Published in final edited form as:

Chem Phys Lett. 2015 October 16; 639: 157–160. doi:10.1016/j.cplett.2015.09.005.

A Computational Model for Predicting Experimental RNA Nearest-Neighbor Free Energy Rankings: Inosine•Uridine Pairs

Elizabeth A. Jolley, Michael Lewis*, and Brent M. Znosko*

Department of Chemistry, Saint Louis University, 3501 Laclede Avenue, Saint Louis, Missouri, 63103, United States

Abstract

A computational model for predicting RNA nearest neighbor free energy rankings has been expanded to include the nonstandard nucleotide inosine. The model uses average fiber diffraction data and molecular dynamic simulations to generate input geometries for Quantum mechanic calculations. This resulted in calculated intrastrand stacking, interstrand stacking, and hydrogen bonding energies that were combined to give total binding energies. Total binding energies for RNA dimer duplexes containing inosine were ranked and compared to experimentally determined free energy ranks for RNA duplexes containing inosine. Statistical analysis showed significant agreement between the computationally determined ranks and the experimentally determined ranks.

Graphical abstract

*Corresponding Author: (M.L.) Lewism5@slu.edu; (B.M.Z.) Znoskob@slu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ASSOCIATED CONTENT

Supporting Information. Supplementary data can be found in the online version.

Notes

The authors declare no competing financial interests.

Keywords

MD simulations; nonstandard nucleotide; RNA; QM calculations; inosine; base stacking

Recently, Johnson et al.¹ developed a method for calculating base stacking and hydrogen bonding energies for Watson-Crick pairs using average fiber diffraction geometries.¹⁻⁴ The study computationally determined energies for all standard nucleotide combinations and successfully compared these computational rankings to experimentally determined nearest neighbor (NN) free energy rankings.^{1,5,6} The method resulted in rankings that agreed well with the experimental rankings, but it has not been tested on non-Watson-Crick pairs.

The thermodynamics of RNA duplexes containing inosine (I) have previously been studied. Wright et al.⁷ experimentally investigated the thermodynamics of duplexes containing I•U pairs and compared the results to the same sequences containing A-U and G-U pairs. Because A-U, G-U, and I•U base pairs are all predicted to have two hydrogen bonds and because I•U is isosteric with G-U (Fig. 1), it was surprising to discover by optical melting experiments that RNA duplexes containing an internal I•U pair were, on average, 2.3 and 1.9 kcal/mol less stable than the same duplexes containing an A-U or G-U pair, respectively.⁷ Although this difference in free energy was observed, optical melting experiments do not provide any evidence suggesting why I•U pairs are less stable. It was hypothesized that I•U pairs have weaker hydrogen bonds and/or do not stack as well as A-U or G-U pairs.⁷

There are many options for generating input geometries to be used for computational chemistry. Using the actual geometry from a three-dimensional structure solved by NMR or X-ray crystallography is one possibility. Unfortunately, a search of the Protein Data Bank (PDB)⁸ resulted in only one entry containing an I•U pair; however, this entry contained tandem I•U pairs, not an I•U pair with Watson-Crick neighbors (as was used for the thermodynamic studies reported by Wright et al.⁷). Another source of input geometries, frequently used by Šponer et al.⁹⁻¹¹, employs molecular dynamics (MD) simulations. This method allows for the generation of an input geometry for a motif that may otherwise have no solved structures.

The study herein utilizes the Johnson et al.¹ computational approach to calculate the hydrogen bonding and base stacking energies for I•U pairs adjacent to all possible Watson-Crick neighbors. In order to expand the Johnson et al.¹ method to nonstandard nucleotides, this study used both average fiber diffraction data for Watson-Crick pairs and stacks, as well as MD simulations to create input geometry coordinates for RNA duplexes containing I•U pairs. Quantum mechanical energy calculations of the resulting structures were combined to give a total binding energy for each NN combination containing an I•U pair. The total binding energies were then ranked and compared to the experimental free energy rankings previously determined for I•U pairs.⁷ Input geometries include average fiber diffraction geometries for the standard bases as well as MD simulations for pairs and/or stacks involving I.

InsightII (Accelrys) was used to build all 11 possible NN dimer duplexes (Fig. 2) containing at least one A-U pair since only the four standard nucleotides are available in *InsightII*. The A's were replaced by I's, resulting in I•U pairs. When the A in an A-U pair is replaced by I, the resulting I•U conformation is not oriented correctly for a two-hydrogen bond I•U pair (compare Fig. 1 (left) to Fig. 1 (middle)). To address this, duplexes were imported into PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.) and the I base manipulated to be in proximity with the corresponding U base for optimal hydrogen bonding (hydrogen bond donors and acceptors of I and U were positioned within 3 Å of each other), resulting in an I•U conformation similar to Fig. 1 (middle). These manipulated structures were then minimized using AMBER.¹² Topology and coordinate files¹³ for the I containing duplexes were created using xLEaP.¹² Sodium ions were added to structures to neutralize charge, and the structures were placed in a truncated octahedral TIP3P solvation box with a periodic boundary of 8.0 Å. Minimizations were performed using the SANDER module of AMBER¹² using the ff10 force field.¹⁴ Minimizations were carried out with the solute having position restraints on every residue except the I base. The final dimer structures then contain all atoms restrained to the average fiber diffraction geometry, with the exception of the I. The I stacks from the resulting structures were used as the input geometries for QM calculations.

As previously described by Johnson et al.¹, the sugar-phosphate backbone was deleted and replaced with a hydrogen atom. Hydrogen bond energies, interstrand binding energies, and intrastrand binding energies for pairs or stacks containing I (Fig. 2) were obtained by calculating the individual monomer and dimer energies and then subtracting the monomer energies from the dimer energies. Calculations were performed at the MP2(full)/6-311G** level of theory using the Gaussian 09 software package.¹⁵ All binding energies were corrected for basis set superposition error (BSSE) using the counter-poise method.¹⁶ Because all nucleotides other than I were restrained to their average fiber diffraction geometries during the generation of input geometries, binding energies for any stacking interactions and any hydrogen bonding pairs not containing I were taken from Johnson et al.¹ The I•U hydrogen bond energy from each NN dimer duplex was calculated and then averaged to give the final I•U hydrogen bond energy, -2.47 kcal/mol per I•U pair. Once calculated, the appropriate hydrogen bonding energies, interstrand binding energies, and intrastrand binding energies were combined to give the total binding energy (denoted E_{total} in Table 1) of a NN dimer duplex, as given in eq 1.¹ B_1 - B_2 and B_3 - B_4 are hydrogen bond energies, B_1B_3 and B_4B_2 are intrastrand binding energies, and $5'B_1-3'/3'_-B_45'$ and $5'_-B_23'/3'B_3-5'$ are interstrand binding energies. To be consistent with the process for determining experimental NN values, the total hydrogen bonding energy for a base pair is equally distributed between the two nearest neighbor parameters that base pair is involved in; therefore, in eq 1, the hydrogen bonding energies are halved.¹⁷

$$E \left(\begin{array}{cc} 5'B_1 & B_23' \\ 3'B_3 & B_45' \end{array} \right) = 0.5 \cdot [E(B_1-B_2) + E(B_3-B_4)] + E(B_1B_3) + E(B_4B_2) + E \left(\begin{array}{c} 5'B_1-3' \\ 3'_-B_45' \end{array} \right) + E \left(\begin{array}{c} 5'_-B_23' \\ 3'B_3-5' \end{array} \right) \quad (1)$$

Table 1 shows the results of the calculated base stacking energies for the intrastrand and interstrand binding energies for the stacks containing I. Table 2 shows the total energies for

all NN dimer duplexes containing I, resulting from the summation of two hydrogen bonding terms, two intrastrand stacking terms, and two interstrand stacking terms, as described in eq 1. Table 2 also shows that the NN dimer duplexes containing one G-C pair are most stabilizing, and so ranked higher, than those with one A-U pair or two I•U pairs. Similarly, the NN dimer duplexes containing one A-U pair are more stabilizing, and so ranked higher, than those with two I•U pairs.

The total binding energies were given a rank order and compared with the experimental rank order previously determined for nearest neighbor combinations containing I•U pairs⁷ (Table 2). The ranks were compared by using mean absolute deviation (MAD) values as well as Spearman rank correlation coefficient (r_s) values. The MAD value (1.3) represents the deviation from the mean of the computationally determined ranks and the experimentally determined ranks, thus smaller MAD values mean less difference between the calculated rank and experimental rank. The Spearman rank correlation coefficient (0.89) is a measure of the association between ranked data sets to test a hypothesis of no association between data sets. For an 11 data point rank set, an r_s value 0.818 shows the no association hypothesis can be rejected at the 99.5% confidence level.¹⁸ Thus, there is a statistically significant association between the computationally determined ranks and the experimentally determined ranks. This is in agreement with the results from the Johnson et al.¹ study which reports a MAD value of 1.0 and a Spearman rank correlation coefficient of 0.88 for the calculated total binding energies for RNA dimer duplexes containing Watson-Crick pairs. A limitation of the approach used here to generate input geometries for the I stack/pairs is the rigidity of the other three nucleotides. The effects of this limitation may result in a conformation that is not truly representative of the conformation in solution. Although more rigorous approaches could be used, statistical analysis of the ranks shows that this simplified approach agrees quite well with the experimental values, giving a good correlation.

In order to further understand if the model used here captures the changes that occur when an A-U pair is replaced with an I•U pair, it is useful to compare the experimental difference between these pairs to the computational difference between these pairs. Table 3 shows the experimentally determined free energies and ranks for nearest neighbor dimer duplexes containing A-U pairs and for the nearest neighbor dimer duplexes containing I•U pairs. Table 3 also gives the computationally determined E_{total} energies and ranks for the same nearest neighbor dimer duplexes. The NN dimer duplexes included in the table are those containing one G-C pair with one I•U pair and those containing one A-U pair with one I•U pair. The NN dimer duplexes containing tandem I•U pairs were not included as these pairs contain a compounding of the effect seen by the replacement of an A-U pair with an I•U pair. The MAD value for the agreement between the experimentally determined free energy ranks of the A-U NN dimer duplexes and the experimentally determined free energy ranks of the I•U NN dimer duplexes is 1.1. The MAD value for the agreement between the computational energy ranks for A-U NN dimer duplexes and I•U NN dimer duplexes is 0.9. The overall difference in the experimentally determined rank orders of the A-U and I•U dimer duplexes is relatively the same as the overall difference in the computationally determined rank orders.

This study successfully expanded the previously published computational approach¹ derived for Watson-Crick pairs to dimer duplexes containing at least one non-Watson-Crick pair, in particular, an I•U pair. However, further work will need to be done to explore the decrease in stability when an I•U pair replaces an A-U or G-U pair in an RNA duplex. For example, the current study computed energies based on dimer duplexes, which may not account for all of the interactions involved in longer duplexes. Computing energies for longer sequences may be necessary to fully understand the stacking or hydrogen bonding interactions that may be responsible for the decreased stability of a duplex due to an I•U pair.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was funded by the National Institute of General Medical Sciences of the National Institutes of Health via Grant R15GM085699.

REFERENCES

1. Johnson CA, Bloomingdale RJ, Ponnusamy VE, Tillinghast CA, Znosko BM, Lewis M. Computational model for predicting experimental RNA and DNA nearest-neighbor free energy rankings. *J. Phys. Chem. B*. 2011; 115:9241–9255.
2. Fiethen A, Jansen G, Hesselmann A, Schütz M. Stacking energies for average B-DNA structures from the combined density functional theory and symmetry-adapted perturbation theory approach. *J. Am. Chem. Soc.* 2008; 130:1802–1803. [PubMed: 18201088]
3. Kamy PRN, Muchall HM. Revisiting the effects of sequence and structure on the hydrogen bonding and π -stacking interactions in nucleic acids. *J. Phys. Chem. A*. 2011; 115:12800–12808. [PubMed: 21721560]
4. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolberger C, Berman HM. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 2001; 331:229–237. [PubMed: 11601858]
5. Šponer J, Morgado CA, Svozil D. Comment on “Computational model for predicting experimental RNA and DNA nearest-neighbor free energy rankings”. *J. Phys. Chem. B*. 2012; 116:8331–8332. [PubMed: 22686484]
6. Johnson CA, Bloomingdale RJ, Ponnusamy VE, Tillinghast CA, Znosko BM, Lewis M. Reply to “Comment on ‘Computational model for predicting experimental RNA and DNA nearest-neighbor free energy rankings’”. *J. Phys. Chem. B*. 2011; 116:8333–8332.
7. Wright DJ, Rice JL, Yanker DM, Znosko BM. Nearest neighbor parameters for inosine-uridine pairs in RNA duplexes. *Biochemistry*. 2007; 46:4625–4634. [PubMed: 17378583]
8. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
9. Krepl M, Otyepka M, Banáš P, Šponer J. Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: molecular dynamic Thermodynamics integration study. *J. Phys. Chem. B*. 2013; 117:1872–1879. [PubMed: 23350675]
10. Svozil D, Hobza P, Šponer J. Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations. *J. Phys. Chem. B*. 2010; 114:1191–1203. [PubMed: 20000584]
11. Morgado CA, Svozil D, Turner DH, Šponer J. Understanding the role of base stacking in nucleic acids. MD and QM analysis of tandem GA base pairs in RNA duplexes. *Phys. Chem. Chem. Phys.* 2012; 14:12580–12591. [PubMed: 22722325]

12. Case, DA.; Darden, TA.; Cheatham, TE., III; Simmerling, CL.; Wang, J.; Duke, RE.; Luo, R.; Walker, RC.; Zhang, W.; Merz, KM.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, AW.; Kolossváry, I.; Wong, KF.; Paesani, F.; Vanicek, J.; Wolf, RM.; Liu, J.; Wu, X.; Brozell, SR.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, MJ.; Cui, G.; Roe, DR.; Mathews, DH.; Seetin, MG.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, PA. AMBER 12. San Francisco, CA: University of California; 2012.
13. Aduri R, Psciuk BT, Saro P, Taniga H, Schlegel HB, SantaLucia J Jr. AMBER force field parameters for the naturally occurring modified nucleosides in RNA. *J. Chem. Theory Comput.* 2007; 3:1464–1475.
14. Zgarbová M, Otyepka M, Šponer J, Mládek A, Banás P, Cheatham III TE, Jurek P. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem Theory Comput.* 2011; 7:2886–2902. [PubMed: 21921995]
15. Frisch, MJ., et al. *Gaussian 09*, revision A.1.. Wallingford, CT: Gaussian, Inc; 2009.
16. Boys SF, Bernardi F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with errors. *Mol. Phys.* 1970; 19:553–566.
17. Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner D. H. *Biochemistry.* 1998; 37:14719–14735. [PubMed: 9778347]
18. Mendenhall, W. *Introduction to Probability and Statistics.* 2nd ed.. Belmont, CA: Wadsworth Publishing Company, Inc.; 1968. p. 314-317.

Highlights

- Computational model for predicting RNA nearest neighbor free energy rankings.
- Expanded to include nonstandard RNA nucleotides.
- Comparison of computational energy rankings to experimental energy rankings.

Figure 1.

Structure of an A-U (left), I•U (middle), and G-U (right) pair. Notice that all three pairs have two hydrogen bonds and that the I•U pair is isosteric with the G-U pair.

Figure 2. Dimer duplex of sequence 5'-B₁B₃-3'/3'-B₂B₄-5'. Red lines represent intrastrand stacking, green lines represent interstrand stacking, and blue lines represent hydrogen bonding interactions. (Figure taken from Ref. 1)

Table 1

Intrastrand and interstrand binding energies for stacks containing inosine.

Intrastrand		Interstrand	
Stacks with I	E_{bind} (kcal/mol)	Stacks with I	E_{bind} (kcal/mol)
5'A 3'I	-1.77	5'I_3' 3'_A5'	-1.21
5'I 3'A	-4.70	5'_I3' 3'_A_5'	-1.75
5'C 3'I	-0.07	5'I_3' 3'_C5'	2.46
5'I 3'C	-3.83	5'_I3' 3'_C_5'	-1.45
5'G 3'I	-1.63	5'I_3' 3'_G5'	-1.22
5'I 3'G	0.35	5'_I3' 3'_G_5'	1.81
5'U 3'I	1.23	5'I_3' 3'_U5'	1.09
5'I 3'U	-1.98	5'_I3' 3'_U_5'	0.81
5'I 3'I	-1.67	5'I_3' 3'_I5'	2.96
		5'_I3' 3'_I_5'	0.01

Table 2

Experimentally determined nearest neighbor free energies and ranks (from most stable to least stable) and computationally determined total energies and ranks.

NN Dimer duplexes	Experimental G_{37}° (kcal/mol)	Rank	Computational E_{total} (kcal/mol)	Rank
5'GC3' 3'IU5'	-1.34	1	-13.73	2
5'IU3' 3'GC5'	-1.22	2	-11.58	3
5'IU3' 3'CG5'	-1.03	3	-20.12	1
5'CG3' 3'IU5'	-0.77	4	-11.35	4
5'IU3' 3'UA5'	-0.50	5	-7.82	6
5'AU3' 3'IU5'	-0.41	6	-3.82	7
5'UA3' 3'IU5'	0.37	7	-3.15	8
5'IU3' 3'AU5'	0.43	8	-8.55	5
5'UI3' 3'IU5'	2.23	9	3.11	11
5'IU3' 3'IU5'	2.66	10	-0.04	9
5'IU3' 3'UI5'	3.58	11	0.09	10
MAD ^a				1.3
r_s^b				0.89

^a Mean absolute deviation

^b Spearman rank correlation coefficient

Table 3
Experimental A-U and I•U NN comparison and computational A-U and I•U NN comparison.

Experimental					Computational				
NN	$G_{37}^{\circ,17}$ (kcal/mol)	Rank	NN	$G_{37}^{\circ,7}$ (kcal/mol)	Rank	NN	E_{total}^1 (kcal/mol)	Rank	E_{total} (kcal/mol)
5'GC3' 3'AU5'	-2.35	1	5'GC3' 3'AU5'	-1.34	1	5'GC3' 3'AU5'	-20.19	4	5'GC3' 3'IU5'
5'AU3' 3'GC5'	-2.08	4	5'IU3' 3'GC5'	-1.22	2	5'AU3' 3'GC5'	-23.10	2	5'IU3' 3'GC5'
5'AU3' 3'CG5'	-2.24	2	5'IU3' 3'CC5'	-1.03	3	5'AU3' 3'CG5'	-23.12	1	5'IU3' 3'CG5'
5'CG3' 3'AU5'	-2.11	3	5'CG3' 3'IU5'	-0.77	4	5'CG3' 3'AU5'	-21.52	3	5'CG3' 3'IU5'
5'AU3' 3'UA5'	-1.10	6	5'IU3' 3'UA5'	-0.50	5	5'AU3' 3'UA5'	-10.59	5	5'IU3' 3'UA5'
5'AU3' 3'AU5'	-0.93	7.5	5'AU3' 3'IU5'	-0.41	6	5'AU3' 3'AU5'	-7.68	6.5	5'AU3' 3'IU5'
5'UA3' 3'AU5'	-1.33	5	5'UA3' 3'IU5'	-0.37	7	5'UA3' 3'AU5'	-4.78	8	5'UA3' 3'IU5'
5'AU3' 3'AU5'	-0.93	7.5	5'IU3' 3'AU5'	-0.43	8	5'AU3' 3'AU5'	-7.68	6.5	5'IU3' 3'AU5'
MAD ^a					1.1				
									0.9

^a Mean absolute deviation between experimental NN pairs for A-U and I•U and between computational NN pairs for A-U and I•U