CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.200800961

Theoretical Study on the Factors Controlling the Stability of the Borate Complexes of Ribose, Arabinose, Lyxose, and Xylose

Judit E. Šponer,*[a] Bobby G. Sumpter,[b] Jerzy Leszczynski,[c] Jiří Šponer,[a] and Miguel Fuentes-Cabrera*[b]

Abstract: Recent experimental studies suggest that complexation with borate minerals stabilizes ribose, and that the borate complex of ribose is more stable than those of related aldopentoses, that is, arabinose, lyxose, and xylose. These findings have revived the debate on the plausibility of the RNA-world theory, because they provide an explanation for the stabilization and selection of ribose in prebiotic conditions. In this paper we unravel the factors that make the ribose–borate complex the most stable one. For this purpose, we have investigated the structure and stability of the ribose–, arabinose–, lyxose–, and xylose–borate complexes using density functional theory and a continuum solvent approach. The computed results reveal that in the aldopentose–borate complexes, the electrostatic field of the borate is strong enough to change the orientation of the nearby hydroxyl groups compared to noncomplexed aldopentoses. In addition, we show that the distinct stability of the ribose–borate 2:1 complex can be attributed to 1) a strong hydrogen bond between the ribose 3-OH and one of the negatively charged borate oxygen atoms, and 2) a favorable contact between the aqueous medium and the 5-CH_2OH group due to the space separation between the 5-CH_2OH group and the borate anion.

Keywords: borates · carbohydrates · density functional calculations · genetic material · RNA

Introduction

The term "RNA world" stands for the theory that presumes that RNA played a critical role in the origin of life,[1,2] and it hypothesizes that, on early earth, RNA functioned both as a carrier of genetic information and as an enzyme that catalyzed the synthesis of other RNAs. Although the theory received considerable impetus after catalytic RNAs were discovered,[3,4] its plausibility remains controversial. For example, one of the main objections is that D-ribose, a component of RNA, could not have accumulated under prebiotic conditions.

Ribose and related aldopentoses (arabinose, lyxose, and xylose), are formed during the formose reaction,[5] but it is debatable whether this reaction could have led to the accumulation of ribose on early earth.[6,7] New findings, however, have cast a different light on this debate: in reference [8], Ricardo et al. found that borate minerals do not interfere with the formose reaction and, more importantly, that they stabilize ribose. As borate minerals are not excluded from early earth, Ricardo et al. argued that the accumulation of ribose could not be ruled out either. Subsequently, Li et al.[9] found that ribose has the strongest affinity for boron, followed by lyxose, arabinose, and xylose. This implies that in the presence of borate minerals, nature would have a preference for accumulating ribose. The findings in references [8,9] provide a way to stabilize and select ribose in prebiotic conditions, and in turn remove one of the major objections to the RNA-world theory. However, questions remain,

[a] Dr. J. E. Šponer, Prof. J. Šponer
Institute of Biophysics
Academy of Sciences of the Czech Republic
Královopolská 135, 61265, Brno (Czech Republic)
Fax: (+420) 541-211-293
E-mail: judit@ncbr.chemi.muni.cz

[b] Dr. B. G. Sumpter, Dr. M. Fuentes-Cabrera
Center for Nanophase Materials Sciences and
Computer Sciences and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, P. O. Box 2008, Oak Ridge, TN 37831–6494 (USA)
Fax: (+1) 865-241-0381
E-mail: fuentescabma@ornl.gov

[c] Prof. J. Leszczynski
Department of Chemistry
Computational Center for Molecular Structure and Interactions
Jackson State University, Jackson, Mississippi 39217 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200800961.

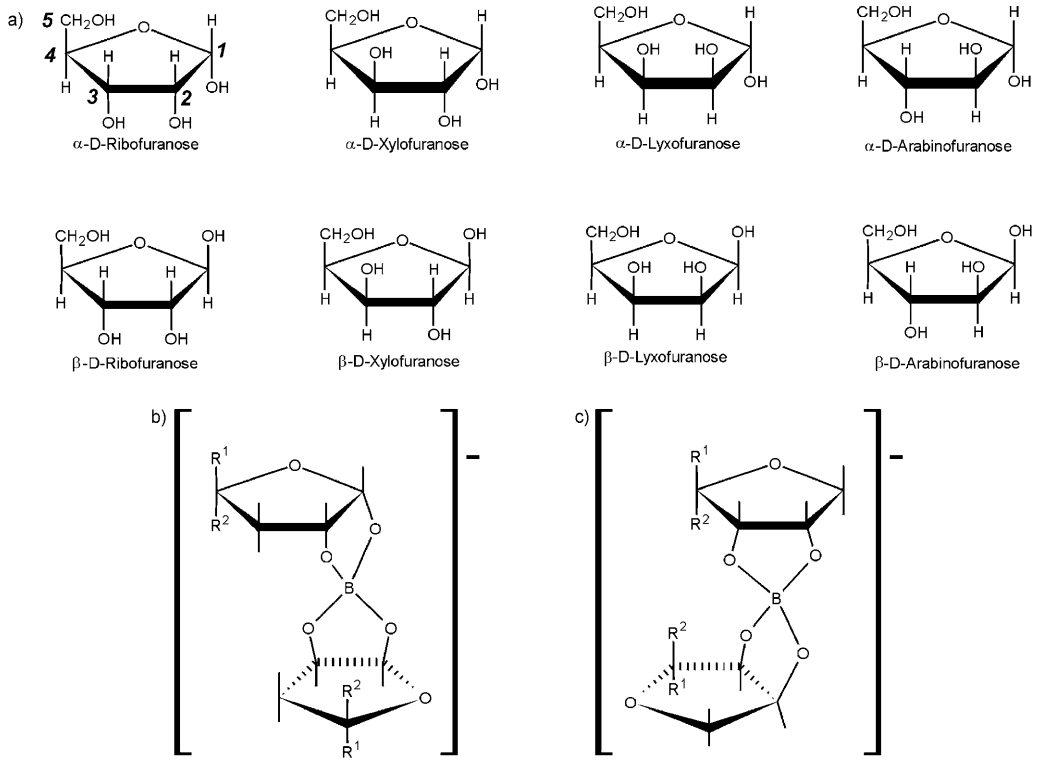


Figure 1. Aldopentoses of the α- and β-series (a) may bind to borate either in the 1,2 (b) or 2,3 (c) fashion.[9,10] $R^1=CH_2OH$, $R^2=H$ for ribose and xylose. $R^1=H$, $R^2=CH_2OH$ for arabinose and lyxose. The numbers in bold italics on the top left structure show the numbering of the ring positions according to the carbohydrate nomenclature.

the most apparent one being why ribose has the strongest affinity for boron.

Figure 1a shows the furanose form of the alpha and beta isomers of ribose, arabinose, lyxose, and xylose. The binding of two aldopentose molecules and B through the oxygen atoms denoted as O1,O2 as well as O2,O3 are shown in Figure 1b and c, respectively. Ribose and lyxose can bind to B through O1,O2 or O2,O3. The beta isomer of arabinose, as well as the alpha isomer of xylose, can bind B through O1,O2. These two aldopentose molecules do not bind to borate through O2,O3, because their O2 and O3 atoms are in the *trans* position. In Figure 1b and c the aldopentose complexes are homogeneous, that is, these complexes involve only one type of aldopentose.

In reference [8], Ricardo et al. assumed that the aldopentoses are in the furanose form, but then they provided only tentative structures for their homogeneous borate complexes. The study by Chapelle and Verchere,[10] on the other hand, contains more structural information on homogeneous aldopentose–borate complexes (no information on heterogeneous complexes is provided). They investigated the structure and stability constants of these complexes in aqueous solution, and found that the aldopentose is always present in a furanose form; for ribose and lyxose, the complexes appear as a mixture of species, bound either through O1,O2 or O2,O3. They suggest that ribose and xylose form considerably more stable borate complexes than arabinose and lyxose. Although references [9,10] report different relative stabilities of the aldopentose–borate complexes, they agree that the borate complex of ribose is the most stable one.

The experimental information is consistent in that the ribose–borate complex is the most stable one, but there is no explanation as to why this is the case. In addition, instead of the *absolute* stability of the borate complexes, both studies measure the stabilites *relative* to the free pyranose form of the pentose molecules. It is, however, necessary to point out that the primary product of the borate-assisted pentose synthesis suggested by Ricardo et al. is not the free pentose, but the pentose–borate complex. As all complexes are formed from the same starting material, the free-energy change upon pentose–borate complex formation is dictated by the absolute stability of the product, that is, the pentose–borate complex.

In the current work, for the first time, we attempt to determine the order of absolute stabilities of the pentose–borate 2:1 complexes. For this purpose we use quantum-chemical calculations, which derive the absolute stability of these complexes directly from their wavefunction. In contrast, experimental methods, based on the evaluation of thermodynamic equilibria, always measure relative energies. Thus, unless using the same reference compound, which, in most cases, is not affordable, experiments do not measure directly the order of absolute stabilities. A further advantage of the theoretical calculations is that, in addition to the energies, they simultaneously provide tentative structural models for these compounds, thus directly relating structures and

energies. In this way a panoramic understanding of the structure and stability of the aldopentose–borate 2:1 complexes can be achieved. The objective of the current paper is thus to provide a complex theoretical characterization of the pentose–borate 2:1 complexes with a special emphasis on the stabilities and structures of these compounds.

Computational Methods

All computations were carried out at the B3LYP level of theory[11–13] with the Gaussian 03[14] computer code. The B3LYP functional is a widely used density functional in quantum chemistry, which performs very well for a broad spectrum of chemical problems. The 6-31G** basis set was used for geometry optimization. Relative energies were evaluated utilizing the considerably larger 6-31++G(2d,2p) basis set on the 6-31G** optimized geometries. A similar approach was used in reference [15]; reference [16] shows that single-point calculations at the B3LYP/6-31++G(2d,2p)//B3LYP/6-31G** level provide very good energy estimates for carbohydrates, because of the very effective reduction of the intramolecular basis set superposition error[17] artifact. Thus, due to compensation of errors, the level of calculations used here is optimally tuned for the purpose of our investigation.

We use the following energy units throughout the paper: total electronic energies are given in Hartree, while relative energies are in kcal mol^{-1}. The two energy units can be interconverted as follows: 1 Hartree = 627.509 kcal mol^{-1}.

The presence of the aqueous medium ($\varepsilon = 78.4$) was accounted for with the COSMO dielectric continuum method.[18,19] The united atom topological model and a scaling factor of 1.2 were used to define the atomic radii. As the geometries of the aldopentose–borate complexes were optimized in solution, we could not carry out zero-point and entropy corrections of the computed energy data.

Analysis of the molecular orbitals was performed on the B3LYP/6-31++G(2d,2p) wavefunction with the NBO3.0 program[20] implemented in the Gaussian 03 code.

Results and Discussion

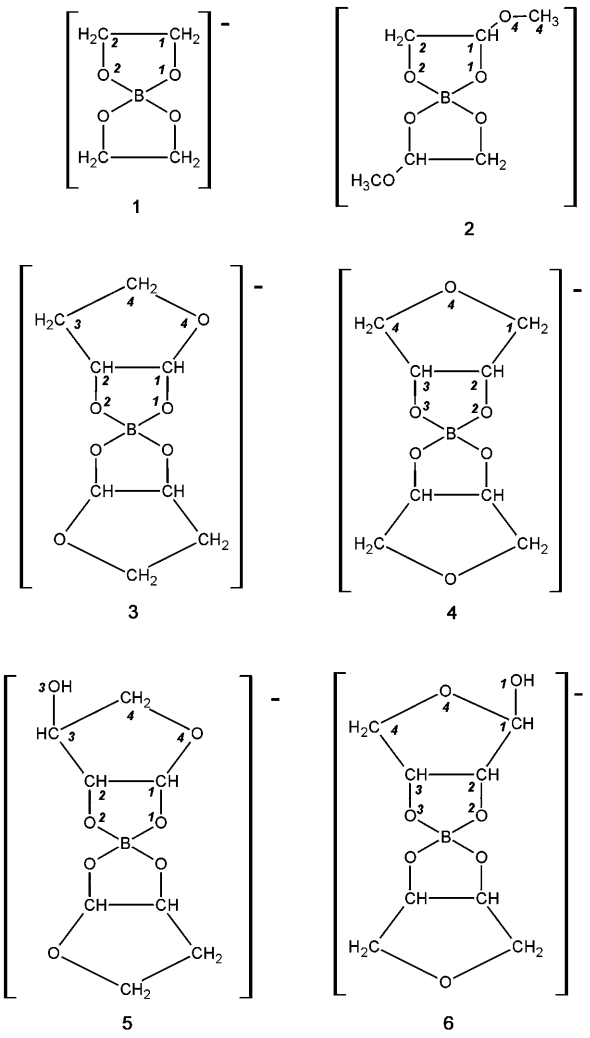
To provide reliable energy estimates, one needs correct models of the aldopentose–borate complexes in solution. The precise molecular structures of these complexes are not known. Therefore, we first had to propose plausible model geometries for these systems. This was done in a stepwise manner by exploiting the available knowledge on the stability of carbohydrates in the gas phase and in solution.

Anomeric and steric effects are the most important factors in determining the equilibrium conformation of the furanose ring of gas-phase aldopentoses.[21,22] We first investigate how these factors affect the conformation of the furanose ring and its mode of binding to the borate anion. These calculations were carried out in the gas phase.

Then we consider the effect of hydrogen-bonding interactions, which is of primary importance in determining the stability of carbohydrates both in aqueous solution[15] and in the gas phase.[21] Whereas in the gas phase the carbohydrates tend to establish as many internal hydrogen bonds as possible,[21] in aqueous solutions the water molecules break the intramolecular hydrogen bonds.[15] If no other effects are present, the OH groups of the carbohydrate are oriented to ensure a favorable contact with the solvent water molecules.[15] Nevertheless, in the aldopentose–borate complexes the negatively charged borate anion may compete with the water molecules and thus change the orientation of the hydroxyl groups of the aldopentoses as compared to their orientation adopted in noncomplexed form. All these alternatives will be tested and accounted for upon constructing the models.

Thus, to suggest plausible models for the aldopentose–borate 2:1 complexes we will consider stereoelectronic and steric effects in the gas phase as well as the effect of hydrogen bonding in solution. We begin with the description of the stereoelectronic effects, since they are molecular orbital effects and therefore will be largely unchanged even in the presence of a polar solvent.[23]

Hyperconjugation—determining the conformation of the furanose ring in the aldopentose–borate complexes: The six models that we used to investigate how the conformation of the furanose ring is affected by stereoelectronic and steric effects are depicted in Scheme 1. Starting from the 2:1



Scheme 1.

borate complex of ethylene glycol (**1**), we gradually extended our model up to the furanose–borate 2:1 complexes. At each step of the model extension, we selected a model that enabled the maximum extent of hyperconjugative stabilization of the complex. Models with the maximum extent of hyperconjugation could be easily proposed, knowing that in aliphatic R-O-C-X linkages, in which X is an electronegative atom or functional group (e.g., OH), the C–X group is *gauche* with respect to R–O. This is because of the favorable stereoelectronic orbital interaction between the antibonding C–X and nonbonding O orbitals.[24]

In this way the following complexes were derived from the 2:1 borate complex of ethylene glycol (**1**). Replacing one of the hydrogen atoms of each ethylene glycol moiety of **1** with a methoxy group, but keeping the C_2 symmetry, resulted in model **2**. From the optimized geometry of **1** and **2**, we constructed two series of models representing binding of the furanose rings to the borate anion through O1,O2 or O2,O3. In **3**, the etheric O is geminal to one of the borate O atoms, and it describes the situation when B and the pentose are bound through O1,O2. In **4**, the etheric O is vicinal to both borate oxygen atoms and it represents the O2,O3 binding. From the most stable geometries of **3** and **4**, we constructed models **5** and **6**, respectively. Model **5** contains an OH group connected to C3; **6** has the OH group connected to C1. All the models were optimized in the gas phase at the B3LYP/6-31G** level. The optimized geometries are depicted in Figure 2.

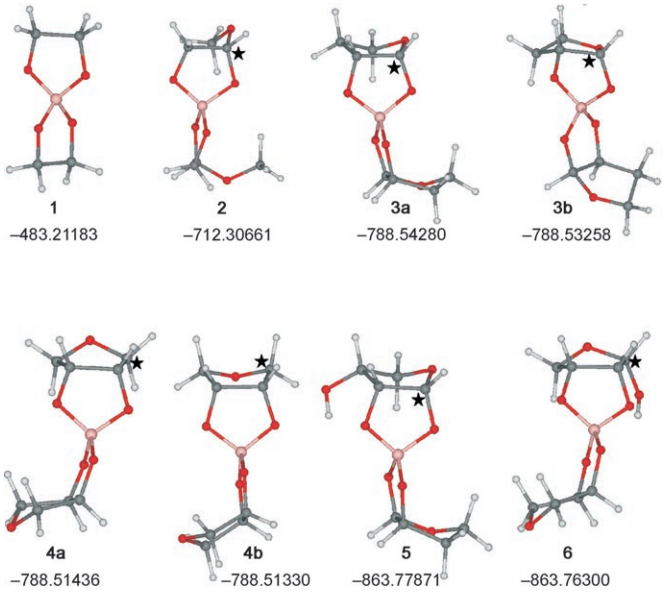


Figure 2. Gas-phase B3LYP/6–31G** optimized geometries of complexes **1**–**6**. Atom colors: C-dark grey; O-red; H-light grey; B-pale pink. The numbers refer to the computed total electronic energies (Hartree) obtained at the B3LYP/6-31++G(2d,2p) level using the B3LYP/6-31G** optimized geometries. 1 Hartree = 627.509 kcal mol^{-1}. Black asterisks indicate the positions of the C1 atoms.

After the geometrical optimizations we evaluated the stereoelectronic effects with the natural bond orbitals (NBO, see reference [20]) analysis. This method allows the tracking down of stereoelectronic effects in complex systems in an accurate way.[25] The multicentered orbitals were projected into localized bond orbitals, and the stabilization energy brought about by donor–acceptor interactions between localized bond orbitals was quantified by the second-order stabilization energies. These energies are derived from the nondiagonal elements of the Fock matrix expressed on the basis of natural bond orbitals.[26] Table 1 summarizes the computed second-order stabilization energies for the most important orbital interactions in compounds **1**–**6**.

In the optimized geometry of **1**, the dihedral angle between the vicinal B–O1 and C1–C2

Table 1. The second-order stabilization energies due to the delocalization of the donor orbital into the acceptor orbital (E_{ij}, kcal mol^{-1}) characterizing the main donor–acceptor interactions between localized NBO orbitals.[26]

Model	Donor→acceptor[a]	E_{ij}	Model	Donor→acceptor[a]	E_{ij}
1	n(O1)→σ*(C1–H1)	10.9	**4b**	n(O3)→σ*(C3–C4)	8.8
				n(O3)→σ*(C3–H3)	7.2
2	n(O1)→σ*(C1–H1)	4.3, 1.2[b]		n(O4)→σ*(C4–C3)	2.9, 1.8[b]
	n(O1)→σ*(C1–O4)	19.2		n(O4)→σ*(C4–H4)	7.3
	n(O4)→σ*(C1–O1)	9.3		n(O2)→σ*(C1–C2)	8.7
	n(O2)→σ*(C2–H2)	10.2		n(O2)→σ*(C2–H2)	7.7
				n(O4)→σ*(C1–C2)	2.9, 1.8[b]
3a	n(O1)→σ*(C1–H1)	4.6, 1.2[b]		n(O4)→σ*(C1–H1)	7.4
	n(O1)→σ*(C1–O4)	20.2			
	n(O4)→σ*(C1–O1)	12.7	**5**	n(O1)→σ*(C1–H1)	5.0, 1.1[b]
	n(O2)→σ*(C2–C3)	9.0		n(O1)→σ*(C1–O4)	18.7
	n(O2)→σ*(C2–H2)	5.9, 0.8 [b]		n(O4)→σ*(C1–O1)	12.9
				n(O2)→σ*(C2–C3)	8.8
3b	n(O1)→σ*(C1–H1)	6.9, 0.9[b]		n(O2)→σ*(C2–H2)	3.9, 1.6[b]
	n(O1)→σ*(C1–O4)	17.2		n(O3)→σ*(C2–C3)	3.1, 2.0[b]
	n(O4)→σ*(C1–O1)	9.1		n(O3)→σ*(C3–C4)	9.9
	n(O2)→σ*(C2–C3)	11.3			
	n(O2)→σ*(C2–H2)	3.5, 1.4[b]	**6**	n(O3)→σ*(C3–C4)	11.5
				n(O3)→σ*(C3–H3)	1.5, 3.3[b]
4a	n(O3)→σ*(C3–C4)	11.7		n(O4)→σ*(C3–C4)	1.4, 3.4[b]
	n(O3)→σ*(C3–H3)	3.9, 1.2[b]		n(O4)→σ*(C4–H4)	6.8
	n(O4)→σ*(C4–C3)	3.8, 1.2[b]		n(O2)→σ*(C1–C2)	8.7
	n(O4)→σ*(C4–H4)	6.6, 0.6[b]		n(O2)→σ*(C2–H2)	5.7, 1.2[b]
	n(O2)→σ*(C1–C2)	9.4		n(O4)→σ*(C1–C2)	2.1, 2.1[b]
	n(O2)→σ*(C2–H2)	6.5, 0.7[b]		n(O4)→σ*(C1–O1)	14.0
	n(O4)→σ*(C1–C2)	3.1, 1.6[b]		n(O1)→σ*(C1–O4)	17.0
	n(O4)→σ*(C1–H1)	6.9		n(O1)→σ*(C1–H1)	4.8, 0.7[b]
				n(O1)→σ*(C1–C2)	3.5

[a] For atom numbering refer to Scheme 1. [b] Two numbers refer to two different lone pair (donor) orbitals centered on the same oxygen atom.

bonds (for atom numbering refer to Scheme 1) is 26°, which enables a stabilizing n→σ* hyperconjugation between the O1 lone pair and the vicinal C1−H1 bond located in the antiperiplanar position. This orbital interaction has a second-order stabilization energy of 10.9 kcal mol^{-1}.[27]

Model **2** was obtained from **1** by replacing one of the H atoms[28] at C1 with a methoxy group (see Scheme 1). In **2**, the methyl group is antiperiplanar to H1 in order to ensure an orientation of the C1-O4-C4 linkage similar to that in the furanose ring. As a result of the attachment of the methoxy group, a new hyperconjugation arises between the C1−O4 antibond and the lone pair of O1; this hyperconjugation contributes 19.2 kcal mol^{-1} to the stability of **2** (see Table 1). In the reverse direction, there is an electron donation from O4 to the C1−O1 antibond, which results in a slightly weaker (9.3 kcal mol^{-1}) second-order stabilization energy.

By replacing one of the hydrogen atoms of the terminal C4 methyl group in **2** with a $-CH_2-$ moiety we arrive at a model representing the binding of the furanose ring to the borate through O1,O2 (structure **3a** in Figure 2). In **3a**, the C4−O4 bond is synclinal to the C1−C2 bond and all stabilizing effects discussed above are preserved. In addition, a new (albeit less stabilizing) orbital interaction arises between the lone pair of O2 and the C2−C3 antibond (see Table 1) as a consequence of ring closure. Due to the rigidifying effect of the strained chelate ring formed with the borate, the conformation of **3a** is the only one that enables optimum conditions for a strong n(O4)→σ*(C1−O1) hyperconjugation. We tested this by creating another furanose–borate 2:1 complex, denoted in Figure 2 as model **3b**, by moving the C4−O4 bond into synperiplanar position with respect to the vicinal C1−C2 bond. As shown in Table 1, on going from **3a** to **3b** the n(O1)→σ*(C1−O4) contribution is reduced from 20.2 to 17.2 kcal mol^{-1}, while the n(O4)→σ*(C1−O1) term decreases from 12.7 to 9.1 kcal mol^{-1}. In line with this, **3b** is destabilized with respect to **3a** by about 6.4 kcal mol^{-1}.

Model **4** represents the situation in which the borate and the furanose rings are bound through O2,O3. We created two models (**4a** and **4b**) for this type of binding considering the following points. In models in which the borate anion and furanose rings are bound through O2,O3 the hyperconjugation between the borate O (either O2 or O3) and O4 is absent, because these two atoms are separated by two carbon atoms. On the other hand, remarkable stereoelectronic orbital interactions between the borate O (O2,O3) lone pairs and the adjacent C−H and C−C bonds are expected, as the steric hindrance introduced by B causes the lone pair of the borate oxygen atoms (O2,O3) to be in an optimal position to donate electrons to the adjacent C−H or C−C antibonds. Favorable donor–acceptor orbital interactions between the lone pair of the etheric O (O4) and the adjacent C−C and C−H bonds in **4a** and **4b** requires a staggered conformation between the C−O4 and the geminal C−H and C−C bonds. Because of the geometrical restraints imposed by the five-membered chelate ring formation with borate, the staggered conformation can be achieved in only two different ways: by placing the etheric O above or below the plane defined by the other four ring atoms (i.e., *anti* or *syn* to the borate, respectively). Models **4a** and **4b**, the optimized structures of which are shown in Figure 2, represent the *anti* and *syn* orientations, respectively. As expected, the orbital interactions in **4a** and **4b** are not very different from each other, and this is also reflected in their very similar total energies. In both complexes the leading stabilization term is the n→σ* hyperconjugation between the lone pair of borate oxygens (O2,O3) and the adjacent C−C bonds, contributing around 9–12 kcal mol^{-1} to the stability of the complex. This value, however, is far less than the n→σ* hyperconjugation between the lone pair of the borate O (O1) and the geminal C1−O4 bond in models **3a** and **3b** (20.2 and 17.2 kcal mol^{-1}, respectively, see Table 1). Similarly, hyperconjugation between the lone pair of O4 and the adjacent C1−O1 bond in **3a** and **3b** is considerably stronger (12.7 and 9.1 kcal mol^{-1}, respectively) than that between the O4 lone pair and the geminal C−C bonds in **4a** and **4b** (4.7–5.0 kcal mol^{-1}). This weakening of the donor–acceptor orbital interactions is reflected by the stability of **4a** and **4b** being approximately 18 kcal mol^{-1} lower than that of **3a**.

We have also considered the stereoelectronic effect of the free (i.e., noncomplexed) hydroxyl groups of the aldopentoses. In the 1,2-borate complexes of aldopentoses the additional OH is bound to C3 and thus it is not in direct contact with C−O bonds or oxygen lone pairs. In contrast, due to the presence of the anomeric hydroxyl group in the 2,3-complexes, one has to count on hyperconjugation between the lone pair of O1 and the C1−O4 bond. This orbital interaction is analogous to the strongly stabilizing n→σ* hyperconjugation between the lone pair of O1 and the geminal C1−O4 bond in **3a** and **3b**. Thus, it is reasonable to expect that addition of the OH groups to the studied pentose–borate complexes results in a more substantial stabilization of **4** (representing borate binding through O2,O3) than **3** (binding through O1,O2). To check this, we constructed two models (**5**, **6**) by adding an OH group to C3 and C1 of **3a** and **4a**, respectively. In **6**, the initial position of the added hydroxyl was selected in such a way as to enable the maximum extent of hyperconjugation between the lone pair of the added hydroxyl and the adjacent C1−O4 bond. Upon geometry optimization, however, a hydrogen bond was formed between one of the borate oxygen atoms and the added OH, slightly displacing the OH group from its initial position. In **5**, the added OH was oriented in a similar position, to enable hydrogen-bond formation with the adjacent borate oxygen. In this complex the added OH is not in contact with any oxygen atoms separated by a single C atom; therefore here the 3-OH group does not participate to a significant extent in stereoelectronic interactions. In contrast, this model truly captures the neat effect of the hydrogen bonding with the borate oxygen on the stability of **5** and **6**. Thus, the energy difference between **5** and **6** reflects the effect brought about by the hyperconjugation between the lone pair of O1 and the C1−O4 bond. The optimized geometries of complexes **5** and **6** are depicted in Figure 2. Indeed, the energy difference between **5** and **6** is 9.9 kcal mol^{-1},

which is half of the energy difference between **3a** and **4a** (17.8 kcal mol^{-1}). As anticipated, in **6** the main stabilizing effect is the hyperconjugation between the lone pair of the hydroxyl oxygen and the adjacent C1–O4 bond, which amounts to 17.0 kcal mol^{-1}. This structure is further stabilized by another hyperconjugation between the lone pair of O4 and the C1–O1 antibond, contributing 14.0 kcal mol^{-1} to the stability (see Table 1). These two effects are quite substantial, but in total are still noticeably smaller than those resulting from the analogous orbital interactions of the O1 oxygen of **3a** (20.2 and 12.7 kcal mol^{-1} in **3a**). Overall, **5** is still 9.9 kcal mol^{-1} more stable than **6**.

Since the 5-OH of the aldopentoses is separated from the nearest oxygen atoms by at least two C atoms, it is reasonable to expect that hyperconjugation between the CH_2OH group and the furanose ring is fairly weak. To check this we added a CH_2OH group to the C4 of **3a** both in *syn* and *anti* orientation with respect to the borate. In the optimized models the strongest orbital interaction was found between the lone pair of O4 and the C4–C5 bond of the *syn* model, which contributes to the stabilization of the complex by about 6.2 kcal mol^{-1}. All other hyperconjugation effects are even weaker. Since the analogous hyperconjugation effects brought about by the OH groups are about 2–3 times stronger (about 12–20 kcal mol^{-1}), ignoring the hyperconjugation between the 5-CH_2OH group and the furanose rings seems to be a reasonable approach.

In summary, the gas-phase results above show that for the furanose–borate complexes bound through O1,O2 and O2,O3 the most stable furanose conformations are those given by models **3a** and **4a**, respectively. On the other hand, the furanose–borate complex bound through O1,O2 is more stable than that bound through O2,O3. In what follows, we use model **3a** to build the O1,O2-aldopentose–borate complexes. Specifically, to make the O1,O2-aldopentose–borate complexes, we connect an OH group to C3 of **3a** and a CH_2OH group to C4.

Solvent effects—determining the structure of the aldopentose–borate complexes: In the previous section we found that models **3a** and **4a** contain the most stable furanose ring conformations. In this section, we describe how we used models **3a** and **4a** to build the aldopentose–borate complexes of ribose, arabinose, lyxose, and xylose. These complexes were then optimized in water within the continuum solvent approximation.

In the O1,O2 borate complexes of aldopentoses the positions of the 3-OH and 5-CH_2OH groups are as follows: in ribose and lyxose the 3-OH is *syn* to the borate; in xylose and arabinose the 3-OH is in *anti* position with respect to the borate. In ribose and xylose the 5-CH_2OH is *anti* to the borate; in lyxose and arabinose the 5-CH_2OH group is *syn* to the borate. Nevertheless, this does not say anything about the conformation of the 3-OH and 5-OH groups with respect to the bonds connected to C3 and C5, respectively. Since the stereoeletronic effects associated with the 3-OH and 5-OH groups are insignificant, the directionality of these groups is determined primarily by the solvation effects. Kirschner and Woods have shown that the CH_2OH group of methyl α-D-glucopyranoside and methyl α-D-galactopyranoside does not form an internal hydrogen bond with the adjacent etheric O, but is instead fully exposed to the solvent water molecules, similar to all other hydroxyl groups present in carbohydrates.[15] This applies to neutral solutes. The borate complexes of aldopentoses carry a charge of −1, which dramatically changes the situation. In the neutral aldopentoses the strength of the hydrogen bonds formed with neutral water molecules is similar to that of the internal hydrogen bonds in the neutral carbohydrates. In contrast, in the borate complexes the equivalence of the internal and intermolecular hydrogen bonds is lost, due to the presence of the negatively-charged borate oxygens. Thus, it is reasonable to expect that in the borate complexes, hydroxyl groups lying in the vicinity of the negative charge concentration will prefer intramolecular hydrogen bonds rather than intermolecular ones with the solvent. To justify this assumption we carried out two sets of calculations.

In the first set of calculations, we replaced one of the H atoms at C3 of **3a** with an OH group, and created four models (**T1**–**T4**). These models differ in: 1) the configuration of the 3-OH group with respect to the borate (i.e., *syn* or *anti* for **T3**,**T4** and **T1**,**T2**, respectively) and 2) the exposure of the 3-OH to the solvent. The optimized geometries of **T1**–**T4** are depicted in Figure 3. The two *anti* complexes (**T1**,**T2**) are very close in energy, but still weakly favor (by

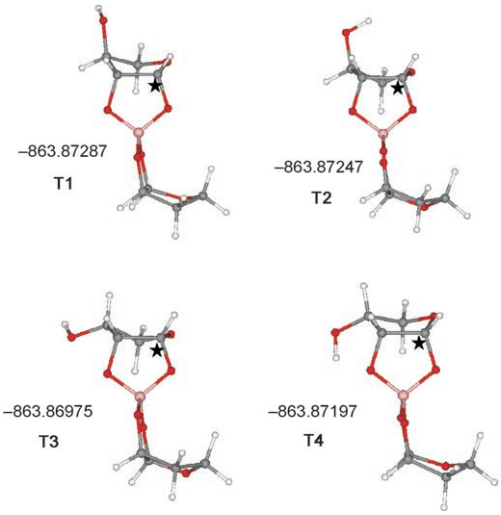


Figure 3. Optimized geometries of the models used to evaluate the effect of the orientation of the 3-OH group. Optimizations were carried out at the B3LYP/6-31G** level of theory using the COSMO dielectric continuum method. Atom colors: C = dark grey; O = red; H = light grey; B = pale pink. In **T1** and **T2** the 3-OH group is *anti* to the borate; in **T3** and **T4** the 3-OH group is *syn* to the borate. In **T1** and **T3** the 3-OH is fully exposed to the solvent; in **T2** the 3-OH points towards O4; in **T4** the 3-OH points towards O2. The numbers refer to the computed total electronic energies (Hartree) obtained at the B3LYP/6-31++G(2d,2p) level using the B3LYP/6-31G** optimized geometries and the COSMO implicit solvent model. 1 Hartree = 627.509 kcal mol^{-1}. Black asterisks indicate the positions of the C1 atoms.

0.3 kcal mol^{-1}) the conformation in which the 3-OH is fully exposed to the solvent. This is in agreement with the findings of reference [15]. In contrast, the *syn* complex with the 3-OH exposed to the solvent (**T3**) is destabilized by 1.4 kcal mol^{-1} with respect to the *syn* complex that has the 3-OH forming a hydrogen bond with O2 (**T4**). This again justifies our assumption that the negative charge of the borate moiety alters the orientation of the vicinal *syn* OH groups.

We then created model **T5** by connecting a CH_2OH group to C4 of **3a**. In **T5**, the OH group of CH_2OH is in *syn* orientation; its H points inwards (i.e., towards the borate); its O is in the energetically favorable *gauche–trans* orientation.[15] We optimized **T5** using the COSMO dielectric continuum approach. The optimized geometry is depicted in Figure 4. Subsequently, we created model **T6** by rotating the

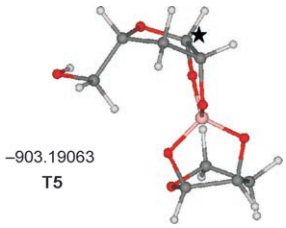


Figure 4. Optimized geometry of the model used to evaluate the effect of the orientation of the 5-OH group (**T5**). (Optimization was carried out at the B3LYP/6-31G** level of theory using the COSMO dielectric continuum method). Atom colors: C = dark grey; O = red; H = light grey; B = pale pink. The number refers to the computed total electronic energy (Hartree) obtained at the B3LYP/6-31++G(2d,2p) level using the B3LYP/6–31G** optimized geometry and the COSMO implicit solvent model. 1 Hartree = 627.509 kcal mol^{-1}. Black asterisk indicates the position of the C1 atom.

OH of **T5** by 180° about the O–C bond. In this way, the OH of **T6** was fully exposed to water. We optimized **T6** in solution by relaxing all the structural parameters except the H-O5-C5-C4 dihedral angle. Comparing the total energies of **T5** and **T6**, we found that **T5** was 1.6 kcal mol^{-1} more stable than **T6**. In other words, the 5-OH group in *syn* position prefers to bind to the borate oxygen rather than to the bulk solvent. This result justifies our assumption for the 5-CH_2OH group.

Thus, in the aldopentose–borate complexes, the conformation of the 3-OH and 5-OH groups with respect to the C4–C5 bond was selected as follows: 1) OH groups *syn* to the borate O were rotated to enable interaction with the borate anion and 2) *anti* OH groups were exposed to the bulk water. In all the models, O5 was *gauche* with respect to O4 and *trans* with respect to C3; this conformation of the CH_2OH group is the most energetically favored in carbohydrate solutions.[15] Then, models constructed in this way were optimized at the B3LYP/6-31G** level of theory in aqueous solution using the COSMO dielectric continuum method.

Relative stability of the aldopentose–borate complexes: The optimized geometries of the most stable 2:1 complexes for ribose, xylose, lyxose, and arabinose are depicted in Figure 5. Their relative energies (kcal mol^{-1}) obtained at the

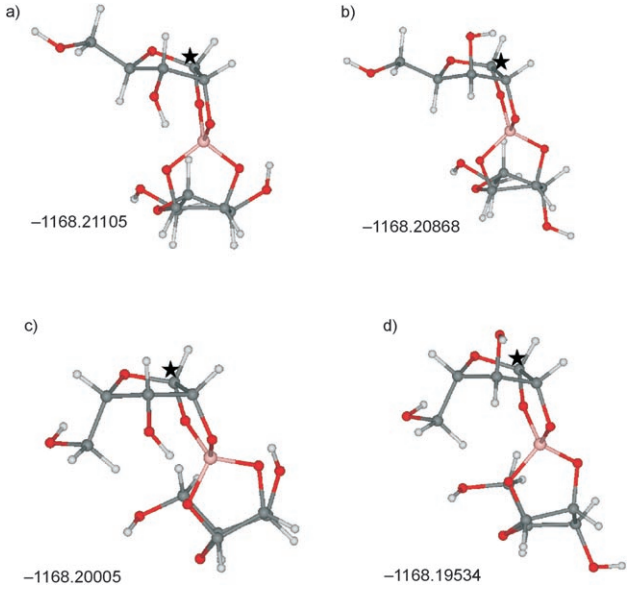


Figure 5. Predicted B3LYP/6–31G** optimized geometries of the most stable aldopentose–borate 2:1 complexes. Optimizations were carried out in solution using the COSMO dielectric continuum model to mimic the aqueous environment. a) ribose, b) xylose, c) lyxose, and d) arabinose. Atom colors: C = dark grey; O = red; H = light grey; B = pale pink. The numbers refer to the computed total electronic energies (Hartree) obtained at the B3LYP/6-31++G(2d,2p) level using the B3LYP/6-31G** optimized geometries and the COSMO implicit solvent model. 1 Hartree = 627.509 kcal mol^{-1}. Black asterisks indicate the positions of the C1 atoms.

B3LYP/6-31++G(2d,2p)//B3LYP/6-31G** level vary as follows: ribose (0) > xylose (1.5) > lyxose (6.9) > arabinose (9.9). This order agrees with that reported by Chapelle and Verchere.[10] Although the stability order suggested by Li et al.[9] differs slightly from that of Chapelle and Verchere,[10] both experimental studies agree that the most stable complex is formed with ribose, which is also supported by our computational results.

In general, the complexes with *anti* mutual orientation of the borate and the 5-CH_2OH (ribose and xylose, see Figure 5a and b) are more stable than their *syn* counterparts (lyxose and arabinose, Figure 5c and d). The energy difference between the complexes carrying the 5-CH_2OH group in *syn* and *anti* position is more than 6.0 kcal mol^{-1}. This is not surprising in view of the electrostatic potential of the furanose–borate 2:1 complexes (Figure 6). In the *syn* form the partially negatively charged O5 is forced to interact with the negatively charged region of the complex, resulting in a repulsive electrostatic contact. In addition to this, steric repulsion between the hydroxymethyl group and the rest of the complex may further destabilize the apparently more crowded *syn* complexes with respect to the *anti* forms.

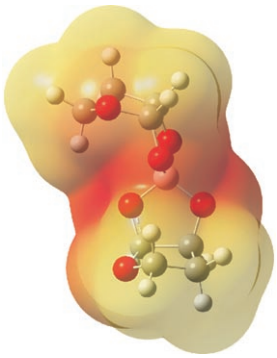


Figure 6. Molecular electrostatic potential calculated for the 0.0004 a.u. isodensity surface of the most stable model composed of two furanose rings and a borate moiety (**3a**). The yellow regions belong to weak and the red regions to strong negative charge accumulation.

For a given configuration of 5-CH_2OH, the stability order is determined by the position of the 3-OH: the *syn* position is the energetically favored one, because of the internal hydrogen-bond formation with the strongly attractive O2; this makes the 2:1 complexes of ribose and lyxose more stable than those of xylose and arabinose, respectively.

For comparison, we computed the stability of the O2,O3 ribose–borate 2:1 complex. Its structure was built up analogously to that of the O1,O2 complex, with the exception that **4a** was used as the furanose ring model instead of **3a**. The 1-OH was added to **3a** in the α-position, due to the energetically favorable hydrogen-bond formation with the borate oxygen at C2. The optimized geometry of the complex is depicted in Figure 7. As expected, the stability of the optimized model is lower than that of the O1,O2 isomer, by 4.4 kcal mol^{-1}.

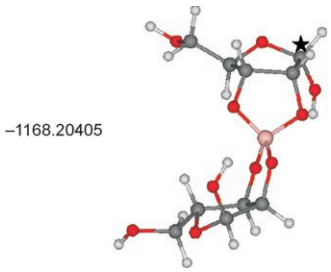


Figure 7. Predicted B3LYP/6–31G** optimized geometry of the most stable ribose–borate 2:1 complex formed through O2,O3 binding. Optimization was carried out in solution using the COSMO dielectric continuum model to mimic the aqueous environment. Atom colors: C = dark grey; O = red; H = light grey; B = pale pink. The number refers to the computed total electronic energy (Hartree) obtained at the B3LYP/6-31++G(2d,2p) level using the B3LYP/6-31G** optimized geometry and the COSMO implicit solvent model. 1 Hartree = 627.509 kcal mol^{-1}. Black asterisk indicates the position of the C1 atom.

Conclusion

In the current paper, for the first time, we have applied electronic structure calculations to predict the structure and determine the stability order of the 2:1 aldopentose–borate complexes of ribose, arabinose, lyxose, and xylose. In contrast to experimental studies[9,10] published prior to this work, our energy data are directly derived from the wavefunction and therefore do not depend on the choice of reference system.

The computed stability order of the aldopentose borate 2:1 complexes is as follows: ribose > xylose > lyxose > arabinose. In general, O1,O2 binding is energetically more favorable than O2,O3 binding, due to the higher ring stability in the O1,O2 form.

Our results clearly show that solution stability of the ribose–borate 2:1 complex formed through O1,O2 binding to the borate is superior to those of other aldopentose complexes. This is attributable to the favorable configuration of the $-CH_2OH$ group, complemented by a hydrogen-bonding contact between the 3-OH group and one of the borate oxygen atoms. Thus, we suggest that the fortuitous interplay of intra- and intermolecular hydrogen bonding, electrostatic, and steric interactions present in the hydrated ribose–borate 2:1 complexes also contributes to the fact that, among the four aldopentoses, ribose had the greatest potential to survive in prebiotic conditions and serve as a building unit for the first RNA architectures.[29]

Acknowledgements

This contribution was supported by the Grant Agency of the Academy of Sciences of the Czech Republic, grants IAA400550701, IAA400040802, 1QS500040581, by the Academy of Sciences of the Czech Republic, grants AV0Z50040507 and AV0Z50040702, the Ministry of Education of the Czech Republic, grant LC06030, as well as NSF Chemical Bonding Center, grant 0739189. Work at Oak Ridge National Laboratory (ORNL) was supported by the Center for Nanophase Materials Sciences, sponsored by the Division of Scientific User Facilities, U.S Department of Energy (USDOE) (BGS, MFC).

[1] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381–393.
[2] W. Gilbert, *Nature* **1986**, *319*, 618.
[3] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.
[4] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
[5] R. Breslow, *Tetrahedron Lett.* **1959**, *1*, 22–26.
[6] R. Shapiro, *Origins Life Evol. Biospheres* **1988**, *18*, 71–85.
[7] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8158–8160.
[8] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 196.
[9] Q. Li, A. Ricardo, S. A. Benner, J. D. Winefordner, D. H. Powell, *Anal. Chem.* **2005**, *77*, 4503–4508.
[10] S. Chapelle, J.-F. Verchere, *Tetrahedron* **1988**, *44*, 4469–4482.
[11] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[12] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[13] B. Miehlich, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* **1989**, *157*, 200–206.
[14] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K.

Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.

[15] K. N. Kirschner, R. J. Woods, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10541–10545.

[16] J.-H. Lii, B. Ma, N. L. Allinger, *J. Comput. Chem.* **1999**, *20*, 1593–1603.

[17] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–566.

[18] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

[19] M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681.

[20] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899–926.

[21] L. P. Guler, Y.-Q. Yu, H. I. Kenttämaa, *J. Phys. Chem. A* **2002**, *106*, 6754–6764.

[22] W. K. Olson, *J. Am. Chem. Soc.* **1982**, *104*, 278–286.

[23] J. E. Šponer, J. Leszczynski, F. Glahe, B. Lippert, J. Šponer, *Inorg. Chem.* **2001**, *40*, 3269–3278.

[24] A. J. Kirby, *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer, Berlin, **1983**.

[25] F. Weinhold, C. Landis, *Valency and Bonding*, Cambridge University Press, Cambridge, **2005**, pp. 215–275.

[26] A. E. Reed, P. von R. Schleyer, *J. Am. Chem. Soc.* **1987**, *109*, 7362–7373.

[27] In Table 1 we have listed the second order stabilization energies for one of the two rings only. For symmetrical structures (**1**–**4**) the corresponding data are almost identical for both rings. In larger complexes (models **5**–**6**), we have added substituents only to one of the rings. This, however, affected the energetics of the main donor–acceptor orbital interactions on the nonsubstituted ring by less than 8% relative to its symmetrical counterpart. All these stabilization energies are negative values; in the text and Table 1 we give them in absolute values.

[28] In what follows, H atoms are numbered according to the number of the C atoms they are connected to; for example, if there are two H atoms connected to C2 then they are labeled as H2 and H2′.

[29] Note added in proof: Stabilization of ribose by complexation with borate was suggested for the first time by B. E. Prieur, *C. R. Chim.* **2001**, *4*, 667–670.

Received: May 20, 2008
Published online: September 22, 2008

