THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

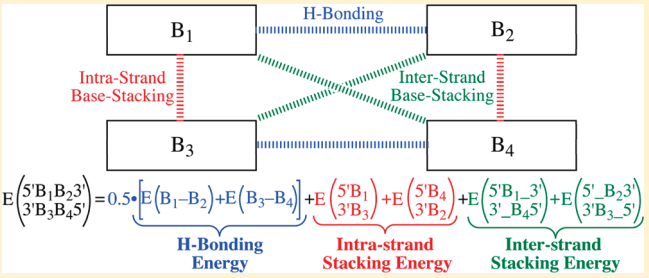
Computational Model for Predicting Experimental RNA and DNA Nearest-Neighbor Free Energy Rankings

Charles A. Johnson, Richard J. Bloomingdale, Vikram E. Ponnusamy, Conor A. Tillinghast, Brent M. Znosko,* and Michael Lewis*

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

Supporting Information

ABSTRACT: Hydrogen-bonding, intrastrand base-stacking, and interstrand base-stacking energies were calculated for RNA and DNA dimers at the MP2(full)/6-311G** level of theory. Standard A-form RNA and B-form DNA geometries from average fiber diffraction data were employed for all base monomer and dimer geometries, and all dimer binding energies were obtained via single-point calculations. The effects of water solvation were considered using the PCM model. The resulting dimer binding energies were used to calculate the 10 unique RNA and 10 unique DNA computational nearest-neighbor energies, and the ranking of these computational nearest neighbor energies are in excellent agreement with the ranking of the experimental nearest-neighbor free energies. These results dispel the notion that average fiber diffraction geometries are insufficient for calculating RNA and DNA stacking energies.



INTRODUCTION

The three-dimensional structure, conformational flexibility, and overall stability of RNA and DNA are dictated primarily by hydrogen bonding[1] and base-stacking interactions;[2] however, base-phosphate group interactions[3] and base-ribose sugar interactions (in RNA)[3a] also play a role. Although the nature of hydrogen bonding has been widely studied and well documented,[1] the most important factor in RNA/DNA stabilization is base-stacking interactions, yet significant work remains before they are fully understood.[4] The literature contains lively debate on the appropriate input geometries to use when computationally predicting relative base-stacking energies for either RNA or DNA. Perhaps the biggest current debate centers on the appropriateness of using RNA or DNA geometries derived from average fiber diffraction data to investigate RNA or DNA base-stacking interactions. The use of average fiber diffraction data to investigate nucleic acid base-stacking has a long history,[5] and a recent study used B-DNA geometries obtained from average fiber diffraction data to probe the contribution of electrostatics, induction, exchange, and dispersion to the overall base-stacking binding energies via symmetry-adapted perturbation theory.[2b] This work has received significant criticism on the basis that geometry averaging can result in repulsive interactions that are not found in nature, and it has been suggested that other methods for geometry selection are superior, such as employing MD simulations.[2a,4] The given reason for the supposed inferiority of RNA or DNA base-stacking geometries obtained from average fiber diffraction data is that they may contain non-natural, repulsive intermolecular contacts and that they give different relative base-stacking energies than other geometry selection methods.[4] Of course, a much more satisfactory way to evaluate computational approaches is via comparison to experiment. The widely used RNA/DNA nearest-neighbor (NN) free energies[6] offer the experimental data to evaluate approaches to calculating relative base-stacking energies. Quite surprisingly, however, the authors are unaware of any studies that have justified the use of certain RNA or DNA input geometries by benchmarking the resulting base-stacking energies to the relative NN free energies. In fact, it has been suggested that such a comparison is not even possible and that there is no correlation between calculated base-stacking energies and the experimental NN free energies.[2a] This is a sentiment we disagree with for reasons outlined below. Here, we report calculated A-form RNA and B-form DNA base-stacking and hydrogen-bonding energies that employed input geometries obtained from average fiber diffraction data. The resulting base-stacking and hydrogen-bonding energies were used to generate NN energy rankings that are in excellent agreement with the experimental free energy rankings. Furthermore, the agreement with experiment is better than it is for computational approaches that employ MD simulations to obtain base-stacking input geometries.

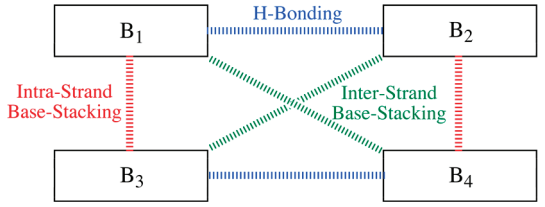
COMPUTATIONAL APPROACH

Although there are only 10 unique RNA and 10 unique DNA NN combinations, there are 16 possible intrastrand and 20

Received: February 8, 2011
Revised: May 26, 2011
Published: May 30, 2011

dx.doi.org/10.1021/jp2012733 | J. Phys. Chem. B 2011, 115, 9244–9251

Scheme 1. H-Bonding, Intrastrand Stacking, and Interstrand Stacking Dimers for a General Four Nucleotide System Containing Monomers B_1-B_4[a]



[a] B_1 and B_3 are in one strand and B_2 and B_4 are in the other strand.

possible interstrand base-stacking dimers for each biopolymer, along with the two possible H-bonding dimers. Scheme 1 graphically illustrates these three types of dimers, and the binding energies are shown in Tables 1 and 2. The geometries of the RNA and DNA base monomers and the 38 base dimers in Tables 1 and 2 were obtained from the *InsightII* (Accelrys, San Diego, CA) RNA/DNA visualization program, which employs average fiber diffraction data to generate the monomer and dimer structures. In each case, the sugar−phosphate backbone was omitted, and the $N-C_{sugar}$ bond was substituted with either an N−H or $N-CH_3$ bond, yielding what is termed here RNA-H/DNA-H and RNA-Me/DNA-Me monomers and dimers, respectively. The position(s) of the N−H hydrogen atom and the $N-CH_3$ methyl group atoms were optimized for each monomer and dimer at the MP2(full)/6-311G** level of theory while the rest of the RNA/DNA base atoms were constrained to their *InsightII* position. The dimer total energies ($E_{Tot,Dim}$) were corrected for basis set superposition error (BSSE) via the counter-poise method,[7] and the binding energy (E_{bind}) for each dimer was determined by subtracting the relevant MP2(full)/6-311G** monomer total energies ($E_{Tot,Mono}$) from the BSSE-corrected MP2(full)/6-311G** $E_{Tot,Dim}$ values, as illustrated in eq 1 where B_m and B_n are any RNA/DNA base. Note that the general $B_m{\bullet}B_n$ dimers in eq 1 represent either H-bonding dimers, intrastrand base-stacking dimers, or interstrand base-stacking dimers. Binding energies were determined both for RNA-H/DNA-H dimers ($E_{bind,H}$) and RNA-Me/DNA-Me dimers ($E_{bind,Me}$), using the general formula shown in eq 1, and the values are given in Tables 1 and 2. The MP2(full)/6-311G** level of theory was employed because it has been shown to yield reliable stacking energies for arene−arene interactions.[8] Finally, all of the $E_{bind,H}$ and $E_{bind,Me}$ energies were recalculated taking into consideration water solvation via the polarized continuum model (PCM),[9] which has been shown to perform well in modeling solvation effects,[10] and these results are given in parentheses in Tables 1 and 2. The PCM calculations employed the default solute radii scheme and solvent dielectric constant at 310.15 K.[9] It should be noted that PCM was parametrized for solvation free energies and formally should only be employed to correct for the effects of solvation on gas-phase free energies.[11] Of course, obtaining the thermodynamic data required to calculate free energies requires geometry optimization; however, one hallmark of the approach presented here is employing geometries from average diffraction data.

$$E_{bind}(B_m{\bullet}B_n) = E_{Tot,Dim}(B_m{\bullet}B_n) - (E_{Tot,Mono}(B_m) + E_{Tot,Mono}(B_n)) \quad (1)$$

As it is ill advised to calculate free energies using geometries from average diffraction data, which are not gas-phase optimized, we have refrained from doing so. Still, PCM calculations were performed since

Table 1. MP2(full)/6-311G Calculated A-Form RNA Hydrogen Bonding, Intrastrand Base-Stacking, and Interstrand Base-Stacking Energies (kcal/mol)[a]**

hydrogen-bonding energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
A−U	−9.78	−11.46	C−G	−25.97	−25.86
	(−11.36)	(−12.29)		(−19.35)	(−20.61)
intrastrand base-stacking energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
5′A	−0.14	−3.69	5′G	−5.88	−6.35
3′A	(−5.23)	(−8.87)	3′A	(−8.79)	(−9.59)
5′A	−1.95	−2.56	5′G	−8.80	−9.24
3′C	(−7.81)	(−8.49)	3′C	(−9.19)	(−10.09)
5′A	−5.48	−4.71	5′G	−0.12	−0.56
3′G	(−9.93)	(−8.02)	3′G	(−7.71)	(−8.52)
5′A	−2.06	−3.81	5′G	−0.36	−2.03
3′U	(−6.70)	(−7.56)	3′U	(−7.32)	(−8.36)
5′C	−1.83	−2.38	5′U	0.38	−3.49
3′A	(−4.46)	(−5.11)	3′A	(−1.83)	(−4.65)
5′C	−0.46	−1.07	5′U	1.51	−3.23
3′C	(−5.66)	(−6.24)	3′C	(−2.06)	(−5.96)
5′C	−3.17	−3.70	5′U	0.04	−1.90
3′G	(−3.69)	(−4.39)	3′G	(−3.13)	(−4.69)
5′C	−2.85	−4.52	5′U	2.20	−0.65
3′U	(−4.61)	(−5.49)	3′U	(−3.08)	(−4.14)
interstrand base-stacking energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
5′A_3′	−0.24	−0.27	5′_A3′	+1.14	+1.08
3′_A5′	(−1.12)	(−1.02)	3′A_5′	(−3.50)	(−3.55)
5′A_3′	+0.42	+0.42	5′_A3′	0.00	−0.14
3′_C5′	(−0.38)	(−0.24)	3′C_5′	(−3.18)	(−3.22)
5′A_3′	−2.44	−2.39	5′_A3′	−2.95	−3.06
3′_G5′	(−1.19)	(−4.45)	3′G_5′	(−4.33)	(−4.45)
5′A_3′	+1.06	−0.54	5′_A3′	−1.01	−2.70
3′_U5′	(+0.55)	(−0.26)	3′U_5′	(−2.11)	(−3.03)
5′C_3′	+1.14	+0.94	5′_C3′	+7.80	+0.93
3′_C5′	(−0.11)	(−0.11)	3′C_5′	(+4.29)	(−0.10)
5′C_3′	−2.29	−3.00	5′_C3′	−3.37	−3.53
3′_G5′	(+0.04)	(−0.44)	3′G_5′	(−3.26)	(−3.45)
5′C_3′	+1.09	−0.57	5′_C3′	−0.49	−2.27
3′_U5′	(+0.72)	(−0.13)	3′U_5′	(−0.46)	(−1.37)
5′G_3′	+2.79	+15.28	5′_G3′	−0.48	−0.64
3′_G5′	(−1.06)	(−4.45)	3′G_5′	(−3.95)	(−4.22)
5′G_3′	+2.05	+0.52	5′_G3′	+2.68	+1.12
3′_U5′	(+0.68)	(−0.19)	3′U_5′	(−1.37)	(−2.35)
5′U_3′	+3.11	+0.75	5′_U3′	+3.56	+0.74
3′_U5′	(−2.43)	(+0.67)	3′U_5′	(+0.91)	(−0.18)

[a] Nonsolvated $E_{bind,H}$ and $E_{bind,Me}$ values on top and not in parentheses. Solvated $E_{bind,H}$ and $E_{bind,Me}$ values on bottom and in parentheses. Solvation accounted for via the PCM method.

dx.doi.org/10.1021/jp2012733 |J. Phys. Chem. B 2011, 115, 9244–9251

Table 2. MP2(full)/6-311G Calculated B-Form DNA Hydrogen Bonding, Intrastrand Base-Stacking, and Interstrand Base-Stacking Energies (kcal/mol)[a]**

hydrogen-bonding energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
A−T	−13.53	−13.92	C−G	−27.31	−27.40
	(−14.54)	(−15.05)		(−20.35)	(−20.40)

intrastrand base-stacking energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
5′A	−1.60	−4.81	5′G	−6.66	−5.33
3′A	(−7.31)	(−11.52)	3′A	(−10.75)	(−12.19)
5′A	−3.03	−0.41	5′G	−8.04	−8.34
3′C	(−8.46)	(−8.73)	3′C	(−9.01)	(−10.17)
5′A	−5.48	−6.63	5′G	−1.05	−2.39
3′G	(−9.85)	(−11.35)	3′G	(−10.27)	(−11.86)
5′A	−3.81	−3.38	5′G	−2.64	−3.87
3′T	(−9.00)	(−10.51)	3′T	(−9.17)	(−10.40)
5′C	−1.82	−3.35	5′T	−2.62	−4.66
3′A	(−5.80)	(−7.40)	3′A	(−5.55)	(−7.47)
5′C	0.30	−0.68	5′T	−2.43	−3.39
3′C	(−5.86)	(−6.98)	3′C	(−6.37)	(−7.39)
5′C	−4.24	−5.29	5′T	−3.16	−3.47
3′G	(−5.75)	(−6.94)	3′G	(−6.03)	(−9.38)
5′C	−3.61	−5.00	5′T	−6.26	−2.94
3′T	(−6.91)	(−8.45)	3′T	(−11.10)	(−12.41)

interstrand base-stacking energies (kcal/mol)					
bases	$E_{bind,H}$	$E_{bind,Me}$	bases	$E_{bind,H}$	$E_{bind,Me}$
5′A_3′	+3.72	−0.42	5′_A3′	−1.49	−1.51
3′_A5′	(−2.31)	(−2.37)	3′A_5′	(−4.28)	(−4.39)
5′A_3′	+3.99	+0.27	5′_A3′	+0.79	+0.72
3′_C5′	(+2.27)	(−1.07)	3′C_5′	(−1.83)	(−1.94)
5′A_3′	−2.99	−3.22	5′_A3′	−3.40	−3.39
3′_G5′	(−2.70)	(−2.81)	3′G_5′	(−5.15)	(−5.08)
5′A_3′	−2.29	−2.13	5′_A3′	−0.12	−2.87
3′_T5′	(−2.80)	(−2.45)	3′T_5′	(−1.17)	(−3.33)
5′C_3′	+1.31	+1.15	5′_C3′	+3.05	+2.8
3′_C5′	(−0.37)	(−0.43)	3′C_5′	(−0.16)	(−0.20)
5′C_3′	−4.07	−4.24	5′_C3′	−2.80	−3.06
3′_G5′	(−1.00)	(−1.13)	3′G_5′	(−2.45)	(−2.65)
5′C_3′	−2.19	−2.12	5′_C3′	−1.89	−1.94
3′_T5′	(−2.16)	(−1.93)	3′T_5′	(−2.26)	(−2.11)
5′G_3′	−3.01	+1.36	5′_G3′	−3.01	−3.65
3′_G5′	(−5.79)	(−3.18)	3′G_5′	(−5.79)	(−6.13)
5′G_3′	−3.01	−1.66	5′_G3′	−1.61	−1.8
3′_T5′	(−2.56)	(−2.27)	3′T_5′	(−3.99)	(−3.81)
5′T_3′	−1.34	−1.42	5′_T3′	−1.14	−1.30
3′_T5′	(−2.03)	(−2.01)	3′T_5′	(−2.12)	(−2.11)

[a] Nonsolvated $E_{bind,H}$ and $E_{bind,Me}$ values on top and not in parentheses. Solvated $E_{bind,H}$ and $E_{bind,Me}$ values on bottom and in parentheses. Solvation accounted for via the PCM method.

this was the best available option to account for solvation. All calculations were performed using the Gaussian 09 suite of programs.[12]

RESULTS AND DISCUSSION

Dimer Binding Energies and Solvation. The hydrogen-bonding, intrastrand base-stacking, and interstrand base-stacking energies in Tables 1 and 2 show that solvation, at least when accounted for via the PCM model, usually makes the dimer binding energies, $E_{bind,H}$ and $E_{bind,Me}$, more stable with respect to the nonsolvated values. Specifically, 33 of the 38 RNA dimer $E_{bind,H}$ values, 32 of the 38 RNA dimer $E_{bind,Me}$ values, 32 of the 38 DNA dimer $E_{bind,H}$ values, and 33 of the 38 DNA dimer $E_{bind,Me}$ values are more stable when solvation is considered (Tables 1 and 2). All of the RNA and DNA intrastrand base-stacking dimers where the bases are directly on top of one another, those with the following general formula in Scheme 1,

$$\begin{matrix} 5'B_1 \\ 3'B_3 \end{matrix} \quad \text{and} \quad \begin{matrix} 5'B_2 \\ 3'B_4 \end{matrix}$$

are more stable when solvated by water. The five or six instances for each case ($E_{bind,H}$ and $E_{bind,Me}$, both gas-phase/vacuum and solvated) where the nonsolvated gas-phase/vacuum E_{bind} values are more stable than the PCM water solvated E_{bind} values are either G−C H-bonding dimers (B_1-B_2 and B_3-B_4 in Scheme 1) or interstrand base-stacking dimers with the following general formula in Scheme 1.

$$\begin{matrix} 5'B_1\_3' \\ 3'\_B_4 5' \end{matrix} \quad \text{and} \quad \begin{matrix} 5'\_B_2 3' \\ 3'B_3\_5' \end{matrix}$$

Solvation leading to more stable dimer binding energies is certainly not always the case. For instance, recent work on amino acid pair interaction energies showed that the introduction of ether or water solvent decreased the binding energy by about half.[13] In fact, the calculated amino acid binding energies decreased in going from gas-phase/vacuum to ether to water, showing that the greater the solvent polarity the less the binding. In contrast, computational work on the $\pi-\pi$ stacking interactions between DNA bases and members of the indenoisoquinoline class of topoisomerase I inhibitors showed that water solvation stabilized the complexes.[14] Thus, gas-phase/vacuum base-stacking interactions where the monomers are directly on top of each other appear to be stabilized when solvated by water. In contrast, interactions where the monomers are not directly on top of one another, such as the H-bonding dimers and the interstrand base-stacking dimers, appear less likely to be stabilized by water since these are the only instances where a decrease in dimer stability is observed when going from gas-phase/vacuum to water. It is worth noting here that although we employ the term "stacking" for the interstrand stacking dimers, the monomers are not in the same strand, and although there is more overlap than suggested by the general block diagram view in Scheme 1, the overlap between the monomers is quite small compared to the intrastrand dimers.

The differences in how much water stabilizes, or destabilizes, the dimers compared to gas-phase/vacuum can be explained, to a large degree, by the differential effect of water solvation on the dimer dipole moments. For the purposes of the following discussion, the difference in dipole moment in going from gas-phase/vacuum to water, for the monomers or dimers, will be referred to as ΔD_{sol}. The average ΔD_{sol} values for the RNA-H and RNA-Me

monomers, $\Delta D_{sol,H}$ and $\Delta D_{sol,Me}$, are 27.5% and 29.2% respectively, whereas the average $\Delta D_{sol,H}$ and $\Delta D_{sol,Me}$ values for the RNA-H and RNA-Me intrastrand dimers are 129.4% and 135.9% (all calculated RNA and DNA monomer and dimer dipole moments are in the Supporting Information). The fact that the intrastrand base stacking dimers increase in polarity to a much greater extent than do the monomers when going from gas-phase/vacuum to water explains why the intrastrand base stacking dimers are greatly stabilized in water. By comparison, the average $\Delta D_{sol,H}$ and $\Delta D_{sol,Me}$ values for the RNA-H and RNA-Me interstrand dimers are 50.7% and 83.7%, respectively. This shows that the interstrand dimers increase in polarity to a slightly greater degree than do the monomers when going from gas-phase/vacuum to water, and it is thus not surprising that some of these dimers are destabilized by water solvation while most of them are stabilized. The change in dipole moment upon water solvation for the RNA-H and RNA-Me H-bonding dimers is exactly as expected based on the E_{bind} values from Table 1. The RNA-H and RNA-Me A−U H-bonding dimer $E_{bind,H}$ and $E_{bind,Me}$ values are more stabilized in water compared to gas-phase/vacuum, and the dipole moment of these dimers increases by 42.7% and 47.7%, respectively. Conversely, the RNA-H and RNA-Me G−C H-bonding dimer $E_{bind,H}$ and $E_{bind,Me}$ values are destabilized in water, and the dipole moments decrease by 8.9% and 11.0%, respectively. While the average ΔD_{sol} values work very well in explaining the overall trend in how water solvation affects the E_{bind} values in Table 1, one should be cautioned against viewing this as a comprehensive approach for explaining how solvation affects trends in binding energies. For instance, three of the RNA-H and RNA-Me intrastrand dimers either decrease in polarity upon water solvation or increase in polarity to a smaller extent than the constituent monomers, yet water solvation stabilizes the $E_{bind,H}$ values. The same trend occurs for four of the RNA-H interstrand dimers and five of the RNA-Me interstrand dimers. Still, almost all of the RNA-H and RNA-Me dimers that are destabilized upon water solvation become less polar in going from gas-phase/vacuum to water. Furthermore, the E_{bind} values for 30 of the 38 RNA-H dimers and 28 of the 38 RNA-Me dimers can be rationalized via the ΔD_{sol} values.

Determining Nearest-Neighbor (NN) Energy Rankings. The best way to benchmark the H-bonding and base-stacking energies in Tables 1 and 2 is through comparison with experiment, and the most appropriate experimental data are the well-documented NN thermodynamic parameters.[6] As noted in the introduction, it has been suggested that this is not a viable approach for evaluating experimental base-stacking calculations;[2a,4] however, we disagree with this sentiment for the reasons outlined here. The approach we propose to evaluate the H-bonding and base-stacking energies in Tables 1 and 2 is to computationally determine NN energies (E_{NN}) and compare the rankings to the $\Delta G_{NN,exp}$ rankings, as shown in Table 3. The E_{NN} values are calculated as the sum of the intrastrand base-stacking energies, interstrand base-stacking energies, and one-half the H-bonding energies. The H-bonding energies were halved due to the manner in which the experimental NN free energies were determined.[6] Scheme 2 shows an example of how E_{NN} is calculated for the following NN four nucleotide system.

$$5'C - G3'$$
$$3'U - A5'$$

Although this calculation does not include a contribution from entropy, the rankings of the E_{NN} values are compared to the $\Delta G_{NN,exp}$ rankings. The $\Delta H_{NN,exp}$ rankings were not used due to the large errors associated with enthalpy values derived from optical melting experiments.[6] In addition to ignoring entropy, the method described here has another limitation. The $\Delta G_{NN,exp}$ values were determined via optical melting experiments,[6] which from a molecular standpoint involve heating a double-stranded base-paired segment of RNA or DNA in buffered aqueous solution until the segment unwinds and becomes single-stranded.[6] It is important to recognize that the experimental values are ΔG values or the difference between the double-stranded state and the single-stranded state. The computational values that we determine, however, calculate the stabilizing energetics involved within the double-stranded state and do not account for the single stranded state. Therefore, there is a significant difference in the magnitude of the computational and experimental numbers. In comparing the relative rankings of the computational and experimental numbers, we are assuming that forces present in the single-stranded state (i.e., decreased base stacking) are relatively independent of sequence.

Using the approach outlined in Scheme 2, computational NN energies (E_{NN}) were calculated for the 10 unique RNA and 10 unique DNA NN four nucleotide systems using the $E_{bind,H}$ and $E_{bind,Me}$ values in Tables 1 and 2, and the results are presented in Table 3. The $E_{bind,H}$ and $E_{bind,Me}$ values in Tables 1 and 2 allow for four different approaches to the calculation of E_{NN} values. The $E_{bind,H}$ and $E_{bind,Me}$ values without consideration of solvation can be employed, and the nearest-neighbor energies that result from these values are termed $E_{NN,Calc\text{-}H}$ and $E_{NN,Calc\text{-}Me}$, respectively, in Table 3. In addition, the $E_{bind,H}$ and $E_{bind,Me}$ values that account for solvation via the PCM method can be used, and the resulting nearest-neighbor energies in Tables 3 are termed $E_{NN,Calc\text{-}H,Sol}$ and $E_{NN,Calc\text{-}Me,Sol}$. Of note in Table 3, the RNA and DNA $E_{NN,Calc\text{-}H}$, $E_{NN,Calc\text{-}H,Sol}$, $E_{NN,Calc\text{-}Me}$, and $E_{NN,Calc\text{-}Me,Sol}$ values are all much larger in magnitude than the corresponding experimental values ($\Delta G_{NN,exp}$). The discrepancy between the magnitude of the calculated and experimental values in Table 3 is likely due to the reasons outlined above: the E_{NN} do not account for entropy, nor do they account for the enthalpic/energetic stability of the single-stranded RNA or DNA segments. Other minor issues may be that the $\Delta G_{NN,exp}$ values also contain information on the reorganization of the sugar−phosphate backbone and on the differential solvation of the double- and single-stranded RNA or DNA segments, whereas the E_{NN} values would clearly not account for these factors. Of course, the suggestion here is that these factors would affect the magnitude of the E_{NN} values in Table 3 and not the relative rank. Thus, despite its limitations, the comparison of the relative E_{NN} values with the relative $\Delta G_{NN,exp}$ values provides a reasonable approach to evaluating the H-bonding and base-stacking energies in Tables 1 and 2.

The numbers in parentheses in Table 3 are the relative rankings of the NN four nucleotide systems using the four different approaches. The least stable NN four nucleotide system is given a ranking of 1 and the most stable is given a ranking of 10. The $E_{NN,Calc\text{-}H}$, $E_{NN,Calc\text{-}H,Sol}$, $E_{NN,Calc\text{-}Me}$, and $E_{NN,Calc\text{-}Me,Sol}$ columns end with mean absolute deviation (MAD) and Spearman rank correlation coefficient (r_s)[15] values for the computational approach. The MAD values were determined by taking the mean of the absolute difference between the NN four nucleotide system ranking of the calculated value and the ranking of the corresponding experimental value, $\Delta G_{NN,exp}$. For RNA, the $E_{NN,Calc\text{-}H}$, $E_{NN,Calc\text{-}Me}$, and $E_{NN,Calc\text{-}Me,Sol}$ numbers each have a

MAD of 1.0, whereas the $E_{NN,Calc\text{-}H,Sol}$ values have a MAD of 1.6. For DNA, the $E_{NN,Calc\text{-}H}$, $E_{NN,Calc\text{-}H,Sol}$, $E_{NN,Calc\text{-}Me}$, and $E_{NN,Calc\text{-}Me,Sol}$ values give MADs of 0.4, 1.4, 1.2, and 2.0, respectively. Most of these MAD values are quite good, and the agreement between the DNA $E_{NN,Calc\text{-}H}$ and $\Delta G_{NN,exp}$ values is excellent (MAD of 0.4). The calculated gas-phase/vacuum E_{NN} values are either better than, or equal to, the PCM calculated water solvated E_{NN} values; the MAD values for the RNA $E_{NN,Calc\text{-}Me}$ and $E_{NN,Calc\text{-}Me,Sol}$ are both 1.0, and in all other cases, the MADs are better for the gas-phase/vacuum E_{NN} values than for the water solvated E_{NN} values (Table 3). This may seem somewhat surprising since the $\Delta G_{NN,exp}$ values were obtained in water; however, recall that there are numerous differences in how the E_{NN} and $\Delta G_{NN,exp}$ values are determined. Thus, there is no reason to expect calculated water solvated E_{NN} values would be any better or worse than the calculated gas-phase/vacuum E_{NN} values.

A statistical approach to test for how well two rankings correlate with each other is the r_s value. The Spearman rank correlation test was applied to the association between the experimental and calculated rankings with a null hypothesis of no association between the rankings. For rankings with 10 data points, an r_s value greater than 0.794 shows the null hypothesis can be rejected at the 99.5% confidence level,[15] and thus there is a correlation between the two rankings. Thus, for all four calculated RNA E_{NN} values, we can say that the calculated rankings are very strongly correlated with the experimental $\Delta G_{NN,exp}$ ranking. For the calculated DNA E_{NN} values, only the $E_{NN,Calc\text{-}H}$ and $E_{NN,Calc\text{-}Me}$ rankings have r_s values above 0.794 and thus have strong correlations with the experimental $\Delta G_{NN,exp}$ ranking. For the $E_{NN,Calc\text{-}H,Sol}$ ranking, the r_s value is just below the cutoff for association with the $\Delta G_{NN,exp}$ ranking at the 99.5% confidence level, and instead meets the 0.745 standard of the 99% confidence level.[15] The DNA $E_{NN,Calc\text{-}Me,Sol}$ ranking has an r_s value of 0.37, and the null hypothesis holds at all confidence levels; there is no correlation with the experimental ranking. Given the strong correlation between all of the RNA E_{NN} rankings and the experimental $\Delta G_{NN,exp}$ ranking, and between two of the DNA E_{NN} rankings and the $\Delta G_{NN,exp}$ ranking, it is fair to say the approach presented in Scheme 2 has high predictive value in determining relative RNA and DNA $\Delta G_{NN,exp}$ rankings.

Table 3. Relative Rank Comparison between RNA and DNA Experimental Nearest-Neighbor (NN) Free Energies and Computational E_{NN} Values[a]

	RNA nearest-neighbor energies (kcal/mol)					
NN	$\Delta G_{NN,exp}$[b] (rank)	$E_{NN,Calc\text{-}H}$[c] (rank)	$E_{NN,Calc\text{-}H,Sol}$[d] (rank)	$E_{NN,Calc\text{-}Me}$[e] (rank)	$E_{NN,Calc\text{-}Me,Sol}$[f] (rank)	ref 2a Rank
5′AU3′	−0.93	−7.68	−21.24	−19.04	−28.60	2
3′AU5′	(1)	(2)	(2)	(3)	(2.5)[g]	
5′AU3′	−1.10	−10.59	−24.98	−18.61	−28.60	3
3′UA5′	(2)	(3)	(3)	(2)	(2.5)[g]	
5′UA3′	−1.33	−4.78	−20.96	−11.46	−24.47	8
3′AU5′	(3)	(1)	(1)	(1)	(1)	
5′CG3′	−2.08	−23.10	−31.64	−26.36	−32.56	4
3′UA5′	(4)	(6)	(7)	(4)	(5)	
5′CG3′	−2.11	−21.52	−26.56	−26.57	−30.84	5
3′AU5′	(5)	(5)	(4)	(5)	(4)	
5′AU3′	−2.24	−23.12	−32.13	−27.91	−39.12	6
3′CG5′	(6)	(7)	(8)	(7)	(8)	
5′GC3′	−2.35	−20.19	−28.70	−27.85	−35.40	7
3′AU5′	(7)	(4)	(5)	(6)	(7)	
5′CG3′	−2.36	−31.65	−30.80	−32.96	−33.72	9
3′GC5′	(8)	(8)	(6)	(9)	(6)	
5′GC3′	−3.26	−32.21	−35.94	−34.02	−39.26	1
3′GC5′	(9)	(9)	(10)	(10)	(9)	
5′GC3′	−3.42	−32.97	−34.50	−28.12	−45.34	10
3′CG5′	(10)	(10)	(9)	(8)	(10)	
MAD[h]		**1.0**	**1.6**	**1.0**	**1.0**	**1.6**
r_s[i]		**0.88**	**0.82**	**0.90**	**0.90**	**0.44**

Table 3. Continued

DNA nearest-neighbor energies (kcal/mol)						
NN	$\Delta G_{NN,exp}$[b] (rank)	$E_{NN,Calc\text{-}H}$[c] (rank)	$E_{NN,Calc\text{-}H,Sol}$[d] (rank)	$E_{NN,Calc\text{-}Me}$[e] (rank)	$E_{NN,Calc\text{-}Me,Sol}$[f] (rank)	ref 2a rank
5′TA3′	−0.60	−21.60	−31.96	−26.16	−36.39	7
3′AT5′	(1)	(2)	(1)	(2)	(1)	
5′AT3′	−0.73	−18.58	−36.98	−22.40	−40.55	6
3′TA5′	(2)	(1)	(5)	(1)	(2)	
5′AT3′	−1.02	−23.80	−36.92	−26.67	−44.76	5
3′AT5′	(3)	(3)	(4)	(3)	(10)	
5′AT3′	−1.16	−27.13	−35.93	−33.82	−42.39	2
3′GC5′	(4)	(4)	(2)	(7)	(7)	
5′CG3′	−1.38	−30.99	−36.59	−32.99	−41.53	3
3′AT5′	(5)	(6)	(3)	(6)	(5)	
5′AT3′	−1.43	−30.97	−40.03	−30.10	−41.77	8
3′CG5′	(6)	(5)	(9)	(4)	(6)	
5′GC3′	−1.46	−31.72	−38.95	−30.32	−41.51	4
3′AT5′	(7)	(7)	(7)	(5)	(4)	
5′CG3′	−1.77	−34.92	−39.93	−37.77	−43.03	1
3′CG5′	(8)	(8)	(8)	(8)	(8)	
5′CG3′	−2.09	−37.49	−38.01	−40.48	−40.84	9
3′GC5′	(9)	(9)	(6)	(10)	(3)	
5′GC3′	−2.28	−43.35	−44.30	−39.92	−44.11	10
3′CG5′	(10)	(10)	(10)	(9)	(9)	
MAD[h]		**0.4**	**1.4**	**1.2**	**2.0**	**2.8**
r_s[i]		**0.98**	**0.78**	**0.87**	**0.37**	**0.24**

[a] The rank for each approach is in parentheses under the E_{NN} values, with the least energetically favorable nearest-neighbor combination ranked 1 and the most favorable ranked 10. [b] $\Delta G_{NN,exp}$ are the experimental nearest-neighbor energies from reference 6. [c] $E_{NN,Calc\text{-}H}$ is the calculated nearest-neighbor energies where the sugar was replaced by an H atom and solvation is not considered. [d] $E_{NN,Calc\text{-}H,Sol}$ is the calculated nearest-neighbor energies where the sugar was replaced by an H atom and solvation is considered via the PCM method. [e] $E_{NN,Calc\text{-}Me}$ is the calculated nearest-neighbor energies where the sugar was replaced by a methyl group and solvation is not considered. [f] $E_{NN,Calc\text{-}Me,Sol}$ is the calculated nearest-neighbor energies where the sugar was replaced by a methyl group and solvation is considered via the PCM method. [g] The RNA $E_{NN,Calc\text{-}Me,Sol}$ values are the exact same for the nearest-neighbor four nucleotide systems that would be ranked second and third, and thus they were each given a rank of 2.5. [h] Mean absolute difference (MAD) is determined by taking the mean of the absolute difference between the NN four nucleotide system ranking of the calculated value and the ranking of the corresponding experimental value. [i] The Spearman rank correlation coefficient (r_s) was determined as shown in ref 15.

There are two important results in Table 3. First, the MAD and r_s values in Table 3 validate the assumptions that went into using relative E_{NN} values, calculated using the equation in Scheme 2, as a means for predicting relative $\Delta G_{NN,exp}$ values. Second, and perhaps most important, the very good agreement between the E_{NN} and $\Delta G_{NN,exp}$ rankings validates the use of average diffraction data to generate RNA and DNA geometries for calculating H-bonding and base-stacking energies. Of note, this is not the first computational study to rank the 10 unique RNA and 10 unique DNA nearest-neighbor four nucleotide systems. Sponer and co-workers also calculated and ranked the NN four nucleotide system energies using RNA and DNA geometries obtained from MD simulations,

Scheme 2. Approach Used to Calculate Computational Nearest-Neighbor Energies (E_{NN})

$$E\begin{pmatrix}5'CG3'\\3'UA5'\end{pmatrix} = 0.5\bullet\underbrace{\left[E(C\text{–}G)+E(U\text{–}A)\right]}_{\text{H-Bonding Energy}} + \underbrace{E\begin{pmatrix}5'C\\3'U\end{pmatrix} + E\begin{pmatrix}5'A\\3'G\end{pmatrix}}_{\text{Intra-strand Stacking Energy}} + \underbrace{E\begin{pmatrix}5'C\_3'\\3'\_A5'\end{pmatrix} + E\begin{pmatrix}5'\_G3'\\3'U\_5'\end{pmatrix}}_{\text{Inter-strand Stacking Energy}}$$

and the results are shown in the last column of Table 3.[2a] Sponer and co-workers did not determine MAD and r_s values for their calculated rankings, and thus we did so and included them in Table 3. The RNA MAD value of 1.6 for the Sponer and co-workers ranking is

dx.doi.org/10.1021/jp2012733 |J. Phys. Chem. B 2011, 115, 9244–9251

not that bad, it is the same as for the poorest RNA E_{NN} value reported in this study, the $E_{NN,Calc\text{-}H,Sol}$ value; however, the r_s value of 0.44 is significantly worse than any of the RNA r_s values reported for the E_{NN} rankings calculated via the approach outlined in Scheme 2, and it suggests there is no correlation between the Sponer and co-workers ranking and the experimental $\Delta G_{NN,exp}$ ranking. The DNA MAD value of 2.8 and the r_s value of 0.24 for the Sponer and co-workers ranking are both very weak, and they too suggest no correlation with the experimental $\Delta G_{NN,exp}$ ranking. Sponer and co-workers "conclude that there is no quantitative correlation between the QM gas phase stacking data, irrespective of the accuracy of the calculations, and the nucleic acids stability".[2a] Although this is certainly true using the approach Sponer and co-workers employed, the results of the present study show this statement is not general; there most definitely is a quantitative correlation between the quantum mechanical (QM) gas-phase stacking data and nucleic acid stability, in the form of the comparison between the relative calculated E_{NN} values and the relative experimental $\Delta G_{NN,exp}$ values. It is worth mentioning that the format used here to name NN four nucleotide systems is different than the format employed in the papers reporting the experimental values ($\Delta G_{NN,exp}$),[6] but it is the same as the format used by Sponer and co-workers.[2a] Therefore, the four nucleotide system denoted here as the following

$$5'A - U3'$$
$$3'A - U5'$$

is termed the following

$$5'A - A3'$$
$$3'U - U5'$$

in the experimental work.[6] It was determined that being consistent with the work of Sponer and co-workers[2a] was more important due to the comparisons made with the study.

Table 3 reveals that the key to attaining a good correlation between calculated relative E_{NN} values and relative $\Delta G_{NN,exp}$ values is to at least predict that the four nucleotide systems with four H-bonds are least stable, the four nucleotide systems with five H-bonds are of intermediate stability, and the four nucleotide systems with six H-bonds are most stable. That is, four nucleotide systems with only A and U(T) bases need to be ranked 1−3, four nucleotide systems with one of each base need to be ranked 4−7, and four nucleotide systems with only C and G bases need to be ranked 8−10. For the work presented here, of the five instances in Table 3 where the MAD values are 1.2 and below, four of them meet this standard. Only for the RNA $E_{NN,Calc\text{-}Me,Sol}$ values, where the MAD value is 1.0, does this not hold; the following four nucleotide system

$$5'A - U3'$$
$$3'C - G5'$$

is ranked 8 and the following four nucleotide system

$$5'C - G3'$$
$$3'G - C5'$$

is ranked 6. Experimentally, these two four nucleotide systems are ranked 6 and 8, respectively, and since the calculated and experimental rankings are still quite close, it does not negatively affect the MAD value. Not surprisingly, the r_s values for the five computational E_{NN} rankings with MAD values ≤ 1.2 show the strongest correlation with the respective $\Delta G_{NN,exp}$ rankings. The two cases from the work presented here with four nucleotide system ranking MAD values of 1.6 and 2.0 (Table 3), along with the Sponer and co-workers calculated RNA four nucleotide system ranking (MAD value of 1.6, Table 3),[2a] all have two instances where the four nucleotide system rankings do not meet the standard based on number of H-bonds. In the case of the RNA $E_{NN,Calc\text{-}H,Sol}$ ranking, with MAD value of 1.6, the r_s value of 0.82 still shows a very strong correlation with the $\Delta G_{NN,exp}$ rankings. However, the DNA $E_{NN,Calc\text{-}Me,Sol}$ ranking, with MAD of 2.0, and the Sponer and co-workers RNA ranking, with MAD of 1.6, have very weak r_s values of 0.37 and 0.44 respectively, suggesting there is no correlation with the corresponding $\Delta G_{NN,exp}$ rankings. Surprisingly, the DNA $E_{NN,Calc\text{-}H,Sol}$ values have a MAD value of 1.4, yet six of the four nucleotide systems do not meet the standard based on number of H-bonds. Not surprisingly, the slightly weaker r_s value of 0.78 captures the deterioration in correlation with the corresponding $\Delta G_{NN,exp}$ ranking. A quick look at the $E_{NN,Calc\text{-}H,Sol}$ column in Table 3 shows that it is quite fortuitous that the MAD and r_s values remain decent; the four nucleotide systems that are in the wrong group based on number of H-bonds are still quite close to the experimental $\Delta G_{NN,exp}$ four nucleotide system ranking. As would be expected, the Sponer and co-workers DNA four nucleotide system ranking,[2a] with a MAD value of 2.8 and an r_s value of 0.24, has seven of the ten calculated four nucleotide systems ranked in the wrong group according to number of H-bonds (Table 3).

Briefly, it is important to note the differences between the approach presented here and the approach employed by Sponer and co-workers.[2a] Of course, Sponer and co-workers employed MD simulations for the selection of RNA/DNA base geometries. Although the geometries do not differ dramatically from the ones employed in this study, relatively minor changes in geometry can significantly impact base-stacking binding energies.[16] In addition, Sponer and co-workers determined four nucleotide system binding energies as the sum of the intrastrand and interstrand binding energies, along with a many body correction term, which is slightly different than the approach outlined in Scheme 2.

CONCLUSIONS

The work presented here outlines an approach for determining relative RNA and DNA nearest neighbor stability via the calculation of nucleic acid base H-bonding dimer energies, intrastrand base-stacking binding energies, and interstrand base-stacking binding energies. Combining these terms via the equation shown in Scheme 2 for the 10 unique RNA and 10 unique DNA four nucleotide systems (Table 3) and ranking the resulting E_{NN} values gives very good agreement with the $\Delta G_{NN,exp}$ rankings. For both RNA and DNA, the best agreement with experiment was obtained when the gas-phase/vacuum H-bonding and base-stacking energies were employed to determine the E_{NN} values. Taking into account solvation via the PCM method always led to more attractive bonding and base-stacking dimers, and more attractive E_{NN} values, than the corresponding gas-phase vacuum values. This can be explained, to a large degree, by taking into account the change in monomer and dimer dipole moments upon PCM water solvation. One of the most important aspects of this study is that the RNA and DNA base monomer

and dimer geometries were taken from average fiber diffraction data via the *InsightII* RNA/DNA visualization program. Using average fiber diffraction data to calculate RNA and DNA base-stacking energies has been heavily criticized, on quite reasonable grounds, and other approaches to choosing RNA and DNA base monomer and dimer geometries, such as MD simulations, have been touted as being far superior.[2a] However, such comparisons ring hollow in the absence of comparison with experiment. The authors are not aware of any studies that use the well documented $\Delta G_{NN,exp}$ four nucleotide system rankings,[6] or any other experimental values, to justify the calculation of base-stacking energies using a particular approach to obtaining RNA or DNA base monomer and dimer geometries. The findings here show that using RNA and DNA base monomer and dimer geometries obtained from average fiber diffraction data to calculate base-stacking and H-bonding energies and using these energies to determine relative four nucleotide system E_{NN} rankings gives very good agreement with relative $\Delta G_{NN,exp}$ four nucleotide system rankings. In fact, the agreement with experiment is better than for studies that used RNA and DNA base monomer and dimer geometries obtained from MD simulations, although the MD simulation studies used a different approach to determine relative four nucleotide system stability rankings.[2a] This dispels the notion that RNA and DNA base monomer and dimer geometries obtained from average fiber diffraction data are inferior to other methods for geometry selection and suggests, at the very least, that the chosen method for calculating relative four nucleotide system stability, relative E_{NN} values (Scheme 2) for the work reported here, is as important as the method for selecting the nucleic acid base monomer and dimer geometries. Finally, it should be noted that the calculated E_{NN} values ignore cooperativity effects between bases. Given the excellent agreement between E_{NN} and $\Delta G_{NN,exp}$ rankings, this was obviously not a major problem; however, it is possible that properly addressing this issue could lead to improved correlations. Future work will include addressing the assumptions that went into using relative E_{NN} values to predict relative $\Delta G_{NN,exp}$ values to determine if a better correlation can be obtained and employing this approach to help understand experimental thermodynamic data, such as why replacing an adenine with inosine gives less stable $\Delta G_{NN,exp}$ values.[17]

■ ASSOCIATED CONTENT

Supporting Information. Computational data and complete citations for refs 5a and 12. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: (M.L.) LewisM5@slu.edu; (B.M.Z.) Znoskob@slu.edu.

■ ACKNOWLEDGMENT

This work was supported by Research Corporation (M.L.: CC7804; B.M.Z.: CC7621), the National Science Foundation through TeraGrid resources provided by the National Center for Supercomputing Applications under Grant Nos. TG-CHE050039N (M.L.) and TG-CHE070046N (B.M.Z.), and the National Institutes of Health (B.M.Z.: 1R15GM085699-01A1). We thank the Reviewer who suggested the addition of the Spearman rank correlation coefficient analysis; the manuscript is much improved for the addition. V.E.P. and C.A.T. performed research as part of the Students and Teachers As Research Scientists (STARS) program administered by the University of Missouri — Saint Louis.

■ REFERENCES

(1) (a) Szatylowicz, H.; Sadlej-Sosnowska, N. *J. Chem. Inf. Model* **2010**, *50*, 2151–2161. (b) Gil, A.; Branchadell, V.; Bertran, J.; Oliva, A. *J. Phys. Chem. B* **2009**, *113*, 4907–4914. (c) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.

(2) (a) Svozil, D.; Hobza, P.; Sponer, J. *J. Phys. Chem. B* **2010**, *114*, 1191–1203. (b) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schultz, M. *J. Am. Chem. Soc.* **2008**, *130*, 1802–1803. (c) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *J. Chem. Theory Comput.* **2008**, *4*, 1996–2000. (d) Cysewski, P.; Czyznikowska, Z.; Zalesny, R.; Czelen, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2665–2672. (e) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2008**, *130*, 1304–1308. (f) Sedlak, R.; Jurecka, P.; Hobza, P. *J. Chem. Phys.* **2007**, *127*, 075104. (g) Hesselmann, A.; Jansen, G.; Schultz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730–11731. (h) Oostenbrink, C.; van Gunsteren, W. F. *Chem.—Eur. J.* **2005**, *11*, 4340–4348.

(3) (a) Mladek, A.; Sponer, J. E.; Jurecka, P.; Banas, P.; Otyepka, M.; Svozil, D.; Sponer, J. *J. Chem. Theory Comput.* **2010**, *6*, 3817–3835. (b) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 4898–4918.

(4) Sponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610.

(5) (a) Olson, W. K.; Bansal, M.; Burley, S. K.; Dikerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakked, Z.; et al. *J. Mol. Biol.* **2001**, *313* (229), 237. (b) Alhambra, C.; Luque, F. J.; Gago, F.; Orozco, M. *J. Phys. Chem. B* **1997**, *101*, 3846–3853.

(6) (a) Xia, T.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *34*, 14719–14735. (b) SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* **1996**, *35*, 3555–3562.

(7) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(8) Watt, M.; Hardebeck, L. K. E.; Kirkpatrick, C. C.; Lewis, M. *J. Am. Chem. Soc.* **2011**, *133*, 3854–3862.

(9) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151–5158.

(10) (a) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094. (b) Tomasi, J.; Cammi, R.; Mennucci, B.; Cappelli, C.; Corni, S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697–5712. (c) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.

(11) Ho, J.; Klamt, A.; Coote, M. L. *J. Phys. Chem. A* **2010**, *114*, 13442–13444.

(12) Frisch, M. J. et al. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.

(13) Riley, K. E.; Merz, K. M. *J. Phys. Chem. B* **2006**, *110*, 15650–15653.

(14) Song, Y.; Cushman, M. *J. Phys. Chem. B* **2008**, *112*, 9484–9489.

(15) Mendenahll, W. *Introduction to Probability and Statistics*, 5th ed.; Duxbury Press: Belmont, CA, 1979; pp 506−512.

(16) Sponer, J.; Jurecka, P.; Marchan, I.; Luque, F. J.; Orozco, M.; Hobza, P. *Chem.—Eur. J.* **2006**, *12*, 2854–2865.

(17) Wright, D. J.; Rice, J. L.; Yanker, D. M.; Znosko, B. M. *Biochemistry* **2007**, *46*, 4625–4634.