J|A|C|S COMMUNICATIONS

Published on Web 03/26/2010

Role of Unsatisfied Hydrogen Bond Acceptors in RNA Energetics and Specificity

Nathan A. Siegfried,*,†,§ Ryszard Kierzek,‡ and Philip C. Bevilacqua*,†

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802 and Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Noskowskiego 12/14, Poland

Received December 21, 2009; E-mail: pcb@chem.psu.edu; nsiegfried@unc.edu

Specificity and energetics are key determinants of many biochemical phenomena including substrate selection by enzymes and gene regulation of metabolic activity.[1] RNA plays important roles in much of biology where it serves both as an enzyme (ribozyme) and as a switch (riboswitch) to regulate the activity of genes.[2] These properties parallel those of proteins, but the molecular bases for specificity and energetic determinants are less well understood in RNA. Of central importance to the folding and energetics of RNA are hydrogen bonding, stacking, electrostatics, and steric interactions.[3−5] In an effort to uncover determinants of RNA energetics and specificity, we examine contributions that nucleobase functional groups make, including the consequence of having an unsatisfied hydrogen bonding group in a secondary structure. The results have implications in the specificity of amino acid encoding during protein synthesis.

Previously, we examined energetic coupling along RNA and DNA helices using double mutant cycles and a cooperativity formalism.[6] Folding free energies of DNA and RNA hairpins were evaluated in the absence (WT), presence of one (SM_A or SM_B) or presence of two (DM) base modifications (Figure 1). Thermodynamic impact of modification is represented adjacent to the arrow connecting corners of the box. Cooperativity between interactions is provided by the coupling constant (δ_{AB}).

$$\delta_{AB} = \Delta G_{AB} - [\Delta G_A + \Delta G_B] \quad (1)$$

If interactions do not couple, then the thermodynamic impact of SM_A plus SM_B is the same as that of DM and $\delta_{AB} = 0$. Conversely, if two interactions couple positively, δ_{AB} is negative, and if they couple negatively, δ_{AB} is positive.[6,7]

Standard base pair formation in nucleic acids is accompanied by hydrogen bonding along the Watson−Crick face. We reported that a G to inosine (I) change in an *internal* GC has a large effect ($\Delta G°_{37} = 3.44$ kcal/mol) and positively couples to a neighboring AU (δ_{AB} −1.73 ± 0.1 kcal/mol);[6] this free energy value is nearly twice those reported for an *external* GI (i.e., at the end of the helix),[6,8] and similar position trends have been reported for LNA[9] and base mismatches.[10] The G to I modification removes the exocyclic amine, a functional group that donates a hydrogen bond to the carbonyl of C. The value of 3.44 kcal/mol is much larger than *net* hydrogen bonding values reported for RNA and DNA of only −0.25 to −0.5 kcal/mol[5,6,11] and likely represents coupling from stacking and hydration.[5,11] One goal of the present study is to try to uncouple these effects.

Amino-carbonyl hydrogen bonding in the minor groove of a GC base pair compensates for unfavorable dehydration of hydrogen bond donors and acceptors. Without this compensation, it is likely that (1) unshed waters disrupt local structure (strain effects) and (2) shed but uncompensated waters destabilize the helix (dehydration effects). One strategy to determine if dehydration of the carbonyl of C provides this magnitude of destabilization is to modify the carbonyl to eliminate hydrogen bonding with solvent. To this end, we incorporated 2-thiocytosine (s^2C), in which the carbonyl is changed to a thiocarbonyl, into RNA hairpins and determined stabilities (Figure 2).[12] Sulfur is a poor-hydrogen bond acceptor[13] owing to its size and diffuse electron cloud, factors used to explain its scarcity in globular proteins.[14]

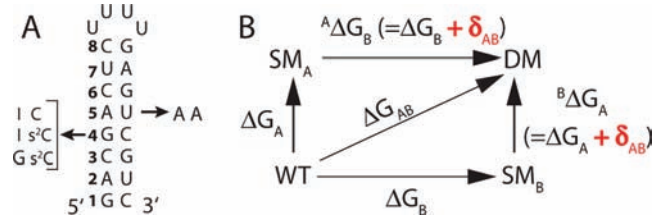

Figure 1. (A) RNA hairpin studied. Arrows indicate mutations. (B) Thermodynamic cycle. Differences in free energy between the corners are plotted on edges. The coupling constant (δ_{AB}) is in red and is the degree to which the two mutations A and B are energetically coupled. WT is wild-type or unmodified RNA; SM_A and SM_B are the single mutants; and DM is the double mutant in which both A and B are changed.

To evaluate effects of s^2C base modification on stability, double mutant cycles were prepared (Figure 2A−D, Supporting Information).[15] If destabilization from a GC to IC mutation (3.44 kcal/mol) (Figure 2A, left edge) is representative of the energy associated with one hydrogen bond, then disrupting this interaction by a GC to Gs^2C change (Figure 2A, bottom edge) should induce a similar energetic penalty. Instead, this led to a destabilization of only 0.60 kcal/mol, a value in agreement with above net energetics for nucleic acid hydrogen bonds. It has been noted previously that the van der Waals radius of a thiocarbonyl is 0.45 Å longer than that for a carbonyl.[16,17] Because of this it is possible that steric clashes could occur upon introducing the thiocarbonyl opposite guanine (Figure 2A, bottom edge). Any such clashes that might occur would serve to destabilize the Gs^2C hairpin, suggesting that the 0.60 kcal/mol is an upper limit to the amount of destabilization (i.e., the intrinsic worth for hydrogen bond deletion in Figure 2A, bottom edge, might be even smaller in magnitude). Moreover, lack of a significant thermodynamic effect upon removing the 2-amino group in going from G to I in the s^2C background (Figure 2A, right edge) suggests that any steric strain from the thiocarbonyl is minimal in the Gs^2C hairpin.

Observation of only a slight destabilization upon introduction of the thiocarbonyl to the GC suggests that most of the GC to IC disruption results from sources other than loss of a single Watson−Crick hydrogen bond. This conclusion is strengthened by

† The Pennsylvania State University.
‡ Polish Academy of Sciences.
§ Present Address: Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina, 27599.

10.1021/ja9107726

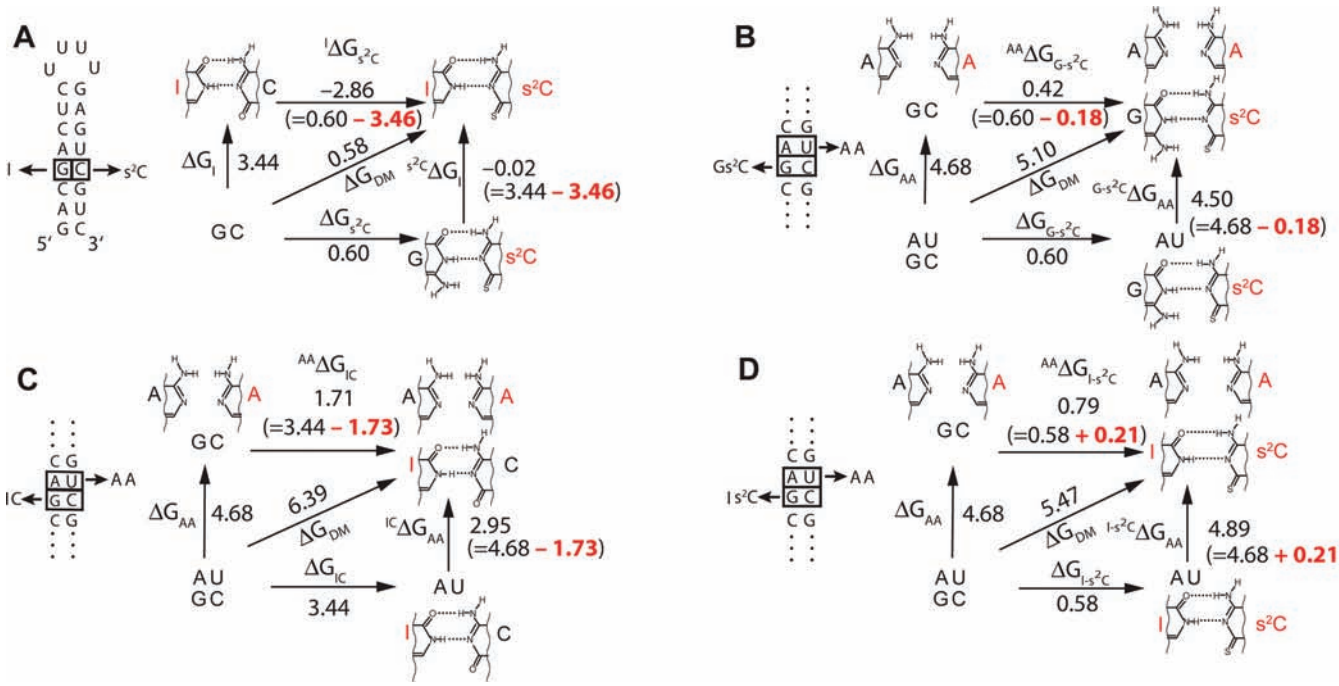

Figure 2. Thermodynamic cycle data. Mutant species are indicated by the arrows emanating from the hairpins on the left side of each panel and on the subscripts on connecting arrows that indicate the directionality of the mutations. (A) Thermodynamic cycle for a GC to Is^{2}C change. (B) Thermodynamic cycle for a 5′GA3′/3′CU5′ to GA/(s^{2}C)A change. (C) Thermodynamic cycle for a GA/CU to IA/CA change, as previously reported.[6] (D) Thermodynamic cycle for a GA/CU to IA/(s^{2}C)A change.

observation that introduction of I across from s^{2}C gives no further energetic disruption (Figure 2A, right edge). This lack of effect is remarkable in that it suggests not only that the hydrogen bonding interaction has already been disrupted in Gs^{2}C but also that stacking and other changes from a G to I change play a minor role in the disruption observed on the opposite edge. Similarity in stacking interactions of G and I in RNA has also been provided by Turner et al. who demonstrated that 3′- (or 5′-) dangling G and I make very similar contributions to duplex stability.[8] Stacking of C and s^{2}C is also similar, as can be inferred from experiments in the present thermodynamic cycles: when s^{2}C incorporation across from G is performed with different nearest neighbors, an AU base pair or AA mismatch, the energetic impact is essentially unchanged at 0.60 and 0.42 kcal/mol, respectively (Figure 2B, bottom and top edges), suggesting that C and s2C have similar stacking preferences.

The above analysis, suggesting that hydrogen bonding and stacking are relatively minor contributors in this system, leaves the unsatisfied carbonyl group on C as the likely source of the disruption on the left edge of Figure 2A. Further support of a major energetic role for the unsatisfied O2 of C is found by considering the top edge of Figure 2A, where an IC to Is^{2}C change significantly *stabilizes* the hairpin ($\Delta G = -2.86$ kcal/mol). If the energetic change associated with a GC to IC change (+3.44 kcal/mol) is approximately the sum of hydrogen bonding (ΔG_{HB}) and hydration (ΔG_{DEHYD}) and ΔG_{HB} is only ~0.60 kcal/mol, then ΔG_{DEHYD} is ~2.84 kcal/mol or the approximate stabilization found in $^{I}\Delta G_{s2C}$ (Figure 2A, top edge). In fact, the overall disruption following these opposing changes (ΔG_{DM}) is just 0.58 kcal/mol (Figure 2A, diagonal), similar to the destabilization observed for incorporation of s^{2}C and in good agreement with literature net energetic values for a single hydrogen bond. Cooperativity in this thermodynamic cycle is very large at −3.46 kcal/mol (Figure 2A), consistent with it representing interplay between two bases in a single base pair.[18]

Results presented to this point support a key energetic role of an unsatisfied hydrogen bond acceptor in a base pair. Broadening the scope of the study, the cycles in Figure 2B−D probe interactions between two different, neighboring base pairs. In Figure 2B, interaction between base pairs 4 and 5 was probed by Gs^{2}C and AA changes, and cooperativity was weak ($\delta = -0.18$ kcal/mol). A similar cycle probing the interaction of base pairs 4 and 5 by IC and AA changes gave greater cooperativity ($\delta = -1.73$ kcal/mol, Figure 2C).[6] Comparison of the thermodynamic cycles in Figure 2B and 2C is consistent with significantly reduced structural disruption in the absence of an unsatisfied carbonyl on cytosine. This supports the general idea that an appropriate structural model for the left edge of Figure 2A includes some dragging of water into the RNA helix. Additional support for this hypothesis is lent by the data in Figure 2D where slight *negative* coupling is observed between neighboring base pairs when Is^{2}C, which has no steric clash, is incorporated adjacent to an AA ($\delta = +0.21$ kcal/mol). From a different perspective the energetic penalty associated with introduction of an AA mismatch (4.68 kcal/mol; Figure 2B−D, left edges) is reduced in the background of an IC (to 2.95 kcal/mol) but not the background of Gs^{2}C or Is^{2}C, consistent with helical disruption from water incorporation in the background of IC.

It is noteworthy that incorporation of s^{2}C across from G (Figure 2A, bottom edge), a mutation that removes the hydrogen bonding partner of the amino group on G, does not exhibit a large destabilization, giving a free energy difference of only 0.6 kcal/mol. This contrasts sharply with the left edge of Figure 2A, where removal of a hydrogen bonding functionality led to the very large 3.4 kcal/mol penalty. This difference can be due to (1) weak hydrogen bonding between the amino on G and the thiocarbonyl on C[13] or (2) extensive hydration of helical RNA allowing the amino group to remain partially hydrated in the helix.[4,19]

Unsatisfied hydrogen bond acceptors likely play major roles in RNA and DNA thermodynamics; moreover, this role is underappreciated since it is typically not recorded in structural studies. A common example is formation of AU and AT base pairs where the O2 group on the pyrimidine does not have a hydrogen bonding partner across

from it and thus experiences a desolvation penalty. Accordingly, replacing U with 2-thiouridine (s^2U) in RNA, or 2-thiothymidine (s^2T) in DNA, should relieve the desolvation penalty, as observed herein for replacing C with s^2C opposite I (Figure 2A, top edge), which gave the remarkable stabilization of −2.9 kcal/mol. Indeed, in RNA duplexes substitution of s^2U in an internal AU provides a stabilization of ∼1 kcal/mol,[20] while introduction of s^2T into DNA duplexes in an internal AT provides stabilization of a 1.4 °C increase in melting temperature, T_m.[17] Likewise, substitution of s^4U in an internal G•U wobble (also a non-hydrogen bonding functionality) in RNA provides an exceptional energetic stabilization of ∼3 kcal/mol,[20] which is similar in magnitude to that in Figure 2A, top edge. A very large energetic stabilization is also observed for a similar change in DNA, where s^4T substitution in a G•T wobble increases T_m by 10 °C.[17] The O4 resides in the major groove of A-form RNA, which is narrow and deep, where retained water might be particularly disruptive,[4,19] consistent with a large stabilization of the RNA fold upon shedding waters. An alternative model in which the s^4U or s^4T base forms an enethiol tautomer and engages in a three-hydrogen bond nonwobble G-s^4U or G-s^4T base pair[17] cannot be excluded. The energetic cost of making an uncompensated hydrogen bond (i.e., two functionalities) in proteins has been estimated at ∼5 kcal/mol;[21] results herein indicate that preparing one unsatisfied hydrogen bonding functionality in RNA costs 2.9 kcal/mol (Figure 2A, top edge), or roughly half of 5, consistent with similar energetics in RNA.

Incorporation of s^2U across from an A in RNA has been observed to be ∼4 times more stabilizing in an internal context than an external one.[20] Likewise, incorporation of s^4U opposite G in a GU mismatch is ∼10 times more stabilizing internally than externally.[20] The studies of related nucleobase analogues by Sintim and Kool in DNA gave large effects, and in their case the substitutions were internally positioned in the helix.[17] Our previous studies and those of others[6,10,22] suggest that positioning of functionalities internally in the helix is a major contributing factor to RNA energetics and specificity and, therefore, that unsatisfied hydrogen bond acceptors will play more significant roles when located internally rather than externally.

It appears that nature has developed ways to use unsatisfied hydrogen bonding groups to tune specificity. Thionucleosides occur exclusively in tRNA, where they have been found to substitute for any of the carbonyls in the pyrmidines: s^2C, s^2U, and s^4U.[23,24] In certain tRNAs, s^2U derivatives occur in the first position of the anticodon, where they decode codons ending with A but not those ending with G.[25] One possibility is that the s^2U derivative favors AU pairing by both destabilizing noncognate GU and stabilizing cognate AU. The s^2U modification has also been shown to increase the affinity of tRNA for its cognate tRNA synthetase.[26]

Observation that thermodynamic penalties in general are larger for internal than external helical positions[6,8,10] suggests that tightly packed tertiary regions may be especially sensitive to unsatisfied functionalities. That similar trends are likely in internal secondary structures, tertiary structure, and both RNA and DNA suggests that effects of unsatisfied hydrogen bond acceptors on energetics and specificity will be a general phenomenon.

Unsatisfied hydrogen bond acceptors contribute exceptionally large magnitudes to the free energies measured herein. While the general nature of this trend is supported by indifference to the nearest neighbors found between this and other studies, further investigations will be necessary to determine the magnitude of this effect in wide-ranging contexts. Considering the vast array of modified bases in nature,[24] the presence of noncanonical interactions in most RNAs,[27] and the large magnitude of energetic effects observed herein, unsatisfied hydrogen bond acceptors are likely to play key roles in the energetics and specificity of many RNA and DNA systems.

Acknowledgment. Supported by NSF MCB-0527102 (P.C.B.) and by the Ministry of Science and Higher Education, Grant PZB-MNiSW-07/1/2007 (R.K.).

Supporting Information Available: List of RNA sequences studied, mass spectrometry data, thermodynamic parameters for hairpin folding, and supporting references. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fersht, A. R.; Shi, J. P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M.; Winter, G. *Nature* **1985**, *314*, 235. Street, I. P.; Armstrong, C. R.; Withers, S. G. *Biochemistry* **1986**, *25*, 6021. Joyce, C. M.; Benkovic, S. J. *Biochemistry* **2004**, *43*, 14317. Rodnina, M. V.; Wintermeyer, W. *Curr. Opin. Cell Biol.* **2009**, *21*, 435.
(2) Doudna, J. A.; Lorsch, J. R. *Nat. Struct. Mol. Biol.* **2005**, *12*, 395. Fedor, M. J. *Annu. Rev. Biophys.* **2009**, *38*, 271. Henkin, T. M. *Genes Dev.* **2008**, *22*, 3383. Roth, A.; Breaker, R. R. *Annu. Rev. Biochem.* **2009**, *78*, 305. Montange, R. K.; Mondragón, E.; van Tyne, D.; Garst, A. D.; Ceres, P.; Batey, R. T. *J. Mol. Biol.* **2010**, *396*, 761.
(3) Turner, D. H.; Bevilacqua, P. C. *Thermodynamic considerations for evolution by RNA. In The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1993; p 447. Turner, D. H.; Mathews, D. H. *Nucleic Acids Res.* **2009**, *38*, D280.
(4) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000; p 794.
(5) Kool, E. T. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 1.
(6) Siegfried, N. A.; Metzger, S. L.; Bevilacqua, P. C. *Biochemistry* **2007**, *46*, 172.
(7) Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. *Cell* **1984**, *38*, 835. Di Cera, E. *Chem. Rev.* **1998**, *98*, 1563. Siegfried, N. A.; Bevilacqua, P. C. *Methods Enzymol.* **2009**, *455*, 365.
(8) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 3783.
(9) Kierzek, E.; Pasternak, A.; Pasternak, K.; Gdaniec, Z.; Yildirim, I.; Turner, D. H.; Kierzek, R. *Biochemistry* **2009**, *48*, 4377.
(10) Kierzek, R.; Burkard, M. E.; Turner, D. H. *Biochemistry* **1999**, *38*, 14214.
(11) Kato, Y.; Conn, M. M.; Rebek, J., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1208. Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 16285.
(12) RNA was synthesized and purified as previously described.[20]
(13) Donohue, J. *J. Mol. Biol.* **1969**, *45*, 231.
(14) Gregoret, L. M.; Rader, S. D.; Fletterick, R. J.; Cohen, F. E. *Proteins* **1991**, *9*, 99.
(15) Folding free energies were determined experimentally through UV-monitored thermal denaturation. Thermal denaturation experiments were carried out in 10 mM sodium phosphate buffer (pH 7.0), 0.1 mM Na_2EDTA, and performed on a Gilford Response II spectrophotometer equipped with a temperature controller.[6]
(16) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; p 556.
(17) Sintim, H. O.; Kool, E. T. *J. Am. Chem. Soc.* **2006**, *128*, 396.
(18) Moody, E. M.; Feerrar, J. C.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 7992.
(19) Auffinger, P.; Westhof, E. *J. Biomol. Struct. Dyn.* **1998**, *16*, 693.
(20) Testa, S. M.; Disney, M. D.; Turner, D. H.; Kierzek, R. *Biochemistry* **1999**, *38*, 16655.
(21) Mitchell, J. B. O.; Price, S. L. *Chem. Phys. Lett.* **1991**, *180*, 517. Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623.
(22) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719.
(23) Carbon, J.; David, H.; Studier, M. H. *Science* **1968**, *161*, 1146. Baczynskyj, L.; Biemann, K.; Hall, R. H. *Science* **1968**, *159*, 1481. Agris, P. F.; Crain, P. F.; Rozenski, J.; Fabris, D.; Vendeix, F. A. P. The RNA Modification Database. http://biochem.ncsu.edu/RNAmods/.
(24) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1994**, *22*, 2183. Rozenski, J.; Crain, P. F.; McCloskey, J. A. *Nucleic Acids Res.* **1999**, *27*, 196.
(25) Sekiya, T.; Takeishi, K.; Ukita, T. *Biochim. Biophys. Acta* **1969**, *182*, 411. Ajitkumar, P.; Cherayil, J. D. *Microbiol. Rev.* **1988**, *52*, 103.
(26) Gustilo, E. M.; Dubois, D. Y.; Lapointe, J.; Agris, P. F. *RNA Biol.* **2007**, *4*, 85.
(27) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497. Holbrook, S. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 302. Mukherjee, S.; Bansal, M.; Bhattacharyya, D. *J. Comput. Aided Mol. Des.* **2006**, *20*, 629. Olson, W. K.; Esguerra, M.; Xin, Y.; Lu, X. J. *Methods* **2009**, *47*, 177. Schroeder, K. T.; Lilley, D. M. *Nucleic Acids Res.* **2009**, *37*, 7281.

JA9107726