Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides

GRAHAM A. HUDSON, RICHARD J. BLOOMINGDALE, and BRENT M. ZNOSKO[1]

Department of Chemistry, Saint Louis University, Saint Louis, Missouri 63103, USA

ABSTRACT

Pseudouridine (Ψ) is the most common noncanonical nucleotide present in naturally occurring RNA and serves a variety of roles in the cell, typically appearing where structural stability is crucial to function. Ψ residues are isomerized from native uridine residues by a class of highly conserved enzymes known as pseudouridine synthases. In order to quantify the thermodynamic impact of pseudouridylation on U-A base pairs, 24 oligoribonucleotides, 16 internal and eight terminal Ψ-A oligoribonucleotides, were thermodynamically characterized via optical melting experiments. The thermodynamic parameters derived from two-state fits were used to generate linearly independent parameters for use in secondary structure prediction algorithms using the nearest-neighbor model. On average, internally pseudouridylated duplexes were 1.7 kcal/mol more stable than their U-A counterparts, and terminally pseudouridylated duplexes were 1.0 kcal/mol more stable than their U-A equivalents. Due to the fact that Ψ-A pairs maintain the same Watson-Crick hydrogen bonding capabilities as the parent U-A pair in A-form RNA, the difference in stability due to pseudouridylation was attributed to two possible sources: the novel hydrogen bonding capabilities of the newly relocated imino group as well as the novel stacking interactions afforded by the electronic configuration of the Ψ residue. The newly derived nearest-neighbor parameters for Ψ-A base pairs may be used in conjunction with other nearest-neighbor parameters for accurately predicting the most likely secondary structure of A-form RNA containing Ψ-A base pairs.

Keywords: RNA thermodynamics; secondary structure prediction; modified nucleotides; nonstandard nucleotides

INTRODUCTION

RNA is an important class of biomolecules ubiquitous to all forms of life. RNA plays crucial roles in virtually all steps of protein synthesis, including, but not limited to, acting as the transcript of DNA (mRNA), gene regulation (siRNAs and snRNA), the recognition and delivery of amino acids to the ribosome (tRNA), and the creation of peptide bonds in nascent proteins (rRNA). RNA also may act as the genetic material for viruses that utilize reverse transcription to create complementary DNA copies that can subsequently commandeer cellular machinery for viral synthesis and propagation. Like its more stable, genomic counterpart, DNA, RNA is typically composed of four canonical nucleotides that pair via hydrogen bonding: guanosine, cytidine, adenosine, and uridine. However, unlike DNA, RNA may contain a wide variety of modified, noncanonical bases that may contribute special structural or catalytic properties, allowing RNA to serve in the variety of aforementioned processes. These modified nucleotides are the product of RNA-editing enzymes that chemically modify specific bases as part of the process of producing mature RNA (Grosjean et al. 1998). Of the many modified nucleotides found in nucleic acids, pseudouridine (Ψ) is the most common and serves a variety of roles within the cell (Grosjean et al. 1998; Charette and Gray 2000; Ofengand et al. 2001; Del Campo et al. 2007).

Pseudouridine, as the name suggests, is a modification of the canonical uridine residue that is formed by a class of enzymes collectively known as Ψ synthases (Del Campo et al. 2007; Vaidyanathan et al. 2007). These highly conserved enzymes often modify a single specific uridine residue and function by detaching the residue's base from its sugar and "rotating" it 180° along the N3-C6 axis followed by reattachment of the base's 5-carbon to the 1′-carbon of the sugar (Fig. 1; Charette and Gray 2000).

The resultant Ψ residue features the same Watson-Crick hydrogen bonding capabilities as the parent uridine residue; however, an imino group (N1H) now resides at the position
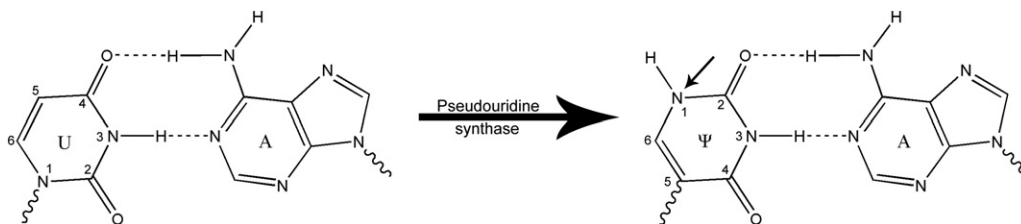
FIGURE 1. Pseudouridine (Ψ) synthases isomerize uridine residues, resulting in the creation of a new hydrogen bond donor (N1H) in the major groove of the Ψ-A pair (indicated by arrow).

that C5 previously occupied, resulting in a new hydrogen donor that is situated in the major groove when in the preferred *anti* conformation (Charette and Gray 2000). This modification results in increased stability, attributed to the enhanced stacking between adjacent RNA bases (Davis 1995; Charette and Gray 2000), as well as the novel hydrogen bonding capabilities afforded by the N1H imino proton (Charette and Gray 2000; Ofengand et al. 2001). Although the N1 imino proton does not directly interact with Ψ's pairing partner, it has been shown to coordinate a water molecule to the phosphodiester backbone of the RNA. The rigidity imparted by this novel interaction acts as an additional support to the backbone of the RNA, giving rise to a stabilizing effect that is propagated throughout the entire molecule (Hall and McLaughlin 1992; Charette and Gray 2000; Schroeder et al. 2005). Due to Ψ's highly stabilizing nature, it is not surprising that this modified nucleotide very commonly appears in biochemical roles where structure is vital to function.

Ψ is found in virtually all organisms spanning all kingdoms of life, with a strong correlation between Ψ frequency and organism complexity (Ofengand et al. 2001). For example, prokaryotic *Escherichia coli* large subunit (LSU) rRNA contains 10 Ψ residues, whereas eukaryotic *Homo sapiens* LSU rRNA has 55 Ψ residues (Ofengand et al. 2001). In LSU rRNA, these residues have been shown to cluster at catalytically important sites such as the peptidyltransferase center, where nascent peptides are elongated (Bakin and Ofengand 1993; Bakin et al. 1994; Ofengand et al. 2001). In stark contrast, Ψ has not been found to cluster in functionally important domains in small subunit (SSU) rRNA and instead may act in the capacity of quality control factors (Ofengand et al. 2001; Vaidyanathan et al. 2007). For example, previous studies have shown that Ψ may be used to mark correctly folded precursory SSU rRNA as it proceeds through the stages of synthesis and assembly (Vaidyanathan et al. 2007). This suggests that Ψ may only act in structural or quality control roles in SSU rRNA, while LSU rRNA may employ Ψ for both structural and catalytic purposes. Irrespective of purpose or point of modification, deletion of certain Ψ residues on the ribosome has been shown to produce dominantly lethal mutants, likely due to improper manufacture or assembly of ribosomal subunits (Vaidyanathan et al. 2007). It is therefore vital to understand how pseudouridylation is able to impart stability and viability on a ribosome simply by isomerizing a native uridine residue.

Ψ is also thought to play an integral part in the stabilization of tRNA structure due to its ubiquitous presence in the TΨC loop of tRNAs, as well as its frequent presence in the anticodon-stem–loop (ASL) (Charette and Gray 2000). Experiments have shown that while *E. coli* lacking Ψ in the ASL are viable and therefore are able to maintain a functional structure, deletion results in retarded organism growth, increased translation errors, and lowered rates of polypeptide elongation (Davis and Poulter 1991). This is thought to be due to stabilization imparted by Ψ's novel hydrogen bonding capabilities, resulting in more accurate and stable recognition interactions between tRNA and the ribosome (Davis and Poulter 1991). Notably, this property is exploited by the HIV genome, which relies on the stabilization afforded by the pseudouridylation of tRNA[Lys3] to act as a reverse transcriptase primer (Bilbille et al. 2009). Ψ may also play a prominent role in spliceosomal RNA responsible for gene regulation. Oftentimes, Ψ is present in regions involved in RNA–RNA or RNA–protein interactions that promote the assembly and reaction of a spliceosome to yield viable mRNA such as in AU/AC intron splicing. For example, Ψ has been shown to assist in the stabilization of snRNA/intron transition states during splicing, which allows a nucleophilic residue to assume a less-transient near-attack conformation (Charette and Gray 2000). Such splice site–specific Ψ residues are highly conserved in related species but vary between genera within a family (Charette and Gray 2000; Ofengand et al. 2001).

Given that structure–function relationships form the basis for current biochemical schools of thought, it is necessary to understand how base modifications affect RNA structure, stability, and function. Free energy minimization using the nearest-neighbor model is a common method of predicting RNA thermodynamics and the most likely secondary structure adopted by an RNA sequence (Xia et al. 1998; Wright et al. 2007). The nearest-neighbor model uses linearly independent thermodynamic parameters for consecutive base pairs in an RNA duplex to predict the change in enthalpy, entropy, and Gibbs' free energy parameters. These nearest-neighbor parameters have been generated for each set of adjacent canonical RNA base pairs (Xia et al. 1998; Chen et al. 2012); however, parameters for adjacent base pairs containing modified bases are lacking. Recent efforts have yielded novel parameters for inosine (RNA nucleotide with hypoxanthine as the base) paired with uridine, which suggested that significant

changes in thermodynamic stability are imparted by conversion of an A-U pair to an I-U pair (Wright et al. 2007). While previous studies have thermodynamically characterized molecules containing Ψ residues (Hall and McLaughlin 1991; Davis 1995; Sipa et al. 2007), linearly independent parameters for use in nearest-neighbor predictions have yet to be experimentally derived.

The availability of Ψ nearest-neighbor parameters would aid in the prediction of stability and structure for RNA containing Ψ residues. In particular, the availability of parameters characterizing Ψ-A base pairs would be of value because many uridine residues are paired with adenosine prior to pseudouridylation. Predictive Ψ-A nearest-neighbor parameters would also assist in the investigation of their suspected role in tRNA stabilization and recognition as well as the catalysis of spliceosomal RNA; the prediction of how secondary structure is dependent on pseudouridylation may elucidate Ψ's role in these critical biological reactions and whether its role is catalytic, structural, or perhaps both. Additionally, comparison of Ψ-A to U-A base pairs could reveal how the isomerization of a uridine residue imparts enhanced hydrogen bonding, stacking, and major groove interactions, thus increasing our understanding of nucleic acid structure and stability as a whole.

This study reports linearly independent nearest-neighbor parameters for adjacent base pairs containing a Ψ-adenosine (Ψ-A) base pair. Oligoribonucleotides comprised of all possible Ψ-A nearest-neighbor pairs have been chemically synthesized, purified, and thermodynamically characterized by optical melting. The derived thermodynamic parameters were analyzed using linear regression to generate nearest-neighbor parameters for Ψ-A base pairs. These parameters aim to improve the efficacy of secondary structure prediction for RNA containing Ψ-A pairs and thus aid in elucidating the thermodynamic implications of adding Ψ-A base pairs to natural or synthetic RNA.

RESULTS

Thermodynamic parameters of duplexes containing Ψ-A pairs

The thermodynamic parameters of duplex formation derived by fitting observed optical melting curves to a two-state model as well as from their corresponding van't Hoff (T_M vs. $\log(C_T/4)$) plots are shown in Table 1. All van't Hoff plots exhibited a positive correlation coefficient of at least 0.95, with most being greater than 0.99, and all enthalpy values derived from the average of curve fits were within 15% of the enthalpy values derived from the van't Hoff plot, suggesting that all duplexes melted in a two-state manner (Schroeder and Turner 2009).

Comparison of Ψ-A experimental free energies to U-A predicted free energies

All of the duplexes containing Ψ-A pairs were more stable than what is predicted for the same duplexes containing

U-A pairs (Table 2). Specifically, duplexes containing Ψ-A pairs are on average 1.7 and 1.0 kcal/mol more stable than what is predicted for the same duplexes containing internal and terminal U-A pairs, respectively. Similar calculations and comparisons for enthalpy and entropy are shown in Supplemental Tables S1 and S2.

Thermodynamic contribution of Ψ-A base pairs and derivation of nearest-neighbor parameters

The thermodynamic contribution of Ψ-A base pairs to duplex thermodynamics, calculated by subtracting Watson-Crick nearest-neighbor contributions from the measured duplex thermodynamics (as shown in Equation 4), is shown in Supplemental Table S3. As described in Materials and Methods, these values were used to generate linearly independent nearest-neighbor parameters for Ψ-A pairs (Table 3). All of the generated Ψ-A nearest-neighbor free energy contributions are more stabilizing than their canonical U-A equivalents (Table 3), with a general trend of increased stability afforded by 3′-stacked Ψ-A pairs compared with 5′-stacked Ψ-A pairs, as evidenced by the ranking order of

$$\frac{G\Psi}{CA} > \frac{A\Psi}{UA} \approx \frac{C\Psi}{GA} \approx \frac{\Psi U}{AA} > \frac{\Psi C}{AG} > \frac{\Psi G}{AC} \approx \frac{\Psi A}{AU} > \frac{U\Psi}{AA}$$

(Supplemental Table S4). Specifically, the average value of 3′-stacked Ψ-A parameters was −2.62 kcal/mol, and the average for 5′-stacked Ψ-A parameters was −2.38 kcal/mol. This is in contrast with canonical U-A doublets where 3′-stacked U-A pairs contribute on average −1.59 kcal/mol and 5′-stacked U-A pairs contribute on average −1.68 kcal/mol (Xia et al. 1998). Additionally, the terminal Ψ-A penalty of 0.31 kcal/mol is less destabilizing than the 0.45 kcal/mol penalty assigned to terminal U-A pairs (Xia et al. 1998).

The newly generated nearest-neighbor parameters for Ψ-A, in conjunction with previously generated parameters for canonical base pairs, can be used to predict thermodynamic parameters for any RNA duplex containing nonadjacent Ψ-A pairs using the nearest-neighbor model as described in Materials and Methods. These new parameters were used to predict the experimental thermodynamics of the duplexes studied here (Table 1), and average deviations between the experimental and predicted thermodynamics were 1.7% (0.15 kcal/mol), 6.7% (4.00 kcal/mol), and 8.0% (13.06 entropy units [eu]) for $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ respectively. These deviations are comparable to previous models reported for canonical base pairs (3.2%, 6.0%, and 6.8%) (Xia et al. 1998), and inosine–uridine base pairs (5.1%, 4.6%, and 5.1%) (Wright et al. 2007).

DISCUSSION

Helix to coil transition of Ψ-A oligoribonucleotides

As stated in the Results, a positive correlation was observed between melting temperature and RNA concentration,

TABLE 1. Thermodynamic parameters of duplex formation

RNA duplex[a]	T_M^{-1} vs log $(C_T/4)$ plots				Average of curve fits				Predicted using new Ψ-A nearest-neighbor parameters[b]			
	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG°$_{37}$ (kcal/mol)	T_M (°C)[c]	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG°$_{37}$ (kcal/mol)	T_M (°C)[c]	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG°$_{37}$ (kcal/mol)	T_M (°C)[c]
Internal												
CGAΨACG GCUAUGC	72.9 ± 7.0	203.0 ± 21.5	9.93 ± 0.34	52.2	70.0 ± 4.9	193.8 ± 15.4	9.84 ± 0.21	52.4	76.1	212.6	10.12	52.4
CGCΨACG GCGAUGC	69.1 ± 7.9	187.6 ± 24.2	10.96 ± 0.51	58.2	65.4 ± 1.8	176.2 ± 5.5	10.77 ± 0.18	58.5	72.6	198.2	11.17	58.1
CGGΨACG GCCAUGC	72.9 ± 1.2	197.1 ± 3.7	11.71 ± 0.08	60.7	74.0 ± 0.4	200.5 ± 1.5	11.78 ± 0.03	60.7	79.0	217.6	11.53	57.9
CGUΨACG GCAAUGC	77.6 ± 1.9	220.9 ± 6.0	9.10 ± 0.07	47.6	77.3 ± 0.3	219.9 ± 1.0	9.09 ± 0.05	47.6	73.8	209.3	8.83	47.0
CGAΨCCG GCUAGGC	80.0 ± 1.6	219.4 ± 4.7	11.92 ± 0.10	59.4	78.4 ± 0.4	214.7 ± 1.3	11.82 ± 0.03	59.5	82.9	230.0	11.53	56.9
CGCΨCCG GCGAGGC	72.1 ± 1.7	190.8 ± 5.1	12.93 ± 0.13	67.2	68.1 ± 0.6	178.8 ± 1.8	12.61 ± 0.03	67.4	79.5	215.7	12.57	62.5
CGGΨCCG GCCAGGC	94.9 ± 2.6	264.7 ± 7.9	12.76 ± 0.15	58.8	85.8 ± 1.4	237.2 ± 4.6	12.23 ± 0.03	59.1	85.8	235.0	12.94	62.0
CGUΨCCG GCAAGGC	84.8 ± 3.0	242.0 ± 9.4	9.76 ± 0.11	49.3	84.6 ± 0.7	241.2 ± 2.3	9.75 ± 0.06	49.3	80.6	226.8	10.24	51.9
CGAΨGCG GCUACGC	86.5 ± 1.6	241.8 ± 4.9	11.45 ± 0.08	55.7	87.4 ± 0.4	244.8 ± 1.4	11.49 ± 0.03	55.7	78.3	215.5	11.40	57.6
CGCΨGCG GCGACGC	68.1 ± 9.6	179.8 ± 28.4	12.35 ± 0.83	66.0	68.5 ± 2.5	181.0 ± 7.4	12.37 ± 0.26	65.9	74.8	201.2	12.45	63.6
CGGΨGCG GCCACGC	82.0 ± 1.7	223.7 ± 5.0	12.59 ± 0.12	61.7	81.2 ± 0.4	221.4 ± 1.3	12.53 ± 0.03	61.8	81.2	220.5	12.81	63.0
CGUΨGCG GCAACGC	69.5 ± 9.6	190.8 ± 29.5	10.34 ± 0.57	55.0	69.0 ± 3.7	189.2 ± 12.0	10.33 ± 0.15	55.1	75.9	212.3	10.11	52.4
CGAΨUCG GCUAAGC	95.4 ± 1.0	274.1 ± 3.1	10.42 ± 0.04	50.2	95.2 ± 0.6	273.2 ± 2.0	10.41 ± 0.02	50.2	91.6	260.2	10.86	52.4
CGCΨUCG GCGAAGC	97.5 ± 1.0	275.6 ± 3.1	12.06 ± 0.06	55.6	96.4 ± 0.5	272.0 ± 1.6	12.00 ± 0.02	55.7	88.2	245.9	11.91	57.2
CGGΨUCG GCCAAGC	93.8 ± 0.7	262.2 ± 2.0	12.51 ± 0.04	58.1	93.5 ± 0.3	261.3 ± 0.9	12.49 ± 0.02	58.1	94.5	265.2	12.27	57.1
CGUΨUCG GCAAAGC	86.8 ± 0.4	249.2 ± 1.3	9.51 ± 0.01	48.0	85.9 ± 0.5	246.5 ± 1.7	9.49 ± 0.01	48.0	89.3	256.9	9.58	48.0
Terminal												
GCGCAΨ CGCGUA	65.4 ± 0.6	178.8 ± 1.9	9.90 ± 0.03	53.9	64.6 ± 0.6	176.4 ± 1.8	9.87 ± 0.02	53.9	71.3	198.8	9.71	51.4
GCGCCΨ CGCGGA	76.7 ± 0.8	213.0 ± 2.4	10.63 ± 0.04	54.5	76.7 ± 0.5	213.1 ± 1.5	10.63 ± 0.01	54.5	68.5	185.8	10.84	57.7
GCGCGΨ CGCGCA	70.6 ± 0.9	193.9 ± 2.7	10.43 ± 0.05	55.1	70.6 ± 0.6	194.0 ± 1.9	10.43 ± 0.01	55.1	73.5	203.3	10.46	54.6
GCGCUΨ CGCGAA	70.9 ± 0.7	201.0 ± 2.1	8.55 ± 0.02	46.1	70.8 ± 0.4	200.8 ± 1.3	8.55 ± 0.02	46.1	70.1	198.7	8.50	46.0
ΨAGCGC AUCGCG	70.8 ± 0.8	199.3 ± 2.5	8.93 ± 0.03	47.9	71.0 ± 0.5	200.0 ± 1.7	8.94 ± 0.01	47.9	61.8	170.2	8.99	49.8
ΨCGCGC AGCGCG	63.7 ± 0.4	174.6 ± 1.3	9.56 ± 0.02	52.5	63.6 ± 0.3	174.3 ± 1.1	9.55 ± 0.01	52.5	66.8	184.1	9.66	52.3
ΨGGCGC ACCGCG	67.6 ± 1.2	184.7 ± 3.6	10.30 ± 0.06	55.3	66.9 ± 0.8	182.6 ± 2.6	10.27 ± 0.02	55.3	63.4	171.4	10.27	56.3
ΨUGCGC AACGCG	66.1 ± 0.5	181.6 ± 1.4	9.77 ± 0.02	53.0	65.6 ± 0.3	180.2 ± 1.0	9.75 ± 0.01	53.0	76.2	214.6	9.65	50.2

[a]Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na$_2$EDTA (pH 7.0).
[b]Predicted using newly derived nearest-neighbor parameters for Ψ-A base pairs.
[c]Calculated for 10^{-4} M total strand concentration.

TABLE 2. Difference in free energy of duplex formation for Ψ-A duplexes compared with predicted U-A counterparts

Duplex	$-\Delta G°_{37}$ (kcal/mol)[a]	$-$NN U-A (kcal/mol)[b]	ΔU-A (kcal/mol)[c]
Internal			
CGAΨACG GCUAUGC	9.93	7.65	−2.28
CGCΨACG GCGAUGC	10.96	9.70	−1.26
CGGΨACG GCCAUGC	11.71	9.70	−2.01
CGUΨACG GCAAUGC	9.10	7.37	−1.73
CGAΨCCG GCUAGGC	11.92	9.69	−2.23
CGCΨCCG GCGAGGC	12.93	11.74	−1.19
CGGΨCCG GCCAGGC	12.76	11.74	−1.02
CGUΨCCG GCAAGGC	9.76	9.41	−0.35
CGAΨGCG GCUACGC	11.45	9.61	−1.84
CGCΨGCG GCGACGC	12.35	11.66	−0.69
CGGΨGCG GCCACGC	12.59	11.66	−0.93
CGUΨGCG GCAACGC	10.34	9.33	−1.01
CGAΨUCG GCUAAGC	10.42	7.36	−3.06
CGCΨUCG GCGAAGC	12.06	9.41	−2.65
CGGΨUCG GCCAAGC	12.51	9.41	−3.10
CGUΨUCG GCAAAGC	9.51	7.08	−2.43
Average			−1.74
Terminal			
GCGCAΨ CGCGUA	9.90	7.87	−2.03
GCGCCΨ CGCGGA	10.63	10.00	−0.63
GCGCGΨ CGCGCA	10.43	9.26	−1.17
GCGCUΨ CGCGAA	8.55	7.67	−0.88
ΨAGCGC AUCGCG	8.93	8.07	−0.86
ΨCGCGC AGCGCG	9.56	9.37	−0.19
ΨGGCGC ACCGCG	10.30	10.03	−0.27
ΨUGCGC AACGCG	9.77	7.70	−2.07
Average			−1.01

[a]Measured value for $-\Delta G°_{37}$ using T_M vs. log $C_T/4$ plots.
[b]Predicted using nearest-neighbor parameters by substituting U-A for Ψ-A base pairs (Xia et al. 1998).
[c]Difference between observed $-\Delta G°_{37}$ values and the predicted $-\Delta G°_{37}$ using U-A pairs in place of Ψ-A pairs, where negative values indicate that duplexes containing Ψ-A pairs are more stable than the corresponding duplexes containing U-A pairs.

suggesting duplex formation occurred instead of unimolecular folding. Additionally, all melt curves exhibited sharp first derivative plots with a single maximum, indicative of a single helix-to-coil transition. Lastly, all of the enthalpy values derived via van't Hoff analyses were within 15% of the enthalpy values derived from individual melt curve fits, also suggesting that the duplexes melted in a two-state manner (Schroeder and Turner 2009).

Comparison of Ψ-A stability to predicted U-A counterparts

Ubiquitously, duplexes containing Ψ-A base pairs were more stable than predicted for their canonical counterparts containing U-A pairs (Table 2). This is in agreement with previous studies that have shown that pseudouridylation almost always results in imparting stability that is not present with canonical U-A base pairs (with notable exception being pseudouridylation of nonclosing base pairs on some hairpins) (Meroueh et al. 2000; Ofengand et al. 2001). The source of this increased stability has been attributed to several different

TABLE 3. Experimentally derived nearest-neighbor parameters for Ψ-A pairs

Nearest neighbors[a]	No. of occurrences[b]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°_{37}$ (kcal/mol)	$\Delta\Delta G°_{37}$ U-A (kcal/mol)[c]
AΨ UA	5	−22.08	−62.21	−2.80	−1.70
CΨ GA	5	−16.23	−43.46	−2.77	−0.69
GΨ CA	5	−24.07	−66.99	−3.29	−1.05
UΨ AA	5	−20.81	−61.94	−1.62	−0.69
ΨA AU	5	−12.47	−33.47	−2.10	−0.77
ΨC AG	5	−17.29	−47.72	−2.49	−0.14
ΨG AC	5	−11.19	−29.03	−2.20	−0.09
ΨU AA	5	−26.94	−78.09	−2.74	−1.81
Terminal Ψ · A	8	−2.04	−7.65	0.31	−0.14

[a]For each nearest-neighbor pair, the top sequence is written 5′ to 3′, and the bottom sequence is written 3′ to 5′. The error values for all nearest-neighbor values are 4.00 kcal/mol, 13.06 eu, and 0.15 kcal/mol for $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$, respectively. The error values for the terminal penalty are 2.78 kcal/mol, 9.09 eu, and 0.11 kcal/mol for $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$, respectively.
[b]The number of times the nearest-neighbor pair appears in the duplexes studied by optical melting experiments.
[c]Calculated by subtracting the value of $\Delta G°_{37}$ for the canonical parameter from the value of $\Delta G°_{37}$ for the analogous Ψ-A parameter, where negative values indicate that Ψ-A parameters are more stabilizing than the corresponding U-A parameters.

sources (Davis 1995; Charette and Gray 2000). Perhaps the most obvious source of stability is the creation of the new hydrogen bond donor at the relocated N1H imino group. Several studies have confirmed via temperature-dependent NMR studies that the N1H imino group coordinates a water molecule to the phosphodiester backbone of RNA. Furthermore, the structure of A-form RNA appears to be ideally suited to protect this coordinated water molecule from exchanging with bulk solvent by limiting access to the major groove, thus conferring exceptional stability due to this hydrogen bond (Hall and McLaughlin 1992; Schroeder et al. 2005).

Since the novel hydrogen bonding capabilities are unlikely to be the only contributor of stabilization in duplexes containing Ψ, an additional and likely source is enhanced stacking interactions. Previous studies involving NMR spectra of RNAs containing Ψ suggest that pseudouridylation causes the residue's sugar to favor the C3′-endo conformation by constraining the conformational freedom of the RNA (Davis 1998; Charette and Gray 2000), resulting in negentropy, which is consistent with our data (Supplemental Table S2). This entropic tradeoff is countered by the fact that the C3′-endo conformation is strongly correlated to enhanced stacking between the attached nucleotide and neighboring bases, with literature reporting a more pronounced enhancement on preceding, i.e., 5′, nucleotides (Davis 1995). This is mostly consistent with the derived Ψ-A nearest-neighbor parameters, which demonstrate stronger stabilization afforded by $\begin{smallmatrix} N\Psi \\ NA \end{smallmatrix}$ (3′-stacking Ψ-A) parameters compared with the stability imparted by $\begin{smallmatrix} \Psi N \\ AN \end{smallmatrix}$ (5′-stacking Ψ-A) parameters (Supplemental Table S4). This effect seems exclusive to Ψ-A stacking; when base pair order (and thus stacking) is reversed (i.e., going from the 5′ $\begin{smallmatrix} \Psi N \\ AN \end{smallmatrix}$ doublet to the 3′ $\begin{smallmatrix} N\Psi \\ NA \end{smallmatrix}$ doublet), there is a significant difference in stabilities of up to 1.1 kcal/mol (Equation 1):

$$\Delta G^{\circ}_{37}\left(\begin{smallmatrix} G\Psi \\ CA \end{smallmatrix} \right) - \Delta G^{\circ}_{37}\left(\begin{smallmatrix} \Psi G \\ AC \end{smallmatrix} \right) = 1.09 \text{ kcal/mol.} \quad (1)$$

This is strikingly different from canonical doublets containing U-A, which only display increased stability with more hydrogen bonding (i.e., a G-C base pair) and do not differ significantly (<0.30 kcal/mol) when the same stacking reversal is performed (Equation 2):

$$\Delta G^{\circ}_{37}\left(\begin{smallmatrix} GU \\ CA \end{smallmatrix} \right) - \Delta G^{\circ}_{37}\left(\begin{smallmatrix} UG \\ AC \end{smallmatrix} \right) = -0.13 \text{ kcal/mol.} \quad (2)$$

Nearest-neighbor parameters for Ψ-A doublets

As can be seen in Table 3, the linearly independent nearest-neighbor parameters for internal Ψ-A base pairs contribute anywhere from −1.62 to −3.29 kcal/mol to duplex stability. This is in stark contrast to the analogous canonical U-A doublets, which only contribute −0.93 to −2.35 kcal/mol to du-

plex stability (Xia et al. 1998). Unexpectedly, Ψ-A doublets containing G-C pairs were not always more stable than those containing A-U pairs (as is the case with canonical bases) (Table 3; Supplemental Table S4), indicating that the stability pseudouridylation imparts is highly dependent on sequence and cannot simply be attributed to water coordination. This observation also suggests that pseudouridylation significantly alters Ψ's stacking interactions. The penalty added for terminal Ψ-A pairs, 0.31 kcal/mol, is slightly less destabilizing than U-A pairs, suggesting that Ψ is not as susceptible to terminus-based destabilization (fraying) as U-A pairs are.

As stated previously, an unexpected feature of the derived Ψ-A parameters is that there is a significant difference between 5′- and 3′-stacked Ψ-A doublets (e.g., comparing $\begin{smallmatrix} \Psi U \\ AA \end{smallmatrix}$ vs. $\begin{smallmatrix} U\Psi \\ AA \end{smallmatrix}$) in stability contributed to duplex formation. This feature is not present in canonical doublets, which differ by <0.30 kcal/mol when the nucleotide order is reversed. Since a water-mediated hydrogen bond is likely to be constant and unaffected by neighboring intrastrand nucleotides, such a difference in energetics is likely due to novel stacking interactions as a result of the relocation of the N1 during pseudouridylation. Nucleotide stacking can be thought to be composed of multiple components, including electrostatics (dipole–dipole interactions) and electrodynamics (induced dipoles, van der Waals interactions, etc.) (Hill et al. 2003). It is particularly revealing to compare the electrostatic maps of uridine and Ψ, which were generated using atomic substitution in conjunction with previously optimized geometries for U (Johnson et al. 2011), operating under the assumption that Ψ assumes the same geometry as the parent U residue (Fig. 2). The relocation of the imino group during pseudouridylation likely leads to different stacking interactions between the resultant Ψ residue and neighboring bases due to the change in dipole and charge localization. For instance, the new electron-deficient imino group becomes situated exactly above the C5-C6 double bond of pyrimidines and directly above the electron rich ring system of purines when Ψ is 3′-stacked on these bases (Fig. 3A,B), possibly resulting in dipole interactions or increased polarizability and thus stronger van der Waals interactions, which are not present in uridine. Such an interaction would not be possible for 5′-stacked Ψ as the N1 is located in "dead space" and cannot directly stack with the neighboring nucleotide (Fig. 3C,D);
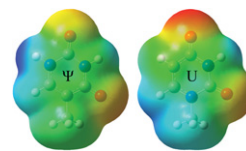
FIGURE 2. Electrostatic maps for Ψ and uridine using a methyl group to simulate the electronic contribution of the ribose ring. These maps were generated in *Gaussian 09* (Frisch et al. 2009) using previously computed geometries for uridine (Johnson et al. 2011) with the corresponding atoms being substituted to produce Ψ.
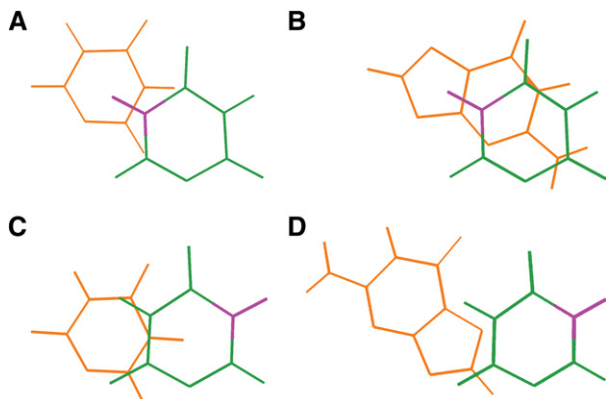
FIGURE 3. The ability of the relocated imino group (purple) on Ψ (green) to impact stacking interactions is likely decided by the positioning (5′ or 3′) of Ψ relative to the adjacent base (orange). Stacks shown are as follows: (A) 5′-UΨ; (B) 5′-GΨ; (C) 5′-ΨU; and (D) 5′-ΨG. These stacks were generated using *InsightII* (Accelrys) for geometries and PyMOL (Schrodinger 2010) for visualization.

this offers a convenient explanation for why 3′-stacked Ψ-A parameters tend to offer additional stability to duplex formation compared to their 5′-stacked counterparts. Another plausible mechanism of stabilization is that the relocated imino group is able to pull electron density away from the O4 of Ψ and thus reduce clashing with the O6 of a guanosine or O4 of a uridine in either direction. It should be noted that these preliminary electronic arguments are based on an assumption that Ψ adopts a similar geometry to U in an A-form duplex. It is unknown how correct this assumption is due to the fact that no structures in the Protein Data Bank contain Ψ without nearby motifs, ligands, etc., which may significantly perturb standard A-form geometry. Due to the unknown validity of this assumption and the fact that a thorough survey of Ψ electronic properties is beyond the scope of this investigation, such explanations for Ψ's stability must be taken as purely hypothetical; however, electronic arguments offer a convenient starting point in justifying the large range of differences in stability observed for Ψ-A compared with U-A.

Biochemical impact of Ψ-A nearest-neighbor parameters

Given the pervasiveness of Ψ residues in biological systems, the ability to predict the stability of pseudouridylated RNA may aid in both modeling RNA folding as well the design of synthetic RNA for use in laboratory or clinical settings. The newly derived parameters for Ψ-A pairs may be used to determine the stability of any RNA molecule containing nonconsecutive Ψ-A base pairs; specifically, free energy minimization can be used to predict how base modification affects RNA secondary structure, the strength of siRNA's base-pairing, or the stability of tRNA recognition interactions.

MATERIALS AND METHODS

Sequence selection, synthesis, and purification

Oligoribonucleotides containing Ψ-A pairs were designed to encompass all possible combinations of surrounding canonical base pairs, 16 internal Ψ-A duplexes, and eight terminal Ψ-A duplexes, including both 5′ and 3′ terminal Ψ-A pairs. The Ψ-A pair and adjacent base pairs were placed within a stem sequence that consists of G-C pairs. The G-C rich stem was chosen to ensure that the melting temperature would exceed 50°C to avoid issues with misalignment and nonduplex motif formation and to avoid fraying of duplexes during optical melting experiments. Predictive melting temperatures were calculated using nearest-neighbor values by substituting a U-A pair in place of the intended Ψ-A pair. Oligonucleotides containing Ψ were purchased from the Keck Lab at Yale University (New Haven, CT), and complementary strands containing standard nucleotides were ordered from Integrated DNA Technologies. Deprotection and purification of the oligoribonucleotides were performed using previously described, standard procedures (Wright et al. 2007).

Duplex formation, lyophilization, and reconstitution

Single-strand concentrations of the RNA were calculated according to the method previously described (Richards 1975) using the application *RNACalc*, except that the extinction coefficient for uridine (1.0×10^4 $M^{-1} \cdot cm^{-1}$) was used instead of Ψ (8.1×10^3 $M^{-1} \cdot cm^{-1}$) (Hall 1971). Absorbance readings were performed at 85°C in order to disrupt any undesired single-strand folding. Equimolar amounts of complementary single-stranded RNA were mixed to form double-stranded RNA (dsRNA). The total concentration of RNA was then assayed again at 85°C using the average of the single-strand extinction coefficients to calculate the total concentration of RNA. An appropriate volume was dried and reconstituted in 100 μL melt buffer (1 M NaCl, 20 mM sodium cacodylate, 0.5 mM disodium EDTA at pH 7.0); the volume of the aliquot was chosen to ensure that maximum absorbance would be ∼2 absorbance units during melting experiments.

Optical melting experiments

An optical melting scheme was devised that featured three melting experiments using three different concentrations for each melting experiment; successive experiments used the previous experiment's solutions mixed with more buffer to modify the concentration of dsRNA while keeping the salt concentration constant. This scheme resulted in nine different dsRNA concentrations that spanned an approximately 50-fold gamut. Each optical melting experiment was performed by monitoring the absorbance at 280 nm using a Beckman-Coulter DU800 spectrometer equipped with a Beckman-Coulter high performance temperature controller using a heating rate of 1°C/min from 10°C to 90°C.

Thermodynamic analysis of optical melting experiments

The optical melting curves obtained for each duplex were analyzed using *MeltWin* (McDowell and Turner 1996), which fits the sigmoidal melt curves to a two-state model by assuming linear sloping

baselines, as well as temperature independent values for the standard entropy and enthalpy (Petersheim and Turner 1983; McDowell and Turner 1996). Thermodynamic parameters for the melt curves were calculated using the melting temperature (T_M) values at various concentrations according to the method described by Borer et al. (1974):

$$T_M^{-1} = \left(\frac{2.303R}{\Delta H^\circ}\right)\log\left(\frac{C_T}{4}\right) + \left(\frac{\Delta S^\circ}{\Delta H^\circ}\right), \qquad (3)$$

where R is the gas constant 1.987 cal/mol*K. The change in Gibbs' free energy at 37°C was calculated using the following equation:

$$\Delta G^\circ_{37} = \Delta H^\circ - (310.15\ K)\Delta S^\circ. \qquad (4)$$

Values calculated using the melt curve fits were found to be in excellent agreement with the T_M versus log C_T (van't Hoff) plots; however, only the thermodynamic parameters derived from the van't Hoff plots were used for later analyses.

Isolation of thermodynamic contribution of Ψ-A pairs

The observed thermodynamic parameters derived from the optical melting experiments were subjected to analysis using the nearest-neighbor model. Each duplex's thermodynamic contribution can be written in the following form, which sums the contributions of all nearest-neighbor interactions in the RNA duplex:

$$\Delta G^\circ_{37}\begin{pmatrix} CGA\Psi ACG \\ GCUAUGC \end{pmatrix} = \Delta G^\circ_{37,i} + \Delta G^\circ_{37}\begin{pmatrix} CG \\ GC \end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix} GA \\ CU \end{pmatrix}$$
$$+ \Delta G^\circ_{37}\begin{pmatrix} A\Psi \\ UA \end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix} \Psi A \\ AU \end{pmatrix}$$
$$+ \Delta G^\circ_{37}\begin{pmatrix} AC \\ UG \end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix} CG \\ GC \end{pmatrix}, \qquad (5)$$

where $\Delta G^\circ_{37}\begin{pmatrix} CGA\Psi ACG \\ GCUAUGC \end{pmatrix}$ is the observed Gibbs free energy change associated with the duplex (determined via optical melting), $\Delta G^\circ_{37,i}$ is the free energy change penalty for duplex initiation (4.09 kcal/mol) (Xia et al. 1998), and all other parameters are individual nearest-neighbor contributions to the free energy change (Xia et al. 1998). The thermodynamic contribution of Ψ-A nearest-neighbor interactions were isolated for each duplex by subtracting off nearest-neighbor parameters for all canonical base pair doublets:

$$\Psi - A\ \text{Contributions} = \Delta G^\circ_{37}\begin{pmatrix} A\Psi \\ UA \end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix} \Psi A \\ AU \end{pmatrix}$$
$$= \Delta G^\circ_{37}\begin{pmatrix} CGA\Psi ACG \\ GCUAUGC \end{pmatrix} - \Delta G^\circ_{37,i}$$
$$- \Delta G^\circ_{37}\begin{pmatrix} CG \\ GC \end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix} GA \\ CU \end{pmatrix}$$
$$- \Delta G^\circ_{37}\begin{pmatrix} AC \\ UG \end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix} CG \\ GC \end{pmatrix}. \qquad (6)$$

For the duplexes with internal Ψ-A pairs, the value for Ψ-A contribution yielded by this operation contains two nearest-neighbor parameters, in this example, both $\begin{smallmatrix} A\Psi \\ UA \end{smallmatrix}$ and $\begin{smallmatrix} \Psi A \\ AU \end{smallmatrix}$. The same operation performed on duplexes containing terminal Ψ-A pairs resulted in only one Ψ-A nearest neighbor. All of these Ψ-A nearest-neighbor values were then subjected to linear regression in order to determine linearly independent nearest-neighbor parameters. Similar analyses were performed for all duplexes to isolate the entropic and enthalpic contribution of the Ψ-A nearest-neighbor pairs.

Linear regression and derivation of linearly independent Ψ-A nearest-neighbor parameters

The isolated Ψ-A contributions for each duplex were assembled into a matrix in Microsoft Excel and subjected to linear regression using the LINEST function. The variables used in this analysis were all possible Ψ-A nearest-neighbor combinations and an additional terminal parameter if the thermodynamic value was contributed by a terminal Ψ-A pair. For instance, the Ψ-A contributions for the previously used example of $\begin{smallmatrix} CGA\Psi ACG \\ GCUAUGC \end{smallmatrix}$ would be as follows:

$$\Psi - A\ \text{Contributions} = \Delta G^\circ_{37}\begin{pmatrix} A\Psi \\ UA \end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix} \Psi A \\ AU \end{pmatrix}. \qquad (7)$$

These same analyses were performed for values for ΔH° and ΔS°. Linear regression of the entire data set yielded linearly independent nearest-neighbor parameters for all eight possible Ψ-A nearest-neighbor combinations as well as a penalty for a terminal Ψ-A pair.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article. Tables showing enthalpy and entropy comparisons, the isolated Ψ-A contribution to thermodynamics, and the rank order of canonical U-A nearest-neighbor pairs and newly derived Ψ-A nearest-neighbor pairs are available as supporting information in Microsoft Word format.

ACKNOWLEDGMENTS

REFERENCES

Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. *Biochemistry* **32:** 9754–9762.

Bakin A, Lane BG, Ofengand J. 1994. Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* **33:** 13475–13483.

Bilbille Y, Vendeix FAP, Guenther R, Malkiewicz A, Ariza X, Vilarrasa J, Agris PF. 2009. The structure of the human tRNA^Lys3 anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* **37:** 3342–3353.

Borer PN, Dengler B, Tinoco I, Uhlenbeck O. 1974. Stability of ribonucleic-acid double-stranded helices. *J Mol Biol* **86:** 843–853.

Charette M, Gray MW. 2000. Pseudouridine in RNA: What, where, how, and why. *Life* **49:** 341–351.

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.

Davis DR. 1995. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23:** 5020–5026.

Davis DR. 1998. Biophysical and conformational properties of modified nucleosides in RNA (nuclear magnetic resonance studies). In *Modification and editing of RNA* (ed. Grosjean H, Benne R), pp. 85–102. ASM Press, Washington, DC.

Davis DR, Poulter CD. 1991. ^{1}H-^{15}N NMR studies of *Escherichia coli* tRNA^{Phe} from hisT mutants: A structural role for pseudouridine. *Biochemistry* **30:** 4223–4231.

Del Campo M, Kaya Y, Ofengand J. 2007. Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA* **7:** 1603–1615.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al. 2009. *Gaussian 09, revision A.1.* Gaussian, Inc., Wallingford, CT.

Grosjean H, Motorin Y, Morin A. 1998. RNA-modifying and RNA-editing enzymes: Methods for their identification. In *Modification and editing of RNA* (ed. Grosjean H, Benne R), pp. 21–45. ASM Press, Washington, DC.

Hall RH. 1971. *The modified nucleosides in nucleic acids.* Columbia University Press, New York.

Hall KB, McLaughlin LW. 1991. Properties of a U1/mRNA 5′ splice site duplex containing pseudouridine as measured by thermodynamic and NMR methods. *Biochemistry* **30:** 1795–1801.

Hall KB, McLaughlin LW. 1992. Properties of pseudouridine N1 imino protons located in the major groove of A-form RNA duplex. *Nucleic Acids Res* **20:** 1883–1889.

Hill G, Forde G, Hill N, Lester WA Jr, Sokalski WA, Leszczynski J. 2003. Interaction energies in stacking DNA bases? How important are electrostatics? *Chem Phys Lett* **381:** 729–732.

Johnson CA, Bloomingdale RJ, Ponnusamy VE, Tillinghast CA, Znosko BM, Lewis M. 2011. Computational model for predicting experimental RNA and DNA nearest-neighbor free energy rankings. *J Phys Chem B* **115:** 9244–9251.

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.

Meroueh M, Grohar PJ, Qiu J, SantaLucia J Jr, Scaringe SA, Chow CS. 2000. Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res* **28:** 2075–2083.

Ofengand J, Malhotra A, Remme J, Gutgsell NS, Del Campo M, Jean-Charles S, Peil L, Kaya Y. 2001. Pseudouridine and pseudouridine synthases of the ribosome. *Cold Spring Harb Sym* **66:** 147–159.

Petersheim M, Turner DH. 1983. Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, mCCGGAp, ACGGp, CCGGUp, and ACCGGUp. *Biochemistry* **22:** 256–263.

Richards EG. 1975. In *Handbook of biochemistry and molecular biology: Nucleic acids*, 3rd ed. (ed. Fasman GD), Vol. I, pp. 596–599. CRC Press, Cleveland, OH.

Schrodinger LLC. 2010. *The PyMOL Molecular Graphics System, version 1.3r1.*

Schroeder SJ, Turner DH. 2009. Optical melting measurements of nucleic acid thermodynamics. In *Methods in Enzymology* (ed. Herschlag D), Vol. 468, pp. 371–387. Academic Press, Burlington, MA.

Schroeder KT, Skalicky JJ, Greenbaum NL. 2005. NMR spectroscopy of RNA duplexes containing pseudouridine in supercooled water. *RNA* **11:** 1012–1016.

Sipa K, Sochacka E, Kazmierczak-Baranska J. 2007. Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA. *RNA* **13:** 1301–1316.

Vaidyanathan PP, Deutscher MP, Malhotra A. 2007. RluD, a highly conserved pseudouridine synthase, modifies 50S subunits more specifically and efficiently than free 23S rRNA. *RNA* **13:** 1868–1876.

Wright DJ, Rice JL, Yanker DM, Znosko BM. 2007. Nearest neighbor parameters for inosine•uridine pairs in RNA duplexes. *Biochemistry* **46:** 4625–4634.

Xia T, SantaLucia 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.