80.30	12.31	256.2	13.3	231.4	39.3
CGGGCGUUCG	11.55	0.15	11.65	1.58	101.66	2.78	101.34	24.34	290.5	8.5	289.2	73.4
CGGUGCAUCG	14.76	0.21	14.18	0.21	102.42	2.82	94.72	6.68	282.6	8.4	259.7	18.2
CUGGAUUCAG	10.15	0.17	9.97	0.33	97.81	4.16	92.86	6.25	282.7	12.9	267.3	19.2
GAGAGCUUUC[d]	8.82	0.04	8.72	0.17	86.57	9.96	82.81	1.42	250.6	32.0	238.9	5.1
GAGGAUCUUC	9.83	0.35	9.40	0.51	93.86	8.25	83.92	12.33	270.9	25.5	240.3	38.3
GAGUGGAGAG/	9.87	0.08	9.79	0.17	96.93	2.82	93.52	6.18	280.7	8.8	270.0	19.4
GGUUCGGGCC	13.59	0.13	13.56	1.97	115.71	2.08	113.36	28.05	329.3	6.3	321.8	84.1
GUGAAUUUAC[d]	4.78	0.17	4.60	0.01	62.63	8.24	72.40	4.66	186.4	27.3	218.6	15.0
GUGUGCAUAC	8.90	0.18	9.20	0.18	58.60	4.25	66.70	3.86	160.1	13.2	185.6	12.3
GUUAGCUGAC	8.60	0.10	8.50	0.10	69.60	3.70	66.90	4.63	196.7	11.7	188.5	14.8

[a] Listed in order of length of the oligoribonucleotide and then alphabetically for sequences of the same length. All non-self-complementary sequences have a slash and only one sequence shown. GU base pairs are underlined. [b] Calculated from $1/T_M$ vs. $\ln(C_T/a)$ fit. [c] Calculated from average of curve fits. [d] Errors were calculated from the standard deviation of multiple sets of measurements of their respective thermodynamic parameters for these sequences. Errors for other sequences were calculated from the fits of their respective thermodynamic parameters with MeltWin 3.5 (*8*).

Table S3: Component free energies of GU pairs.

	T_M^{-1} vs log C_T			
Oligomer	$\Delta G°_{37}$ ($1/T_M$ vs. $\ln(C_T/a)$) (kcal/mol)	Experimental $\Delta G°_{37}$(component) (kcal/mol)	Predicted $\Delta G°_{37}$(component)[a] (kcal/mol)	Predicted $\Delta G°_{37}$(component)[b] (kcal/mol)
CGGCUG	−5.55	−6.65	−6.04	−7.04
CUGCGG	−4.31	−5.41	−6.04	−7.04
GCCGGUp	−9.17	−4.81	−4.30	−5.02
GCGUGC	−5.11	−2.79	−1.78	−1.53
GCUGGC	−6.47	−4.15	−4.11	−3.92
GGCGCU	−8.42	−3.74	−3.54	−4.22
GGCGUC	−4.67	−6.83	−7.90	−8.08
GUGCAU	−5.10	−1.98	−1.80	−2.72
UCCGCC/	−6.71	−2.38	−3.40	−3.92
UCCGGGp	−7.44	−3.08	−3.60	−3.06
UGGCCGp	−8.56	−3.14	−2.50	−2.82
UUGCAG	−4.20	−1.08	−0.70	−1.10
CUCGCUC/	−7.78	−3.01	−3.40	−3.92
GCGGGAC/	−9.00	−2.72	−3.57	−3.64
AGUCGAUU	−6.00	−4.36	−2.50	−3.82
AUGCGCGUp	−9.31	−5.53	−4.30	−5.54
AUGCGUAUp	−5.27	−6.13	−5.08	−7.02
AUGUGCAUp	−6.17	−5.17	−4.87	−4.72
CAGGGCUC/	−11.10	−3.15	−3.57	−3.64
CCAGUUGG	−5.70	0.52	0.02	0.19
CCAUGUGG	−7.80	−1.58	−2.37	−2.42
CCUGUAGG	−6.81	−0.65	−0.06	−0.71
CGGAUUCG	−6.56	−5.26	−4.62	−5.60
CGUUGACG	−6.93	−2.25	−1.59	−2.24
CUCGGCUC/	−8.22	−3.45	−3.65	−3.45
CUGGCUAG	−7.10	−4.04	−4.32	−6.22
GACGCCAG/	−10.50	−2.55	−3.40	−3.92
GACGCGUU	−9.50	−1.40	−1.02	−2.54
GAGGUGAG/	−7.63	−2.86	−2.33	−1.65
GAGUGCUC	−9.40	−5.06	−4.87	−4.72
GAGUGGAG/	−9.66	−4.89	−4.49	−4.32
GAUGCAUUp	−6.82	−1.50	−1.02	−2.54
GCAGCUGU	−10.30	−3.02	−4.30	−5.02
GCAGUUGC	−5.90	0.64	0.02	0.19
GCAUGUGC	−8.40	−1.86	−2.37	−2.42

Table S3 (continued): Component free energies of GU pairs.

	T_M^{-1} vs log C_T			
Oligomer	$\Delta G°_{37}$ ($1/T_M$ vs. $\ln(C_T/a)$) (kcal/mol)	Experimental $\Delta G°_{37}$(component) (kcal/mol)	Predicted $\Delta G°_{37}$(component)[a] (kcal/mol)	Predicted $\Delta G°_{37}$(component)[b] (kcal/mol)
GCUGGUGC/	−7.60	−0.66	−1.54	−1.89
GGAGCUCU	−10.50	−2.74	−3.54	−4.22
GGAGUUCC	−6.43	0.27	0.02	0.19
GGAUGUCC	−8.39	−1.69	−2.37	−2.42
GGCGGGGC/	−13.80	−4.53	−3.27	−3.05
GGCGUGCC	−9.72	−0.88	−1.78	−1.53
GGCUGGCC	−13.10	−4.26	−4.11	−3.92
GGUUGACC	−8.30	−1.82	−1.59	−2.24
GUAGCUAU	−7.30	−1.58	−1.80	−2.72
GUCGGGCC/	−15.00	−5.46	−3.57	−3.64
GUCGUGAC	−6.05	−1.39	−1.78	−1.53
GUCUAGAU	−7.70	−2.03	−1.80	−2.72
UACCGGUG	−9.70	−0.86	−0.78	−2.00
UAUGCAUGp	−6.44	−1.12	−0.78	−2.00
UCACGUGG	−8.40	−1.86	−3.60	−3.06
UGACGUCG	−10.40	−3.38	−2.50	−2.82
UUACGUAG	−6.20	−1.22	−0.70	−1.10
CAGAGGAGAC/	−9.43	−0.31	−1.11	−1.35
CAGCGCGUUG	−12.31	−4.47	−5.00	−6.12
CAGUCGAUUG	−8.70	−1.94	−2.50	−3.82
CCAGCGUCCU/	−11.60	−3.08	−4.65	−5.83
CCGAAUUUGG	−6.99	−2.03	−3.52	−5.36
CGGAAUUUCG	−7.88	−4.72	−4.62	−5.60
CGGAUAUUCG	−8.78	−5.05	−4.62	−5.60
CGGGCGUUCG	−11.55	−8.99	−8.40	−7.14
CGGUGCAUCG	−14.76	−6.92	−5.40	−5.78
CUGGAUUCAG	−10.15	−5.19	−4.62	−5.60
GAGAGCUUUC	−8.82	−1.06	−1.72	−3.64
GAGGAUCUUC	−9.83	−3.85	−4.24	−5.32
GAGUGGAGAG/	−9.87	−0.78	−1.15	−1.80
GGUUCGGGCC	−13.59	−9.23	−8.40	−7.14
GUGAAUUUAC	−4.78	−1.86	−1.80	−4.54
GUGUGCAUAC	−8.90	−1.30	−2.58	−4.72
GUUAGCUGAC	−8.60	−1.06	−1.80	−4.54
UCGCCAGAGG/	−15.32	−2.37	−3.40	−3.92

[a] Calculated with free energies in Table 3. [b] Calculated with free energies in ref. (*9*).

Table S4: Component enthalpies of GU pairs.

	T_M^{-1} vs log C_T			
Oligomer	ΔH° (1/T_M vs. ln(C_T/a)) (kcal/mol)	Experimental ΔH°(component) (kcal/mol)	Predicted ΔH°(component)[a] (kcal/mol)	Predicted ΔH°(component)[b] (kcal/mol)
CGGCUG	−43.20	−31.93	−30.00	−35.44
CUGCGG	−41.40	−30.13	−30.00	−35.44
GCCGGUp	−58.20	−24.39	−22.18	−25.18
GCGUGC	−46.18	−20.03	−24.95	−25.81
GCUGGC	−59.10	−32.95	−31.52	−33.48
GGCGCU	−56.40	−19.61	−18.88	−24.22
GGCGUC	−38.10	−31.07	−36.24	−41.84
GUGCAU	−47.50	−15.35	−14.78	−17.62
UCCGCC/	−57.00	−25.11	−16.65	−18.20
UCCGGGp	−47.70	−13.89	−14.06	−16.66
UGGCCGp	−53.00	−14.95	−11.12	−11.22
UUGCAG	−37.20	−5.05	−7.92	−6.42
CUCGCUC/	−64.20	−21.97	−16.65	−18.20
GCGGGAC/	−45.20	0.55	−16.47	−20.44
AGUCGAUU	−53.30	−28.83	−22.70	−24.04
AUGCGCGUp	−54.90	−25.55	−25.90	−28.84
AUGCGUAUp	−46.80	−28.45	−24.10	−39.16
AUGUGCAUp	−57.10	−28.51	−34.82	−34.44
CAGGGCUC/	−62.80	−7.69	−16.47	−20.44
CCAGUUGG	−61.10	−17.05	−21.75	−21.01
CCAUGUGG	−70.50	−26.45	−27.42	−26.88
CCUGUAGG	−71.10	−26.97	−15.75	−28.57
CGGAUUCG	−72.60	−45.55	−34.82	−42.32
CGUUGACG	−73.50	−33.03	−33.40	−34.92
CUCGGCUC/	−73.90	−31.67	−34.47	−31.67
CUGGCUAG	−60.38	−28.15	−20.80	−38.20
GACGCCAG/	−63.80	−9.26	−16.65	−18.20
GACGCGUU	−62.20	−6.85	−20.76	−25.66
GAGGUGAG/	−78.40	−36.17	−26.42	−28.53
GAGUGCUC	−83.00	−40.77	−34.82	−34.44
GAGUGGAG/	−82.30	−40.07	−33.17	−33.96
GAUGCAUUp	−62.90	−11.99	−20.76	−25.66
GCAGCUGU	−72.30	−19.19	−22.18	−25.18
GCAGUUGC	−64.80	−17.77	−21.75	−21.01
GCAUGUGC	−72.40	−25.37	−27.42	−26.88

	T_M^{-1} vs log C_T			
Oligomer	$\Delta G°_{37}$ ($1/T_M$ vs. $\ln(C_T/a)$) (kcal/mol)	Experimental $\Delta G°_{37}$(component) (kcal/mol)	Predicted $\Delta G°_{37}$(component)[a] (kcal/mol)	Predicted $\Delta G°_{37}$(component)[b] (kcal/mol)
GCUGGUGC/	−7.60	−0.66	−1.54	−1.89
GGAGCUCU	−10.50	−2.74	−3.54	−4.22
GGAGUUCC	−6.43	0.27	0.02	0.19
GGAUGUCC	−8.39	−1.69	−2.37	−2.42
GGCGGGGC/	−13.80	−4.53	−3.27	−3.05
GGCGUGCC	−9.72	−0.88	−1.78	−1.53
GGCUGGCC	−13.10	−4.26	−4.11	−3.92
GGUUGACC	−8.30	−1.82	−1.59	−2.24
GUAGCUAU	−7.30	−1.58	−1.80	−2.72
GUCGGGCC/	−15.00	−5.46	−3.57	−3.64
GUCGUGAC	−6.05	−1.39	−1.78	−1.53
GUCUAGAU	−7.70	−2.03	−1.80	−2.72
UACCGGUG	−9.70	−0.86	−0.78	−2.00
UAUGCAUGp	−6.44	−1.12	−0.78	−2.00
UCACGUGG	−8.40	−1.86	−3.60	−3.06
UGACGUCG	−10.40	−3.38	−2.50	−2.82
UUACGUAG	−6.20	−1.22	−0.70	−1.10
CAGAGGAGAC/	−9.43	−0.31	−1.11	−1.35
CAGCGCGUUG	−12.31	−4.47	−5.00	−6.12
CAGUCGAUUG	−8.70	−1.94	−2.50	−3.82
CCAGCGUCCU/	−11.60	−3.08	−4.65	−5.83
CCGAAUUUGG	−6.99	−2.03	−3.52	−5.36
CGGAAUUUCG	−7.88	−4.72	−4.62	−5.60
CGGAUAUUCG	−8.78	−5.05	−4.62	−5.60
CGGGCGUUCG	−11.55	−8.99	−8.40	−7.14
CGGUGCAUCG	−14.76	−6.92	−5.40	−5.78
CUGGAUUCAG	−10.15	−5.19	−4.62	−5.60
GAGAGCUUUC	−8.82	−1.06	−1.72	−3.64
GAGGAUCUUC	−9.83	−3.85	−4.24	−5.32
GAGUGGAGAG/	−9.87	−0.78	−1.15	−1.80
GGUUCGGGCC	−13.59	−9.23	−8.40	−7.14
GUGAAUUUAC	−4.78	−1.86	−1.80	−4.54
GUGUGCAUAC	−8.90	−1.30	−2.58	−4.72
GUUAGCUGAC	−8.60	−1.06	−1.80	−4.54
UCGCCAGAGG/	−15.32	−2.37	−3.40	−3.92

[a] Calculated with free energies in Table 3. [b] Calculated with free energies in ref. (*9*).

Table S4 (continued): Component enthalpies of GU pairs.

	T_M^{-1} vs log C_T			
Oligomer	ΔH° (1/T_M vs. ln(C_T/a)) (kcal/mol)	Experimental ΔH°(component) (kcal/mol)	Predicted ΔH°(component)[a] (kcal/mol)	Predicted ΔH°(component)[b] (kcal/mol)
GCUGGUGC/	−69.40	−22.33	−26.17	−29.27
GGAGCUCU	−66.57	−9.46	−18.88	−24.22
GGAGUUCC	−73.10	−25.05	−21.75	−21.01
GGAUGUCC	−73.00	−24.95	−27.42	−26.88
GGCGGGGC/	−76.50	−23.57	−32.82	−31.19
GGCGUGCC	−73.40	−20.47	−24.95	−25.81
GGCUGGCC	−87.20	−34.27	−31.52	−33.48
GGUUGACC	−78.30	−32.33	−33.40	−34.92
GUAGCUAU	−50.30	−2.69	−14.78	−17.62
GUCGGGCC/	−96.00	−36.86	−16.47	−20.44
GUCGUGAC	−69.10	−25.03	−24.95	−25.81
GUCUAGAU	−70.00	−20.08	−14.78	−17.62
UACCGGUG	−51.70	4.91	−1.92	−13.98
UAUGCAUGp	−62.30	−11.39	−1.92	−13.98
UCACGUGG	−46.90	3.81	−14.06	−16.66
UGACGUCG	−63.80	−9.09	−11.12	−11.22
UUACGUAG	−44.60	0.61	−7.92	−6.42
CAGAGGAGAC/	−98.95	−34.88	−32.16	−29.51
CAGCGCGUUG	−77.02	−23.59	−30.10	−31.60
CAGUCGAUUG	−92.33	−39.54	−22.70	−24.04
CCAGCGUCCU/	−87.90	−35.09	−28.00	−30.22
CCGAAUUUGG	−76.76	−30.57	−31.88	−36.88
CGGAAUUUCG	−90.52	−49.83	−34.82	−42.32
CGGAUAUUCG	−88.20	−44.08	−34.82	−42.32
CGGGCGUUCG	−101.66	−73.35	−71.88	−68.78
CGGUGCAUCG	−102.42	−48.99	−28.84	−34.28
CUGGAUUCAG	−97.81	−50.20	−34.82	−42.32
GAGAGCUUUC	−86.57	−29.46	−28.68	−32.08
GAGGAUCUUC	−93.86	−38.33	−26.80	−30.64
GAGUGGAGAG/	−96.93	−32.82	−29.16	−33.29
GGUUCGGGCC	−115.71	−81.90	−71.88	−68.78
GUGAAUUUAC	−62.63	−20.42	−22.68	−39.64
GUGUGCAUAC	−58.60	−3.65	−16.70	−31.60
GUUAGCUGAC	−69.60	−14.57	−22.68	−39.64
UCGCCAGAGG/	−93.46	−17.73	−16.65	−18.20

[a] Calculated with enthalpies in Table 3. [b] Calculated with enthalpies in ref. (*9*).

.

Table S5: Comparison of measured and predicted $\Delta G°_{37}$ values for terminal GU pairs. Component $\Delta G°_{37}$ calculated by subtracting the $\Delta G°_{37}$ of each non-GU INN parameter from the $1/T_M$ vs. $\ln(C_T/a)$-fitted $\Delta G°_{37}$ of a duplex with a terminal GU pair. In brackets are INN parameters fitted from linear regression. In parentheses are the component $\Delta G°_{37}$ calculated by subtracting the $1/T_M$ vs. $\ln(C_T/a)$-fitted $\Delta G°_{37}$ of each non-GU-containing sequence from the $1/T_M$ vs. $\ln(C_T/a)$-fitted $\Delta G°_{37}$ of a duplex with a terminal GU pair.

Doublet motif:
5´WY
3´XZ

WX/YZ	GU	UG
AU	−0.58±0.05 [−0.35] (not available)	−0.93±0.12 [−0.90] (−1.38)
CG	−1.63±0.08 [−1.25] (−1.56±0.05)	−1.62±0.35 [−1.77] (−1.58±0.46)
GC	−1.24±0.43 [−1.80] (−1.18±0.38)	−1.96±0.63 [−2.15] (−2.06±0.35)
UA	−0.50±0.09 [−0.39] (−0.60)	−0.73±0.04 [−0.51] (not available)

The standard deviation of the differences between the component $\Delta G°_{37}$ and that obtained from linear regression was calculated to be 0.30 for all motifs.

Table S6: The number of occurrences of each triplet, the component $\Delta G°_{37}$, and in brackets, the sum of the $\Delta G°_{37}$'s of each INN in triplet. $\Delta G°_{37}$ for each INN triplet was taken from Table 3. No error limit is listed when only a single duplex was measured.

Triplet motif:
5´WGY
3´XUZ

WX/YZ	AU	CG	GC	UA
AU	2 −0.53 [−0.86]	2 −2.24 [−2.50]	2 −1.91 [−2.12]	4 −1.58±0.86 [−1.25]
CG	2 −1.02 [−1.76]	4 −2.58±0.30 [−3.40]	4 −3.02±0.44 [−3.02]	2 −2.77 [−2.15]
GC	8 −2.48±0.13 [−2.31]	2 −3.42 [−3.95]	3 −3.78±1.47 [−3.57]	3 −3.21±0.35 [−2.70]
UA	4 −0.74±0.29 [−0.90]	2 −3.07 [−2.54]	2 −2.01 [−2.16]	2 −0.65 [−1.29]

Table S7: Free energy differences when GC pairs are replaced with AU pairs.

GC duplex	$-\Delta G^\circ_{37}$ GC expt.	$-\Delta G^\circ_{37}$ GC pred.	Ref.	AU duplex	$-\Delta G^\circ_{37}$ AU expt.	$-\Delta G^\circ_{37}$ AU pred.	Ref.	$\Delta\Delta G^\circ_{37}$ AU expt.	$\Delta\Delta G^\circ_{37}$ AU pred.	$\Delta\Delta G^\circ_{37}$ expt. per AU pair
CGCGCGp	9.12	9.40	(*10*)	CACGUG	6.59	6.54	(*1*)	2.53	2.86	1.27
CCAUGG	7.30	7.32	(*11*)	UCAUGA	4.31	4.60	(*1*)	2.99	2.72	1.50
CCGCGGp	9.84	10.14	(*1*)	UCGCGA	6.85	7.42	(*1*)	2.99	2.72	1.50
CCGCGGp	9.84	10.14	(*1*)	CUGCAGp	7.11	7.28	(*12*)	2.73	2.86	1.37
CCGCGGp	9.84	10.14	(*1*)	CCAUGG	7.30	7.32	(*11*)	2.54	2.82	1.27
CCUAGG	7.80	7.49	(*13*)	UCUAGA	4.95	4.77	(*1*)	2.85	2.72	1.43
CGCGCGp	9.12	9.40	(*10*)	UGCGCA	8.22	8.00	(*10*)	0.90	1.40	0.45
CGCGCGp	9.12	9.40	(*10*)	CACGUG	6.59	6.54	(*1*)	2.53	2.86	1.27
CGGCCGp	9.90	10.14	(*14*)	CAGCUGp	6.68	7.28	(*12*)	3.22	2.86	1.61
GCGCCG/	10.91	10.73	(*1*)	GCAUCG/	7.26	7.25	(*1*)	3.65	3.48	1.83
GCCGCG/	10.88	10.73	(*1*)	GCUACG/	7.56	7.34	(*1*)	3.32	3.39	1.66
GCCGGCp	11.20	11.20	(*14*)	GUCGAC	7.09	7.02	(*15*)	4.11	4.18	2.06
GCCGGCp	11.20	11.20	(*14*)	ACCGGUp	8.51	7.94	(*16*)	2.69	3.26	1.35
GCCGGCp	11.20	11.20	(*14*)	GCUAGC	7.92	7.81	(*17*)	3.28	3.39	1.64
GCGCCG/	10.91	10.73	(*1*)	GUGUCG/	7.18	7.21	(*1*)	3.73	3.52	1.87
GCGCGG/	11.38	10.73	(*1*)	GUGCAG/	7.67	7.87	(*1*)	3.71	2.86	1.86
GCGCGCp	10.62	10.46	(*14*)	GCAUGC	7.38	7.64	(*12*)	3.24	2.82	1.62
GCGGCG/	10.40	10.73	(*18*)	GCAACG/	7.01	6.97	(*1*)	3.39	3.76	1.70
GCGGCG/	10.40	10.73	(*18*)	GUGGUG/	7.67	7.87	(*1*)	2.73	2.86	1.37
GCGUCG/	8.76	8.64	(*1*)	GUGUCG/	7.18	7.21	(*1*)	1.58	1.43	1.58
GCGUCG/	8.76	8.64	(*1*)	GCAUCG/	7.26	7.25	(*1*)	1.50	1.39	1.50
GGCGCCp	11.33	11.20	(*10*)	GGUACC	7.35	7.81	(*19*)	3.98	3.39	1.99
GGCGCCp	11.33	11.20	(*10*)	AGCGCU	7.99	7.94	(*10*)	3.34	3.26	1.67
GGCGCCp	11.33	11.20	(*10*)	GACGUC	7.35	7.02	(*15*)	3.98	4.18	1.99
UCCGGAp	7.79	8.16	(*20*)	UCUAGA	4.95	4.77	(*1*)	2.84	3.39	1.42
GUCUAGAC	10.11	10.15	(*12*)	AUCUAGAU	7.20	6.97	(*15*)	2.91	3.18	1.46
AUGCGCAUp	10.17	10.20	(*17*)	AUACGUAU	6.53	6.28	(*17*)	3.64	3.92	1.82
CAUGCAUGp	9.67	9.54	(*21*)	UAUGCAUAp	7.27	7.08	(*21*)	2.40	2.46	1.20
GAUGCAUCp	10.12	10.02	(*21*)	AAUGCAUUp	7.18	6.28	(*21*)	2.94	3.74	1.47
GAUGCAUCp	10.12	10.02	(*21*)	GAUAUAUC	6.09	6.14	(*12*)	4.03	3.88	2.02
GCUGCAGC	14.50	14.34	(*4*)	ACUAUAGU	6.98	6.98	(*15*)	7.52	7.36	1.88
GUCUAGAC	10.11	10.15	(*12*)	AUCUAGAU	7.20	6.97	(*15*)	2.91	3.18	0.73
Average										1.54±0.35

Table S8: INN parameters for GU base pairs in 1 M NaCl with a separate parameter for terminal GU pairs.

INN	Occurrence	$\Delta G°_{37}$ (kcal/mol)	$\Delta G°_{37}$ error (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta H°$ error (kcal/mol)	$\Delta S°$[a] (eu)	$\Delta S°$ error (eu)
5´GU3´ 3´UG5´	8	0.72	0.19	−12.76	4.25	−43.5	13.1
5´GG3´ 3´UU5´	9	−0.25	0.17	−17.01	3.77	−54.0	11.6
5´AG3´ 3´UU5´	22	−0.35	0.08	−4.69	1.77	−14.0	5.4
5´UG3´ 3´AU5´	18	−0.39	0.09	−1.42	1.81	−3.3	5.6
5´UU3´ 3´AG5´	26	−0.50	0.08	−10.57	1.79	−32.5	5.5
5´UG3´ 3´GU5´	10	−0.58	0.19	−11.96	4.02	−36.7	12.4
5´AU3´ 3´UG5´	24	−0.89	0.08	−7.97	1.67	−22.8	5.1
5´CG3´ 3´GU5´	26	−1.25	0.09	−5.96	1.69	−15.2	5.2
5´CU3´ 3´GG5´	21	−1.77	0.10	−9.70	1.77	−25.6	5.4
5´GG3´ 3´CU5´	24	−1.80	0.09	−7.62	1.78	−18.8	5.4
5´GU3´ 3´CG5´	25	−2.14	0.10	−11.34	1.78	−29.6	5.4
5´GGUC3´[b] 3´CUGG5´	3	−4.12	0.54	−30.8	8.87	−86.0	23.7
terminal GU penalty	34	−0.02	0.06	2.34	1.17	7.6	3.6

[a] Values for $\Delta S°$ were derived from $\Delta S° = (\Delta H° - \Delta G°_{37})/310.15$. [b] Ref. (*9*).

Table S9: Probability density function (PDF) of Student's t-distribution for $\Delta G°_{37}$ and $\Delta H°$ for each GU INN motif with a separate parameter for terminal GU pairs.[a] The 5´GGUC/3´CUGG motif is not included.

Motif	St(t;f) $\Delta G°_{37}$	St(t;f) $\Delta H°$	Motif	St(t;f) $\Delta G°_{37}$	St(t;f) $\Delta H°$
5´GU3´ 3´UG5´	5.8×10^{-4}	5.6×10^{-3}	5´AU3´ 3´UG5´	9.2×10^{-16}	2.3×10^{-5}
5´GG3´ 3´UU5´	1.3×10^{-1}	5.6×10^{-5}	5´CG3´ 3´GU5´	6.8×10^{-20}	1.3×10^{-3}
5´AG3´ 3´UU5´	8.8×10^{-5}	1.4×10^{-2}	5´CU3´ 3´GG5´	6.5×10^{-25}	1.8×10^{-6}
5´UG3´ 3´AU5´	1.0×10^{-4}	2.9×10^{-1}	5´GG3´ 3´CU5´	1.3×10^{-27}	1.2×10^{-4}
5´UU3´ 3´AG5´	1.0×10^{-7}	3.7×10^{-7}	5´GU3´ 3´CG5´	4.0×10^{-29}	6.3×10^{-8}
5´UG3´ 3´GU5´	4.9×10^{-3}	6.0×10^{-3}	terminal GU penalty	3.8×10^{-1}	5.6×10^{-2}

[a] Refs. (*1*) and (*22*).

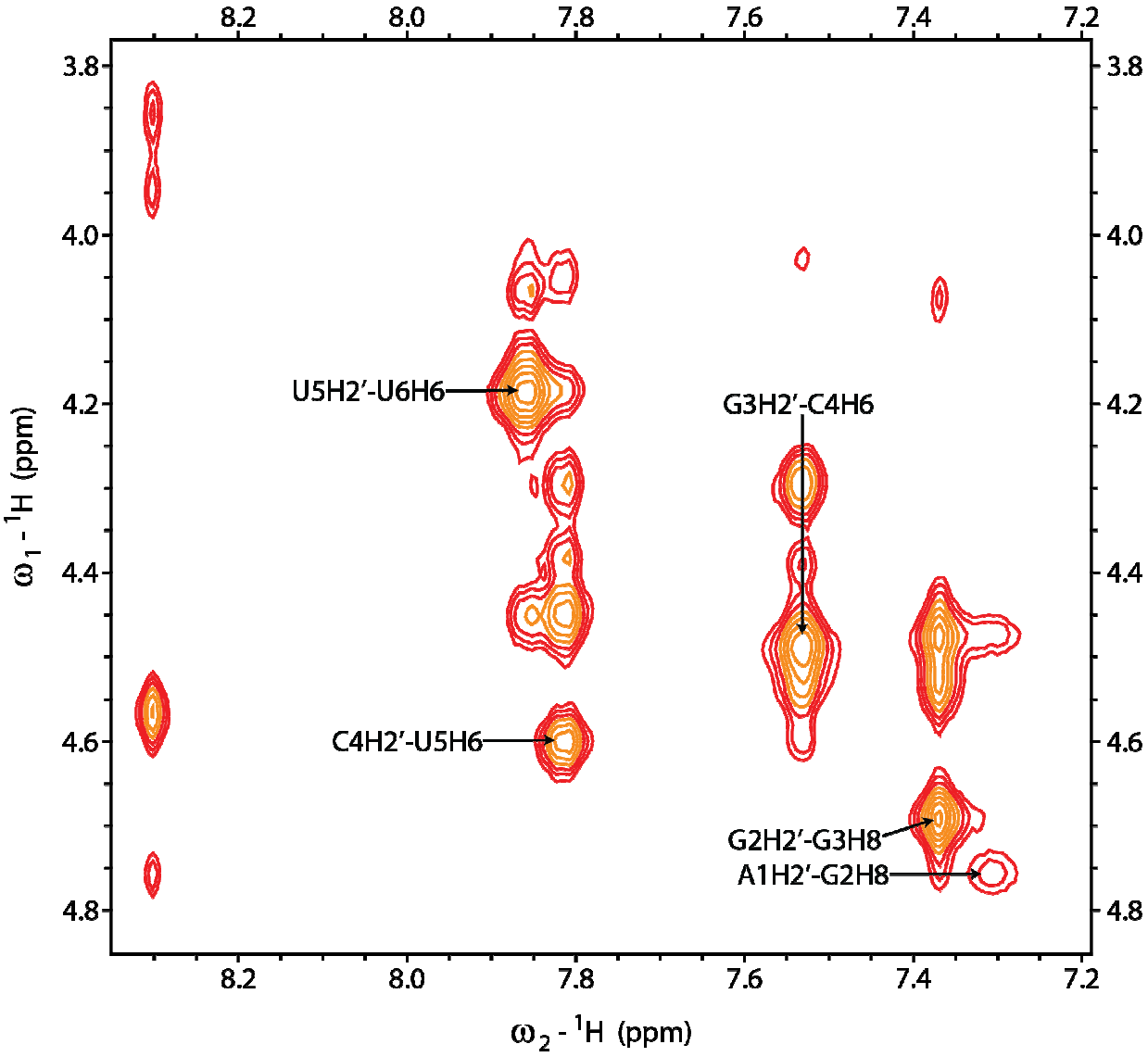

Figure S1: WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 100 ms mixing time at 0 °C showing the H2´-H6/H8 region.

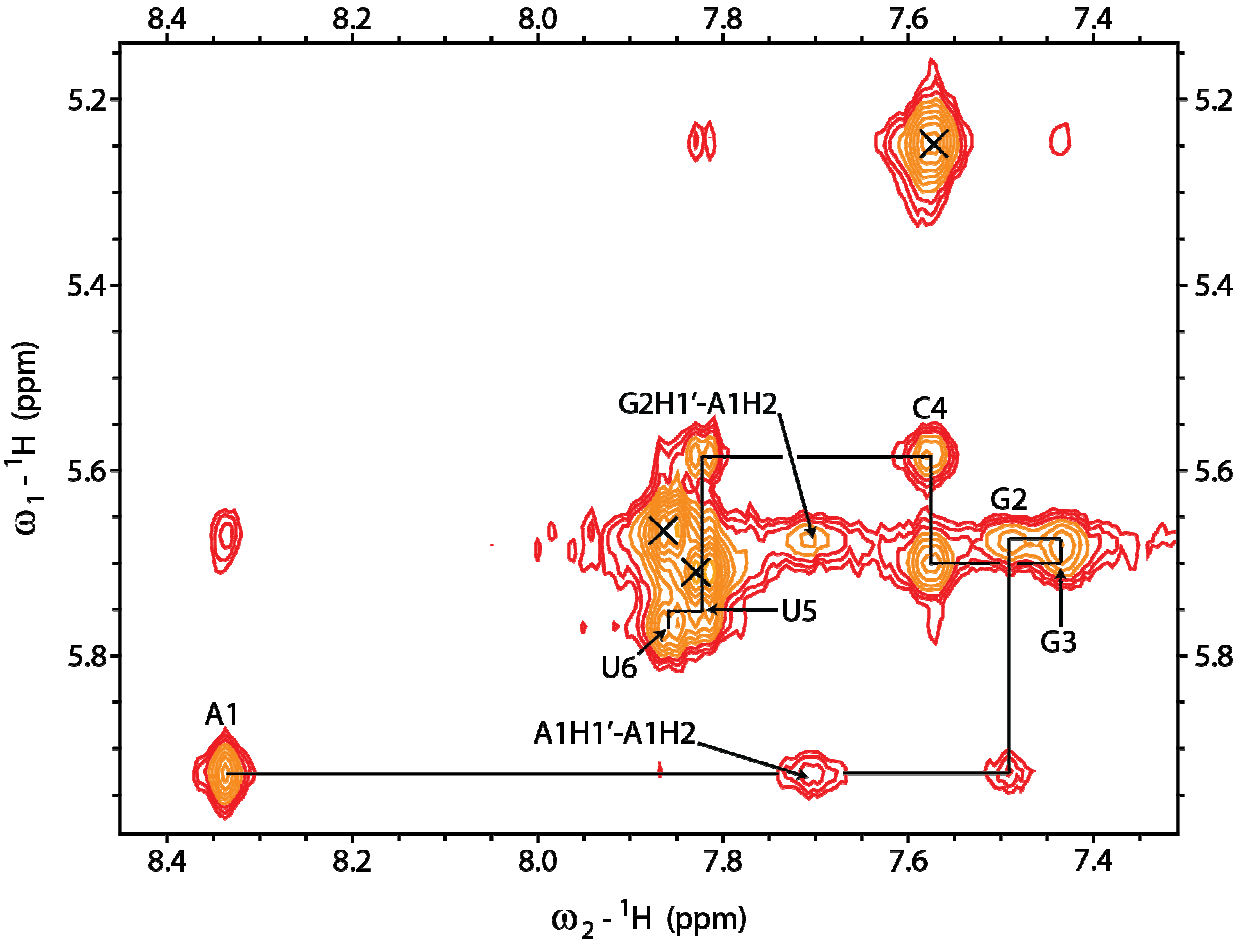

Figure S2: WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 400 ms mixing time at 8 °C showing the H6/H8/H2/H5/H1´ region. C and U H5-H6 NOEs are labeled with crosses.

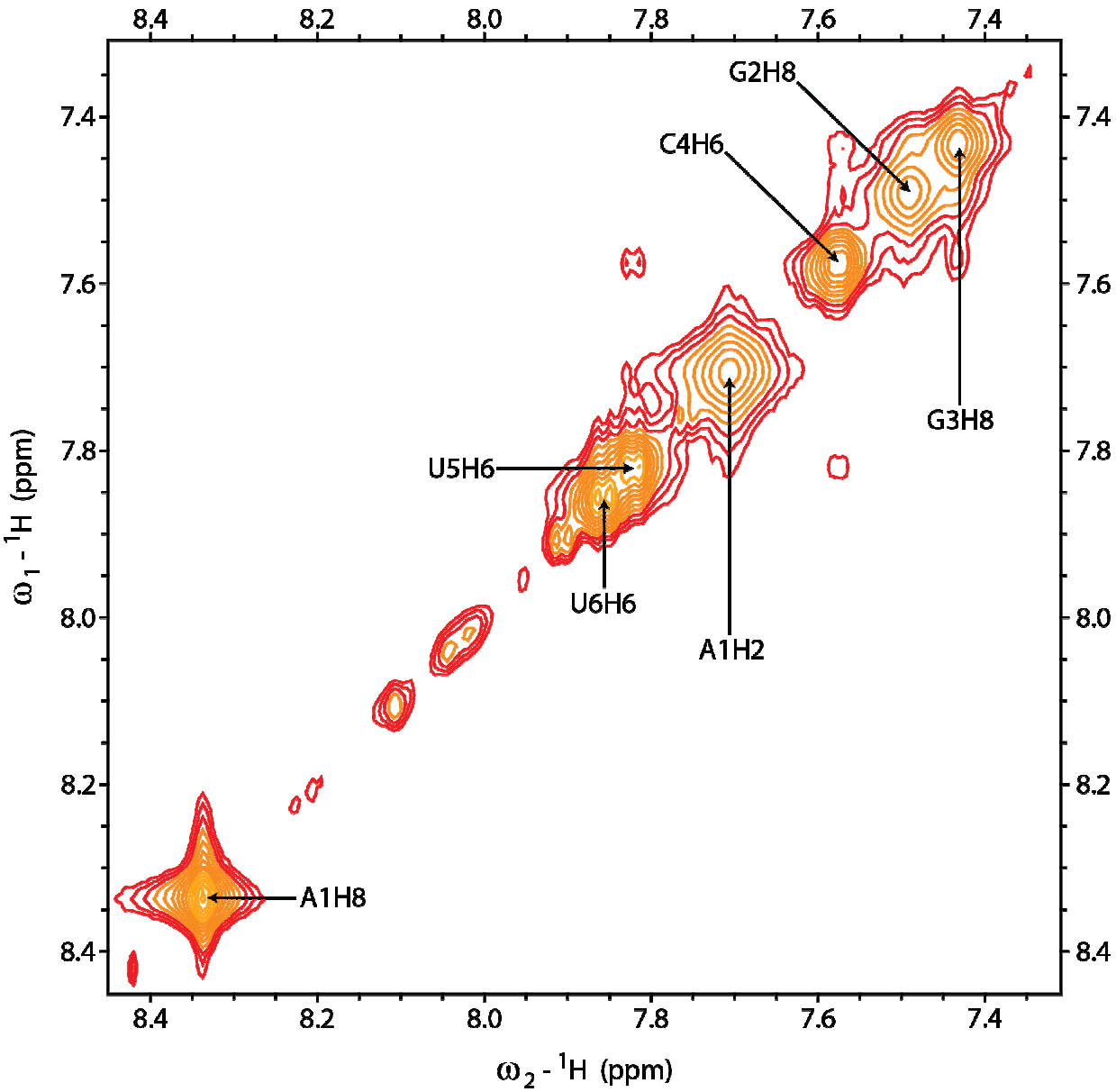

Figure S3: WATERGATE NOESY spectrum of r(A<u>G</u>GC<u>U</u>U)$_2$ with 400 ms mixing time at 8 °C showing the H2/H6/H8 region.

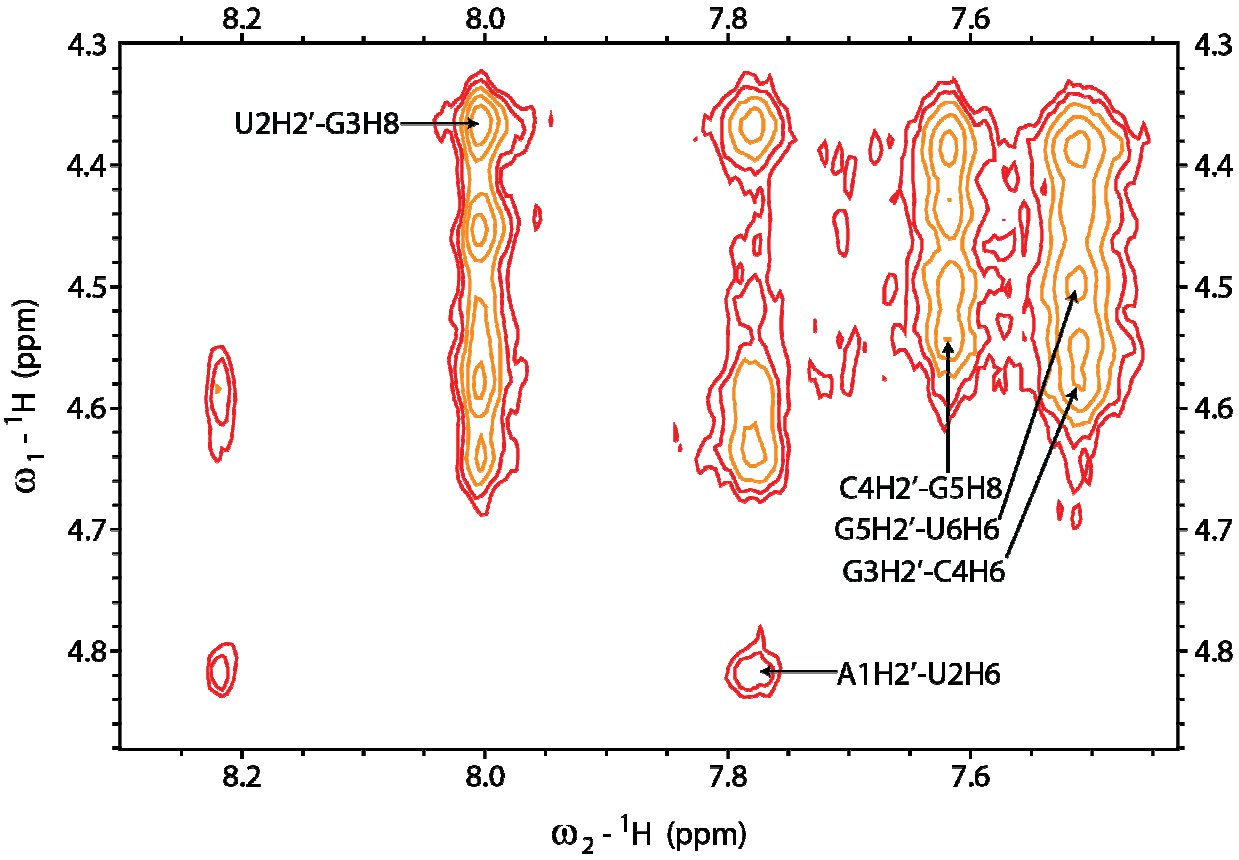

Figure S4: NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H2´-H6/H8 region.

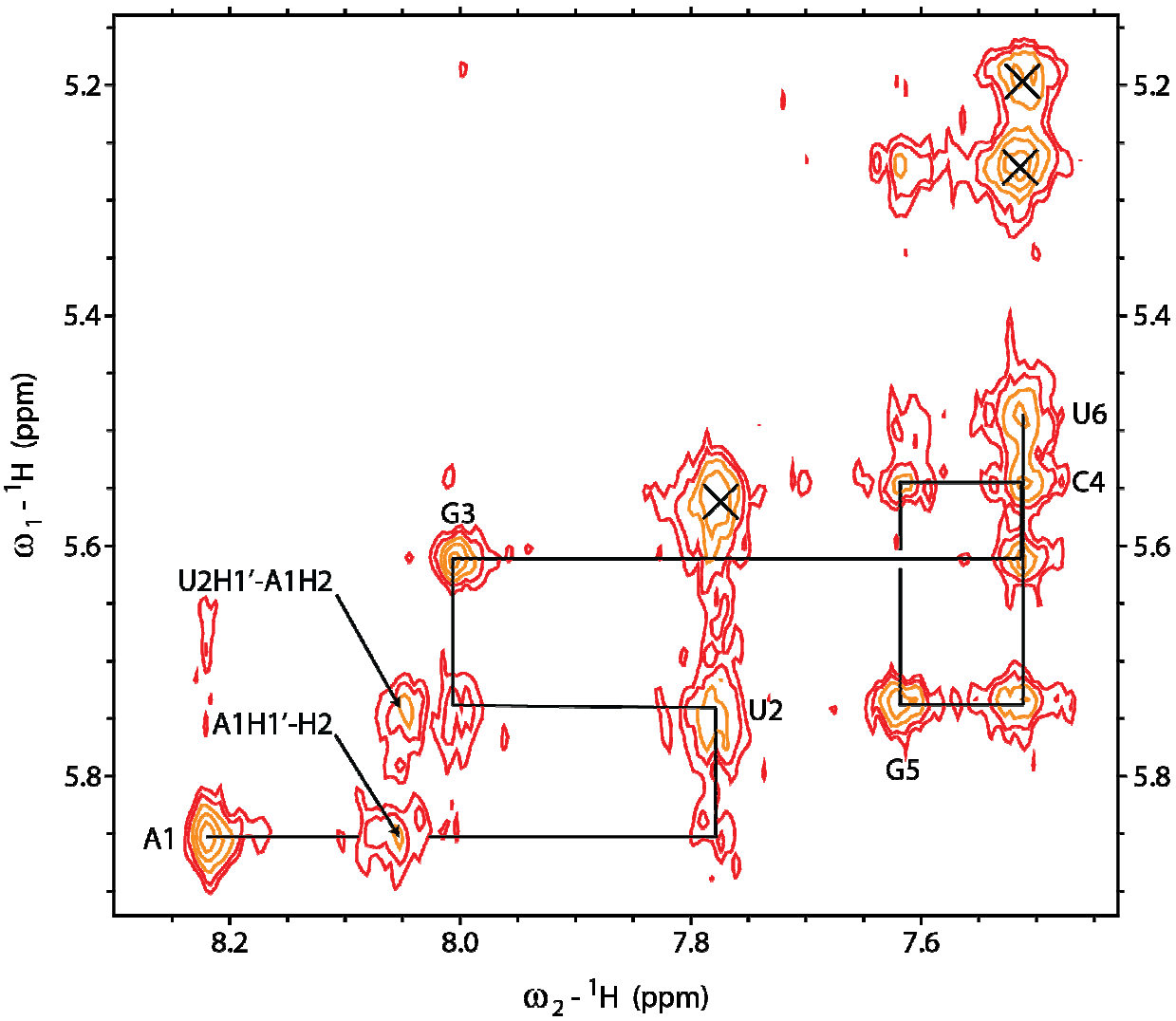

Figure S5: NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H6/H8/H2/H5/H1´ region. C and U H5-H6 NOEs are labeled with crosses.

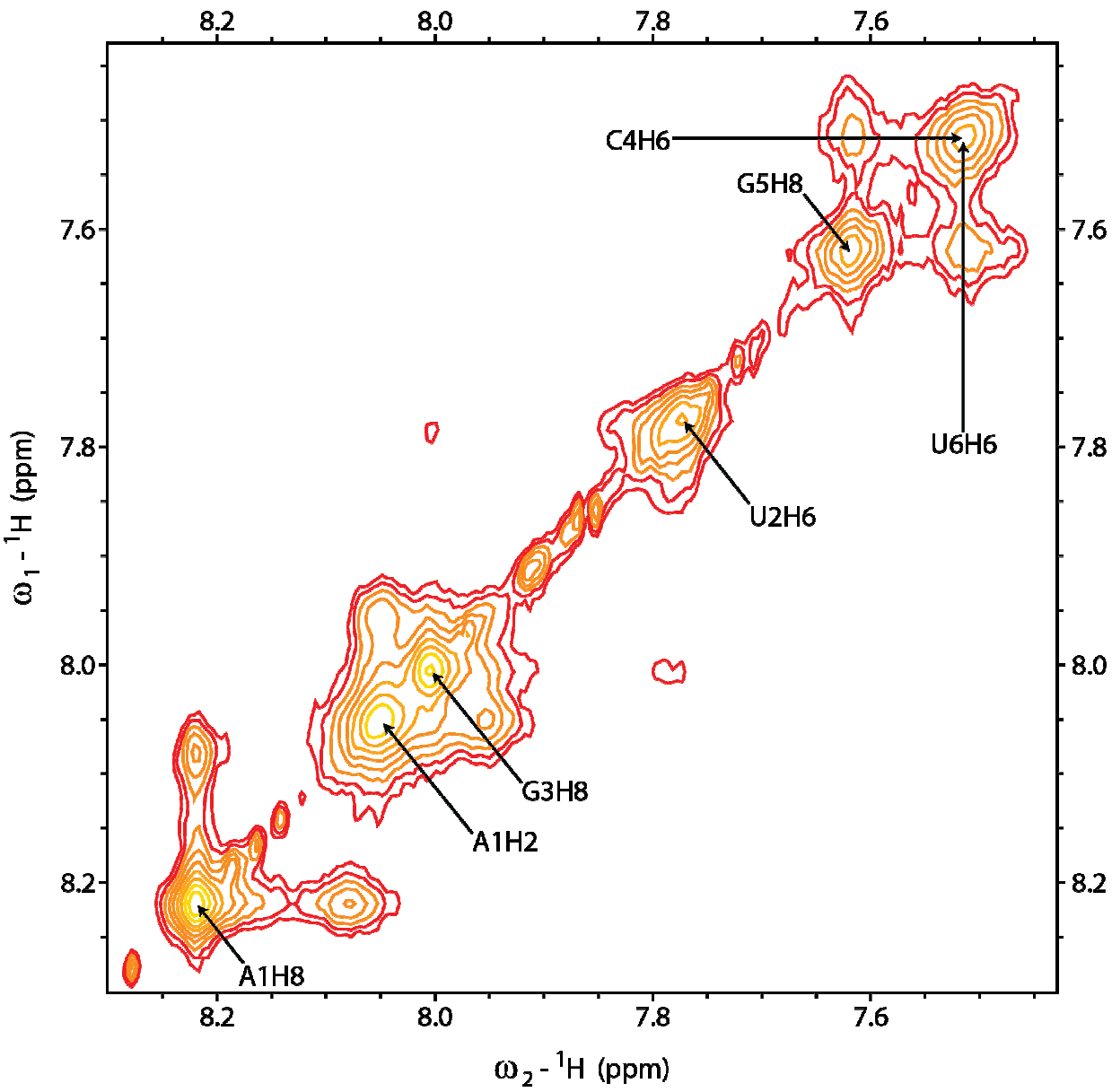

Figure S6: NOESY spectrum of r(A<u>U</u>GC<u>G</u>U)$_2$ with 400 ms mixing time at 0 °C showing the H2/H6/H8 region.

Sep-Pak desalting of RNA oligomers. To desalt each of those sequences, a Sep-Pak C18 cartridge was washed with 5 mL of stock CH_3CN, followed by 2 mL of 20X buffer (0.1 M ammonium bicarbonate, adjusted to pH 4.5 with acetic acid) and three 2 mL aliquots of 20X buffer diluted 11-fold. The RNA sample was dissolved in 2 mL of 20X buffer and added to the column. The filtrate from the column was collected and checked for RNA content with silica gel plates under UV illumination. The column was washed with 2 mL of the 20X buffer that was diluted 11-fold; the RNA was eluted with 30% acetonitrile and dried under vacuum.

REFERENCES

1. Xia, T. B., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X. Q., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs, *Biochemistry 37*, 14719-14735.
2. Burkard, M. E., Xia, T., and Turner, D. H. (2001) Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another noncanonical pair, *Biochemistry 40*, 2478-2483.
3. Testa, S. M., Disney, M. D., Turner, D. H., and Kierzek, R. (1999) Thermodynamics of RNA-RNA duplexes with 2-or 4-thiouridines: Implications for antisense design and targeting a group I intron, *Biochemistry 38*, 16655-16662.
4. Burkard, M. E., and Turner, D. H. (2000) NMR structures of r(GCA(G)GC(G)UGC)(2) and determinants of stability for single guanosine-guanosine base pairs, *Biochemistry 39*, 11748-11762.
5. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257-9274.
6. Chen, X. Y., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G·U pairs: One hydrogen bond for each G·U Pair in r(GGCGUGCC)2 and two for each G·U Pair in r(GAGUGCUC)2, *Biochemistry 39*, 8970-8982.
7. Chen, G., Znosko, B. M., Jiao, X. Q., and Turner, D. H. (2004) Factors affecting thermodynamic stabilities of RNA 3x3 internal loops, *Biochemistry 43*, 12865-12876.
8. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)(2) by two-dimensional NMR and simulated annealing, *Biochemistry 35*, 14077-14089.
9. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure, *J. Mol. Biol. 288*, 911-940.
10. Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., and Turner, D. H. (1986) Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids, *Biochemistry 25*, 3214-3219.
11. McDowell, J. A., He, L. Y., Chen, X. Y., and Turner, D. H. (1997) Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUCC)(2) and (rGGAUGUCC)(2), *Biochemistry 36*, 8030-8038.

12. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proc. Natl. Acad. Sci. USA 83*, 9373-9377.
13. He, L., Kierzek, R., SantaLucia, J., Walter, A. E., and Turner, D. H. (1991) Nearest-neighbor parameters for G·U mismatches - 5'GU3'/3'UG5' is destabilizing in the contexts CGUG/GUGC, UGUA/AUGU, and AGUU/UUGU but stabilizing in GGUC/CUGG, *Biochemistry 30*, 11124-11132.
14. Freier, S. M., Sinclair, A., Neilson, T., and Turner, D. H. (1985) Improved free energies for G·C base-pairs, *J. Mol. Biol. 185*, 645-647.
15. Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., and Freier, S. M. (1986) Polymer-supported RNA synthesis and its application to test the nearest-neighbor model for duplex stability, *Biochemistry 25*, 7840-7846.
16. Petersheim, M., and Turner, D. H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry 22*, 256-263.
17. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helixes, *Biochemistry 25*, 5755-5759.
18. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278-285.
19. Wu, M., McDowell, J. A., and Turner, D. H. (1995) A periodic table of symmetric tandem mismatches in RNA, *Biochemistry 34*, 3204-3211.
20. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes, *Biochemistry 24*, 4533-4539.
21. Sugimoto, N., Kierzek, R., and Turner, D. H. (1987) Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid, *Biochemistry 26*, 4554-4558.
22. Crawley, M. J. (2007) *The R Book*, 1 ed., John Wiley & Sons, West Sussex.