JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Imino Hydrogen Positions in Nucleic Acids from Density Functional Theory Validated by NMR Residual Dipolar Couplings

Alexander Grishaev,[†] Jinfa Ying,[†] and Ad Bax*

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

Supporting Information

ABSTRACT: Hydrogen atom positions of nucleotide bases in RNA structures solved by X-ray crystallography are commonly derived from heavy-atom coordinates by assuming idealized geometries. In particular, N1−H1 vectors in G and N3−H3 vectors in U are commonly positioned to coincide with the bisectors of their respective heavy-atom angles. We demonstrate that quantum-mechanical optimization of the hydrogen positions relative to their heavy-atom frames considerably improves the fit of experimental residual dipolar couplings to structural coordinates. The calculations indicate that deviations of the imino N−H vectors in RNA U and G bases result from H-bonding within the base pair and are dominated by the attractive interaction between the H atom and the electron density surrounding the H-bond-acceptor atom. DFT optimization of H atom positions is impractical in structural biology studies. We therefore have developed an empirical relation that predicts imino N−H vector orientations from the heavy-atom coordinates of the base pair. This relation agrees very closely with the DFT results, permitting its routine application in structural studies.

Imino groups are of key importance in solution NMR studies of nucleic acid structure. Their ^{15}N−^{1}H correlations are typically far better dispersed than base or ribose ^{13}C−^{1}H resonances, and they play a pivotal role in the study of larger structures, akin to backbone amide ^{15}N−^{1}H pairs in protein NMR studies. Unambiguous identification of the imino H-bond-acceptor atom often can be made through H-bond $^{2h}J_{NN}$ couplings,[1,2] and assignments frequently can be obtained by analysis of nuclear Overhauser effect patterns,[3] in particular when supplemented by residual dipolar coupling (RDC) information.[2,4] Moreover, imino group ^{15}N−^{1}H RDCs ($^{1}D_{NH}$) provide important restraints on the relative orientations of helical fragments in larger RNAs.[5] A common assumption in such work is that N1−H1 vectors in G and N3−H3 vectors in U reside in the planes of their bases and coincide with the bisectors of their respective heavy-atom angles (C2−N1−C6 in G and C2−N3−C4 in U). Here we present the results of density functional theory (DFT) calculations[6] showing that the imino ^{15}N−^{1}H vectors can deviate significantly from these idealized orientations, particularly with regard to the assumption that they are located in the base plane. We validate this result using experimental RDC measurements and show that the deviation between the imino proton's computed lowest-energy position and its idealized in-plane position closely follows the H-bond-acceptor atom but is also impacted by the presence of nearby protons of the H-bonded base.

H atom coordinates in RNA/DNA structures are generally difficult to determine experimentally. On the one hand, very few X-ray structures of RNA have been solved at atomic resolution, as required for H atoms to be observed directly. On the other, NMR structures are primarily derived from ^{1}H signals but are never determined at the accuracy and precision needed to provide independent quantification of small deviations of these atoms from their idealized positions.

We recently described a method for facile measurement of imino $^{1}D_{NH}$ RDCs in larger nucleic acids and demonstrated this so-called ARTSY experiment for the 71 nucleotide adenine riboswitch RiboA.[7] With a pairwise root-mean-square difference (rmsd) of 1.6 Hz (Pearson's correlation coefficient R_p = 0.954; Q = 0.108),[7] the observed $^{1}D_{NH}$ RDCs fit well to the coordinates of its 2.1 Å resolution X-ray model (PDB entry 1Y26)[8] when the imino protons were built into the X-ray structure with standard in-plane geometry using the program REDUCE (Figure 1).[9] However, this rmsd of 1.6 Hz is more than 2-fold higher than the estimated uncertainty in the RDC measurements. Such an increased $^{1}D_{NH}$ rmsd is also commonly

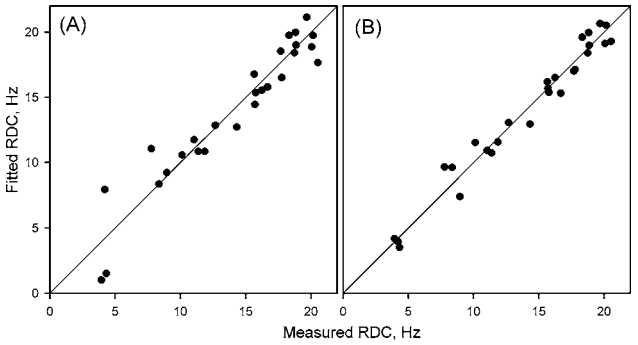

Figure 1. Fits of the experimental imino ^{15}N−^{1}H RDCs to the X-ray structure of RiboA (PDB entry 1Y26) with (A) imino protons added to the X-ray structure in idealized positions using the program REDUCE (rmsd = 1.6 Hz; Q = 0.108) and (B) imino N−H vector orientations derived from DFT calculations carried out on the isolated base pairs and subsequently grafted onto the X-ray structure (rmsd = 0.92 Hz; Q = 0.062).

Received: February 22, 2012
Published: April 11, 2012

observed for proteins and has been attributed to "structural noise",[10] representing the difficulty in determining both the precise orientation of the peptide plane from the X-ray electron density and the true out-of-plane orientation of the amide $^{15}N-^{1}H$ vector.[11] For RNA and DNA structures studied at high crystallographic resolution, the typically well-defined electron density of the large nucleic acid bases greatly reduces their orientational uncertainty in comparison with peptide planes in protein structures solved at equivalent crystallographic resolution. In view of the high quality of the X-ray structure of RiboA, the elevated rmsd between the experimental RDCs measured in solution and the values predicted by the X-ray structure therefore must result from either small differences between these vectors in the crystalline and solution states or deviations of the N−H vectors from their idealized in-plane positions. Here we demonstrate for RiboA that optimization of the imino proton positions by DFT calculations improves the agreement between the RDCs and the X-ray structure by >40% (Figure 1).

To generate the input for the DFT calculations, the heavy-atom coordinates of the base pairs were taken from the high-resolution X-ray structure of RiboA (PDB entry 1Y26).[8] The H atoms were added using the program REDUCE,[9] and the ribose was replaced by a methyl group. With the heavy atoms frozen at the input coordinates, the H atom positions were then optimized at the B3LYP/6-31G level using Gaussian 09.[12] The B3LYP method uses Becke's three-parameter hybrid exchange functional[13] and the correlation functional of Lee, Yang, and Parr, which includes both local and nonlocal terms.[14] The 6-31G basis set[15] was used in all of the calculations, unless noted otherwise. To assess the adequacy of this method, calculations also were performed on the RiboA base pairs using second-order Møller−Plesset perturbation theory with the same basis set (MP2/6-31G)[16] or B3LYP with a larger basis set (B3LYP/6-311++G**).[17] The results showed minimal dependence of the imino N−H vector orientations on the quantum theory level or the basis set size (rmsd's of 0.13 and 0.23° for the in-plane and out-of-plane angles, respectively, between the MP2/6-31G and B3LYP/6-31G structures and rmsd's of 0.18 and 0.33° for these angles between the B3LYP/6-311++G** and B3LYP/6-31G structures). Moreover, the experimental $^{1}D_{NH}$ values fit equally well to the X-ray structure of RiboA using imino hydrogens obtained from the B3LYP/6-31G geometries (Q = 0.062) and those obtained from the more computationally intensive calculations (Q = 0.060 and 0.063 for MP2/6-31G and B3LYP/6-311++G**, respectively).

Histograms of the in-plane and out-of-plane deviations from the idealized geometry in RiboA reveal that out-of-plane deviations can be quite large, while the range of in-plane variations is ca. 2-fold smaller (Figure 2). A similar conclusion was previously reached for the amide N−H vector relative to the peptide plane in proteins.[11]

To gain insight into the physical interactions that cause the significant deviations from ideality for the imino bond vectors, we investigated the impact of H-bond-acceptor and other nearby atoms using DFT calculations. In our model, the imino hydrogen within the base pair is affected by three types of forces: (1) a force that attempts to restore its position to the idealized geometry of the isolated base; (2) an attractive interaction with the H-bond-acceptor atom(s) of the paired base; and (3) repulsive terms between the imino H^N atom and the proximal H atoms of the paired base. The attractive interaction involves the electron distribution(s) surrounding

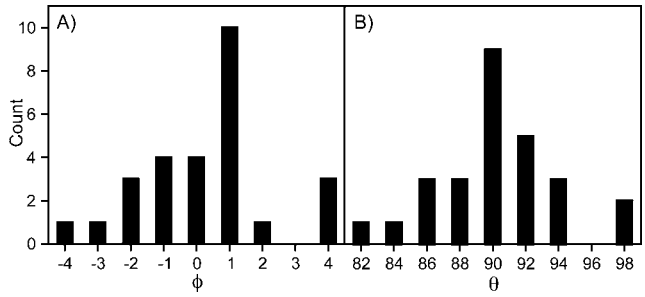

Figure 2. Histograms of the differences between standard idealized geometry (ϕ = 0°; θ = 90°) and DFT-optimized orientations for the H-bonded imino groups in RiboA: (A) in-plane angle, ϕ; (B) angle θ relative to the base normal. For the signs of the angles, see Figure 3.

the H-bond-acceptor atom(s). We approximated the center of the attractive potential by a virtual point (VP) within the base plane of the paired base rather than using the N (or O) nucleus itself. The exact position of this VP was optimized during the fitting of the DFT results. The H−H repulsive potential was parametrized as a power function, an approximation valid within the small range (ca. ±0.5 Å) of the H/H distances sampled by common interacting base geometries in the crystallographic database. Using a large number of geometries taken from this database [Figure S1 in the Supporting Information (SI)] as input for DFT optimization of the imino H^N positions, we then aimed to parametrize the DFT-optimized H^N positions in terms of the following potential function:

$$U(r, \theta, \phi) = k_r(r - r_0)^2 + k_\theta(\theta - \theta_0)^2 + k_\phi(\phi - \phi_0)^2 - \frac{k_{el}}{r_{H-VP}} + k_{H-H} \sum_{r_{H-H \le 3Å}} r_{H-H}^{\alpha_{H-H}} \quad (1)$$

where r, θ, and ϕ are the spherical coordinates describing the position of H^N relative to the plane of its base (Figure 3) and k_r, k_θ, and k_ϕ represent the resistance of the N−H vector to changes in length and out-of-plane and in-plane bending, respectively, relative to the calculated bond length r_0 and angles θ_0 = 90° and ϕ_0 for the isolated base in vacuum (see the Figure 3 caption). The Coulombic potential for the interaction with the acceptor atom's electrons is described by the $-k_{el}/r_{H-VP}$ term, where r_{H-VP} is the distance between H^N and the VP and the force constant k_{el} encapsulates the values of the effective charges of the H and acceptor atoms. The final term in eq 1 includes the H−H distances r_{H-H} with an exponent α_{H-H} and corresponds to the repulsive interactions of the imino proton with proximate ($r_{H-H} \le 3$ Å) protons on the acceptor base. This term is particularly important in A:U, U:U, and G:U wobble base pairs, where close proximity between the protons occurs (Figure S2).

We limited the parametrization of eq 1 to geometries that are commonly found in nature by considering only those RNA structures that are observed in X-ray structures deposited in the RCSB PDB database or small variations thereof (Figure S1). Considering that the imino H^N generally is NMR-visible only when it is involved in a stable H-bonding interaction, we selected from the database a set of base pairs for which the H1 atom of G or the H3 atom of U is H-bonded to another base. All of the significantly populated base-pair types where this applies, specifically the Watson−Crick G:C (226), Watson−Crick A:U (202), G:U wobble (79), U:U asymmetric ($N3_i-O2_j$ and $N3_j-O4_i$; 60), and reverse-Hoogstein A:U (39) base pairs, were included in our analysis; the number of each type of base

dx.doi.org/10.1021/ja301775j | J. Am. Chem. Soc. 2012, 134, 6956−6959

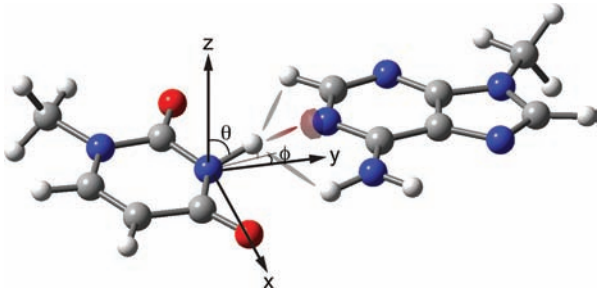

Figure 3. Schematic representation of the coordinate frame of the N−H vector in the frame of the H-bond-donor base. The z axis is normal to the base plane and defined as parallel to the cross product of the C2−N3 and C4−N3 vectors (for U) or C2−N1 and C6−N1 vectors (for G); the y axis is opposite to the bisector of the C−N−C angle. θ and ϕ are the polar angles of the N−H vector within this coordinate frame, with $\phi = 0°$ when the N−H projection on the xy plane coincides with the y axis and $\phi > 0°$ when the angle relative to the x axis is >90°. The DFT-calculated values for the free G (U) base in vacuum are r_0 = 1.0123 (1.0127) Å; $\phi_0 = -3.7$ (+0.5)°, and $\theta_0 = 90°$. The interactions of the H^N atom with the H-bond-acceptor and proximal H atoms are shown as shaded ovals. The semitransparent brown oval represents the lone-pair electron density on the H-bond-acceptor atom as obtained from NBO analysis[18] (plotted at an iso value of 0.42 e/Å^3) and correlates with the location of the virtual point (VP) of attraction. Definitions of the axis systems for the positions of the VPs are presented in Table S2 and Figure S5.

pair for which DFT calculations were carried out is given in parentheses.

For all of these base pairs, idealized isolated-base coordinates, with a methyl group substituted for the ribose, were first best-fit to the crystallographic coordinates of the heavy atoms of each base and subsequently used to replace them. Next, H atom coordinates were optimized by DFT energy minimization, keeping all of the heavy-atom positions fixed. After generation of the base pairs with DFT-optimized H atom positions, a simultaneous fit of the parameters of eq 1 by nonlinear least squares (NLS) minimization of the rmsd between the DFT-optimized r, θ, and ϕ values and those corresponding to the minimum of the trial potential in eq 1 was carried out over the full set of DFT-optimized coordinates. The optimized parameters included the H−H repulsion descriptors k_{H-H} and α_{H-H}; donor-specific values for the force constants k_r, k_θ, and k_ϕ; and acceptor-specific force constants k_{el} for the electrostatic attractive potential (Tables 1 and 2). Because of the arbitrary scale of the energy function in eq 1, all of the reported values are relative to the force constant for the length of the U:N3−H3 bond, which was assumed to be unity. The values obtained for k_{el} and the locations of the VP relative to the H-bond-acceptor atoms varies slightly with the acceptor atom type (Table 1). We found that on average, the VP falls 0.3−0.4 Å closer to the imino proton than the nucleus of the H-bond acceptor (Table 1). Additional details are presented in the SI. The VP locations obtained from eq 1 fall remarkably close to canonical positions of sp^2-hybridized lone-pair electrons for the N and O atoms, consistent with the H-bonding geometries observed in high-resolution X-ray structures of small molecules.[19] However, natural bond orbital (NBO) analysis of H-bonding to the carbonyl-containing bases showed a similarity to a recent report on H-bonds in proteins,[20] where carbonyl lone pairs were shown to exist as nondegenerate "s-rich" and "p-rich" orbitals coaligned and oriented perpendicularly, respectively, to the C−O bond vector. For that reason, we simply refer to the center of electrostatic attraction as a virtual point.

Table 1. Parameters of the Interaction Potential (eq 1) Best-Fit to the DFT-Optimized U and G Imino H Coordinates in RNA[a]

H-bond acceptor	r_{VP} (Å)	ϕ_{VP} (deg)	k_{el} (Å)
A:N1[b]	0.383	−2.1	0.167
A:N7[c]	0.438	7.4	0.147
C:N3[d]	0.372	−6.2	0.143
G:O6[e]	0.267	67.5	0.134
U:O2, U:O4[f]	0.424	65.9	0.110

[a]The parameters r_{VP} and ϕ_{VP} describe the distance and in-plane angle of the attractive virtual point of the H-bond-acceptor atom, which is assumed to be in the plane of the acceptor base. Positive signs of the in-plane angles correspond to right-handed rotations around the z axis of the molecular frame, as defined in Table S2. [b]Watson−Crick U:A. [c]Reverse-Hoogstein U:A. [d]Watson−Crick G:C. [e]U:G wobble. [f]U:U, G:U wobble.

Table 2. Force Constants for the Harmonic Stiffness and H−H Repulsion Terms in eq 1[a]

bond vector	k_r (Å^{-2})	k_θ (deg^{-2})	k_ϕ (deg^{-2})	k_{H-H}	α_{H-H}
U:N3−H3	1.000	1.514×10^{-5}	5.742×10^{-5}	0.132	−2.43
G:N1−H1	1.391	1.420×10^{-5}	5.276×10^{-5}		

[a]Values are defined relative to the stretching force constant for U:N3−H3.

Given the heavy-atom coordinates, the parametrized form of eq 1 was then used to predict the deviation of the N−H vectors from their ideal positions in a structure by means of site-specific NLS optimization of the r, θ, and ϕ parameters against eq 1. Some of the protons proximal to the imino proton in question are amino protons on the paired base, whose positions are affected by their own H-bonding. Fitted potentials of the form of eq 1 account for this, with their updated coordinates returned by the NLS procedure. The routine can be accessed through our webserver (http://spin.niddk.nih.gov/bax/nmrserver/pdbutil/rnahn.html). The N−H orientations obtained are in excellent agreement with the DFT-derived polar angles (Figures 3 and 4), and the N−H bond length obtained

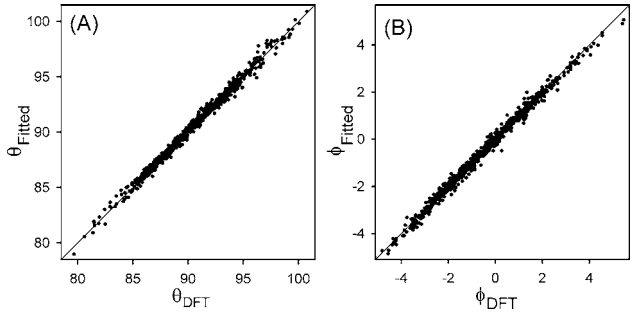

Figure 4. Correlation between the DFT-optimized polar angles of the N−H vectors of G and U for all fitted base pair types and those predicted by application of eq 1.

from eq 1 closely mirrors its very steep dependence on the distance between the H-bond-donor and -acceptor atoms (Figures S3 and S4). Similarly, the quality of the fit of the experimental RiboA RDCs to the X-ray structure with the

dx.doi.org/10.1021/ja301775j | J. Am. Chem. Soc. 2012, 134, 6956−6959

imino ^{1}H positions obtained from eq 1 (rmsd 0.94 Hz) approaches that obtained when using the DFT-optimized imino ^{1}H positions (0.92 Hz), both of which are much better than the fit obtained for the standard geometry (1.59 Hz). Interestingly, other NMR parameters such as H^N chemical shifts or through-H-bond J couplings ($^{2h}J_{NN}$) are rather insensitive to small changes in the orientation of the N−H bond and instead are dominated by the H-bond donor−acceptor distance (Figure S6).

The parametrization of eq 1 indicates that for both U and G, the potential for in-plane bending of the N−H vector from its idealized position is ca. 3.7 times steeper than for out-of-plane angular deviations (Table 2), while the individual k_ϕ and k_θ values are similar for the two nucleotide types. The impact of adjacent, stacked bases on the deviations of the N−H vectors from their ideal positions was evaluated by a series of DFT calculations in which the central base pair was surrounded on both sides by stacked base pairs taken from the X-ray structures. Comparison of the resulting θ and ϕ values for the N−H vector in the central pair with those of the corresponding isolated base pair showed rmsd's for the out-of-plane angle θ of 0.85° for G and 0.70° for U and rmsd's for the in-plane angle ϕ of 0.11° for G and 0.22° for U. Clearly, the impact of these neighboring bases is much smaller than that from the paired base, and for all practical purposes it can be safely neglected. A fit of the experimental RDCs to DFT-derived N−H orientations that included the effect of adjacent stacked base pairs indeed showed a minimal impact on the fit quality (rmsd of 0.88 Hz vs 0.92 Hz when stacked bases were not considered; Table S1).

Our results clearly demonstrate that the assumption of idealized geometry for the positions of imino H atoms in nucleic acids can limit the agreement obtainable between experimental RDCs and vector orientations obtained from heavy-atom coordinates. When the idealized geometry is assumed during NMR structure calculation, forcing the RDCs to fit within the experimental error can therefore result in overfitting of the data and cause small distortions of the structure. However, because the deviation from the idealized geometry can be accurately calculated by DFT and simple parametrization in terms of attractive and repulsive terms allows for rapid and accurate positioning of these key atoms in structural studies, iterative construction of a self-consistent structural model that fits both the experimental RDCs and the small deviations from idealized ^{1}H positions becomes possible. Improved fits of RDCs to X-ray coordinates of subunits in larger RNA structures also will yield higher precision for the alignment tensor, which is a critical factor when extracting global motions from RDCs in such structures.[21]

■ ASSOCIATED CONTENT

Supporting Information

Complete ref 12; details of DFT calculations and the computational procedure for parametrizing eq 1; and figures showing the impact of H−H repulsion on the in-plane orientation of the N−H vector and comparisons of chemical shifts and $^{2h}J_{NN}$ couplings for idealized and DFT-optimized N−H orientations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

bax@nih.gov

Author Contributions

†A.G. and J.Y. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Fernando Clemente of Gaussian, Inc. for helpful discussions and Dr. Yun-Xing Wang for the RiboA sample. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Maryland (http://biowulf.nih.gov). This work was funded by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health (NIH), and the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH.

■ REFERENCES

(1) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293. Dingley, A. J.; Masse, J. E.; Peterson, R. D.; Barfield, M.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 6019.

(2) Wang, Y.-X.; Zuo, X.; Wang, J.; Yu, P.; Butcher, S. E. *Methods* **2010**, *52*, 180.

(3) Hare, D. R.; Reid, B. R. *Biochemistry* **1982**, *21*, 5129.

(4) Wang, J. B.; Zuo, X. B.; Yu, P.; Xu, H.; Starich, M. R.; Tiede, D. M.; Shapiro, B. A.; Schwieters, C. D.; Wang, Y. X. *J. Mol. Biol.* **2009**, *393*, 717.

(5) Stelzer, A. C.; Frank, A. T.; Bailor, M. H.; Andricioaei, I.; Al-Hashimi, H. M. *Methods* **2009**, *49*, 167. Kim, N.-K.; Zhang, Q.; Zhou, J.; Theimer, C. A.; Peterson, R. D.; Feigon, J. *J. Mol. Biol.* **2008**, *384*, 1249. Zhang, Q.; Kim, N.-K.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 20325.

(6) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(7) Ying, J.; Wang, J.; Grishaev, A.; Yu, P.; Wang, Y.-X.; Bax, A. *J. Biomol. NMR* **2011**, *51*, 89.

(8) Serganov, A.; Patel, D. J. *Nat. Rev. Genet.* **2007**, *8*, 776.

(9) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735.

(10) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127.

(11) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179.

(12) Frisch, M. J.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2010.

(13) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(14) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(15) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(16) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618. Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.

(17) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650. Davidson, E. R.; Feller, D. *Chem. Rev.* **1986**, *86*, 681.

(18) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211.

(19) Steiner, T. *Angew. Chem., Int. Ed.* **2002**, *41*, 48.

(20) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. *Nat. Chem. Biol.* **2010**, *6*, 615.

(21) Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M. *Nature* **2007**, *450*, 1263.