Carbohydrate Research, 206 (1990) 183–191
Elsevier Science Publishers B.V., Amsterdam

Ab initio molecular orbital calculations on furanose sugars: a study with the 6-31G* basis set

Eugenia C. Garrett and Anthony S. Serianni†
Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556 (U.S.A.)

(Received January 17th, 1990; accepted with revisions, March 14th, 1990)

ABSTRACT

Ab initio molecular orbital calculations were performed on 2-deoxy-β-D-*glycero*-tetrofuranose (**1**) using the 6-31G* basis set to evaluate the effect of ring conformation on the molecular parameters (bond lengths, angles, and torsions). Geometric optimizations were conducted on the planar and ten envelope conformers of **1**, and these data were compared to those obtained from previous calculations using the STO-3G and 3-21G basis sets. Conformational energy profiles derived from 3-21G and 6-31G* data were found to be qualitatively comparable. The effect of furanose ring conformation on key bond lengths (*e.g.*, C–H, C–O), bond angles (*e.g.*, COC), and bond torsions (*e.g.*, the exoanomeric C-1–O-1 torsion) was examined, and a qualitative agreement was observed between the 3-21G and 6-31G* analyses. The results indicate that, for semi-quantitative *ab initio* studies of intact carbohydrates, the 3-21G basis set is sufficient, and that the STO-3G basis set should not be employed unless crude structural approximations are desired. The observed concerted behavior of C–O bond lengths in the vicinity of the anomeric carbon of the aldofuranose ring has suggested a possible role of C-1–O-1 bond orientation in affecting the mechanism of glycoside bond hydrolysis.

INTRODUCTION

Furanose sugars are important components of many biologically important compounds, including the nucleic acids, DNA and RNA. Despite their critical role in determining the conformations and dynamics of compounds that contain them, the conformational properties of furanoses are not fully understood. The inherent flexibility of the furanose ring significantly restricts the interpretation of physical constants (*e.g.*, n.m.r. coupling constants) commonly used to characterize the structural properties of molecules in solution[1–3]. In order to gain further insight into the factors that affect the conformational behavior of furanose rings, we have been conducting *ab initio* molecular orbital calculations on intact furanoses of different structures and configurations. In an initial study[4], we performed geometric optimizations on the anomers of the aldotetrofuranoses, α- and β-D-erythrofuranoses, and α- and β-D-threofuranoses, in both their planar and envelope conformations using the STO-3G basis set, with some single-point energy refinement using the 3-21G split-valence basis set. This study

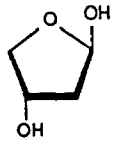
† Author for correspondence.

0008-6215/90/$ 03.50

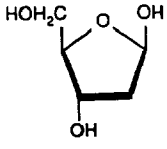
revealed several interesting correlations between molecular parameters (bond lengths, angles and torsions) and ring conformation and showed that *ab initio* calculations may be a valuable tool in studies of furanose ring behavior. However, due to the use of the minimal STO-3G basis set, the observed trends could not be considered firm. Furthermore, some evidence has suggested that the STO-3G basis set may not give reliable conformational energy profiles for some furanose rings (*e.g.*, β-D-erythrofuranose[4]).

To further explore the assets and limitations of these calculations, we recently compared results from geometric optimizations on the anomers of the deoxytetrofuranoses, 2-deoxy-α- and β-D-*glycero*-tetrofuranose and 3-deoxy-α- and β-D-*glycero*-tetrofuranose, using the STO-3G and 3-21G basis sets[5]. While similar overall correlations between structural parameters (*e.g.*, bonds lengths, angles and torsions) and conformation were observed with both basis sets, some significant disparities were observed. Based on the results of this study, we concluded that *ab initio* molecular orbital calculations on intact carbohydrate molecules based on the 3-21G basis set, although still semi-quantitative, appear to be more reliable than those obtained with the STO-3G basis set, and that the former basis set should be considered the smallest reliable basis set for routine molecular orbital calculations on sugars.

Carbohydrates are poly-oxygenated compounds, and thus our previous calculations employing the STO-3G and 3-21G basis sets, which neglect *d*-orbital contributions, may be viewed as approximate. To what extent does the inclusion of *d*-orbitals affect calculated furanose ring geometries and energies? We have addressed this question by studying the geometrically optimized planar and envelope conformers of 2-deoxy-β-D-*glycero*-tetrofuranose (**1**) determined with the 6-31G* basis set, and comparing these structures with those previously[5] determined with the STO-3G and 3-21G basis sets. Compound **1** was chosen for this comparative study because it is biologically relevant [**1** is structurally related to the 2-deoxy-β-D-*erythro*-pentofuranose ring (**2**) of DNA], and because the C-1 and C-3 hydroxyl groups in **1** are *trans*, thus obviating potential intramolecular hydrogen bonding between these groups that might complicate the interpretation of results.



1



2

EXPERIMENTAL

All calculations were conducted as previously[4,5] described with the GAUSSIAN88 program[6] implemented on a Convex C-220 computer at the University of Notre Dame Computing Center, and the data were interpreted within the conformational model defined by the pseudorotational itinerary[7,8] (Fig. 1). Computations were performed on the ten envelope (*E*) forms and planar form of **1** by fixing one (envelope

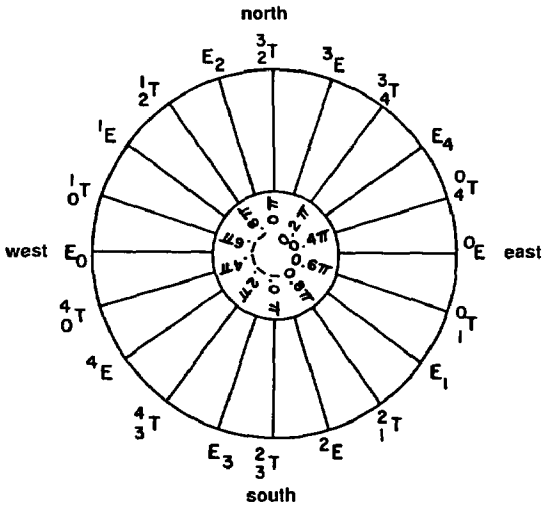


Fig. 1. The pseudorotational itinerary (refs. 7 and 8) describing a pathway for the continuous interconversion of non-planar conformers of aldofuranose rings. Envelope (*E*) conformers are defined as having one ring atom out-of-plane and the remaining four ring atoms co-planar. Twist (*T*) conformers have three ring atoms coplanar, and two ring atoms out-of-plane, with one above and one below the plane. Out-of-plane atoms, and their orientations with respect to the ring plane, are denoted by superscripts (above the plane) and subscripts (below the plane).

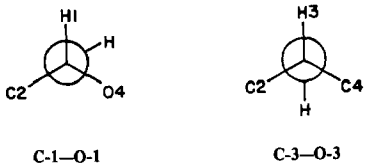


Fig. 2. Initial C-1–O-1 and C-3–O-3 bond torsions of 2-deoxy-β-D-*glycero*-tetrofuranose (1).

forms) or two (planar form) endocyclic torsion angle(s) at 0° (refs. 4 and 5). The C-1–O-1 and C-3–O-3 bond torsions were set initially as shown in Fig. 2, and these initial torsions were permitted to optimize during the calculation. The initial C-1–O-1 torsion was chosen to optimize the "exoanomeric effect"[9a].

The relationship between furanose ring conformation and the phase angle of pseudorotation (*P*)[7,8] is illustrated in Fig. 1, where, for example, the ^{3}E conformation corresponds to $P = 0.1\pi$. To simplify the presentation of data, conformers are identified as P/π, with the ^{3}E conformer corresponding to a value of $P/\pi = 0.1$, E_4 to a value of 0.3, and so forth. As in previous work[4,5], computations were not performed on the twist conformers of **1**. Thus, plots of data acquired on the envelope forms are assumed in this study to be smooth and continuous (*i.e.*, a twist parameter will lie between those of adjacent envelope forms).

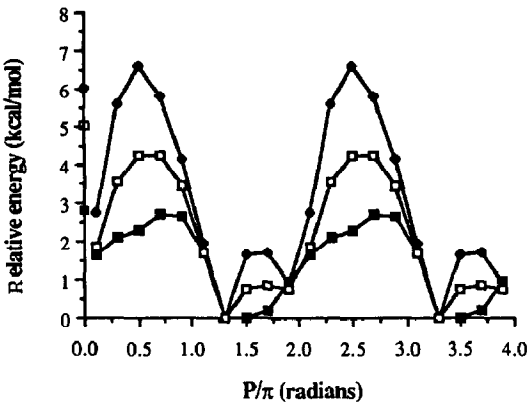


Fig. 3. Conformational energy profiles determined from *ab initio* calculations of **1** using the STO-3G (■), 3-21G (◆) and 6-31G* (□) basis sets. Planar energies are indicated on the *y*-axis.

RESULTS AND DISCUSSION

Conformational energy profiles of **1** obtained with the use of three different basis sets (Fig. 3) are qualitatively similar, but the 3-21G and 6-31G* data reveal additional fine structure (*i.e.*, local minima) not observed in the STO-3G analysis. In addition, the conformational energy curves differ in amplitude, the value of which may be used to approximate the energy barriers of conformer interconversion. The STO-3G and 3-21G data give minimal and maximal amplitudes, respectively, with the 6-31G* amplitude being intermediate in magnitude. Based on the 6-31G* results, the energy difference between the least stable envelope conformer (^0E) and the most stable envelope conformer (^4E) of **1** is 4.25 kcal·mol^{-1}. A local minimum occurs at E_2 (Fig. 3). In addition, ^4E–E_2 interconversion occurs *via* the E_0 and ^1E conformers (west conformers), and the energy barrier is low (<0.9 kcal·mol^{-1}). In contrast, interconversion *via* the east (*e.g.*, ^0E) conformers is characterized by a ~4-fold greater energy barrier (3.4 kcal·mol^{-1}). It is interesting to note that the total energy of the planar form is greater than that of all envelope forms of **1** in STO-3G and 6-31G* calculations. In contrast, the planar form is not found to be the least stable form when the 3-21G basis set is used (Fig. 3).

The effect of basis set on the calculated puckering amplitudes of **1** (Fig. 4) shows that the 3-21G and 6-31G* results are comparable, and that these two basis sets generally yield larger puckerings for each envelope conformer of **1** than the STO-3G basis set. The effect of ring conformation on puckering amplitude is smallest in the 6-31G* data, suggesting a more circular pseudorotational pathway than would be concluded from the STO-3G and 3-21G data. Like the aldotetrofuranoses[4], the puckering amplitudes of **1** are smaller than those observed for the pentofuranosyl rings of nucleosides by X-ray crystallography[9b]. Thus, the bulkier hydroxymethyl group of the pentofuranosyl ring is better accommodated in a more puckered ring[4].

Previous *ab initio* calculations at the STO-3G and 3-21G levels indicated that C–H bonds in the vicinity of the furanose ring oxygen vary in length, depending on their

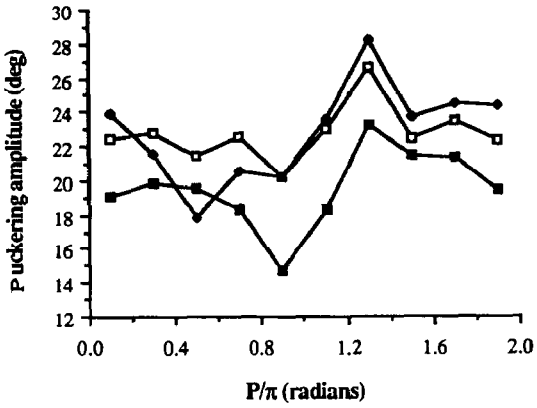


Fig. 4. Puckering amplitudes determined from *ab initio* calculations of **1**. Data were obtained using the STO-3G (■), 3-21G (◆), and 6-31G* (□) basis sets.

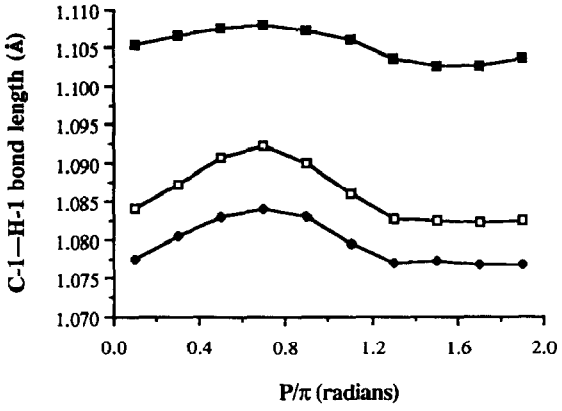


Fig. 5. The effect of furanose ring conformation on the C1-H1 bond length in **1**. Data were obtained using the STO-3G (■), 3-21G (◆), and 6-31G* (□) basis sets.

disposition with respect to the lone-pair orbitals of the ring oxygen[4,5]. The general effect of ring conformation on C-1–H-1 bond length (Fig. 5) is conserved in the three basis sets, with the 3-21G and 6-31G* calculations showing a greater overall change in length than the STO-3G calculations. In terms of absolute bond length, the STO-3G and 3-21G data yield longest and shortest C-H bond lengths, respectively, with the 6-31G* treatment giving lengths intermediate in magnitude. Similar results are obtained for the C-4–H-4*R* and C-4–H-4*S* bonds of **1** (Fig. 6).

The C–O bond lengths in the vicinity of the anomeric carbon (*e.g.*, C-4–O-4, C-1–O-4 and C-1–O-1) have previously been shown to depend in a highly coordinated fashion on furanose ring conformation[5]. From calculations with the 3-21G basis set[5], the C-1–O-1 and C-1–O-4 bonds were found to be maximal and minimal in length, respectively, when the C-1–O-1 bond is *quasi*-axial, whereas the opposite was observed

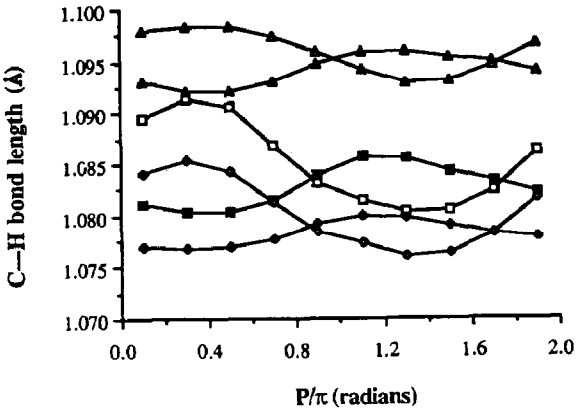


Fig. 6. The effect of furanose ring conformation on the C-4–H-4*R* (closed symbols) and C-4–H-4*S* (open symbols) bond lengths in **1**. Data were obtained using the STO-3G (△), 3-21G (◇), and 6-31G* (□) basis sets.

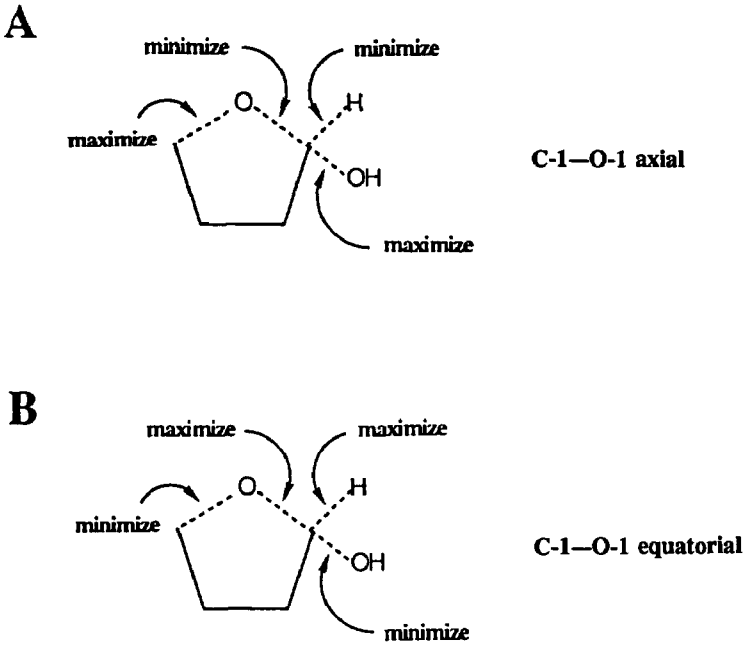


Fig. 7. Bond length behavior in the vicinity of the anomeric carbon when (A) the C-1–O-1 bond is *quasi*-axial, and (B) the C-1–O-1 bond is *quasi*-equatorial.

when the C-1–O-1 bond is *quasi*-equatorial (Fig. 7 and Fig. 8A). This result is confirmed in calculations of **1** using the 6-31G* basis set (Fig. 8B).

In furanoses, the C-1–O-1 bond is able to assume a large number of orientations ranging from pure *quasi*-axial to pure *quasi*-equatorial, in contrast to the more conformationally constrained pyranoses in which the C-1–O-1 bond is either axial or equa-

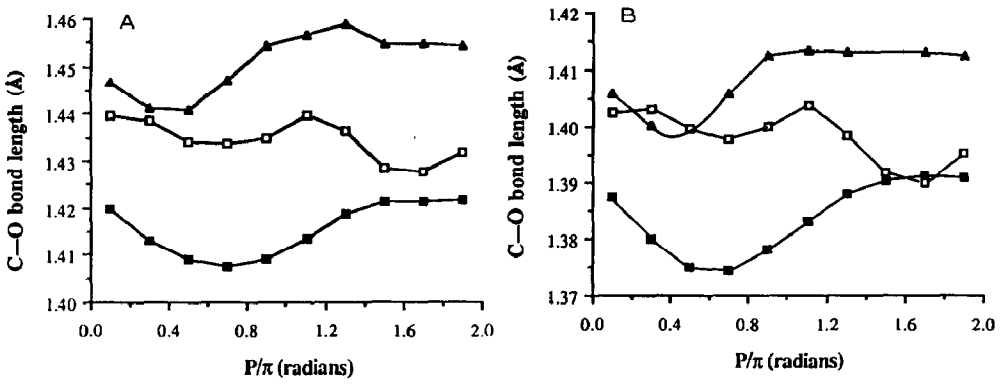
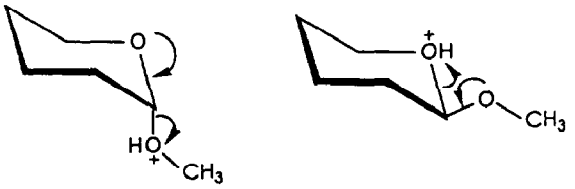


Fig. 8. Effect of furanose ring conformation on the C-4–O-4 (▲), C-1–O-4 (□), and C-1–O-1 (■) bond lengths. (A) 3-21G data. (B) 6-31G* data.



α-D-pyranoside: exocyclic cleavage

β-D-pyranoside: endocyclic cleavage

Fig. 9. Glycosidic bond cleavage *via* exo- or endo-cyclic mechanisms.

torial†. Thus, furanoses provide a good model system to evaluate how C–O bonds in the vicinity of the anomeric carbon change in length in response to conformation. These changes in C–O bond length may influence the mechanism of glycoside bond hydrolysis, which is currently of concern, particularly to enzymologists[10]. The results shown in Fig. 8 suggest that cleavage of the C-1–O-1 bond may be facilitated when the C-1–O-1 bond is *quasi*-axial, whereas cleavage of the C-1–O-4 bond may be facilitated when the C-1–O-1 bond is *quasi*-equatorial. The glycosidic bond of pyranosides may hydrolyze *via* exocyclic (C-1–O-1) or endocyclic (C-1–O-5) C–O bond cleavage mechanisms (Fig. 9). The former may be assisted when the C-1–O-1 bond is axial, whereas an equatorial C-1–O-1 bond may assist the latter. Thus, for example, endocyclic cleavage would be assisted in β-D-pyranosides, and exocyclic cleavage would be assisted in α-D-pyranosides (Fig. 9). Implicit in this argument are the assumptions that the elongated C–O bond is more predisposed towards cleavage, and that C–O bond length changes similar to those found in furanose reducing sugars (*e.g.*, in **1**) also occur in glycosides and protonated glycosides, the latter being the presumed reactive intermediate in the hydrolysis of glycosides (Fig. 9). The second assumption will require validation by

† This argument is valid for pyranoses that assume an idealized chair conformation. The less common boat or skew forms allow C-1–O-1 orientations intermediate between axial and equatorial position.

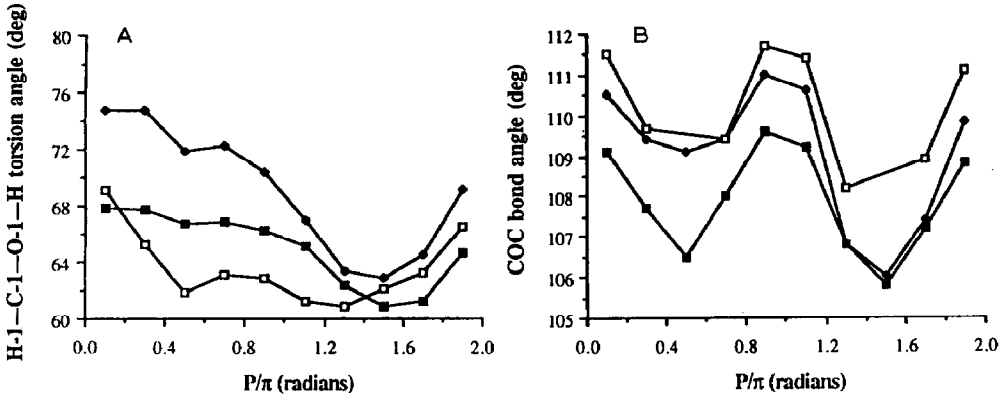


Fig. 10. (A) The effect of furanose ring conformation of the H-1–C-1–O-1–H torsion angle (A) and the C-4–O-4–C-1 bond angle (B) of **1**. Data were obtained using the STO-3G (■), 3-21G (◆), and 6-31G* (□) basis sets.

further calculations. Experimental data indicate that simple glycopyranosides hydrolyze *via* an exocyclic protonation mechanism regardless of anomeric configuration[10]. While the C–O bond length factor alone may favor one mechanism over the other, its effect may be reduced or negated by other competing factors that dominate in solution. Interestingly, Post and Karplus[11] recently proposed that the enzyme-catalyzed hydrolysis of β-D-pyranosides may proceed via endocyclic C–O bond cleavage.

The effect of basis set on two key furanose ring structural parameters of **1**, namely, the exoanomeric H-1–C-1–O-1–H torsion angle and the C-4–O-4–C-1 endocyclic bond angle (Fig. 10) shows the 6-31G* results to be in closer accord with 3-21G data than with STO-3G data.

In conclusion, *ab initio* molecular orbital calculations on furanose rings with the 3-21G and 6-31G* basis sets appear to yield qualitatively similar results. The 6-31G* data, however, is likely to be more quantitatively reliable. It is unclear at the present time whether the inclusion of electron correlation in 6-31G* calculations will significantly affect the calculated structural parameters and conformational energies. However, considering the significantly greater time and expense to conduct calculations at the 6-31G* level or higher, a 3-21G analysis may be more attractive, especially when only semi-quantitative results are needed. More importantly, the observations reported earlier[4,5] on the effect of furanose ring conformation on C–H and C–O bond lengths, especially those in the vicinity of the anomeric carbon, and on other furanose structural parameters, have been validated by calculations conducted with the 6-31G* basis set. It remains to be determined how the inclusion of electron correlation energy corrections (Moller-Plesset functions) in computations with the 6-31G* basis set will affect the results of *ab initio* calculations of furanose rings.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant No. GM-33791). The authors thank the University of Notre Dame Computing Center for use of the Convex C-220 computer, Ken Grant of the Department of Mathematics for assistance in the use of the Convex C-220 computer, and Daniel M. Chipman of the University of Notre Dame Radiation Laboratory for helpful discussions.

REFERENCES

1 E. Westhof and M. Sundaralingam, *J. Am. Chem. Soc.*, 105 (1983) 970–976.
2 A. S. Serianni and R. Barker, *J. Org. Chem.*, 49 (1984) 3292–3300.
3 O. Jardetsky, *Biochem. Biophys. Acta*, 621 (1980) 227–232.
4 A. S. Serianni and D. M. Chipman, *J. Am. Chem. Soc.*, 109 (1987) 5297–5303.
5 E. C. Garrett and A. S. Serianni, A. French and J. Brady (Eds.), *Computer Modeling of Carbohydrate Molecules*, ACS Symposium Series, Vol. 430, American Chemical Society, Washington DC, 1990, chapter 6.
6 M. J. Frisch, M. Head-Gordon, H. B. Schlegel, K. Raghavachari, J. S. Binkley, C. Gonzales, D. J. Defrees, D. J. Fox, R. A. Whiteside, R. Seeger, C. F. Melius, J. Baker, R. Martin, L. R. Kahn, J. J. P. Steward, E. M. Fluder, S. Topiol, and J. A. Pople, GAUSSIAN88, GAUSSIAN, Inc., Pittsburgh, PA, 1988.
7 C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 94 (1972) 8205–8212.
8 C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 95 (1973) 2333–2344.
9 (a) R. U. Lemieux, *Pure Appl. Chem.*, 25 (1971) 527–548. (b) H. P. M. De Leeuw, C. A. G. Haasnoot, and C. Altona, *Isr. J. Chem.*, 20 (1980) 108–126.
10 M. L. Sinnot, in M. I. Page (Ed.), *The Chemistry of Enzyme Action*, Elsevier, Amsterdam, 1984, pp. 389–431.
11 C. B. Post and M. Karplus, *J. Am. Chem. Soc.*, 108 (1986) 1317–1319.