A Comparison of Semiempirical and *Ab Initio* Methods for the Study of Structural Features of Relevance in Carbohydrate Chemistry

Robert J. Woods, Walter A. Szarek* and Vedene H. Smith, Jr.*

Department of Chemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

Semiempirical and *ab initio* molecular orbital calculations have been performed on methanediol, and indicate that the former are capable of reproducing the trends in geometries, conformational energies and proton affinities associated with the anomeric and related effects.

The most dominant conformation-controlling feature in carbohydrates is known as the anomeric effect.[1] The effect refers to the tendency of an electronegative substituent at C-1 of a pyranoid ring to assume the axial rather than equatorial orientation, in contrast to predictions based solely on steric grounds. The *exo*-anomeric effect has the same electronic origin as the anomeric effect, but it is an orientation effect on the aglycone of a glycopyranoside; specifically it refers to the preference for a *gauche* orientation of the aglycone-O–R bond with respect to the endocyclic C–O bond.[2] Probably the most important consequence of the *exo*-anomeric effect concerns the relative disposition of the contiguous sugars of oligo- and poly-saccharides. Since both of these stereoelectronic effects play an important part in the conformational properties of carbohydrates, they have been studied extensively, both experimentally and theoretically.[3] It is the goal here to highlight the abilities of semiempirical molecular orbital methods to reproduce the structural and energetic properties associated with these effects in a model system, namely, methanediol. Semiempirical calculational methods offer the possibility of accurately predicting the molecular properties of systems that otherwise would be too large to study with more rigorous *ab initio* approaches. The complexity of the interactions between vicinal hydroxy groups as well as the stereoelectronic effects associated with the anomeric centre in carbohydrates require a sophisticated calculational approach; however, the relatively large size of such systems makes accurate calculations extremely challenging. A compromise between efficiency and accuracy appears to be offered through semiempirical methods.

It has been reported[4] that the semiempirical method, AM1,[5] accurately reproduces the molecular geometries of carbohydrates, both in terms of sugar-ring conformation and intramolecular hydrogen bonding. Here we compare the results obtained with three semiempirical methods, namely, MNDO,[6] AM1 and the most recent congener, PM3,[7] and those obtained from high-level *ab initio* calculations. The results are discussed in terms of the molecular geometries and conformational energies, as well as the dependence of the basicities of the hydroxy groups on conformation.

It is known[8] that certain C–O bond length variations are observed in molecules containing the O–C–O atomic arrangement. The variations have been explained as arising from interactions between the nonbonded electrons on oxygen (n_p) and the antibonding C–O bond orbital (σ^*_{c-o}), and form the basis of the anomeric effect.[9] These interactions are maximized when the nonbonded or lone-pair electrons on one oxygen adopt an antiperiplanar orientation with respect to the C–O bond involving the other oxygen. This arrangement leads to a shortening of the C–O (donor oxygen) bond and a lengthening of the C–O (acceptor σ*) bond. Thus, by examining the staggered conformers of methanediol (see Fig. 1) it is possible to estimate the extent of these interactions.† The results of *ab initio* and semiempirical calculations are presented in Table 1.

The *ab initio* calculations were perfomed at the Hartree–Fock (HF) level using the split-valence 6-31G*[10] basis set, which incorporates polarization functions only on the non-hydrogen atoms. The effects of electron correlation were examined using Møller–Plesset perturbation theory[11] at the second order (MP2) level, with the 6-311 + +G**[12] basis set.‡ The *ab initio* and PM3 calculations were performed using the GAUSSIAN 86[13] and MOPAC[14] software packages, respectively, whereas, AMPAC[15] was employed for the MNDO and AM1 calculations. The keywords PRECISE and PULAY were included in all semiempirical calculations. Full bond length and bond angle geometry optimizations were performed at each calculational level with the torsion angles constrained to the staggered orientations. A value of 367.2 kcal mol^{-1},[16] (1 cal = 4.184 J) was used for the heat of formation of H^+, as recommended by Dewar and Dieter.[17]

An examination of Table 1 indicates that, regardless of the method, the *sc,sc* conformer is the most stable and the *ap,ap* is the least stable. This feature has been observed previously,[18] and may be attributed to the presence of two anomeric effects in the former conformer and none in the latter. The *sc,ap* and *sc,−sc* conformers have similar energies, the former conformer exhibiting one anomeric effect, and the latter exhibiting two; however, the latter conformer contains an unfavourable steric interaction between the hydroxy protons. While the relative energy calculated for the *ap,ap* conformer using AM1 is higher than the corresponding *ab initio* values,§ the relative

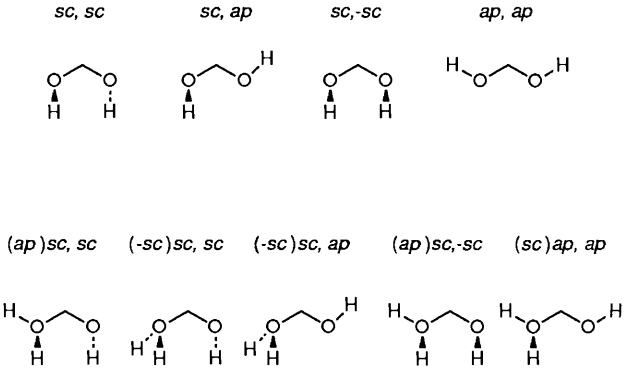

Fig. 1 The staggered conformers of neutral and protonated methanediol

---

† The symmetry of methanediol does not permit a separation of the contributions to the energy arising from the anomeric and *exo*-anomeric effects.

‡ It has recently been reported that the 6-311G and 6-311G** (R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650) basis sets are not of triple-split valence quality, but rather, are of double-zeta quality (R. S. Grev and H. F. Schaefer, III, *J. Chem. Phys.*, 1989, **91**, 7305). The substantial effect of adding diffuse (+) functions to the 6-311G** basis set, as reported by Pople *et al.* (J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari and L. A. Curtiss, *J. Chem. Phys.*, 1989, **90**, 5622), has been related (G. E. Scuseria and H. F. Schaefer, III, *J. Chem. Phys.*, 1989, **90**, 3629) to the less-extended 2p-space of the 6-311G basis set.

§ The omission of polarization functions from the 6-31G* basis set, *i.e.* 6-31G, raises the relative energy of the *ap,ap* conformer by *ca.* 2 kcal mol^{-1} (R. J. Woods, Ph. D. Thesis, Queen's University, Kingston, ON, 1990).

Table 1 Geometries[a] and relative energies[b] for each conformer of methanediol[c]

Parameter	HF/6-31G*	MP2/6-311 ++G**	MNDO	AM1	PM3
Conformer			*sc,sc*		
Relative energy[d]	0.00	0.00	0.00	0.00	0.00
R(C–O)	1.386	1.405	1.399	1.403	1.393
R(O–H)	0.949	0.962	0.947	0.965	0.950
θ(O–C–O)	112.2	112.5	110.3	106.7	107.9
θ(C–O–H)	108.8	106.7	110.4	108.0	109.0
Conformer			*sc,ap*		
Relative energy	3.97	3.89	1.78	4.18	1.92
R(*sc*-C–O)	1.373	1.389	1.390	1.399	1.384
R(*ap*-C–O)	1.396	1.417	1.405	1.410	1.396
R(O–H)	0.948	0.961	0.947	0.964	0.950
θ(O–C–O)	108.5	108.1	107.1	103.2	101.0
θ(*sc*-C–O–H)	108.9	106.9	111.5	107.7	109.2
θ(*ap*-C–O–H)	110.0	107.9	113.2	107.7	108.8
Conformer			*sc,−sc*		
Relative energy	4.51	4.33	2.75	3.56	3.47
R(C–O)	1.389	1.407	1.398	1.403	1.394
R(O–H)	0.947	0.960	0.946	0.964	0.950
θ(O–C–O)	113.7	114.0	111.5	108.4	110.0
θ(C–O–H)	110.2	108.1	113.8	108.1	109.4
Conformer			*ap,ap*		
Relative energy	8.59	8.50	4.91	10.50	5.04
R(C–O)	1.382	1.400	1.396	1.407	1.390
R(O–H)	0.947	0.960	0.947	0.965	0.948
θ(O–C–O)	105.4	104.6	104.3	100.8	96.5
θ(C–O–H)	109.6	107.3	110.4	106.0	107.4

[a] Å and degrees. [b] kcal mol^{-1}. [c] At each calculational level the values for the C–O bond length in methanol are: 1.511, 1.420, 1.391, 1.410 and 1.395. [d] The corresponding *ab initio* total energies (au) are: −189.9005 and −190.5928; the semiempirical heats of formation (kcal mol^{-1}) are: −107.32, −114.18 and −100.91.

energies for the other conformers are in good agreement with the *ab initio* values. Both MNDO and PM3 appear to underestimate the relative energies of both the *ap,ap* and *sc,ap* conformers. All of the methods, with the exception of AM1, predict the *sc,ap* conformer to be slightly more stable than the *sc,−sc* conformer.

As expected,[8] in the case of the *sc,ap* conformer, the *sc*-C–O bond is shorter than the *ap*-C–O bond. Moreover, the *sc*-C–O bond is shorter than the C–O bonds in the *sc,sc* conformer, and the *ap*-C–O bond is longer than the C–O bonds in the *ap,ap* conformer. There does not appear to be a significant difference between the C–O bond lengths in the *sc,sc* and *ap,ap* conformers, a feature that may indicate that the two anomeric interactions in the former conformer have approximately equal but opposite effects on the C–O bond lengths. A shortening of the C–O bond lengths in the *ap,ap* conformer, relative to that in methanol, is predicted by each method, with the exception of MNDO, and may be due largely to coulombic interactions. A close agreement between the AM1- and MP2-calculated C–O bond lengths is observed.

As reported for the case of dimethoxymethane,[19] the value of the O–C–O bond angle exhibits a conformational dependence, and may be ranked according to the following sequence of conformers: *sc,−sc* > *sc,sc* > *sc,ap* > *ap,ap*. This trend is reproduced at each computational level; however, the AM1- and PM3-derived values underestimate the value of this parameter by *ca.* 5°.

The geometries and relative energies of protonated methanediol are presented in Table 2. A total of five staggered conformers may be generated from protonation of the staggered conformers of methanediol. In agreement with predictions based on the reverse-anomeric effect,[3e] the lowest-energy conformer was found to be the (*−sc*)*sc,ap* conformer. This species may be viewed as arising from protonation of the *sc,ap* conformer, in which the proton adopts a *−sc* orientation with respect to the opposite *ap*-C–O bond. Each calculational method predicts the same ordering of conformational energies, namely, (*−sc*)*sc,ap* > (*ap*)*sc,sc* > (*−sc*)*sc,sc* > (*ap*)*sc,−sc* > (*sc*)*ap,ap*, with the exception that MNDO and PM3 predict the (*sc*)*ap,ap* conformer to be slightly more stable than the (*ap*)*sc,−sc* conformer.

Protonation leads to a lengthening of the C–O(+) bond and, in conformers in which the $\sigma^*_{\text{C–O}(+)}$ participates in an anomeric effect with the nonprotonated oxygen, this lengthening may be noticeably greater than predicted for protonated methanol. This feature has been postulated to be of relevance in the hydrolysis of glycosides, and, in particular, as regards the relative rates of hydrolysis of α- and β-glycosides.[8] The close agreement between the C–O bond lengths calculated at the PM3 and MP2 levels is noteworthy; however, both PM3 and AM1 again appear to underestimate the values of the O–C–O bond angles.

In order to examine the effect of conformation on hydroxy group basicity, the proton affinities of each conformer were calculated, and the results are presented in Table 3. Only the proton affinities arising from the direct protonation of specific conformers of methanediol have been included. Since the geometries employed are not the exact equilibrium geometries, no estimate of the vibrational energy contributions to the *ab initio* values have been included. The MNDO- and AM1-derived proton affinities are lower than the corresponding *ab initio* values, by *ca.* 10 kcal mol^{-1}; however, the PM3-derived values are lower by more than 20 kcal mol^{-1}. The low values of the PM3-derived proton affinities appear to result from an underestimation of the stability of the protonated species (see Tables 1 and 2). For the lowest-energy conformer of methanediol (*sc,sc*) each method predicts that the highest proton affinity is found for the formation of the (*ap*)*sc,sc* species. However, the *sc,sc* conformer does not have the highest proton affinity, but rather, the *sc,ap* conformer. It

Table 2 Geometries[a] and relative energies[b] for each conformer of protonated methanediol[c]

Parameter	HF/6-31G*	MP2/6-311 ++G**	MNDO	AM1	PM3
Conformer			(*ap*)*sc*,*sc*		
Relative energy	1.69	0.44	1.56	2.12	2.02
R(C–O)	1.323	1.334	1.353	1.357	1.335
R[C–O(+)]	1.542	1.559	1.515	1.508	1.544
R(O–H)	0.954	0.968	0.953	0.977	0.955
R[O(+)–H]	0.967	0.978	0.969	0.998	0.973
θ(O–C–O)	109.2	109.7	107.9	101.9	101.0
θ(C–O–H)	114.1	111.8	117.1	111.4	113.8
θ[C–O(+)*sc*–H]	111.1	110.4	114.8	110.7	111.0
θ[C–O(+)H]	113.9	113.6	112.6	111.9	111.0
Conformer			–(*sc*)*sc*,*sc*		
Relative energy	3.01	1.69	3.53	3.50	4.46
R(C–O)	1.328	1.340	1.358	1.360	1.340
R[C–O(+)]	1.536	1.551	1.509	1.496	1.543
R(O–H)	0.952	0.966	0.951	0.974	0.954
R[O(+)–H]	0.965	0.977	0.970	0.996	0.972
θ(O–C–O)	113.6	114.8	110.7	106.8	109.0
θ(C–O–H)	115.6	113.1	117.9	112.1	114.0
θ[C–O(+)*sc*–H]	111.1	111.5	114.0	111.1	109.7
θ[C–O(+)H]	113.9	113.6	115.2	112.1	110.6
Conformer			(–*sc*)*sc*,*ap*		
Relative energy[d]	0.00	0.00	0.00	0.00	0.00
R(C–O)	1.352	1.369	1.378	1.381	1.367
R[C–O(+)]	1.505	1.504	1.491	1.501	1.492
R(O–H)	0.954	0.966	0.950	0.973	0.951
R[O(+)–H]	0.966	0.978	0.972	0.999	0.976
θ(O–C–O)	102.9	103.0	102.5	97.4	95.5
θ(C–O–H)	113.2	110.9	115.4	111.1	112.4
θ[C–O(+)*sc*–H]	111.8	111.2	113.9	110.1	109.4
Conformer			(*ap*)*sc*,–*sc*		
Relative energy	6.34	4.83	5.00	6.98	6.25
R(C–O)	1.320	1.332	1.352	1.356	1.331
R[C–O(+)]	1.556	1.574	1.519	1.513	1.569
R(O–H)	0.952	0.966	0.951	0.974	0.954
R[O(+)–H]	0.964	0.976	0.968	0.993	0.970
θ(O–C–O)	111.3	111.8	109.3	104.0	104.2
θ(C–O–H)	116.0	113.5	117.8	112.1	114.6
θ[C–O(+)*sc*–H]	112.8	112.6	115.9	111.9	111.1
θ[C–O(+)H]	113.4	112.5	111.2	111.4	109.0
Conformer			(*sc*)*ap*,*ap*		
Relative energy	6.71	5.87	4.63	9.94	5.50
R(C–O)	1.346	1.362	1.374	1.386	1.369
R[C–O(+)]	1.479	1.485	1.478	1.475	1.474
R(O–H)	0.952	0.965	0.951	0.975	0.951
R[O(+)–H]	0.967	0.979	0.971	0.999	0.973
θ(O–C–O)	103.6	103.1	103.0	98.4	93.6
θ(C–O–H)	112.5	110.3	114.3	108.8	111.1
θ[C–O(+)*sc*–H]	110.7	110.0	114.5	109.8	108.5
θ[C–O(+)H]	113.6	113.0	112.1	109.9	108.7

[a] Å and degrees. [b] kcal mol^{-1}. [c] At each calculational level the values for the C–O bond lengths in protonated methanol are: 1.511, 1.508, 1.474, 1.495 and 1.476. [d] The corresponding *ab initio* total energies (au) are: −190.1955 and −190.8833; the semiempirical heats of formation (kcal mol^{-1}) are: 85.82, 80.18 and 102.90.

should be noted that high proton affinities are found also for the *ap*,*ap* conformer. While the high proton affinity associated with the formation of the (*ap*)*sc*,*sc* conformer is in agreement with Deslongchamps' proposal that the presence of an anomeric effect enhances the basicity of the σ^*_{C-O} oxygen atom,[8] the high proton affinities associated with formation of the (–*sc*)*sc*,*ap* and (*sc*)*ap*,*ap* conformers, in which no n-$\sigma^*_{C-O(+)}$ interactions are present, are not. It should be noted that Deslongchamps' hypothesis was based on a consideration of the properties of dimethoxymethane, a compound that we have not explicitly examined.

Despite the simplifications inherent in the semiempirical methods, the trends in the geometries and conformational energies of methanediol and its protonated derivative are accurately reproduced. AM1 appears to perform well on the neutral species, while PM3 predicted values for the C–O bond lengths in the protonated conformers that were in good agreement with the *ab initio* values. Interestingly, the precursor to AM1 and PM3, namely MNDO, gave the most accurate description of the O–C–O angles; however, it performed poorly on the relative conformational energies. The trends in proton affinities are accurately predicted by AM1. It appears that the error associated with the relative energy of the *ap*,*ap* conformer, as calculated by AM1, is present also in the protonated species, and has little effect on the corresponding proton affinities.

Table 3 Calculated proton affinities of methanediol[a,b]

	Protonated species				
Neutral species	(*ap*)*sc,sc*	(−*sc*)*sc,sc*	(−*sc*)*sc,ap*	(*ap*)*sc,*−*sc*	(*sc*)*ap,ap*
HF/6-31G*					
sc,sc	183.4	182.1	—	—	—
sc,ap	187.4	—	189.1	182.7	182.4
sc,−*sc*	—	186.6	—	183.3	—
ap,ap	—	—	—	—	187.0
MP2/6-311++G**					
sc,sc	181.8	180.6	—	—	—
sc,ap	185.7	—	186.2	181.4	180.3
sc,−*sc*	—	185.0	—	181.8	—
ap,ap	—	—	—	—	185.0
MNDO					
sc,sc	172.5	170.5	—	—	—
sc,ap	174.3	—	175.8	170.8	171.2
sc,−*sc*	—	173.3	—	171.8	—
ap,ap	—	—	—	—	174.3
AM1					
sc,sc	170.7	169.3	—	—	—
sc,ap	174.9	—	177.0	170.0	167.1
sc,−*sc*	—	172.9	—	169.4	—
ap,ap	—	—	—	—	173.4
PM3					
sc,sc	161.4	158.9	—	—	—
sc,ap	163.3	—	165.3	159.1	159.8
sc,−*sc*	—	162.4	—	160.6	—
ap,ap	—	—	—	—	162.9

[a] kcal mol^{-1}. [b] A value of 367.2 kcal mol^{-1} was used for the heat of formation of H^+; see ref. 17.

We thank the Natural Sciences and Engineering Research Council of Canada for the awards of grants (to V. H. S. and W. A. S.) and a scholarship (to R. J. W.).

Received, 23rd Ocotober 1990; Com. 0/04772D

References

1 (*a*) *Anomeric Effect, Origin and Consequences*, eds. W. A. Szarek and D. Horton, ACS Symposium Series 87, American Chemical Society, Washington, DC, 1979; (*b*) A. J. Kirby, *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer-Verlag, New York, 1983.

2 R. U. Lemieux, S. Koto and D. Voisin, in ref. 1*a*, ch. 2.

3 For example, (*a*) G. A. Jeffrey, J. A. Pople and L. Radom, *Carbohydr. Res.*, 1972, **25**, 117; (*b*) G. A. Jeffrey, J. A. Pople and L. Radom, *Carbohydr. Res.*, 1974, **38**, 81; (*c*) G. A. Jeffrey, J. A. Pople, J. S. Binkley and S. Vishveshwara, *J. Am. Chem. Soc.*, 1978, **100**, 373; (*d*) H. Paulsen, P. Luger and F. R. Heiker, in ref. 1*a*, ch.5; (*e*) S. Wolfe, M.-H. Whangbo and D. J. Mitchell, *Carbohydr. Res.*, 1979, **69**, 1; (*f*) G. A. Jeffrey and J. H. Yates, *Carbohydr. Res.*, 1981, **96**, 205; (*g*) M. C. Krol, C. J. M. Huige and C. Altona, *J. Comput. Chem.*, 1990, **11**, 765.

4 M. Khalil, R. J. Woods, D. F. Weaver and V. H. Smith, Jr., *J. Comput. Chem.*, in the press.

5 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902.

6 M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, 1977, **99**, 4899.

7 J. J. P. Stewart, *J. Comput. Chem.*, 1989, **10**, 221.

8 See, P. Deslongchamps, *Stereoelectronic Effects in Organic Chemistry*, Pergamon Press, New York, 1983.

9 See ref. 3*e*, and references cited therein.

10 P. C. Hariharan and J. A. Pople, *Theoret. Chim. Acta*, 1973, **28**, 213.

11 C. Moller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618; J. A. Pople, J. S. Binkley and R. Seeger, *Int. J. Quantum Chem. Symp.*, 1976, **10**, 1; R. Krishnan and J. A. Pople, *Int. J. Quantum Chem.*, 1976, **14**, 91.

12 G. W. Spitznagel, T. Clark, J. Chandrasehar and P. v. R. Schleyer, *J. Comput. Chem.*, 1983, **3**, 363.

13 M. J. Frisch, J. S. Binkley, H. B. Schlegel, K. Raghavachari, C. F. Melius, R. L. Martin, J. J. P. Stewart, F. W. Bobrowicz, C. M. Rohlfing, L. R. Kahn, D. J. DeFrees, R. Seeger, R. A. Whiteside, D. J. Fox, E. M. Fluder and J. A. Pople, GAUSSIAN 86, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh, PA, 1984.

14 J. J. P. Stewart, MOPAC 5.0 (QCPE Program No. 581), 1989.

15 Dewar Research Group and J. J. P. Stewart, AMPAC (QCPE Program No. 527), 1988.

16 D. R. Stull and J. Prophet, JANAF Thermochemical Tables, NSRDS-NBS37, 1971.

17 M. J. S. Dewar and K. M. Dieter, *J. Am. Chem. Soc.*, 1986, **108**, 8075.

18 See, for example, ref 3*a*.

19 K. B. Wiberg and M. A. Murcko, *J. Am. Chem. Soc.*, 1989, **111**, 4821.