Advancing viral RNA structure prediction: measuring the thermodynamics of pyrimidine-rich internal loops

ANDY PHAN,[1,3] KATHERINE MAILEY,[1,3] JESSICA SAEKI,[1,3] XIAOBO GU,[1,2] and SUSAN J. SCHROEDER[1,2]

[1]Department of Chemistry and Biochemistry, [2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, Oklahoma 73019, USA

ABSTRACT

Accurate thermodynamic parameters improve RNA structure predictions and thus accelerate understanding of RNA function and the identification of RNA drug binding sites. Many viral RNA structures, such as internal ribosome entry sites, have internal loops and bulges that are potential drug target sites. Current models used to predict internal loops are biased toward small, symmetric purine loops, and thus poorly predict asymmetric, pyrimidine-rich loops with >6 nucleotides (nt) that occur frequently in viral RNA. This article presents new thermodynamic data for 40 pyrimidine loops, many of which can form UU or protonated CC base pairs. Uracil and protonated cytosine base pairs stabilize asymmetric internal loops. Accurate prediction rules are presented that account for all thermodynamic measurements of RNA asymmetric internal loops. New loop initiation terms for loops with >6 nt are presented that do not follow previous assumptions that increasing asymmetry destabilizes loops. Since the last 2004 update, 126 new loops with asymmetry or sizes greater than 2 × 2 have been measured. These new measurements significantly deepen and diversify the thermodynamic database for RNA. These results will help better predict internal loops that are larger, pyrimidine-rich, and occur within viral structures such as internal ribosome entry sites.

Keywords: internal ribosome entry site; viral RNA; thermodynamics; pyrimidine-rich internal loops; RNA secondary structure prediction

INTRODUCTION

Amidst a flood of sequence information, lightening-fast development of RNA therapeutics, whirlwind discoveries of novel RNA gene regulation functions, and the thunderous threat of outbreaks from viral RNA pathogens such as Ebola and Zika, RNA thermodynamics provide a compass to navigate the storm. Accurate thermodynamic measurements of RNA motifs form the core for a wide array of approaches to predicting RNA structure and function from sequence (for example, Xu et al. 2014; Eggenhofer et al. 2016; Xu and Mathews 2016). These prediction tools are a critical resource for responding to the deluge of sequence information generated by ever-advancing technology. For example, a conference on the challenges of RNA structure prediction lists as the top priority improvements in thermodynamics measurements: "Nearest-neighbor interactions are robust parameters that represent a central anchor for all of secondary structure prediction. However, the context dependence of these terms and their conditional variation make it important to continually improve them in order to advance the power of prediction algorithms" (Pyle and Schlick 2016). Accurate thermodynam-

ic parameters also enable better three-dimensional structure prediction. For example, only the prediction tool that included the most updated thermodynamic parameters for GU pairs accurately predicted the structure of four consecutive terminal GU pairs (Gu et al. 2015). Thus, we present 40 new thermodynamic parameters for asymmetric loop motifs that are underrepresented in the thermodynamic database but frequently occur in viral RNA. We also present an updated database and prediction rules for all 194 free energy measurements of RNA asymmetric internal loops. The database of thermodynamic parameters is a foundation upon which prediction algorithms build to provide a launch pad for hypotheses on novel noncoding RNA function, screening of viral RNA drug target sites, and design of RNA therapeutics.

Although rigorous measurement of the thermodynamic stabilities for all possible internal loop sequences is not realistic, a diverse and deep sampling of possible sequences can provide a foundation for estimating the thermodynamic stabilities of loop sequences and thus aid RNA structure prediction tools. Accurately predicting secondary structure from sequence is an integral step in understanding RNA structure and function. Internal loops occur frequently within RNA

---

structure and often form part of enzyme active sites, protein interactions, metal ion binding sites, or drug targets. The current model used to predict internal loops and RNA secondary structure was last updated in 2004 and has an accuracy of ~76% when compared to known RNA structures that have been determined through chemical modification or phylogenetic analysis (Mathews et al. 2004). This model predicts well small loops of <6 nt, but inaccuracies increase with loops of larger sizes that had not yet been studied in the 2004 database. The 2004 thermodynamic database is biased toward smaller, purine-rich internal loops. Symmetric, 2 × 2 loops (a loop that has 2 nucleotides [nt] opposite 2 nt) are overrepresented in the thermodynamic database, whereas loops that are pyrimidine-rich or larger than size 3 × 3 are underrepresented. Opposite trends are observed in loops that naturally occur in internal ribosome entry sites (IRES) (Mokrejs et al. 2010). The 2004 model surveyed RNA secondary structures from different species of bacteria, yeast, and mammals, but did not include any viral RNA structures (Mathews et al. 2004). Thus, new viral RNA structures and new measurements on more diverse loops can improve RNA structure prediction tools.

Many pathogenic viruses such as HIV, polio, foot and mouth, and hepatitis use internal ribosome entry sites (IRES) to produce viral proteins (Plank and Kieft 2012; Lozano and Martinez-Salas 2015). Internal loops and bulges that occur within the Hepatitis C IRES have been extensively studied as known drug target sites (Kikuchi et al. 2005; Davis and Seth 2011; Dibrov et al. 2014). In vitro studies and crystal structures demonstrate that small molecules such as benzimidazole bind to the loop motifs within the HCV IRES, induce a conformational change, and effectively inhibit viral translation (Paulsen et al. 2010; Dibrov et al. 2012). Current thermodynamic parameters poorly predict these loop motifs within viral IRESs, and thus more accurate RNA structure predictions can guide better design of viral therapeutics and selection of drug target sites.

The differences between the thermodynamic database and the IRES database may be a contributing factor in the inaccuracies of the current prediction model (Mathews et al. 2004; Mokrejs et al. 2010). This article presents new thermodynamic parameters for 40 loops and significantly diversifies the thermodynamic database. The NMR data show that pyrimidine-rich loops are stabilized by multiple UU pairs. A revised model for the prediction of asymmetric loops is proposed that includes bonus terms for UU pairs in the middle of loops and protonated CC pairs. In addition, new initiation terms for loops consisting of 6, 7, and 8 nt are presented from analysis of an updated database of 507 internal loop measurements.

RESULTS AND DISCUSSION

Database analysis

All the thermodynamic data incorporated in the 2004 prediction model (Mathews et al. 2004) were compiled and updated

with new data for 126 RNA loops (Chen et al. 2004, 2006, 2009; Bourdelat-Parks and Wartell 2005; Badhwar et al. 2007; Christiansen and Znosko 2008, 2009; Hausmann and Znosko 2012; Zhong et al. 2015). A complete spreadsheet of loop thermodynamic parameters is available in the Supplemental Information. Prior to this study, the thermodynamic database contained a total of 469 internal loops. The IRES database consists of 107 internal loops from viruses and eukaryotic mRNAs whose RNA structures have been determined through chemical modification and phylogenetic analysis (Mokrejs et al. 2010). Analysis of the databases compared the size of the loops, the nucleotide content within the loops, and loop symmetry (Fig. 1). Forty-two percent of the thermodynamic database consisted of 2 × 2 loops. In contrast, 2 × 2 loops only comprised 15% of the IRES database. The IRES database had a more even distribution in terms of loop size. Loops with >6 nt were the largest category in the IRES database (17% $n \times n$ loops where n is greater than three). Fifty percent of the loops within the thermodynamic
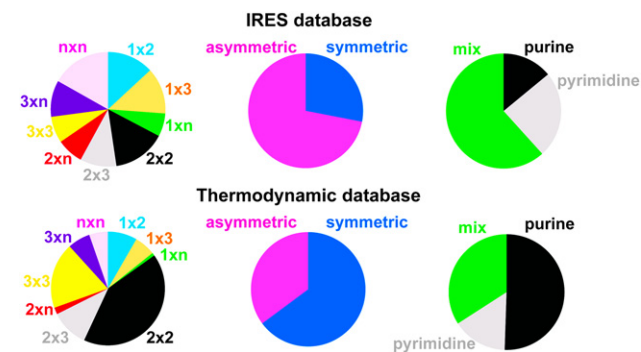

FIGURE 1. A comparison of the internal loop characteristics in viral IRES and the thermodynamic database for RNA structure prediction. Data for RNA internal loops are from an IRES database (Mokrejs et al. 2010) (*top* row). The RNA secondary structures in the IRES database were experimentally determined from phylogeny and chemical or enzymatic probing. Analysis for RNA internal loops in the thermodynamic database (Santa Lucia et al. 1990, 1991; Peritz et al. 1991; Walter et al. 1994; Wu et al. 1995; Schroeder et al. 1996; Xia et al. 1997; Schroeder and Turner 2000, 2001; Burkard et al. 2001; Chen et al. 2004, 2005, 2006, 2009; Bourdelat-Parks and Wartell 2005; Chen and Turner 2006; Badhwar et al. 2007; Christiansen and Znosko 2008, 2009; Hausmann and Znosko 2012) (*bottom* row). Loop sizes are described by the number of nucleotides on each side of a loop. For example, a 1 × 3 loop has 1 nt opposite 3 nt. *N* is any number >3 nt. When the number of nucleotides on each side of the loop is different, then the loop is described as asymmetric. A loop with both pyrimidine and purine nucleotides is described as a mix. The IRES database and thermodynamic database contain 107 and 469 total loops, respectively. Note that the thermodynamic database does not include the new measurements presented in this work. For comparisons of loop size, 1 × 2 loops are shown in blue; 1 × 3 loops in peach; 1 × *n* loops in green; 2 × 2 loops in black; 2 × 3 loops in gray; 2 × *n* loops in bright pink; 3 × 3 loops in yellow; 3 × *n* loops in purple; and *n* × *n* loops in pastel pink. For comparisons of loop symmetry, symmetric loops are shown in blue, and asymmetric loops are shown in pink. For comparisons of nucleotide content, purine-only loops are shown in black, pyrimidine-only loops in gray; and loops with a mix of purine and pyrimidines in green.

database contained only purines. In contrast, only 14% of the loops within the IRES database were purines. Sixty-two percent of the loops were a mixture of purines and pyrimidines in the IRES database. Lastly, the distribution of symmetric and asymmetric loops in both databases was completely opposite. Sixty-five percent of the loops within the thermodynamic database were symmetric, primarily due to the substantial number of 2×2 loops. In contrast, 72% of the IRES loops were asymmetric. The differences between the two databases may contribute to the limited accuracy of predictions for viral RNA structures (Mathews et al. 2004). Additional thermodynamic data for large, asymmetric pyrimidine-rich loops will improve predictions for viral IRES RNA.

New thermodynamic data diversifies the database

In order to expand the thermodynamic database to better represent loops that occur in virus IRES, additional thermodynamic data on RNA duplexes containing large, asymmetric, pyrimidine loops were collected. Table 1 shows the optical melting data of internal loops that are pyrimidine rich and that have sequences modeled on viral IRES structures. Both the linear van't Hoff plot and melt curve fits data in Table 1 are shown in order to validate two-state behavior. Thermodynamic data for internal loops studied at pH 5.5, a pH at which cytidines become protonated, are listed below the data for pH 7 buffer conditions.

As shown in Table 1 column 3, loops with the potential to form UU pairs are more stable than those without possible UU pairs at pH 7. Furthermore, loops that contain more than one possible UU pair are more stable than loops that only contain one possible pair. The range of free energies for all the new loops studied was −0.1 kcal/mol to 5.4 kcal/mol. In general, loop stability decreased with increasing loop size, and uracil loops were discovered to be more stable than cytidine loops of the same size. For example, the measured free energy at 37°C of the 3×4 all-uracil loop was found to be 2.16 kcal/mol in contrast to the 3×4 all-cytidine loops that had a free energy value of 4.0 kcal/mol. The 2×4 uracil loop (duplex xviii) was the most stable of all the loops studied, with a free energy of −0.1 kcal/mol. Thus, even though loop stability generally becomes less favorable with increasing loop size and asymmetry, UU pairs provide adequate stability to overcome the energetic costs of larger, asymmetric loops. For example, the largest all-uracil 4×4 loop ($\Delta G_{loop} = 1.6$ kcal/mol, duplex vi) is more stable than the smallest all-cytidine 2×2 loop ($\Delta G_{loop} = 2.6$ kcal/mol, duplex xl).

The loop free energies determined experimentally were then compared to current predicted values (Supplemental Table 1; Mathews et al. 2004). The current prediction model generally overpredicts the stabilities of pyrimidine loops at pH 7. The predicted loop free energies for 13 of the 40 new loops measured at pH 7 are different from experimental values by >1.0 kcal/mol, which is greater than the experimental

error range of ±0.5 kcal/mol. The largest prediction error is 2.5 kcal/mol for an all cytidine 4×5 loop (duplex iii), which is a larger loop than any previous loops measured. Thus, these new measurements of internal loops will improve the accuracy of future predictions and diversify the thermodynamic database.

Uracil base-pairing contributes to loop stability

Although the 2004 prediction model (Mathews et al. 2004) gives a bonus for UU pairing that occurs adjacent to closing base pairs, it does not account for UU mismatches that can occur across or within the middle of asymmetric internal loops. In order to determine whether UU base-pairing occurs within internal loops, sequences containing potential UU mismatches within internal loops were designed and studied through optical melting and one-dimensional imino proton NMR. For example, some of the loops tested if UU noncanonical pairing would occur across asymmetric internal loops. The duplex 5′UGAC UUU CUGA/3′ACUG UUUU GACU (duplex xix) has the potential for a total of three noncanonical UU pairs and a free energy of 2.2 kcal/mol, whereas 5′UGAC CUC CUGA/3′ACUG CUUC GACU (duplex xx) has the potential for one UU pair and has a free energy of 3.0 kcal/mol.

NMR indicates UU noncanonical pairs occur across asymmetric internal loops

One-dimensional imino proton experiments were then performed to test if UU nucleotides form hydrogen-bonded pairs across asymmetric internal loops. All of the oligonucleotides studied by NMR have identical stems, and thus four GC and two AU imino protons are expected in all of the NMR spectra. UU imino protons resonate around 10.4–11.3 ppm (Schroeder and Turner 2000) and thus peaks that occur within this region can be attributed to UU base-pairing. In Supplemental Figure 1, four of the five duplexes that have the potential to form UU pairs show peaks in this region. The loop sequence 5′CUUC/3′GCCCCG (duplex xxxv) has no potential UU pairs and thus serves as the control and helps assign imino proton resonances in the stem Watson–Crick pairs. At least seven of the eight expected UU imino peaks were observed in the spectra of the 4×4 all-uridine loop (duplex xxviii) (Supplemental Fig. S1), which suggests multiple UU hydrogen-bonded pairs form in this loop. This 4×4 all-uridine loop is also among the most thermodynamically stable loops. The imino proton spectra demonstrate the formation of stable hydrogen-bonded UU pairs that can occur adjacent to and also nonadjacent to the closing Watson–Crick pair of the loop. These NMR data support the free energy bonus term for loops with the potential to form UU pairs (see Discussion below).

TABLE 1. RNA thermodynamic data

Duplex	pH	$\Delta G°_{37}$ loop (kcal/mol)	Linear fit of T_m^{-1} versus $\ln(C_T/4)$				Average of melt curve fits			
			$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)
5 × 5 loop										
i 5′ UGAC **CCCCC** CUGA 3′ ACUG **CCCCC** GACU	7.0 5.5	4.4 2.0	3.9±0.7 6.3±0.1	55.3±10.6 72.2±4.4	166.0±35.7 212.6±14.3	22.7 35.8	3.8±0.5 6.2±0.3	57.1±8.7 79.2±7.1	171.9±29.3 235.3±23.3	22.8 35.6
4 × 6 loop										
ii 5′ UGAC **CCCC** CUGA 3′ ACUG **CCCCCC** GACU	5.5	2.1	6.1±0.2	70.7±6.0	208.0±19.7	35.3	6.1±0.3	80.4±7.3	239.8±23.8	35.1
4 × 5 loops										
iii 5′UGAC **CCCC** CUGA 3′ACUG **CCCCC** GACU	7.0 5.5	5.4 2.0	3.2±0.6 6.2±0.1	63.1±9.0 60.8±2.2	193.1±30.1 176.0±7.2	21.6 35.4	3.5±0.4 6.2±0.2	57.9±8.9 67.9±8.1	175.4±29.7 199.0±26.4	21.6 35.4
iv 5′ CCGA **UUCU** UCGCC 3′ GGCU **UUCAC** GGGCG[e]	5.5	3.4	7.3±0.1	63.9±4.6	182.5±14.8	40.7	7.3±0.3	59.6±20.8	168.5±67.2	41.2
3 × 6 loop										
v 5′ UGAC **CCC** CUGA 3′ACUG**CCCCCC**GACU	7.0 5.5	3.7 2.3	4.6±0.4 5.9±0.2	37.1±4.6 63.1±6.8	104.7±15.9 184.2±22.4	21.5 34.0	4.4±1.0 5.9±0.3	43.4±12.8 72.4±11.7	125.8±44.0 214.6±37.6	22.4 34.1
4 × 4 loops										
vi 5′UGAC**UUUU**CUGA 3′ACUG**UUUU**GACU	7.0	1.6	6.6±0.2	88.3±8.1	263.3±25.8	37.3	6.7±0.1	77.6±4.3	228.6±13.9	37.5
vii 5′UGAC**UUUU**CUGA 3′ACUG**CUCU**GACU	7.0	1.6	6.6±0.1	92.4±4.4	276.8±14.0	37.2	6.6±0.1	83.6±1.6	248.1±5.3	37.4
viii 5′UGAC**UCUC**CUGA 3′ACUG**UUUU**GACU	7.0	2.1	6.0±0.1	86.0±6.2	257.8±20.2	35.1	6.1±0.1	75.7±2.0	224.2±6.6	35.4
ix 5′UGAC**UUUU**CUGA 3′ACUG**CUUC**GACU	7.0	2.9	5.3±0.2	73.2±5.1	218.8±16.9	32.0	5.3±0.1	72.5±1.8	216.57±6.1	31.9
x 5′ UGAC**UCUC**CUGA 3′ ACUG**CUCU**GACU[a,b]	7.0	3.3	5.2±0.2	50.0±4.5	144.4±15.3	28.9	5.0±0.1	53.8±1.7	157.5±5.7	28.4
xi 5′ UGAC**UCCU**CUGA 3′ ACUG**CUUC**GACU	7.0	3.4	4.9±0.1	54.5±1.6	160.2±5.3	27.8	4.8±0.3	57.2±6.5	168.9±21.6	27.9
xii 5′ UGAC**UCUC**CUGA 3′ ACUG**CUUC**GACU[a,b]	7.0	3.5	4.8±0.2	56.4±4.8	166.4±16.5	27.5	4.7±0.1	55.8±1.1	164.6±3.7	27.3
xiii 5′ UGAC**UUUU**CUGA 3′ ACUG**CCCC**GACU	7.0	3.7	4.4±0.2	50.9±2.9	149.8±9.8	24.6	3.9±0.4	59.5±5.6	179.1±19.0	24.0
xiv 5′ UGAC**UCUC**CUGA 3′ ACUG**CCCC**GACU	7.0	3.9	4.0±0.3	64.5±3.8	195.0±13.0	25.3	4.4±0.1	57.0±2.5	169.5±8.5	25.8
xv 5′ UGAC**CCCC**CUGA 3′ ACUG**CCCC**GACU[c,d]	7.0 5.5	4.0 1.5	4.4±0.4 6.8±0.1	55.2±7.2 74.6±4.0	163.9±24.6 218.5±13.1	25.4 38.1	4.3±0.3 6.9±0.2	58.9±3.2 85.5±16.4	176.3±11.0 253.4±52.8	23.4 38.3
3 × 5 loop										
xvi 5′ UGAC **CCC** CUGA 3′ ACUG **CCCCC** GACU	7.0 5.5	5.0 2.0	3.3±0.4 6.3±0.1	64.2±6.4 60.3±2.8	196.5±21.9 174.0±9.2	22.0 35.8	3.6±0.5 6.3±0.2	59.2±6.3 70.3±10.4	179.4±21.5 206.3±33.8	22.2 35.9
2 × 6 loop										
xvii 5′UGAC **CC** CUGA 3′ACUG**CCCCCC**GACU[c]	7.0 5.5	2.9 2.2	4.3±0.5 6.1±0.3	46.2±7.1 48.9±5.8	135.9±24.1 138.1±18.9	22.5 34.1	4.5±0.4 6.1±0.4	44.0±6.3 49.0±6.9	127.4±21.6 138.5±23.1	23.1 34.1

Continued

TABLE 1. Continued

Duplex	pH	$\Delta G°_{37\,loop}$ (kcal/mol)	Linear fit of T_m^{-1} versus ln(C_T/4)				Average of melt curve fits			
			$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)
3 × 4 loops										
xviii 5′ UGAC **UUU** CUGA / 3′ ACUG**CUCU**GACU	7.0	2.1	6.1 ± 0.1	64.2 ± 2.4	187.2 ± 8.0	35.1	6.0 ± 0.1	73.9 ± 5.4	219.0 ± 17.3	34.8
xix 5′UGAC **UUU** CUGA / 3′ACUG**UUUU**GACU	7.0	2.2	6.1 ± 0.3	78.9 ± 9.3	234.7 ± 29.8	35.2	6.1 ± 0.3	72.91 ± 6.2	215.44 ± 20.3	35.1
xx 5′ UGAC **CUC** CUGA / 3′ ACUG**CUUC**GACU[b]	7.0	3.0	5.3 ± 0.1	51.8 ± 2.0	150.0 ± 6.8	29.6	5.0 ± 0.3	62.1 ± 9.5	184.2 ± 31.5	29.3
xxi 5′ UGAC **CUC** CUGA / 3′ ACUG**UUUU**GACU	7.0	3.1	5.1 ± 0.2	58.2 ± 3.4	171.1 ± 13.1	29.7	5.1 ± 0.5	61.7 ± 11.6	182.4 ± 38.8	30.1
xxii 5′ UGAC **UCU** CUGA / 3′ ACUG**CUUC**GACU	7.0	3.2	5.1 ± 0.1	54.9 ± 1.6	160.8 ± 5.3	28.8	4.9 ± 0.3	59.4 ± 7.6	175.8 ± 25.5	28.7
xxiii 5′ UGAC **UUU** CUGA / 3′ ACUG**CCCC**GACU	7.0	3.7	4.5 ± 0.1	52.3 ± 2.2	153.8 ± 7.5	25.7	4.0 ± 0.5	61.6 ± 5.3	185.7 ± 18.4	24.8
xxiv 5′ UGAC **CUC** CUGA / 3′ ACUG**CCCC**GACU	7.0	3.8	4.4 ± 0.2	55.3 ± 3.3	164.1 ± 11.2	25.5	4.0 ± 0.5	63.0 ± 8.9	190.1 ± 30.1	25.2
xxv 5′ UGAC **CCC** CUGA / 3′ACUG **CCCC** GACU[d]	7.0	4.0	4.3 ± 0.3	50.3 ± 3.9	148.2 ± 13.5	23.8	3.8 ± 0.5	59.1 ± 8.7	178.3 ± 29.6	23.2
	5.5	1.6	6.7 ± 0.1	64.3 ± 2.3	186.0 ± 7.3	37.6	6.7 ± 0.1	70.2 ± 5.4	204.8 ± 17.5	37.7
2 × 5 loop										
xxvi 5′UCAG **CC** GUCA / 3′AGUC**CCCCCC**CAGU	7.0	3.8	4.6 ± 0.2	53.3 ± 4.6	156.9 ± 15.6	26.0	4.5 ± 0.4	53.3 ± 6.1	157.2 ± 20.9	25.8
	5.5	1.9	6.3 ± 0.1	57.1 ± 2.7	163.9 ± 9.0	35.9	6.3 ± 0.2	59.5 ± 4.1	171.4 ± 13.6	35.8
3 × 3 loop										
xxvii 5′ UGAC **CCC** CUGA / 3′ ACUG **CCC** GACU[d]	7.0	3.7	4.5 ± 0.1	48.6 ± 1.8	142.1 ± 6.0	24.7	4.2 ± 0.3	57.0 ± 6.6	170.1 ± 22.2	24.8
	5.5	0.8	7.5 ± 0.1	68.7 ± 6.0	197.4 ± 19.1	41.2	7.6 ± 0.4	80.7 ± 13.7	235.8 ± 43.1	41.1
2 × 4 loops										
xxviii 5′UGAC **UU** CUGA / 3′ACUG**UUUU**GACU	7.0	−0.1	6.0 ± 0.2	72.1 ± 5.1	213.4 ± 17.0	34.5	5.8 ± 0.0	83.8 ± 2.6	251.7 ± 8.6	34.1
xxix 5′UGAC **UU** CUGA / 3′ACUG**CUUC**GACU	7.0	0.8	5.2 ± 0.2	66.8 ± 4.7	198.8 ± 15.3	30.7	5.4 ± 0.1	57.8 ± 3.0	169.0 ± 10.1	31.1
xxx 5′ UGAC **UC** CUGA / 3′ ACUG**UUUU**GACU	7.0	2.4	5.9 ± 0.1	56.1 ± 2.0	161.8 ± 6.8	33.6	5.8 ± 0.3	67.3 ± 12.8	198.5 ± 41.8	33.6
xxxi 5′UGAC **UU** CUGA / 3′ACUG**CUCU**GACU	7.0	2.4	5.5 ± 0.2	62.3 ± 4.4	183.1 ± 14.5	32.1	5.8 ± 0.1	52.8 ± 1.8	151.6 ± 5.9	32.8
xxxii 5′ UGAC **UC** CUGA / 3′ ACUG**CUUC**GACU	7.0	2.9	5.4 ± 0.1	49.7 ± 1.7	142.9 ± 5.6	29.9	5.1 ± 0.2	57.4 ± 5.6	168.5 ± 18.7	29.4
xxxiii 5′ UGAC **UC** CUGA / 3′ ACUG**CCCC**GACU	7.0	3.3	4.9 ± 0.2	51.0 ± 2.9	148.7 ± 9.7	27.6	4.6 ± 0.2	60.6 ± 11.3	180.7 ± 36.7	27.3
xxxiv 5′ UGAC **CC** CUGA / 3′ ACUG**CCCC**GACU[d]	7.0	3.8	4.4 ± 0.3	58.3 ± 5.1	173.7 ± 17.1	26.2	4.2 ± 0.3	64.0 ± 7.4	192.8 ± 24.6	26.1
	5.5	2.0	6.3 ± 0.1	59.1 ± 2.9	170.3 ± 9.3	35.8	6.3 ± 0.1	68.4 ± 4.6	200.2 ± 14.6	35.8
xxxv 5′UGAC **UU** CUGA / 3′ACUG**CCCC**GACU	7.0	4.0	4.4 ± 0.4	55.2 ± 7.2	163.9 ± 24.6	25.4	4.2 ± 0.2	58.9 ± 3.2	176.2 ± 11.0	25.4

Continued

TABLE 1. *Continued*

Duplex	pH	$\Delta G°_{37\ loop}$ (kcal/mol)	Linear fit of T_m^{-1} versus $\ln(C_T/4)$				Average of melt curve fits			
			$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (cal/mol K)	T_m (°C)
1 × 5 loop										
xxxvi 5′ UGAC **C**　　CUGA / 3′ ACUG **CCCC** GACU	7.0	3.8	4.4 ± 0.5	55.4 ± 8.1	164.4 ± 27.4	25.6	4.9 ± 0.8	47.1 ± 10.1	135.9 ± 34.9	26.7
	5.5	3.0	5.2 ± 0.2	52.5 ± 4.0	152.4 ± 13.2	29.4	5.3 ± 0.6	55.3 ± 10.3	161.4 ± 34.8	30.0
2 × 3 loops										
xxxvii 5′UCAC **UC** GUCA / 3′AGUG**CCUC**AGU[d]	7.0	2.4	5.9 ± 0.1	60.4 ± 2.7	176.0 ± 8.9	33.6	5.8 ± 0.2	68.6 ± 5.9	202.7 ± 19.2	33.5
	5.5	1.3	7.0 ± 0.0	65.1 ± 2.7	187.5 ± 8.7	39.2	7.0 ± 0.1	70.2 ± 4.4	203.6 ± 14.0	39.3
xxxviii 5′ UGAC **CC** CUGA / 3′ACUG **CCC** GACU[d]	7.0	3.4	4.8 ± 0.3	49.0 ± 5.6	142.4 ± 18.6	26.6	4.6 ± 0.3	55.1 ± 8.6	162.8 ± 28.4	26.7
	5.5	1.5	6.8 ± 0.1	61.0 ± 3.8	174.9 ± 12.4	38.3	6.8 ± 0.2	72.7 ± 6.7	212.5 ± 21.7	38.3
1 × 4 loop										
xxxix 5′UCAG **C**　GUCA / 3′AGUCC**CCCC**CAGU	7.0	3.3	5.0 ± 0.3	52.1 ± 6.0	152.0 ± 19.9	28.0	5.0 ± 0.2	53.4 ± 7.3	156.0 ± 23.8	28.1
	5.5	1.8	6.4 ± 0.1	56.3 ± 2.6	160.9 ± 8.5	36.4	6.5 ± 0.2	66.0 ± 9.8	192.0 ± 31.9	36.6
2 × 2 loop										
xl 5′UGAC **CC** CUGA / 3′ACUG **CC** GACU[d]	7.0	2.6	5.6 ± 0.2	60.5 ± 4.9	176.8 ± 16.2	32.4	5.5 ± 0.3	70.6 ± 8.5	210.1 ± 27.5	32.4
	5.5	0.2	8.1 ± 0.0	72.6 ± 2.4	208.0 ± 7.7	43.8	8.2 ± 0.1	76.4 ± 5.2	220.0 ± 16.4	43.9

Loops are listed by decreasing size (i.e., the total number of nucleotides) and then increasing asymmetry. Within each type of loop, loops are listed by decreasing stability. Melts were performed in standard buffer containing 1 M NaCl, 10 mM sodium cacodylate, 0.5 mM Na₂EDTA at pH 7.01 or 1 M NaCl, 50 mM MES, 0.5 mM Na₂EDTA at pH 5.54. Bold text indicates the nucleotides in the internal loop.

[a]Duplexes were analyzed using one-dimensional imino proton NMR in 10 mM NaCl, 10 mM sodium phosphate, 0.5 mM Na₂EDTA, and 10% D₂O at pH 7.

[b]Duplexes were analyzed using one-dimensional imino proton NMR in 10 mM NaCl, 10 mM sodium phosphate, 0.5 mM Na₂EDTA, and 10% D₂O at pH 6.

[c]Duplex was analyzed using one-dimensional imino proton NMR at 1°C in 10 mM NaCl, 10 mM NaH₂PO₄, 0.5 mM Na₂EDTA at pH 5.2.

[d]Loop sequence was based upon trends within IRES database.

[e]Loop sequence from Hepatitis A IRES.

Protonated cytosine pairs stabilize internal loops

As shown in Figure 2, an acidic environment stabilizes internal cytidine loops. The range of free energies for all cytidine loops is 5.2 to 2.6 kcal/mol and 2.3 to 0.2 kcal/mol at pH 7 and 5.5, respectively. The stability of the cytidine loops at low pH may be due to the greater potential for protonated cytosine pairs to form an additional hydrogen bond and/or changes in base stacking. For protonated cytosine pairs, the expected resonance frequencies occur between 10.4 and 11.3 ppm (Santa Lucia et al. 1991). Other than 2×2 loops, no additional imino protons were observed within the regions where protonated cytidines resonate for loops. Thus, the added stability of cytosine base pairs at low pH may be due more to base stacking interactions than hydrogen bonding.

Symmetric and asymmetric loops show different stabilizing effects at pH 5.5. In general, symmetric loops are more stable than asymmetric loops, and smaller loops are more stable than larger loops at pH 7, which is consistent with previous studies of adenine loops (Peritz et al. 1991) and cytidine loops (Weeks and Crothers 1993). The asymmetric cytidine loops, however, show much less dependence on size and asymmetry at pH 5.5. The range of loop free energies for asymmetric cytidine loops is only 2.2 to 1.5 kcal/mol. A difference of only 0.7 kcal/mol is near the experimental error in optical melting measurements and calculation of loop free energies. Thus, the weighted average free energy of the asymmetric cytidine loops at pH 5.5 is 2.0 kcal/mol and provides a reasonable prediction value. The proposed prediction rule for symmetric cytidine loops at pH 5.5 is to subtract 2.5 kcal/mol from the loop free energy at pH 7, which maintains the dependence on loop size. The symmetric loops also show a >10 kcal/mol change in loop enthalpy at lower pH, which
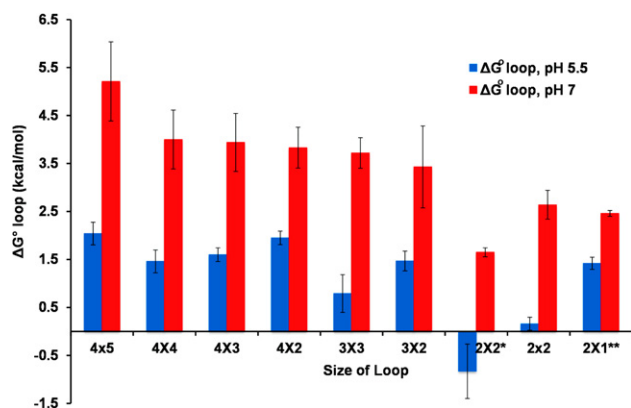

FIGURE 2. Loops with all cytidine nucleotides are more stable at pH 5. Comparison of cytidine loop free energies at pH 7 (red bars) and pH 5 (blue bars). Standard error was determined by comparing the free energy errors from the linear and melt curve fits. (*) Data obtained from Santa Lucia et al. (1991). The duplex sequence is 5′CGC CC GCG/ 3′GCG CC CGC and has only 3 bp in the stems. The difference in the length of the surrounding stems may explain the difference in stabilities for this loop (Schroeder and Turner 2000). (**) Data obtained from Schroeder and Turner (2000).

would be consistent with changes in base stacking and hydrogen bonding.

Interestingly, a 4×5 asymmetric loop, 5′AUUCUU3′/ 3′UUUCACG5′ (duplex iv), based on a loop sequence in the Hepatitis A IRES (Mokrejs et al. 2010), shows a clear two-state transition only at pH 5.5 and not at pH 7. Similarly, a 4×6 all-cytidine loop (duplex ii) shows two-state melting behavior only in pH 5.5 conditions. The IRES loop has possible protonated C^+C^+ pairs and C^+U pairs as well as UU pairs. This 4×5 hepatitis IRES loop (3.4 kcal/mol, duplex iv) is less stable than the 4×5 all-cytidine loop (2.0 kcal/mol, duplex iii), however. Thus, the additional stability of the possible U–U pairs may be less than the protonated cytosine pairs at pH 5.5. The different stabilities for IRES loops at different pH conditions may facilitate conformational changes in viral RNA that are triggered by pH changes, such as virus disassembly.

Proposed model for predicting the free energies of asymmetric internal loops

Optical melting data were analyzed with the linest function in Excel to obtain new prediction rules for asymmetric loops. Factors considered in the prediction rules include loop size, symmetry, or asymmetry, AU or GU closing base pairs, and potential GA, GG, UU, or protonated C^+C^+ pairs (Gralla and Crothers 1973). Other models for loop stability with different parameters were analyzed but did not produce better results. The current prediction model (2004 model, Mathews et al. 2004) is depicted in Equation 1 below:

$$\Delta G^{\circ}_{\text{loop}} = \Delta G^{\circ}_{\text{loop initiation}} + \Delta G^{\circ}_{\text{asymmetry penalty}}$$
$$+ \Delta G^{\circ}_{\text{AU/GU penalty}} + \Delta G^{\circ}_{\text{GG/GA/UU bonus}}. \quad (1)$$

The last free energy term is used to account for GG, GA, or UU pairs that occur adjacent to the closing Watson–Crick pairs. Analysis of the new data suggests the current model should be revised to include UU, GG, and GA base pairs that occur across internal loops and an additional term for protonated cytosine base pairs. For loops that were capable of either forming a GA or GG pair, preference was given to GA pairs due to their increased thermodynamic stability in comparison to GG pairs on average over all loop types. A revised model (2016) is proposed in Equation 2 and Table 2:

$$\Delta G^{\circ}_{\text{loop}} = (\Delta G^{\circ}_{\text{loop initiation}} + \Delta G^{\circ}_{\text{asymmetry penalty}})$$
$$+ \Delta G^{\circ}_{\text{AU/GU penalty}} + \Delta G^{\circ}_{\text{GG/GA/UU/C}^+\text{C}^+ \text{ bonus}}. \quad (2)$$

Table 2 lists one parameter for loop initiation and asymmetry because each type of loop was analyzed separately using linear regression. Thus, both the loop initiation and asymmetry are a single constant value in the linear regression analysis. For example, in order to predict the stability of the 3×4 loop 5′CCUCC/3′GCUUCG at pH 7 (duplex xx), add together 3.7 kcal/mol ($\Delta G_{\text{loop}} + \Delta G_{\text{asymmetry}}$), 0.0 kcal/mol

TABLE 2. RNA loop prediction rules

Loop	Model	ΔG° init + ΔG° asym (kcal/mol)[a]	ΔG° GA bonus (kcal/mol)[b]				ΔG° GG (kcal/mol)	ΔG° UU (kcal/mol)	Type Specific terms	ΔΔG° predict-experimental (kcal/mol)[c]	#Loops	pur[d]	pyr[e]	mix[f]
			5'RG 3'YA	5'YG 3'RA	5'YA 3'RG	5'RA 3'YG								
4 × 6	**2016**	**4.1 ± 0.4**	**−1.0 ± 0.2 for each possible GA pair +1.0 ± 0.4 for each possible GG pair**					**−0.7 ± 0.1**	**No GU[h]**	**±0.2**	**5**	**4**	**1[j]**	**0**
	2004	*3.6 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±0.8*	nda[g]			0
4 × 5	**2016**	**5.3 ± 0.4**	**−1.3 ± 0.1 for each possible GA or GG pair**					**−0.7 ± 0.1**	**No GU**	**±0.2**	**3**	**2**	**1**	**0**
	2004	*2.9 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±2.1*	nda			
3 × 6	**2016**	**3.4 ± 0.4**	**1 or 2 GA or GG pairs: ΔG° loop =−0.7 ± 0.2 3 GA or GG pairs: ΔG° loop = −1.9 ± −0.9**					**−0.7 ± 0.1**	**No GU**	**±0.3**	**9**	**8**	**1**	**0**
	2004	*4.1 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±1.0*	nda			
3 × 5	**2016**	**4.9 ± 0.2**	**−0.4 ± 0.1 for each possible GA or GG pair**					**−0.7 ± 0.1**	**No GU**	**±0.2**	**7**	**6**	**1**	**0**
	2004	*3.4 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±0.7*	nda			
3 × 4	**2016**	**3.7 ± 0.2**	**−1.1 ± 0.1**	**−1.7 ± 0.1**	**0**	**−0.6 ± 0.2**	**+1.4 ± 0.4**	**−0.5 ± 0.2**	**No GU**	**±0.3**	**24**	**13**	**8**	**3**
	2004	*2.7 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±1.0*	nda			
2 × 6	**2016**	**2.9 ± 0.5**	**−0.5 ± 0.1**	**−1.0 ± 0.1**	**−0.8 ± 0.1**	**−0.8 ± 0.1**	**−1.2 ± 0.1**	**−0.7 ± 0.1**		**0**	**2**	**1**	**1**	**0**
	2004	*4.6 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±1.2*	nda			
2 × 5	**2016**	**3.7 ± 0.2**	**−1.6 ± 0.1**	**−0.4 ± 0.1**	**−0.8 ± 0.1**	**0**	**−1.2 ± 0.1**	**−0.7 ± 0.1**		**0**	**3**	**1**	**1**	**1**
	2004	*3.9 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±0.4*	nda			
2 × 4	**2016**	**3.6 ± 0.2**	**−1.0 ± 0.1**	**−0.9 ± 0.3**	**−0.8 ± 0.1**	**0**	**−2.0 ± 0.4**	**−1.5 ± 0.2**	**No stack[i] 1.9 ± 0.4**	**±0.3**	**15**	**5**	**9**	**1**
	2004	*3.1 ± 0.1*	*−1.0 ± 0.1*	*−1.0 ± 0.1*	*−0.8 ± 0.1*	*−0.8 ± 0.1*	*−1.2 ± 0.1*	*−0.7 ± 0.1*		*±0.6*	2	2	0	0
2 × 3	**2016**	**2.9 ± 0.2**	**−1.4 ± 0.2**	**−1.5 ± 0.1**	**−0.6 ± 0.3**	**0**	**−1.0 ± 0.4**	**−0.6 ± 0.2**		**±0.4**	**52**	**32**	**6**	**14**
	2004	*2.7 ± 0.1*	*−1.2 ± 0.1*	*−1.1 ± 0.1*	*−0.5 ± 0.2*	*0*	*−0.8 ± 0.2*	*−0.4 ± 0.1*		*±0.6*	32	23	4	5
1 × 5	**2016**	**3.7 ± 0.2**	**0**				**0**	**0**		**±0.2**	**3**	**2**	**1**	**0**
	2004	*4.3 ± 0.1*	*0*				*0*	*0*		*±0.6*	2	2	0	0
1 × 4	**2016**	**3.3 ± 0.3**	**0**				**0**	**0**		**±0.2**	**3**	**2**	**1**	**0**
	2004	*3.9 ± 0.1*	*0*				*0*	*0*		*±0.6*	2	2	0	0
1 × 3	**2016**	**2.5 ± 0.2**	**0**				**0**	**0**		**±0.4**	**29**	**17**	**7**	**5**
	2004	*2.3 ± 0.1*	*0*				*0*	*0*		*±0.5*	29	17	7	5
1 × 2	**2016**	**2.2 ± 0.1**	**−1.2 ± 0.3**	**−1.2 ± 0.2**	**−1.1 ± 0.5**	**0**	**−1.4 ± 0.4**	**−1.0 ± 0.2**		**±0.3**	**39**	**23**	**12**	**4**
	2004	*2.2 ± 0.1*	*−1.1 ± 0.2*	*−1.1 ± 0.2*	*−1.1 ± 0.2*	*0*	*−1.1 ± 0.2*	*−0.7 ± 0.2*		*±0.3*	26	13	9	4

Bold font indicates new rules based on analysis in this work. Italic font indicates that rules are based on extrapolations from smaller loops in the analysis presented in the 2004 model rather than fitting experimental data. The 2004 model is presented in Mathews et al. (2004) and the 2016 model is presented in this work.

[a]The combined value of the free energies for initiation and asymmetry.

[b]For some loop types, the bonus terms for GA pairing depend on the orientation of the stacking on the closing base pairs. Y represents a pyrimidine nucleotide. R represents a purine nucleotide.

[c]The average difference between the predicted and experimentally measured loop free energy.

[d]Number of loops that contain only purine nucleotides.

[e]Number of loops that contain only pyrimidine nucleotides.

[f]Number of loops that contain a mixture of purine and pyrimidine nucleotides.

[g](nda) No data available for this type of loop in 2004. The prediction rules were extrapolated from measurements on smaller loops.

[h]The "no GU" label means that for this type of loop, there is no experimental evidence for a free energy penalty for a loop with closing GU pairs.

[i]The "no stack" label indicates that there is an additional penalty when stabilizing pairs occur but cannot stack on the closing base pairs due to intervening nucleotides, as in the loop 5'CUUC/3'GCUCUG.

[j]The only pyrimidine 4 × 6 loop was measured at pH 5.5, which is not included in prediction rules for pH 7 loops.

($\Delta G_{AU/GU\ penalty}$), and -0.5 kcal/mol ($\Delta G_{UU\ bonus}$). Using the new parameters for stabilizing pairs across an internal loop and revised initiation terms for larger sized loops, the prediction of the experimental free energies for asymmetric loops is within 0.3 kcal/mol on average. Sequence-dependent prediction rules for symmetric loops are still under development in collaboration with the Turner laboratory and alumni. Table 2 summarizes the improvements in the prediction parameters for asymmetric loops, database size, and database diversity between the 2004 model and the 2016 model.

New initiation terms for asymmetric loops are proposed in Table 2. Linear regression analysis of purine and pyrimidine loops separately revealed there are no biases in the loop initiation and asymmetry terms. The 2004 model used an asymmetry penalty of 0.7 kcal/mol per degree asymmetry and the Jacobson–Stockmayer approximation to estimate the free energies of initiation for larger, unmeasured loops. This approximation does not predict well the new loop measurements presented in Table 1 and Supplemental Table 1. Figure 3 shows that this approximation both overpredicts and underpredicts the cost of loop initiation. For example, in loops with 8 nt, the 2016 and 2004 parameters are 2.9 and 4.6 kcal/mol for 2×6 loops and 4.9 and 3.4 kcal/mol for 3×5 loops, respectively. The new initiation terms for 2×6 and 3×5 loops, which have asymmetries of four and two, respectively, do not follow the previous assumption that increasing asymmetry destabilizes a loop. Similarly for loops of 9 nt, the 3×6 and 4×5 loop initiation terms are 3.4 and 5.3 kcal/mol, respectively, for loops with three and one asymmetries, respectively. Thus, the occurrence of large, highly asymmetric internal loops in virus IRES may not be as energetically unfavorable as previously predicted.

The 2016 model retains the value of -0.7 kcal/mol for an AU closing base pair. This parameter repeatedly gave the same value within error for all linear regression analyses. The value of -0.7 kcal/mol is consistent with measurements of the free energy of a hydrogen bond and accounts for the difference in the number of hydrogen bonds in AU and GC pairs (Xia et al. 1998). The same value has been used for loops with closing GU pairs (Schroeder and Turner 2001). In this analysis of 4×6, 4×5, 3×6, 3×5, and 3×4 loops, however, there was

little evidence for any penalty for loops with GU pairs. For example, the free energies of 3×5 loops 5'UGAAG/3'GAAGGAC and 5'CGAAG/3'GAAGGAC are both 1.8 kcal/mol, and the free energies of 3×4 loops 5'CGAAG/3'GAAGGC and 5'UGAAG/3'GAAGGC are -0.7 and -0.5 kcal/mol, respectively (Chen et al. 2006). The thermodynamic stabilities of GU pairs are highly idiosyncratic (Schroeder and Turner 2001; Schroeder et al. 2003; Nguyen and Schroeder 2010; Chen et al. 2012), and in some contexts, the stacking interactions with GU pairs may outweigh the loss of a hydrogen bond.

Table 2 shows how the values for the bonus terms for GA, GG, and UU pairs vary with different loop sizes and asymmetries. The bonus values for GA pairs depend on the orientation and stacking of the GA pair. For example, in 2×3 loops, a 5'RG/3'YA pair adds -1.4 kcal/mol while a 5'RA/3'YG pair adds no stability. (Note that R and Y indicate a purine or a pyrimidine in a Watson–Crick pair, respectively.) A 5'RA/3'YG in a 2×4 loop and a 5'YA/3'RG pair in a 3×4 loop also add no stability to a loop. The range of additional stabilities provided by GA pairs in an asymmetric loop ranges from zero to -1.8 kcal/mol depending on the loop size, asymmetry, and base stacking. For 5×5, 4×6, 4×5, 3×6, and 3×5 loops, the 2004 model clearly does not predict well the experimental values, but there are less than 10 loop measurements for each type of loop. In these cases, there are too few measurements to parse statistically significant differences between different GA stacking interactions, and an average value is used for the prediction rule.

The GG bonus varies widely between different loops. In 2×4 loops, a GG pair adds -2.0 kcal/mol while a GG pair in a 3×4 loop is destabilizing by 1.4 kcal/mol. Guanine nucleotides have many hydrogen bond donors and acceptor sites and are able to form many different types of pairs (Leontis et al. 2002). Guanine nucleotides also have the largest dipole moment (Bloomfield et al. 2000) and thus are capable of forming strong stacking interactions. Thus, the thermodynamic stabilities of GG pairs are highly idiosyncratic (Burkard and Turner 2000).

Uridine nucleotides have weaker stacking interactions than guanine nucleotides and show less variation in the bonus for UU pairs (Freier et al. 1983). The values for UU bonus terms range from -0.4 kcal/mol to -1.0 kcal/mol for all loops except 2×4 loops. The 2×4 loop prediction rule also includes a penalty for when the formation of possible noncanonical stabilizing pairs occurs in ways that would force nucleotides to bulge out of the loop if the noncanonical pair stacked on the closing Watson–Crick pair. Individual bulge nucleotides also have an unfavorable parameter in prediction rules, and similar unfavorable energetic effects of backbone distortion may occur in small highly asymmetric loops such as 2×4 loops.
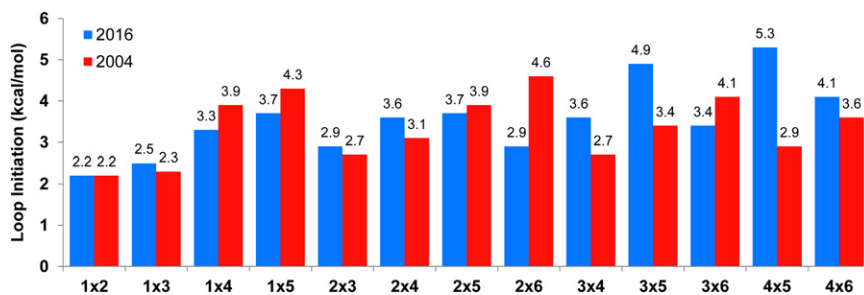

FIGURE 3. Loop initiation terms in the 2016 and 2004 models. Loop initiation parameters include the terms for the free energy penalties for loop asymmetry. The parameters are the constant value in the linear regression analysis of each type of loop.

The 2016 model now includes a bonus term for protonated CC pairs in symmetric loops at pH 5 that would account for stabilizing hydrogen bonding and stacking interactions. Some viruses, such as enteroviruses, can survive in acidic environments and the release of viral genomes into the host cell is mediated by a change in pH. For these reasons, the cytidine loops were studied at both pH 7.0 and 5.5. The new prediction rule for loops with protonated cytosine pairs depends on whether the loop is symmetric or asymmetric. The symmetric cytidine loops show more dependence on loop size. Recent crystal structures of tandem CC pairs (2×2 loop) at neutral pH show stacking interactions and a single hydrogen bond between the N3 and amino groups or a bifurcated hydrogen bond from the amino group to the N3 and the carbonyl (Dodd et al. 2016). This type of pairing would not be possible if the N3 position were protonated, but possible C^+C^+ pairs could form with two hydrogen bonds. There are several possible conformations for CC and CU pairs that would be disrupted by cytosine protonation (Kiliszek et al. 2012; Rypniewski et al. 2016). Cytidine protonation would also change stacking interactions. A cytidine tetramer shows A-form stacking interactions by NMR (Tubbs et al. 2013). If the cytidines were protonated, then different stacking interactions would likely form. The stabilities of stacking interactions for cytidines in the unfolded single strand state are likely similar to stacking interactions for adenines because the similarities of loop stabilities for all-adenine loops and all-cytidine loops at neutral pH do not have any stabilizing bonus terms.

Conclusion

The thermodynamic database for asymmetric internal loops has been curated and expanded. The new initiation terms for loops containing >6 nt greatly improved accuracy in predicting the experimental free energies. The stabilities of large, highly asymmetric internal loops that occur in virus IRES may be more favorable than previously predicted. The bonus terms for GA, GG, UU, and protonated CC pairs vary with the size and asymmetry of the loop. The thermodynamics obtained from optical melting experiments suggests that protonated cytosines and uracil base-pairing at any position in internal loops contribute to loop stability. One-dimensional imino proton NMR spectra validate the formation of UU base pairs, even when the pairs may not stack directly on the closing Watson–Crick pairs. The bonus terms for cytidine loops at low pH will improve predictions for viral RNA that undergo conformational changes in different pH environments. These results will contribute to the ongoing effort to expand and diversify the RNA thermodynamic database and improve RNA structure prediction.

MATERIALS AND METHODS

The IRES structures studied were obtained from iresite.org (Mokrejs et al. 2010), which contains a database of experimentally determined IRES structures. The database contains IRES structures from 11 viruses and 11 eukaryotic cellular mRNAs that contain a total of 107 IRES internal loops. Although IRESs are also rich in bulge loops, hairpins, and pseudoknots, only internal loops were examined for this study.

RNA oligos were obtained from Dharmacon and then deblocked according to the manufacturer's protocol. The concentrations of the oligos were then determined by measuring the absorbance at 80°C and 260 nm using a Beckman DU800 spectrophotometer with temperature control. The RNA oligos were then dissolved in standard optical melting buffer, which consists of 1 M NaCl, 0.5 mM Na_2EDTA, and 10 mM sodium cacodylate or 50 mM MES at pH 7.01 or 5.54, respectively. One-dimensional imino proton NMR data were collected with samples at 0.5 mM RNA concentrations in NMR buffer containing 10 mM NaCl, 10 mM potassium phosphate in 90% H_2O, and 10% D_2O at pH's 6.0 and 4.5 with Watergate solvent suppression on a 500 MHz Varian NMR spectrometer (Piotto et al. 1992; Lukavsky and Puglisi 2001).

Single-stranded melts were performed to check if stable homodimers form. For both single strand and heteroduplex melts, the absorbance at 260 nm was measured as a function of temperature from 0.1°C to 80°C at a heating rate of 1°C/min. The individual melt curves were fit using Meltwin software (McDowell and Turner 1996). A plot of the Tm^{-1} versus the natural log of the duplex concentration was fit according to the equation:

$$1/T_m = -(R/\Delta H°)\ln(\text{conc.}) + \Delta S°/\Delta H°, \quad (3)$$

where R is the gas constant (Schroeder and Turner 2009). If the enthalpy terms agree within 15% between the two different fits, then the heteroduplex is considered to be two-state. Using the equation from the Van't Hoff plot in conjunction with the Gibbs free energy Equation 1, the thermodynamic parameters ΔG, ΔH, and ΔS can be determined for the duplex. Using the nearest-neighbor approximations (Xia et al. 1998), the free energy of the stems can be calculated and subtracted from the duplex to get the free energy of the loop itself.

$$\Delta G°_{\text{loop}} = \Delta G°_{\text{duplex}} - \Delta G°_{\text{stems}} + \Delta G°_{\text{interrupted nearest neighbor}}. \quad (4)$$

For example, in order to calculate the experimental free energy for the 3×4 loop 5′CCUCC/3′GCUUCG at pH 7, add −5.3 kcal/mol (ΔG_{duplex} 5′UGAC CUC CUGA/3′ACUG CUUC GACU), −11.5 (ΔG_{stems} 5′UGACCUGA/3′ACUGGACU), and −3.3 kcal/mol ($\Delta G_{\text{interrupted nearest neighbor}}$ 5′CC/3′GG).

Data from optical melting experiments were fit using the linest function in Excel. The loops at pH 7 from this work were combined with all previous measurements of internal loop stabilities (Santa Lucia et al. 1990, 1991; Peritz et al. 1991; Walter et al. 1994; Wu et al. 1995; Schroeder et al. 1996; Xia et al. 1997; Schroeder and Turner 2000, 2001; Burkard et al. 2001; Chen et al. 2004, 2005, 2006, 2009; Bourdelat-Parks and Wartell 2005; Chen and Turner 2006; Badhwar et al. 2007; Christiansen and Znosko 2008, 2009; Hausmann and Znosko 2012; Zhong et al. 2015). Thermodynamic data were manually curated, and data with non-two-state behavior, highly probable alternate duplex structures, and three or more consecutive guanine nucleotides without single-strand optical melting data were not included in the database for analysis. The prediction model was assessed by the number of loops that were predicted within 0.3 kcal/mol and 0.5 kcal/mol; low standard deviation in the model parameters; and Students' t and P tests.

Loops were grouped by size and asymmetry, and new prediction rules were tested for each loop type.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Badhwar J, Karri S, Cass CK, Wunderlich E, Znosko B. 2007. Thermodynamic characterization of RNA duplexes containing naturally occurring 1×2 nucleotide internal loops. *Biochemistry* **46:** 14715–14724.

Bloomfield VA, Crothers DM, Tinoco I Jr. 2000. *Nucleic acids: structures, properties, and functions.* University Science Books, Sausalito, CA.

Bourdelat-Parks BN, Wartell RM. 2005. Thermodynamics of RNA duplexes with tandem mismatches containing a uracil-uracil pair flanked by C·G/G·C or G·C/A·U closing base pairs. *Biochemistry* **44:** 16710–16717.

Burkard ME, Turner DH. 2000. NMR structures of r(GCAGGCGUGC)₂ and determinants of stability for single guanosine-guanosine base pairs. *Biochemistry* **39:** 11748–11762.

Burkard ME, Xia T, Turner DH. 2001. Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another noncanonical pair. *Biochemistry* **40:** 2478–2483.

Chen G, Turner DH. 2006. Consecutive GA pairs stabilize medium-size internal loops. *Biochemistry* **45:** 4025–4043.

Chen G, Znosko BM, Jiao X, Turner DH. 2004. Factors affecting the thermodynamic stabilities of RNA 3×3 internal loops. *Biochemistry* **43:** 12865–12876.

Chen G, Znosko BM, Kennedy SD, Krugh TR, Turner DH. 2005. Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry* **44:** 2845–2856.

Chen G, Kennedy SD, Qiao J, Krugh TR, Turner DH. 2006. An alternating sheared AA pair and elements of stability for a single sheared purine-purine pair flanked by sheared GA pairs in RNA. *Biochemistry* **45:** 6889–6903.

Chen G, Kennedy SD, Turner DH. 2009. A CA⁺ pair adjacent to a sheared GA or AA pair stabilizes size-symmetric RNA internal loops. *Biochemistry* **48:** 5738–5752.

Chen JL, Dishler AL, Kennedy SD, Yildirim I, Liu B, Turner DH, Serra MJ. 2012. Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry* **51:** 3508–3522.

Christiansen ME, Znosko BM. 2008. Thermodynamic characterization of the complete set of sequence symmetric tandem mismatches in RNA and an improved model for predicting the free energy contribution of sequence asymmetric tandem mismatches. *Biochemistry* **47:** 4329–4336.

Christiansen ME, Znosko BM. 2009. Thermodynamic characterization of tandem mismatches found in naturally occurring RNA. *Nucleic Acids Res* **37:** 4696–4706.

Davis DR, Seth PP. 2011. Therapeutic targeting of HCV internal ribosomal entry site RNA. *Antivir Chem Chemother* **21:** 117–128.

Dibrov SM, Ding K, Brunn ND, Parker MA, Bergdahl BM, Wyles DL, Hermann T. 2012. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proc Natl Acad Sci* **109:** 5223–5228.

Dibrov SM, Parsons J, Carnevali M, Zhou S, Rynearson KD, Ding K, Garcia Sega E, Brunn ND, Boerneke MA, Castaldi MP, et al. 2014. Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site. *J Med Chem* **57:** 1694–1707.

Dodd DW, Tomchick DR, Corey DR, Gagnon KT. 2016. Pathogenic C9ORF72 antisense repeat RNA forms a double helix with tandem C:C mismatches. *Biochemistry* **55:** 1283–1286.

Eggenhofer F, Hofacker IL, Höner Zu, Siederdissen C. 2016. RNAlien - Unsupervised RNA family model construction. *Nucleic Acids Res* **44:** 8433–8441.

Freier SM, Burger B, Alkema D, Neilson T, Turner DH. 1983. Effects of 3′ dangling stacking on the stability of GGCC and CCGG double helices. *Biochemistry* **22:** 6198–6206.

Gralla J, Crothers DM. 1973. Free energies of imperfect nucleic acid helices III. Small internal loops resulting from mismatches. *J Mol Biol* **78:** 301–309.

Gu X, Mooers BH, Thomas LM, Malone J, Harris S, Schroeder SJ. 2015. Structures and energetics of four adjacent G·U pairs that stabilize an RNA helix. *J Phys Chem B* **119:** 13252–13261.

Hausmann NZ, Znosko BM. 2012. Thermodynamic characterization of RNA 2×3 nucleotide internal loops. *Biochemistry* **51:** 5359–5368.

Kikuchi K, Umehara T, Fukuda K, Kuno A, Hasegawa T, Nishikawa S. 2005. A hepatitis C virus (HCV) internal ribosome entry site (IRES) domain III-IV-targeted aptamer inhibits translation by binding to an apical loop of domain IIId. *Nucleic Acids Res* **33:** 683–692.

Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. 2012. Crystallographic characterization of CCG repeats. *Nucleic Acids Res* **40:** 8155–8162.

Leontis N, Stombaugh J, Westof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30:** 3497–3531.

Lozano G, Martinez-Salas E. 2015. Structural insights into viral IRES-dependent translation mechanisms. *Curr Opin Virol* **12:** 113–120.

Lukavsky P, Puglisi J. 2001. RNAPack: an integrated NMR approach to RNA structure determination. *Methods* **25:** 316–332.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101:** 7287–7292.

McDowell JA, Turner DH. 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.

Mokrejs M, Masek T, Vopálensky V, Hlubucek P, Delbos P, Pospísek M. 2010. IRESite—a tool for the examination of viral and cellular internal ribosome entry sites. *Nucleic Acids Res* **38:** D131–D136.

Nguyen MT, Schroeder SJ. 2010. Consecutive terminal GU pairs stabilize RNA helices. *Biochemistry* **49:** 10574–10581.

Paulsen RB, Seth PP, Swayze EE, Griffey RH, Skalicky JJ, Cheatham TE III, Davis DR. 2010. Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc Natl Acad Sci* **107:** 7263–7268.

Peritz AE, Kierzek R, Sugimoto N, Turner DH. 1991. Thermodynamic study of internal loops in oligoribonucleotides: symmetric loops are more stable than asymmetric loops. *Biochemistry* **30:** 6428–6436.

Piotto M, Saudek V, Sklenár V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2:** 661–665.

Plank TD, Kieft JS. 2012. The structures of nonprotein-coding RNAs that drive internal ribosome entry site function. *Wiley Interdiscip Rev RNA* **3:** 195–212.

Pyle AM, Schlick T. 2016. Challenges in RNA structural modeling and design. *J Mol Biol* **428:** 733–735.

Rypniewski W, Banaszak K, Kuliński T, Kiliszek A. 2016. Watson-Crick-like pairs in CCUG repeats: evidence for tautomeric shifts or protonation. *RNA* **22:** 22–31.

Santa Lucia J Jr, Kierzek R, Turner DH. 1990. Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry* **29:** 8813–8819.

Santa Lucia J Jr, Kierzek R, Turner DH. 1991. Stabilities of consecutive A·C, C·C, G·G, U·C, and U·U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U·U and C·C$^+$ pairs. *Biochemistry* **30:** 8242–8251.

Schroeder SJ, Turner DH. 2000. Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry* **39:** 9257–9274.

Schroeder SJ, Turner DH. 2001. Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry* **40:** 11509–11517.

Schroeder SJ, Turner DH. 2009. Optical melting measurements of nucleic acid thermodynamics. *Methods Enzymol* **468:** 371–387.

Schroeder SJ, Kim J, Turner DH. 1996. GA and UU mismatches can stabilize 1×2 internal loops of three nucleotides. *Biochemistry* **35:** 16105–16109.

Schroeder SJ, Fountain MA, Kennedy SD, Lukavsky PJ, Puglisi JD, Krugh TR, Turner DH. 2003. Thermodynamic stability and structural features of the J4/5 loops in a pneumocystis carini group I intron. *Biochemistry* **42:** 14184–14196.

Tubbs JD, Condon DE, Kennedy SD, Hauser M, Bevilacqua PC, Turner DH. 2013. The nuclear magnetic resonance of CCCC RNA reveals a right-handed helix, and revised parameters for AMBER force field torsions improve structural predictions from molecular dynamics. *Biochemistry* **52:** 996–1010.

Walter AE, Wu M, Turner DH. 1994. The stability and structure of tandem GA mismatches in RNA depend on closing base pairs. *Biochemistry* **33:** 11349–11354.

Weeks KM, Crothers DM. 1993. Major groove accessibility of RNA. *Science* **261:** 1574–1577.

Wu M, McDowell JA, Turner DH. 1995. A periodic table of symmetric tandem mismatches in RNA. *Biochemistry* **34:** 3204–3211.

Xia T, McDowell JA, Turner DH. 1997. Thermodynamics of nonsymmetric tandem mismatches adjacent to G·C base pairs in RNA. *Biochemistry* **36:** 12486–12497.

Xia T, Santa Lucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37:** 14719–14735.

Xu ZZ, Mathews DH. 2016. Experiment-assisted secondary structure prediction with RNAstructure. *Methods Mol Biol* **1490:** 163–176.

Xu X, Zhao P, Chen SJ. 2014. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One* **9:** e107504.

Zhong Z, Soh LH, Lim MH, Chen G. 2015. A UU pair to UC pair mutation-induced RNA native structure destabilisation and stretching-force-induced RNA misfolding. *ChemPlusChem* **80:** 1267–1278.

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2017/02/17/rna.059865.116.DC1
References	This article cites 52 articles, 5 of which can be accessed free at: http://rnajournal.cshlp.org/content/23/5/770.full.html#ref-list-1
Open Access	Freely available online through the *RNA* Open Access option.
Creative Commons License	This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions