THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

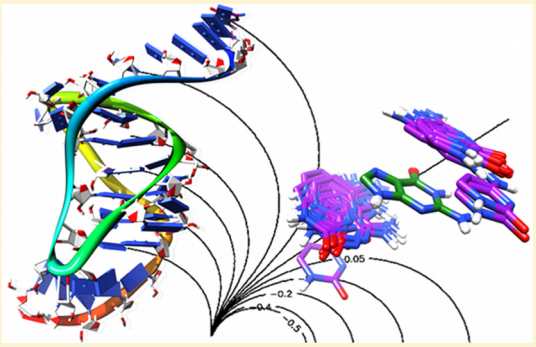
Analysis of the Contributions of Ring Current and Electric Field Effects to the Chemical Shifts of RNA Bases

Aleksandr B. Sahakyan and Michele Vendruscolo*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

Supporting Information

ABSTRACT: Ring current and electric field effects can considerably influence NMR chemical shifts in biomolecules. Understanding such effects is particularly important for the development of accurate mappings between chemical shifts and the structures of nucleic acids. In this work, we first analyzed the Pople and the Haigh−Mallion models in terms of their ability to describe nitrogen base conjugated ring effects. We then created a database (DiBaseRNA) of three-dimensional arrangements of RNA base pairs from X-ray structures, calculated the corresponding chemical shifts via a hybrid density functional theory approach and used the results to parametrize the ring current and electric field effects in RNA bases. Next, we studied the coupling of the electric field and ring current effects for different inter-ring arrangements found in RNA bases using linear model fitting, with joint electric field and ring current, as well as only electric field and only ring current approximations. Taken together, our results provide a characterization of the interdependence of ring current and electric field geometric factors, which is shown to be especially important for the chemical shifts of non-hydrogen atoms in RNA bases.



INTRODUCTION

The use of NMR chemical shifts for protein structure determination is becoming a tool of increasing importance in structural biology.[1−8] However, the same approach is not yet widely applied to nucleic acids. This situation arises from the challenges in developing structure-based chemical shift predictors for nucleic acids that possess a sufficient accuracy in predicting structure-induced chemical shift variation for individual atom types. Progress in this area can be expected on the basis of a series of seminal studies that have demonstrated that the ideas established to explain the conformational dependence of chemical shifts in proteins are, at least in part, transferrable to nucleic acids.[9−12]

In order to increase the accuracy of nucleic acid chemical shift predictors, our understanding of the factors that modulate chemical shifts in these biomolecules should be further developed. Ring current effects can considerably affect chemical shift values, and this influence is particularly pronounced in nucleic acids since they are abundant in conjugated rings. Therefore, it is important to study ring current effects in nucleic acids, in particular to investigate their influence on non-hydrogen atoms and to examine their coupling with conformational and electric field effects, which are also relevant in nucleic acids because of the highly charged nature of the polynucleotide chains.

Ring current effects on ^{1}H chemical shifts in proteins and nucleic acids have been thoroughly studied by Case and coworkers[1,13,14] and parametrized for nucleic acid bases[14] through the Haigh−Mallion[15,16] and Waugh−Fessenden−Johnson−Bovey[17,18] models. Most current chemical shift predictors adopt the Pople[19] and Haigh−Mallion models, with the preference being toward the Pople model owing to the simplicity of its implementation in molecular dynamics simulations.

In this work, we study ring current effects by comparing the Pople and Haigh−Mallion models. We propose equations for mapping one model onto the other in order either to increase the computational efficiency of existing parametrizations that already utilize the more complex Haigh−Mallion model or to increase the accuracy of the simpler Pople model. We then focus on RNA, by extending the study of ring current effects on the chemical shifts of heavy nuclei (^{13}C, ^{15}N, and ^{17}O) and investigating the interdependence of ring current and electric field effects in modulating chemical shift values. As model systems for such studies, inter-ring arrangements in RNA bases are generated by compiling a dibase atlas of conformations for all the possible pairs found in high-resolution RNA structure database.[20] Taking into account the different sign conventions used in previous studies where ring current effects have been discussed, we also review the most widely used ring current models in a unified notation and sign convention, in order to prevent possible confusion in future work.

The present study is aimed at describing the specific effects of RNA conjugated rings on RNA chemical shifts. We provide

Received: June 12, 2012
Revised: January 10, 2013
Published: February 11, 2013

ACS Publications © 2013 American Chemical Society

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989−1998

such a description through geometric factors and parameters that model the effects of ring currents and electric fields of the conjugated rings on the chemical shifts of neighboring RNA atoms. With the majority of the results that we present being transferrable to any 6- and 5-membered rings, this study can be useful for solving a wide range of problems that involve the analysis of ring current contributions in nuclear shielding phenomena. Furthermore, additional studies, which will consider the contributions to chemical shifts of the effects of other RNA moieties, will enable the major factors that determine the chemical shifts in RNAs to be brought together in a high-quality structure-based predictor of nucleic acid chemical shifts. We expect that this type of predictions will enable the refinement of RNA structures based on chemical shifts measured via solid- or solution-state NMR spectroscopy, in a procedure similar to the one successfully used for proteins.[6,21]

METHODS

All the quantum mechanical calculations are done using the Gaussian03 suite of programs.[22] All the scripting and the linear model fitting in the work are done using the *R* programming language for statistical computing.[23] The scripts can be obtained from the authors upon request. The RingPar web server for accessing ring current and electric field parameters is created using *Rwui*, an interface to generate web servers based on *R*-scripts.[24] The RingPar server and the DiBaseRNA database can be accessed at http://www-vendruscolo.ch.cam.ac.uk/software.html.

Further details on the calculations and database analyses are presented along with the discussion in the following sections.

RESULTS AND DISCUSSION

Ring Current Models. Since the successful estimation of diamagnetic anisotropy of crystalline benzene by Lonsdale's and Pauling's assumption of π-electron precession along the ring atoms,[25,26] ring current effects have become one of the most important aspects of NMR spectroscopy. Theoretical models of ring current effects emerged from both classical and quantum mechanical approaches,[16] which were also accompanied by empirical look-up tables that widened the usage of ring current evaluations in the proximity of simple conjugated systems.[15,18] Of the numerous theoretical models,[16] three have received a considerable attention owing to the ease of their implementation[19] and the availability of the derived empirical tables.[15,18] Because of the often confusing mismatch of signs in different publications, we provide a brief description of these three models using a common sign convention.

The assumption of the secondary magnetic field created by the electric current circulating in a benzene ring[25,26] was soon followed by the Pople simplification of the magnetic field description,[19] with the source of the magnetic field approximated by a magnetic dipole positioned at the ring center. The magnetic dipole has a magnitude of $ne^2a^2B_0/4\pi mc^2$, where n is the number of circulating electrons, a is the radius of the ring (usually, 1.39 Å is taken for benzene ring, which is equal to the carbon−carbon bond length), B_0 is the applied uniform magnetic field, and e, m, and c have their conventional meaning and, in part, emerge from the expression for precession frequency ($\omega_L = -eB_0/2mc$). The secondary or induced magnetic field B_{ind} at any point around that dipole is determined by the expression $B_{ind} = ne^2a^2B_0(1 - 3\cos^2\theta)/4\pi mc^2r^3$ with θ being the angle between the query point and the ring normal, and r being the distance from the ring center (Figure 1a). Taking into account that $B_{ind} = -\sigma B_0$, where σ is the isotropic nuclear shielding constant, the expression for the change in σ_{ring} isotropic nuclear shielding constant in ppm originated by the ring current effect according to the Pople point dipole model can be written as

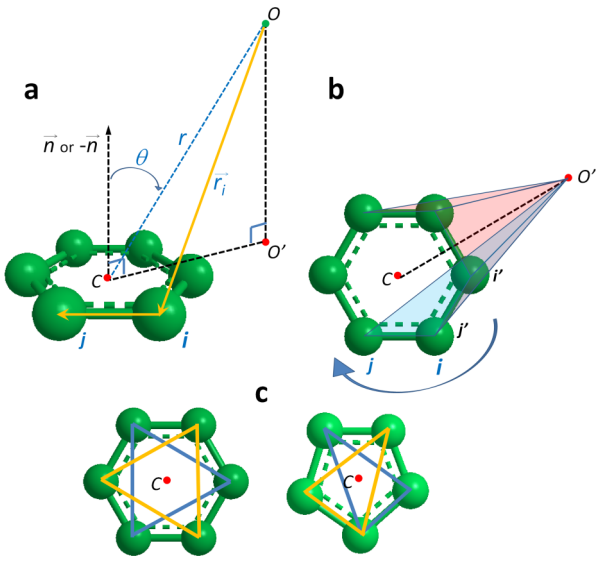


Figure 1. Schematic representation of the geometric concepts used in the Pople (a) and Haigh−Mallion (a,b) models of ring currents, along with our suggested convention for determining the ring normals (c) for 6- and 5-membered rings, which results in geometric factors more stable and robust against out-of-plane geometric fluctuations of the constituent ring atoms.

$$\Delta\sigma_{ring}^{P} = 10^6 \times \frac{ne^2a^2}{4\pi mc^2} \times \frac{3\cos^2\theta - 1}{r^3} \quad (1)$$

Since the chemical shift is the negative of the nuclear shielding constant, the geometric factor $(1 - 3\cos^2\theta)/r^3$, especially when geometric terms of different models are outlined in comparison, should be used only for chemical shifts, not for the nuclear shielding constants. Furthermore, the interconnection between the local or effective B_{loc}, the applied B_0 magnetic fields and the isotropic nuclear shielding constant is given by the expression $B_{loc} = B_0(1 - \sigma)$. The interconnection between the local and applied magnetic fields can be described as $B_{loc} = B_0 + B_{ind}$. In this context, the expression $B_{loc} = B_0 - B_{ind}$ can often be found because the induced magnetic field usually opposes the external one. However, we favor the first expression, where the negative sign is embedded within B_{ind} and is in intrinsic correspondence with the sign and concept of a nuclear shielding constant.

A more detailed classical description of the ring currents was suggested by Waugh and Fessenden[17] and then improved and parametrized for benzene by Johnson and Bovey.[18] In this approach, the complete classical description of the electric current circulating in a loop of radius a is considered. The ring current model was also extended to account the nature of the π-orbitals by assigning two loops, above and below the ring plane and separated by $\Delta z = z_2 - z_1$. In this case, each of the two loops possesses $n/2$ circulating electrons, and thus the general form of the equation for $\Delta\sigma_{ring}$ in ppm is given by

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989−1998

$$\Delta\sigma_{\text{ring}}^{\text{WFJB}} = 10^6 \times \frac{ne^2}{12\pi mc^2 a} \times \sum_{p=1}^{2}\left(\frac{1}{\sqrt{(1+\rho)^2 + {z_p}^2}}\left(K(k) + \frac{1-\rho^2-{z_p}^2}{(1-\rho)^2+{z_p}^2}E(k)\right)\right) \quad (2)$$

which is expressed in a cylindrical coordinate system centered at the ring center with z and ρ given in the unit defined by the loop radius a. K and E are the complete elliptic integrals of the argument k, which is defined by the expression ($k = (4\rho/[(1+\rho)^2 + {z_p}^2])^{1/2}$). A 1.28 Å (0.918$a$, with radius a taken to be equal to the benzene ring radius) separation between the loops was found to be optimal to represent the hydrogen shielding in benzene.[18] To escape possible confusion, it should be noted that, although the chemical shift δ notation was used in the original articles,[17,18] the expression reflects the $\Delta\sigma_{\text{ring}}$ change in isotropic nuclear shielding constant, and a minus sign should be used to obtain $\Delta\delta_{\text{ring}}$.

The third ring current model discussed here was proposed by Haigh and Mallion, based on the London and McWeeny approximations and Hückel molecular orbital theory.[15,16] Thus, the method can be regarded as an empirical quantum mechanical (QM) model. In its simplified representation, the secondary magnetic field follows the $B_{\text{ind}} \approx -J_{\text{ring}}\sum_{ij}[S_{ij}(1/r_i^3 + 1/r_j^3)]$ proportionality. In this expression, J_{ring} is a quantum mechanical quantity calculated from the Coulson bond orders P_{ij} and the so-called mutual bond–bond polarizabilities $\bar{\pi}_{(ij)(kl)}$ within the Hückel formalism. S_{ij} is the algebraic (signed) triangle area formed by the O' projection of the query point O onto the ring plane and the ring atoms i and j (Figure 1a,b). Denoting $\mathbf{T}_{O'i}$ and $\mathbf{T}_{ij}$ as vectors joining O' to the ring atom i and ring atom i to j, respectively, the sign of the triangle is positive if the vector product $\mathbf{T}_{O'i} \times \mathbf{T}_{ij}$ has the same direction as the ring normal with ring atoms counted in $i \rightarrow j$ direction. In addition, r_i and r_j are the distances between O and atoms i and j, respectively. The summation goes over all the adjacent ij atom pairs forming the ring, thus with the number of constituents being equal to the number of bonds in the conjugated ring. The expression for the ring current contribution to the change in the isotropic nuclear shielding constant is

$$\Delta\sigma_{\text{ring}}^{\text{HM}} = 10^6 \times KJ_{\text{ring}} \times \sum_{ij} S_{ij}\left(\frac{1}{r_i^3} + \frac{1}{r_j^3}\right) \quad (3)$$

where K is a proportionality constant, and the minus sign is discarded to follow the $\sigma = -B_{\text{ind}}/B_0$ definition for the nuclear shielding constant. The minus sign was present in ref 15, which was canceled by the negative parameter, calculated for benzene. The comparative mismatch of the signs is present in ref 27 where the Haigh–Mallion geometric factor, in a form consistent with nuclear shielding constant, was used to calculate chemical shift difference ($\Delta\delta = -\Delta\sigma$). However, the further reverse definition of the algebraic sign of the S_{ij} triangle areas changed the sign of the expression making consistent with chemical shifts.

Paying attention to the last factors in eqs 1–3, one can see that, in each of the described M models, the $\Delta\sigma_{\text{ring}}^{\text{M}}$ term can be represented as a geometric factor $G^{\text{M}}(\vec{\mathbf{r}})$ multiplied by a proportionality constant K^{M}. The geometric factor itself only describes the geometric arrangement of the query point, where the shielding needs to be evaluated, relative to the conjugated ring. An exception is the geometric factor $G^{\text{WFJB}}(\vec{\mathbf{r}})$ of the Waugh–Fessenden–Johnson–Bovey model, which also includes an adjustable parameter Δz, describing the separation between the two loops with circulating electrons above and below the ring plane.[17,18]

In a more general case of the conjugated system comprised of multiple cycles, ring current effects on nuclear shielding constant can be represented as a sum of the effects from each cycle c

$$\Delta\sigma_{\text{ring}}^{\text{M}} = \sum_{c} K_c^{\text{M}} G_c^{\text{M}}(\vec{\mathbf{r}}) \quad (4)$$

The geometric factor in eq 4 has the same sign as the change in the nuclear shielding constant $\Delta\sigma_{\text{ring}}^{\text{M}}$ and the reverse sign of the change in chemical shifts $\Delta\delta_{\text{ring}}^{\text{M}}$. The $G_c^{\text{M}}(\vec{\mathbf{r}})$ is determined by the corresponding geometric factors in eqs 1–3, which describe intramolecular geometric relation, independent of overall molecular orientation and tumbling motion.

Comparative Analysis of Pople and Haigh–Mallion Ring Current Models: The Case of Benzene. Since we are interested in the study of ring current models that, upon the inclusion of other major factors that determine the values of chemical shifts in RNA, will eventually enable the use of these NMR parameters as structural restraints in molecular dynamics simulations, we will continue our discussion by focusing on the Pople[19] and Haigh–Mallion[15,16] models. Although the Waugh–Fessenden–Johnson–Bovey[17,18] model is differentiable, its geometric factor, unlike the ones in other models, contains an adjustable parameter that should be optimized for different conjugated systems. The Haigh–Mallion model has become the most applied framework for ring current induced chemical shift change evaluation. It should be added, however, that, as demonstrated before, if thoroughly parametrized, the precision of the point dipole model is comparable to Waugh–Fessenden–Johnson–Bovey model and is only slightly worse than the performance of the Haigh–Mallion model.[28]

As mentioned before, ring current effects are significant modulators of chemical shifts in nucleic acids. Nucleic acids are constructed of different conjugated rings with conformational distribution covering highly shielded and deshielded regions, from very close, stacked, to planar, hydrogen bonded, states. Therefore, the quality of structure-based chemical shift predictors for nucleic acids is expected to be highly dependent on the accuracy of the description of ring currents. We thus return to the problem of the comparison of the ring current models, this time by also assessing the errors that are to be expected from different inter-ring arrangements.

We implemented the Pople and Haigh–Mallion models by following the conventions described above. Our results indicate that the large fluctuations in the direction of the ring normal, which should be calculated in both models, is the main reason for the instabilities in chemical shift predictions and restrained molecular dynamics simulations. The fluctuations of the ring normal orientation occur because of slight out-of-plane displacements of the ring atoms. A sophisticated approach for eliminating this problem would be to fit a plane via least-squares fitting by using all non-hydrogen atoms of the ring. Although proposed and successfully implemented before for ring current evaluations,[29] the fitting procedure can increase the computational time required for chemical shift restrained molecular dynamics simulations. To this end, we suggest the usage of two noninterconnected sets of atoms for defining two

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989–1998

planes (Figure 1c), so that two ring normals, instead of one, can be calculated from the mentioned atom sets for then taking the average of the two vectors.

We begin from the simplest and the most studied case, the benzene molecule. The molecular structure of benzene was geometry optimized using hybrid-DFT (density functional theory[30]) with B3LYP exchange-correlation functional[31−33] and 6-311+G(d,p) split-valence basis set.[34] The geometry optimization was carried out with D_{6h} symmetry constraints and tight convergence criteria. This procedure resulted in C−C bond length, further taken as ring radius a, equal to 1.3946 Å, which is quite close to the widely accepted zero-point average C−C distance for benzene, 1.395 Å.[35]

The ring current geometric factors, from both Pople and Haigh−Mallion models, are then calculated for 100,000 points uniformly distributed around the benzene ring in the orthogonal plane corresponding to $\phi = 0$ within the cylindrical coordinates ranging from 0 to 4 units of benzene radius a for both z and ρ coordinates relative to the benzene ring (Figure 2).

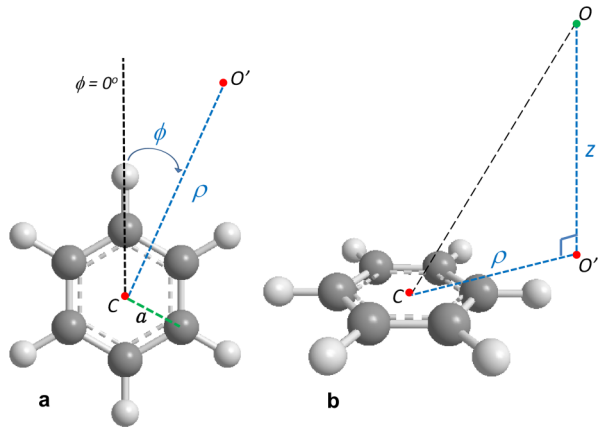


Figure 2. Cylindrical coordinates and notation associated with a point O in the proximity of the benzene ring.

The resulting maps are shown for both Pople (Figure 3a) and Haigh−Mallion (Figure 3b) geometric factors. The white-colored area in the maps (Figure 3) corresponds to a discarded region with extremely large absolute values of the geometric factor.

The correlation between the Pople and Haigh−Mallion geometric factors is shown in Figure 4, where the data are broken down into four different regions of the proton position around the benzene ring. By drawing an analogy with dibase arrangement in nucleic acids, A and B would correspond to the stacking inter-ring arrangement, C is for the diagonal, and D is for hydrogen-bonded planar arrangements. The border specification of the zones A, B, C, and D are $\rho \leq (1.7 + a)$ Å and $2.9 > z > 1.7$ Å; $\rho \leq (1.7 + a)$ Å and $z \geq 2.9$ Å; $\rho > (1.7 + a)$ Å and $z > 1.7$ Å; $\rho > (1.7 + a)$ Å and $z \leq 1.7$ Å, respectively, where 2.9 Å is the sum of the carbon and hydrogen van der Waals radii and 1.7 Å is the van der Waals radius of the hydrogen atom.

The correlation, in general, has a consistent shape, suggesting that the geometric factors can be translated into each other, hence eliminating the small underperformance of the simple point dipole model as compared to the Haigh−Mallion one.[28] The only exception is the region A (Figure 4, red). It should be noted, however, that, in practice, the distance between a proton

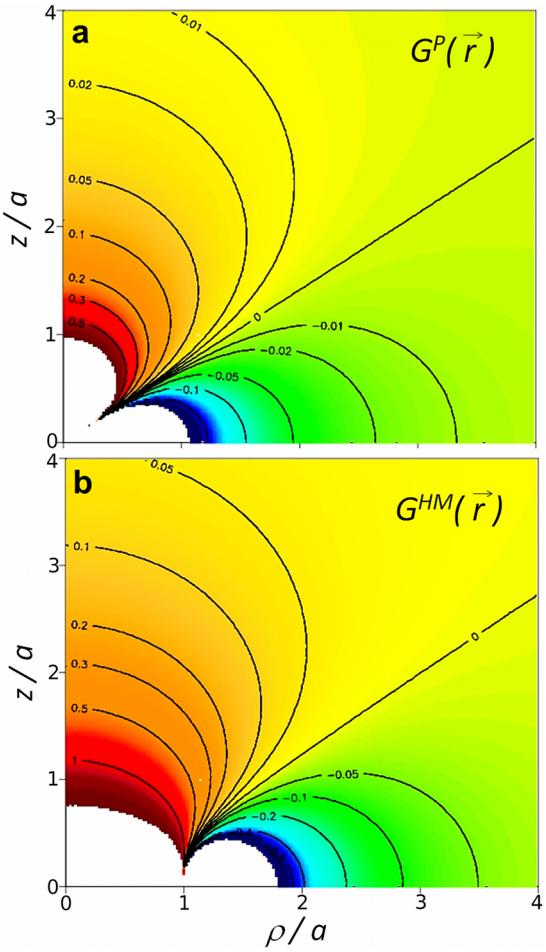


Figure 3. Maps of the values of the geometric factors in the proximity of a benzene ring as defined by Pople (a) and Haigh−Mallion (b) models.

of one nucleic acid base and the ring center in another nucleic acid base is almost always greater than 3.0 Å for the stacked arrangement and more than $1.7 + a$ Å for the coplanar arrangement when the protons as close as the ones participating in hydrogen bonding are considered. Therefore, the divergence between the Pople and Haigh-Mallion models at the A region can only be influential for very strong stacking interactions, perhaps happening in nucleic acid and intercalative drug complexes.

The equations of the interconversion from Pople to Haigh−Mallion geometric factors and vice versa is determined below, using the DiBaseRNA database of dibase arrangements in RNAs (see the next section) to allow the usage of simple and intuitive point dipole model, while enhancing its performance up to the level of the Haigh−Mallion model.

Generation of the DiBaseRNA Database of Inter-Ring Arrangements in RNAs. To study the ring current effects in molecular structures that closely resemble nucleic acids, we created a structural database that reflects the observed interbase arrangement in high-resolution X-ray structures of RNAs. As RNAs possess conformations that are more variable than those of DNAs, focusing initially on RNAs rather than DNAs provides an opportunity to study the ring current effects more generally. The initial RNA structures were taken from the RNA05 database of Richardson and co-workers,[20] which contains 171 coordinate files of RNA X-ray structures with

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989−1998

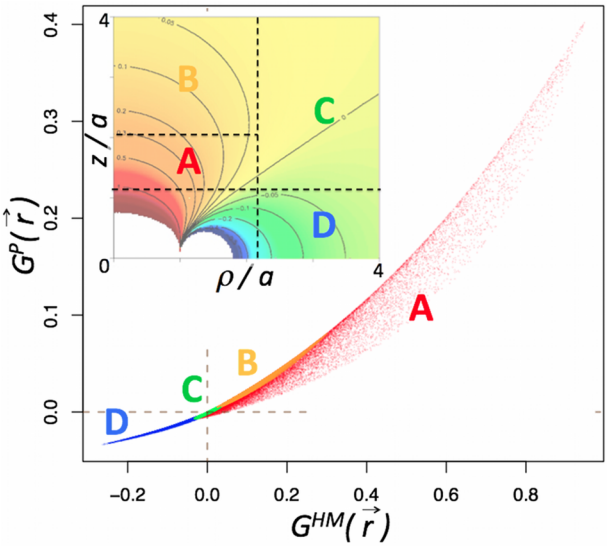


Figure 4. Relationship between the geometric factors of the Haigh–Mallion (*x*-axis) and Pople (*y*-axis) models for ring current effects. The correlations corresponding to four different spatial regions around the ring are differentiated by red, orange, green and blue colors, also denoted by letter A, B, C, and D and clarified in the built-in graph. In the case of two nucleic acid bases, the regions A and B correspond to the stacked arrangement, C is for the diagonal, and D is for coplanar (hydrogen-bonded) arrangements. For the full specification of the borders for the separate spatial regions, see the text.

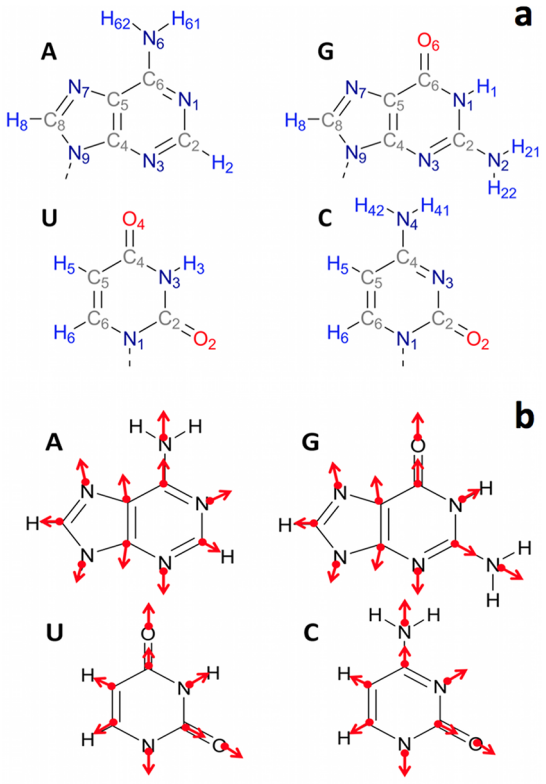


Figure 5. Scheme of the four nitrogen bases of RNA: (a) numbering scheme and (b) the outline of the points where the electric field values, generated by the second base, are calculated, with arrows showing the direction for the considered projections.

9486 nucleotide content and 3.0 Å or better resolution. Then, all the structures with equal to or better than 1.8 Å resolution were scanned, and all the possible dibase arrangements between any pairs among the conjugated rings of adenine (A), guanine (G), cytosine (C), and uracil (U) bases (Figure 5a) were retrieved.

The dibase geometries are classified into three categories: adjacent (ADJ), spatial (SPT), and hydrogen bonded (HBD) arrangements. In this context, ADJ indicates that, in the XY arrangement of the bases X and Y, the conjugated rings belong to the neighboring nucleotides within the same chain. The adjacent dibases are scanned in both 5′ to 3′ and 3′ to 5′ directions for the retrieval. In cases where one of the dibase members is common in both directions (for instance, if we search for the adjacent arrangement of GA in the GAG sequence, A will be common in GA moieties retrieved from both directions), the fragment from the 3′ to 5′ scan is discarded. HBD refers to the arrangement where the bases are nearly coplanar with hydrogen bonds (either classical Watson–Crick or of other types) between them and can belong to the same or different polynucleotide chains. The SPT arrangement is defined as the one where the two bases are not coplanar and belong either to different chains or the same chain but separated by at least 3 nucleotides. They would usually represent the diagonal arrangement of the base-rings for the situations where the RNA molecule is self-assembled into local double helical structures. Further, hydrogens are added to the N9 and N1 positions (Figure 5a) of the purine and pyrimidine rings as a replacement of the glycosidic bond, so that the resulting coordinate files correspond to a complete and closed-shell system of two bases, ready for quantum chemical calculations. An example of the resulting geometries for the GG pair is shown in Figure 6, with the full set presented in Figures S1–S10, Supporting Information.

The DiBaseRNA database was further refined by a partial geometry optimization with frozen core atoms via the semiempirical AM1 Hamiltonian[36] as implemented in the MOPAC2009 package.[37] The AM1 was selected, taking into account its accuracy in representing the amide bond lengths[38] and the geometries of amino groups attached to conjugated systems.[39] Furthermore, the known issue of nitrogen pyramidality overestimation is the least pronounced for AM1[38] in NDDO-type (neglect of diatomic differential overlap) semiempirical methods. In order to remove the residual pyramidality, the N–H bonds were further frozen to be within the same plane as the corresponding conjugated ring. The resulting coordinate files in PDB format represent the variant 1 of the proposed DiBaseRNA database that features experimentally determined positions of the core atoms of each of the conjugated systems. However, small structural variations in the experimental bond lengths within the ring, caused by the experimental errors, are inevitable. Hence, another variant of the DiBaseRNA database was generated (variant 2), where only the relative arrangement of the two rings were learned from the X-ray data and the final files were constructed by rotating and translating the standard (predefined) geometries of the constituent bases to match the experimental arrangement. To obtain the standard geometries, A, G, C, and U bases were geometry optimized without any constraints and with tight convergence criteria. Hybrid density functional theory[30] via the Becke three-parameter exchange functional and the Lee, Yang, and Parr correlation functional (B3LYP)[31−33] is used with the split-valence 6-311+G(2d,p) basis set.[34] After the geometry optimization, amino groups in adenine, guanine, and cytosine do appear to be slightly out of plane of the conjugated system;

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989–1998

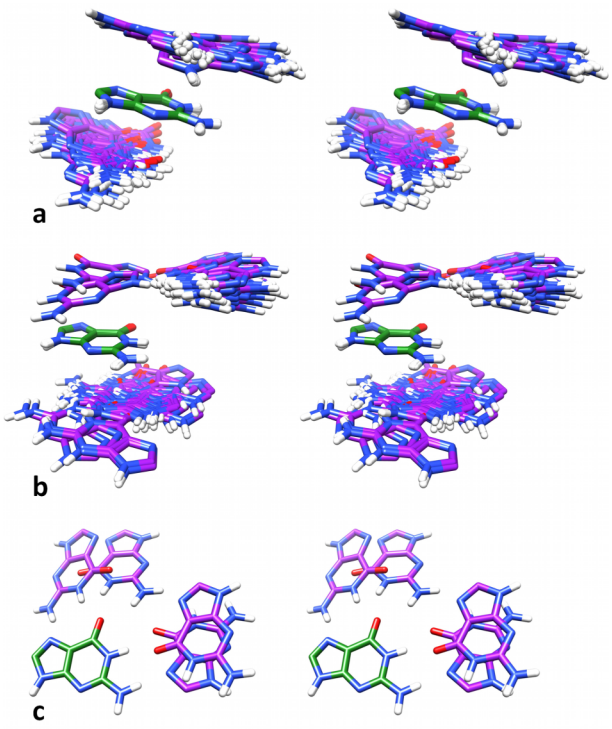


Figure 6. Example of the inter-ring arrangement pattern from the DiBaseRNA database. Guanine−guanine (GG) dibases are presented in their adjacent (a, ADJ), spatial (b, SPT), and hydrogen bonded (c, HBD) states. Some hydrogen atoms are hidden for clarity. For an explanation of this classification for the arrangements, please see the text.

however, a certain level of the out-of-plane displacement is not an artifact of the calculation and does represent the observations by both experimental and more sophisticated theoretical methods.[40,41]

The content and the number of structures, present in each variant of the database, are summarized in Table 1. The complete set of stereoviews representing the dibase arrangements in DiBaseRNA database is presented in Figures S1−S10, Supporting Information.

Interconversion between Pople and Haigh−Mallion Ring Current Geometric Factors. We observed a consistent correlation between the geometric factors of the Pople and Haigh−Mallion models at around the benzene 6-membered ring (Figure 6), if only the regions around the ring that are populated in usual biomolecular structures are accounted. This result indicates that simple equations can be devised for the interconversion of the Pople and Haigh−Mallion geometric factors. Such equations can be useful either for increasing the accuracy of the Pople model or for converting the Haigh−Mallion geometric factors into the Pople ones since the Haigh−Mallion model is the most implemented one in the existing chemical shift predictors,[7,8,27] whereas the Pople model is much simpler and convenient for implementing as restraints in molecular dynamics simulations or geometry optimization routines.

To develop the conversion equations, the ring current geometric factors (both the Pople and Haigh−Mallion ones) calculated for all the hydrogen atoms of all the dibase entries in DiBaseRNA database were used. Only the geometric factors originated from the neighboring ring in the dibase couple for each DiBaseRNA entry were accounted for. The 5- and 6-membered rings were considered separately.

The obtained factors thus account for the regions around the 5- and 6-membered rings that are populated in RNA structures. A simple mathematical model was found using the *Eureqa* automatic technique to search for hidden dependences in data.[42] The resulting equations are

$$G^{P,5}(\vec{\mathbf{r}}) = \frac{(1.000243G^{HM,5}(\vec{\mathbf{r}}) - 0.000146)}{(3.525744 - 5.873933G^{HM,5}(\vec{\mathbf{r}}))} \quad (5)$$

$$G^{P,6}(\vec{\mathbf{r}}) = \frac{(0.325964G^{HM,6}(\vec{\mathbf{r}}) - 0.000466)}{(1.743379 - 2.391111G^{HM,6}(\vec{\mathbf{r}}))} \quad (6)$$

where $G^{M,5}(\vec{\mathbf{r}})$ and $G^{M,6}(\vec{\mathbf{r}})$ are geometric factors for the ring current model M and 5- and 6-membered rings correspondingly. The $G^{P}(\vec{\mathbf{r}})$ versus $G^{HM}(\vec{\mathbf{r}})$ dependences determined by the eqs 5 and 6 are shown in Figure 7 as dotted lines.

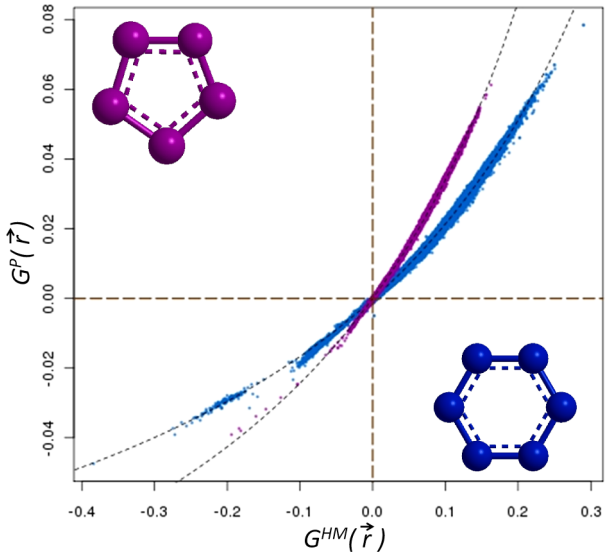


Figure 7. Correlation between the Pople and Haigh−Mallion ring current geometric factors for 5- (violet points) and 6-membered (dark blue points) rings in the RNA structures of the DiBaseRNA database. The correlations fitted by eqs 5 and 6 are shown as dotted lines.

Density Functional Theory Calculations of the Ring Current and Electric Field Effects on Nuclear Shielding Constants of Nucleic Acid Bases. We used the PBE1PBE[43]

Table 1. Number of Entries in the DiBaseRNA Database of the Dibase Arrangement As Observed in High-Resolution X-ray Structures of RNAs

base pairs	AA	AC	AG	AU	GC	GG	GU	CC	CU	UU
adjacent	37	39	64	24	79	114	55	72	39	23
H-bonded		3	6	38	95	4	9	5		
spatial	21	21	24	7	20	81	26	10	13	4

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989−1998

density functional theory,[30] which was proven to be particularly effective for studying the NMR shieldings of a wide range of nuclei, in many cases outperforming the results from the low-order perturbation studies, such as the ones using the correlated second order Møller–Plesset perturbation method.[43] The split-valence 6-311+G(2d,p) basis set[34] with the gauge-invariant atomic orbital (GIAO) method[44,45] was used because of its known good performance[46] for nuclear shielding calculations. All the calculations were done using the Gaussian03 suite of programs[22] with self-consistent field convergence criteria increased by the built-in *tight* mode.

The DiBaseRNA database was used to calculate (1) the ring current induced nuclear shielding changes and (2) the electric fields on a given base originated by its neighboring base. We also performed an additional test to verify that, in the DiBaseRNA configurations, the replacement of the ribose carbon attached to the conjugated ring by hydrogen atom does not significantly change the aromaticity of the base. In the explored set of adenine–methane test configurations, the proton-to-methyl group substitution at the adenine N9 position changed the chemical shift of all the tried positions of methane proton by at most 0.07 ppm (Figure S11, Supporting Information). This change, at least in part, can be attributed to a direct effect from the methyl substituent itself, rather than through the change in the core effects and the aromaticity of adenine.

For each of the entries in the DiBaseRNA database, three calculations were done. The first one was done for the complete pair of bases to obtain the nuclear shielding constant values of the system, the constituent atoms of which are already under the influence of ring currents and electric field effects (Figure 8a). The other two calculations were carried out for each of the two bases in isolation, where we kept the positions of the neighboring base by changing its atoms into dummy placeholders (Figure 8b,c), hence enabling further retrieval of the electric field values generated by the isolated base but acting on the exact positions where the atoms of the second base are located in the complete base pair. The electric field projections were taken along the approximate local symmetry axes at each of the atom locations as defined in Figure 5b.

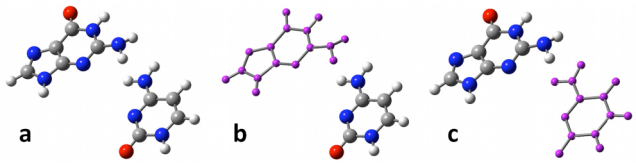


Figure 8. Example of a geometry breakdown for the three DFT calculations done for each entry of the DiBaseRNA database. The system a represents the complete pair of bases. One of the two bases is replaced by a dummy placeholder in systems b and c for the retrieval of the electric field values at the corresponding locations imposed by the retained base.

Influence of Structural Fluctuations on the ^{1}H, ^{15}N, ^{13}C, and ^{17}O Chemical Shifts of Nucleic Acid Bases. We at first performed nuclear shielding calculations on the variant 1 of the DiBaseRNA database, where the core structures come from the X-ray structures with 1.8 Å resolution or better. The hydrogen atom positions in this DiBaseRNA variant were optimized via semiempirical quantum chemistry in a planar constraint. Those structures were additionally geometry optimized, this time with B3LYP/6-31G(d,p) model chemistry, again allowing only the hydrogen atoms to change their position. Then, the nuclear shielding computations were performed as described above. Our results indicate that the small structural fluctuations of the core of the bases, which are in place for even the high-resolution X-ray structures, cause significant fluctuations of the ^{1}H, ^{15}N, ^{13}C, and ^{17}O chemical shifts because of the changes in the aromaticity of the conjugated ring. Histograms of the fluctuations for the N1, H1, C6, and O6 atoms in guanine base are shown in Figure 9 as examples, with the full set presented in Figures S12–S15, Supporting Information.

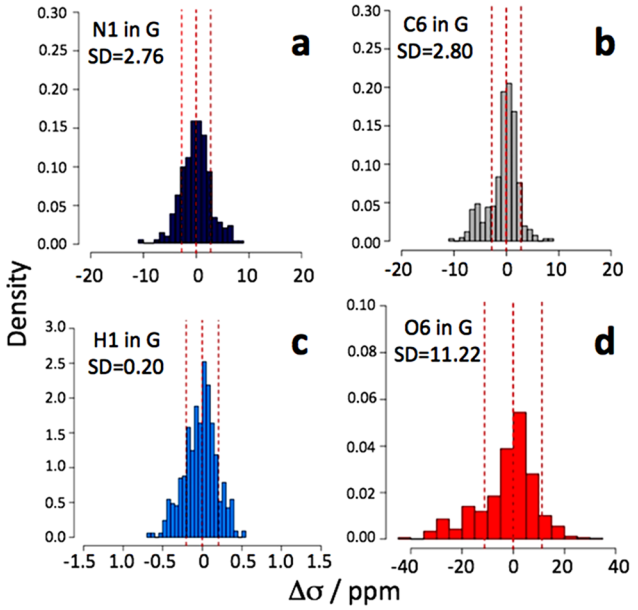


Figure 9. Histograms of the fluctuations of the nuclear shielding constants of the ^{15}N, ^{13}C, ^{1}H, and ^{17}O nuclei in guanine, shown as examples. The fluctuations are referenced by the median value of the nuclear shielding constant of each type. The standard deviations of the fluctuations are also shown.

Since the variation of nuclear shielding constants upon small structural fluctuations is greater than the changes caused by ring current or electric field effects, this result outlines the necessity of using a data set constructed via the fixed standard structures of individual nucleic acid bases, placed in the same dibase arrangement patterns (variant 2 of the DiBaseRNA database) as observed in the fragments extracted from the X-ray structure database. The real dynamical fluctuations of the bond lengths mostly occur around their equilibrium values with their influence, whether linear or nonlinear, averaged in the observed chemical shifts.

Hierarchy of Ring Current and Electric Field Effects for Hydrogen and Heavy Nuclei in RNA Bases. We used the calculated electric fields and changes in nuclear shielding constants imposed by the spatially neighboring base-rings (by going from Figure 8b or c to a) to perform a linear model fitting using various frameworks for deriving different models and associated coefficients. The fittings were performed in multiple schemes, applying both the Pople and Haigh–Mallion models with and without accounting the electric field effects.

The electric field effects were taken into account individually for each atom type. For a given atom *a* in a base *P*, the electric field value $F_{\parallel}^{a,N}$ imposed by the neighboring base *N* and acting on the position of the atom *a* along the direction indicated in

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989–1998

Figure 5b was incorporated in the fitting procedure. For hydrogen atoms, the field acting on the directly bonded heavy atom was taken into account (Figure 5b). Therefore, the electric field contribution to the change in absolute nuclear shielding constant of atom a can be expressed as[47]

$$\Delta\sigma_{\mathrm{ef}}^{a,P} = k_{\mathrm{ef}}^{a,N} F_{\parallel}^{a,N} \quad (7)$$

Hence, the fitting procedure with the inclusion of the electric field component optimizes additional $k_{\mathrm{ef}}^{a,N}$ coefficients, which for the atom a and the neighboring base N describe the linear nuclear shielding dependence on electric field[47] from the neighboring base.

The calculations were done for both hydrogen and non-hydrogen atoms, treating both the whole set of geometries in the DiBaseRNA database, and only concentrating on different classes of geometric arrangements of bases. Such combination of physical phenomena and geometric classes enabled us to investigate the hierarchy of different effects in explaining chemical shift changes for each type of base−base interaction. Examples are presented in Figure 10 for all types of ^{1}H and ^{13}C atoms, except the ones that directly take part in hydrogen bonding (i.e., excluding A, H, B, and C from all the molecular moieties capable of forming A−H···B−C hydrogen bonds). The complete set of the results is presented in Figures S16−S23, Supporting Information; the fitted parameters can be downloaded from the RingPar server. In the case of the joint treatment of ring current effects, we assumed the coefficient of the ring current geometric term from a given ring type to be the same for all the atoms of a given type (i.e., ^{1}H, ^{13}C, and ^{15}N). By contrast, in the case of the separate treatment of the ring current effects, the coefficients were assumed to be different for the individual atom types (for instance, individual ring current coefficients for adenine-C5, adenine-C6, guanine-C5, uracil-C2, etc., for ^{13}C nuclei), even if the neighboring ring was of the same type (for instance, adenine). Although physically the change in the local magnetic field value induced by the ring currents depend solely on the position of the query point O and the type of the conjugated ring, in cases in which the ring current geometric factor also covers electric and other types of interactions, the corresponding coefficients will also contain the response of the nuclear shielding constant of the query atom toward such interactions. Since this response will be dependent on the electronic environment for each atom type, by enabling individual coefficients for the ring current geometric term acting on each query atom type, we separately accounted the expected different responses of the query atoms toward the outer changes. Indeed, in the case of the separate treatment of the ring current effects, an improvement in the agreement between the conjugated ring induced nuclear shielding constants predicted from the fitted model and calculated via hybrid-DFT can be noted, which was even more apparent for non-hydrogen atoms.

We found that the aromaticity of RNA bases inferred from the fitted coefficients have a type of hierarchy noted before.[14] However, our coefficients are not comparable with previously published ones since they were parametrized to represent also the electric field effects. Overall, neglecting the dibase systems with hydrogen-bonded configurations, the overall effect of the neighboring conjugated rings on ^{1}H, ^{15}N, and ^{13}C nuclei in RNAs can be as high as 4.84, 34.49, and 27.18 ppm, respectively, with the average of the absolute changes in chemical shifts being 0.35, 1.75, and 2.60 ppm.

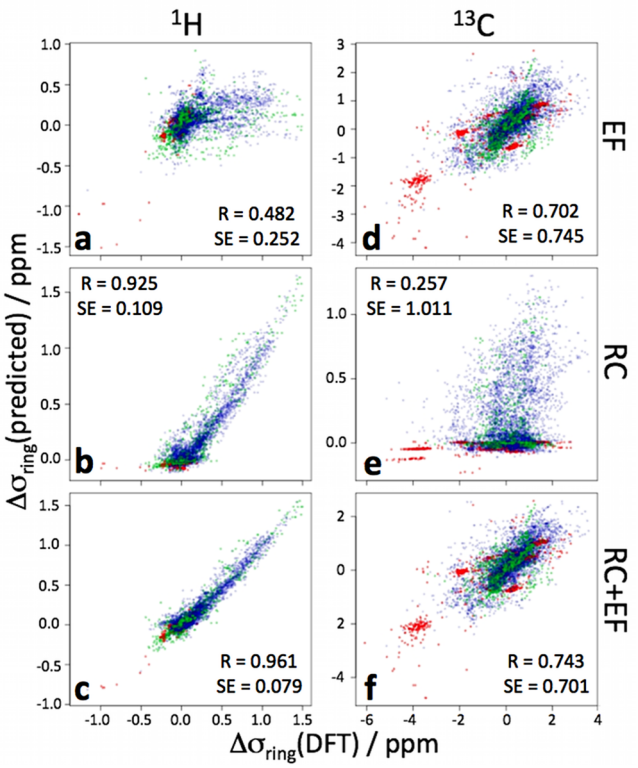


Figure 10. Correlations between the changes in nuclear shielding constants predicted by the phenomenological models and those calculated using the hybrid-DFT method. Results are shown for all the ^{1}H (a,b,c) and ^{13}C (d,e,f) nuclei that do not directly participate in hydrogen bonding, from all RNA bases. Three different cases are presented, using only electric field terms (eq 7, EF, a and d), only ring current terms (eq 1, RC, b and e), and both effects (RC + EF, c and f). Here, the ring current effects are accounted for via a joint treatment scheme, where, for a given ring type, the coefficients for the ring current geometric factor are assumed to be the same for all the atoms of single type (for all H, all C, all N, and all O atoms), regardless of their chemical state. Blue, green, and red points come from the inter-ring arrangements of the ADJ, SPT, and HBD classes in the DiBaseRNA database (see the text). The Pearson correlation coefficients and the standard errors of the predictions are shown on the plots. The complete set of the fitting results for all the nuclei and model variants is presented in Figures S16−S23, Supporting Information.

For ^{1}H nuclei, our results indicate that the conjugated ring-induced chemical shift alterations in the hydrogen-bonded complexes are well explained both by the electric-field-only treatment and by the ring current geometric factor. In general, it seems the ring current geometric terms can, in many cases, replace the required hydrogen-bonding geometric terms for nucleic acid chemical shift predictor development. This procedure will greatly simplify the chemical shift predictions for base atoms that are involved in hydrogen bonding, where the ring current geometric factor can simultaneously account for both ring current and hydrogen bonding effects via a single joint coefficient. It is also clear that the SPT and ADJ arrangements of the conjugated rings affect the ^{1}H chemical shifts via mostly the ring current effect.

Although conjugated ring effects on ^{13}C nuclei are almost always explained by electric field effects, regardless of the inter-ring arrangement type, it is interesting to note that the ring current geometric terms can capture electric field effects if the

dx.doi.org/10.1021/jp3057306 | J. Phys. Chem. B 2013, 117, 1989−1998

ring current coefficients are calculated separately for each carbon atom type. Similar results are found for ^{15}N nuclei, but not for ^{17}O, for which ring current geometric terms or their addition to the electric field terms improve only the description of the ^{17}O chemical shifts that take part in hydrogen bonding. For the other (SPT, ADJ, and their combination) inter-ring arrangements, the electric field term is the dominant factor affecting ^{17}O nuclear shielding constants. ^{17}O nuclei possess a remarkably high sensitivity to electric field effects, which can change the chemical shift values by up to 80 ppm.

The use of the Haigh−Mallion geometric term, instead of the Pople model, has almost always slightly improved the quality of the linear models, however, with the difference still being not essential.

CONCLUSIONS

We have performed an extensive analysis of the Pople model of ring current effects against the more accurate Haigh−Mallion model. We have thus proposed general equations for 5- and 6-membered rings to convert the geometric factors of one model into those of the other. These results offer an opportunity either to switch to the Pople model when a chemical shift predictor uses the Haigh−Mallion model or to increase the accuracy of the Pople model while maintaining its simplicity.

We have then generated an atlas of dibase arrangements observed from high-resolution RNA structures, on which hybrid-DFT calculations were done to estimate the change in local electric fields and nuclear shielding constants in the vicinity of conjugated rings. Besides the widely studied ^{1}H nuclei, we have also considered ^{13}C, ^{15}N, and ^{17}O nuclei.

We have further used linear model fitting to derive ring current and electric field based models for explaining the chemical shift changes induced by neighboring conjugated rings. This combined fitting procedure has enabled us to assess the hierarchy of ring current and electric field effects on both hydrogen and non-hydrogen nuclei. In particular, we have demonstrated that the chemical shift changes of non-hydrogen atoms are mostly determined by electric field, rather than ring current, effects. However, we found that hydrogen-bonding-induced electric field effects are very well captured by the geometric factors of ring current models, a property that will be useful to account for modeling hydrogen bonding effects on chemical shifts in nucleic acid bases.

Finally, we have set up a web server, RingPar, which enables users to extract the fitting coefficients and resulting correlation plots, after defining the type of fitting to be done and the structural classes to be used. Besides nuclear shielding constants, RingPar also reports models and coefficients for the nuclear shielding anisotropies, useful for future developments of models for chemical shift anisotropies.

In summary, we have presented a study of the effects of ring currents and electric fields on the chemical shifts of RNA bases. We anticipate that, by including the other factors that influence RNA chemical shifts, it will be possible to develop accurate chemical shift predictors that will work for both hydrogen and non-hydrogen nuclei.

ASSOCIATED CONTENT

Supporting Information

Ten figures of inter-ring arrangement patterns in DiBaseRNA database for all combinations of RNA base-types, one figure presenting the results for testing the effect of the reduction of the ribose substituent to the RNA base up to a hydrogen atom on the aromaticity of the base, four figures with the histograms of chemical shift fluctuations upon small structural changes in crystallographic structures of RNA bases, and eight figures with all the plots from various model fittings to represent the DFT calculated nuclear shielding constants. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: mv245@cam.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Herchel Smith Fund, Darwin College, Schlumberger Research (to A.B.S.), and the BBSRC (to M.V.). We thank Professor Roger Mallion for a helpful discussion.

REFERENCES

(1) Case, D. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 624−630.
(2) Oldfield, E. *Annu. Rev. Phys. Chem.* **2002**, *53*, 349−378.
(3) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615−9620.
(4) Shen, Y.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685−4690.
(5) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. *J. Am. Chem. Soc.* **2009**, *131*, 13894−13895.
(6) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. *Structure* **2010**, *18*, 923−933.
(7) Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M. *J. Biomol. NMR* **2011**, *50*, 331−346.
(8) Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 9620−9623.
(9) Wijmenga, S. S.; Kruithof, M.; Hilbers, C. W. *J. Biomol. NMR* **1997**, *10*, 337−350.
(10) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287−387.
(11) Cromsigt, J. A. M. T. C.; Hilbers, C. W.; Wijmenga, S. S. *J. Biomol. NMR* **2001**, *21*, 11−19.
(12) Lam, S. L.; Chi, L. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56*, 289−310.
(13) Ösapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444.
(14) Case, D. A. *J. Biomol. NMR* **1995**, *6*, 341−346.
(15) Haigh, C. W.; Mallion, R. B. *Org. Magn. Reson* **1972**, *4*, 203−228.
(16) Haigh, C. W.; Mallion, R. B. *Prog. Nucl. Magn. Reson. Spectrosc.* **1980**, *13*, 303−344.
(17) Waugh, J. S.; Fessenden, R. W. *J. Am. Chem. Soc.* **1957**, *79*, 846−849.
(18) Johnson, C. E.; Bovey, F. A. *J. Chem. Phys.* **1958**, *29*, 1012−1014.
(19) Pople, J. A. *J. Chem. Phys.* **1956**, *24*, 1111−1111.
(20) Murray, L. J. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13904−13909.
(21) Camilloni, C.; Robustelli, P.; De Simone, A.; Cavalli, A.; Vendruscolo, M. *J. Am. Chem. Soc.* **2012**, *134*, 3968−3971.
(22) Frisch, M. J.; et al. *Gaussian 03*, revision E.01; Gaussian Inc.: Wallingford, CT, 2004.
(23) R Development Core Team. *R: a language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011.
(24) Newton, R.; Wernisch, L. *R News* **2007**, *7*, 32−35.
(25) Pauling, L. *J. Chem. Phys.* **1936**, *4*, 673−677.
(26) Lonsdale, K. *Proc. R. Soc. A* **1937**, *159*, 149−161.
(27) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215−240.

(28) Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702−709.
(29) Haigh, C. W.; Mallion, R. B. *Mol. Phys.* **1971**, *22*, 955−970.
(30) Kohn, W.; Sham, L. J. *Phys. Rev. A* **1965**, *140*, 1133−1138.
(31) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(32) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200−206.
(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(34) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.
(35) Tamagawa, K.; Iijima, T.; Kimura, M. *J. Mol. Struct.* **1976**, *30*, 243−253.
(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(37) Stewart, J. J. P. *MOPAC2009*; Stewart Computational Chemistry: Colorado Springs, CO, 2008.
(38) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(39) Yatsenko, A. V.; Paseshnichenko, K. A. *J. Mol. Struct.* **1999**, *492*, 277−283.
(40) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161−3164.
(41) Sychrovsky, V.; Foldynova-Trantirkova, S.; Spackova, N.; Robeyns, K.; Meervelt, L. V.; Blankenfeldt, W.; Vokacova, Z.; Sponer, J.; Trantirek, L. *Nucleic Acids Res.* **2009**, *37*, 7321−7331.
(42) Schmidt, M.; Lipson, H. *Science* **2009**, *324*, 81−85.
(43) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113−119.
(44) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789−807.
(45) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251−8260.
(46) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497−5509.
(47) Buckingham, A. D. *Can. J. Chem.* **1960**, *38*, 300−307.