Ground State Properties of the Nucleic Acid Constituents Studied by Density Functional Calculations. I. Conformational Features of Ribose, Dimethyl Phosphate, Uridine, Cytidine, 5′-Methyl Phosphate−Uridine, and 3′-Methyl Phosphate−Uridine

Nicolas Leulliot and Mahmoud Ghomi*

Laboratoire de Physicochimie Biomoléculaire et Cellulaire, UPRESA CNRS 7033, Case courrier 138, Université Pierre et Marie Curie, 4 Place Jussieu, F-75252 Paris Cedex 05, France

Giovanni Scalmani

Service de Chimie des Matériaux Nouveaux, Centre de Recherche en Electronique et Photonique Moléculaires, Université de Mons-Hainaut, 20 Place du Parc, B-7000 Mons, Belgium

Gaston Berthier

Laboratoire d'Etude Théorique des Milieux Extrêmes, Ecole Normale Supérieure, 24 rue Lhomond, F-75231 Paris Cedex 05, France

Received: May 14, 1999

Ground state energies and geometries have been determined at the DFT/B3LYP level for different model compounds such as ribose, dimethyl phosphate, uridine, cytidine, 3′-methyl phosphate−uridine, and 5′-methyl phosphate−uridine as a function of the most prominent conformations adopted by each of them. The counterion used for neutralizing the phosphate negative charge was an ammonium ion (NH_4^+). This systematic study allowed us to analyze the stability of a ribonucleotide (base+ribose+phosphate) which is the chemical repeating unit of RNA. In the dimethyl phosphate model, the lowest energy corresponds to the gauche⁻-gauche⁻ conformation, as also predicted by previous calculations on this motif at different theoretical levels. In the ribose model, the C2′-endo (S-type) conformer has a lower energy than the C3′-endo (N-type) one. When a pyrimidine base (uracil or cytosine) is added to the ribose to form a ribonucleoside, the electronic energies of the three optimized conformers with the C3′-endo and C2′-endo sugar puckers as well as the anti and syn orientations of the base with respect to the sugar are located in the following order: C3′-endo/anti < C2′-endo/anti < C3′-endo/syn. However, the energy difference between these conformers depends on the type of the pyrimidine base connected to the ribose. The optimization of the ribonucleotides confirms the stability of the conformers containing A- and Z-form conformational angles. The role of the intramolecular O−H···O and C−H···O hydrogen bonds in the overall stability of ribose, nucleosides, and ribonucleotides has been discussed.

I. Introduction

RNA is a single-stranded polymer, formed by building block monomers called ribonucleotides. It generally folds onto itself to give very complex tertiary structures. This folding generally appears along particular intramolecular RNA structures called hairpins. The resolution of the tRNA[1] and ribozyme[2,3] structures from X-ray diffraction patterns provided interesting information on the secondary structures of hairpins, mainly arising from the conformational properties of the nucleotides involved in them. During recent years, synthetic oligoribonucleotides, which mimic the sequences of short and unusually stable hairpins in aqueous phase, have been prepared and analyzed by a large variety of spectroscopic techniques: UV absorption,[4−9] NMR,[8−13] and vibrational[8,9,15] spectroscopy. Structural data arising from these analyses are consistent with characteristic nucleotide conformations appearing in the short loops of highly stable and conserved hairpins. Well-known examples are the tetraloops belonging to the UNCG,[4−5,7−11,15] GNRA,[6,12−14] and CUUG[16] families (where

N is any of the four major RNA nucleotides and R is a purine). In fact, some of the nucleosides in these tetraloops adopt the unusual C2′-endo/anti or C3′-endo/syn conformations, instead of the C3′-endo/anti one which is observed, for instance, in the canonical A-form single-stranded or double-stranded helix. Of course, other experimental data[17,18] had previously confirmed the possibility for RNA to form a Z-form double helix with unusual nucleoside conformations, i.e., C2′-endo/anti pyrimidine nucleosides and C3′-endo/syn purine nucleosides, such as those included in the Z-form DNA double helices.

As far as the previous theoretical investigations on the nucleic acid constituents are concerned, one can recall the pioneering calculations of Saran and Pullman[19] based on the perturbative configuration interaction over localized orbitals (PCILO) method for constructing the conformational energy map of nucleosides and nucleotides by single-point calculations. Other calculations based on particular conformations of nucleosides, modified nucleosides, or their constituents performed at the Hartree−Fock (HF),[20−25] density fuctional theory (DFT)[26] and semiempirical[27−29] levels are also worth mentioning. Some other recent

* To whom all correspondence should be addressed. Phone: +33-1-44277560. Fax: +33-1-44277560. E-mail: ghomi@lpbc.jussieu.fr

Ground State Properties of Nucleic Acid Constituents

J. Phys. Chem. A, Vol. 103, No. 43, 1999 **8717**

results at the HF[25] and MP2[30] levels, based on the model compounds whose chemical structures are close to the nucleosides and nucleotides involved in nucleic acid chains, should also be mentioned. In DNA, base-stacking and base-pairing effects have also been examined by quantum mechanical calculations including electronic correlation.[31]

The main aim of these calculations was to analyze the dependence of sugar geometrical parameters on its well-known phase angle of pseudorotation (P) and to show that the C2′-endo and C3′-endo sugar puckers correspond to the conformational energy global minima of DNA and RNA nucleosides, respectively.

The present paper is the first of a new series of investigations devoted to the analysis of the ground state properties (conformation and vibration) of the RNA constituents. Here, our effort has been devoted to a systematic investigation of the role and influence of each of these constituents (phosphate, ribose, and base) on the energy and the overall conformation of an isolated pyrimidine ribonucleotide. Recently, we have published a series of results obtained at the Møller−Plesset second-order perturbation (MP2) level concerning the geometry optimization and vibrational modes of pyrimidine nucleic acid bases.[32−36] It has been shown that in the case of uracil, the DFT method[37] provides similar vibrational modes to those obtained at the MP2 level.[32] We should note here that vibrational calculations through their comparison with experimental results (observed wavenumbers and intensities) are considered to be an excellent test for analyzing both the molecular geometry and the shape (curvature) of the potential energy surface. All other results obtained up to now on the nucleic acid constituents have clearly shown that the DFT method, which takes into account electronic correlation by means of nonlocal exchange and correlation functional,[38,39] can be undoubtedly used as a cost-effective alternative to the sophisticated and time-consuming MP2 treatment, provided that a reasonable sized basis set is used.[40,41] Thus, the present results can complete all of the calculated results existing in the literature on the ribose, phosphate group and nucleosides. In addition, to our knowledge, no full geometry optimization including correlation effects has been performed up to now on the different conformers of a ribonucleotide.

II. Theoretical Details

II.A. Choice of Molecular Models. *Ribose and Nucleosides.* Preliminary calculations have been performed at the HF,[21,24,25] DFT,[26] and MP2[25−26,30,42] levels on the molecular models containing a furanose ring. The molecular model called *ribose* in this work (Figure 1) is in fact a subunit of the nucleoside containing three hydroxyl groups at the 2′, 3′, and 5′ positions. The base, which is generally connected to the C1′ atom in a ribonucleotide, is replaced by a hydrogen atom called H1″ in this molecular model. As far as the pyrimidine nucleoside models are concerned, they correspond, in fact, to 1-β-D-furanosyl-uracil and 1-β-D-furanosyl-cytosine,[43] hereafter referred to as uridine and cytidine, respectively (Figure 1). Uracil is the structurally simplest RNA base, and both DFT and MP2 calculations suggested a planar geometry (C_s symmetry) for this molecule (when considered isolated).[32] Full geometry optimization of cytosine base by MP2 and DFT methods has revealed the pyramidalization of the amino group leading to a slight perturbation of the ring (see ref 34 and references therein).

Phosphate Group and Nucleotides. The phosphate group is one of the most important subunits of an RNA (or DNA) chain. A correct modeling of this compound obviously needs the explicit consideration of a counterion neutralizing the negative
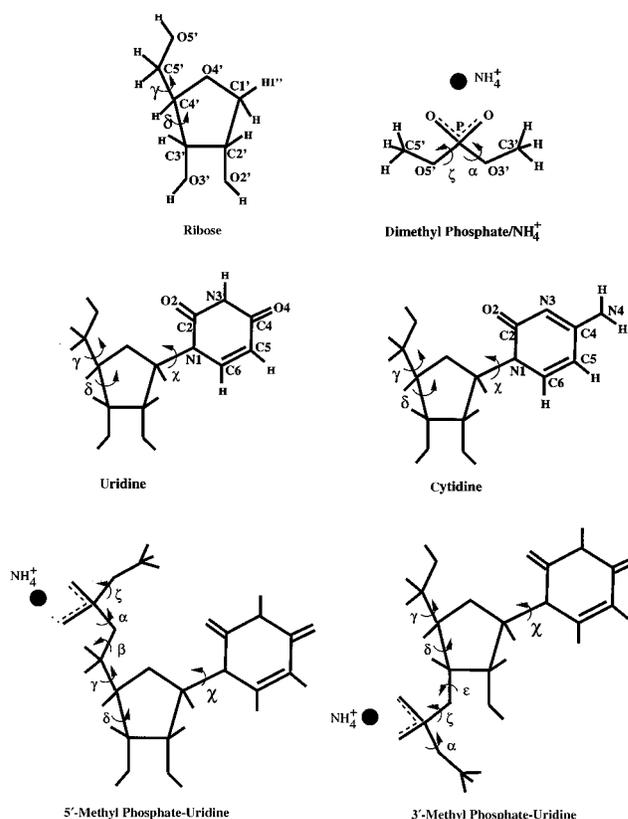
Figure 1. Atom numbering and chemical structure of the RNA constituents used as model compounds in the present calculations. The conformational angles of the ribonucleotides, i.e., α, β, γ, δ, ϵ, ζ, and χ are also shown on the chemical bonds.

charge of phosphate. On the basis of experimental observations, it is impossible to obtain an ordered nucleic acid chain without screening the electrostatic repulsion between the adjacent phosphate groups with counterions. We can also emphasize the conformational transitions of nucleic acids that can be induced by salt concentration, i.e., B−Z or A−Z transitions in DNA and RNA, respectively. The most common counterion in nucleic acids is obviously Na$^+$, but this cation interacts with the phosphate group through a sophisticated complexation with water molecules.[43] Some recent calculations are, however, based on the isolated dimethyl phosphate model compound (anion or neutralized by Na$^+$ counterion) at the HF and DFT levels of theory.[44] Other calculations treat this model compound in its anionic form (without counterion) for estimating the solvation free energies by using the Langevin dipole method to model solvent.[45] Recent results obtained at the HF level[46] on the dimethyl phosphate molecular model with NH$_4^+$ have led us to use this counterion which has the advantage of possessing high symmetry and forming two directional intermolecular N−H···O=P hydrogen bonds with the phosphate group (Figure 1). Although it has been shown that in the dimethyl phosphate model compound, the presence of counterion does not influence the energy order of different conformers,[44] it is, however, absolutely necessary to neutralize the negative charge on the phosphate in order to correctly model the nucleotides because of long-range electrostatic interactions between this phosphate group and other remaining elements such as base and ribose. To give to the phosphate group and its counterion an environment similar to the one found in a RNA chain, we have inserted the phosphate group between a nucleoside and a methyl group, thus leading to either 5′-methyl phosphate-uridine or 3′-methyl phosphate-uridine (Figure 1) molecular models. These two

compounds can be compared to the nucleotides located at the 5′- or 3′-terminals of a poly- or oligoribonucleotide.

II.B. Initial Conformations of the Molecular Models. Instead of performing single-point energy calculations for constructing the energy map of the molecular models shown in Figure 1 by a continuous variation of their conformational (torsion) angles, we have focused our effort on the complete geometry optimization of these models in order to find directly the lowest energy conformers in the vicinity of the most interesting starting conformations. The choice of these starting conformations has been made according to its relevance with respect to the local geometry of these motifs in RNA chains, i.e., the conformations observed experimentally in single-stranded and double-stranded helices as well as in the folded single-stranded chains (see the Introduction of this paper). No conformational degree of freedom (bond-lengths, valence angles, and torsion angles) was frozen in the course of the geometry optimization.

Ribose, Uridine, and Cytidine. Two opposite twist geometries, corresponding to the C3′-endo and C2′-endo local conformations,[43] have been selected as starting conformations for ribose. These two types of sugar puckering can also be analyzed by means of the δ torsion angle (Figure 1): $\delta = \phi(O3′-C3′-C4′-C5′)$.

As far as the nucleoside (uridine and cytidine) molecular models are concerned, three possible different starting points, i.e., C3′-endo/anti, C3′-endo/syn, and C2′-endo/anti conformations, have been considered. Anti and syn refer to the orientation of the base with respect to the sugar ring and are determined by the value of the glycosyl torsion angle: $\chi = \phi(O4′-C1′-N1-C2)$ (Figure 1). The initial values of the χ angle were approximately 180° and 60° for the anti and syn orientations, respectively. In all cases, the initial value of the γ torsion angle, where $\gamma = \phi(C3′-C4′-C5′-O5′)$ (Figure 1), corresponds to a gauche$^+$ ($\gamma = +60°$) orientation of the O5′ atom with respect to the C3′ atom of the sugar moiety. Generally speaking, the gauche$^+$ orientation is the most favorable one in nucleosides including pyrimidine bases.[43]

Dimethyl Phosphate, 3′-Methyl Phosphate-Uridine, and 5′-Methyl Phosphate- Uridine. In dimethyl phosphate, the initial conformations correspond to the values associated with the couple of torsion angles ($\alpha-\zeta$, see below) defined around the phosphate group P–O single bonds (Figure 1). Three different initial couples, namely gauche$^-$-gauche$^-$, gauche$^-$-trans, and trans-trans, have been considered, where gauche$^-$ and trans correspond to −60° and 180° initial torsion angles, respectively.

The attachment of the phosphate group in the nucleoside at its 5′ and 3′ ends to the uridine nucleoside to construct the 5′-methyl phosphate or 3′-methyl phosphate nucleotides leads to a notable increase of the number of torsion angles to be considered, i.e., α, β, ϵ, and ζ, where $\alpha = \phi(O3′-P-O5′-C5′)$, $\beta = \phi(P-O5′-C5′-C4′)$, $\epsilon = \phi(C4′-C3′-O3′-P)$, and $\zeta = \phi(C3′-O3′-P-O5′)$ (Figure 1). The starting values of these torsion angles have been chosen in order to obtain: (i) a nucleotide conformation as observed in the canonical A-form of RNA, (ii) a nucleotide conformation as observed in the Z-form of RNA (or Z-DNA), (iii) a nucleotide conformation having all the canonical A-form torsion angles, except the δ angle which gives rise to a C2′-endo sugar puckering. Hereafter these three conformations will be referred to as A, Z, and A*, respectively.

Through the choice of the above-mentioned starting conformations, we have attempted to first confirm the stability of the nucleotides with all A- and Z-form torsion angles and second,

to analyze the effect of the C3′-endo to C2′-endo conformational transition (i.e., the change in the δ torsion angle) on the overall geometry of a nucleotide in which all other torsion angles correspond to those of an A-form RNA chain.

II.C. Theoretical Methods and Computational Details. To estimate the geometrical parameters of the above-mentioned molecular models (§II.A) considered as isolated molecules, quantum mechanical calculations have been carried out at the DFT level by means of the hybrid Becke−Lee−Yang−Parr (B3LYP) type nonlocal exchange and correlation functional[47−49] and double-ζ, split valence 6-31G Gaussian basis sets enlarged with d-type polarization functions. The exponents of the polarization functions were 0.55, 0.75, 0.80, and 0.85 for phosphorus, carbon, nitrogen, and oxygen, respectively.[50,51] Hereafter, this special atomic basis set will be referred to as 6-31G$^{(*)}$. Moreover, to model as accurately as possible the hydrogen bonds between the phosphate group bound to the NH_4^+ cation, semidiffuse functions have been added. For the sake of computational cost, these additional functions have been considered only for the atoms directly involved in the H-bonds, i.e., the phosphate oxygens and their partners (hydrogens) involved N−H···O═P hydrogen bonds. The exponents of the semidiffuse functions were those reproducing electronic polarizabilities in simple molecules, i.e., $\zeta_{s,p} = 0.05$ and $\zeta_d = 0.30$ for oxygen as well as $\zeta_{s,p} = 0.10$ for hydrogen. Hereafter, as in our previous papers based on intermolecular hydrogen bonds, this atomic basis set will be referred to as 6-31G$^{(†)}$.[40−41]

Consequently, the total number of basis functions used in the calculations on the six molecular models of Figure 1 was 146 (for ribose), 179 (for dimethyl phosphate/ NH_4^+), 262 (for uridine), 264 (for cytidine), and 396 (for 3′- or 5′-methyl phosphate-uridine/NH_4^+). Because of the magnitude of these numbers, numerical computations on the model compounds containing less than 300 basis functions have been carried out on Cray C94 and C98 supercomputers, using the *Gaussian 94* package.[52] For the calculations on the 3′- or 5′-methyl phosphate-uridine/NH_4^+, an in-house modified version of *Gaussian 94* with enhanced parallel performance[53] has been used on MPP (massively parallel processors) platforms such as Silicon Graphics Cray Origin2000 and Cray T3E. This version of the program takes advantage of message passing interface (MPI)[54] to accomplish interprocessor communications and of ScaLA-PACK[55] to carry out various matrix operations in parallel (Fock matrix diagonalization and extrapolation, eigenvector rotations, etc.). In the case of complex geometry optimizations such as those reported in this paper (which involved sometimes more than one hundred steps), the reduction in computational time attributable to the enhanced parallel capabilities of our modified code has been important.

III. Results and Discussion

Before carrying out a detailed examination of the geometrical parameters and energies of the different conformers considered in this paper, we should mention that harmonic vibrational analyses at the B3LYP/6-31G$^{(*)}$ level for ribose, uridine, and cytidine and at the B3LYP/6-31G$^{(†)}$ level for dimethyl phosphate/ NH_4^+ have been performed in order to check the reliability of the optimized conformations. The absence of imaginary frequencies for the optimized geometries led us to conclude that they correspond well to ground state energy minima for the considered molecular models. We have also verified that the addition of the zero-point vibrational energy (ZPV) does not change the energy order of the different conformers. However, discussion on all of these points is beyond the scope of this

TABLE 1: Energy Values Obtained by Geometry Optimization at the DFT Level for the Different Model Compounds

model compd	conformation	E_e (Hartrees)^c	ΔE_e (kcal/mol)^c	theor level
ribose	C2′-endo	−497.383118	0.	B3LYP/6-31G$^{(*)}$
	C3′-endo	−497.382398	+0.45	
uridine	C3′-endo /anti	−910.998590	0.	B3LYP/6-31G$^{(*)}$
	C2′-endo /anti	−910.998586	+2. 10^{-3}	
	C3′-endo /syn	−910.994692	+2.45	
cytidine	C3′-endo /anti	−891.111439	0.	B3LYP/6-31G$^{(*)}$
	C2′-endo /anti	−891.109443	+1.25	
	C3′-endo /syn	−891.107050	+2.75	
dimethyl phosphate	gauche$^-$-gauche$^{-\,d}$	−779.330394	0.	B3LYP/6-31G$^{(\dagger)}$
	gauche$^-$-trans^d	−779.328840	+0.97	
	trans-trans	−779.325154	+3.29	
5′-MPU^a	A^e	−1574.615209	0.	B3LYP/6-31G$^{(\dagger)}$
	Z^f	−1574.617954	−1.72	
	A*^{g}	−1574.614542	+0.42	
3′-MPU^b	A	−1574.613092	0.	B3LYP/6-31G$^{(\dagger)}$
	A*	−1574.613788	−0.44	

^{a} 5′-MPU: 5′-methyl phosphate-uridine. ^{b} 3′-MPU: 3′-methyl phosphate-uridine. ^{c} E_e and ΔE_e are the electronic energy and the differences between the electronic energies. ^{d} Gauche$^-$ and trans refer to the values of α and ζ torsion angles defined around the phosphate group P−O single bonds. ^{e} A refers to the conformation of a nucleotide involved in the canonical A-form of RNA. ^{f} Z refers to the conformation of a pyrimidine nucleotide involved in Z-form of RNA (or Z-DNA), i.e., with a C2′-endo conformation for ribose. ^{g} A* corresponds to the conformation of a nucleotide having the A-form torsion angles, except for the δ angle which corresponds to a C2′-endo sugar puckering.

paper which is mainly focused on geometry and energy. The detailed vibrational analyses of uridine and cytidine will be reported in the forthcoming publications of this series.

III.A. Conformational Energies and Conformational Angles. In Table 1 are given the electronic energies (E_e) obtained after full geometry optimization. To compare the energies of the different conformers of the same molecular modes, we have also reported the electronic energy difference (ΔE_e) calculated as follows:

(i) In the case of the ribose, uridine, cytidine, and dimethyl phosphate molecular models, the electronic energy of the conformer relative to the lowest energy minima has been taken as reference ($\Delta E_e = 0$), and the energy difference between this conformer and the others (positive values) has been reported consequently.

(ii) In the case of the nucleotides (3′- or 5′-methyl phosphate-uridine) the energy of the A-conformer (with the initial torsion angles corresponding to the A-RNA) has always been taken as reference ($\Delta E_e = 0$) for calculating the energy difference between conformers.

As mentioned above, the overall conformation of a nucleotide within a nucleic acid chain is determined by means of seven torsion angles (α, β, γ, δ, ϵ, ζ, and χ, see §II.B). Each of the molecular models used in this work allows us to analyze a certain number of these conformational angles (Table 2). A graphic representation of the optimized geometries is displayed in Figures 2, 3, 4, 5, and 6 for the ribose, uridine, cytidine, 5′-methyl phosphate, and 3′-methyl phosphate molecular models, respectively.

A glance at the ΔE_e values of Table 1 shows that the optimized C2′-endo and C3′-endo sugar conformations of ribose are separated in the energy scale by 0.45 kcal/mol. Note that this difference is well below the kT value at room temperature (∼0.6 kcal/mol). Thus, the C2′-endo ribose has a lower energy if isolated conformers are considered.

Although in uridine the C3′-endo/anti conformer corresponds to the lowest energy, the calculated ΔE_e between this conformer and the C2′-endo/anti one is so low (0.002 kcal/mol) that we can assume that they constitute a couple of energetically degenerate conformers. In contrast, the rotation of the base to a syn orientation costs a considerable amount of energy
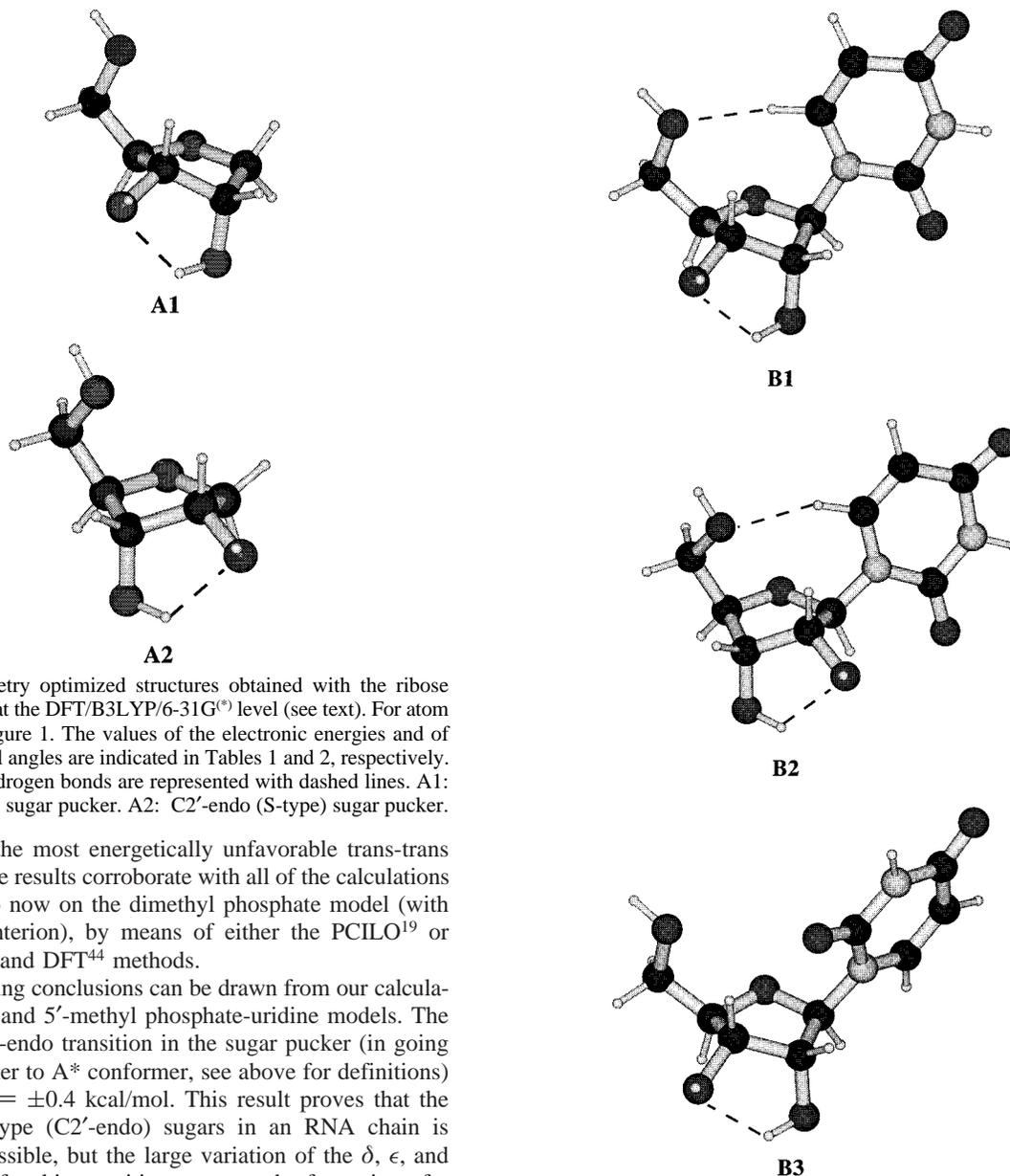
TABLE 2: Conformational Angles Obtained by Geometry Optimization at the DFT Level for the Different Model Compounds^a

model compds	conformation	α	β	γ	δ	ϵ	ζ	χ
ribose	C2′-endo			47	144			
	C3′-endo			49	87			
uridine	C3′-endo/anti			51	85			202
	C2′-endo/anti			51	146			235
	C3′-endo/syn			49	85			76
cytidine	C3′-endo/anti			52	85			197
	C2′-endo/anti			51	145			233
	C3′-endo/syn			47	84			78
dimethyl phosphate	gauche$^-$-gauche$^-$	292					292	
	gauche$^-$-trans	295					174	
	trans-trans	180					180	
5′-MPU	A	304	161	56	85			211
	A*	302	179	56	144			240
	Z	89	140	55	143			249
3′-MPU	A			53	87	215	296	202
	A*			53	146	194	267	230
RNA	A	292	178	54	84	207	289	202
DNA	B	314	213	36	156	155	264	262
DNA (pyrimidine)	Z	146	164	66	147	260	74	213

^{a} For notations used here, see the caption of Table 1. Molecular energies are indicated in Table 1. ^{b} For the definition of the torsional angles, see text and Figure 1. ^{c} The conformational angles observed in canonical A, B and Z forms are indicated for comparison.[43]

(∼2.5 kcal/mol, C3′-endo/syn conformer). In cytidine, in addition to the fact that the NH$_2$ group remains pyramidal as in the case of cytosine (see §II.A), the optimized conformers keep the same electronic energy order as in uridine. However, there is a larger energy difference between the C3′-endo/anti and C2′-endo/anti conformers (∼1.2 kcal/mol). The C3′-endo/syn conformer is higher in energy by ∼2.7 kcal/mol with respect to the C3′-endo/anti conformer. For this reason, in our calculations on the methyl phosphate nucleotides we have not considered the syn orientation for the uracil base (see below).

In dimethyl phosphate, the gauche$^-$-gauche$^-$ is the lowest energy conformer. The gauche$^-$-trans conformation is higher by ∼1 kcal/mol. Three times more energy is needed for this

A1

A2

Figure 2. Geometry optimized structures obtained with the ribose model compound at the DFT/B3LYP/6-31G$^{(*)}$ level (see text). For atom numbering see Figure 1. The values of the electronic energies and of the conformational angles are indicated in Tables 1 and 2, respectively. Intermolecular hydrogen bonds are represented with dashed lines. A1: C3′-endo (N-type) sugar pucker. A2: C2′-endo (S-type) sugar pucker.

motif to adopt the most energetically unfavorable trans-trans conformer. These results corroborate with all of the calculations performed up to now on the dimethyl phosphate model (with or without counterion), by means of either the PCILO[19] or HF,[21,46] MP2,[44] and DFT[44] methods.

Some interesting conclusions can be drawn from our calculations on the 3′- and 5′-methyl phosphate-uridine models. The C3′-endo to C2′-endo transition in the sugar pucker (in going from A conformer to A* conformer, see above for definitions) presents a $\Delta E_e = \pm 0.4$ kcal/mol. This result proves that the presence of S-type (C2′-endo) sugars in an RNA chain is energetically possible, but the large variation of the δ, ϵ, and ζ angle needed for this transition prevents the formation of a regular helix. For the Z-conformer, we have carried out calculations only on the 5′-methyl phosphate-uridine molecular model (Table 1, $\Delta E_e \sim -1.7$ kcal/mol).

III.B. Location of the NH$_4^+$ Counterion on the Phosphate Group. In the gauche$^-$-gauche$^-$ conformer of dimethyl phosphate with NH$_4^+$ counterion (corresponding to the lowest energy), the cation interacts through two symmetrical N−H···O hydrogen bonds with the oxygens of the phosphate group. The geometrical optimization on this conformer leads to a structure having full C_2 symmetry and characterized by the following geometrical parameters: N···P distance = 3.044 Å, N−H bond length = 1.085 Å (when H-bonded with phosphate), and N−H bond length = 1.018 Å (when free), H···O distance = 1.594 Å, O···N distance = 2.590 Å and N−H···O angle = 150 degrees. Other bond lengths and valence angles of the optimized phosphate group are not reported here, because their values are very close to those previously published.[44] It is also worth mentioning that the local symmetry of the phosphate group is lowered to C_1 upon its inclusion in the nucleotides because of intramolecular interactions involving the phosphate oxygens. Consequently, N−H bond lengths and N···O distances become slightly asymmetrical compared to the values found in

B1

B2

B3

Figure 3. Geometry optimized structures obtained with the uridine model compound at the DFT/B3LYP/6-31G$^{(*)}$ level (see text). For atom numbering see Figure 1. The values of the electronic energies and of the conformational angles are indicated in Tables 1 and 2, respectively. Intermolecular hydrogen bonds are represented with dashed lines. B1: C3′-endo/anti conformer. B2: C2′-endo/anti conformer. B3: C3′-endo/syn conformer.

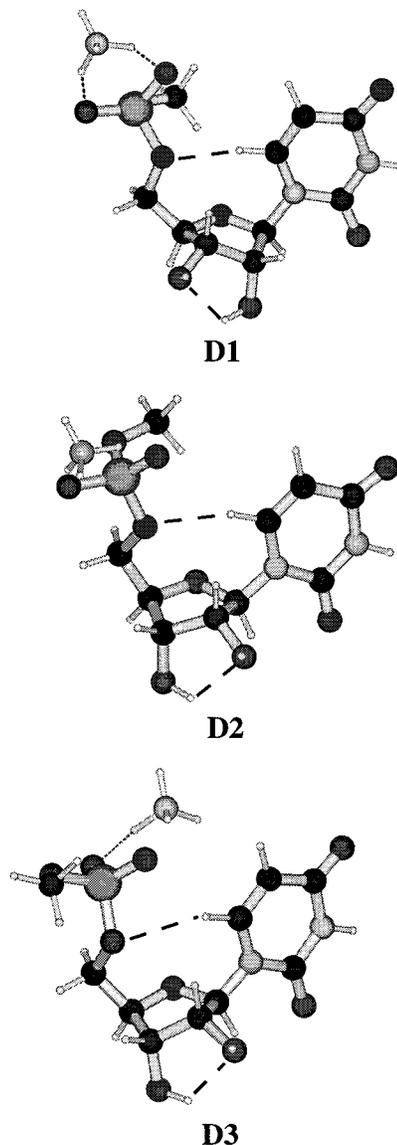dimethyl phosphate. Therefore, a slight dispersion around the above-mentioned values is obtained.

III.C. Intramolecular Hydrogen Bond Network Stabilization of the Optimized Structures. Recent reviews have indicated the important role of intramolecular H-bonds, especially those arising from the weak O−H···O and C−H···O interactions, in stabilizing the conformations adopted by biomolecules.[55−58] In fact, because of the absence of strong hydrogen bonds in the phosphodiester chain, the above-mentioned weaker hydrogen bonds contribute to the conformational stability of nucleic acids. From the extensive analysis of crystallographic data, the C6−H···O5′ has been the most documented nonstandard H-bond of RNA structures.[57] The present quantum mechanical calculations performed on the different RNA constituents provide us a preliminary insight not

Ground State Properties of Nucleic Acid Constituents

J. Phys. Chem. A, Vol. 103, No. 43, 1999 **8721**
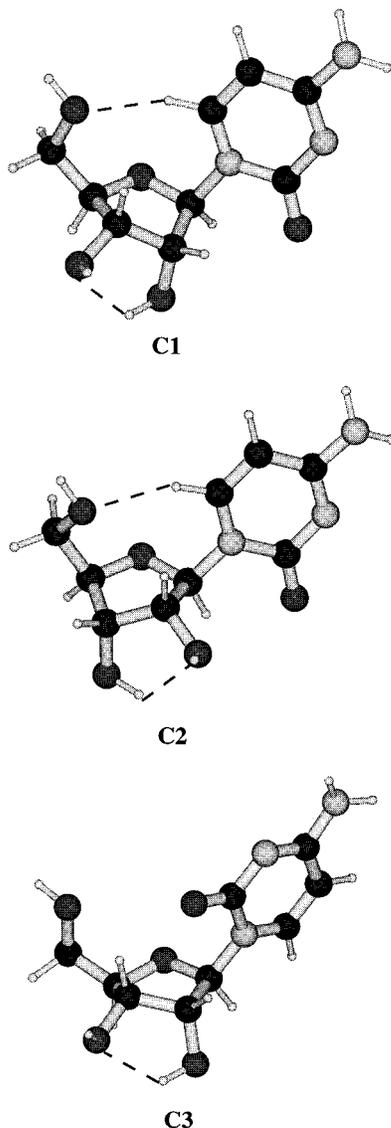
Figure 4. Geometry optimized structures obtained with the cytidine model compounds at the DFT/B3LYP/6-31G[(*)] level (see text). For atom numbering see Figure 1. The values of the electronic energies and of the conformational angles are indicated in Tables 1 and 2, respectively. Intermolecular hydrogen bonds are represented with dashed lines. C1: C3′-endo/anti conformer. C2: C2′-endo/anti conformer. C3: C3′-endo/syn conformer.
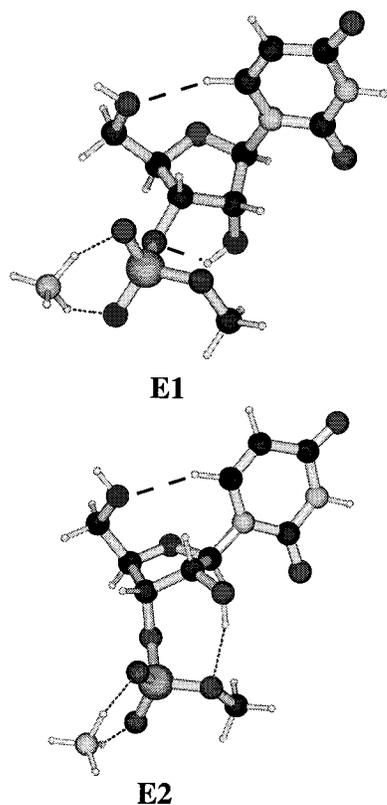
Figure 5. Geometry optimized structures obtained with the 5′-methyl phosphate-uridine/NH$_4^+$ model compound at the DFT/B3LYP/6-31G[(†)] level (see text). For atom numbering see Figure 1. The values of the electronic energies and of the conformational angles are indicated in Tables 1 and 2, respectively. Intermolecular hydrogen bonds are represented with dashed lines. D1: A conformer. D2: A* conformer. D3: Z conformer. For the definition of these conformers see text and caption of Table 1.

only to confirm the role of the experimentally observed hydrogen bonding but also to analyze the influence of all possible C−H···O and O−H···O interactions on the structure of the building blocks of nucleic acids. We have selected in Table 3 three representative H-bonds involving C6−H6, O3′−H, and O2′−H bonds which should be considered as important stabilizing factors in ribose, nucleosides, and nucleotides. For each case, the C−H and O−H bond lengths as well as the hydrogen−acceptor, donor−acceptor distances, and donor−hydrogen−acceptor angles have been reported. The most prominent results from this analysis can be summarized as follows:

(i) Since in ribose, uridine, cytidine, and 5′-methyl phosphate-uridine the two hydroxyl groups are on the two adjacent C2′ and C3′ carbons, the O−H···O hydrogen bonds in which they are involved show a symmetrical character in terms of hydrogen bond lengths and donor−hydrogen−acceptor angle (Table 3): C3′-endo sugars are stabilized by a O2′−H···O3′ hydrogen

bond, whereas C2′-endo sugars are accompanied by a O3′−H···O2′ hydrogen bond (Table 3, Figures 2−5).

(ii) Upon the introduction of a phosphate group at the 3′-end of the nucleoside (3′-methyl phosphate-uridine molecular model), the lack of the 3′-hydroxyl as a result of the substitution of the hydrogen atom of this group by the phosphate motif makes a O3′−H···O2′ hydrogen bond impossible (Figure 6). Thus, as in the cases of ribose, uridine, cytidine, and 5′-methyl phosphate-uridine (see above and Figures 2-5), the presence of a C3′-endo sugar conformation in 3′-methyl phosphate-uridine leads to the formation of the O2′−H···O3′ hydrogen bond (Table 3, Figure 6). On the other hand, with a C2′-endo sugar the geometry optimization leads to the formation of a O2′−H···O5′ hydrogen bond because of the proximity of the O5′ atom to the O2′−H bond (Figure 6). This hydrogen bond formation is correlated with important changes in the ϵ and ζ torsion angles (Table 2).

(iii) As it has been observed experimentally,[56,57] the formation of the C6−H···O5′ hydrogen bond stabilizes the anti pyrimidine

E1

E2

Figure 6. Geometry optimized structures obtained with the 3′-methyl phosphate-uridine/NH_4^+ model compounds at the DFT/B3LYP/6-31G[(†)] level (see text). For atom numbering see Figure 1. The values of the electronic energies and of the conformational angles are indicated in Tables 1 and 2, respectively. Intermolecular hydrogen bonds are represented with dashed lines. E1: A conformer. E2: A* conformer. For the definition of these conformers see text and caption of Table 1.

bases in nucleosides and nucleotides (Figures 3, 5, 6). The present calculations satisfactorily confirm this fact (Table 3). Moreover, the C3′-endo to C2′-endo conformational transition of the sugar pucker in a nucleoside or a nucleotide increases the glycosyl torsion angle χ by +30 degrees (Table 2). The geometry optimization reveals that this angle variation is correlated with the reorientation of the base in order to keep intact the C6−H···O5′ hydrogen bond geometry (see superimposed structures in Figure 7).

(iv) Upon the anti to syn rotation of the uracil base, and consequently the loss of the C6−H···O5′ hydrogen bond, the C3′-endo/anti nucleoside conformation is, however, stabilized by interactions between the uracil O2 atom of the base and the C3′−H and C2′−H bonds (Figure 3).
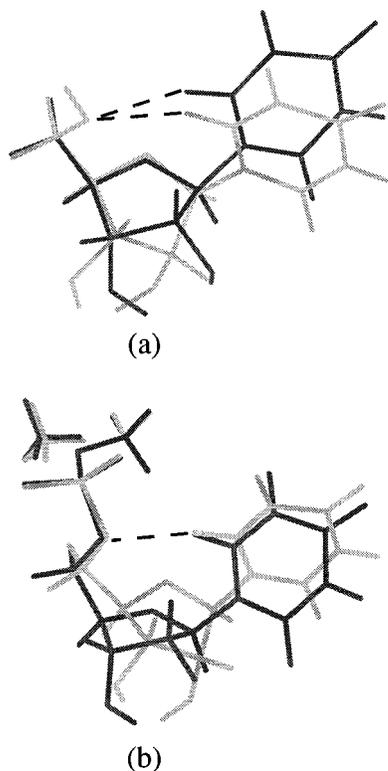
IV. Concluding Remarks

This paper is the first attempt to study the conformational properties of all of the RNA constituents (ribose, phosphate, nucleoside, and nucleotide) by means of full geometry optimization based on a reliable quantum mechanical method. It should be emphasized here that the geometrical analysis of the above-mentioned constituents needs the consideration of correlation effects. The neglect of these effects invariably leads to unreliable calculated results, especially those related to the vibrational mode analysis. This fact will be discussed in the next papers of this series. We have shown in this paper that the DFT approach using hybrid B3LYP functional and double-ζ split valence basis sets appears to be a promising method for modeling accurately at a reasonable computational cost the conformational features

TABLE 3: Geometrical Data for the Three Representative H-bonds Involving C6−H6, O3′−H and O2′−H Bonds[a,b]

| | hydrogen bonds | | | | | | | | | | | |
| | O2′-H···O3′ | | | | O3′-H···O2′ | | | | C6-H···O5′ | | | |
model compd	d(O2′−H)	d(H···O3′)	d(O2′···O3′)	θ(O2′−H···O3′)	d(O3′−H)	d(H···O2′)	d(O3′···O2′)	θ(O3′−H···O2′)	d(C6−H)	d(H···O5′)	d(C6···O5′)	θ(C6−H···O5′)
ribose (C3′-endo)	0.975	2.060	2.689	120	(0.969)							
ribose (C2′-endo)	(0.969)c				0.975	2.045	2.683	121				
uridine (C3′-endo/anti)	0.975	2.073	2.690	120	(0.969)				1.084	2.283	3.314	158
uridine (C2′-endo/anti)	(0.969)				0.975	2.055	2.680		1.086	2.376	3.418	160
uridine (C3′-endo/syn)	0.976	2.087	2.698	119	(0.968)							
cytidine (C3′-endo/anti)	0.974	1.996	2.647	122	(0.969)				1.084	2.33	3.329	151
cytidine (C2′-endo/anti)	(O.969)				0.976	2.038	2.680	121	1.086	2.425	3.458	158
cytidine (C3′-endo/syn)	0.976	1.997	2.649	122	(0.968)							
5′-MPU (A)	0.975	2.058	2.681	120	(0.969)				1.083	2.255	3.331	172
5′-MPU (A*)	(0.968)				0.975	2.047	2.673	120	1.086	2.397	3.429	158
5′-MPU (Z)	(0.969)				0.975	2.059	2.677	119	1.085	2.660	3.627	148
3′-MPU (A)	0.973	2.068	2.687	120					1.083	2.255	3.331	172
3′-MPU (A*)									1.085	2.486	3.516	158

[a] For each case, the C−H and O−H bond lengths and the hydrogen−acceptor distances (in Å), donor−acceptor distances (in Å), and donor−hydrogen−acceptor angle (in degrees) have been reported. [b] For notations used here, see the caption of Table 1. [c] In paretheses the O−H bond lengths not involved in hydrogen bonds are indicated in order to show their elongation upon H-bonding.

Ground State Properties of Nucleic Acid Constituents

J. Phys. Chem. A, Vol. 103, No. 43, 1999 **8723**

(a)

(b)

Figure 7. Superposition of the geometry optimized structures of the uridine (a) and the 5′-methyl phosphate-uridine/NH$_4^+$ (b) model compounds with C3′-endo sugar pucker (gray line drawing) and C2′-endo sugar pucker (bold line drawing). Note that whatever the ribose conformation is, the intermolecular C6−H···O5′ (dashed lines) is conserved (see also text).

of the RNA constituents and probably those of other biomolecules. From the results of our calculations, we have drawn the following conclusions:

(i) The modeling of the sugar, phosphate, and nucleosides with different conformations for each of these constituents allowed us to compare the molecular energies associated with different conformers. It has been revealed that in order to take into account all of the stabilizing factors of the structure of the nucleic acid building blocks, it is necessary to consider at least a nucleotide model compound. In the present calculations we have checked the level of accuracy of the theoretical method and basis functions used in reproducing the geometrical features of the nucleotides involved in the A-form, which is the most favorable single-stranded or double-stranded RNA conformation in physiological conditions. The agreement between the experimental and calculated results in this case has been discussed extensively above.

(ii) The geometry optimization on nucleotide model compounds with the B-form torsion angles showed drastic changes in the phosphodiester torsion angles, thus is unacceptable for the formation of a regular helix. The reason behind this is the steric hindrance between the 2′-OH and the adjacent 3′-phosphate group, and not the preference of RNA to adopt C3′-endo sugar puckering as emphasized in several papers devoted to the RNA structural features. It should be mentioned that the B-form corresponds to the preferred double-helical conformations of DNA (containing 2′-deoxyribose in its structure) in solution. However, the nonexistence of the B-form RNA does not mean that the sugar puckers in RNA cannot adopt a C2′-endo conformation. The first striking proof of this fact is the existence of the Z-form RNA containing C2′-endo sugar puckers in the pyrimidine nucleotides as proven experimentally

by several spectroscopic methods.[17−18,59−60] We have also attempted to model a pyrimidine nucleotide involved in Z-form RNA. The results of these preliminary calculations will be extended in further theoretical investigations.

(iii) More surprisingly, our calculations seem to indicate that the change of the sugar pucker from C3′-endo to C2′-endo in single-stranded RNA is energetically possible and can also be seen as releasing the 2′-hydroxyl group to form an intramolecular (or possibly an intermolecular) hydrogen bond with another acceptor. The importance of 2′-hydroxyl in hydrogen bonds and its contribution to the specificity of RNA interactions along with the presence of C2′-endo sugars in these same structures has to be emphasized: the extraordinary stable and highly conserved UUCG tetraloop hairpins contain interstrand 2′-hydroxyl hydrogen bonds stabilizing the U...G mismatched base pairing in this motif.[11] The *nucleotide triple* structure of the P4/P6 region of the self-splicing group I is stabilized through 2′-hydroxyl interactions with several sugars adopting C2′-endo sugars.[61] The critical role of some 2′-hydroxyls in the docking of the P1 helix into the catalytic core of the *Tetrahymena* ribozyme has been recently proved.[62] Moreover, the activation of the catalytic properties of hammerhead ribozymes is thought to involve a C3′-endo to C2′-endo conformational change of one of the nucleotide sugars, thus favorably orienting the 2′-hydroxyl for an attack on the opposite phosphate.[63]

A thorough understanding of the conformational features which can explain or predict such experimental results is a primary task in elucidating the structure−function relationship of RNA chains. Further theoretical works on the larger size motifs in the presence of explicit solvent and counterions would be necessary for this purpose.

References and Notes

(1) Quigley, G. J.; Rich, A. *Science* **1976**, *194*, 796.
(2) Pley, H. W.; Flaherty, K. M.; McKay, D. B. *Nature* **1994**, *372*, 111.
(3) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stodddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065.
(4) Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901.
(5) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819.
(6) SantaLucia, J., Jr.; Kierzek, R.; Turner, D. H. *Science* **1992**, *256*, 217.
(7) Molinaro, M.; Tinoco, I., Jr. *Nucleic Acids Res.* **1995**, *23*, 3056.
(8) Abdelkafi, M.; Ghomi, M.; Turpin, P. Y.; Baumruk, V.; Hervé du Penhoat, C.; Lampire, O.; Bouchemal-Chibani, N.; Goyer, P.; Namane, A.; Gouyette, C.; Huynh-Dinh, T.; Bednárová, L. *J. Biomol. Struct. Dyn.* **1997**, *14*, 579.
(9) Abdelkafi, M.; Leulliot, N.; Ghomi, M.; Hervé du Penhoat, C.; Namane, A.; Gouyette, C.; Huynh-Dinh, T.; Baumruk, V.; Turpin, P. Y. *J. Mol. Struct.* **1997**, *409*, 241.
(10) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1991**, *30*, 3280.
(11) Allain, F. H. T.; Varani, G. *J. Mol. Biol.* **1995**, *250*, 333.
(12) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191.

(13) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H. M.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968.

(14) Varani, G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379.

(15) Abdelkafi, M.; Leulliot, N.; Baumruk, V.; Bednárová, L.; Turpin, P.-Y.; Namane, A.; Gouyette, C.; Huynh-Dinh, T.; Ghomi, M. *Biochemistry* **1998**, *37*, 7878.

(16) Jucker, F. M.; Pardi, A. *Biochemistry* **1995**, *34*, 14416.

(17) Nakamura, Y.; Fujii, S.; Urata, H. Uesugi, S.; Ikehara, M.; Tomita, K. *Nucleic Acids Res. Symp. Ser.* **1985**, *16*, 29.

(18) Davis, P. W.; Adamiak, R. W.; Tinoco, I., Jr. *Biopolymers* **1990**, *29*, 109.

(19) Pullman, B.; Saran, A. *Progress in Nucleic Acid Research and Molecular Biology* 18; W. E. Cohn; Academic Press: New York, 1976; pp 215−325.

(20) Pullman, A.; Berthod H *Int. J. Quantum Chem: Quantum Biol. Symp.* **1977**, *4*, 327.

(21) MacKerell, A. D., Jr.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946.

(22) Cysewski, P.; Jeziorek, D. *J. Mol. Struct. (THEOCHEM)* **1998**, *430*, 219.

(23) Hernandez, B.; Elass, A.; Navarro, R.; Vergoten, G.; Hernanz, A. *J. Phys. Chem. B* **1998**, *102*, 4233.

(24) Gruza, J.; Koca, J.; Pérez, S.; Imberty, M. *J. Mol. Struct. (THEOCHEM)* **1998**, *424*, 269.

(25) Brameld, K. A.; Goddard, W. A., III *J. Am. Chem. Soc.* **1999**, *121*, 985.

(26) Dejaegere, A. P.; Case, D. A. *J. Phys. Chem. A* **1998**, *102*, 5280.

(27) Graña, A. M.; Ríos, M. A. *J. Mol. Struct. (THEOCHEM)* **1995**, *334*, 37.

(28) Gavira, J. M.; Campos, M.; Diaz, G.; Hernanz, A.; Navarro, R. *Vibrat. Spectrosc.* **1997**, *15*, 1.

(29) Estévez, C. M.; Graña, A. M.; Ríos, M. A. *J. Mol. Struct. (THEOCHEM)* **1993**, *288*, 207.

(30) Foloppe, N.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **1998**, *102*, 6669.

(31) Sponer, J.; Leszczynski, J.; Hobza P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117.

(32) Aamouche, A.; Ghomi, M.; Coulombeau, C.; Jobic, H.; Grajcar, L.; Baron, M. H.; Baumruk, V.; Turpin, P. Y.; Henriet, C.; Berthier, G. *J. Phys. Chem.* **1996**, *100*, 5224.

(33) Aamouche, A.; Ghomi, M.; Coulombeau, C.; Grajcar, L.; Baron, M. H.; Jobic, H.; Berthier, G. *J. Phys. Chem. A* **1997**, *101*, 1801.

(34) Aamouche, A.; Ghomi, M.; Grajcar, L.; Baron, M. H.; Romain, F.; Baumruk, V.; Stepánek, J.; Coulombeau, C.; Jobic, H.; Berthier, G. *J. Phys. Chem. A* **1997**, *101*, 10063.

(35) Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1998**, *102*, 6010.

(36) Smets, J.; Adamowicz, L.; Maes, G. *J. Phys. Chem.* **1996**, *100*, 6434.

(37) Berthier, G.; Cadioli, B.; Gallinella, E.; Aamouche, A.; Ghomi M. *J. Mol. Struct. (THEOCHEM)* **1997**, *390*, 11.

(38) Parr, R. G.; Yang, W. *Annu. Rev. Chem.* **1995**, *46*, 701.

(39) St-Amant, A. Density Functional Methods in Biomolecular Modeling, *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds; VCH Publishers: New York, 1996; Vol 7.

(40) Aamouche, A.; Berthier, G.; Cadioli, B.; Gallinella, E.; Ghomi, M. *J. Mol. Struct. (THEOCHEM)* **1998**, *426*, 307.

(41) Ghomi, M.; Aamouche, A.; Cadioli, B.; Berthier, G.; Grajcar, L.; Baron, M. H.; *J. Mol. Struct.* **1997**, *411*, 323.

(42) Cadioli, B.; Gallinella, E.; Coulombeau, C.; Jobic, H.; Berthier, G. *J. Phys. Chem.* **1993**, *97*, 7844.

(43) Saenger, W. *Principles of Nucleic Acid Structure*; Cantor, C. R., Ed; Springer-Verlag: New York, 1984.

(44) Florián, J.; Baumruk, V.; Strajbl, M.; Bednárová, L.; Stepánek, J. *J. Phys. Chem.* **1996**, *100*, 1559.

(45) Florián, J.; Stajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959.

(46) Guan, Y.; Choy, G. S. C.; Glaser, R.; Thomas, G. J., Jr. *J. Phys. Chem.* **1995**, *99*, 12054.

(47) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

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

(49) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(50) Zeiss, G. D.; Scott, W. R.; Suzuki, N.; Chong, D. P.; Langhoff, S. R. *J. Mol. Phys.* **1979**, *37*, 1543.

(51) El Bakali Kassimi, N.; Tadjeddine, M.; Flament, J. P.; Berthier, G.; Gervais, H. P. *J. Mol. Struct. (THEOCHEM)* **1992**, *254*, 177.

(52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1995.

(53) Scalmani, G.; Moro, G.; Cosentino, U.; Pitea, D. Parallel execution of Gaussian 94 using PVM and MPI as message passing tools, *Proceedings of the III National Congress of Computer Science in Chemistry*; Naples (Italy), 1997.

(54) (a) *The International Journal of Supercomputing Applications and High Performance Computing*, **1994**, vol. 8. (b) For further information, see the MPI homepage at http://www.mcs.anl.gov/mpi/.

(55) (a) Blackford, L. S. et al. *ScaLAPACK Users' Guide*; SIAM: Philadelphia, PA, 1997. (b) For further information, see the ScaLAPACK homepage at http://www.netlib.org/scalapack/.

(56) Desiraju, G. R. *Acc. Chem. Res.* **1996**, *29*, 441.

(57) Wahl, M. C.; Sundaralingam, M. *Trends. Biochem. Sci.* **1997**, *22*, 97.

(58) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *274*, 54.

(59) Nishimura, Y.; Tsuboi, M.; Uesugi, S.; Ohkubo, M.; Ikehara, M. *Nucleic Acids Res. Symp. Ser.* **1985**, *16*, 25.

(60) Trulson, M. O.; Cruz, P.; Puglisi, J. D.; Tinoco, I., Jr.; Maties, R. A. *Biochemistry* **1987**, *26*, 8624.

(61) Chastain, M.; Tinoco, I., Jr. *Biochemistry* **1993**, *32*, 14220.

(62) Strobel, S. A.; Cech, T. R. *Biochemistry* **1993**, *32*, 13593.

(63) Setlik, R. F.; Shibata, M.; Sarma, R. H.; Sarma, M. H.; Kazim, A. L.; Ornstein, R. L.; Tomasi, T. B.; Rein, R. *J. Biomol. Struct. Dyn.* **1995**, *13*, 515.