A geometrical parametrization of C1′-C5′ RNA ribose chemical shifts calculated by density functional theory

Reynier Suardíaz,[a] Aleksandr B. Sahakyan, and Michele Vendruscolo[b]
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

(Received 17 February 2013; accepted 5 June 2013; published online 15 July 2013)

It has been recently shown that NMR chemical shifts can be used to determine the structures of proteins. In order to begin to extend this type of approach to nucleic acids, we present an equation that relates the structural parameters and the ^{13}C chemical shifts of the ribose group. The parameters in the equation were determined by maximizing the agreement between the DFT-derived chemical shifts and those predicted through the equation for a database of ribose structures. Our results indicate that this type of approach represents a promising way of establishing quantitative and computationally efficient analytical relationships between chemical shifts and structural parameters in nucleic acids.
 [http://dx.doi.org/10.1063/1.4811498]

INTRODUCTION

The functional importance of ribonucleic acids (RNAs) is increasingly recognized in molecular biology.[1–5] In addition to their key role in the transmission and control of genetic information, RNAs constitute fundamental structural components of ribosomes and other macromolecular assemblies, and catalyze many biochemical reactions.[1–5] The determination of the conformations of RNAs at atomic level is therefore essential to reveal the structural basis of the biological functions of these important molecules. In this respect, together with X-ray crystallography, one of the most useful techniques is nuclear magnetic resonance (NMR) spectroscopy, as indicated by the fact that nearly half of the RNA structures deposited in Protein Data Bank (PDB)[6] have been solved by NMR methods.[7]

Macromolecular structure determination by NMR spectroscopy is primarily based on the use of inter-proton distances, derived from nuclear Overhauser effects, and dihedral angles, derived from scalar couplings.[8] More recently, information provided by residual dipolar couplings has also been used as a source of structural constraints.[9–11] By contrast, chemical shifts, which are the most readily and accurately measurable NMR parameters, have only recently been exploited in NMR studies of macromolecules, because of the complications in defining their relationship with structural parameters.[12–14] Approaches based on the use of chemical shifts, at least in principle, can be very powerful since the values of these parameters depend in an extremely sensitive manner on the molecular conformation.[12–18] The use of chemical shifts for structure determination would be particularly advantageous in the case of RNAs, because of the great challenges that these molecules present for NMR spectroscopy,[19–28] as the proton density is generally lower in these molecules than in proteins, and hence fewer structural restraints from inter-proton distances can usually be obtained.

As any method aimed at elucidating the structural properties of biological macromolecules using the information provided by chemical shifts requires the possibility of calculating chemical shifts from a given structure, it is essential to establish such procedures for RNAs. In this respect, phenomenological approaches for the calculation of chemical shifts have been very successful for proteins,[27,29–34] and they have also been proposed for nucleic acids.[35–38] Compared to proteins, however, the available structures of RNA molecules have been determined at lower resolution and they are fewer in numbers. Therefore, it has been particularly difficult to reveal the connection between structural parameters and chemical shifts based on the information contained in currently existing structural datasets. Establishing firmly such a connection would be particularly timely, as suggested by a recent study in which proton chemical shifts were used to facilitate the calculation of RNA structures,[39] thereby indicating that, upon further development of structure-based chemical shift predictors, approaches based on the use of chemical shifts can lead to the determination of the high-resolution structures of RNA molecules.

Our goal in this work is to investigate a particular aspect of the relationship between chemical shifts and structural parameters in RNAs by focusing on the ^{13}C chemical shifts of the ribose ring. This ring occurs in two types of conformations, usually referred to as S (or C2′-endo) and N (or C3′-endo) forms,[40] which can be described in terms of two parameters, the angle of pseudorotation, P, and the degree of pucker, ν_{max}.[40] Our work builds upon previous studies that have been concerned with the relationship between these ribose puckering parameters and the C1′-C5′ chemical shifts. Dejaegere and Case have studied the variation of these chemical shifts with P by density functional theory (DFT) calculations and have reported an upfield shift in the C3′ and C5′ chemical shifts when the ribose is in the N state.[41] Xu *et al.* reported that ribose groups experience more pronounced sugar puckering effects than deoxyribose groups[42] and identified different

[a] Present address: Departament de Química and Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
[b] Author to whom correspondence should be addressed. Electronic mail: mv245@cam.ac.uk

behaviors depending on the backbone γ angle. They described a downfield shift for the C3′ chemical shifts when the ribose is in the S state and the γ angle is in the gauche conformation.[43] Ebrahimi *et al.* developed a method that uses solid-state chemical shifts of C1′, C4′, and C5′ to define two "canonical" coordinates capable of discriminating between N and S ribose conformations.[44] Subsequently, Ohlenschläger *et al.*[45] improved that method using solution NMR data of RNAs deposited in the PDB[7] and BMRB[46] databases. More recently, an update of the expression for the canonical coordinate has been proposed by Cherepanov *et al.*[47] All these approaches have been concerned with the classification of the ribose group in terms of the two major conformations (N and S) rather than with a detailed link between the ribose conformation and chemical shifts.

Here we present an equation that relates ^{13}C chemical shifts with the P and ν_{max} structural parameters of the RNA ribose group based on quantum mechanical calculations. In order to find the functional mapping between structures and chemical shifts in the absence of enough high-resolution RNA structures, we use quantum chemistry methods to calculate chemical shifts on a given database of structures. The success of this type of approach, especially by DFT, has been shown to be remarkable.[48] These calculations have been extensively applied to biological systems, including proteins and nucleic acids.[36,49–57] DFT calculations have been proven to be very useful in rationalizing the results of chemical shift anisotropy measurements[48] and for the direct evaluation of ^{13}C chemical shifts in ribose groups.[35] The ShiftS method,[11] which is in part based on such calculations, enables accurate structure-based calculations of chemical shifts in proteins and is among the most accurate chemical shift predictors currently available. By following this strategy, the parameters in the equations presented here are derived from quantum mechanical calculations of chemical shifts on a number of structures making up a database sufficiently large to establish a quantitative relationship between chemical shifts and structural parameters in ribose groups.

METHODS

Calculation of the ribose chemical shifts

In order to calculate the chemical shifts corresponding to given ribose structures, we used a DFT approach. The ribose moiety chosen to study the influence of the sugar pucker conformation on the ribose ^{13}C chemical shifts is illustrated in Figure 1 as a part of an RNA molecule. For the geometry optimization we used the B3LYP hybrid functional[58,59] with the 6-311G(d,p) basis set, and, for the chemical shielding calculations, the PBE0 (or PBE1PBE)[60–62] functional in combination with the 6-311+G(2d,p) basis set.[35] This level of theory is generally deemed to be sufficiently accurate to yield good predictions of ^{13}C chemical shifts.[63–65] We performed a benchmark for this procedure using a set of 12 small molecules (Table S1 and Figure S1 in the supplementary material[66]). These molecules were either in a single conformation or they evenly populated different conformations; in this latter case, nuclear shielding constants, denoted as σ, were averaged across different conformations. For comparison, experimental chemical shifts were taken from previously published studies.[44,67–72] A linear regression resulted in a coefficient of correlation of 0.996 between experimental and calculated chemical shifts (Figure S1 in the supplementary material[66]). Ideally, the linear regression should give a -1 slope and an intercept equal to the isotropic nuclear shielding constant of the reference used for experimental chemical shifts, in this case tetramethylsilane (TMS), calculated at the same level of theory ($\sigma_{\mathrm{TMS}} = 187.76$ ppm). As we found a slope of -1.044 and an intercept of 187.52 ppm, the predicted nuclear shielding constants at this level of theory can be considered as accurate. Results of this benchmark are very similar to those recently reported for a more general ^{13}C dataset.[57,73]

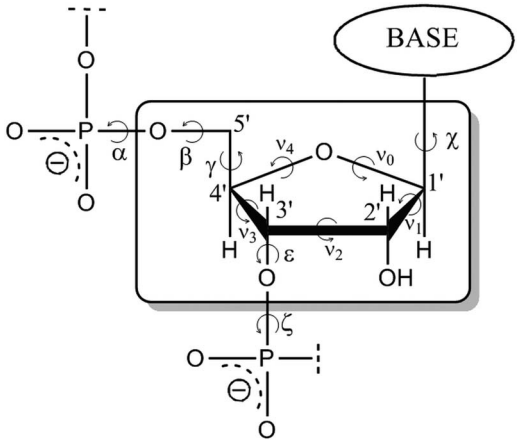

FIG. 1. Schematic illustration of the structural model of the ribose used in the quantum mechanical calculations described in this work, and its relation to the overall RNA structure. The model is enclosed in the square, without the hydrogen caps.

Generation of a structural database and corresponding chemical shifts

A set of 500 ribose conformations was generated with random values of the P and ν_{max} parameters by using the following procedure:

(i) 1000 sets of P and ν_{max} values were randomly generated via Eqs. (1) and (2) (see Figure 1 for the definition of the dihedral angles),

$$P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2\nu_2(\sin(36^\circ) + \sin(72^\circ))}, \quad (1)$$

$$\nu_{\max} = \frac{\nu_2}{\cos(P)}, \quad (2)$$

with a uniform distribution in the range from 0° to 360° for P and from 0° to 60° for ν_{max};

(ii) O4′-C1′-C2′-C3′ (ν_1) and C1′-C2′-C3′-C4′ (ν_2) dihedral angles (Figure 1) were calculated using the equations

$$\nu_1 = 1.0009\nu_{\max}\cos\left(\frac{(P - 0.90 + 16)\pi}{5}\right), \quad (3)$$

$$\nu_2 = 0.9770\nu_{\max}\cos(P - 0.48), \quad (4)$$

which were taken from de Leeuw *et al.*;[74]

(iii) 1000 structures were then generated with the dihedral angles resulting from Eqs. (3) and (4);

(iv) these structures were further geometry optimized by keeping frozen the ν_1 and ν_2 dihedral angles and optimizing the remaining degrees of freedom within the DFT framework described above.

These two dihedral angles quite accurately define the overall conformation of the ribose five-membered ring. Since during the partial geometry optimization (step iv), bond lengths and bond angles can undergo small changes, we recalculated P and ν_{max} after geometry optimizations to refine the values and verify that the range defined initially was maintained. This procedure was applied to three versions of the ribose model, each with a different value of the exocyclic dihedral angle γ (O5′-C5′-C4′-C3′, Figure 1), in order to take into account its effects. These three values were set to −60°, 60°, and 180° in the geometry optimizations, which are the values that approximately correspond to the most populated γ conformers in RNA structures.[75] Since some of the partial optimizations did not reach convergence, the number of converged structures was different for each γ conformer. We therefore considered 500 converged structures of each γ conformation to have the same number in all cases. Nuclear shielding constants were calculated for all NMR-active isotopes of elements in the ribose model (^{1}H, ^{13}C, ^{15}N, and ^{17}O) over those optimized geometries. This procedure provided a dataset of 1500 chemical shifts for each atom of the model (Figure 1) along with the corresponding P and ν_{max} puckering parameters. The full data are provided in the supplementary material.[66]

Definition of a relationship between structures and chemical shifts

We used the Eureqa software[76] to detect the relationship between ^{13}C chemical shifts and ring pucker parameters. For this purpose, we analyzed the conformations and corresponding DFT-calculated chemical shifts described above. The Eureqa software uses a symbolic regression method that optimizes simultaneously the parameters and the analytical form of the regression equation in order to identify mathematical relationships in a dataset. The search for an accurate regression equation was performed using the mean absolute error as the optimization score. The addition, subtraction, multiplication, multiplication by a constant, division, exponentiation, logarithm, sine, and cosine terms were considered in the equation space sampling procedure. At each step in the procedure, the new equations are formed by stochastically altering their constituent terms, which are then stored or discarded, depending on their optimization score. In order to validate the performance of the equation resulting from this optimization procedure, we followed the standard criterion adopted by the Eureqa software, which consists in splitting the available data into a training set (75% of the data) and a validation set (25% of the data) and in checking that the statistical error is similar in the two sets to avoid overfitting.

RESULTS AND DISCUSSION

A geometrical parametrization of the ^{13}C chemical shifts in ribose

The main result presented in this work is an equation that relates the nuclear shielding constants of the ^{13}C atoms of the ribose group to the P and ν_{max} puckering parameters

$$\sigma_\gamma = a + b\nu_{max} + \sum_{n=1}^{4} c_n \cos(d_n + e_n P + f_n \nu_{max}) + \sum_{n=1}^{5} g_n \nu_{max} \cos(h_n + i_n P + j_n \nu_{max}), \quad (5)$$

where the coefficients a to j_n are fitted to maximize the agreement with the nuclear shielding constants in the structural database that we used here. The parameters P and ν_{max} should be taken in the range 0°–360° and 0° –60°, respectively (see Methods section). Equation (5) reproduces in a computationally efficient manner the nuclear shielding constants calculated by the PBE0 (or PBE1PBE)[60–62] functional with the 6-311+G(2d,p) basis set used in the model development. With the nuclear shielding constant of the reference compound (TMS) calculated at the same level of theory ($\sigma_{\mathrm{TMS}} = 187.76$ ppm), one can also obtain the chemical shifts of the C1′-C5′ atoms as $\delta = \sigma_{\mathrm{TMS}} - \sigma(P, \nu_{max})$.

Eq. (5) can be considered as a generalized procedure for expressing DFT-obtained ribose NMR chemical shifts as a function of structural parameters. This approach is similar in spirit to that used for fitting coupling constants to dihedral angles (the well known Karplus equations[77]), which gives equations of the type

$$^3J = A\cos^2\theta + B\cos\theta + C, \quad (6)$$

where θ is a dihedral angle. This equation is equivalent to

$$^3J = C_1 + C_2\cos\theta + C_3\cos 2\theta, \quad (7)$$

which can be considered as the real part of a Fourier series truncated at $n = 2$. Equation (5) thus represents a functional expansion, which represents in a compact form the 15 equations that we consider in our work, as we have five atom types and three equations for atom type depending on the rotameric state of the dihedral angle γ. Although, in principle, Eq. (5) has 38 coefficients, most of them are zero (Tables I–III), so the effective number of parameters in each individual equation is relatively small. For instance, the equation for C1′ when $\gamma = 180°$ is

$$\sigma_{C1'(\gamma=180)} = 105 + 10.6\nu_{max} + \cos(2.14\nu_{max}) + 1.38\cos(\pi/2 - 5.19P) + 0.5\cos(\pi/2 - 8.77P + 1.05\nu_{max}) + 0.5\cos(-\pi/2 + 8.77P + 3.23\nu_{max}). \quad (8)$$

As another example, Eq. (5) for C2′ reduces to

$$\sigma_{C2'(\gamma=180)} = 112 + 1.56\cos(4.04 - 4.18P - \nu_{\max}) - \nu_{\max}\cos(1.93P). \quad (9)$$

In this case, only a, c_1–f_1, g_1, and i_1 are not zero. In the case of C3′, Eq. (5) acquires a particularly simple form when

TABLE I. Parameters for Eq. (5) when $\gamma \approx 60^\circ$.

	C1′	C2′	C3′	C4′	C5′
a	101.81	110.68	111.85	93.29	119.36
b	13.60	1.00	0.24	0.00	2.54
c_1	1.00	1.26	− 1.65	7.16	0.36
d_1	$\pi/2$	0.10	− 2.31	$\pi/2$	
e_1	− 1.00	− 1.00	1.00		1.00
f_1	0.49	− 1.93	− 2.27	− 2.00	
c_2	− 2.04	1.26	− 1.65	1.00	
d_2	$\pi/2$	0.90	− 5.45	− 1.39	
e_2	− 1.58	− 1.00	1.00	2.09	
f_2		− 5.11	2.27		
c_3		− 0.25			
d_3		1.00			
e_3		2.89			
f_3		− 5.57			
c_4		− 0.25			
d_4		− 1.00			
e_4		2.89			
f_4		5.57			
g_1		− 0.25	− 2.22	2.27	1.54
h_1		1.00		$\pi/2$	$\pi/2$
i_1		2.89	2.14	− 0.86	− 2.26
j_1		1.47		− 2.56	
g_2		− 0.25		2.27	4.40
h_2		− 1.00		$-\pi/2$	1.21
i_2		2.89		− 0.86	− 1.00
j_2		− 1.47		2.56	1.00
g_3					1.00

$\gamma = 180^\circ$

$$\sigma_{C3'(\gamma=180)} = 110 + \cos(-2.19\nu_{\max}). \quad (10)$$

It should also be noted that it is rather common to have a relatively large number of parameters in empirical expressions used in the literature for predicting chemical shift values from the structures of macromolecules.[2–6,8–11,35,36] For this reason it is important to the obtained analytical expressions against overfitting. The major advantage of developing parametrized empirical models to account for the conformational dependence of chemical shifts is that these models can provide prediction results in milliseconds of CPU time, i.e., several orders of magnitude faster than typical *ab initio* or DFT calculations.

The coefficients (a to j_n) of Eq. (5) for each carbon nucleus type are reported in Tables I–III. The chemical shifts calculated using DFT methods (Figure 2, top row) and those predicted using the geometrical parametrization derived in this work (Eq. (5)) with the coefficients of Tables I–III (Figure 2, bottom row) are in good agreement (Table IV). The root mean square deviations (RMSDs) between DFT and Eq. (5)-based chemical shifts range between 0.17 and 0.38 ppm (Table IV). For comparison, in proteins the typical deviations between experimental and calculated chemical shifts for carbon atoms are three- to fivefold larger,[4–6,8–11] suggesting that the geometrical parametrization in Eq. (5) captures very well the conformational dependence of the chemical shifts of the ribose carbon atoms on the internal coordinates of the ribose itself.

TABLE II. Parameters for Eq. (5) when $\gamma \approx -60^\circ$.

	C1′	C2′	C3′	C4′	C5′
a	107.62	114.15	110.04	96.26	121.15
b		− 1.81	1.00	3.29	1.00
c_1	− 4.34	− 1.00	− 0.53	1.64	0.54
d_1	0.29	$\pi/2$	56.26	1.93	− 0.79
e_1	3.00	− 0.66	− 1.00	− 1.00	2.16
f_1		− 3.66	3.02	− 2.17	
c_2		1.00	− 2.35	− 1.64	
d_2		$\pi/2$		1.93	
e_2		− 0.66	− 0.17	− 1.00	
f_2		3.66	1.00	2.17	
c_3		− 0.34	− 2.35	1.00	
d_3		$\pi/2$	− 1.15	− 2.35	
e_3		2.12	2.12	2.00	
f_3		− 3.66	− 3.66	2.17	
c_4		0.34		1.00	
d_4		$\pi/2$		− 2.35	
e_4		2.12		2.00	
f_4		3.66		− 2.17	
g_1	1/2		− 5.57		3.62
h_1	0.20				60.64
i_1	3.03		1.00		− 1.00
j_1					− 1.00
g_2	1/2				− 3.62
h_2	− 0.78				− 57.50
i_2	− 1.03				1.00
j_2					− 1.00
g_3	4.34				− 3.42
h_3	0.49				− 0.79
i_3	2.03				2.16
j_3					
g_4	−1/2				
h_4	0.49				
i_4	4.06				
j_4					
g_5	− 2.52				
h_5	− 0.29				
i_5	1.00				

Table IV also reports values from a further validation test in which we generated 250 additional geometries and their corresponding DFT-calculated chemical shifts following the same procedure used above. Equation (5) and the P, ν_{max} values of these conformers were then used to calculate a new set of predicted chemical shift values. The errors in both training and validation datasets are comparable.

Application to small nucleotides and nucleosides in the solid state

The geometrical parametrization described above (Eq. (5)) was derived in order to find out whether the chemical shifts calculated by DFT can be approximated using an expression dependent only on the coordinates of the ribose nuclei. In the following, however, we explored whether this geometrical parametrization could be useful as a starting point in the prediction of experimentally measured C1′-C5′ chemical shifts in RNAs. For this purpose we considered

TABLE III. Parameters for Eq. (5) when $\gamma \approx 180°$.

	C1′	C2′	C3′	C4′	C5′
a	105.00	112.00	110.00	90.30	117.16
b	10.60				2.56
c_1	1.00	1.56	1.00	9.82	
d_1		4.04		$\pi/2$	
e_1		−4.18		−1.75	
f_1	2.14	−1.00	−2.19		
c_2	1.38				
d2	$\pi/2$				
e2	−5.19				
f2	1.00				
c3	0.50				
d3	$\pi/2$				
e3	−8.77				
f3	1.05				
c4	0.50				
d4	$-\pi/2$				
e4	8.77				
f4	3.23				
g1		−1.00		2.63	2.23
h1				$\pi/2$	$3\pi/2$
i1		1.93		−2.27	−2.15
j1					−1.00
g2				6.18	2.23
h2				0.45	$-\pi/2$
i2				−0.89	2.15
j2					−1.00
g3					4.46
h3					$\pi/2$
j3					−1.00
g4					−3.64
h4					$\pi/2$
i4					−1.06

TABLE IV. Root mean square deviations (RMSDs, in ppms) between the chemical shifts calculated by DFT and by Eq. (5). Results from both training and validation data sets are presented.

Atom type	C1′	C2′	C3′	C4′	C5′
RMSD (ppm) in training data set	0.200	0.295	0.169	0.396	0.175
RMSD (ppm) in validation data set	0.188	0.305	0.166	0.386	0.179

a set of ten molecules (Figure S2 and Table S2 in the supplementary material[66]), consisting of eight nucleosides (cytidine, 5-methyluridine hemihydrate, 2-thiocytidine dihydrate, 5-bromouridine, 3-deazauridine, 5-hydroxyuridine, 8-bromoadenosine, and xanthosine dehydrate) and two nucleotides (adenosine-3′-phosphate dihydrate and cytidine-3′-phosphate). The crystal structures of these molecules[44] were obtained from the Cambridge Crystallographic Data Center (CCDC).[78] Using the crystal geometries, the P and ν_{max} values were calculated for each molecule. Taking into account the γ dihedral angle in the corresponding crystal structure, the appropriate coefficients for Eq. (5) were selected to calculate chemical shifts. In the case of C1′, a shift of approximately 15 ppm was applied to account for the fact that this atom is bonded to a nitrogen atom in the selected molecules, which is missing in the model used to obtain Eq. (5). The 15 ppm shift is used to reproduce approximately the difference between the average chemical shift of C1′ in the BMRB (91 ppm) and the 73.7 ppm chemical shift of the corresponding carbon in tetrahydrofuran-2,3,4-triol.[79] A linear regression between predicted and experimental chemical shifts resulted in a correlation coefficient of 0.94 (Figure 3). However, the comparison between predicted and experimental chemical shifts for individual carbon atoms is not very good (Figures 4(a)–4(e)). To check whether the poor correlations observed for individual carbon atoms are related to the use of Eq. (5), we calculated the same chemical shifts using DFT. The agreement between experimental and DFT-calculated chemical shifts (Figures 4(f)–4(j)) is similar to that obtained for the chemical shifts predicted using Eq. (5). The same level of agreement is obtained between DFT-calculated chemical shifts and those predicted by Eq. (5) (Figures 4(k)–4(o)). These results show between experimental and calculated chemical shifts arise because of the factors other than the ribose internal geometry (Figure 1), which affect the ribose carbon chemical shifts in

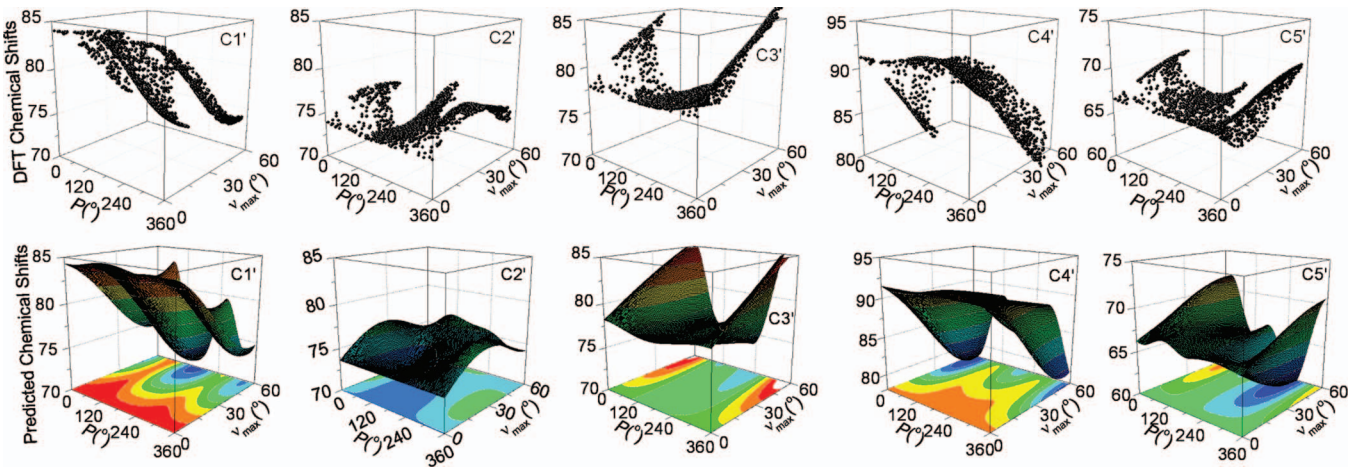

FIG. 2. Comparison between the chemical shifts calculated by DFT (point surfaces, top row) and predicted using Eq. (5) (continuous surfaces, bottom row) for the C1′-C5′ atoms in ribose (in ppm).

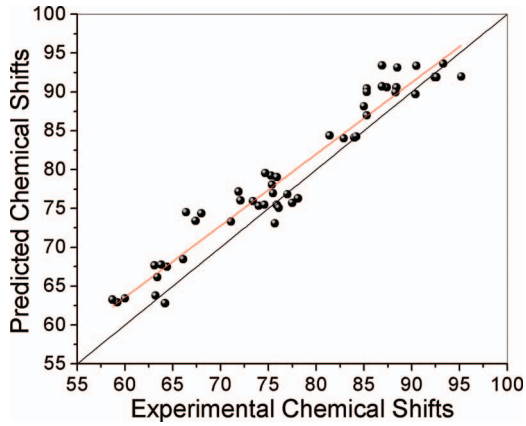

FIG. 3. Comparison between predicted and experimental C1′-C5′ chemical shifts (in ppm) for a set of eight nucleosides and two nucleotides. The ideal behavior (black line) is compared with that of the predictions (red line), which has a coefficient of correlation of 0.94.

the system that we considered, such as ring current effects from the nitrogen bases of spatially close rings from the other constituents of the crystal etc. The relevance of this type of effects has been extensively studied in the literature.[80,81] In this respect, the results are poorer for C1′ and C2′ chemical shifts as these atoms are very close to the bases and are appreciably affected by ring current effects.

Application to a small RNA molecule in solution

In order to further assess whether the geometrical parametrization presented in this work (Eq. (5)) could represent a useful starting point for deriving a structure-based predictor of RNA chemical shifts, we applied this expression to two small RNA hairpin molecules containing a UUCG tetraloop. The first one is a 14-mer RNA hairpin whose structure was solved by NMR at high resolution in the solution[82,83] and solid[47] states. The experimentally measured chemical shifts were available and the values of P and ν_{max} were calculated from the corresponding structure in Protein Data Bank (PDB ID: 2KOC) (Figure 5(a)). The second molecule is a

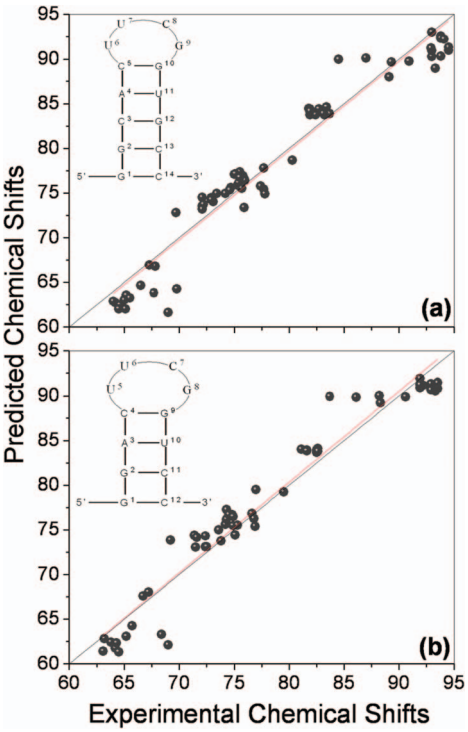

FIG. 5. Comparison between chemical shifts predicted using Eq. (5) and experimental ones (in ppm) for two small RNA hairpins in solution. The ideal behavior (black line) is compared with the predicted one (red line). (a) 14-mer GGCAC(UUCG)GUGCC, coefficient of correlation 0.99, slope 1.02, intercept −0.93. (b) 12-mer GGAC(UUCG)GUCC, coefficient of correlation 0.98, slope 1.00, intercept −0.28).

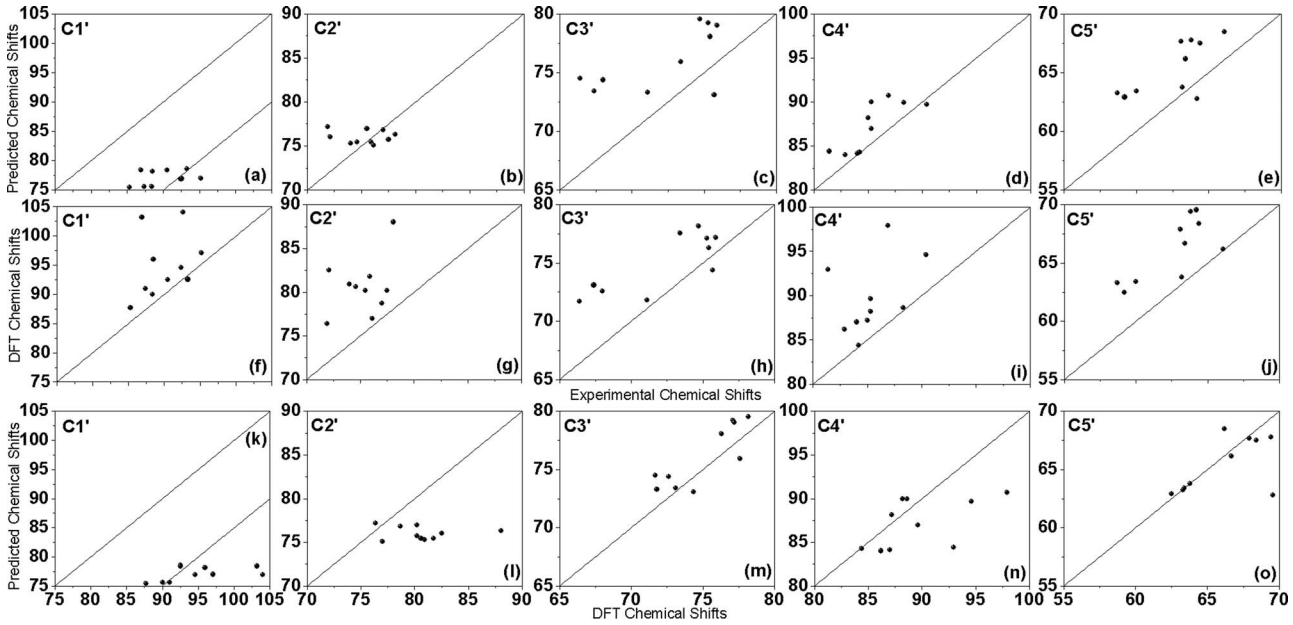

FIG. 4. Comparison between predicted, DFT-calculated and experimental C1′-C5′ chemical shifts for the set of eight nucleosides and two nucleotides. (a)–(e) Predicted chemical shifts using Eq. (5) against the experimental chemical shifts; (f)–(j) DFT-calculated chemical shifts against the experimental chemical shifts; and (k)–(o) predicted chemical shifts using Eq. (5) against the DFT-calculated chemical shifts.

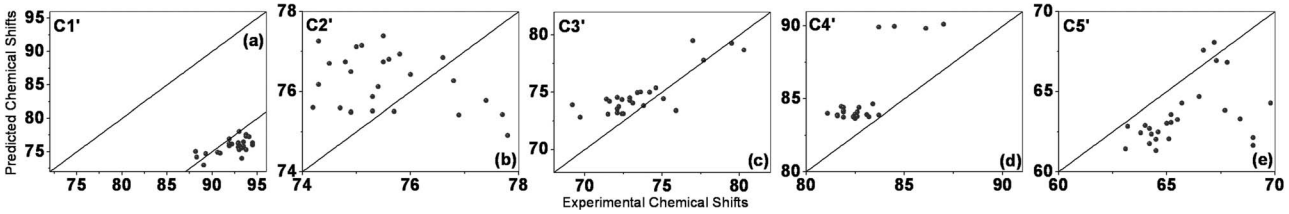

FIG. 6. Comparison between chemical shifts predicted using Eq. (5) and experimental ones (in ppm) for two small RNA hairpins in solution, presented separately for the C1′-C5′ nuclei.

12-mer RNA hairpin whose NMR structure was reported by Varani *et al.*[84] and the data about chemical shifts and puckering parameters were taken from the supporting information of Xu *et al.*[43] (Figure 5(b)). We found that the use of Eq. (5) results in a promising correlation between predicted and experimental chemical shifts (the C1′ chemical shifts were shifted by 15 ppm as described above). It should be stated again, however, that the model that we discuss in this paper describes only the core effect of the ribose conformation and does not account a number of important effects relevant in the whole RNA molecule. In particular, as in the case of the nucleosides and nucleotides discussed above, the calculations for the C1′ and C2′ chemical shifts (Figures 6(a) and 6(b)) suggest that they are strongly affected by other factors, such as conjugated ring and χ dihedral angle effects, not taken into account in the present model. The results for the C3′-C5′ chemical shifts are slightly better, in particular for the C3′ chemical shifts, which are less dependent on factors other than the sugar pucker effect.

CONCLUSIONS

We have investigated the relationship between the atomic coordinates of the RNA ribose ring and the chemical shifts of its carbon atoms calculated by density functional theory on a reduced model. We have found that it is possible to derive a geometrical parametrization that enables the prediction of these chemical shifts from the knowledge of two structural parameters, P and ν_{max}, which describe the conformation of the ribose ring. Our results indicate that through the approach followed in this work, subject to further development and extension of the system size, it may become possible to establish general phenomenological relationships between the structures and the chemical shifts of nucleic acids and exploit them for structure determination.

ACKNOWLEDGMENTS

We acknowledge support from the Ánimo-Chévere Project of the Erasmus-Mundus Program (R.S.), the Schlumberger Research and Darwin College (A.B.S.), and BBSRC (M.V.).

[1]D. Elliot and M. Ladomery, *Molecular Biology of RNA* (Oxford University Press, Oxford, 2011).
[2]D. P. Bartel, Cell **136**, 215 (2009).
[3]L. S. Waters and G. Storz, Cell **136**, 615 (2009).
[4]R. W. Carthew and E. J. Sontheimer, Cell **136**, 642 (2009).
[5]R. C. Friedman, K. K. H. Farh, C. B. Burge, and D. P. Bartel, Genome Res. **19**, 92 (2009).
[6]P. W. Rose, B. Beran, C. X. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlic, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman, and P. E. Bourne, Nucleic Acids Res. **39**, D392 (2011).
[7]H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).
[8]K. Wuthrich, Science **243**, 45 (1989).
[9]N. Tjandra and A. Bax, Science **278**, 1111 (1997).
[10]J. R. Tolman, J. M. Flanagan, M. A. Kennedy, and J. H. Prestegard, Proc. Natl. Acad. Sci. U.S.A. **92**, 9279 (1995).
[11]J. R. Bothe, E. N. Nikolova, C. D. Eichhorn, J. Chugh, A. L. Hansen, and H. M. Al-Hashimi, Nat. Methods **8**, 919 (2011).
[12]D. S. Wishart, Prog. Nucl. Magn. Reson. Spectrosc. **58**, 62 (2011).
[13]A. Cavalli, X. Salvatella, C. M. Dobson, and M. Vendruscolo, Proc. Natl. Acad. Sci. U.S.A. **104**, 9615 (2007).
[14]Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, and A. Bax, Proc. Natl. Acad. Sci. U.S.A. **105**, 4685 (2008).
[15]G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker, and L. E. Kay, Nature (London) **477**, 111 (2011).
[16]D. M. Korzhnev, T. L. Religa, W. Banachewicz, A. R. Fersht, and L. E. Kay, Science **329**, 1312 (2010).
[17]P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundstrom, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, and L. E. Kay, Science **336**, 362 (2012).
[18]C. Camilloni, P. Robustelli, A. De Simone, A. Cavalli, and M. Vendruscolo, J. Am. Chem. Soc. **134**, 3968 (2012).
[19]G. Varani, F. Aboulela, and F. H. T. Allain, Prog. Nucl. Magn. Reson. Spectrosc. **29**, 51 (1996).
[20]S. E. Butcher, F. H. T. Allain, and J. Feigon, Biochemistry **39**, 2174 (2000).
[21]Z. H. Liu, I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Kramer, and M. Sattler, Science **294**, 1098 (2001).
[22]B. Fürtig, C. Richter, J. Wöhnert, and H. Schwalbe, ChemBioChem **4**, 936 (2003).
[23]P. J. Lukavsky and J. D. Puglisi, Methods Enzymol. **394**, 399 (2005).
[24]J. Flinders, and T. Dieckmann, Prog. Nucl. Magn. Reson. Spectrosc. **48**, 137 (2006).
[25]M. P. Latham, D. J. Brown, S. A. McCallum, and A. Pardi, ChemBioChem **6**, 1492 (2005).
[26]C. Dominguez, M. Schubert, O. Duss, S. Ravindranathan, and F. H. T. Allain, Prog. Nucl. Magn. Reson. Spectrosc. **58**, 1 (2011).
[27]J. Lehtivarjo, K. Tuppurainen, T. Hassinen, R. Laatikainen, and M. Perakyla, J. Biomol. NMR **52**, 257 (2012).
[28]K. Hall, Methods Mol. Biol. **875**, 67 (2012).
[29]X. P. Xu and D. A. Case, J. Biomol. NMR **21**, 321 (2001).
[30]K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, and M. Vendruscolo, J. Am. Chem. Soc. **131**, 13894 (2009).
[31]B. Han, Y. F. Liu, S. W. Ginzinger, and D. S. Wishart, J. Biomol. NMR **50**, 43 (2011).
[32]A. B. Sahakyan, W. F. Vranken, A. Cavalli, and M. Vendruscolo, J. Biomol. NMR **50**, 331 (2011).
[33]A. B. Sahakyan, W. F. Vranken, A. Cavalli, and M. Vendruscolo, Angew. Chem., Int. Ed. **50**, 9620 (2011).
[34]Y. Shen and A. Bax, J. Biomol. NMR **48**, 13 (2010).
[35]P. Rossi and G. S. Harbison, J. Magn. Reson. **151**, 1 (2001).
[36]A. B. Sahakyan and M. Vendruscolo, J. Phys. Chem. B **117**, 1989 (2013).

[37]S. L. Lam, Nucl. Acids Res. **35**, W713 (2007).
[38]S. L. Lam and L. M. Chi, Prog. Nucl. Magn. Reson. Spectrosc. **56**, 289 (2010).
[39]A. T. Frank, S. Horowitz, I. Andricioaei, and H. Al-Hashim, J. Phys. Chem. B **117**, 2045 (2013).
[40]C. Altona and M. Sundaral, J. Am. Chem. Soc. **94**, 8205 (1972).
[41]A. P. Dejaegere and D. A. Case, J. Phys. Chem. A **102**, 5280 (1998).
[42]X. P. Xu and S. C. F. Au-Yeung, J. Phys. Chem. B **104**, 5641 (2000).
[43]X.-P. Xu, W.-L. A. K. Chiu, and S. C. F. Au-Yeung, J. Am. Chem. Soc. **120**, 4230 (1998).
[44]M. Ebrahimi, P. Rossi, C. Rogers, and G. S. Harbison, J. Magn. Reson. **150**, 1 (2001).
[45]O. Ohlenschläger, S. Haumann, R. Ramachandran, and M. Görlach, J. Biomol. NMR **42**, 139 (2008).
[46]E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, and J. L. Markley, Nucleic Acids Res. **36**, D402 (2008).
[47]A. V. Cherepanov, C. Glaubitz, and H. Schwalbe, Angew. Chem., Int. Ed. **49**, 4747 (2010).
[48]D. L. Bryce, A. Grishaev, and A. Bax, J. Am. Chem. Soc. **127**, 7387 (2005).
[49]D. Sitkoff and D. A. Case, Prog. Nucl. Magn. Reson. Spectrosc. **32**, 165 (1998).
[50]J. A. Vila, J. M. Aramini, P. Rossi, A. Kuzin, M. Su, J. Seetharaman, R. Xiao, L. Tong, G. T. Montelione, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **105**, 14389 (2008).
[51]J. A. Vila, Y. A. Arnautova, O. A. Martin, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **106**, 16972 (2009).
[52]J. A. Vila, Y. A. Arnautova, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **105**, 1891 (2008).
[53]J. Abi-Ghanem, B. Heddi, N. Foloppe, and B. Hartmann, Nucleic Acids Res. **38**, e18 (2010).
[54]A. M. Belostotskii, J. Org. Chem. **73**, 5723 (2008).
[55]J. A. Vila, M. E. Villegas, H. A. Baldoni, and H. A. Scheraga, J. Biomol. NMR **38**, 221 (2007).
[56]Q. Gao, S. Yokojima, T. Kohno, T. Ishida, D. G. Fedorov, K. Kitaura, M. Fujihira, and S. Nakamura, Chem. Phys. Lett. **445**, 331 (2007).
[57]R. Jain, T. Bally, and P. R. Rablen, J. Org. Chem. **74**, 4017 (2009).
[58]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[59]C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[60]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[61]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).
[62]C. Adamo and V. Barone, Chem. Phys. Lett. **298**, 113 (1998).
[63]J. R. Cheeseman, G. W. Trucks, T. A. Keith, and M. J. Frisch, J. Chem. Phys. **104**, 5497 (1996).
[64]F. Mocci, M. Usai, and G. Cerioni, Magn. Reson. Chem. **47**, 31 (2009).
[65]Y. Zhang, X. Xu, and Y. J. Yan, J. Comput. Chem. **29**, 1798 (2008).
[66]See supplementary material at http://dx.doi.org/10.1063/1.4811498 for Tables S1 and S2, Figures S1 and S2, and three files with the P and ν_{max} parameters.
[67]S. Berger, W. Bock, G. Frenking, V. Jonas, and F. Mueller, J. Am. Chem. Soc. **117**, 3820 (1995).
[68]A. F. Casy, Magn. Reson. Chem. **26**, 351 (1988).
[69]V. I. Dostovalova, S. S. Zlotsky, D. L. Rakhmankulov, and A. B. Terent'ev, Magn. Reson. Chem. **25**, 1 (1987).
[70]G. E. Maciel and G. B. Savitsky, J. Phys. Chem. **69**, 3925 (1965).
[71]K. Kakinuma, B. I. Milavetz, and K. L. Rinehart, J. Org. Chem. **41**, 1358 (1976).
[72]H. E. Gottlieb, V. Kotlyar, and A. Nudelman, J. Org. Chem. **62**, 7512 (1997).
[73]M. W. Lodewyk, M. R. Siebert, and D. J. Tantillo, Chem. Rev. **112**, 1839 (2012).
[74]F. A. A. M. de Leeuw, P. N. Van Kampen, C. Altona, E. Diez, and A. L. Esteban, J. Mol. Struct. **125**, 67 (1984).
[75]A. Gelbin, B. Schneider, L. Clowney, S.-H. Hsieh, W. K. Olson, and H. M. Berman, J. Am. Chem. Soc. **118**, 519 (1996).
[76]M. Schmidt and H. Lipson, Science **324**, 81 (2009).
[77]M. Karplus, J. Chem. Phys. **30**, 11 (1959).
[78]F. H. Allen, Acta Crystallogr., Sect. B: Struct. Sci. **58**, 380 (2002).
[79]E. Breitmaier and W. Voelter, *Carbon-13 NMR Spectroscopy* (VHC, Weinheim, 1986).
[80]K. Ösapay and D. A. Case, J. Am. Chem. Soc. **113**, 9436 (1991).
[81]D. A. Case, Curr. Opin. Struct. Biol. **8**, 624 (1998).
[82]B. Fürtig, C. Richter, W. Bermel, and H. Schwalbe, J. Biomol. NMR **28**, 69 (2004).
[83]S. Nozinovic, B. Fürtig, H. R. A. Jonker, C. Richter, and H. Schwalbe, Nucleic Acids Res. **38**, 683 (2010).
[84]G. Varani, C. J. Cheong, and I. Tinoco, Biochemistry **30**, 3280 (1991).