

UNIVERSITY OF HAWAII LIBRARY

COMPUTATIONAL STRATEGY FOR PREDICTING THE SPECIFIC  
OPTICAL ROTATION VALUES OF LARGE FLEXIBLE MOLECULES

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE  
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

CHEMISTRY

DECEMBER 2006

By  
Anna Melnichuk

Thesis Committee:

John Head, Chairperson  
Thomas Hemscheidt  
Kristin Kumashiro

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Chemistry.

DISSERTATION COMMITTEE

John Head

Chairperson

H. G. G. G.

M. V. K.

© Copyright 2006

by

Anna Melnichuk

## **ACKNOWLEDGEMENTS**

**I wish to thank Dr. John Head for the guidance I've received while doing this project. I also thank the Maui High Performance Computing Center and the University of Hawai'i Dell Cluster for providing me with computer time so that I could complete my work.**

## ABSTRACT

The topic discussed in this thesis concerns the challenges faced when one needs to find a theoretical value for the specific optical rotation ( $[\alpha]_D$ ) for a relatively large and flexible molecule. A method was developed to find all of the low-energy conformers that may be present in solution and to calculate the  $[\alpha]_D$  values for these conformers without exceeding a reasonable computer time.

Initially, the effects of changes in molecular geometry on  $[\alpha]_D$  values were investigated using several small and rigid molecules. (2S)-2-methyl oxirane was used to investigate the effect of rotating the methyl group on calculated  $[\alpha]_D$  values. The  $[\alpha]_D$  values were calculated using different basis sets which were found to produce similar  $[\alpha]_D$  values at the optimized geometry. However, for a non-equilibrium geometry the different basis sets gave very different  $[\alpha]_D$  values. This variation was attributed to the diffuse p-orbital basis functions on hydrogen atoms which are present in the aug-cc-pVDZ basis set and absent in the 6-31G\* basis set. 2-Propanol was used to show that even though it is an achiral molecule and thus has an  $[\alpha]_D$  value of zero, there are several low-energy conformers of 2-propanol which have non-zero calculated  $[\alpha]_D$  values. The three low-energy conformers were obtained by rotating the hydroxy group about the HOCH dihedral angle: the *trans* conformer has  $C_s$  symmetry and an  $[\alpha]_D$  value of zero. However, the two *gauche* conformers have  $C_1$  symmetry and a non-zero  $[\alpha]_D$  values which have the same magnitude but opposite sign. The calculated  $[\alpha]_D$  values for each conformer were averaged using Boltzmann statistics to produce an average  $[\alpha]_D$  value of zero.

The Monte Carlo (MC) program was written to search for all of the low-energy conformers of flexible molecules by randomly varying user-specified dihedral angles along the

molecular backbone. The MC program used AM1 semiempirical method in the GAMESS software for fast energy screening. These AM1 low-energy conformers found using the MC program were re-optimized at the B3LYP/6-31G\* level of theory and then their  $[\alpha]_D$  values were calculated at the B3LYP/aug-cc-pVDZ level of theory.

The effectiveness of the MC program at correctly finding the low-energy conformers was tested using the achiral molecule: 3-hydroxypentadiol. The average  $[\alpha]_D$  value for this molecule converged to zero after two MC runs. The magnitude of error for the calculated  $[\alpha]_D$  value when the MC routine was used to find the low-energy conformers was found to be  $\pm 40^\circ$  based on six MC runs using a chiral molecule: (*S*)-(+)-ethyl 3-hydroxybutanoic acid with a known experimental  $[\alpha]_D$  value. The absolute stereochemistry of 3,5-dimethoxyoctan-1-ol with two chiral centers is currently unknown. However, the relative stereochemistry of the two chiral centers is known to be *syn*. After seven MC runs, the average  $[\alpha]_D$  value converged to  $61^\circ$  for an (*R,R*) conformer. The experimental  $[\alpha]_D$  value is still needed to evaluate whether if the calculated value is correct within the expected error of  $\pm 40^\circ$  and to assign the absolute stereochemistry for 3,5-dimethoxyoctan-1-ol.

Using the methods developed in this work, it appears feasible to compute the  $[\alpha]_D$  values for even larger flexible molecules with multiple chiral centers. This tactic will be very useful in cases where other methods for elucidating the absolute stereochemistry, such as NMR and X-ray diffraction, do not work well.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>LIST OF TABLES</b>	<b>ix</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>CHAPTER 1. INTRODUCTION</b>	<b>1</b>
1.1 Statement of the problem of finding the absolute stereochemistry of a molecule	1
1.2 Review of the theory used in the calculation of $[\alpha]_D$	3
1.3 Obtaining experimental specific optical rotation value	9
1.4 Review of the challenges in calculating $[\alpha]_D$ for a structurally rigid molecule	9
1.5 Using Boltzmann averaging to find $[\alpha]_D$ for a molecule	12
1.6 Review of the previous efforts to determine $[\alpha]_D$ for a natural product	14
1.7 Challenges concerning flexible molecules	18
1.8 Outline of the thesis	19

<b>CHAPTER 2. COMPUTATION OF AVERAGE SPECIFIC OPTICAL ROTATION VALUE: EVALUATION OF BASIS SETS AND THEORY LEVEL</b>	<b>23</b>
2.1 Introduction	23
2.2 (2 <i>S</i> )-2-Methyl oxirane: effect of HCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets	24
2.3 Hexafluoro-methyl oxirane: effect of FCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets	31
2.4 2-Propanol: effect of HCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets; calculating the average $[\alpha]_D$ value	34
2.5 Summary of Chapter 2	40

<b>CHAPTER 3. THE DEVELOPMENT AND USAGE OF A MONTE CARLO CONFORMER SEARCH PROCEDURE FOR THEORETICALLY DETERMINING THE AVERAGE SPECIFIC OPTICAL ROTATION VALUE</b>	<b>41</b>
3.1 Statistics and $[\alpha]_D$ in large molecules (Introduction)	41
3.2 Outline of the Monte Carlo program	43
3.3 Test of computational procedure using an achiral molecule	47

3.4 Basis set selection for $[\alpha]_D$ calculations	59
3.5. Testing the predicative power of the MC based $[\alpha]_D$ value calculations	67
3.6 Determination of absolute stereochemistry of 3,5-dimethoxyoctan-1-ol via $[\alpha]_D$ calculation	74
3.7 Recommended procedure for finding the absolute stereochemistry of the fatty acid	83
CHAPTER 4. CONCLUSIONS	84
APPENDIX	87
REFERENCES	132



## LIST OF TABLES

2.1a	$[\alpha]_D$ calculations using HF method for (2S)-2-methyl oxirane using HF/6-31G* geometries	25
2.1b	$[\alpha]_D$ calculations using B3LYP method for (2S)-2-methyl oxirane using B3LYP/6-31G* geometries	25
2.2	$[\alpha]_D$ calculations for hexafluoro-methyl oxirane analog using B3LYP method	33
2.3	Calculated average specific optical rotation value for 2-propanol	39
3.1a	Lowest energy conformers 1-29 from a MC run with nseed = 111324 of 3-hydroxypentadiol.	50
3.1b	Lowest energy conformers 30-59 from a MC run with nseed = 111324 of 3-hydroxypentadiol.	50
3.1c	Lowest energy conformers 30-59 from a MC run with nseed = 111324 of 3-hydroxypentadiol.	50
3.1d	Lowest energy conformers 30-59 from a MC run with nseed = 111324 of 3-hydroxypentadiol.	51
3.2a	Lowest energy conformers 1-29 from a MC run with nseed = 171982 of 3-hydroxypentadiol.	52
3.2b	Lowest energy conformers 30-59 from a MC run with nseed = 171982 of 3-hydroxypentadiol.	52
3.2c	Lowest energy conformers 1-59 from a MC run with nseed = 171982 of 3-hydroxypentadiol.	54
3.3	Combined results for conformers 1 – 59 from MC run with nseed = 111324 and conformers 1 – 59 of MC run with nseed = 171982 of 3-hydroxypentadiol.	55
3.4	Trend in the average specific rotation value of 3-hydroxypentadiol	56
3.5	Basis set evaluation for $[\alpha]_D$ calculation of (2S)-methyl oxirane	60
3.6	$[\alpha]_D$ value correlation between aug-cc-pVDZ and 6-31G* basis sets for 3-hydroxypentadiol	61
3.7	$[\alpha]_D$ value correlation between aug-cc-pVDZ, 6-31G*, and 6-31++G(d,p) basis sets for 3-hydroxypentadiol	64
3.8a	$[\alpha]_D$ value correlation between aug-cc-pVDZ and 6-31++G* basis sets for 3-hydroxypentadiol	65
3.8b	$[\alpha]_D$ value correlation between aug-cc-pVDZ, 6-31G* and 6-31++G* basis sets for 3-hydroxypentadiol	65
3.9	Lowest energy conformers 1-60 from a MC run with nseed = 39679 of (S)-(+)-ethyl-3-hydroxybutanol	69
3.10	Lowest energy conformers 1-30 from a MC run with nseed = 3276251 of (S)-(+)-ethyl-3-hydroxybutanol	70

3.11	Lowest energy conformers 1-50 from a MC run with nseed = 8476125 of ( <i>S</i> )-(+)-ethyl-3-hydroxybutanol	71
3.12	Combined results for ( <i>S</i> )-(+)-ethyl-3-hydroxybutanol	72
3.13	Lowest energy conformers 1-40 from a MC run with nseed = 549078 of ( <i>R,R</i> )-3,5-dimethoxyoctan-1-ol	76
3.14	Lowest energy conformers 1-100 from a MC run with nseed = 1527699 of ( <i>R,R</i> )-3,5-dimethoxyoctan-1-ol	78
3.15	Lowest energy conformers 1-33 from a MC run with nseed = 3723001 of ( <i>R,R</i> )-3,5-dimethoxyoctan-1-ol	79
3.16	Lowest energy conformers 1-33 from a MC run with nseed = 2789 of ( <i>R,R</i> )-3,5-dimethoxyoctan-1-ol	80
3.17	Combined results for ( <i>R,R</i> )-3,5-dimethoxyoctan-1-ol	81
3.18	Specific optical rotation as a function of temperature	82

## LIST OF FIGURES

1.1	xz-Plane polarized light propagating in the x-direction	4
1.2	r-Polarized light which has a constant magnitude electric and magnetic field	5
1.3	2,7,8-Trioxabicyclo[3.2.1]octane	10
1.4	Hennoxazole A divided into 3 fragments	15
1.5	Table of results of calculated molar rotations from fragments of Hennoxazole A	16
1.6	Bistramide C divided into 4 fragments	17
1.7	Fatty acid	19
1.8	3,5-Dimethoxyoctan-1-ol	20
1.9	( <i>S</i> )-(+)-Ethyl 3-hydroxybutanoic acid	21
1.10	3-Hydroxypentadiol	21
2.1	(2 <i>S</i> )-2-Methyl oxirane; with a Newmann projection of the HCCO dihedral angle	23
2.2	2-Propanol	24
2.3a	PES of (2 <i>S</i> )-2-methyl oxirane along HCCO dihedral angle	26
2.3b	[ $\alpha$ ] <sub>D</sub> values for (2 <i>S</i> )-2-methyl oxirane using B3LYP Functional and different basis sets	26
2.3c	[ $\alpha$ ] <sub>D</sub> values for (2 <i>S</i> )-2-methyl oxirane using HF/SCF Method and different basis sets	26

2.4	Comparison of $[\alpha]_D$ values using B3LYP/6-31G* and B3LYP/aug-cc-pVDZ for (2S)-2-methyl oxirane	28
2.5a	Comparison of relative energies at HF/6-31G* and B3LYP/6-31G* geometries for (2S)-2-methyl oxirane	29
2.5b	Comparison of $[\alpha]_D$ values at HF/6-31G* and B3LYP/6-31G* using HF/6-31G* geometries for (2S)-2-methyl oxirane	29
2.6	PES and $[\alpha]_D$ values for butane and butene from Wiberg et al.	31
2.7	Hexafluoro-methyl oxirane energy and $[\alpha]_D$ values as functions of FCCO dihedral angle	33
2.8	Newman projection of 2-propanol conformers	34
2.9a	PES of 2-propanol gauche(+,-) using HF/6-31G* geometry	36
2.9b	$[\alpha]_D$ of 2-propanol gauche(+,-) using HF/SCF method and different basis sets	36
2.9c	PES of 2-propanol trans using HF/6-31G* geometry	38
2.9d	$[\alpha]_D$ of 2-propanol trans using HF/SCF method and different basis sets	38
3.1	Flow chart summarizing the steps in the MC program	46
3.2	3-Hydroxypentadiol; the center axes for each of the five dihedral axes are labeled	47
3.3	3-Hydroxypentadiol; average $[\alpha]_D$ Boltzmann normalization trend as a function of the Boltzmann normalization factor	56
3.4	Eight lowest energy conformers found for 3-hydroxypentadiol	58
3.5	$[\alpha]_D$ value correlation between using aug-cc-pVDZ and 6-31G* basis sets for 3-hydroxypentadiol	62
3.6a	Absolute error in $[\alpha]_D$ value correlation using aug-cc-pVDZ and 6-31G* basis sets for 3-hydroxypentadiol	63
3.6b	Relative error in $[\alpha]_D$ value correlation using aug-cc-pVDZ and 6-31G* basis sets for 3-hydroxypentadiol	63
3.7	$[\alpha]_D$ value correlation between using aug-cc-pVDZ and 6-31++G* basis sets for 3-hydroxypentadiol	66
3.8	(S)-(+)-Ethyl 3-hydroxybutanol; showing the five dihedral angles that were varied in the MC run	67
3.9	Eight lowest energy conformers of (S)-(+)-ethyl 3-hydroxybutanol	72
3.10	3,5-Dimethoxyoctan-1-ol; showing the nine dihedral angles that were varied in the MC run	74
3.11	Specific optical rotation as function of temperature for 3,5-dimethoxyoctan-1-ol	82
3.12	Fatty acid; showing the 21 dihedral angles to be used in the MC run	83

# CHAPTER 1

## INTRODUCTION

### **1.1 Statement of the problem of finding the absolute stereochemistry of a molecule**

The determination of the absolute stereochemistry of natural products is often one of the more challenging problems in a molecular structure elucidation. Natural products are often large molecules and can have multiple chiral centers. There are several experimental techniques to choose from, including Nuclear Magnetic Resonance (NMR), X-ray crystallography, optical rotary dispersion (ORD) and circular dichroism (CD). Each of these methods have strengths and weaknesses.

X-ray crystallography is the method with the least room for error in the determination of the 3-D structure. However, relatively few natural products form a crystalline structure which can be analyzed using this method. Furthermore, the amount of the isolated pure material needed for this experiment is sometimes much higher than the volume that can be practically obtained. The presence of a heavy atom in the sample is also required.<sup>1</sup>

Proton NMR can be used to determine the absolute stereochemistry of a compound. The amount of material required is on the order of a few hundred micrograms which is far less than the milligram quantities required for the X-ray diffraction experiment. However, the restrictions on the types of compounds that can be analyzed by the NMR method are severe. Firstly, the proton NMR spectrum must be correctly assigned and there can be no overlap of the signals in the spectral region of the protons associated with any of the chiral centers. Secondly, there must be a way to determine the absolute stereochemistry of at least one chiral center. In the case where multiple chiral centers are a part of the same spin system or in close spatial proximity to each

other, the stereochemistry of the other chiral centers can then be determined relative to the chiral center of known absolute stereochemistry. However, there are some molecules which have multiple chiral centers which are neither a part of the same spin system nor in close proximity to one another. The relative stereochemistry for this type of a molecule cannot be determined and the absolute stereochemistry needs to be determined for each chiral center on an individual basis. To determine the absolute stereochemistry of a chiral center by NMR, one must manipulate the molecule in question to give different chemical shifts based on its stereochemistry. If there are functional groups (such as hydroxy groups) present at the chiral center with which one can make an ester, a chiral carboxylic acid of known configuration can be used to selectively shield or deshield protons in the vicinity of the chiral center. In this way, the absolute stereochemistry of the chiral center can be determined. A discussion of the NMR method involving ester synthesis can be found in a paper by J. A. Dale and H. S. Mosher<sup>2</sup>. This method however popular, still has many limitations, the chief one being that there must be a functional group at the chiral center and the proton spectrum must be "clean". In addition, there are also all the risks and errors associated with performing chemical reactions with the natural product and isolating the correct final product for analysis. These problems include getting unwanted by-products from the reaction, product loss due to a low yielding reaction or due to purification, and the relatively large amount of starting material that is needed to perform the chemical reaction.

The spectroscopic rotation method such as ORD (Optical Rotary Dispersion) where the rotation of the plane of polarized light is measured can be performed on a small scale (microgram quantities) on any compound that is or can be made optically active. This prerequisite is seldom a problem when it comes to natural products. The interpretation of the results is the most challenging aspect of this experiment. There is no reliable method for predicting the absolute stereochemistry of a compound just by looking at the results from an ORD or a CD experiment.

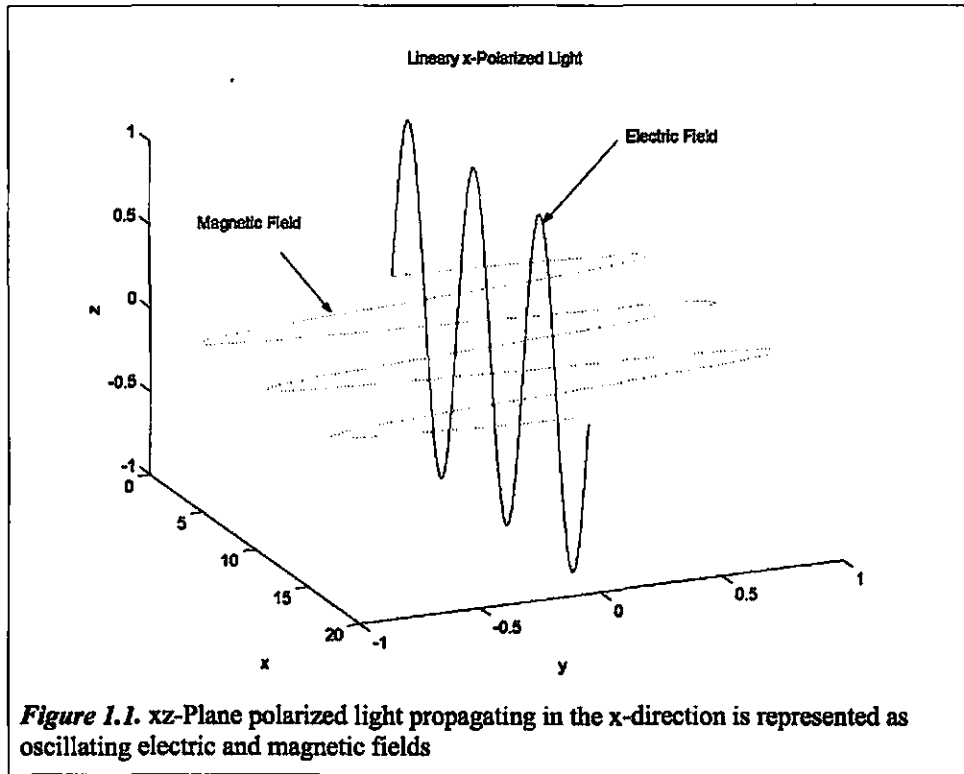
There have been attempts to classify and catalog the optical rotation data for a variety of compounds and functional groups to create a database for comparison to newly isolated substances. But this method of comparison does not always work and becomes entirely useless when a new compound is unlike any other previously cataloged.<sup>1</sup>

Only with recent advances in computational power has it become possible to use Quantum Mechanical (QM) theory to describe the optical birefringence due to the interaction of polarized light with chiral molecules. This theory was developed by Rosenfeld and Kirkwood and refined by Applequist<sup>3-5</sup>. A recent review article by T. D. Crawford<sup>6</sup> offers a good summary and discussion on the current state of theoretical progress in elucidating the absolute stereochemistry of a compound via optical spectroscopic methods. The goal of the research described in this thesis is to predict specific optical rotation values for natural products using QM.

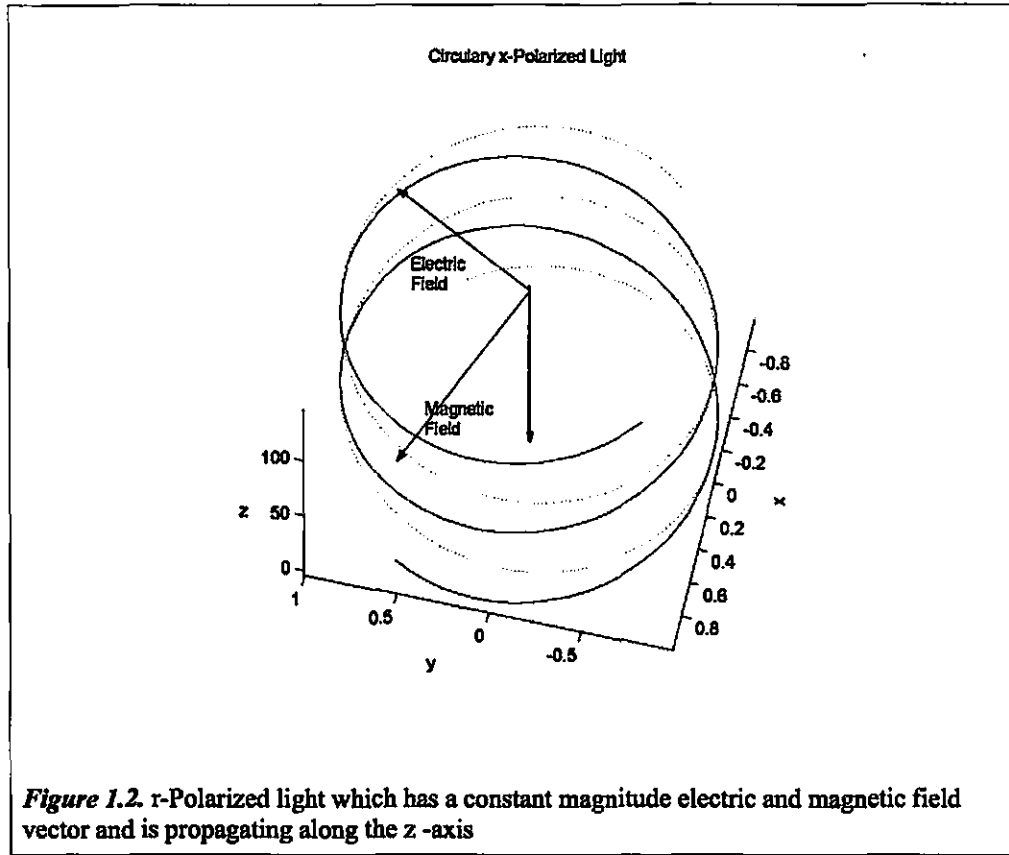
In the remainder of this chapter, we first outline the QM theory used to find the calculated  $[\alpha]_D$  values as well as give a brief introduction in to the way that the  $[\alpha]_D$  value is measured experimentally. Then, the method by which the average  $[\alpha]_D$  value can be calculated for molecules which have more than one low-energy conformer will be discussed along with the recently published results which serve as examples of how one can predict the specific optical rotation value for a natural product. In the final section of this chapter an outline for the rest of the thesis will be presented.

## **1.2 Review of the theory used in the calculation of specific optical rotation**

Linearly xz-plane polarized light propagating in the x-direction can be described as an electric field oscillating in the xz plane with a perpendicular magnetic field oscillating in phase at the same frequency, as shown on Figure 1.1.



Circularly polarized light can be described as a plane polarized light where an *xz*-plane polarized light is rotating in the *yz*-plane as it travels in the *x*-direction. If the magnitude of the electric and magnetic fields associated with the light were fixed, the propagating electromagnetic radiation would be as represented in Figure 1.2.



The left-circularly polarized (L-polarized) and the right-circularly polarized (R-polarized) light waves can be taken as the components of linearly polarized light. For example, the energy ( $E$ ) of  $xz$ -plane polarized light can be expressed as a sum of L-polarized and R-polarized light:

$$E = R + L = E_0(\hat{x} e^{i(kx - \omega t)} - i \hat{y} e^{i(kx - \omega t)}) + E_0(\hat{x} e^{i(kx - \omega t)} + i \hat{y} e^{i(kx - \omega t)}) \quad (1)$$

where  $k = \frac{2\pi}{\lambda}$ ,  $\lambda$  being the wavelength of the light and  $E$  is the energy of light:  $E = \hbar \omega$ ,

$\omega$  being the frequency of the light.

When this light travels through a chiral medium, the indices of refraction are different for



the L-polarized and the R-polarized light. At normal incidence, the difference in the index of refraction causes the L-polarized light to travel at a speed from that of the R-polarized light:

$$E = R + L = E_0 (\hat{x} e^{i(k(x-d)n_R) - \omega t} - i \hat{y} e^{i(k(x-d)n_R) - \omega t}) + E_0 (\hat{x} e^{i(k(x-d)n_L) - \omega t} + i \hat{y} e^{i(k(x-d)n_L) - \omega t}) \quad (2)$$

where  $d$  is the pathlength and  $n_L$  and  $n_R$  are the indices of refraction for the L-polarized and the R-polarized light. This difference in speed of propagation is responsible for the phase shift between the the L-polarized wave and the R-polarized wave, which causes the light to exit the sample at a different plane of polarization than the one it entered the sample with. The phase shift ( $\Delta\theta$ ) is expressed as:

$$\Delta\theta = d \omega \left[ \frac{1}{c_{medL}} - \frac{1}{c_{medR}} \right] = \frac{d \omega}{c} (n_L - n_R) = \frac{d \omega \Delta n}{c} \quad (3)$$

where  $\Delta n$  is the difference between the  $n_L$  and  $n_R$  and  $d$  is the distance traveled by the light along the  $x$  axes. The specific rotation measurement is directly related to the degree of the phase shift including experimental factors of temperature, concentration of the sample, and the type of solvent used.

In order to calculate the specific rotation, one must estimate the index of refraction of both L-polarized and R-polarized light for the molecule in question. The index of refraction ( $n$ ) is defined by the following equation:

$$n = \frac{c}{c_{med}} = \epsilon_r^{1/2} \quad (4)$$

where  $\epsilon_r$  is the relative electric permittivity,  $c$  is the speed of light in a vacuum and  $c_{med}$  is the speed of light through the medium. The  $\epsilon_r$  value depends on the total (static + dynamic) dipole

moment of the medium through which the light travels. Quantum theory is used to predict the time dependent dipole moment as a combination of the static electric dipole moment of the molecule and the induced electric dipole moment due to the electric field of the incident light and the polarizability of the medium.

The static electric dipole moment ( $p_0$ ) depends strictly on the distance between charges in space and it can be expressed in the following way:

$$p_0 = e \left[ \sum_{\alpha=1}^N z_{\alpha} r_{\alpha} + \int \Psi_0^* \sum_{i=1}^n r_i \Psi_0 d\tau \right] \quad (5)$$

where  $\sum_{\alpha=1}^N z_{\alpha} r_{\alpha}$  is the nuclear charge contribution ( $z$  is the charge of each nucleus and  $r$  is the

distance between charges); and  $\int \Psi_0^* \sum_{i=1}^n r_i \Psi_0 d\tau$  is the electron charge contribution where

$\Psi_0$  is the electronic wavefunction for the ground state of the molecule.

Using the Dirac notation, the static electric dipole moment ( $p_0$ ) can be expressed as:

$$p_0 = \langle 0 | q \hat{R} | 0 \rangle \quad \text{where } | 0 \rangle = \Psi_0 \text{ and} \quad (6)$$

$$q \hat{R} \text{ is the total dipole operator: } \sum_{\alpha=1}^N z_{\alpha} r_{\alpha} + \sum_{i=1}^n e r_i$$

The static magnetic dipole moment ( $m_0$ ) can also be expressed with Dirac notation:

$$m_0 = \langle 0 | \hat{M} | 0 \rangle \quad \text{where } \hat{M} \text{ is related to spin ( } \hat{S} \text{ )} \quad (7)$$

$$\text{and angular momentum ( } \hat{L} \text{ ) operators: } \hat{M} = -(e/2m_e)(\hat{L} + g_e \hat{S})$$

where  $m_e$  is the mass of an electron,  $g_e$  is the electron g-value and  $e$  is the charge of an electron.

Using QM we can calculate the frequency-dependent electric dipole-magnetic dipole

polarizability tensor  $\beta(\omega)$ :

$$\beta(\omega) = \frac{-2}{\hbar} \Im \sum_{m \neq 0} \frac{P_{m0} m_{0m}}{\omega_{0m}^2 - \omega^2} \quad (8)$$

where the electronic transition matrix element is  $P_{0m} = \langle 0 | q \hat{R} | m \rangle$  ; and the magnetic transition matrix element is  $m_{m0} = \langle m | \hat{M} | 0 \rangle$  . The 0-subscript indicates the ground state and the m-subscript indicates excited states. The  $\omega_{0m}$  is the natural frequency of the energy of the transition between the ground state and the  $m$  excited state and  $\omega$  is the frequency of the induced radiation. The average of the trace of  $\beta(\omega)$  provides information regarding the local dipole moment in the vicinity of the molecule in response to the perturbation by the incident light:

$$\frac{\text{tr}(\beta(\omega))}{3} = \frac{\beta_{xx}(\omega) + \beta_{yy}(\omega) + \beta_{zz}(\omega)}{3} \quad (9)$$

The specific rotation can then be directly related to the trace of  $\beta(\omega)$  by the following expression:

$$[\alpha]_D = \frac{72.0E6 \hbar^2 N_A \omega [\text{tr}(\beta(\omega))]}{M (cm_e)^2} \quad (10)$$

where  $N_A$  is Avogadro's number,  $m_e$  is the electron rest mass in kg,  $\omega$  is the frequency of the induced radiation in atomic units,  $M$  is the molecular weight in atomic units,  $c$  is the speed of light in (m/s). The subscript D on specific rotation defines the frequency of the light that which is used in the experiment (in this case, a sodium D-line of 589.3 nm).

Computationally, the specific rotation is actually calculated using time-dependent perturbation theory via the linear response theory and the use of propagators.<sup>6</sup>

### 1.3 Obtaining experimental specific optical rotation value

A polarimeter experimentally measures the angle of rotation of plane-polarized light at a specified frequency ( $\omega$ ) as it passes through a solution containing a chiral sample. This measurement is the “observed rotation”  $\alpha$ , and it is converted to specific rotation  $[\alpha]_{\omega}$  by the following formula:

$$[\alpha]_{\omega} = \frac{\alpha}{c \cdot l} \quad (11)$$

where  $c$  is concentration in g/mL and  $l$  is the path length of cuvette in decimeters

The measurements are typically made using dilute samples. For example, if the observed rotation for a sample is  $3^{\circ}$  and the concentration of this sample is 10% (0.1g/mL), the specific rotation is reported as  $30^{\circ}$ .

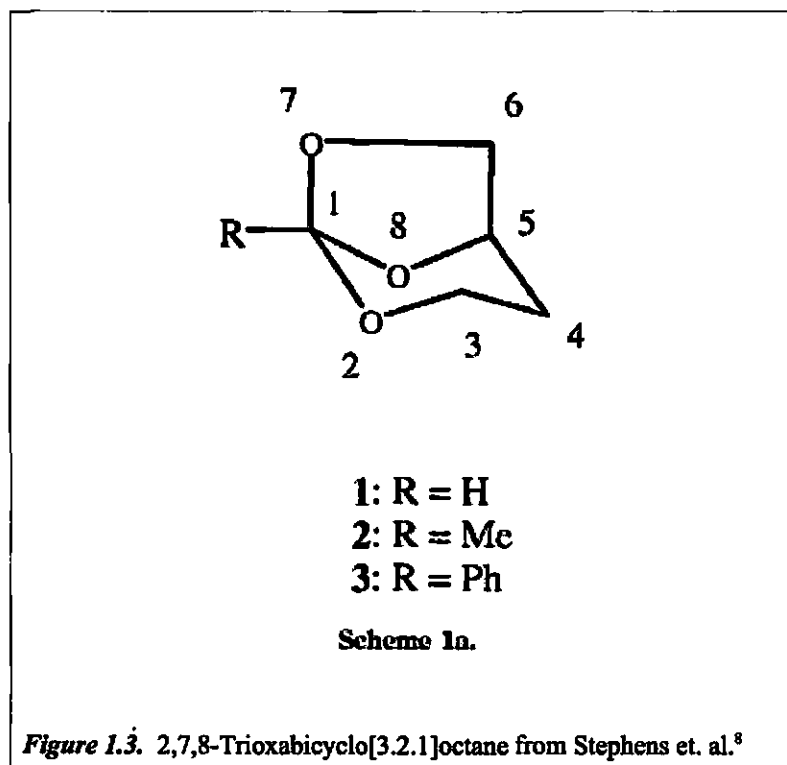
The full range of the polarimeters generally range from  $-350^{\circ}$  to  $350^{\circ}$ , but the concentrations are made such that the observed rotation falls in the range of  $\pm 10^{\circ}$  to  $\pm 35^{\circ}$ . There are a variety of wavelengths at which the specific rotation can be measured in the UV and visible range. The measurements can also be done at different temperatures. This information was obtained through *Rudolph Instruments*.<sup>7</sup>

### 1.4 Review of the challenges in calculating $[\alpha]_{\omega}$ for a structurally rigid molecule

The first challenge in calculating the specific rotation of a molecule is to determine which basis set and method of computation to use. In addition, one needs to understand what kind of errors one can expect due to: solvent effects, the method and basis set used to optimize the geometry of the molecule, and the influence of introducing Gauge Independent Atomic Orbitals into the calculation. Once these issues are resolved using a structurally rigid molecule as an example, one can move on to the additional problems associated with flexible molecules such as

searching for different molecular conformers and Boltzmann averaging.

There have been several studies done using a variety of basis sets and calculation methods to determine a procedure which would yield a result that agrees the most with experimental values for  $[\alpha]_D$ . One of the more extensive works is that by Stephens et. al.<sup>8</sup> who performed calculation on the derivatives of 2,7,8-trioxabicyclo[3.2.1]octane, shown in Figure 1.3.



These molecules are geometrically rigid and depending on which derivative was used, there was only one or two low-energy conformers.

The authors concluded that the best  $[\alpha]_D$  results were obtained by using a DFT<sup>9</sup> functional such as B3LYP<sup>10,11</sup> and a large basis set such as aug-cc-pVDZ<sup>12,13</sup> or 6-311++G(2d,2p)<sup>14,15,16</sup>. The equilibrium geometry should be calculated using the B3LYP functional and 6-31G\* basis set<sup>14,15</sup>.

Their results furthermore showed that the calculations should be performed using Gauge

Independent Atomic Orbitals (GIAOs), whereas the default AOs give  $[\alpha]_D$  values which change when the molecule is translated in Cartesian space. To rid the calculation of this problem, the AOs are transformed to GIAOs by multiplying them by a complex phase factor ( $k$ ):<sup>17</sup>

$$k = \exp\left(\frac{-i}{2c}[B \times (R_A - R_G)] \cdot r\right) \quad (12)$$

where  $R_A$  is the position of the nucleus which is the center,  $r$  is the distance variable,  $R_G$  is the gauge origin or the origin of the coordinate system and  $B$  is the curl of the magnetic vector potential. The resulting Gauge Independent AO is:

$$X_A(r, R_A) = k \cdot N \cdot e^{-\alpha(r-R_A)^2} \quad (13)$$

where  $N \cdot e^{-\alpha(r-R_A)^2}$  is the default AO.

The results from calculation that involve simulating an applied magnetic field are dependent on the position of the molecule in the Cartesian reference frame, specifically, how the orbitals are placed with respect to the z-axis. In the GIAO approximation, each atomic orbital has a local origin placed at its center so the GIAOs also include a magnetic and/or electric vector potential factor.<sup>18,19</sup>

Stephens et. al.<sup>8</sup> concluded that the absolute error for  $[\alpha]_D$  values between the calculated values and experimental values was calculated to be  $\pm 30^\circ$  for the derivatives of 2,7,8-trioxabicyclo[3.2.1]octane.<sup>8</sup> This error was attributed to solvent effects, vibrational effects, errors associated with the basis set and DFT functional, and experimental error. The instrumental and human error in the experimental value is something that a theoretical chemist has no control over, generally the experimental values for  $[\alpha]_D$  have uncertainties of  $\pm 1$  to  $\pm 5$  degrees. The errors associated with either the basis set or with the method of calculation can be minimized to less than  $10^\circ$  provided one uses a large basis set with diffuse functions and a DFT method. The errors

due to the solvent effects and vibrational effects are likely the most significant contributors to the overall error. Currently, there is a lot of effort being directed into finding a solvent model that works well with this type of calculation. Research is also being done into trying to account for the perturbations in geometry caused by vibration and how to include them in the final computed  $[\alpha]_D$  value.

Wiberg et al.<sup>20</sup> have further shown that the B3LYP functional paired with a large basis set is the best way to get an accurate theoretical  $[\alpha]_D$  value. Furthermore, they have explored the effect of the diffuse basis functions in these large basis sets (such as aug-cc-pVDZ) on the correctness of the calculated value. According to this work, the diffuse p-orbital on the hydrogen atom is critical in finding the correct sign and magnitude for the  $[\alpha]_D$  value which makes the aug-cc-pVDZ basis set preferable to the 6-311++G(2d,2p) basis set, since the latter does not include a diffuse p-orbital on the protons. The authors have determined this fact by including the hydrogen diffuse p-orbital basis functions from the aug-cc-pVDZ basis set in the 6-311++G(2d,2p) basis set. This has resulted in a better agreement of the  $[\alpha]_D$  value between the 6-311++G(2d,2p) and the aug-cc-pVDZ basis sets as a function of geometry of the molecule they were studying.

Wiberg et al.<sup>20</sup> have also examined how the changes in equilibrium geometry and changes in geometry due to the vibration of a molecule can alter the calculated  $[\alpha]_D$  value. The authors have concluded that the geometry deformations due to the torsional motion of a molecule (butene in this case) have a large effect on the calculated  $[\alpha]_D$  value. The geometries due to the positive and negative distortions from the ground state were found and the  $[\alpha]_D$  values were calculated for both of the geometries for several vibrational modes which involve torsional deformations. The  $[\alpha]_D$  values were averaged between for the positive and the negative distortions from the ground state using the relative energies of the two geometries. It was found that the average  $[\alpha]_D$  values for the different torsional modes were in good agreement with the  $[\alpha]_D$  value calculated for the

ground state. Therefore, Wiberg et al.<sup>20</sup> have concluded that the geometry distortions due to the vibration of the molecule is a minor source of error in the calculation of the  $[\alpha]_D$  values.

### 1.5 Using Boltzmann averaging to find $[\alpha]_D$ for a molecule

Since most molecules are not conformationally rigid, some type of conformational search must be performed to obtain all the low-energy conformers which are likely to be present in a sample of molecules under the experimental conditions. Boltzmann statistics can be used to predict the probability with which any given conformer is present in an experimental sample. A molecular property such as the  $[\alpha]_D$  value is determined for each conformer and averaged according to the relative abundance of the conformer.

The Boltzmann distribution equation is used to determine the relative population of each conformer in a set:

$$\frac{n_i}{N} = \frac{e^{\frac{-E_i}{kT}}}{\sum_{i=1}^N e^{\frac{-E_i}{kT}}} \quad (14)$$

where  $E_i$  is the energy of the  $i^{\text{th}}$  conformer,  $n_i$  is the number of molecules that have  $E_i$  energy, and  $N$  is the total number of molecules available.

We can find the probability,  $p_i$ , with which the  $i^{\text{th}}$  conformer can exist at a given temperature by rearranging the above equation:

$$p_i = \frac{e^{\frac{-(E_i - E_0)}{kT}}}{Z} \quad \text{such that} \quad P = \sum_{i=1}^N p_i = 1 \quad (15)$$



where  $E_0$  is the lowest energy in the set and  $Z$  is the partition function:

$$Z = \sum_{i=1}^N e^{\frac{-(E_i - E_0)}{kT}} \quad (16)$$

The property such as  $[\alpha]_D$  is then calculated for each conformer and multiplied by the probability associated with it. Then the ensemble average  $[\alpha]_D$  is found from:

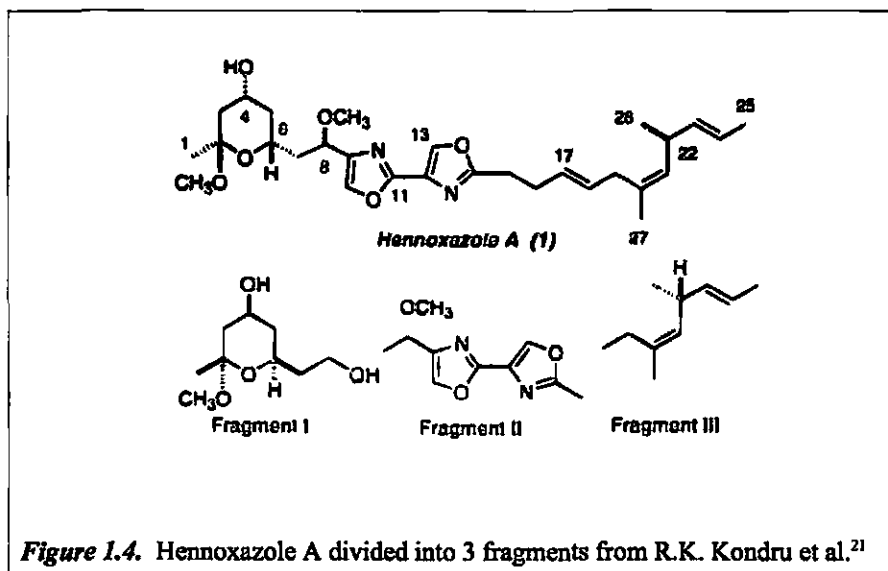
$$\text{Average } [\alpha]_D = \sum_{i=1}^N p_i \cdot [\alpha]_{Di} \quad (17)$$

This procedure is demonstrated more clearly using a small flexible molecule as an example in Section 2.5 of Chapter 2. At room temperature, 298 K, if one wishes to account for 99% of the conformations present in a sample, the relative energy of the low-energy conformers cannot exceed 11.3 kJ/mol relative to the energy of the lowest-energy conformer.

## 1.6 Review of the previous efforts to determine $[\alpha]_D$ for a natural product

Several synthesis and natural products groups have taken advantage of this new computational method to help with elucidating the absolute stereochemistry of large organic compounds. A typical procedure for a natural product with multiple chiral centers called Hennoxazole A is shown below as it appeared in the publication by R.K. Kondru et al.<sup>21</sup> The chiral centers are denoted in the figure below. The chirality associated with double-bonds is also marked.

Hennoxazole A shown in Figure 1.4 is too large a molecule to be dealt with as a single unit due to computational limitations. It was therefore divided into three fragments (Fragment I has three chiral centers, Fragment II has three chiral centers, and Fragment III has four chiral centers):



Kondru et al. justify the division of the molecule into three fragments by stating that van't Hoff's rule applies. That is the stereocenters which are separated by at least two bonds from each other are considered non-interacting. The sum of the individual specific optical rotations of the fragments can be added to approximate the specific optical rotation of the parent molecule, e.g. hennoxazole A.

The relative stereochemistry of fragment I was determined by NMR to be either all-S or all-R and the stereocenters in fragments II and III can have either an R or an S stereochemistry (for a total of 6 fragments that need to be analyzed). A Monte Carlo algorithm was used to find low energy conformers of each of the fragment at the MM2 force field level of theory using the Macromodel software package. The low energy conformations (within 3 kcal/mol from global minimum) were used to calculate  $[M]_D$  values using the 6-31G basis set (it is unclear whether they have used a DFT functional or Hartree-Fock). The  $[M]_D$  value is calculated by multiplying the  $[\alpha]_D$  value by the molecular weight of the molecule and dividing by 100. The  $[M]_D$  values were averaged using a Boltzmann distribution and the results are shown in Figure 1.5:

**Table 1. Calculated Molar Rotations for the Three Fragments in the Specific Enantiomeric Configurations, Averaged from Four Calculations**

fragment	$[M]_D \pm 2\sigma$ (configuration, sign)
I	$162 \pm 11$ (2R,4R,6R, -ve; 2S,4S,6S, +ve)
II	$105 \pm 22$ (8R, +ve; 8S, -ve)
III	$146 \pm 17$ (22R, +ve; 22S, -ve)

**Table 2. Molar Rotations of Different Stereoisomers of Hennoxazole A Calculated from Fragment Increments According to van't Hoff's Rule<sup>a</sup>**

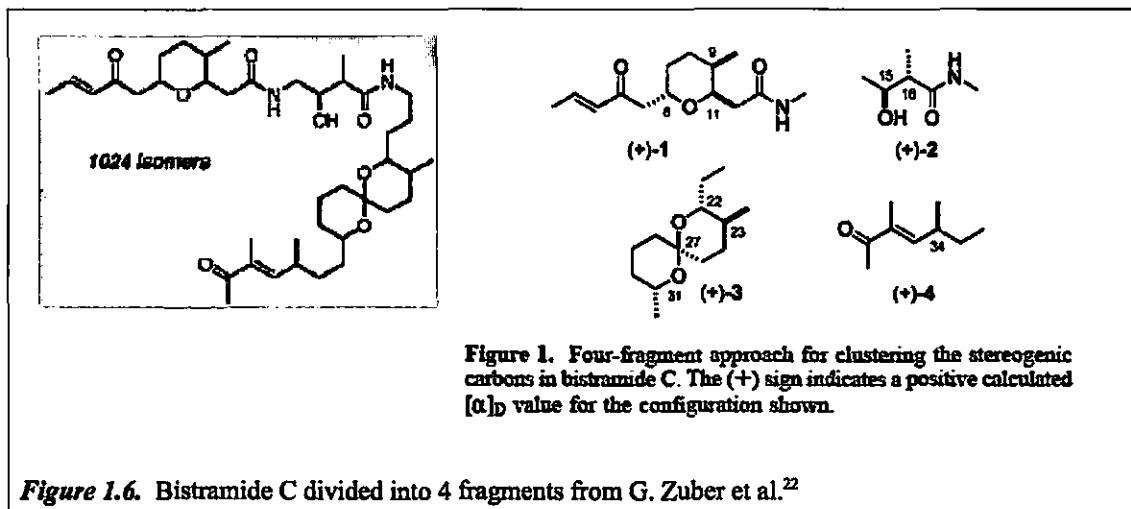
hennoxazole	configuration	$[M]_D \pm 2\sigma$	obsd $[M]_D$
10	(2S,4S,6S,8S,22R)	$+203 \pm 50$	+257
11	(2S,4S,6S,8S,22S)	$-89 \pm 50$	-48
12	(2S,4S,6S,8R,22S)	$+121 \pm 50$	+175
13	(2S,4S,6S,8R,22R)	$+413 \pm 50$	+459
14	(2R,4R,6R,8S,22R)	$-121 \pm 50$	(-175) <sup>b</sup>
15	(2R,4R,6R,8S,22S)	$-413 \pm 50$	(-459) <sup>b</sup>
1 <sup>c</sup>	(2R,4R,6R,8R,22S)	$-203 \pm 50$	-247
16	(2R,4R,6R,8R,22)	$+89 \pm 50$	(+48) <sup>b</sup>

<sup>a</sup> Observed molar rotations refer to data of natural or synthetic isomers. <sup>b</sup> Experimental molar rotation measurements for these stereoisomers are not available. Numbers in parentheses are based on the values measured for the corresponding enantiomers. <sup>c</sup> The natural product. Note that it can be assigned unambiguously by comparing the observed rotation angle with the eight computed values. Note, too, that all computed values correlate well with the observed rotation angles for the respective stereoisomers.

**Figure 1.5.** Table of results of calculated molar rotations for fragments of Hennoxazole A from R.K. Kondru et al.<sup>21</sup>

As it turns out, the experimental value for  $[M]_D$  for Hennoxazole A was  $-247^\circ$  and the closest calculated value was  $-203^\circ \pm 50^\circ$ , which would yield an absolute stereochemistry of (2R,4R,6R, 8R,22S) for this natural product. This result was confirmed via total synthesis of the molecule. It is somewhat surprising that such an accurate result was achieved using a level of theory much lower than that used by Stephens and Wiberg.

In a later work G. Zuber et al.<sup>22</sup> used a similar approach which was used to determine the absolute stereochemistry of Bistramide C which was separated into 4 fragments, shown on Figure 1.6:



The chiral centers are indicated in the same way as in Hennoxazole A excluding the chirality associated with double bonds. There are a total of 10 stereogenic carbons in this molecule: three in Fragment 1, two in Fragment 2, three in Fragment 3, and one in Fragment 4. This time the equilibrium geometries were calculated using BP86<sup>23,24</sup> DFT method and the SVP<sup>25</sup> basis set. The  $[M]_D$  value was calculated at the BP86/aug-cc-pVDZ level of theory. Again, the correlation was apparent enough to draw a conclusion on the nature of the absolute stereochemistry of Bistramide C. The calculated average  $[M]_D$  value for Bistramide C was  $97^\circ$  and the experimental value was  $70^\circ$ . The  $[M]_D$  value in this case has been overestimated by  $27^\circ$ .

## 1.7 Challenges concerning flexible molecules

The natural products in previous studies, some of which are described in Section 1.6 and some of which are mentioned in the review paper by T. D. Crawford<sup>6</sup>, have one common feature which makes computing the average specific optical rotation achievable. These molecules, while reaching a large size, can be broken down into fragments which contain non-interacting chiral centers by the virtue of van't Hoff's Rules.<sup>26</sup> Furthermore, each of these fragments has a relatively rigid structure with a small number of low-energy conformers. Thus

the search for these conformers can be easily accomplished with a variety of commercially available software. When a moderately flexible fragment was encountered, the difficulty of doing a conformational search is always mentioned in the article. For example, fragment III of Hennoxazole A<sup>21</sup> was considered a challenge due to its flexibility; a total of 20 low energy conformers were reported for this fragment.

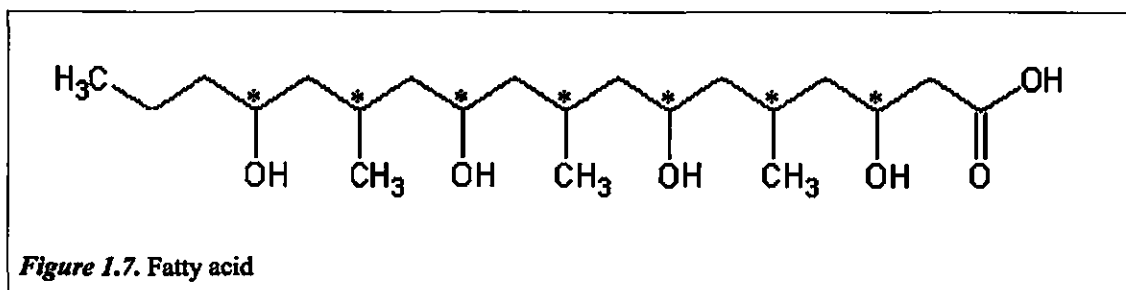
A much more challenging natural product for the theoretical prediction of the absolute stereochemistry will be one which has a lot of flexibility and the chiral centers are spaced in such a way that the whole molecule cannot readily be broken down into smaller fragments. For a flexible ring such as cycloheptadecane (C<sub>17</sub>H<sub>34</sub>) as many as 12,513 conformations have been reported using the MM3 force field method.<sup>27</sup> If one wanted to determine the specific rotation of a molecule similar in size and in flexibility, more than ten thousand conformers would need to be located, optimized at B3LYP/6-31G\* level of theory and sorted to find the low energy conformers. This will then be followed by the linear response calculation at B3LYP/aug-cc-pVDZ level of theory to obtain an  $[\alpha]_D$  value for each conformer. Using semiempirical methods such as AM1<sup>28</sup> or PM3<sup>29</sup> significantly improves the speed for optimizing the many local geometries of large molecules. The AM1 (Austin Model 1) method uses experimental data to fit atomic parameters that are used in the evaluation of the electron interaction integrals. This method is very flexible as the parameters can be adjusted to better fit a system of interest and it is known to yield reasonable results. However, the energy and the geometry at the AM1 level of theory is not reliable enough for calculating an accurate value for specific rotation, which requires that the calculated equilibrium geometry to be as close to the actual equilibrium geometry as possible.

A large number of conformers was encountered with some of the more flexible molecules (described in detail in Section 1.6), using conformational search routines in some of the

commercially available software packages. This indicates that the previous strategy for finding the total specific rotation value as suggested in the literature and outlined in Section 1.5 is going to be a big challenge. The more flexible the molecule, the more one needs to be careful when addressing the problem of searching for the many local minima on the Potential Energy Surface (PES): including cataloging and sorting the conformers, removing any duplicates, and finally sorting the low-energy minima from the high-energy minima. To do these tasks a computer program needed to be written which would be able to search for local minima of a flexible molecule while allowing the user to have as much control as possible over the parameters of the conformer search.

### 1.8 Outline of the thesis

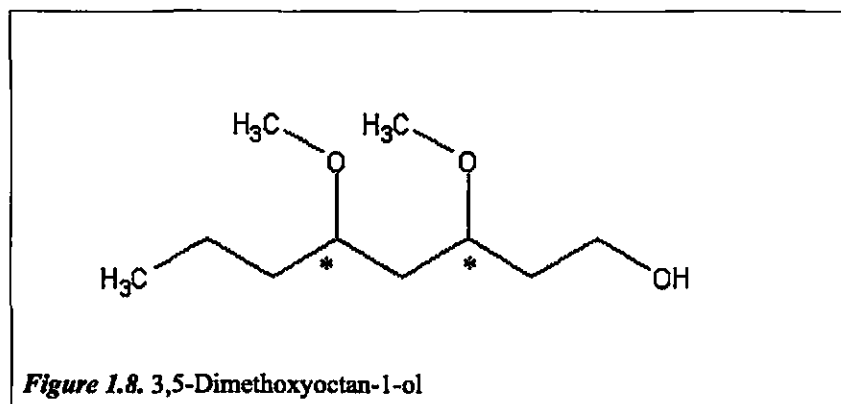
This work describes a method which can be used to perform QM assisted structure elucidation via specific optical rotation of a large and flexible molecule. There are several molecules presented in this study, similar in functionality but varying in the degree of complexity. Figure 1.7 shows a molecule with 7 chiral centers. This fatty acid is a natural product which has been isolated from the marine cyanobacterium *Lyngbya majuscula*. The full IUPAC name for this fatty acid is 3,7,11,15-tetrahydroxy-5,9,13-trimethyl-octadecanoic acid and it will be referred to as “fatty acid” for the remainder of this text.



The absolute stereochemistry of the fatty acid cannot be determined by NMR due to the

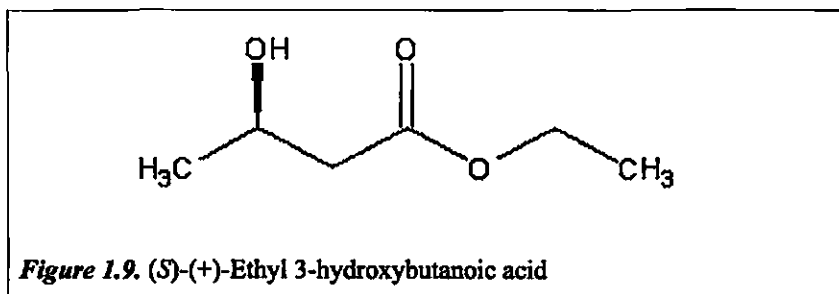
overlap of signals in the proton domain. Neither can X-ray crystallography be used as there is too little of this substance available. The only alternative for the determination of absolute stereochemistry is by an optical method such as specific rotation. In order for this method to work, a computational study similar to the studies described in previous sections must be performed. From a biological standpoint, the fatty acid is expected to have one of two relative stereochemistry configurations: either where all of the chiral centers are *syn* (all-*R* or all-*S*) or where the chiral centers are *anti* (*R,S,R,S,R,S,R* or *S,R,S,R,S,R,S*). This greatly reduces the number of different possibilities which one needs to investigate.

A molecule similar to the fatty acid but considerably smaller in size with only 2 chiral centers is shown on Figure 1.8. It was determined that the relative configuration of this molecule is *syn*, but whether it is the (*R,R*) or the (*S,S*) enantiomer is unknown. A calculation can be done to determine the magnitude on the sign for the  $[\alpha]_D$  of the (*R,R*) isomer and if the magnitude of rotation is in agreement with the experimental value, the absolute stereochemistry can be derived by comparing the sign (Chapter 3, Section 3.6).

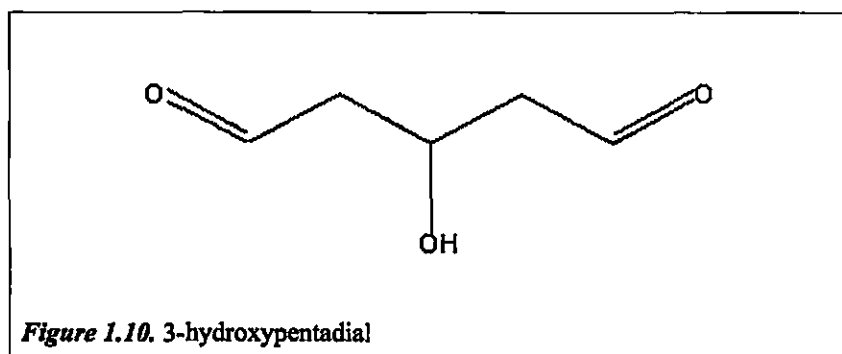


A chiral and flexible molecule with a known specific rotation value can be used to evaluate the ability of any new method to correctly predict the  $[\alpha]_D$  value and also to determine what magnitude of error one should expect with the approach which is being tested. For this

purpose, (*S*)-(+)-ethyl 3-hydroxybutanoic acid with only one chiral center, as shown in Figure 1.9 was used (Chapter 3, Section 3.5). The reported  $[\alpha]_D$  value for (*S*)-(+)-ethyl 3-hydroxybutanoic acid in chloroform at 20°C is +43° (Aldrich).



A flexible and achiral molecule shown on Figure 1.10 can be used to calibrate the conformer search program parameters so as to test whether all possible low energy conformers have been found. The final average weighted  $[\alpha]_D$  for such a molecule should be zero if the program is working correctly (Chapter 3, Section 3.3). Testing a new approach in this way will help to determine the computational time scale by which one should expect convergence on the correct  $[\alpha]_D$  value as well as to determine the most efficient search and sort parameters for molecules of the type and functionality such as the ones on Figures 1.7 and 1.8.



In Chapter 2, several molecules in the oxirane family are used to study the effects of the change in geometry on the calculated specific rotation value. In addition, the effects of basis sets



and methods of calculation were studied using small and rigid oxirane-related molecules.

Propanol was used to explore the effectiveness of Boltzmann statistics in the determination of  $[\alpha]_D$  of a sample.

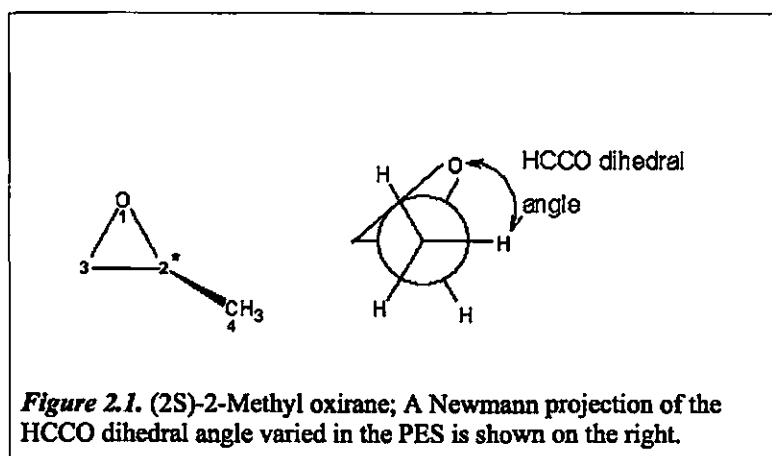
To facilitate calculating an  $[\alpha]_D$  value for flexible molecules such as the ones shown in Figures 1.7 and 1.8 a Monte Carlo (MC) code was written using the Fortran programming language. The GAMESS<sup>30</sup> software package was used for performing the *ab initio* calculations for the MC program described in Chapter 3, Section 3.2. This approach was successfully tested using the achiral and the chiral test cases and a predicted  $[\alpha]_D$  value was calculated for (*R,R*)-3,5-dimethoxyoctan-1-ol. In the last section of Chapter 3 we discuss the feasibility of running the calculations for the large fatty acid molecule.

## CHAPTER 2

### COMPUTATION OF AVERAGE SPECIFIC OPTICAL ROTATION VALUE: EVALUATION OF BASIS SETS AND THEORY LEVEL

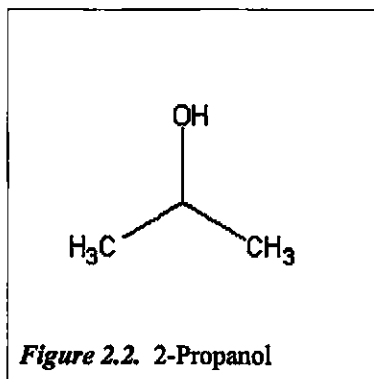
#### 2.1 Introduction

In this chapter, both the variation of the  $[\alpha]_D$  value as a function of molecular geometry and the influence of the level of theory on calculated  $[\alpha]_D$  values will be investigated. (2*S*)-2-methyl oxirane shown in Figure 2.1 is a good molecule to start investigating the influence of chain flexibility in a region close to the chiral center on the  $[\alpha]_D$  value associated with that chiral center. The three membered ring of (2*S*)-2-methyl oxirane is very strained and rigid such that only the rotation of the methyl group about the HC(4)C\*(2)O dihedral angle is thermodynamically accessible. This means that the experimental value for optical rotation of methyl oxirane is only due to this one conformer and no Boltzmann averaging is needed to compare the computational result with the experimental  $[\alpha]_D$  value.



Calculations were also performed where all of the protons of (2*S*)-2-methyl oxirane were replaced with heavy atoms to test the influence of diffuse orbitals on the  $[\alpha]_D$  calculation.

Fluorine was chosen for this purpose so that the sigma bonds as they appear on (2*S*)-2-methyl oxirane were preserved.



2-Propanol (Figure 2.2) was subjected to a similar examination as the members of the oxirane family to show that even in achiral molecules, geometry perturbations can greatly affect the calculated specific rotation value. Also, 2-propanol was used to demonstrate the effectiveness of using Boltzmann statistics in the computation of the average  $[\alpha]_D$  value for a molecule.

## 2.2 (2*S*)-2-Methyl oxirane: effect of HCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets

**Introduction.** Calculating the  $[\alpha]_D$  value as a function of the HC(4)C\*(2)O dihedral angle on (2*S*)-2-methyl oxirane was done to examine how the slight changes in geometry of the methyl group can affect the  $[\alpha]_D$  value. Also, it is a chance to compare different basis sets and methods in more detail as opposed to just determining which method/basis set gives the most accurate value for the optimized geometry. This type of analysis has been done previously by others for butane, ethane, and propane<sup>20</sup>, as well as for 2-chloropropionitrile.<sup>32</sup>

**Method.** The PES was determined as a function of the HC(4)C\*(2)O dihedral angle from 0° to 120° in increments of 10° using HF and B3LYP methods with a 6-31G\* basis set and the

GAMESS software<sup>30</sup>. For each fixed dihedral angle the other coordinates of the molecule were optimized. It provided a set of partially optimized structures. For each of the partially optimized structures the  $[\alpha]_D$  was calculated using the Gaussian03<sup>31</sup> software using HF/6-31G\*, HF/6-31G\*\*, and HF/aug-cc-pVDZ methods as well as the B3LYP/6-31G\* and B3LYP/aug-cc-pVDZ methods.

Table 2.1a

Angle	Energy (a.u)	Rel. E (kcal/mol)	$[\alpha]_D$ 6-31G*	$[\alpha]_D$ aug-cc-pVDZ	$[\alpha]_D$ 6-31G**
0	-191.90633	2.119	-0.46	37.91	8.76
10	-191.90735	1.477	-19.35	35.88	-20.63
20	-191.90843	0.802	-29.07	25.3	-37.79
30	-191.90927	0.270	-28.59	9.95	-42.73
40	-191.90971	0.000	-17.72	-6.01	-33.82
50	-191.90963	0.047	2.29	-18.69	-10.8
60	-191.90907	0.402	27.47	-25.31	22.21
70	-191.90813	0.988	51.34	-24.6	57.47
80	-191.90704	1.670	66.18	-16.84	83.81
90	-191.90609	2.266	66.12	-3.07	91.05
100	-191.90557	2.595	50.76	14.51	75.04
110	-191.90566	2.540	26.3	30.98	43.36
120	-191.90633	2.119	1.48	39.71	8.76
130	-191.90735	1.476	-17.87	37.26	-19.63

Table 2.1b

Angle	Energy (a.u)	Rel. E (kcal/mol)	$[\alpha]_D$ 6-31G*	$[\alpha]_D$ aug-cc-pVDZ
0	-193.10520	2.053	7.66	60.27
10	-193.10617	1.445	-21.89	53.2
20	-193.10721	0.793	-38.38	37.79
30	-193.10804	0.276	-42.06	13.8
40	-193.10847	0.000	-31.6	-11.07
50	-193.10844	0.024	-7.09	-28.6
60	-193.10793	0.342	27.07	-36.92
70	-193.10705	0.895	62.82	-34.63
80	-193.10600	1.553	88.82	-21.78
90	-193.10506	2.144	94.83	-0.73
100	-193.10451	2.490	76.99	25.07
110	-193.10456	2.457	43.44	49.07
120	-193.10520	2.053	7.66	60.27
130	-193.10619	1.433	-21.89	54.52

Table 2.1. Relative energy dependence of (2S)-2-methyl oxirane on the HCCO dihedral angle using (a) HF/6-31G\* and (b) B3LYP/6-31G\* partially optimized energy calculations; and the  $[\alpha]_D$  values at those geometries using different basis sets.

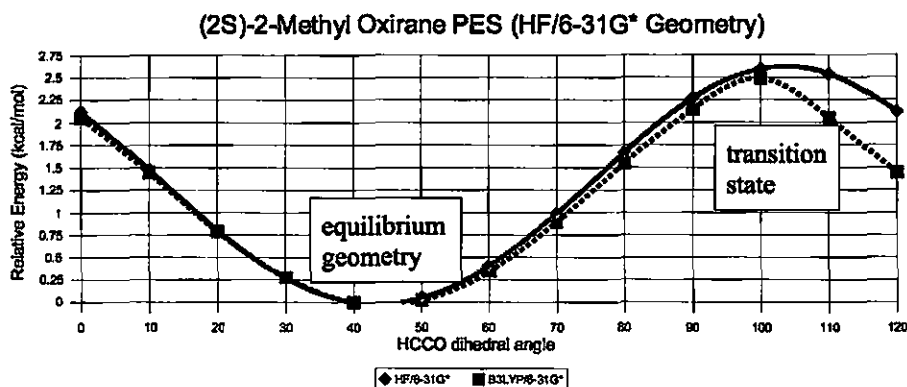


Figure 2.3a. Potential Energy Surface (PES) of (2S)-2-methyl oxirane along the HCCO dihedral angle

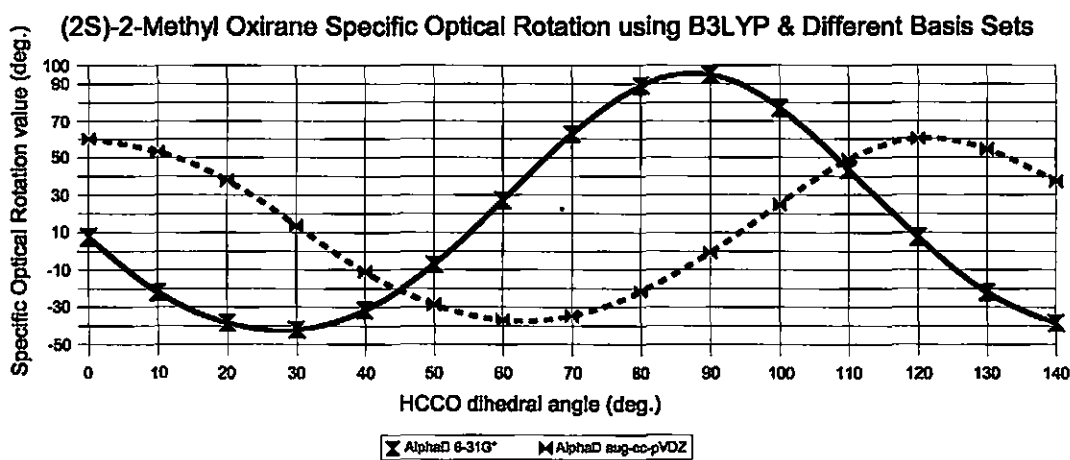


Figure 2.3b. Specific optical rotation values using the B3LYP method and different basis sets

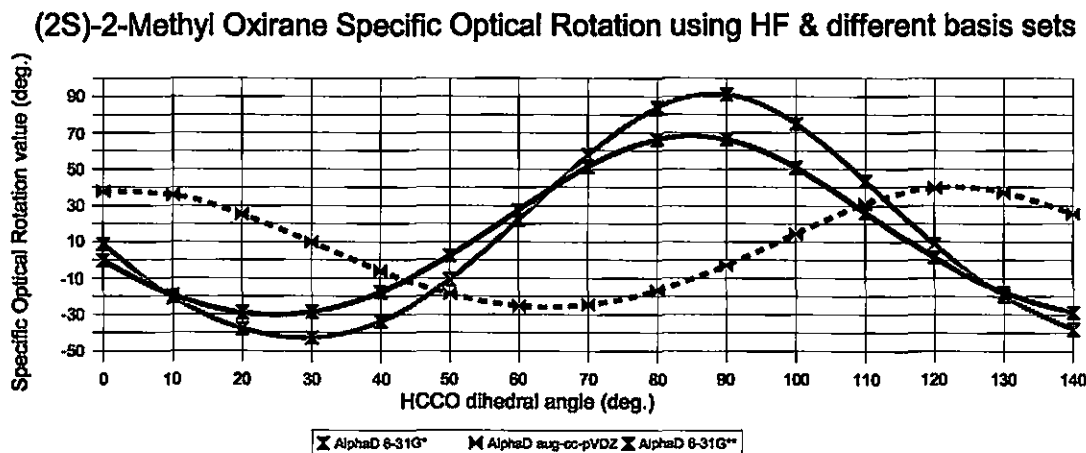


Figure 2.3c. Specific optical rotation values using the HF method and different basis sets

Figure 2.3(a-c). The PES and the calculated  $[\alpha]_D$  variation along the HCCO dihedral angle at each constrained geometry is shown. The optical rotation values for the same geometries are different for different basis sets. The  $[\alpha]_D$  values appear to agree best only at either the equilibrium geometry or at the transition state. The  $[\alpha]_D$  was always calculated using the same method that was used to optimize.

**Results.** Table 2.1 and Figure 2.3 show the variation of relative energy and specific rotation of (2*S*)-2-methyl oxirane as a function of the the HCCO dihedral angle. The specific rotation varies as a function of the dihedral angle and it can vary by as much as a 100° between the staggered and the eclipsed conformers depending on the basis set used to calculate the  $[\alpha]_D$  value. The low-energy staggered conformer has a HCCO dihedral angle of 40°,  $\Delta E_{\min} = 0$  kcal/mol (Table 2.1a) and the high-energy eclipsed conformer has the a HCCO dihedral angle of 100°,  $\Delta E_{\max} = 2.595$  kcal/mol using HF/6-31G\* and  $\Delta E_{\max} = 2.49$  kcal/mol using B3LYP/6-31G\* (Table 2.1a).

The specific optical rotation values are clearly very dependent on the basis set used as well as on the HCCO dihedral angle. Figure 2.4 shows the plot of  $[\alpha]_D$  values using B3LYP/6-31G\* vs.  $[\alpha]_D$  values using B3LYP/aug-cc-pVDZ for each of the HCCO dihedral angles. Figures 2.3(b,c) and 2.4 show that two basis sets to give  $[\alpha]_D$  curves which have large phase and amplitude differences. These differences can be attributed to the diffuse basis functions which are present in the aug-cc-pVDZ basis set but not in the 6-31G\* basis set.

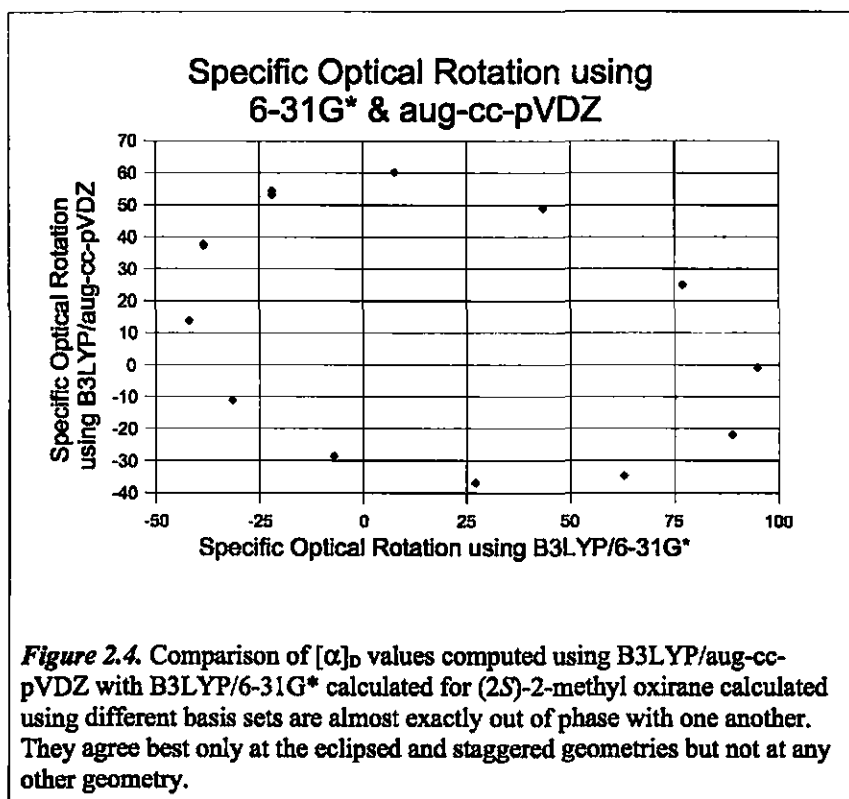
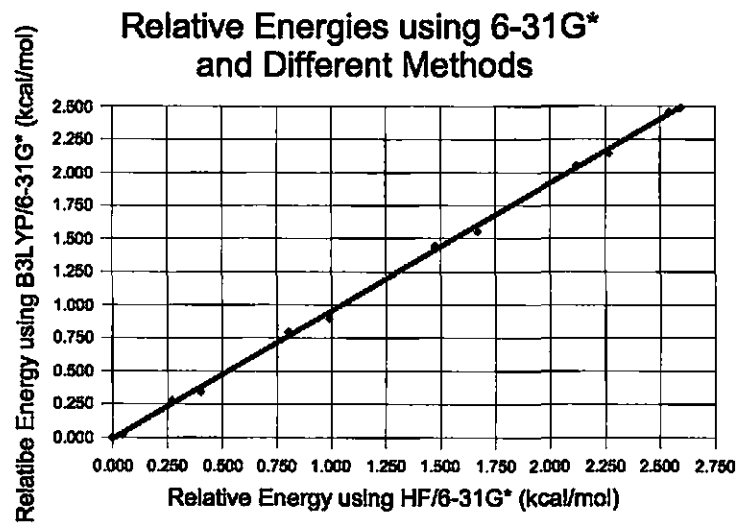
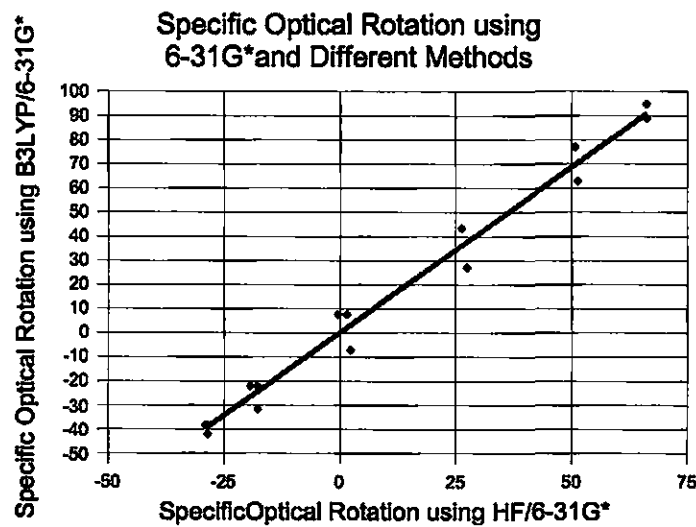


Figure 2.5 compares the relative energies and  $[\alpha]_D$  values computed using HF/6-31G\* and B3LYP/6-31G\* methods. There is excellent correlation between these sets of data. The DFT method in this case has no advantage over the HF method when the same basis set is used, as the  $[\alpha]_D$  values match well between the two as shown on Figures 2.5a and 2.5b. The  $[\alpha]_D$  values computed using the HF method are always smaller in absolute value than the  $[\alpha]_D$  values computed using the DFT method.



**Figure 2.5a.** Comparison of HF/6-31G\* with B3LYP/6-31G\* relative energies for (2*S*)-2-methyl oxirane



**Figure 2.5b.** Comparison of HF/6-31G\* with B3LYP/6-31G\* [ $\alpha$ ]<sub>D</sub> values for (2*S*)-2-methyl oxirane

**Figure 2.5(a,b).** There is good correlation between the relative energies and the specific rotation values between the HF and DFT methods. Only the 6-31G\* basis was used to illustrate the correlation. The correlation between HF and B3LYP methods is just as good with aug-cc-pVDZ basis set.



One key point to note here is that regardless of how different the two curves are in Figures 2.3(b,c), they intersect at the equilibrium geometry with zero relative energy (where the HCCO dihedral angle is about 45°) to give an  $[\alpha]_D$  value of approximately -12°. The experimental value for  $[\alpha]_D$  of (2*S*)-2-methyl oxirane as a neat solution is -14° (Aldrich). This suggests that an accurate calculated  $[\alpha]_D$  value can be achieved using the HF/6-31G\* geometry and either of the two basis sets at the equilibrium geometry. As the partially optimized geometry proceeds to distort and move away from the equilibrium, the  $[\alpha]_D$  values become vastly different for the different basis sets. An interesting note is that the  $[\alpha]_D$  curves as calculated with both basis sets cross each other in the proximity of equilibrium geometry as well at the transition state. This hints that there is a  $dE/d\theta$  dependence in the  $[\alpha]_D$  calculations (where  $\theta$  is the HCCO dihedral angle). The  $[\alpha]_D$  values are in better agreement when that derivative is zero regardless of whether the second order derivative,  $d^2E/d\theta^2$ , is positive or negative. If it were possible to experimentally trap the high-energy eclipsed geometry of (2*S*)-2-methyl oxirane and measure its specific rotation value, it would likely be in agreement with the calculated value of approximately 30°.

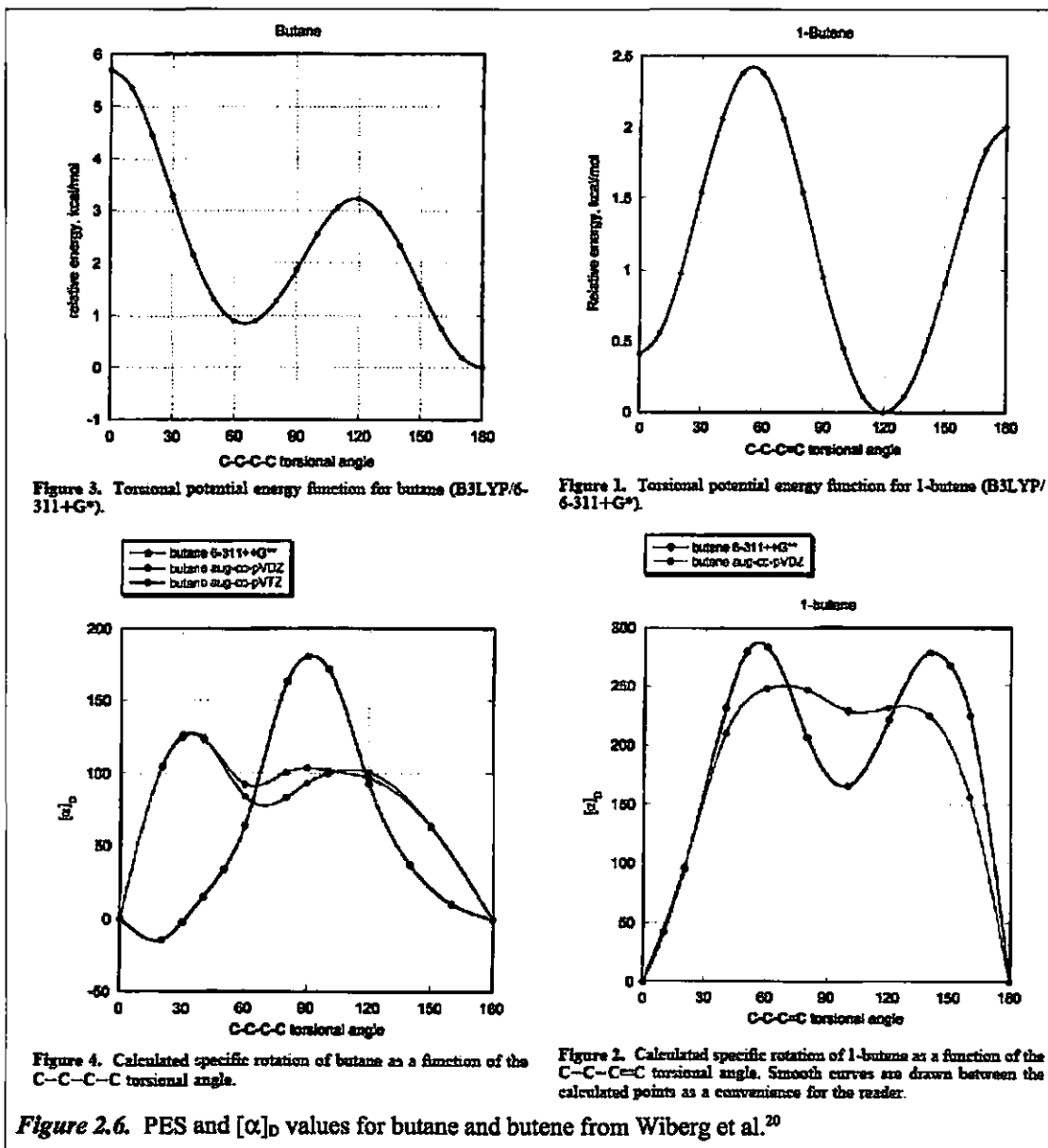
It is also noted that around the equilibrium geometry the  $[\alpha]_D$  curves vary linearly with the dihedral angle. This means that for torsional vibrational modes such as the one shown in Figure 2.3a, the first order vibrational average correction term to the  $[\alpha]_D$  value at equilibrium geometry given in Equation 18 is zero:

$$\int \Psi_v(\theta) [\alpha]_D(\theta) \Psi_v(\theta) d\theta = 0 \quad (18)$$

where  $\Psi_v(\theta)$  is the  $v^{\text{th}}$  vibrational state for a given normal mode. The probability distribution for the  $v^{\text{th}}$  vibrational state is denoted as  $|\Psi_v(\theta)|^2$  and it is an *even* function. Since the  $[\alpha]_D(\theta)$  varies linearly with  $\theta$ , it is an *odd* function. The resulting average value is thus zero. This suggests that there is no contribution to the  $[\alpha]_D$  value from this vibrational mode.

### 2.3 Hexafluoro-methyl oxirane: effect of FCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets

**Introduction.** The phase shift between the  $[\alpha]_D$  curves shown in Figures 2.3(b,c) and Figure 2.4 was attributed to the diffuse p-orbital basis function on the hydrogen atoms which is present in the aug-cc-pVDZ basis set but is not present in either of the 6-31G\* or 6-31G\*\* basis sets. Figure 2.6 shows a similar results found by Wiberg et al.<sup>20</sup> with butane and butene.



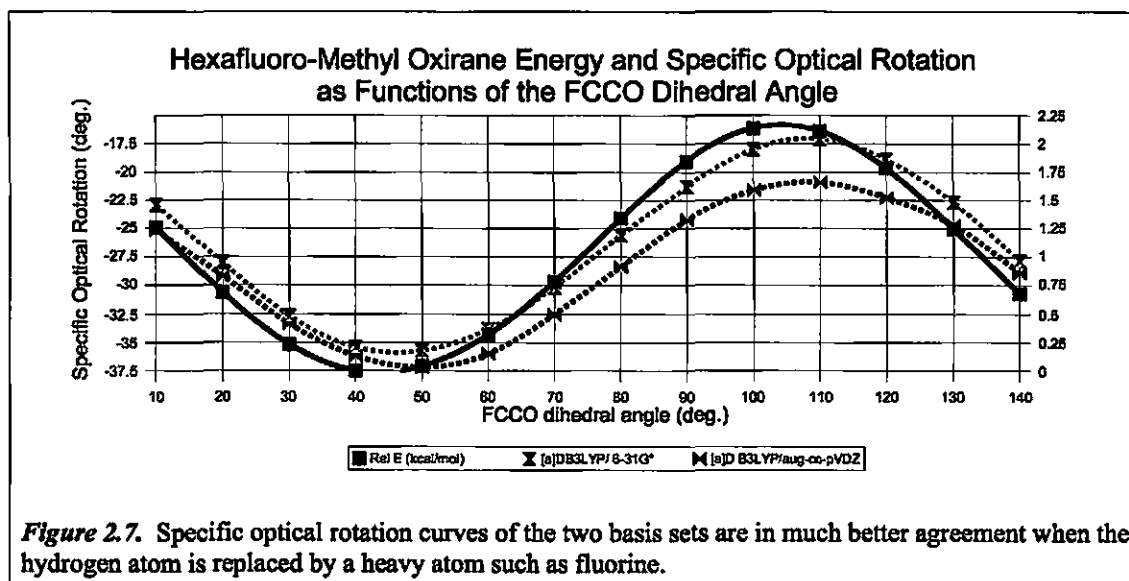
The authors have placed the diffuse p-orbital basis functions on the hydrogen atoms from the aug-cc-pVDZ basis set onto the 6-31++G(2d,2p) basis set and recalculated the  $[\alpha]_D$  values. The  $[\alpha]_D$  values were in much better agreement between the aug-cc-pVDZ basis set and the 6-31++G(2d,2p) (with diffuse p-orbitals) basis set for the geometry which had a 30° CCCC dihedral angle along of butane.

**Method.** To show in a way different from the work by Wiberg et al.<sup>20</sup> that the diffuse p-orbitals on hydrogen atoms play an important role in  $[\alpha]_D$  calculations, all of the hydrogens in (2*S*)-2-methyl oxirane were replaced with fluorine atoms so that the diffuse p-orbitals on hydrogen atoms are no longer an issue. The PES was found by rotating the CF<sub>3</sub> group along the FCCO dihedral angle and the  $[\alpha]_D$  values were calculated using B3LYP/aug-cc-pVDZ and B3LYP/6-31G\* levels of theory.

**Results.** Table 2.2 and Figure 2.7 show the variation of relative energy and specific rotation of hexafluoro-methyl oxirane as a function of the the FCCO dihedral angle. The specific rotation varies as a function of the dihedral angle and it can vary by as much as a 30° between the eclipsed and the staggered conformers depending on the basis set used to calculate the  $[\alpha]_D$  value. The change in the specific rotation value between the eclipsed and the staggered conformation is much less here than in the hydrogen analog. The low-energy staggered conformer has a FCCO dihedral angle of 45°,  $\Delta E_{\text{min}} = 0$  kcal/mol (Table 2.2) and the high-energy eclipsed conformer has a FCCO dihedral angle of 105°,  $\Delta E_{\text{max}} = 2.14$  kcal/mol (Table 2.2). The rotation barrier for this molecule is 0.3 kcal/mol less than the rotation barrier for the hydrogen analog.

OCCF dihedral	B3LYP/6-31G* E (au)	Rel E (kcal/mol)	$[\alpha]_D$ B3LYP/6-31G*	$[\alpha]_D$ B3LYP/aug-cc-pVDZ
10	-788.527350	1.26	-22.93	-25.15
20	-788.528267	0.69	-28.03	-29.21
30	-788.528989	0.23	-32.69	-33.36
40	-788.529360	0	-35.43	-36.2
50	-788.529293	0.04	-35.72	-37.18
60	-788.528849	0.32	-33.82	-36
70	-788.528114	0.78	-30.18	-32.53
80	-788.527221	1.34	-25.58	-28.37
90	-788.526414	1.85	-21.3	-24.21
100	-788.525952	2.14	-17.98	-21.6
110	-788.525994	2.11	-17.06	-20.88
120	-788.526516	1.78	-18.87	-22.27
130	-788.527371	1.25	-22.67	-24.71
140	-788.528282	0.68	-27.88	-28.92

**Table 2.2.** Relative energy dependence of hexafluoro-methyl oxirane on the FCCO dihedral angle using B3LYP/6-31G\* partially optimized energy calculations; and the  $[\alpha]_D$  values at those geometries using different basis sets.



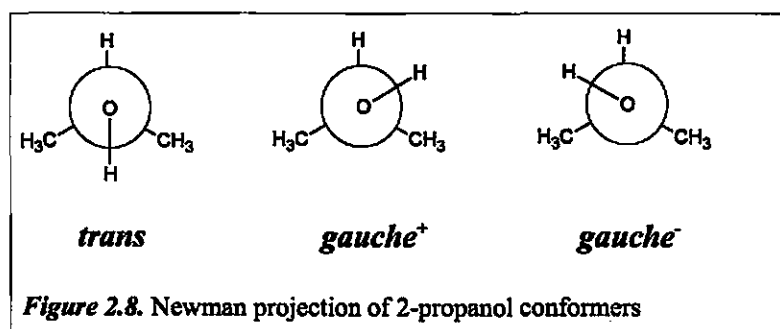
The diffuse p and d orbitals on the heavy atoms still contribute to the phase shift between the two optical rotation functions using different basis sets, but the shift is not as dramatic. Also, both optical rotation curves are phase shifted with respect to the relative energy in such a way that their minima coincide. This means that with an increase in basis set size, the calculation for  $[\alpha]_D$  value is converging much more consistently when the hydrogen atoms are removed. Yet, according to Wiberg et al.<sup>20</sup> the best agreement with the experimental results is obtained by

including the diffuse-p orbitals on hydrogens. This could indicate that there is something intrinsically incorrect with the algorithm for this calculation.

Figure 2.7 also shows that the  $[\alpha]_D$  value for the lowest-energy hexafluoro-methyl oxirane structure is approximately  $-36^\circ$ . It is more negative than the  $[\alpha]_D$  value of  $-12^\circ$  for the (2*S*)-2-methyl oxirane, but it has the same sign.

#### 2.4 2-Propanol: effect of HCCO dihedral angle on $[\alpha]_D$ value using different methods and basis sets; calculating the average $[\alpha]_D$

**Introduction.** Dealing with the effects of the hydroxy group on a chiral center poses several other problems than those for a simple methyl group. The hydroxy functional group does not have a 3-fold symmetry as the methyl functional group and therefore the effect on optical rotation of varying the HOCH dihedral angle from  $0^\circ$  to  $360^\circ$  needs to be examined separately. 2-Propanol was chosen (as opposed to hydroxy-methoxirane) because it is not only small but also because the experimental value for  $[\alpha]_D$  is zero because the molecule is achiral. The next few paragraphs will also demonstrate how breaking the symmetry of the molecule due to the rotation about the HOCH dihedral angle can produce a very substantial non-zero value for  $[\alpha]_D$ .



2-Propanol can form three distinct conformers which are all local minima in a geometry optimization as shown on Figure 2.8. These include the *trans*, *gauche*<sup>+</sup>, and *gauche*<sup>-</sup> conformers

of 2-propanol, and they correspond to the HOCH dihedral angles of  $180^\circ$ ,  $60^\circ$ , and  $-60^\circ$  respectively.

**Method.** For each of the conformers shown in Figure 2.8, the effect of rotation of one of the methyl groups for a fixed HOCH dihedral angle of the hydroxy group on  $[\alpha]_D$  values was studied. The 2-propanol PES was sampled by keeping the OH fixed in one of the three minima shown in Figure 2.8 while rotating one of the methyl groups in the same fashion as was done with (2*S*)-2-methyl oxirane in Section 2.2 but in  $20^\circ$  increments. Thus, three PES curves were constructed using the relative energies obtained with the HF/6-31G\* method and the  $[\alpha]_D$  values were determined using aug-cc-pVDZ and 6-31G\* basis sets.

Boltzmann averaging of the specific optical rotation values was done using the calculated relative energies of the three optimized conformers and their calculated  $[\alpha]_D$  values as described in Chapter 1, Section 1.5.

**Results.** Figure 2.9(a,b) shows the variation of relative energy and specific optical rotation of the *gauche*<sup>+</sup> and the *gauche*<sup>-</sup> conformers of 2-propanol as a function of the the HCCO dihedral angle from one of the methyl groups. The specific optical rotation varies as a function of the dihedral angle and it can vary by as much as a  $100^\circ$  between the staggered and the eclipsed conformers depending on the basis set used to calculate the  $[\alpha]_D$  value. The change in the specific optical rotation value between the staggered and the eclipsed conformation is about the same as in (2*S*)-2-methyl oxirane. The low-energy staggered conformer has a  $60^\circ$  HCCO dihedral angle and  $\Delta E_{\min} = 0$  kcal/mol (Figure 2.9a) and the high-energy eclipsed conformer has a  $120^\circ$  HCCO dihedral angle and  $\Delta E_{\max} = 3.6$  kcal/mol (Figure 2.9a). The rotation barrier for one of the methyl groups on 2-propanol methyl group is 3.6 kcal/mol, about 1 kcal/mol higher than the rotation barrier for the (2*S*)-2-methyl oxirane methyl group. Figure 2.9b shows that the  $[\alpha]_D$  value is approximately  $37^\circ$  for the *gauche*<sup>+</sup> conformer and  $-37^\circ$  for the *gauche*<sup>-</sup> conformer.

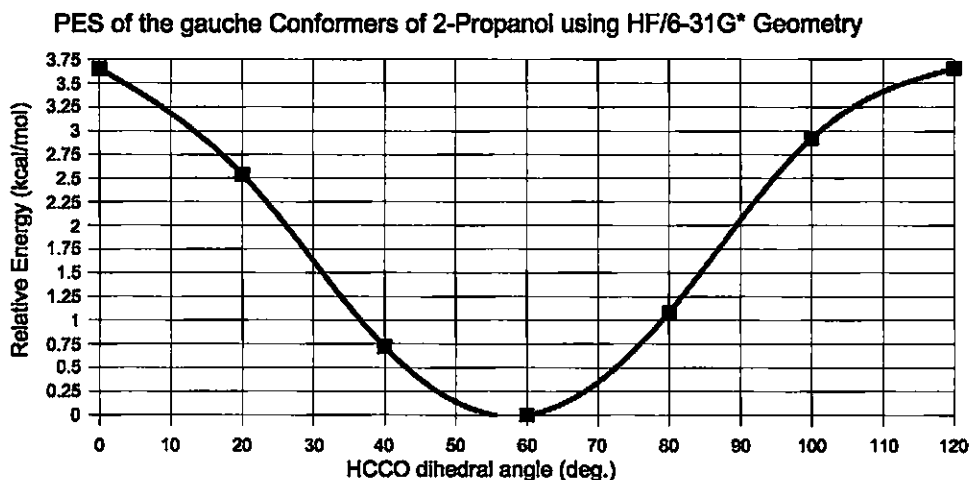


Figure 2.9a. The PES of 2-propanol *gauche*<sup>+</sup> and *gauche*<sup>-</sup> conformers. The PES is the same for the two gauche mirror image isomers. The HCCO dihedral angle of the local minimum for 2-propanol gauche is ~ 60° and of the local maximum is ~ 120°

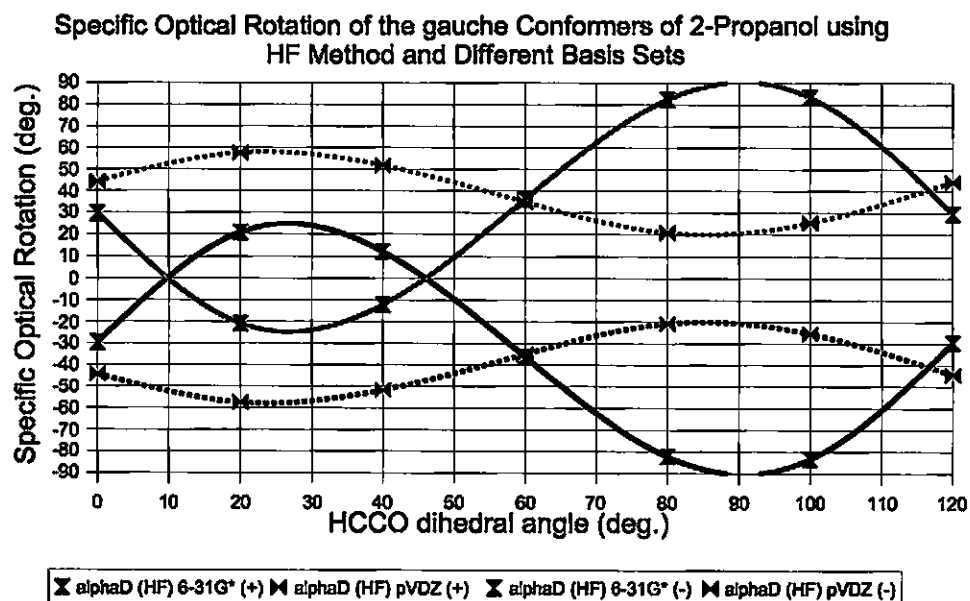
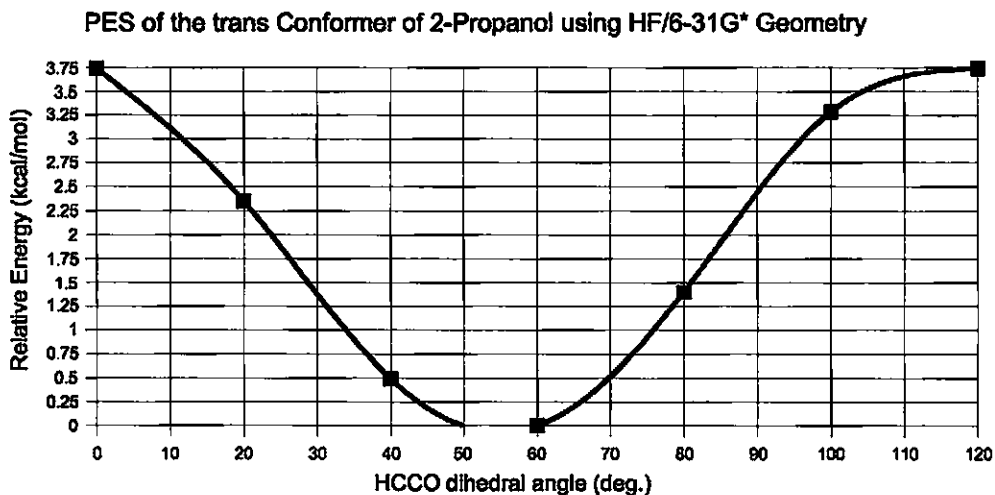


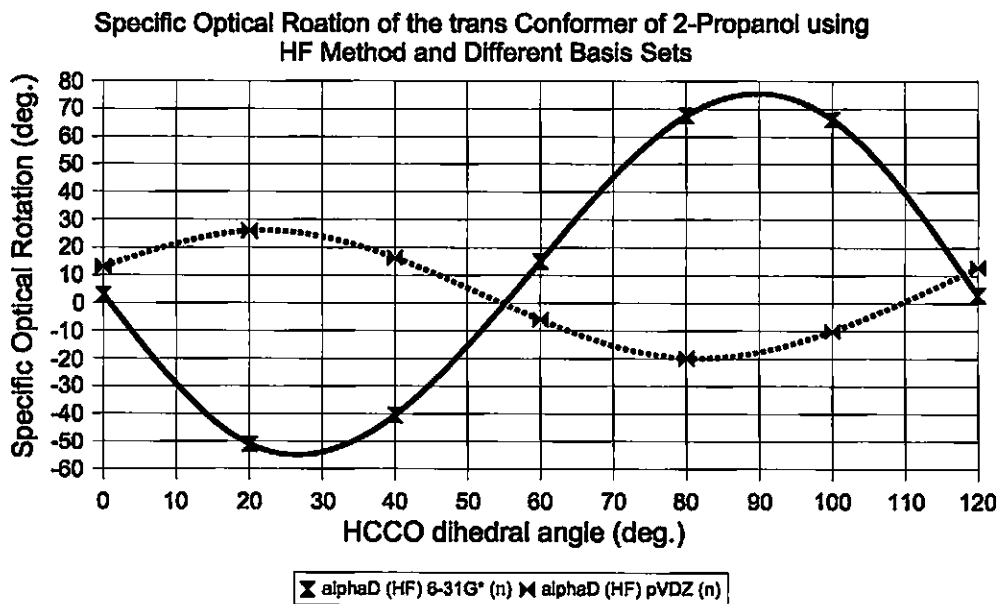
Figure 2.9b. The  $[\alpha]_D$  values for 2-propanol *gauche*<sup>+</sup> and *gauche*<sup>-</sup> conformers. The *gauche*<sup>+</sup> isomer has an  $[\alpha]_D$  value of ~ 37° while the *gauche*<sup>-</sup> isomer has a value of ~ -37°

Figure 2.9(c,d) shows the variation of relative energy and specific optical rotation of the *trans* conformer of 2-propanol as a function of the HCCO dihedral angle. The specific optical rotation varies as a function of the dihedral angle and it can vary by as much as a  $100^\circ$  between the staggered and the eclipsed conformers depending on the basis set used to calculate the  $[\alpha]_D$  value. The  $[\alpha]_D$  value for the *trans* conformer is  $0^\circ$ . The change in the specific optical rotation value between the eclipsed and the staggered conformation is about the same as in (2*S*)-2-methyl oxirane. The low-energy staggered conformer has a  $55^\circ$  HCCO dihedral angle and  $\Delta E_{\min} = 0$  kcal/mol (Figure 2.9c) and the high-energy eclipsed conformer has a  $105^\circ$  HCCO dihedral angle and  $\Delta E_{\max} = 3.75$  kcal/mol (Figure 2.9c). The rotation barrier for one of the methyl groups on 2-propanol methyl group is 3.75 kcal/mol, 1.1 kcal/mol higher than the rotation barrier for the (2*S*)-2-methyl oxirane methyl group.





**Figure 2.9c.** The PES the *trans* conformer of 2-propanol. The HCCO dihedral angle of the local minimum for 2-propanol *trans* is  $\sim 55^\circ$  and of the local maximum is  $\sim 110^\circ$



**Figure 2.9d.** The  $[\alpha]_D$  values for the *trans* conformer of 2-propanol. The  $[\alpha]_D$  of the *trans* isomer is  $0^\circ$  at the equilibrium geometry since it has a  $C_s$  symmetry in this case.

The investigation of the effect on  $[\alpha]_D$  value of the rotating -OH group has shed light on some of the complications that one will encounter when doing conformational averaging of  $[\alpha]_D$  values for a molecule where a hydroxy group is present. In the case of 2-propanol, the experimental  $[\alpha]_D$  value is zero due the fact that the two gauche isomers are mirror images and both are present in solution in equal amounts. This is best shown with a full demonstration of the Boltzmann averaging of  $[\alpha]_D$  values for 2-propanol. The basis of this approach was discussed in Section 1.5. Equations 14 through 17 were used to construct Table 2.3.

	$E$ (a.u.)	$\Delta E$ (kcal/mol)	$p_i$	$[\alpha]_D$	$p_i [\alpha]_D$
<i>gauche</i> <sup>*</sup>	-193.1154	0	0.38	36.12	13.76
<i>gauche</i> <sup>*</sup>	-193.1154	0	0.38	-36.11	-13.76
<i>trans</i>	-193.1149	0.28	0.24	0	0
					<b>sum</b>
					<b>0</b>

**Table 2.3.** The average  $[\alpha]_D$  values for 2-propanol was calculated using Boltzmann statistics. The geometries for the three low-energy conformers were optimized using HF/6-31G\* and the  $[\alpha]_D$  values were calculated using HF/aug-cc-pVDZ.

The average value of  $[\alpha]_D$  adds up to zero as is what is expected for an achiral molecule even though some of the conformations of the molecule have a non-zero  $[\alpha]_D$  value. Note that the gauche conformers are degenerate in energy and also lower in energy than the trans isomer by 0.28 kcal/mol.

In the presence of a chiral center, the mirror image of each conformer will not be present in solution containing only one enantiomer and the optical rotation values will not numerically cancel each other to zero as they do in achiral molecules. This imbalance is what ultimately leads to the final value for the specific rotation.

## 2.5 Summary of Chapter 2

In this chapter, the effects of perturbations to a molecule's geometry on the  $[\alpha]_D$  values was investigated. It was determined that the  $[\alpha]_D$  value is very sensitive to changes in geometry and for the most accurate results one should be as close to the equilibrium geometry as possible. Using B3LYP/6-31G\* or better is recommended for geometry optimizations. The  $[\alpha]_D$  value is best calculated using the aug-cc-pVDZ method. However, it was shown with the example of (2*S*)-2-methyl oxirane and hexafluoro-methyl oxirane molecules that accurate results for the  $[\alpha]_D$  value can also be calculated at the B3LYP/6-31G\* level of theory provided that the geometry is well optimized.

The Boltzmann averaging technique was demonstrated using an achiral molecule 2-propanol which has three low energy conformers. The average  $[\alpha]_D$  value was calculated to be zero for 2-propanol even though two of the low-energy 2-propanol conformers had a non-zero  $[\alpha]_D$  value.

## CHAPTER 3

# THE DEVELOPMENT AND USAGE OF A MONTE CARLO CONFORMER SEARCH PROCEDURE FOR THEORETICALLY DETERMINING THE AVERAGE SPECIFIC OPTICAL ROTATION VALUE

### 3.1 Statistics and $[\alpha]_D$ in large molecules (Introduction)

As was shown in Section 2.5 for the 2-propanol molecule, statistical averaging of the specific optical rotation values of the different conformers present in solution using weights based on the Boltzmann distribution leads to a theoretical  $[\alpha]_D$  value that is close to the experimental  $[\alpha]_D$  value. With the computational resources available at the present time, even though in theory it is possible to find all of the low-energy conformers of large and flexible molecules, this is an unreasonable task given the computational cost and time required. This is especially true for flexible molecules, such as the ones outlined in Section 1.7, since they have many low-energy conformers. The number of low-energy conformers can roughly be estimated to be  $3^n$  where  $n$  is the number of dihedral angles in the molecule. A more practical approach is to try to compute an approximate distribution of the conformers present in the cell of the instrument while keeping in mind that not every single conformer is necessarily present, sampled, or found computationally.

For the purpose of simulating a distribution of conformers in a sample, a Monte Carlo program was developed. This program will use Boltzmann statistics to generate a distribution of low energy conformers which are then used to calculate the average  $[\alpha]_D$  value of the sample. While it is possible to systematically scan the PES by varying each of the dihedral angles in the molecule being investigated, this is not a practical approach. In the case of 2-propanol only the HCCO dihedral angle need be varied to find the three low-energy conformers. The three low-

energy conformers where the hydroxy group is in the staggered orientation is shown in the Newman projection in Figure 2.8. For a molecule with five dihedral angles, there are at least  $3^5$  (243) conformers where the groups on each of the five dihedral angles have a staggered orientation, but only some of these conformers will have a low energy. In addition, there are probably many more conformers which do not have groups in an exactly staggered orientation and yet are still low in energy. Thus, in a systematic search each of the dihedral angles need to be altered in increments of at least  $30^\circ$ ; which means that for a molecule with five dihedral angles  $12^5$  (248,832) conformations would need to be optimized and compared to discard any duplicates and then sorted by energy. More than 200,000 local optimizations of a flexible molecule of modest size would need to be performed to find all of the low-energy conformers. This high a number makes the systematic search approach impractical and consequently we use a Monte Carlo approach.

The Monte Carlo (MC) method works on the principle that a solution to a quantitative problem can be estimated via a statistical sampling of parameters that define the final solution or the expectation value. Expectation values can be calculated using appropriate physical parameters with the addition of random parameters. A random number generator is used to provide for the variation of parameters that the user wants to vary randomly (in this case, the geometry of a molecule is varied randomly). The set of thus obtained expectation values can then be averaged to estimate a theoretical value. Because the expectation values were computed using one or more sets of random parameters, where no two sets are identical, one can tell that the MC routine is working correctly if the result such as a property calculation converges to the same average value. Also, probabilities of any of the expectation values can be determined using this method with any user defined probability function.<sup>27,34</sup> The Monte Carlo Metropolis Sampling method is a well known and a much used algorithm devised by Metropolis et al.<sup>33</sup>

The current problem of finding low-energy conformers can be solved by using a Metropolis Monte Carlo method by randomly changing the geometry of a molecule and calculating its energy. Assuming that the system in question is at thermal equilibrium, the Boltzmann distribution function is used for the purpose of statistical analysis of the energy of different conformers. So a random new conformer can be accepted or rejected based on the probability that the energy of this new conformer is significant in the Boltzmann distribution.

Each time a new geometry is generated, it is accepted if its energy ( $E_2$ ) is less than that of the preceding geometry ( $E_1$ ). If the new energy is greater, the Boltzmann ratio ( $r$ ) for the two geometries is computed. The Boltzmann ratio will be a number greater than zero and less than one.

$$r = \frac{\exp(-E_2/Tk_B)}{\exp(-E_1/Tk_B)} \quad (19)$$

This ratio is compared to a random number between zero and one and if  $r$  is greater than the random number, then the new geometry is also accepted.

In the current project, new geometries were generated by varying the dihedral angles of the molecules of interest at random. The new geometries were partially optimized by keeping the values for the altered dihedral angles constant. Only those dihedral angles which are not three-fold degenerate were chosen to be varied. This includes all of the dihedral angles on a molecule save for those containing a hydrogen on the terminal methyl groups. Furthermore, as discussed in Chapter 2, the rotation barrier for a methyl group is small and in the range of 2-4 kcal/mol, so varying the dihedral angle associated with the methyl group should not produce any new local minima.

### 3.2 Outline of the Monte Carlo program

1. In the MC program, several input parameters are required from the user:

1. An initial internal coordinate GAMESS input file generated for the molecule being investigated
  2. A list of Z-matrix coordinates for the dihedral angles along the backbone of the molecule which are to be varied
  3. Temperature at which the simulation is to take place
  4. Maximum number of geometries to sample (controls the time limit for the simulation)
  5. Number of low energy conformers should this search produce before stopping
  6. Random number generator<sup>35</sup> (RNG) seed
  7. Threshold for comparison of new conformers to determine their uniqueness (in Hartrees for energy and in degrees for dihedral angles).
  8. Level of theory to be used in the calculations for GAMESS (AM1 in this case)
2. An initial full optimization is performed using the GAMESS input file provided and the final energy and geometry are stored as the first conformer.
  3. A new set of coordinates is generated using the RNG to modify the dihedral angles specified by the user.
  4. A partial optimization is done (keeping only the new set of dihedral angles fixed) using the newly generated geometry.
  5. The new energy from (4) is compared to the old energy from (2). If the energy from (4) is lower than the energy from (2), go to (5.1). If the energy from (4) is higher than the energy from (2), calculate  $\exp(-\Delta E/kT)$  and go to (5.2)
    1. The new values for energy and geometry from (4) replace the stored values from (2) and the program moves on to (6).
    2. If the  $\exp(-\Delta E/kT)$  is less than a number from RNG, the new values for energy and

geometry from (4) replace the old values from (2) and the program returns to (3).

3. If the  $\exp(-\Delta E/kT)$  is greater than a number from RNG, the the energy and geometry from (4) are retained and the program returns to (3).
6. The accepted geometry is fully optimized and the new energy and the new set of dihedral angles is recorded.
7. The energy and the dihedral angles of this geometry are compared to the stored values from the previous full optimization runs.
  1. The difference between the new energy and all of the stored energies is determined and compared to a user-defined threshold value for energy difference.
  2. If the difference in energy is greater than the threshold, the new fully optimized conformer is added to the existing array.
  3. If the difference in energy is less than the threshold, the user-specified dihedral angles are compared between the new geometry and the geometries in the array.
  4. If the difference in the dihedral angles is greater than the user-defined threshold, the new conformer is added to the array.
  5. If both the difference in energy and the difference in dihedral angles are less than the user-defined thresholds, the new conformer is considered a duplicate and is discarded.
8. At the end of the simulation, all of the stored conformers are sorted by energy and B3LYP1/6-31G\* geometry optimization input files are generated.

The higher level of theory geometries are compared in a similar fashion to the way the geometries in the Monte Carlo program are compared and duplicate conformers are discarded.

Only the conformers that are energetically accessible at room temperature are used in the calculation of the specific optical rotation. The actual program routines are included in the



appendix along with sample user input parameters. Along with the main code, modifications for each molecule, and copies of supporting files are also included. A flowchart summary of the nine steps of the MC program is shown in Figure 3.1.

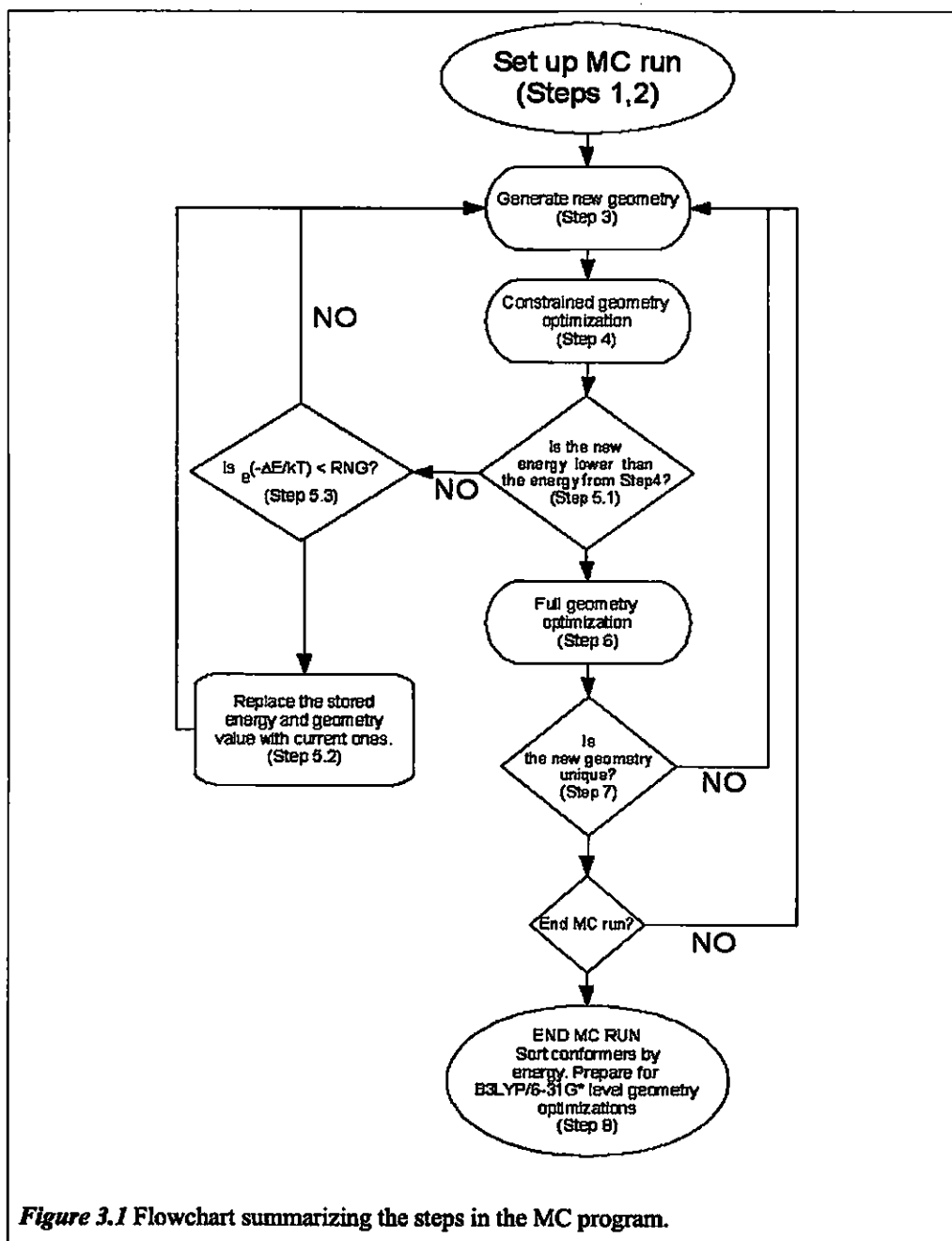
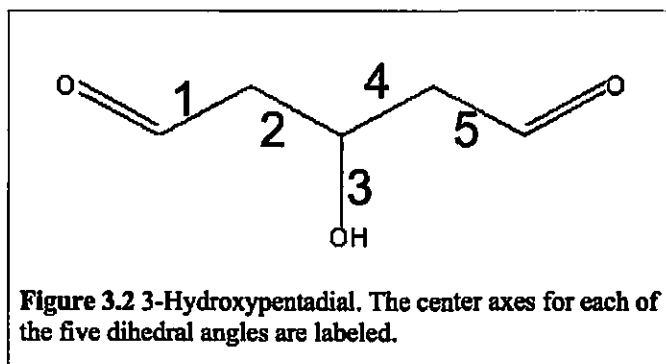


Figure 3.1 Flowchart summarizing the steps in the MC program.

### 3.3 Test of computational procedure using an achiral molecule

**Introduction.** 3-Hydroxypentadiol in Figure 3.2 is an achiral molecule similar in functionality to the target chiral molecules in this study. The experimental  $[\alpha]_D$  value for this and all achiral molecules is zero by definition. In this case, there are individual conformers with non-zero  $[\alpha]_D$  values. Since the molecule is achiral the mirror image of each conformer is present in the sample and thus the weighted  $[\alpha]_D$  value averages out to zero as previously demonstrated using 2-propanol as an example in Chapter 2. 3-Hydroxypentadiol, is larger than 2-propanol and therefore has more conformations present at room temperature. This molecule will be used as a test to make sure that the Monte Carlo (MC) program is finding all of the local minima properly. If the program works correctly, pairs of mirror image conformers will be found and the average  $[\alpha]_D$  value will converge on zero.



**Method.** 3-Hydroxypentadiol has five important dihedral angles along the backbone that were sampled in the MC simulation (shown on Figure 3.2). Only the center axes of the dihedral angles are labeled for clarity. See the GAMESS input file in the appendix for the listing of dihedral angles defined for each molecule. Several MC runs were performed using different temperatures and RNG seeds.<sup>35</sup> The RNG seed is used to generate a random set of dihedral angles for each new conformer in the MC run. Each MC run was allowed a maximum runtime of 24 hours. In a 24hr period 50,000 conformations were generated. For each set, 90 unique conformers

were isolated and sorted by energy (from low-energy to high-energy) at the AM1 level of theory. Before the higher level optimization was done at B3LYP/6-31G\* level of theory, the MC results were divided into sets of 30 conformers each in order of their generation. This was done in order to observe any time-dependent trends that the average specific rotation may have with respect to the MC sampling routine. The results from different MC runs were combined to find the average  $[\alpha]_D$  value. The average specific rotation was computed at 300 K using the Boltzmann distribution to weight of individual values as described in detail in Chapter 1, Section 1.5.

**Results.** The results from the first run are presented in Tables 3.1. This run was performed at T=200K with a RNG seed of 111324. The results from this run were divided into three samples of 30 conformers each according to the order in which the conformers were found.

Several of the different geometries obtained with AM1 have optimized to the same B3LYP/6-31G\* geometry . While the order of energies at the AM1 level was not preserved at the DFT level, the general trend was that the low-energy conformers from the AM1 calculation were also the low-energy conformers at the B3LYP/6-31G\* level. Thus the AM1 can be considered a good method for quick and relatively accurate molecular energy calculations to be used in the MC routine. The specific rotation calculation was done using only the lower energy conformers with a very conservative cutoff limit of approximately 15 kJ/mol or 3 kcal/mol above the lowest energy conformer.

The 1 – 29 sample had an average  $[\alpha]_D$  value of 226° (Table 3.1a), the 30 – 59 sample had an average  $[\alpha]_D$  value of -57° (Table 3.1b), and the 60 – 89 sample had an average  $[\alpha]_D$  value of 264° (Table 3.1c). The data from Tables 3.1(a-c) was then combined and all duplicate conformers removed, giving an overall  $[\alpha]_D$  value for 1 – 89 sample of 140° (Table 3.1d). It is clear from the data that while there were some mirror image pairs sampled, not every conformation yet had a match. In fact, the large positive result is a direct consequence of one low

energy conformer with  $[\alpha]_D$  value of  $300^\circ$  being present while its mirror conformer was still missing from this set. To find more conformers one could increase the temperature in the MC code to get more conformer accepts. One could also change the RNG seed or the starting geometry. One could also increase the time duration of the simulation. However, the most efficient time for the runs for this molecule was determined to be a 24 hour period. Table 3.1d contains all of the conformers found in the MC run with RNG seed of 111234. Also included are the geometries that were found to be unique using the B3LYP/6-31G\* method but were too high in energy to contribute significantly to the average specific optical rotation value. Out of 50,000 conformers that were generated by the MC code, 89 were found to have unique optimized geometries at the AM1 level of theory. Out of the 89 AM1 conformers, 39 conformers optimized to unique geometries at the B3LYP/6-31G\* level, with only 21 conformers being within the 20 kJ/mol range from the lowest-energy conformer. An extremely conservative 20 kJ/mol cutoff was chosen as it was the first run. In the later runs, the cutoff was based more on the percent population value as opposed to the relative energy value. Any conformer which is present in an abundance of less than 1% was not considered to be a significant contributor to the average  $[\alpha]_D$  value and the  $[\alpha]_D$  calculation was not performed for it.

**Table 3.1a**

	1-29 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con437	-420.98699	0	100	62.25	299.53	186.46
2	con797	-420.98574	3.29	26.75	16.65	81.54	13.58
3	con133	-420.98475	5.89	9.42	5.87	361.69	21.22
4	con2	-420.98455	6.42	7.63	4.75	-171.3	-8.14
5	con34	-420.98415	7.45	5.05	3.15	314.7	9.9
6	con1	-420.98397	7.93	4.16	2.59	54.86	1.42
7	con262	-420.98365	8.77	2.97	1.85	-50.2	-0.93
8	con401	-420.98363	8.83	2.9	1.8	-21.4	-0.39
9	con749	-420.98289	10.77	1.33	0.83	436.29	3.62
10	con23	-420.98166	14	0.36	0.23	-206.47	-0.47
11	con873	-420.97986	18.71	0.06	0.03	136.35	0.05
	sum			160.64			226.32

**Table 3.1b**

	30-59 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con1012	-420.98574	0	100	40.34	84.8	34.2
2	con1342	-420.98573	0.02	99.28	40.04	-81.86	-32.78
3	con1303	-420.98478	2.52	36.42	14.69	-362.73	-53.29
4	con1213	-420.98287	7.53	4.89	1.97	-435.26	-8.58
5	con1220	-420.98258	8.28	3.61	1.46	102.88	1.5
6	con1649	-420.98201	9.79	1.97	0.8	182.17	1.45
7	con1299	-420.98113	12.11	0.78	0.31	-225.22	-0.71
8	con1308	-420.98111	12.15	0.77	0.31	225.17	0.7
9	con972	-420.97988	15.37	0.21	0.08	135.53	0.12
	sum			247.92			-57.4

**Table 3.1c**

	60-89 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con5233	-420.98698	0	100	79.89	299.74	239.47
3	con5307	-420.98476	5.83	9.65	7.71	357.66	27.58
4	con5206	-420.98411	7.54	4.86	3.89	-96.17	-3.74
5	con5251	-420.98397	7.91	4.19	3.35	56.58	1.89
6	con5406	-420.98364	8.78	2.96	2.37	-96.17	-2.27
7	con5408	-420.98362	8.82	2.92	2.33	-5.54	-0.13
8	con2991	-420.98201	13.06	0.53	0.43	182.93	0.78
9	con1899	-420.97990	18.59	0.06	0.05	136.34	0.06
	sum			125.17			263.64

**Table 3.1.** The lowest-energy conformers from the MC run with RNG seed of 111234 for 3-hydroxypentadiol. There was a total of 89 conformers isolated at the AM1 level of theory and that set was broken down into 3 subsets: 1-29 (a), 30-59 (b), and 60-89 (c). The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

**Table 3.1d**

	1-89 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con437	-420.98699	0	100	47.42	299.53	142.04
2	con797	-420.98574	3.29	26.75	12.68	81.54	10.34
3	con1342	-420.98573	3.3	26.59	12.61	-81.86	-10.32
4	con1303	-420.98478	5.8	9.76	4.63	-362.73	-16.78
5	con133	-420.98475	5.89	9.42	4.47	361.69	16.16
6	con2	-420.98455	6.42	7.63	3.62	-171.3	-8.2
7	con34	-420.98415	7.45	5.05	2.4	314.7	7.54
8	con5206	-420.98411	7.56	4.82	2.29	-96.17	-2.2
9	con1	-420.98397	7.93	4.16	1.97	54.86	1.08
10	con262	-420.98365	8.77	2.97	1.41	-50.2	-0.71
11	con5406	-420.98364	8.8	2.94	1.39	-96.17	-1.34
12	con401	-420.98363	8.83	2.9	1.37	-21.4	-0.29
13	con5408	-420.98362	8.84	2.89	1.37	-5.54	-0.08
14	con749	-420.98289	10.77	1.33	0.63	436.29	2.76
15	con1213	-420.98287	10.81	1.31	0.62	-435.26	-2.7
16	con1220	-420.98258	11.57	0.97	0.46	102.88	0.47
17	con1649	-420.98201	13.08	0.53	0.25	182.17	0.46
18	con23	-420.98166	14	0.36	0.17	-208.47	-0.36
19	con1299	-420.98113	15.4	0.21	0.1	-225.22	-0.22
20	con1308	-420.98111	15.44	0.21	0.1	225.17	0.22
21	con1899	-420.97990	18.62	0.06	0.03	136.34	0.04
22	con766	-420.97763	24.59	0.01	0.00		
23	con1703	-420.97761	24.63	0.01	0.00		
24	con733	-420.97669	27.05	0.00	0.00		
25	con2325	-420.97658	27.33	0.00	0.00		
26	con743	-420.97624	28.23	0.00	0.00		
27	con1126	-420.97585	29.25	0.00	0.00		
28	con1608	-420.97569	29.66	0.00	0.00		
29	con3239	-420.97568	29.69	0.00	0.00		
30	con1139	-420.97551	30.13	0.00	0.00		
31	con1159	-420.97511	31.2	0.00	0.00		
32	con358	-420.97404	33.99	0.00	0.00		
33	con2151	-420.97387	34.45	0.00	0.00		
34	con2749	-420.97277	37.32	0.00	0.00		
35	con820	-420.97277	37.33	0.00	0.00		
36	con1272	-420.97264	37.66	0.00	0.00		
37	con1960	-420.97263	37.7	0.00	0.00		
38	con1383	-420.96936	46.29	0.00	0.00		
39	con1382	-420.96935	46.31	0.00	0.00		
				sum			sum
				210.88			139.91

**Table 3.1d.** The low-energy conformers from the MC run with RNG seed of 111234 for 3-hydroxypentadiol. This is the entire 1-89 set of unique optimized geometries at the AM1 level of theory. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. The  $[\alpha]_D$  value was calculated only for the first 21 conformers (no more than 20 kJ/mol above the lowest-energy conformer). The higher-energy conformers (22-39) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

The first attempt to find more conformers was to change the RNG seed. The results for the second set are presented in Tables 3.2. This run was performed at T=200K with RNG seed of 171982. It was shown in the previous run that the conformers in the last set of 60 – 89 were just repeats of the conformers in the earlier sets so in the interest of saving computer time, the simulation was stopped after finding 60 conformers.

**Table 3.2a**

	1-29 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con67	-420.98698	0	100	36.69	-299.94	-110.047
2	con23	-420.98697	0.03	98.71	36.22	300.69	108.903
3	con144	-420.98572	3.32	26.38	9.68	82.71	8.005
4	con324	-420.98476	5.85	9.59	3.52	384.25	12.811
5	con35	-420.98455	6.39	7.71	2.83	171.07	4.838
6	con298	-420.98454	6.43	7.6	2.79	-167.76	-4.680
7	con1	-420.98397	7.92	4.18	1.53	54.86	0.842
8	con38	-420.98395	7.97	4.1	1.5	-56.12	-0.843
9	con123	-420.98363	8.8	2.94	1.08	53.57	0.577
10	con405	-420.98363	8.81	2.92	1.07	-41.8	-0.448
11	con463	-420.98361	8.86	2.86	1.05	46.45	0.488
12	con110	-420.98360	8.89	2.83	1.04	-26.11	-0.271
13	con39	-420.98357	8.98	2.73	1	107.53	1.079
			sum	272.56		sum	21.25

**Table 3.2b**

	1-29 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con674	-420.98699	0	100	69.23	299.34	207.224
3	con1060	-420.98569	3.41	25.45	17.62	80.48	14.178
4	con533	-420.98475	5.89	9.42	6.52	361.61	23.594
5	con502	-420.98456	6.39	7.72	5.35	-171.81	-9.184
6	con796	-420.98289	10.77	1.33	0.92	436.88	4.032
7	con1953	-420.98200	13.09	0.52	0.36	184.05	0.669
			sum	144.45		sum	240.513

**Table 3.2.** The lowest-energy conformers from the MC run with RNG seed of 171982 for 3-hydroxypentadiol. There was a total of 59 conformers isolated at the AM1 level of theory and that set was broken down into 2 subsets: 1-29 (a), 30-59 (b). The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Table 3.2c contains all of the conformers found in the MC run with RNG seed of 171982. Also are included the geometries that were found to be unique using the B3LYP/6-31G\* method but were too high in energy to be included in the calculation of the average specific optical rotation value. Out of 50,000 conformers that were generated by the MC code, 89 were found to have unique optimized geometries at the AM1 level of theory. Out of the 89 AM1 conformers, 32 conformers optimized to unique geometries at the B3LYP/6-31G\* level, with only 15 conformers lying within the 15 kJ/mol range from the lowest-energy conformer. The 1 – 29 sample had an average  $[\alpha]_D$  value of  $21^\circ$  (Table 3.2a) and the 30 – 59 sample had an average  $[\alpha]_D$  value of  $-240^\circ$  (Table 3.2b). The data from Table 3.2a and 3.2b was then combined and all repeating conformers removed giving an overall  $[\alpha]_D$  value for 1 – 89 sample of  $24^\circ$  (Table 3.2c).



**Table 3.2c**

	1-59 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con67	-420.98698	0	100	36.29	-299.94	-108.840
2	con23	-420.98698	0	99.81	36.22	299.34	108.416
3	con144	-420.98572	3.32	26.38	9.57	82.71	7.918
4	con533	-420.98475	5.88	9.48	3.44	361.61	12.439
5	con502	-420.98456	6.37	7.77	2.82	-171.81	-4.842
6	con35	-420.98455	6.39	7.71	2.8	171.07	4.785
7	con1	-420.98397	7.92	4.18	1.52	54.86	0.833
8	con38	-420.98395	7.97	4.1	1.49	-56.12	-0.834
9	con123	-420.98363	8.8	2.94	1.07	53.57	0.571
10	con405	-420.98363	8.81	2.92	1.06	-41.8	-0.443
11	con463	-420.98361	8.86	2.86	1.04	46.45	0.483
12	con110	-420.98360	8.89	2.83	1.03	-26.11	-0.269
13	con39	-420.98357	8.98	2.73	0.99	107.53	1.067
14	con796	-420.98289	10.75	1.34	0.49	436.88	2.126
15	con1953	-420.98200	13.08	0.53	0.19	184.05	0.353
16	con1026	-420.97989	18.63	0.06	0.02		
17	con1081	-420.97763	24.56	0.01	0.00		
18	con1180	-420.97763	24.56	0.01	0.00		
19	con256	-420.97712	25.91	0.00	0.00		
20	con864	-420.97673	26.93	0.00	0.00		
21	con87	-420.97665	27.13	0.00	0.00		
22	con1263	-420.97623	28.22	0.00	0.00		
23	con876	-420.97620	28.32	0.00	0.00		
24	con331	-420.97594	29	0.00	0.00		
25	con2108	-420.97587	29.18	0.00	0.00		
26	con1456	-420.97570	29.63	0.00	0.00		
27	con803	-420.97568	29.67	0.00	0.00		
28	con977	-420.97540	30.42	0.00	0.00		
29	con2144	-420.97419	33.59	0.00	0.00		
30	con906	-420.97388	34.42	0.00	0.00		
31	con224	-420.97278	37.3	0.00	0.00		
32	con1192	-420.97278	37.31	0.00	0.00		
				sum			sum
				275.58			23.762

**Table 3.2c.** The low-energy conformers from the MC run with RNG seed of 171982 for 3-hydroxypentadiol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. The  $[\alpha]_D$  value was calculated for the first 15 conformers. The higher-energy conformers (16-32) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

The results from the two MC runs (one with a RNG seed of 111234 and the other with a RNG seed of 171982) were combined to find if the average  $[\alpha]_D$  value would converge to zero as predicted or if more MC runs are necessary. The conformers listed in Table 3.1d and in Table

3.2c were pooled together and repeating conformers were removed. Table 3.3 shows the low-energy conformers from both MC runs. The average  $[\alpha]_D$  was found to be 3°.

	name	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con67	-420.98698	0	100	29.63	-299.94	-88.868
2	con23	-420.98698	0	99.81	29.57	299.34	88.522
3	con797	-420.98574	3.27	28.75	7.93	81.54	6.46
4	con1342	-420.98573	3.29	28.59	7.88	-81.86	-6.45
5	con1303	-420.98478	5.79	9.76	2.89	-362.73	-10.48
6	con533	-420.98475	5.88	9.48	2.81	361.61	10.156
7	con502	-420.98456	6.37	7.77	2.3	-171.81	-3.953
8	con35	-420.98455	6.39	7.71	2.28	171.07	3.907
9	con34	-420.98415	7.43	5.05	1.5	314.7	4.71
10	con5206	-420.98411	7.55	4.82	1.43	-96.17	-1.37
11	con1	-420.98397	7.92	4.18	1.24	54.86	0.680
12	con38	-420.98385	7.97	4.1	1.21	-56.12	-0.681
13	con262	-420.98365	8.75	2.97	0.88	-50.2	-0.44
14	con5406	-420.98364	8.79	2.94	0.87	-96.17	-0.84
15	con123	-420.98363	8.8	2.94	0.87	53.57	0.466
16	con405	-420.98363	8.81	2.92	0.87	-41.8	-0.362
17	con401	-420.98363	8.82	2.9	0.86	-21.4	-0.18
18	con5408	-420.98362	8.82	2.89	0.86	-5.54	-0.05
19	con463	-420.98361	8.86	2.86	0.85	46.45	0.394
20	con110	-420.98360	8.89	2.83	0.84	-26.11	-0.219
21	con39	-420.98357	8.98	2.73	0.81	107.53	0.871
22	con796	-420.98289	10.75	1.34	0.4	436.88	1.736
23	con1213	-420.98287	10.8	1.31	0.39	-435.26	-1.69
24	con1220	-420.98258	11.56	0.97	0.29	102.88	0.29
25	con1649	-420.98201	13.06	0.53	0.16	182.17	0.29
26	con1953	-420.98200	13.08	0.53	0.16	184.05	0.288
27	con23	-420.98166	13.99	0.38	0.11	-206.47	-0.22
28	con1299	-420.98113	15.38	0.21	0.08	-225.22	-0.14
29	con1308	-420.98111	15.42	0.21	0.08	225.17	0.14
30	con1899	-420.97990	18.6	0.06	0.02	136.34	0.02
				sum			sum
				337.51			2.98

**Table 3.3.** The lowest-energy conformers from the MC run with RNG seed of 171982 combined with the lowest-energy conformers from the MC run with RNG seed of 111234 for 3-hydroxypentadiol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Figure 3.3 shows how the average  $[\alpha]_D$  value converges as more conformers are included in the calculation of the average. This graph was constructed from the results shown in Tables 3.1-3.3 and summarized in Table 3.4. The statistical weight of each of the average  $[\alpha]_D$  values is reflected in the Boltzmann normalization factor which is determined by the sum:

$$\sum_i e^{\Delta E_i/kT}$$

Result Reference	Statistical Weight	Average $[\alpha]_D$
Table 3.1a	160.65	226.31
Table 3.1d	210.88	139.91
Table 3.2a	272.56	21.25
Table 3.2c	275.58	23.76
Table 3.3	337.51	2.98

**Table 3.4.** Summary of trends found in the average  $[\alpha]_D$  values listed in table 3.1 to 3.3. The statistical weight or the Boltzmann normalization factor which is denoted by “sum” in Tables 3.1-3.3 and the specific rotation values are listed here. The statistical weight allows one to directly compare the results from individual runs prior to the normalization of the average  $[\alpha]_D$  values.

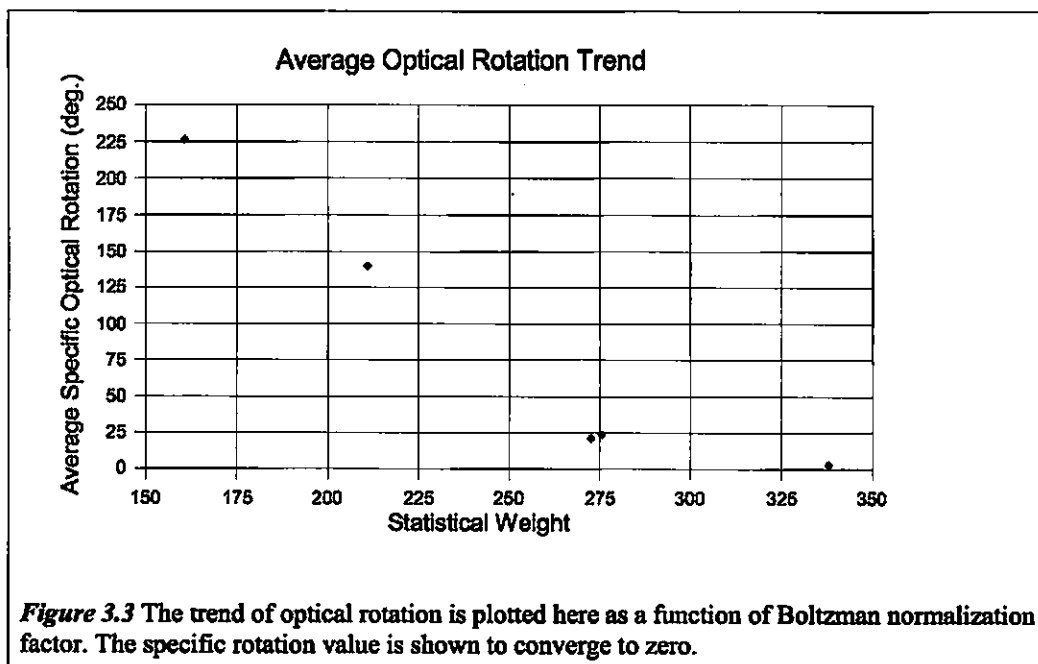
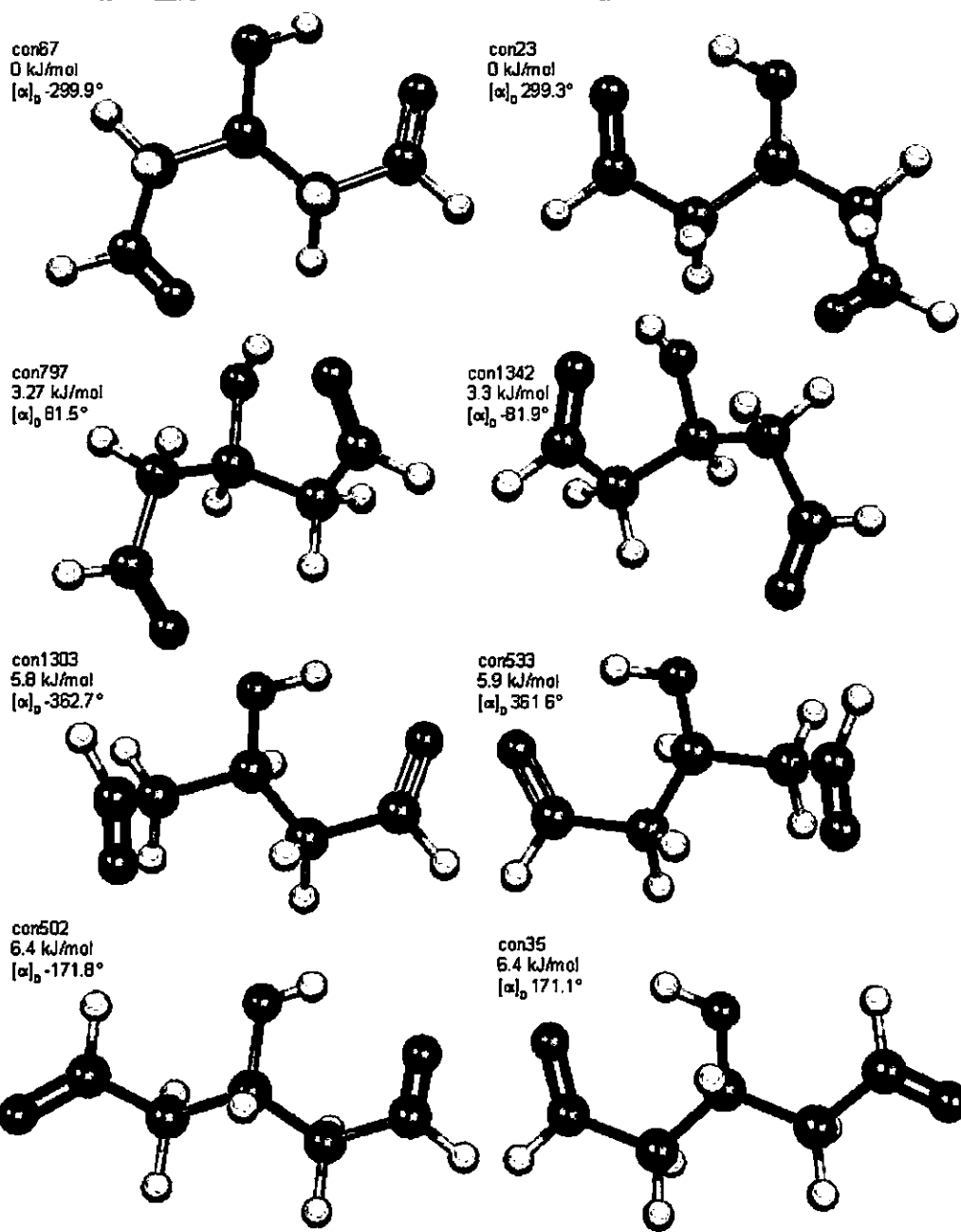


Table 3.4 and Figure 3.3 show that the predicted optical rotation value can be approximated using the Monte Carlo averaging approach as described in the previous section. The average  $[\alpha]_D$  value that carries the most significant statistical weight is also the value that is closest to the expected value of zero for an achiral molecule. It was noted that the low-energy conformers shown in Table 3.3 have a non-zero  $[\alpha]_D$  value individually. However, because this is not a chiral molecule and the mirror image geometries of all conformations with non-zero  $[\alpha]_D$  values are present in solution, the average  $[\alpha]_D$  was expected to be zero. The eight lowest-energy conformers from Table 3.3 are shown in Figure 3.4; note that they are actually four pairs of mirror image geometries which are close in energy and have opposite sign for  $[\alpha]_D$  values. The  $[\alpha]_D$  values for the individual conformers are quite large in magnitude so if one of them was not found with the MC program, the average  $[\alpha]_D$  value would be wrong.



**Figure 3.4** The eight lowest-energy conformers of 3-hydroxypentadiol from Table 3.3 are shown here. Note that they are pairs of mirror image geometries with non-zero specific rotation values of same magnitude and opposite sign.

From the two MC runs performed, only a few higher-energy conformers were found to be missing their mirror images, but all of the major contributors to the average  $[\alpha]_D$  value were accounted for. More runs were done with changes in temperature, RNG, and starting geometry without finding any other conformers that would significantly change the current average of  $3^\circ$ . In theory, it is possible to keep doing runs until that average is zero but that was not the goal of this test. The main goal was to determine a strategy for dealing with chiral molecules similar to the one just analyzed and making sure that the appropriate low-energy conformers are found for a case in which there are no mirror image conformers present in the sample. Based on the data collected for this achiral test molecule, one can say that for a molecule containing 30 – 40 low energy conformers, it is possible to converge to an  $[\alpha]_D$  value with precision of  $\sim 3$  degrees starting from 200 low-energy conformers found in the MC routine at the AM1 level of theory. The optimal parameter to vary is the RNG seed, followed by varying the starting geometry. Varying the temperature of the simulation does not have a great effect on finding new conformers.

### 3.4 Basis set selection for $[\alpha]_D$ calculations

There are several basis sets to choose from for calculating the specific rotation. According to Wiberg et al.,<sup>20</sup> the aug-cc-pVDZ is the basis set that gives the best agreement with experimental values due to the diffuse p-orbital on the hydrogen atoms. According to Stephens et al.,<sup>8</sup> the B3LYP functional is the best way to do these calculations and the DFT method as opposed to the HF method is generally yielding the result closest to the experimental value. In this work, the dependence of  $[\alpha]_D$  value on the inclusion of diffuse p-orbital functions on the hydrogen atoms was demonstrated using a hexafluorine analog of (2*S*)-2-methyl oxirane in Section 2.3.

Calculating the  $[\alpha]_D$  value at the B3LYP/aug-cc-pVDZ level of theory proved to be difficult using 16 processors on an IBM SP/RS6000 P3 node at the Maui High Performance Computing Center (MHPCC). Since more computing power is currently unavailable, an effort has been made to reduce the number of basis functions for this calculation by trying smaller basis sets for the calculation of the  $[\alpha]_D$  value.

Table 3.5 is a list of results comparing the  $[\alpha]_D$  values calculated using the HF method with different basis sets for (2*S*)-2-methyl oxirane. The experimental value of  $[\alpha]_D$  of 2-methyl oxirane (neat) is reported as  $\pm 14^\circ$  (Aldrich), depending on the absolute configuration. Since this molecule is rigid, it only has one low-energy conformer. The geometry optimization for the low-energy conformer was done using HF/6-31G\*.

<i>Basis Set</i>	<i>Number of Basis Functions</i>	<i>HF <math>[\alpha]_D</math></i>
3-21G	48	-33.61
6-31G	48	-35.77
<b>6-31G*</b>	<b>68</b>	<b>-12.27</b>
6-31+G*	84	-9.39
6-311G*	90	-40.53
6-311+G*	106	-7.11
6-311++G**	130	-19.39
<b>6-311++G(2d,2p)</b>	<b>168</b>	<b>-12.54</b>
cc-pVDZ	86	-31.07
<b>aug-cc-pVDZ</b>	<b>146</b>	<b>-13.82</b>

**Table 3.5.** HF  $[\alpha]_D$  values using different basis sets for (2*S*)-2-methyl oxirane. The most accurate results are obtained using the largest basis sets (6-311++G(2d,2p) and aug-cc-pVDZ). The medium-sized 6-31G\* basis set also gives a reasonable result.

This table shows that even at the HF level of theory there are three basis sets which produce results that are very close to the experimental value of the  $[\alpha]_D$  of (2*S*)-2-methyl oxirane. The best result ( $-13.82^\circ$ ) is obtained at the HF/aug-cc-pVDZ level of theory, closely followed by

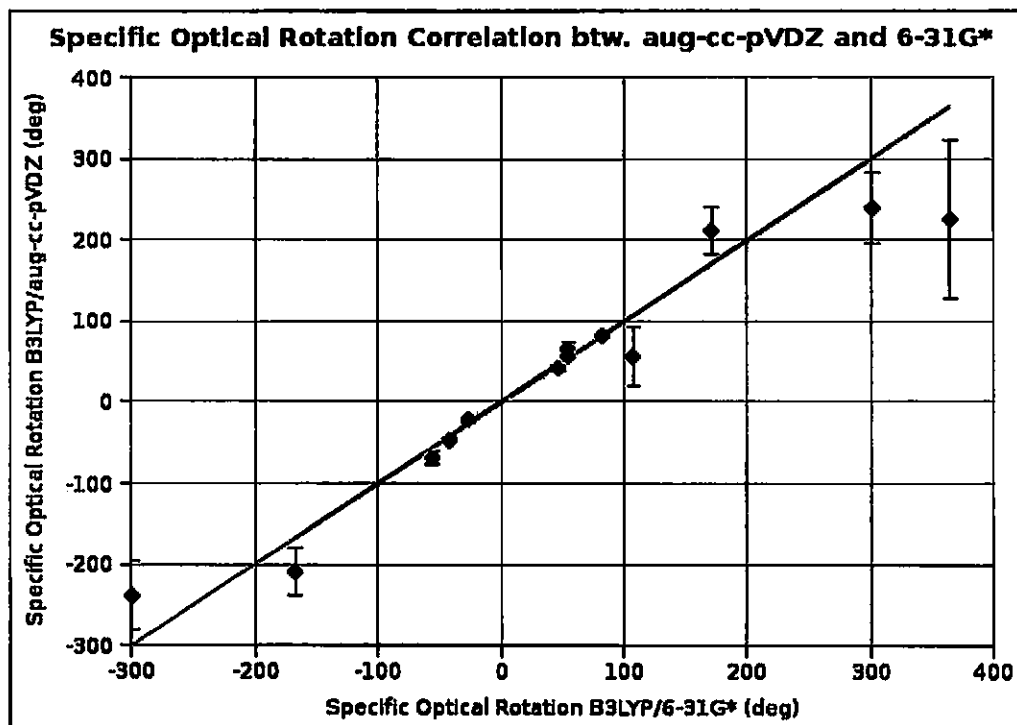
the result of  $-12.54^\circ$  at the HF/6-311++G(2d,2p) level of theory. The HF/6-31G\* level of theory also gives a good result of  $-12.27^\circ$ . The Figures 2.3b and 2.3c in Section 2.2 which show the variation of specific rotation values with HCCO dihedral angle of (2*S*)-2-methyl oxirane show a similar trend that around the equilibrium geometry the calculated  $[\alpha]_D$  value using aug-cc-pVDZ and 6-31G\* basis sets will be in the same range.

A direct comparison was made between  $[\alpha]_D$  values from 6-31G\* and the aug-cc-pVDZ basis sets using the MC 1-30 set (seed=171982) run for 3-hydroxypentadiol. The results from this run were shown in Table 3.2a in Section 3.3. In that run all geometry optimizations were carried out using B3LYP/6-31G\* method and the  $[\alpha]_D$  value was calculated using the B3LYP/6-31G\* and B3LYP/aug-cc-pVDZ.

	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31G*	Abs Error	Standard Dev.	Rel Error (%)
1	con67	-299.94	-238.95	60.99	7.81	-20.33
2	con23	300.69	239.18	61.51	7.84	-20.46
3	con144	82.71	82.4	0.31	0.56	-0.37
4	con324	364.25	226.21	138.04	11.75	-37.9
5	con35	171.07	211.91	40.84	6.39	23.87
6	con298	-167.76	-209.29	41.53	6.44	24.76
7	con1	54.86	66.2	11.34	3.37	20.67
8	con38	-56.12	-67.53	11.41	3.38	20.33
9	con123	53.57	55.51	1.94	1.39	3.62
10	con405	-41.8	-46.23	4.43	2.1	10.6
11	con463	46.45	42.2	4.25	2.06	-9.15
12	con110	-26.11	-22.3	3.81	1.95	-14.59
13	con39	107.53	56.1	51.43	7.17	-47.83

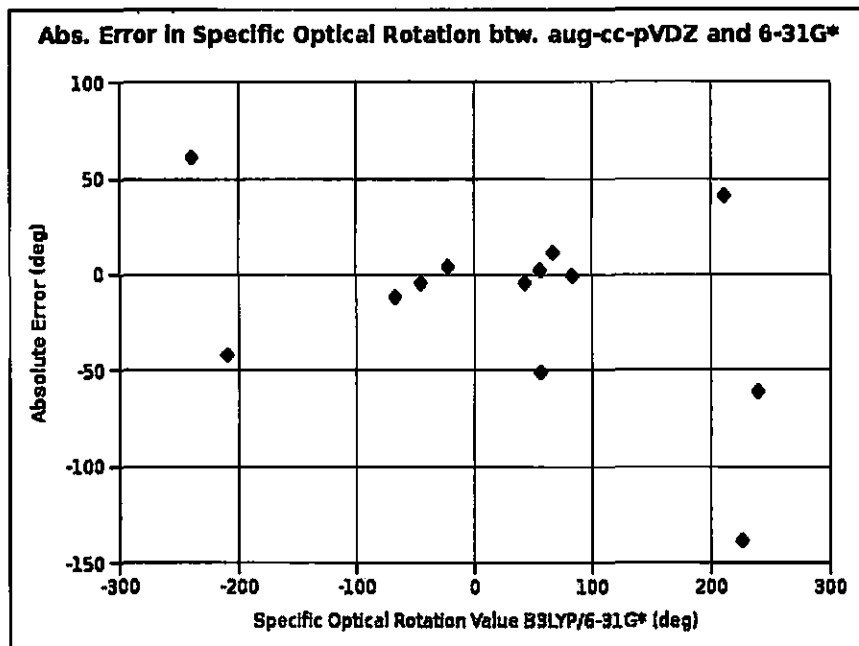
**Table 3.6.** The agreement between specific rotation values for 3-hydroxypentadiol calculated at the B3LYP/6-31G\* level of theory and at the B3LYP/aug-cc-pVDZ level of theory is reasonably good for  $[\alpha]_D$  values of small magnitudes. For a large  $[\alpha]_D$  such as for con324, the agreement is poor. The relative errors are all roughly of the same magnitude.



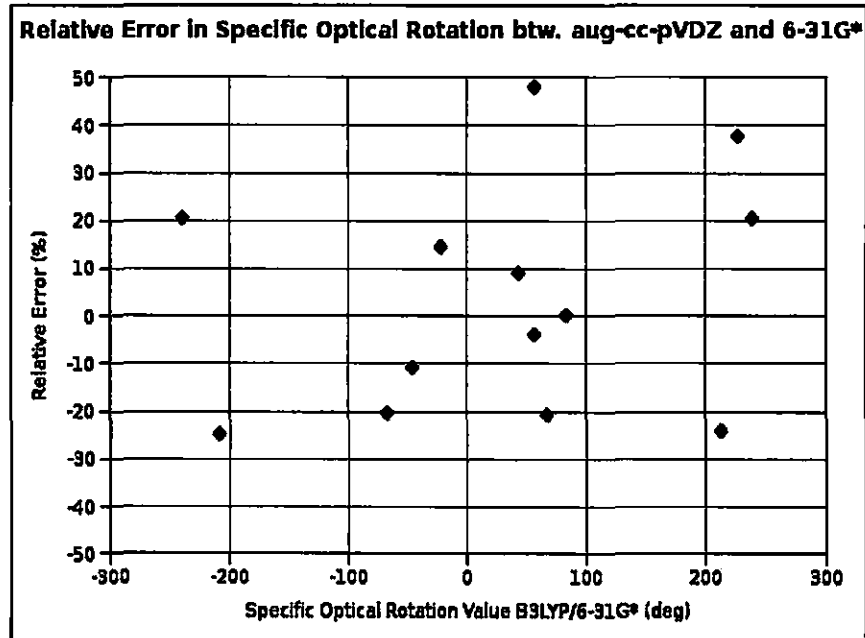


**Figure 3.5.** Correlation of specific rotation values calculated using aug-cc-pVDZ and 6-31G\* basis sets for 3-hydroxypentadiol. The  $R^2$  value is 0.97. The error bars were derived from a two-factor variance statistical analysis with a 95% confidence. The correlation is much better for  $[\alpha]_D$  values close to zero then for values of large magnitude.

Table 3.6 and Figure 3.5 shows that the agreement between the two basis sets (6-31G\* and aug-cc-pVDZ) is excellent for smaller to medium magnitude  $[\alpha]_D$  values. When the  $[\alpha]_D$  values get larger, the agreement worsens. This is especially problematic for a molecule such as 3-hydroxypentadiol where several of the lowest energy conformers have a high optical rotation value. This means that the value with the highest error will carry the greatest statistical weight in the average value. Figures 3.6(a,b) show that there is no simple empirical correction that can be made on the 6-31G\* basis that would shift the erroneous  $[\alpha]_D$  values to match the values from aug-cc-pVDZ basis set as the error appears to be random in both magnitude and sign.



*Figure 3.6a.* Absolute error between the  $[\alpha]_D$  values of 3-hydroxypentadiol calculated using aug-cc-pVDZ and 6-31G\* basis sets. The absolute error appears to be random in magnitude and sign. This lack of trend makes it impossible to determine a simple correction factor.



*Figure 3.6b.* Relative error between the  $[\alpha]_D$  values of 3-hydroxypentadiol calculated using aug-cc-pVDZ and 6-31G\* basis sets. The relative error also appears to be random in magnitude and sign though it varies less than the absolute error.

Table 3.7 shows that the error is significantly decreased if the  $[\alpha]_D$  is calculated using the B3LYP/6-31++G(d,p) basis set for the problematic conformers. However, this method utilizes a large number of basis functions (591) for the fatty acid target molecule making computer time the prohibitive factor in the  $[\alpha]_D$  calculation.

	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31G*	Abs Error	Rel Error (%)
1	con67	-299.94	-238.95	60.99	-20.33
2	con23	300.69	239.18	61.51	-20.46
3	con144	82.71	82.4	0.31	-0.37
4	con324	364.25	226.21	138.04	-37.9
	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31++G(d,p)	Abs Error	Rel Error (%)
1	con67	-299.94	-307	7.06	-2.35
2	con23	300.69	308.14	7.45	-2.48
3	con144	82.71			
4	con324	364.25	351.96	12.29	3.37

**Table 3.7.** Correlation of specific rotation values calculated using aug-cc-pVDZ, 6-31G\* and 6-31++G(d,p) basis sets for 3-hydroxypentadiol is shown for the conformers with the greatest error in  $[\alpha]_D$  values between the aug-cc-pVDZ and 6-31G\* basis sets. The 6-31++G(d,p) basis set is an improvement in the agreement between specific rotation values calculated at the B3LYP/aug-cc-pVDZ level of theory from the values calculated at the B3LYP/6-31G\* level of theory. The absolute error in  $[\alpha]_D$  values is decreased from 60° to 7° and from 138° to 12°. The relative error is reduced by a factor of 10 using the 6-31++G(d,p) basis set.

Even though results using the 6-31++G(d,p) basis set proved to be in good agreement with the results using the aug-cc-pVDZ basis set, there were still too many basis functions for the calculation to have a reasonable running time of less than 10 hours. Therefore, a smaller basis set was tried next. Using the 6-31++G\* basis set, the  $[\alpha]_D$  calculation on 16 processors on an IBM SP/RS6000 P3 node at MHPCC for the fatty acid took 6 hours which is a viable option. The calculation time for 3-hydroxypentadiol was reduced from 40 minutes using the aug-cc-pVDZ basis set to 12 minutes using the 6-31++G\* basis set. The correlation between the calculations at the B3LYP/6-31++G\* level of theory and the calculations at the B3LYP/aug-cc-pVDZ level of theory is shown on Tables 3.8(a,b) and Figure 3.7.

**Table 3.8a**

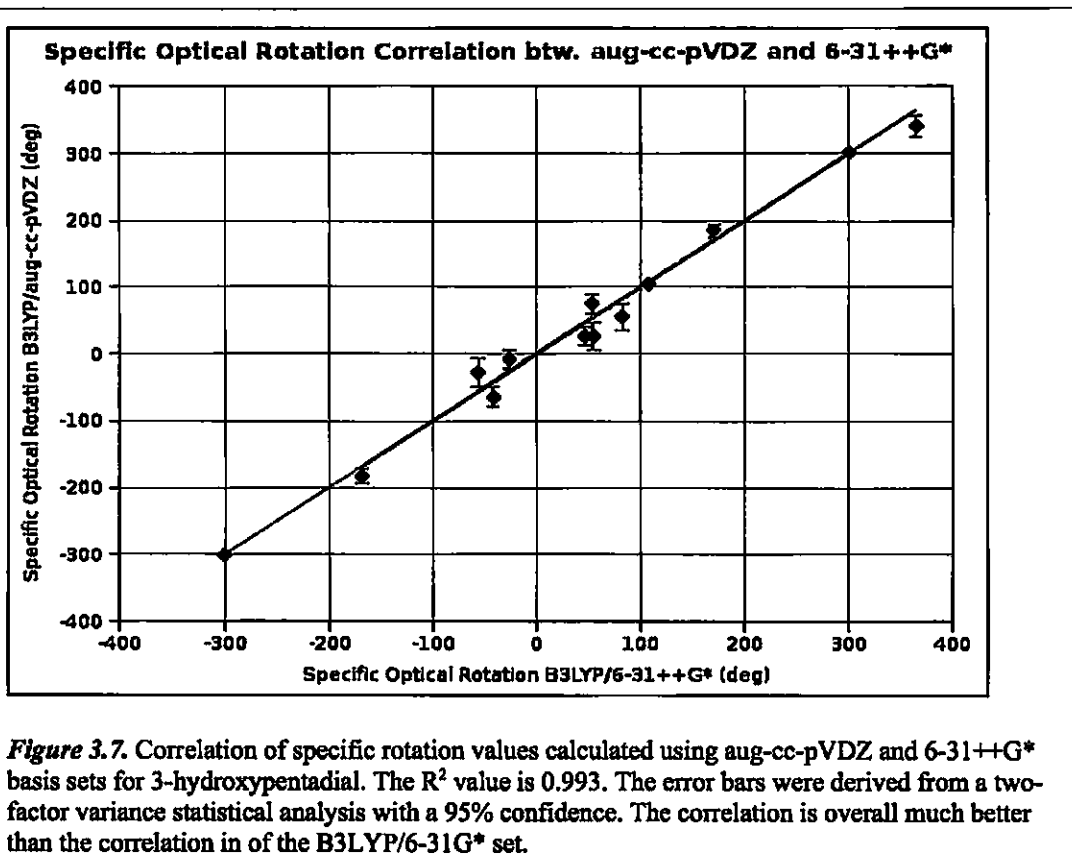
	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31++G*	Abs Error	Standard Dev.	Rel Error (%)
1	con67	-299.94	-300.2	0.26	0.18	0.09
2	con23	300.69	301.41	0.72	0.51	0.24
3	con144	82.71	54.93	27.78	19.64	-33.59
4	con324	364.25	341	23.25	16.44	-6.38
5	con35	171.07	184.99	13.92	9.84	8.14
6	con298	-167.76	-182.97	15.21	10.76	9.07
7	con1	54.86	26.12	28.74	20.32	-52.39
8	con38	-56.12	-27.15	28.97	20.48	-51.62
9	con123	53.57	74.66	21.09	14.91	39.37
10	con405	-41.8	-63.39	21.59	15.27	51.65
11	con463	46.45	27.2	19.25	13.61	-41.44
12	con110	-26.11	-6.99	19.12	13.52	-73.23
13	con39	107.53	105.82	1.71	1.21	-1.59

**Table 3.8a.** The agreement between specific rotation values for 3-hydroxypentadiol calculated at the B3LYP/6-31++G\* level of theory and at the B3LYP/aug-cc-pVDZ level of theory is an improvement over the B3LYP/6-31G\* level of theory. Notably, for a large  $[\alpha]_D$  values such as for con324 where the agreement using the 6-31G\* basis set was poor, the absolute error has been reduced from 138° to 23°.

**Table 3.8b**

	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31G*	Abs Error	Rel Error (%)
1	con67	-299.94	-238.95	60.99	-20.33
2	con23	300.69	239.18	61.51	-20.46
3	con144	82.71	82.4	0.31	-0.37
4	con324	364.25	226.21	138.04	-37.9
	1-30 set	B3LYP/aug-cc-pVDZ	B3LYP/6-31++G*	Abs Error	Rel Error (%)
1	con67	-299.94	-300.2	0.26	0.09
2	con23	300.69	301.41	0.72	0.24
3	con144	82.71			
4	con324	364.25	341	23.25	-6.38

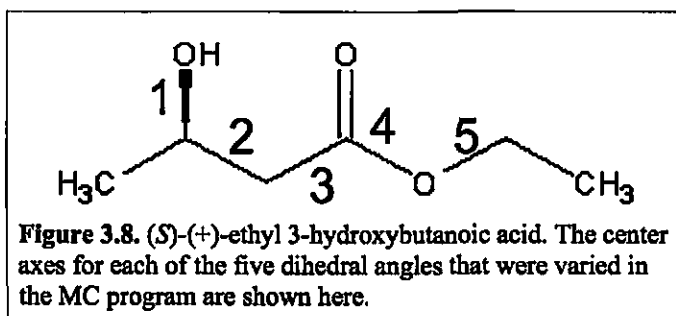
**Table 3.8b.** Correlation of specific rotation values calculated using aug-cc-pVDZ, 6-31G\* and 6-31++G\* basis sets for 3-hydroxypentadiol is shown for the conformers with the greatest error in  $[\alpha]_D$  values between the aug-cc-pVDZ and 6-31G\* basis sets. The 6-31++G\* basis set is an improvement in the agreement between specific rotation values calculated at the B3LYP/aug-cc-pVDZ level of theory from the values calculated at the B3LYP/6-31G\* level of theory. The absolute error in  $[\alpha]_D$  values is decreased from 60° to 0° and from 138° to 23°. The relative errors are also an improvement using the 6-31++G\* basis set.



Based on the correlation results between B3LYP/aug-cc-pVDZ and B3LYP/6-31++G\*, it is proposed that all future  $[\alpha]_D$  calculations will be performed using the B3LYP/6-31++G\* basis set and method in the interest of time and without sacrificing a great deal in terms of accuracy. It does appear that the diffuse p-orbital on the hydrogen atom from the aug-cc-pVDZ basis set does not play as key a role as was thought initially. In fact, given the rather good correlation between B3LYP/aug-cc-pVDZ and B3LYP/6-31G\*, it would appear that the diffuse functions only play a significant role for  $[\alpha]_D$  values of a large magnitude. It was clear from the discussion on Fluorine-substituted oxiranes in Section 2.3 that having a diffuse p-orbital on the hydrogen plays a significant role in getting consistent  $[\alpha]_D$  values as a function of FCCO dihedral angle. The basis sets that yield the most reliable  $[\alpha]_D$  values around the equilibrium geometry are not as reliant on the hydrogen diffuse orbitals as we previously thought.

### 3.5 Testing the predictive power of the MC based $[\alpha]_D$ value calculations

**Introduction.** We have established in Section 3.3 that the Monte Carlo routine is adequate in finding the low-energy conformations of a molecule. The next step is to determine how accurate is the theoretical specific rotation value derived from this approach. This test will be discussed in the present and in the next section using two chiral molecules similar in flexibility and functionality to the target molecule (fatty acid). The molecule used in this section is (*S*)-(+)-ethyl 3-hydroxybutanoic acid shown in Figure 3.8.



The reported  $[\alpha]_D$  value for (*S*)-(+)-ethyl 3-hydroxybutanoic acid in chloroform at 20°C is 43° (Aldrich). This molecule has 5 dihedral angles of interest which were sampled in the MC routine; the methyl groups were not included because they are always minimized. Based on the results obtained using the achiral molecule with five dihedral angles, at least two runs using different starting RNG seeds are needed to get close to the experimental result. The difference between the present calculations and the calculations done using the achiral molecule in Section 3.3, is to see how closely we can expect the theoretical  $[\alpha]_D$  value to agree with the experimental value.

Factors that can contribute to the differences between the theoretical and experimental  $[\alpha]_D$  value, which were not a factor in the first test, include: the fact that the calculated values simulate gas phase whereas the experiment is done in solution and contributions from vibrational effects to the experimental value which are not accounted for in the calculation. In essence, as previously

discussed in the literature and in Section 1.4, errors of  $\pm 50^\circ$  between the calculated and experimental  $[\alpha]_D$  values can be expected due to these factors.

**Method.** Six MC runs were completed at 300 K using different RNG seeds and initial (S)-(+)-ethyl 3-hydroxybutanoic acid starting geometries. The resulting AM1 structures from each of the runs were sorted and then their geometry was optimized at the B3LYP/6-31G\* level of theory. It was noted that while the relative energies of the different conformers in each set was not exactly preserved between the AM1 level of theory and the B3LYP/6-31G\* level of theory, the conformers that were found to have low energies with AM1 were also the ones with low energies using B3LYP/6-31G\*. Thus, rather than re-optimize all of the AM1 conformers that were located by the MC runs, a more efficient, yet still reliable approach was used where only the first 30 – 40 AM1 conformers from each run were re-optimized at a higher level of theory.

**Results.** The first run was done using RNG seed of 39679 until 60 new fully optimized conformers were found at the AM1 level of theory. In Table 3.9 out of 60 AM1 optimized conformers from this MC run, 35 conformers optimized to unique geometries at the B3LYP/6-31G\* level, with only four conformers within the 11 kJ/mol range from the lowest-energy conformer. The relative energy cutoff was moved to 11 kJ/mol because conformers of higher energy have an abundance of less than 1% at 298 K and do not contribute significantly to the average  $[\alpha]_D$  value. The calculated average  $[\alpha]_D$  value for this run was  $81^\circ$ . The run was then extended to find up to 85 new conformers at the AM1 level of theory which were re-optimized at the B3LYP/6-31G\* level of theory. Although new high-energy conformers were found (the results of the extended run are not shown because they are redundant), no new low-energy conformers were found. These four low-energy conformers all come from the 30 lowest energy conformers at the AM1 level of theory. Thus, one can still arrive at the same result with half as many optimizations at a high level of theory as shown in Table 3.9.

	1-60 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con110	-461.55069	0	100	69.01	108.9	75.15
2	con195	-461.54940	3.37	25.05	17.29	-39.3	-6.79
3	con272	-461.54912	4.12	18.4	12.7	93.11	11.83
4	con59	-461.54677	10.3	1.46	1.01	47.58	0.48
5	con249	-461.54565	13.21	0.44	0.30		
6	con92	-461.54420	17.04	0.09	0.06		
7	con305	-461.54415	17.16	0.09	0.06		
8	con962	-461.54413	17.21	0.09	0.06		
9	con465	-461.54396	17.65	0.07	0.05		
10	con436	-461.54391	17.79	0.07	0.05		
11	con481	-461.54371	18.31	0.05	0.04		
12	con498	-461.54306	20.02	0.03	0.02		
13	con1286	-461.54301	20.15	0.03	0.02		
14	con443	-461.54298	20.23	0.02	0.02		
15	con439	-461.54298	20.24	0.02	0.02		
16	con16	-461.54295	20.30	0.02	0.02		
17	con254	-461.54057	26.56	0.00	0.00		
18	con47	-461.54046	26.86	0.00	0.00		
19	con259	-461.54031	27.25	0.00	0.00		
20	con64	-461.53904	30.57	0.00	0.00		
21	con130	-461.53752	34.58	0.00	0.00		
22	con178	-461.53751	34.59	0.00	0.00		
23	con139	-461.53625	37.90	0.00	0.00		
24	con707	-461.53579	39.10	0.00	0.00		
25	con695	-461.53259	47.51	0.00	0.00		
26	con374	-461.53234	48.16	0.00	0.00		
27	con250	-461.53209	48.82	0.00	0.00		
28	con137	-461.53141	50.61	0.00	0.00		
29	con529	-461.53098	51.73	0.00	0.00		
30	con392	-461.53081	52.17	0.00	0.00		
31	con292	-461.53024	53.67	0.00	0.00		
32	con279	-461.52963	55.29	0.00	0.00		
33	con12	-461.52944	55.78	0.00	0.00		
34	con86	-461.52935	56.01	0.00	0.00		
35	con447	-461.52919	56.45	0.00	0.00		
				sum		sum	
				144.91		80.86	

**Table 3.9.** The low-energy conformers from the MC run with RNG seed of 39679 for (*S*)-(+)-ethyl 3-hydroxybutanoic acid. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. The  $[\alpha]_D$  value was calculated only for the first four conformers. The higher-energy conformers (5-35) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

The second run was done using RNG seed of 316251 until 150 new conformers optimized at the AM1 level of theory were isolated and sorted by energy. The 30 lowest-energy conformers were separated for further analysis. In Table 3.10 one can see that out of 30 AM1



optimized conformers from this MC run, 22 conformers optimized to unique geometries at the B3LYP/6-31G\* level, with only six conformers within the 11 kJ/mol range from the lowest-energy conformer. The relative energy cutoff was moved to 11 kJ/mol because conformers of higher energy have an abundance of less than 1% at 298 K and do not contribute significantly to the average  $[\alpha]_D$  value. The calculated average  $[\alpha]_D$  value for this run was 91°.

	1-30 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con5	-461.55041	0	100	53.35	188.05	100.33
2	con19	-461.54941	2.63	34.01	18.14	-41.05	-7.45
3	con50	-461.54913	3.36	25.16	13.42	93.11	12.5
4	con501	-461.54911	3.4	24.73	13.19	-119.82	-15.81
5	con921	-461.54878	9.53	2	1.07	47.58	0.51
6	con597	-461.54854	10.17	1.54	0.82	123.97	1.02
7	con165	-461.54592	11.78	0.79	0.42		
8	con1296	-461.54573	12.28	0.65	0.34		
9	con31	-461.54567	12.44	0.61	0.32		
10	con1681	-461.54419	16.34	0.12	0.07		
11	con368	-461.54416	16.40	0.12	0.06		
12	con422	-461.54396	16.93	0.10	0.05		
13	con537	-461.54396	16.94	0.10	0.05		
14	con1303	-461.54391	17.07	0.09	0.05		
15	con13	-461.54390	17.11	0.09	0.05		
16	con567	-461.54371	17.60	0.07	0.04		
17	con560	-461.54327	18.76	0.05	0.02		
18	con336	-461.54325	18.79	0.04	0.02		
19	con563	-461.54306	19.29	0.04	0.02		
20	con515	-461.54300	19.47	0.03	0.02		
21	con135	-461.53905	29.83	0.00	0.00		
22	con455	-461.52921	55.66	0.00	0.00		
				sum			sum
				187.44			91.1

**Table 3.10.** The low-energy conformers from the MC run with RNG seed of 3176251 for (S)-(+)-ethyl 3-hydroxybutanoic acid. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. The  $[\alpha]_D$  value was calculated only for the first six conformers. The higher-energy conformers (7-22) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

The third run was done using RNG seed of 8476125 until 200 new conformers optimized at the AM1 level of theory were isolated and sorted by energy. The 50 lowest-energy conformers were separated for further analysis. In Table 3.11 one can see that out of 50 AM1 optimized conformers from this MC run, 27 conformers optimized to unique geometries at the B3LYP/6-

31G\* level, with only six conformers within the 10 kJ/mol range from the lowest-energy conformer. The relative energy cutoff was moved to 10 kJ/mol because conformers of higher energy have an abundance of less than 1% and do not contribute significantly to the average  $[\alpha]_D$  value. The calculated average  $[\alpha]_D$  value for this run was 79°.

	1-50 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con27	-461.55069	0	100	32.23	108.9	35.1
2	con11	-461.55042	0.71	74.81	24.11	188.05	45.34
3	con124	-461.55041	0.74	73.86	23.81	14.16	3.37
4	con34	-461.54939	3.4	24.78	7.99	-39.3	-3.14
5	con80	-461.54913	4.09	18.64	6.01	93.11	5.59
6	con33	-461.54911	4.15	18.18	5.86	-119.82	-7.02
7	con114	-461.54660	10.75	1.21	0.39		
8	con86	-461.54654	10.89	1.14	0.37		
9	con7	-461.54573	13.03	0.48	0.15		
10	con77	-461.54565	13.24	0.44	0.14		
11	con14	-461.54449	16.26	0.13	0.04		
12	con514	-461.54421	17.02	0.09	0.03		
13	con4625	-461.54414	17.19	0.09	0.03		
14	con85	-461.54412	17.25	0.08	0.03		
15	con167	-461.54397	17.63	0.07	0.02		
16	con4499	-461.54396	17.68	0.07	0.02		
17	con10	-461.54395	17.68	0.07	0.02		
18	con175	-461.54392	17.76	0.07	0.02		
19	con123	-461.54391	17.80	0.07	0.02		
20	con1414	-461.54391	17.81	0.07	0.02		
21	con6541	-461.54370	18.35	0.05	0.02		
22	con49	-461.54368	18.40	0.05	0.02		
23	con70	-461.54328	19.45	0.03	0.01		
24	con40	-461.54307	20.01	0.03	0.01		
25	con24998	-461.54307	20.01	0.03	0.01		
26	con4	-461.54299	20.21	0.02	0.01		
27	con1417	-461.54297	20.27	0.02	0.01		
				sum		sum	
				310.28		79.24	

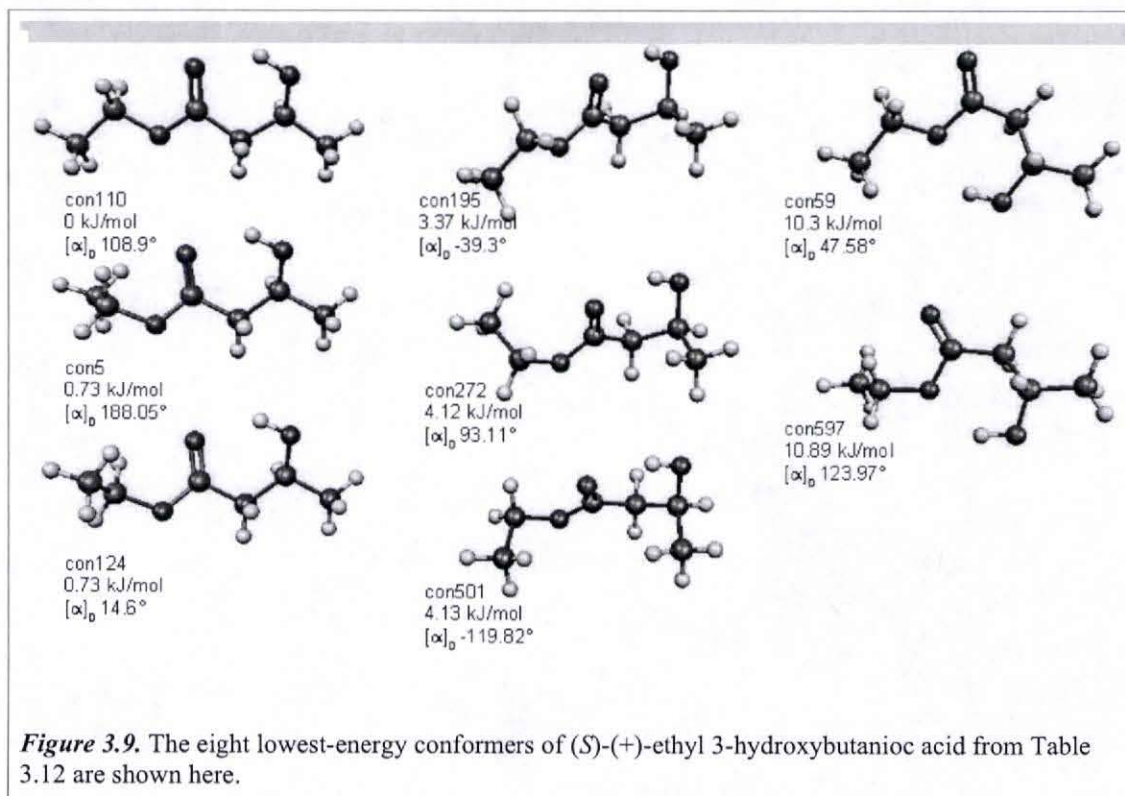
**Table 3.11.** The low-energy conformers from the MC run with RNG seed of 8476125 for (S)-(+)-ethyl 3-hydroxybutanoic acid. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. The  $[\alpha]_D$  value was calculated only for the first six conformers. The higher-energy conformers (7-27) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Three more runs were performed using other RNG seeds and starting geometries, but no additional low-energy conformers were found at the B3LYP/6-31G\* level of theory. Thus we conclude that the conformational search has converged. The theoretical value for  $[\alpha]_D$  was

determined to be 79° after the results from all of the runs listed in Tables 3.9 – 3.11 were combined and any repeating conformers deleted as shown in Table 3.12 and the 3D pictures of the eight conformers listed are shown in Figure 3.9.

	1-140 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con110	-461.55069	0	100	31.98	108.9	34.83
2	con5	-461.55041	0.73	74.25	23.75	188.05	44.66
3	con124	-461.55041	0.73	74.01	23.67	14.16	3.35
4	con195	-461.54940	3.37	25.05	8.01	-39.3	-3.15
5	con272	-461.54912	4.12	18.4	5.89	93.11	5.48
6	con501	-461.54911	4.13	18.36	5.87	-119.82	-7.04
7	con59	-461.54677	10.3	1.46	0.47	47.58	0.22
8	con597	-461.54654	10.89	1.14	0.37	123.97	0.45
	sum			312.69			78.81

**Table 3.12.** The lowest-energy conformers of (*S*)-(+)-ethyl 3-hydroxybutanoic acid from the MC run with RNG seed of 39679 combined with the lowest-energy conformers from the MC run with RNG seed of 3176251 and the lowest-energy conformers from the MC run with RNG seed of 8476125. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/aug-cc-pVDZ level of theory. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.



The number of low-energy conformers (eight conformers with an abundance greater than 1%) found with the MC simulation was different for (*S*)-(+)-ethyl 3-hydroxybutanoic acid than the number of low-energy conformers (21 conformers in abundance of less than 1%) found for 3-hydroxypentadiol in Section 3.2. For (*S*)-(+)-ethyl 3-hydroxybutanoic acid the average  $[\alpha]_D$  value for different runs did not vary as much as it did for 3-hydroxypentadiol.

One reason for the smaller number of low-energy conformers of (*S*)-(+)-ethyl 3-hydroxybutanoic acid is that there is only one stabilizing hydrogen bond between the hydroxy group and the carbonyl group possible, whereas there are two carbonyl groups on 3-hydroxypentadiol which can provide hydrogen bonding for the hydroxy group. One can clearly see this by visually comparing Figures 3.9 and 3.4. In Figure 3.9, the two higher-energy conformers (con59 and con597) are stabilized by hydrogen bonding between the hydroxy group and the  $sp^3$  oxygen of the ester, which is not as stabilizing as the hydrogen bond to the carbonyl oxygen. Thus, even though (*S*)-(+)-ethyl 3-hydroxybutanoic acid should have approximately the same number of conformers as the achiral molecule since they are both similar in size, the bulk of these conformers lies above the 10 kJ/mol energy range where the population contribution from individual conformers is more than 1%.

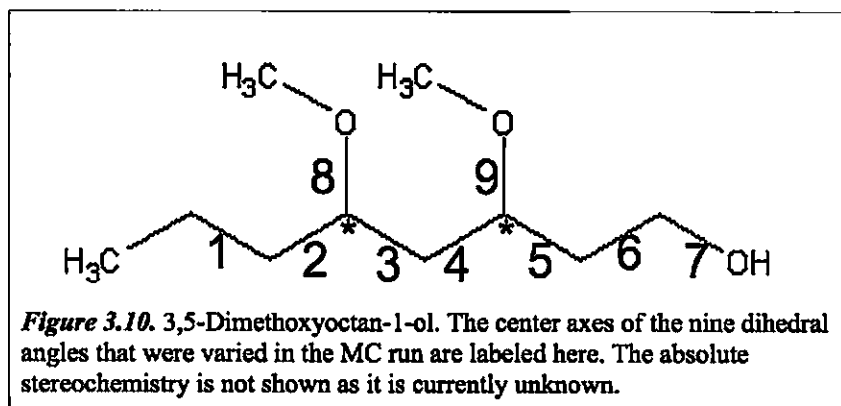
Given the relatively small number of possible low-energy conformers that one needs to find in each run, most of the runs had no difficulty finding them all (the only exception being the first run which was of the shorter than the other one). In addition to the small number of low-energy conformers available for the (*S*)-(+)-ethyl 3-hydroxybutanoic acid, another reason for such a small variation in the average  $[\alpha]_D$  value between the different MC runs is that the three most significant conformers (con110, con5, and con124) all have a positive  $[\alpha]_D$  value and two of them have a large magnitude. At least one of these two conformers was always found (again, the exception being the first run) and thus the average  $[\alpha]_D$  value did not fluctuate as much as it did

in the case of 3-hydroxypentadiol.

The theoretical value for  $[\alpha]_D$  of  $79^\circ$  (Table 3.11) overestimates the experimental value of  $43^\circ$  by  $36^\circ$ . As discussed in Section 1.4 and 1.6, this  $36^\circ$  error is within the  $\pm 50^\circ$  errors typically reported in the literature.

### 3.6 Determination of absolute stereochemistry of 3,5-dimethoxyoctan-1-ol via specific optical rotation calculations

**Introduction.** Based on the results from the two test cases discussed in Sections 3.3 and 3.5, we have established a strategy to for calculating the average the  $[\alpha]_D$  value for a flexible molecule with an error of less than  $\pm 40^\circ$ . The most efficient way to use the MC routine to find the low-energy conformers has been determined. In this section, we use this procedure to find the absolute stereochemistry of 3,5-Dimethoxyoctan-1-ol shown in Figure 3.10. This molecule has two chiral centers and it is larger in size than the chiral molecule studied in Section 3.5.



NMR results have shown that the relative stereochemistry of the two chiral centers in this molecule is either (*R,R*) or (*S,S*) but the absolute configuration still needs to be determined by a specific optical rotation measurement.

**Method.** All of the computations were done using the (*R,R*) isomer to get the magnitude

of  $[\alpha]_D$ . Repeating the conformer search with an (*S,S*) isomer is unnecessary since it was shown in Section 3.2 that the mirror image structures will be found giving an  $[\alpha]_D$  value equal in magnitude and opposite in sign to the (*R,R*) isomer. Once the  $[\alpha]_D$  value is known for the (*R,R*) isomer, it can be compared to the measured value and thus the absolute configuration can be determined. In light of the results from comparing the basis sets for the specific rotation calculation in Section 3.3, as well as in the interest of saving computer time all of the  $[\alpha]_D$  values were calculated at the B3LYP/6-31++G\* level of theory.

There are nine dihedral angles which are indicated in Figure 3.10 that were varied in the MC runs. Seven MC runs were performed for this molecule while varying the starting geometry and the RNG seed. These runs lasted less than 36 hours, depending on how many conformers the run was set to find. After the initial four runs, no more new low-energy conformers were found.

**Results.** The first run was done using RNG seed of 549078 until 40 new conformers optimized at the AM1 level of theory were found and sorted by energy. Table 3.13 shows that out of 40 conformers optimized at the AM1 level of theory, 38 unique conformers were found at the B3LYP/6-31G\* level of theory, and the 17 lowest-energy conformers were within 11 kJ/mol energy range. The relative energy cutoff was taken as 11 kJ/mol because conformers of higher energy have an abundance of less than 1% at 298 K and do not contribute significantly to the average  $[\alpha]_D$  value. The specific rotation value was calculated using B3LYP/6-31++G\* basis set and method. The calculated average  $[\alpha]_D$  value for this run was  $-36^\circ$ .

Note that this was the initial run and therefore short (only 40 AM1-level conformers collected) and one can see from the Table 3.12 that there is virtually no repetition of geometries as is common with more extensive runs. The initial runs should always be short, done mostly for the benefit of debugging the code after modifications.

	1-40 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con105	-619.95206	0	100	28.7	-177.31	-50.89
2	con390	-619.95180	0.69	75.81	21.76	-48.09	-10.46
3	con1050	-619.95113	2.45	37.4	10.73	100.92	10.83
4	con231	-619.95085	3.16	28.13	8.07	38.81	3.13
5	con315	-619.95083	3.23	27.41	7.87	-30.81	-2.42
7	con1026	-619.95042	4.31	17.8	5.11	-57.93	-2.96
9	con845	-619.95033	4.54	16.19	4.65	217.88	10.13
10	con813	-619.95032	4.57	16.03	4.6	200.05	9.2
11	con403	-619.95025	4.76	14.82	4.25	-67.76	-2.88
12	con985	-619.94957	6.54	7.25	2.08	-3.43	-0.07
13	con584	-619.94919	7.55	4.86	1.39	48.52	0.68
16	con987	-619.94800	10.67	1.39	0.4	21.3	0.08
17	con532	-619.94798	10.72	1.36	0.39	27.2	0.11
18	con210	-619.94783	11.11	1.16	0.33		
19	con161	-619.94781	11.17	1.14	0.33		
20	con880	-619.94777	11.26	1.10	0.31		
21	con266	-619.94767	11.52	0.99	0.28		
22	con101	-619.94767	11.54	0.98	0.28		
23	con954	-619.94719	12.78	0.60	0.17		
24	con962	-619.94699	13.32	0.48	0.14		
25	con45	-619.94679	13.83	0.39	0.11		
26	con599	-619.94630	15.12	0.23	0.07		
27	con536	-619.94614	15.54	0.20	0.06		
28	con534	-619.94612	15.6	0.19	0.06		
29	con835	-619.94538	17.55	0.09	0.03		
30	con660	-619.94513	18.19	0.07	0.02		
31	con754	-619.94470	19.33	0.04	0.01		
32	con822	-619.94281	24.29	0.01	0.00		
33	con95	-619.94270	24.58	0.01	0.00		
34	con35	-619.94205	26.29	0.00	0.00		
35	con8	-619.94159	27.5	0.00	0.00		
36	con1	-619.94090	29.29	0.00	0.00		
37	con43	-619.93626	41.48	0.00	0.00		
38	con30	-619.93140	54.23	0.00	0.00		
				sum			sum
				348.45			-35.52

**Table 3.13.** The low-energy conformers from the MC run with RNG seed of 549078 for 3,5-Dimethoxyoctan-1-ol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/6-31++G\* level of theory. The  $[\alpha]_D$  value was calculated only for the first 17 conformers. The higher-energy conformers (18-38) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Table 3.14 shows results from the second run done using RNG seed of 1527699 and the temperature of the simulation was changed from 300 K to 400 K to help the MC conformer search get over some of the higher energy barriers that may be present on the PES. From the 100

conformers optimized at the AM1 level of theory 70 unique conformers were found at the B3LYP/6-31G\* level of theory, out of which 14 were within the 10 kJ/mol range. Higher-energy conformers were found to be present in less than 1% abundance in the sample at 298 K and were considered not to be significant contributors to the average  $[\alpha]_D$  value. The average  $[\alpha]_D$  value was found to be  $-50^\circ$ . It is now easy to appreciate the extent to which a flexible molecule spanning nine dihedral angles of interest is more challenging than the same type of molecule with five dihedral angles of interest. Fortunately, the number of low-energy conformers for which the specific rotation calculation needs to be done is manageable as the majority of the conformers are too high in energy to contribute significantly to the average  $[\alpha]_D$  value.



	1-100 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con2884	-619.95211	0	100	26.89	-177.31	-47.67
2	con2909	-619.95178	0.88	70.34	18.91	-48.09	-9.09
3	con2739	-619.95111	2.63	34.79	9.35	4.75	0.44
4	con2778	-619.95108	2.72	33.61	9.04	99.94	9.03
5	con3014	-619.95089	3.22	27.53	7.4	38.81	2.87
6	con3441	-619.95087	3.28	26.86	7.22	-30.81	-2.22
7	con2761	-619.95070	3.71	22.62	6.08	16.39	1
8	con2741	-619.95041	4.47	16.64	4.48	-57.93	-2.59
9	con3705	-619.95032	4.72	15.06	4.05	-67.76	-2.74
10	con2748	-619.94959	6.61	7.05	1.9	14.6	0.28
11	con344	-619.94931	7.36	5.23	1.41	10.94	0.15
12	con2781	-619.94918	7.72	4.54	1.22	48.52	0.59
13	con2998	-619.94915	7.78	4.43	1.19	96.82	1.15
14	con3318	-619.94886	8.55	3.24	0.87	-54.3	-0.47
15	con362	-619.94826	10.13	1.73	0.46		
16	con950	-619.94817	10.35	1.57	0.42		
17	con234	-619.94804	10.69	1.37	0.37		
18	con3432	-619.94800	10.8	1.32	0.35		
19	con3699	-619.94790	11.07	1.18	0.32		
20	con2850	-619.94787	11.14	1.15	0.31		
21	con2848	-619.94775	11.45	1.02	0.27		
22	con2890	-619.94771	11.56	0.97	0.26		
23	con586	-619.94747	12.18	0.76	0.2		
24	con3312	-619.94742	12.31	0.72	0.19		
25	con1192	-619.94720	12.91	0.56	0.15		
26	con2738	-619.94676	14.06	0.36	0.1		
27	con10	-619.94673	14.13	0.35	0.09		
28	con6	-619.94649	14.77	0.27	0.07		
29	con2916	-619.94613	15.71	0.18	0.05		
SKIPPED TO THE END OF THE LIST							
70	con95	-619.93204	52.69	0	0		
				sum			sum
				371.94			-49.28

**Table 3.14.** The low-energy conformers from the MC run with RNG seed of 1537699 for 3,5-Dimethoxyoctan-1-ol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/6-31++G\* level of theory. The  $[\alpha]_D$  value was calculated only for the first 14 conformers. The higher-energy conformers (15-70) are listed but not included in the average  $[\alpha]_D$ . Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

The third run was done using RNG seed of 37230011 and the temperature of the simulation was changed back to 300 K. 100 new conformers were found at the AM1 level of theory and sorted from low-energy to high-energy. The 33 lowest-energy conformers were re-optimized at B3LYP/6-31G\* level of theory, out of which eight were found to have a relative

energy of less than 10 kJ/mol and the  $[\alpha]_D$  value was found to be 8°. Conformers with a relative energy higher than 10 kJ/mol were found to be present in less than 1% abundance in the sample at 298 K and were considered not to be significant contributors to the average  $[\alpha]_D$  value. The 66 lowest-energy conformers were re-optimized at B3LYP/6-31G\* level of theory but no new low-energy conformers were found. The entire set of 100 AM1-level conformers was re-optimized using B3LYP/6-31G\* and still no new low-energy conformers were found. This result is consistent with the result found in Section 3.3 where all of the low-energy conformers at the B3LYP/6-31G\* level of theory came from the low-energy conformers at the AM1 level of theory. Table 3.15 shows the results from the 1-33 set. Again, one can save computer time by selecting only the first few dozen of low-energy AM1 geometries for high-level optimization, the rest will either be duplicates or too high in energy to significantly contribute to the  $[\alpha]_D$  value.

	1-33 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con12233	-619.95182	0	100	49.13	-48.09	-23.62
2	con12197	-619.95110	1.88	47.12	23.15	99.94	23.13
3	con12206	-619.95085	2.52	36.35	17.86	38.81	6.93
4	con79	-619.94928	6.66	6.93	3.4	10.94	0.37
5	con12200	-619.94916	6.98	6.08	2.99	48.52	1.45
6	con22	-619.94830	9.24	2.46	1.21	-38.66	-0.47
7	con15	-619.94830	9.24	2.46	1.21	-36.25	-0.44
8	con5778	-619.94817	9.57	2.16	1.06	18.07	0.19
				sum			sum
				203.56			7.55

**Table 3.15.** The low-energy conformers from the MC run with RNG seed of 37230011 for 3,5-Dimethoxyoctan-1-ol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/6-31++G\* level of theory. The  $[\alpha]_D$  value was calculated only for the first eight low-energy conformers. The higher-energy conformers are not listed. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Table 3.16 show the results from the fourth run done using RNG seed of 2789 where 100 conformers were found at the AM1 level of theory and sorted from low-energy to high-energy. The first 33 conformers were re-optimized at B3LYP/6-31G\* level of theory and only eight low-energy conformers were found. The  $[\alpha]_D$  value was found to be 124°. No new low-energy

conformers were found when the rest of the 100 AM1-level conformers were re-optimized.

	1-33 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con307	-619.95255	0	100	32.85	90.18	29.63
2	con1778	-619.95230	0.67	76.54	25.15	173.41	43.6
3	con302	-619.95215	1.07	65.09	21.38	176.68	37.78
4	con1012	-619.95149	2.78	32.78	10.77	43.71	4.71
5	con1753	-619.95103	4.01	20.05	6.59	163.42	10.76
6	con451	-619.94949	8.06	3.95	1.3	-108.45	-1.41
7	con1	-619.94929	8.57	3.22	1.06	10.94	0.12
8	con1682	-619.94914	8.96	2.75	0.9	-86.5	-0.78
			sum	304.38			sum
							124.41

**Table 3.16.** The low-energy conformers from the MC run with RNG seed of 2789 for 3,5-Dimethoxyoctan-1-ol. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/6-31++G\* level of theory. The  $[\alpha]_D$  value was calculated only for the first eight low-energy conformers. The higher-energy conformers are not listed. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

Two more long runs were conducted where 300 unique conformers were found for each run at the AM1 level of theory, but no new low-energy conformers were found at the B3LYP/6-31G\* level of theory and thus the conformational search was considered to have converged. The results (listed in Tables 3.13 – 3.16) from all of the MC runs were combined and repeating conformers removed. The final result of this calculation is summarized on Table 3.17 which shows a total of 30 low-energy conformers that were found with an average  $[\alpha]_D$  value of 61°.

	1-206 mc	E (au)	DE (kJ/mol)	relative Dist	Boltz. Dist. (%)	$[\alpha]_D$	Weighted $[\alpha]_D$
1	con307	-619.95	0	100	17.35	90.18	15.65
2	con1778	-619.95	0.67	76.54	13.28	173.41	23.03
3	con302	-619.95	1.07	65.09	11.29	176.68	19.95
4	con105	-619.95	1.3	59.42	10.31	-177.31	-18.28
5	con390	-619.95	1.99	45.05	7.82	-48.09	-3.76
6	con1012	-619.95	2.78	32.78	5.69	43.71	2.49
7	con1050	-619.95	3.75	22.22	3.86	100.92	3.89
8	con2739	-619.95	3.79	21.88	3.8	4.75	0.18
9	con12197	-619.95	3.82	21.66	3.76	99.94	3.76
10	con1753	-619.95	4.01	20.05	3.48	163.42	5.68
11	con231	-619.95	4.46	16.72	2.9	38.81	1.13
12	con315	-619.95	4.53	16.29	2.83	-30.81	-0.87
13	con2761	-619.95	4.86	14.23	2.47	16.39	0.4
14	con1026	-619.95	5.6	10.58	1.83	-57.93	-1.06
15	con739	-619.95	5.66	10.33	1.79	133.1	2.38
16	con845	-619.95	5.84	9.62	1.67	217.88	3.64
17	con813	-619.95	5.86	9.53	1.65	200.05	3.31
18	con403	-619.95	6.06	8.81	1.53	-67.76	-1.04
19	con1008	-619.95	6.68	6.87	1.19	49.14	0.59
20	con2748	-619.95	7.77	4.44	0.77	14.6	0.11
21	con985	-619.95	7.84	4.31	0.75	-3.43	-0.03
22	con451	-619.95	8.06	3.95	0.69	-108.45	-0.74
23	con344	-619.95	8.52	3.29	0.57	10.94	0.06
24	con584	-619.95	8.84	2.89	0.5	48.52	0.24
25	con2998	-619.95	8.93	2.79	0.48	96.82	0.47
26	con1682	-619.95	8.96	2.75	0.48	-86.5	-0.41
27	con121	-619.95	9.62	2.11	0.37	-54.3	-0.2
28	con22	-619.95	11.18	1.13	0.2	-38.66	-0.08
29	con15	-619.95	11.18	1.13	0.2	-36.25	-0.07
30	con5778	-619.95	11.5	0.99	0.17	18.07	0.03
				sum			sum
				576.4			61.15

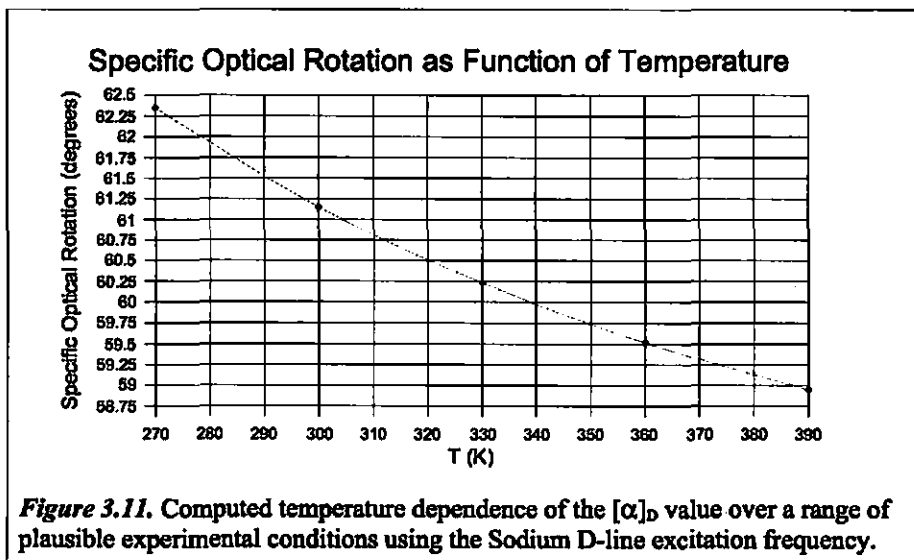
**Table 3.17.** The lowest-energy conformers of 3,5-Dimethoxyoctan-1-ol from the MC run with RNG seed of 549078 combined with the lowest-energy conformers from the MC run with RNG seed of 1527699, RNG seed of 37230011, and RNG seed of 2789. The energies (E) shown are from B3LYP/6-31G\* level of theory and the specific rotation values were calculated using B3LYP/6-31++G\* level of theory. Boltzmann distribution (Boltz. Dist.) was calculated as a percent.

According to the calculations, one would expect the experimental  $[\alpha]_D$  value to be between 100 and 20 degrees if the unknown is the (R,R) isomer or between -20 and -100 degrees in the case that the unknown is the (S,S) isomer. To be more certain of the absolute stereochemistry assigned by this method, several trends can be established. For example, the temperature at which the  $[\alpha]_D$  is measured could be varied. This would slightly change the Boltzmann population of the conformers and thus it will change both experimental and calculated

$[\alpha]_D$  value. The variation of  $[\alpha]_D$  as a function of temperature should be about the same for the experimental and theoretical methods if all of the important low-energy conformers were indeed found. Furthermore,  $[\alpha]_D$  is also a function of incident wavelength. The calculations and the experiment can be repeated using frequencies other than that of the Sodium D-line and again, the trends should be similar. Finding the trend in temperature of the average  $[\alpha]_D$  value at any particular wavelength is easy as the only thing that needs adjusting is the temperature variable in the Boltzmann equation. The results for the (*R,R*) isomer are presented in Table 3.17 and Figure 3.11. Finding the trend in wavelength requires changing one of the parameters in the linear response calculation input file, and while simple to do if needed, these calculations are time consuming and are not presented in this work.

T (K)	Weighted $[\alpha]_D$
270	62.35
300	61.15
330	60.23
360	59.52
390	58.95

**Table 3.18.** This table shows a trend in the average  $[\alpha]_D$  value as a function of temperature. Different temperatures were used to calculate the contributions of conformers to the average  $[\alpha]_D$  value using the Boltzmann statistics.

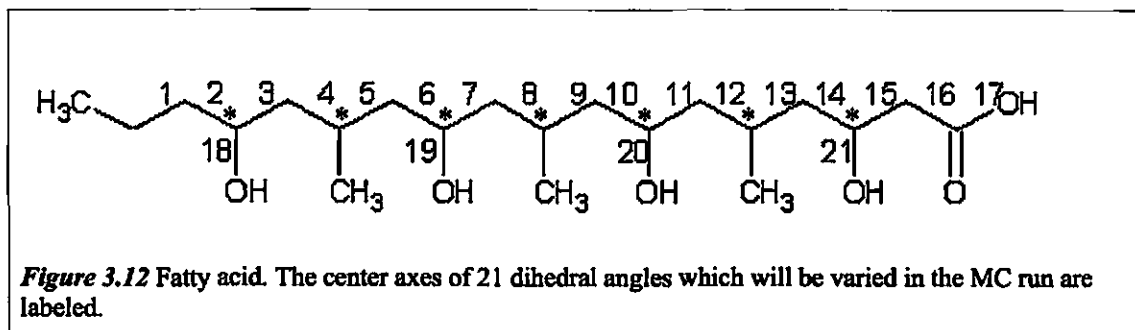


**Figure 3.11.** Computed temperature dependence of the  $[\alpha]_D$  value over a range of plausible experimental conditions using the Sodium D-line excitation frequency.

### 3.7 Recommended procedure for finding the absolute stereochemistry of the fatty acid

Comparing the calculated  $[\alpha]_D$  value of 3,5-dimethoxyoctan-1-ol (Section 3.6) with the experimental value will show whether the new method for calculating the specific rotation of flexible molecules developed in Sections 3.1-3.5 needs further adjusting. Currently, the experimental value for  $[\alpha]_D$  remains unavailable for 3,5-dimethoxyoctan-1-ol.

Assuming that the calculated result of the average  $[\alpha]_D$  value is in good agreement with the experimental result, the computational determination of the absolute stereochemistry of the fatty acid shown in Figure 3.12 should also be possible. This flexible molecule has seven chiral centers and 21 dihedral angles which can be varied using the MC program at the AM1 level of theory. The geometry optimizations at the B3LYP/6-31G\* level of theory can be done in a timely fashion provided that the GAMESS program runs in parallel on at least four processors. The  $[\alpha]_D$  calculation (keeping in mind the results obtained in Section 3.4) can at this time only be executed using the B3LYP/6-31++G\* level of theory using Gaussian03 software running in parallel on at least 16 processors on an IBM SP/RS6000 P3 node at MHPCC.



The starting geometry coordinates for the *syn* and the *anti* isomers of the fatty acid are provided in the Appendix.

## CHAPTER 4

### SUMMARY AND CONCLUSIONS

This work provides an answer to the challenges faced when one needs to find a theoretical value for specific optical rotation ( $[\alpha]_D$ ) for a relatively large and flexible molecule. A method was developed to adequately find all of the low-energy conformers that may be present in solution and calculate the  $[\alpha]_D$  values for these conformers without requiring an unreasonable amount of computer time.

Initially, the computational requirements needed to calculate reliable  $[\alpha]_D$  values for a single conformer were investigated. First, the influence of changes in molecular geometry on  $[\alpha]_D$  values were investigated using several small and rigid molecules. (2*S*)-2-methyl oxirane was used to investigate the effect of rotating the methyl group on calculated  $[\alpha]_D$  values. It was determined that the variation in  $[\alpha]_D$  values was substantial with the variation of the dihedral angle of the methyl group (HCCO) with both HF and B3LYP using either 6-31G\* or aug-cc-pVDZ basis sets. The  $[\alpha]_D$  values calculated with these two basis sets varied differently from each other with geometry but they produced comparable  $[\alpha]_D$  values at the optimized geometry. This difference in variation was attributed to the diffuse p-orbital basis functions on the hydrogen atoms which are present in the aug-cc-pVDZ basis set and absent in the 6-31G\* basis set. All of the hydrogen atoms in (2*S*)-2-methyl oxirane were replaced with fluorine atoms and the variation of calculated  $[\alpha]_D$  values with respect to the FCCO dihedral angle between the 6-31G\* basis set and the aug-cc-pVDZ basis set became more consistent. 2-propanol was used to show that even though it is achiral and thus has an  $[\alpha]_D$  value of zero, there are several low-energy conformers of 2-propanol which have non-zero  $[\alpha]_D$  values. The three low-energy conformers were obtained by rotating the

hydroxy group about the HOCH dihedral angle: the *trans* conformer has  $C_s$  symmetry and consequently an  $[\alpha]_D$  value of zero. However, the two *gauche* conformers had  $C_1$  symmetry and they had  $[\alpha]_D$  values which were of the same magnitude but opposite in sign. The average calculated  $[\alpha]_D$  values obtained using Boltzmann statistics was zero.

The second aspect of this thesis was the development of a Monte Carlo (MC) program to find low-energy conformers of flexible molecules. The MC program was designed to vary the user-specified dihedral angles of the flexible molecules and perform geometry optimizations at the AM1 level of theory. The AM1 low-energy conformers found using the MC program were re-optimized at the B3LYP/6-31G\* level of theory and  $[\alpha]_D$  values were calculated at the B3LYP/aug-cc-pVDZ level of theory.

The effectiveness of the MC program was tested using the achiral molecule 3-hydroxypentadiol where five dihedral angles were varied to make sure that just as in the 2-propanol case the average  $[\alpha]_D$  value would converge to zero. After two MC runs, all of the low-energy conformers that were significantly contributing to the average  $[\alpha]_D$  value were found and even though some of the individual conformers had relatively large non-zero  $[\alpha]_D$  values, the average  $[\alpha]_D$  value converged to zero. The chiral molecule (*S*)-(+)-ethyl 3-hydroxybutanoic acid was used to investigate the magnitude of error which can be expected for the calculated  $[\alpha]_D$  value when five dihedral angles were varied in the MC routine to find the low-energy conformers. After six MC runs, the average  $[\alpha]_D$  value converged to  $79^\circ$  which overestimates the experimental value of  $43^\circ$  by  $36^\circ$ . Thus the error for the calculated average  $[\alpha]_D$  value can be expected to be in the range of  $\pm 40^\circ$  which is consistent with errors found in the literature for the basis sets and methods used in these calculations.

The absolute stereochemistry of larger chiral molecule (3,5-dimethoxyoctan-1-ol) with two chiral centers and nine dihedral angles that were varied in the MC run is currently unknown.



The relative stereochemistry of the two chiral centers was determined to be *syn*. After seven MC runs, the average  $[\alpha]_D$  value converged to  $61^\circ$  for an (*R,R*) conformer. The experimental  $[\alpha]_D$  value is needed to assess if the calculated value is correct within the expected error of  $\pm 40^\circ$  and to determine the absolute stereochemistry of 3,5-dimethoxyoctan-1-ol.

Using the methods developed in this work, the calculated  $[\alpha]_D$  values can be found for even larger flexible molecules such as the fatty acid which has seven chiral centers would require 21 dihedral angles to be varied in the MC conformer search. The absolute stereochemistry of this natural product is unknown. Due to the limitations of other methods such as NMR and X-ray diffraction, comparing the calculated average  $[\alpha]_D$  value against the experimentally measured  $[\alpha]_D$  value may be the only viable option for elucidating the absolute stereochemistry of the fatty acid.

## APPENDIX

### APPENDIX TABLE OF CONTENTS

Example of starting geometry of 3-hydroxypentadiol for MC routine	89
Fortran MC code for 3-hydroxypentadiol	90
Fortran sorting subroutine for MC code for 3-hydroxypentadiol	95
Example of starting geometry of ( <i>S</i> )-(+)-ethyl 3-hydroxybutanoic acid for MC routine	98
Fortran MC code for ( <i>S</i> )-(+)-ethyl 3-hydroxybutanoic acid	99
Fortran sorting subroutine for MC code for ( <i>S</i> )-(+)-ethyl 3-hydroxybutanoic acid	104
Example of starting geometry of (R,R)-3,5-Dimethoxyoctan-1-ol for MC routine	107
Fortran MC code for (R,R)-3,5-Dimethoxyoctan-1-ol	108
Fortran sorting subroutine for MC code for (R,R)-3,5-Dimethoxyoctan-1-ol	114
Starting geometry of <i>syn</i> -fatty acid for MC routine	118
Starting geometry of <i>anti</i> -fatty acid for MC routine	120
Auxiliary subroutines and files for MC code	122

```

!
! 01-jan-2006
! This is the initial part of a gamess input -
! just need to add coordinates and $END record
!
! internal coordinates for 3-hydroxypentadia
! SET UP TO DO FULL GEOMETRY OPTIMIZATION
!
$CONTRL RUNTYP=OPTIMIZE COORD=ZMT NZVAR=42 $END
$SYSTEM TIMLIM=100 MEMORY=1000000 $END
$BASIS GBASIS=AM1 $END
$STATPT NSTEP=40 $END
$SCF DIRSCF=.TRUE. $END
$DATA
3-hydroxypentadia
C1
C1
C2      1      1.5018033
C3      2      1.5289824      1      110.0593520
C4      3      1.5319008      2      111.0116198      1      -173.3202150
C5      4      1.4998097      3      113.6601177      2      -110.1406153
O6      1      1.2308130      2      122.6983772      3           92.4289010
O7      5      1.2319074      4      122.8736750      3      -59.8823890
O8      3      1.4188757      2      106.2421297      4      -122.8968533
H9      1      1.1119719      2      115.1678255      3      -86.3369992
H10     2      1.1211056      3      108.7616299      1      120.6811713
H11     2      1.1201129      3      109.6476045      1      -120.3874760
H12     3      1.1287678      4      108.7071841      8      121.0453268
H13     4      1.1208888      5      109.3790700      3      -123.0232655
H14     4      1.1238491      5      107.8246378      3      120.7092303
H15     5      1.1125060      4      115.6525911      3      121.1710252
H16     8      0.9675670      3      107.5330959      12     -62.5734969
$END

```

```

c
c  program to do monte-carlo searches for low energy minima
c  John D. Head  04-jan-06
c
c  test of reading gamess gam1 output
c  assumes is opt not converged - can still get useful
c  energy, RMS grad and coordinate info

      implicit real*8 (a-h,o-z)
c
parameter (natom=16)
parameter (nzvar=3*natom-6)
dimension geom(3,natom), icor(nzvar), zcor(nzvar)
dimension zold(nzvar), zfull(nzvar)
character*6 at_char(natom)
character*5 char5(natom)
character conf*5
logical accept
integer fill
integer new

dimension results(300,6)
dimension current(1,6)

integer resint(300,2)
integer curint(1,2)
c
c  in gamrun
c  ioptyp=0 for full geometry optimization
c  ioptyp=1 for partial geometry optimization
c
      nseed=57948
      iaccept=0
      iboltz=0
      nsteps=50000
      fill=0
c
c  copy output to gam1234.out
c
c  this is assuming gamstart.out is a constrained optimization
c  and not a local min run
      ioptyp=1
c  if gamstart.out a local min run then use
c  ioptyp=0
      call system('/bin/cp -f gamstart.out gam1234.out')
      call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$      ioptyp,at_char,icor,zcor,char5)
write(6,*) ' initial zcor(6) =',zcor(6)
write(6,*) ' initial zcor(9) =',zcor(9)
write(6,*) ' initial zcor(12) =',zcor(12)
write(6,*) ' initial zcor(15) =',zcor(15)
write(6,*) ' initial zcor(42) =',zcor(42)
write(6,*) ' type of calculation result:',icalc_complete
write(6,*) ' fitness:',fitness

```

```

write(6,*)' grad rms:',rms
eold = fitness

c
c save old coordinates and get local minimum
c
do 90 jzvar=1,nzvar
zfull(jzvar)=zcor(jzvar)
90 zold(jzvar)=zcor(jzvar)
ifull=1

c
c open file for full opt results
c
open(8,file='fulloptres.yeah',status='unknown',
1 form='formatted')
open(9,file='fulloptres.sum',status='unknown',
1 form='formatted')
write(conf,*) ifull
! open(7,file='cartcord'//conf(2:5), status='unknown',
! 1 form='formatted')

1122 format(a5,3f12.7)

ioptyp=0
call gamrun(natom,nzvar,ioptyp,at_char,icor,zfull)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$ ioptyp,at_char,icor,zfull,char5)
call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)

write(6,*)' type of calculation result:',icalc_complete
write(6,*)' local min e:',fulle
write(6,*)' grad rms:',frms
write(6, '(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(42), ' FE:',fulle
write(9, '(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(42), ' FE:',fulle
write(8, '(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(42), ' FE:',fulle

call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$ at_char,icor,zfull,results,current,resint,curint,new)
fulle_old=fulle
zfo6=zfull(6)
zfo9=zfull(9)
zfo12=zfull(12)
zfo15=zfull(15)
zfo42=zfull(42)

```

```

c
c start looping over lots of geometries
c
do 100 i=1,nsteps
del6=50.*(ezrandom(nseed)-0.5)
del9=50.*(ezrandom(nseed)-0.5)
del12=50.*(ezrandom(nseed)-0.5)
del15=50.*(ezrandom(nseed)-0.5)
del42=50.*(ezrandom(nseed)-0.5)
write(6,*) 'delta dihedral'
write(6,*) del6, del9, del12, del15
zcor(6)=zcor(6)+del6
if(zcor(6).gt.360.) then
  zcor(6)=zcor(6)-360.
else if(zcor(6).gt.180.) then
  zcor(6)=zcor(6)+360.
endif
zcor(9)=zcor(9)+del9
if(zcor(9).gt.360.) then
  zcor(9)=zcor(9)-360.
else if(zcor(9).gt.180.) then
  zcor(9)=zcor(9)+360.
endif
zcor(12)=zcor(12)+del12
if(zcor(12).gt.360.) then
  zcor(12)=zcor(12)+360.
else if(zcor(12).gt.180.) then
  zcor(12)=zcor(12)-360.
endif
zcor(15)=zcor(15)+del15
if(zcor(15).gt.360.) then
  zcor(15)=zcor(15)-360.
else if(zcor(15).gt.180.) then
  zcor(15)=zcor(15)+360.
endif
zcor(42)=zcor(42)+del42
if(zcor(42).gt.360.) then
  zcor(42)=zcor(42)-360.
else if(zcor(42).gt.180.) then
  zcor(42)=zcor(42)+360.
endif
write(6, '(a,i5,a,5f13.7)') ' GAMESS run number:',i,' dihedrals:',
1 zcor(6),zcor(9),zcor(12),zcor(15),zcor(42)
ioptyp=1
call gamrun(natom,nzvar,ioptyp,at_char,icor,zcor)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$ ioptyp,at_char,icor,zcor,char5)
write(6,*) ' type of calculation result:',icalc_complete
write(6,*) ' fitness:',fitness
write(6,*) ' grad rms:',rms
deltae=fitness-eold
write(6,*) ' energy change =',deltae
c

```

```

c now see if energy is going down
c
  temp = 0.0006
  if(deltae.lt.0.)then
    accept=.true.
c
  ioptyp=0
  do 35 jzvar=1,nzvar
35  zfull(jzvar)=zcor(jzvar)
    ifull=ifull+1
    write(6,*)' PERFORMING FULL GEOM OPT',ifull
    call gamrun(natom,nzvar,ioptyp,at_char,icor,zfull)
    call system('/bin/mv -f gamrun.usi_out gam1234.out')
    call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$      ioptyp,at_char,icor,zfull,char5)
    write(6,*)' type of calculation result:',icalc_complete
    write(6,*)' local min e:',fulle
    write(6,*)' grad rms:',frms
    write(6,'(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(42),' FE:',fulle
    write(6,*)' YEAH: found local minimum!!'
    write(9,*)' YEAH: found local minimum!!'
    write(8,*)' YEAH: found local minimum!!'
    write(8,'(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(42),' FE:',fulle
    new=0

    call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$      at_char,icor,zfull,results,current,resint,curint,new)
    if (new.eq.1) then

      call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)

      call system('/bin/mv -f cartcord* carts')
      else
      end if

      if (new.eq.2) then
      go to 101
      end if

      write(9,'(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),
1 zfull(15),zfull(42),' FE:',fulle
      fulle_old=fulle
      zfo6=zfull(6)
      zfo9=zfull(9)
      zfo12=zfull(12)
      zfo15=zfull(15)
      zfo42=zfull(42)
c

```

```

else if(exp(-deltae/temp).lt.ezrandom(nseed))then
  write(6,*)' accepted increase in energy'
  accept=.true.
  iboltz=iboltz+1
else
  accept=.false.
end if
eold=fitness
if(accept)then
  iaccept=iaccept+1
  do 120 jzvar=1,nzvar
120   zold(jzvar)=zcor(jzvar)
  else
  do 130 jzvar=1,nzvar
130   zcor(jzvar)=zold(jzvar)
  endif
  write(6,140) i, iaccept, iboltz, fitness, deltae, zcor(6), zcor(9),
1   zcor(12), zcor(15), zcor(42)
140  format(3i5, ' e:', f12.7, ' de:', f11.7, ' dihedral:', 5f12.6)
100  continue
101  write(6,*)' GAMESS runs completed'
  write(6,*)' Out of ', i, ' geoms: no boltzman accepts:', iboltz
  write(6,*)' Number of negative de:', iaccept-iboltz
  write(6,*)' *****'
  stop
end

```



```

subroutine sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$   at_char,icor,zfull,results,current,resint,curint,new)

implicit real*8 (a-h,o-z)

parameter (Ethresh=0.00005)
parameter (Athresh=10.0)
parameter (Athresh2=290.0)
parameter (maxsize=6)

dimension icor(nzvar),zfull(nzvar)
character*6 at_char(natom)

dimension results(300,6)
dimension current(1,6)

dimension temp1(300,maxsize)
integer temp2(300,2)

integer resint(300,2)
integer curint(1,2)

real Echeck
real Acheck
integer fill
integer new

C   ncon is 360/120 ^(# DH), where 120 is the increment
ncon=60

open (unit=7,
file='unsorted_result',status='unknown',form='formatted')

curint(1,1)=ifull
current(1,1)=fulle
current(1,2)=zfull(6)
current(1,3)=zfull(9)
current(1,4)=zfull(12)
current(1,5)=zfull(15)
current(1,6)=zfull(42)
curint(1,2)=icalc_complete
83 format(i5,i5,f12.7,5f8.2)
84 format(i5,i5,i5,f12.7,5f8.2)

C   FILL IN THE FIRST DATA POINT

if (fill.eq.0) then
c       i=fill
       do 10 j=1,maxsize
       results(1,j) = current(1,j)
10 continue
       do 11 j=1,2
       resint(1,j) = curint(1,j)
11 continue

```

```

    fill=fill+1
    go to 50
  else
  end if

C   COMPARE INCOMING DATA TO THE EXSISTING ARRAY

    if (current(1,1).eq.0.) then
      go to 50
    end if
    if (curint(1,2).ne.1.) then
      go to 50
    end if

    ord=1

60  do 20 i=ord,fill

    Echeck = abs(results(i,1)-current(1,1))

    if (Echeck.le.Ethresh) then
C     CHECK THE I-TH DIHEDRALS
      chk=i
      do 30 j=2,maxsize

c       write(7,*) ifull,' checking with # ', chk
      Acheck=abs(results(chk,j)-current(1,j))
c       write(7,*) 'Dangle: ', Acheck
      if
((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.eq.maxsize)) then
        write(7,*) ifull, ' is a match - discarding'
        go to 50
      else if
((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.le.maxsize-1))
then
c       write(7,*) ifull, ' matching in progress'
        go to 30
      else
        if (ord.lt.fill) then
          ord=chk+1
c       write(7,*) 'new conf - check for more'
          go to 60
        else
          write(7,*) ifull, ' is a new conf - adding'
          new=1
          go to 31
        end if
31      fill=fill+1
        do 12 l=1,maxsize
          results(fill,l) = current(1,l)
12      continue
        do 13 l=1,2
          resint(fill,l) = curint(1,l)
13      continue

```

```

30  continue
    else
        if (ord.eq.fill) then
            new=1
            write(7,*) ifull,' is a new conf - adding'
            fill=fill+1
            chk=i
            do 14 l=1,maxsize
                results(fill,l) = current(1,l)
14  continue
            do 15 l=1,2
                resint(fill,l) = curint(1,l)
15  continue
            else
                chk=i
                ord=chk+1
                go to 60
            end if
        end if
20  continue
50  continue

    if (fill.ge.ncon) then
        write(7,*) 'results'
        write(7,83) ((resint(i,l),l=1,2), (results(i,n),n=1,maxsize),
i=1,fill)
        is=1
c      write(7,83)
        ((resint(k,l),l=1,2), (results(k,n),n=1,3),k=1,ncon)
        do 41 j=1,fill
            do 40 k=1,fill
                if (results(k,1).gt.results(k+1,1)) then
                    do n=1,maxsize
                        temp1(k,n)=results(k,n)
                        results(k,n)=results(k+1,n)
                        results(k+1,n)=temp1(k,n)
                    enddo
                    do n=1,2
                        temp2(k,n)=resint(k,n)
                        resint(k,n)=resint(k+1,n)
                        resint(k+1,n)=temp2(k,n)
                    enddo
                else
                    end if
40  continue
41  continue
        write(7,*) 'sorted results'
        write(7,84)
        (k, (resint(k,l),l=1,2), (results(k,n),n=1,maxsize),k=1,fill)
        new=2
c      else
        endif
        return
100  end

```

```

!
! 01-jan-2006
! This is the initial part of a gamess input -
! just need to add coordinates and $END record
!
! internal coordinates for (S)-(+)-ethyl 3-hydroxybutanoic acid
! SET UP TO DO FULL GEOMETRY OPTIMIZATION
!
$CONTRL RUNTYP=OPTIMIZE  COORD=ZMT  NZVAR=57  $END
$SYSTEM TIMLIM=100  MEMORY=1000000  $END
$BASIS  GBASIS=AM1  $END
$STATPT NSTEP=90  $END
$SCF  DIRSCF=.TRUE.  $END
$DATA
(S)-(+)-ethyl 3-hydroxybutanoic acid
C1
C1
C2      1      1.5166977
C3      2      1.5319558      1      111.7047643
C4      3      1.4977385      2      111.0864125      1      -109.2650311
O5      4      1.3712603      3      123.2415517      2      -111.3936090
C6      5      1.4268577      4      120.6425093      3      16.4720728
C7      6      1.5082012      5      112.5063585      4      68.5334522
O8      4      1.2337724      5      111.5539664      3      176.5445553
O9      2      1.4258273      3      105.6342344      1      -121.3773897
H10     9      0.9640144      2      107.1382116      1      46.9416259
H11     1      1.1165179      2      110.2008137      3      177.9780036
H12     1      1.1162999      2      110.4928217      3      -62.0596155
H13     1      1.1164066      2      109.7545761      3      58.2235070
H14     2      1.1285925      3      110.1801930      1      121.8641516
H15     3      1.1230204      4      107.5116299      2      -120.2384692
H16     3      1.1215153      4      111.5727105      2      121.9661681
H17     6      1.1209766      7      111.6131598      5      124.8470458
H18     6      1.1266879      7      110.8211543      5      -113.5512434
H19     7      1.1186073      6      110.7777035      5      -71.1805962
H20     7      1.1165677      6      110.4292833      5      49.2546024
H21     7      1.1162986      6      109.4379945      5      169.0614046
$END

```

```

c
c  program to do monte-carlo searches for low energy minima
c  John D. Head  04-jan-06
c
c  test of reading gamess gam1 output
c  assumes is opt not converged - can still get useful
c  energy, RMS grad and coordinate info

      implicit real*8 (a-h,o-z)

      parameter (natom=21)
      parameter (nzvar=3*natom-6)
      dimension geom(3,natom),icor(nzvar),zcor(nzvar)
      dimension zold(nzvar),zfull(nzvar)
      character*6 at_char(natom)
      character*5 char5(natom)
      character conf*5
      logical accept
      integer fill
      integer new

      dimension results(5000,6)
      dimension current(1,6)

      integer resint(5000,2)
      integer curint(1,2)

c
c  in gamrun
c  iopty=0 for full geometry optimization
c  iopty=1 for partial geometry optimization
c
      nseed=97540032
      iaccept=0
      iboltz=0
      nsteps=90000
      fill=0

c
c  copy output to gam1234.out
c
c  this is assuming gamstart.out is a constrained optimization
c  and not a local min run
      iopty=1
c  if gamstart.out a local min run then use
c  iopty=0
      call system('/bin/cp -f gamstart.out gam1234.out')
      call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$      iopty,at_char,icor,zcor,char5)
      write(6,*)' initial zcor(6) =',zcor(6)
      write(6,*)' initial zcor(9) =',zcor(9)
      write(6,*)' initial zcor(12) =',zcor(12)
      write(6,*)' initial zcor(15) =',zcor(15)
      write(6,*)' initial zcor(24) =',zcor(24)
      write(6,*)' type of calculation result:',icalc_complete
      write(6,*)' fitness:',fitness

```

```

write(6,*)' grad rms:',rms
eold = fitness

c
c save old coordinates and get local minimum
c
do 90 jzvar=1,nzvar
zfull(jzvar)=zcor(jzvar)
90 zold(jzvar)=zcor(jzvar)
ifull=1

c
c open file for full opt results
c
open(8,file='fulloptres.yeah',status='unknown',
1 form='formatted')
open(9,file='fulloptres.sum',status='unknown',
1 form='formatted')
write(conf,*) ifull

ioptyp=0
call gamrun(natom,nzvar,ioptyp,at_char,icor,zfull)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$ ioptyp,at_char,icor,zfull,char5)
call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)

write(9,'(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(24),' FE:',fulle
write(8,'(a,i5,a,5f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(24),' FE:',fulle

call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$ at_char,icor,zfull,results,current,resint,curint,new)
fulle_old=fulle
zfo6=zfull(6)
zfo9=zfull(9)
zfo12=zfull(12)
zfo15=zfull(15)
zfo24=zfull(24)

c
c start looping over lots of geometries
c
do 100 i=1,nsteps
del6=50.*(ezrandom(nseed)-0.5)
del9=50.*(ezrandom(nseed)-0.5)
del12=50.*(ezrandom(nseed)-0.5)
del15=50.*(ezrandom(nseed)-0.5)
del24=50.*(ezrandom(nseed)-0.5)

write(6,*) 'delta dihedral'

```

```

write(6,*) del6, del9, del12, del15, del24
zcor(6)=zcor(6)+del6
if(zcor(6).gt.360.)then
  zcor(6)=zcor(6)-360.
else if(zcor(6).gt.180.)then
  zcor(6)=zcor(6)+360.
endif
zcor(9)=zcor(9)+del9
if(zcor(9).gt.360.)then
  zcor(9)=zcor(9)-360.
else if(zcor(9).gt.180.)then
  zcor(9)=zcor(9)+360.
endif
zcor(12)=zcor(12)+del12
if(zcor(12).gt.360.)then
  zcor(12)=zcor(12)+360.
else if(zcor(12).gt.180.)then
  zcor(12)=zcor(12)-360.
endif
zcor(15)=zcor(15)+del15
if(zcor(15).gt.360.)then
  zcor(15)=zcor(15)-360.
else if(zcor(15).gt.180.)then
  zcor(15)=zcor(15)+360.
endif
zcor(24)=zcor(24)+del24
if(zcor(24).gt.360.)then
  zcor(24)=zcor(24)-360.
else if(zcor(18).gt.180.)then
  zcor(24)=zcor(24)+360.
endif

write(6,'(a,i5,a,5f13.7)')' GAMESS run number:',i,' dihedrals:',
1  zcor(6),zcor(9),zcor(12),zcor(15),zcor(24)
ioptyp=1
call gamrun(natom,nzvar,ioptyp,at_char,icor,zcor)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$ ioptyp,at_char,icor,zcor,char5)
deltae=fitness-eold
write(6,*)' energy change =',deltae

c
c now see if energy is going down
c
temp = 0.0012
if(deltae.lt.0.)then
  accept=.true.

c
ioptyp=0
do 35 jzvar=1,nzvar
35  zfull(jzvar)=zcor(jzvar)
    ifull=ifull+1

write(6,*)' PERFORMING FULL GEOM OPT',ifull

```

```

call gamrun(natom,nzvar,iotyp,at_char,icor,zfull)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$ iotyp,at_char,icor,zfull,char5)

write(8,'(a,i5,a,5f13.7,a,f13.7)') ' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(24),' FE:',fulle

new=0

call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$ at_char,icor,zfull,results,current,resint,curint,new)

write(6,*) 'new = ', new, ' 0 --> skipping, 1 --> adding, 2 -->
matrix complete'

if (new.eq.1) then
call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)

call system('/bin/mv -f cartcord* carts')
else
end if

if (new.eq.2) then
go to 101
end if

fulle_old=fulle
zfo6=zfull(6)
zfo9=zfull(9)
zfo12=zfull(12)
zfo15=zfull(15)
zfo24=zfull(24)

else if (exp(-deltae/temp).lt.ezrandom(nseed)) then
write(6,*) ' accepted increase in energy'
accept=.true.
iboltz=iboltz+1
else
accept=.false.
end if
eold=fitness
if (accept) then
iaccept=iaccept+1
do 120 jzvar=1,nzvar
120 zold(jzvar)=zcor(jzvar)
else
do 130 jzvar=1,nzvar
130 zcor(jzvar)=zold(jzvar)
endif
c write(6,140) i, iaccept, iboltz, fitness, deltae, zcor(6), zcor(9),

```



```
c      1 zcor(12),zcor(15),zcor(18),zcor(27),zcor(30),zcor(33),zcor(36)
140    format(3i5,' e:',f12.7,' de:',f11.7,' dihedral:',9f12.6)
100    continue
101    write(6,*)' GAMESS runs completed'
      write(6,*)' Out of ',i,' geoms: no boltzman accepts:',iboltz
      write(6,*)' Number of negative de:',iaccept-iboltz
      write(6,*)' *****'
      stop
      end
```

```

subroutine sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$   at_char,icor,zfull,results,current,resint,curint,new)

implicit real*8 (a-h,o-z)

parameter (Ethresh=0.00005)
parameter (Athresh=10.0)
parameter (Athresh2=290.0)
parameter (maxsize=6)

dimension icor(nzvar),zfull(nzvar)
character*6 at_char(natom)

dimension results(5000,maxsize)
dimension current(1,maxsize)

dimension temp1(5000,maxsize)
integer temp2(5000,2)

integer resint(5000,2)
integer curint(1,2)

real Echeck
real Acheck
integer fill
integer new

C   ncon is 360/120 ^(# DH), where 120 is the increment
ncon=140

open (unit=7,
file='unsorted_result',status='unknown',form='formatted')

curint(1,1)=ifull
current(1,1)=fulle
current(1,2)=zfull(6)
current(1,3)=zfull(9)
current(1,4)=zfull(12)
current(1,5)=zfull(15)
current(1,6)=zfull(24)

curint(1,2)=icalc_complete
83 format(i5,i5,f12.7,5f8.2)
84 format(i5,i5,i5,f12.7,5f8.2)

C   FILL IN THE FIRST DATA POINT

if (fill.eq.0) then
c       i=fill
do 10 j=1,maxsize
results(1,j) = current(1,j)
10 continue
do 11 j=1,2
resint(1,j) = curint(1,j)

```

```

11         continue
        fill=fill+1
        go to 50
        else
        end if

C    COMPARE INCOMING DATA TO THE EXSISTING ARRAY

        if (current(1,1).eq.0.) then
            go to 50
        end if
        if (curint(1,2).ne.1.) then
            write(7,*) curint(1,1), ' has exit code of: ',curint(1,2), '
- discarding'
            write(6,*) curint(1,1), ' has exit code of: ',curint(1,2), '
- discarding'
            go to 50
        end if

        ord=1

60    do 20 i=ord,fill

        Echeck = abs(results(i,1)-current(1,1))

        if (Echeck.le.Ethresh) then
            write(7,*) curint(1,1), ' and ',resint(i,1),' have Ethresh
of ', Echeck,' - checking DH'
            write(6,*) curint(1,1), ' and ',resint(i,1),' have Ethresh
of ', Echeck,' - checking DH'
C        CHECK THE I-TH DIHEDRALS
            chk=i
            do 30 j=2,maxsize

                Acheck=abs(results(chk,j)-current(1,j))

                if
                (((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.eq.maxsize)) then
                    write(7,*) curint(1,1), ' is a match to ',
resint(chk,1),' - discarding'
                    write(6,*) curint(1,1), ' is a match to ',
resint(chk,1),' - discarding'
                    go to 50
                else if
                (((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.le.maxsize-1))
then
                    go to 30
                else
                    if (ord.lt.fill) then
                        ord=chk+1
                        go to 60
                    else
                        write(7,*) curint(1,1), ' is a new conf - adding
(Athresh)'

```

```

write(6,*) curint(1,1), ' is a new conf - adding
(Athresh)'
new=1
go to 31
end if
31 fill=fill+1
do 12 l=1,maxsize
results(fill,1) = current(1,1)
12 continue
do 13 l=1,2
resint(fill,1) = curint(1,1)
13 continue
end if
30 continue
else

if (ord.eq.fill) then
new=1
write(7,*) ifull,' is a new conf - adding (Ethresh)'
write(6,*) ifull,' is a new conf - adding (Ethresh)'
fill=fill+1
chk=i
do 14 l=1,maxsize
results(fill,1) = current(1,1)
14 continue
do 15 l=1,2
resint(fill,1) = curint(1,1)
15 continue
else
chk=i
ord=chk+1
go to 60
end if
end if

20 continue
50 continue

if (fill.ge.ncon) then
write(7,*) 'results'
write(7,83) ((resint(i,1),l=1,2), (results(i,n),n=1,maxsize),
i=1,fill)
is=1
c write(7,83)
((resint(k,1),l=1,2), (results(k,n),n=1,3),k=1,ncon)
do 41 j=1,fill
do 40 k=1,fill
if (results(k,1).gt.results(k+1,1)) then
do n=1,maxsize
templ(k,n)=results(k,n)
results(k,n)=results(k+1,n)
results(k+1,n)=templ(k,n)

```

```

        enddo
        do n=1,2
            temp2(k,n)=resint(k,n)
            resint(k,n)=resint(k+1,n)
            resint(k+1,n)=temp2(k,n)
        enddo
    else
    end if
40  continue
41  continue
        write(7,*) 'sorted results'
        write(7,84) (k,(resint(k,l),l=1,2),(results(k,n),n=1,maxsize
),k=1,fill)
        new=2
    endif
    return
100 end

```

```

!
! 01-jan-2006
! This is the initial part of a gamess input -
! just need to add coordinates and $END record
!
! internal coordinates for (R,R)-3,5-Dimethoxyoctan-1-ol

```

```

$CONTRL RUNTYP=OPTIMIZE COORD=ZMT NZVAR=99 $END
$SYSTEM TIMLIM=100 MEMORY=1000000 $END
$BASIS GBASIS=AM1 $END
$STATPT NSTEP=90 $END
$SCF DIRSCF=.TRUE. $END
$DATA
(R,R)-3,5-Dimethoxyoctan-1-ol
C1
C1
C2      1      1.5102082
C3      2      1.5124850      1      111.4311431
C4      3      1.5287579      2      118.9382927      1      -219.9165655
C5      4      1.5305464      3      123.7498874      2      -45.2298555
C6      5      1.5228956      4      119.6723460      3      -9.4368090
C7      6      1.5276093      5      111.8959298      4      58.7567292
C8      7      1.5188556      6      112.1697587      5      -164.0583611
O9      4      1.4433845      3      108.1236223      5      120.6080480
O10     6      1.4408066      5      109.2075398      7      115.6664836
O11     8      1.4127560      7      112.8613150      6      -68.3876862
C12     9      1.4141913      4      114.0244945      5      -198.8269885
C13    10      1.4155535      4      89.4674014      7      -169.3233227
H14    11      0.9669727      4      63.6790524      7      126.6705827
H15     1      1.1168164      2      110.7718424      3      63.0747700
H16     1      1.1167317      2      110.1576941      3      -177.0417614
H17     1      1.1167238      2      110.6887088      3      -57.1236309
H18     2      1.1222568      3      110.5457925      4      95.6080986
H19     2      1.1216077      3      109.6900292      4      -23.2743797
H20     3      1.1242072      4      107.3931591      2      122.9443722
H21     3      1.1255582      4      106.5217791      2      -123.0414572
H22     4      1.1325075      3      105.6350937      5      -123.2899366
H23     5      1.1246044      6      108.1477284      4      122.7532528
H24     5      1.1274330      6      108.1889606      4      -123.0616018
H25     6      1.1228252      5      111.4836688      7      -125.1893845
H26     7      1.1220229      8      109.1969671      6      120.7609950
H27     7      1.1196435      8      109.3701797      6      -121.0552832
H28     8      1.1245114      7      110.3298807      6      42.6873846
H29     8      1.1239997      7      110.3845530      6      163.2678623
H30    12      1.1186869      9      110.6484027      4      62.8015733
H31    12      1.1191216      9      104.3884843      4      -178.4429817
$END

```

```

c
c program to do monte-carlo searches for low energy minima
c John D. Head 04-jan-06
c
c test of reading gamess gam1 output
c assumes is opt not converged - can still get useful
c energy, RMS grad and coordinate info

      implicit real*8 (a-h,o-z)
c
      parameter (natom=35)
      parameter (nzvar=3*natom-6)
      dimension geom(3,natom), icor(nzvar), zcor(nzvar)
      dimension zold(nzvar), zfull(nzvar)
      character*6 at_char(natom)
      character*5 char5(natom)
      character conf*5
      logical accept
      integer fill
      integer new

      dimension results(5000,10)
      dimension current(1,10)

      integer resint(5000,2)
      integer curint(1,2)
c
c in gamrun
c ioptyp=0 for full geometry optimization
c ioptyp=1 for partial geometry optimization
c
      nseed=1382
      iaccept=0
      iboltz=0
      nsteps=900000
      fill=0
c
c copy output to gam1234.out
c
c this is assuming gamstart.out is a constrained optimization
c and not a local min run
      ioptyp=1
c if gamstart.out a local min run then use
c ioptyp=0
      call system('/bin/cp -f gamstart.out gam1234.out')
      call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$ ioptyp,at_char,icor,zcor,char5)
      write(6,*) ' initial zcor(6) =',zcor(6)
      write(6,*) ' initial zcor(9) =',zcor(9)
      write(6,*) ' initial zcor(12) =',zcor(12)
      write(6,*) ' initial zcor(15) =',zcor(15)
      write(6,*) ' initial zcor(18) =',zcor(18)
      write(6,*) ' initial zcor(27) =',zcor(27)
      write(6,*) ' initial zcor(30) =',zcor(30)

```

```

write(6,*)' initial zcor(33) =',zcor(33)
write(6,*)' initial zcor(36) =',zcor(36)
write(6,*)' type of calculation result:',icalc_complete
write(6,*)' fitness:',fitness
write(6,*)' grad rms:',rms
eold = fitness

c
c save old coordinates and get local minimum
c
do 90 jzvar=1,nzvar
zfull(jzvar)=zcor(jzvar)
90 zold(jzvar)=zcor(jzvar)
ifull=1

c
c open file for full opt results
c
open(8,file='fulloptres.yeah',status='unknown',
1 form='formatted')
open(9,file='fulloptres.sum',status='unknown',
1 form='formatted')
write(conf,*) ifull
! open(7,file='cartcord'//conf(2:5), status='unknown',
! 1 form='formatted')

ioptyp=0
call gamrun(natom,nzvar,ioptyp,at_char,icor,zfull)
call system('/bin/mv -f gamrun.usi_out gam1234.out')
call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$ ioptyp,at_char,icor,zfull,char5)
call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)

write(9,'(a,i5,a,9f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(18),zfull(27),zfull(30),zfull(33),zfull(36),' FE:',fulle
write(8,'(a,i5,a,9f13.7,a,f13.7)')' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(18),zfull(27),zfull(30),zfull(33),zfull(36),' FE:',fulle

call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$ at_char,icor,zfull,results,current,resint,curint,new)
fulle_old=fulle
zfo6=zfull(6)
zfo9=zfull(9)
zfo12=zfull(12)
zfo15=zfull(15)
zfo18=zfull(18)
zfo27=zfull(27)
zfo30=zfull(30)
zfo33=zfull(33)
zfo36=zfull(36)

```



```

c
c start looping over lots of geometries
c
do 100 i=1,nsteps
del16=50.*(ezrandom(nseed)-0.5)
del19=50.*(ezrandom(nseed)-0.5)
del112=50.*(ezrandom(nseed)-0.5)
del115=50.*(ezrandom(nseed)-0.5)
del118=50.*(ezrandom(nseed)-0.5)
del127=50.*(ezrandom(nseed)-0.5)
del130=50.*(ezrandom(nseed)-0.5)
del133=50.*(ezrandom(nseed)-0.5)
del136=50.*(ezrandom(nseed)-0.5)
write(6,*) 'delta dihedral'
write(6,*) del16, del19, del112, del115, del118, del127, del130, del133,
del136
zcor(6)=zcor(6)+del16
if(zcor(6).gt.360.)then
zcor(6)=zcor(6)-360.
else if(zcor(6).gt.180.)then
zcor(6)=zcor(6)+360.
endif
zcor(9)=zcor(9)+del19
if(zcor(9).gt.360.)then
zcor(9)=zcor(9)-360.
else if(zcor(9).gt.180.)then
zcor(9)=zcor(9)+360.
endif
zcor(12)=zcor(12)+del112
if(zcor(12).gt.360.)then
zcor(12)=zcor(12)+360.
else if(zcor(12).gt.180.)then
zcor(12)=zcor(12)-360.
endif
zcor(15)=zcor(15)+del115
if(zcor(15).gt.360.)then
zcor(15)=zcor(15)-360.
else if(zcor(15).gt.180.)then
zcor(15)=zcor(15)+360.
endif
zcor(18)=zcor(18)+del118
if(zcor(18).gt.360.)then
zcor(18)=zcor(18)-360.
else if(zcor(18).gt.180.)then
zcor(18)=zcor(18)+360.
endif
zcor(27)=zcor(27)+del127
if(zcor(27).gt.360.)then
zcor(27)=zcor(27)-360.
else if(zcor(27).gt.180.)then
zcor(27)=zcor(27)+360.
endif
zcor(30)=zcor(30)+del130
if(zcor(30).gt.360.)then

```

```

        zcor(30)=zcor(30)-360.
    else if(zcor(30).gt.180.) then
        zcor(30)=zcor(30)+360.
    endif
    zcor(33)=zcor(33)+del133
    if(zcor(33).gt.360.) then
        zcor(33)=zcor(33)-360.
    else if(zcor(33).gt.180.) then
        zcor(33)=zcor(33)+360.
    endif
    zcor(36)=zcor(36)+del136
    if(zcor(36).gt.360.) then
        zcor(36)=zcor(36)-360.
    else if(zcor(36).gt.180.) then
        zcor(36)=zcor(36)+360.
    endif
    write(6,'(a,i5,a,9f13.7)') ' GAMESS run number:',i,' dihedrals:',
1
zcor(6),zcor(9),zcor(12),zcor(15),zcor(18),zcor(27),zcor(30),zcor(33),z
cor(36)
    ioptyp=1
    call gamrun(natom,nzvar,ioptyp,at_char,icor,zcor)
    call system('/bin/mv -f gamrun.usi_out gam1234.out')
    call rdfil(natom,nzvar,geom,fitness,rms,icalc_complete,
$      ioptyp,at_char,icor,zcor,char5)
    deltae=fitness-eold
    write(6,*)' energy change =',deltae
c
c now see if energy is going down
c
    temp = 0.0012
    if(deltae.lt.0.)then
        accept=.true.
c
    ioptyp=0
    do 35 jzvar=1,nzvar
35  zfull(jzvar)=zcor(jzvar)
        ifull=ifull+1

    write(6,*)' PERFORMING FULL GEOM OPT',ifull

    call gamrun(natom,nzvar,ioptyp,at_char,icor,zfull)
    call system('/bin/mv -f gamrun.usi_out gam1234.out')
    call rdfil(natom,nzvar,geom,fulle,frms,icalc_complete,
$      ioptyp,at_char,icor,zfull,char5)

    write(8,'(a,i5,a,9f13.7,a,f13.7)') ' FULL LOCAL OPT:',ifull,
1 ' dihedrals:',zfull(6),zfull(9),zfull(12),zfull(15),
1 zfull(18),zfull(27),zfull(30),zfull(33),zfull(36),' FE:',fulle

    new=0

    call sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$      at_char,icor,zfull,results,current,resint,curint,new)

```

```

        write(6,*) 'new = ', new, ' 0 --> skipping, 1 --> adding, 2 -->
matrix complete'

        if (new.eq.1) then
        call
filegen(icalc_complete,fulle,ifull,at_char,icor,zcor,natom,nzvar)
        call system('/bin/mv -f cartcord* carts')
        else
        end if

        if (new.eq.2) then
        go to 101
        end if

        fulle_old=fulle
        zfo6=zfull(6)
        zfo9=zfull(9)
        zfo12=zfull(12)
        zfo15=zfull(15)
        zfo18=zfull(18)
        zfo27=zfull(27)
        zfo30=zfull(30)
        zfo33=zfull(33)
        zfo36=zfull(36)

c
        else if (exp(-deltae/temp).lt.ezrandom(nseed)) then
        write(6,*) ' accepted increase in energy'
        accept=.true.
        iboltz=iboltz+1
        else
        accept=.false.
        end if
        eold=fitness
        if (accept) then
        iaccept=iaccept+1
        do 120 jzvar=1,nzvar
120      zold(jzvar)=zcor(jzvar)
        else
        do 130 jzvar=1,nzvar
130      zcor(jzvar)=zold(jzvar)
        endif

100      continue
101      write(6,*) ' GAMESS runs completed'
        write(6,*) ' Out of ',i,' geoms: no boltzman accepts:',iboltz
        write(6,*) ' Number of negative de:',iaccept-iboltz
        write(6,*) ' *****'
        stop
        end

```

```

subroutine sort(fill,ifull,natom,nzvar,fulle,icalc_complete,
$   at_char,icor,zfull,results,current,resint,curint,new)

implicit real*8 (a-h,o-z)

parameter (Ethresh=0.00005)
parameter (Athresh=10.0)
parameter (Athresh2=290.0)
parameter (maxsize=10)

dimension icor(nzvar),zfull(nzvar)
character*6 at_char(natom)

dimension results(5000,maxsize)
dimension current(1,maxsize)

dimension temp1(5000,maxsize)
integer temp2(5000,2)

integer resint(5000,2)
integer curint(1,2)

real Echeck
real Acheck
integer fill
integer new

C   ncon is 360/120 ^(# DH), where 120 is the increment
ncon=4999

open (unit=7,
file='unsorted_result',status='unknown',form='formatted')

curint(1,1)=ifull
current(1,1)=fulle
current(1,2)=zfull(6)
current(1,3)=zfull(9)
current(1,4)=zfull(12)
current(1,5)=zfull(15)
current(1,6)=zfull(18)
current(1,7)=zfull(27)
current(1,8)=zfull(30)
current(1,9)=zfull(33)
current(1,10)=zfull(36)
curint(1,2)=icalc_complete
83 format(i5,i5,f12.7,9f8.2)
84 format(i5,i5,i5,f12.7,9f8.2)

C   FILL IN THE FIRST DATA POINT

if (fill.eq.0) then
c   i=fill
do 10 j=1,maxsize

```

```

        results(1,j) = current(1,j)
10    continue
        do 11 j=1,2
            resint(1,j) = curint(1,j)
11    continue
        fill=fill+1
        go to 50
        else
        end if

C    COMPARE INCOMING DATA TO THE EXSISTING ARRAY

        if (current(1,1).eq.0.) then
            go to 50
        end if
        if (curint(1,2).ne.1.) then
            write(7,*) curint(1,1), ' has exit code of: ',curint(1,2), '
- discarding'
            write(6,*) curint(1,1), ' has exit code of: ',curint(1,2), '
- discarding'
            go to 50
        end if

        ord=1

60    do 20 i=ord,fill

        Echeck = abs(results(i,1)-current(1,1))

        if (Echeck.le.Ethresh) then
            write(7,*) curint(1,1), ' and ',resint(i,1),' have Ethresh
of ', Echeck,' - checking DH'
            write(6,*) curint(1,1), ' and ',resint(i,1),' have Ethresh
of ', Echeck,' - checking DH'
C    CHECK THE I-TH DIHEDRALS
            chk=i
            do 30 j=2,maxsize

                Acheck=abs(results(chk,j)-current(1,j))

                if
                (((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.eq.maxsize)) then
                    write(7,*) curint(1,1), ' is a match to ',
resint(chk,1),' - discarding'
                    write(6,*) curint(1,1), ' is a match to ',
resint(chk,1),' - discarding'
                    go to 50
                else if
                (((Acheck.le.Athresh).or.(Acheck.ge.Athresh2)).and.(j.le.maxsize-1))
then
                    go to 30
                else
                    if (ord.lt.fill) then

```

```

                                ord=chk+1
                                go to 60
                                else
(Athresh)' write(7,*) curint(1,1), ' is a new conf - adding
(Athresh)' write(6,*) curint(1,1), ' is a new conf - adding
                                new=1
                                go to 31
                                end if
31      fill=fill+1
        do 12 l=1,maxsize
        results(fill,1) = current(1,1)
12     continue
        do 13 l=1,2
        resint(fill,1) = curint(1,1)
13     continue
        end if
30     continue
        else
            if (ord.eq.fill) then
                new=1
                write(7,*) ifull,' is a new conf - adding (Ethresh)'
                write(6,*) ifull,' is a new conf - adding (Ethresh)'
                fill=fill+1
                chk=i
                do 14 l=1,maxsize
                results(fill,1) = current(1,1)
14     continue
                do 15 l=1,2
                resint(fill,1) = curint(1,1)
15     continue
                else
                    chk=i
                    ord=chk+1
                    go to 60
                end if
            end if
20     continue
50     continue

        if (fill.ge.ncon) then
            open (unit=9, file='result300', status='unknown', form='formatted')
            write(9,*) 'results'
            write(9,83) ((resint(i,1),l=1,2), (results(i,n),n=1,maxsize),
i=1,fill)
                is=1
                ((resint(k,1),l=1,2), (results(k,n),n=1,3),k=1,ncon)
                do 401 j=1,fill
                do 400 k=1,fill
                    if (results(k,1).gt.results(k+1,1)) then
                        do n=1,maxsize
                            templ{k,n}=results(k,n)

```

```

        results(k,n)=results(k+1,n)
        results(k+1,n)=temp1(k,n)
    enddo
do n=1,2
    temp2(k,n)=resint(k,n)
    resint(k,n)=resint(k+1,n)
    resint(k+1,n)=temp2(k,n)
enddo
else
end if
400 continue
401 continue
    write(9,*) 'sorted results'
    write(9,84) (k,(resint(k,l),l=1,2),(results(k,n),n=1,maxsize
),k=1,fill)
endif
return
end

```

```

!
! 01-jan-2006
! This is the initial part of a gamess input -
! just need to add coordinates and $END record
!
! internal coordinates for SYN-fatty acid (R,R,R,R,R,R,R)

```

```

$CONTRL RUNTYP=OPTIMIZE COORD=ZMT NZVAR=201 $END
$SYSTEM TIMLIM=100 MEMORY=1000000 $END
$BASIS GBASIS=AM1 $END
$STATPT NSTEP=90 $END
$SCF DIRSCF=.TRUE. $END

```

```
$DATA
```

```
SYN-fatty acid (R,R,R,R,R,R,R)
```

```

C1
C2      1      1.5073853
C3      2      1.5134933      1      112.7078553
C4      3      1.5255036      2      112.8719965      1      77.7237642      0
C5      4      1.5264733      3      115.3438637      2      68.0881924      0
C6      5      1.5201525      4      117.4886195      3      36.2614845      0
C7      6      1.5214950      5      114.8848376      4      61.4666391      0
C8      7      1.5273334      6      114.7760815      5      60.7687693      0
C9      8      1.5268007      7      115.7441776      6      111.0490083      0
C10     9      1.5207500      8      119.0710204      7      37.9163165      0
C11    10      1.5200345      9      114.9924514      8      50.5190006      0
C12    11      1.5267255     10      114.7196178      9      63.7952419      0
C13    12      1.5258175     11      116.0236228     10      99.7205135      0
C14    13      1.5208565     12      118.9477340     11      38.8298479      0
C15    14      1.5190364     13      115.0008326     12      51.8596674      0
C16    15      1.5270504     14      114.9023802     13      61.1059623      0
C17    16      1.5295781     15      112.3125014     14      99.7709425      0
C18    17      1.4981761     16      111.8024419     15      73.7384210      0
O19     4      1.4188134      3      109.9774327      5      127.9007991      0
H20     4      1.1328880      3      108.1514080      5     -120.4762787      0
H21    19      0.9667725      4      108.4655431     20      162.4267305      0
O22     8      1.4205388      7      110.2155172      9      127.9031879      0
H23     8      1.1333084      7      109.2992097      9     -120.5611902      0
H24    22      0.9663348      8      108.6888140     23      170.4166705      0
O25    12      1.4215444     11      110.0818649     13      127.8216559      0
H26    12      1.1341129     11      108.9077257     13     -121.0204891      0
H27    25      0.9653491     12      108.6290926     26      172.9443415      0
O28    16      1.4218070     15      111.1137661     17      123.9166632      0
H29    16      1.1294759     15      110.7010293     17     -121.6554131      0
H30    28      0.9672833     16      108.0810662     29     -165.6089996      0
O31    18      1.3667903     17      118.8819815     16      70.3691059      0
H32    31      0.9701759     18      109.9908113     17      2.5534382      0
O33    18      1.2303759     17      127.1963165     31     -179.0492449      0
C34     6      1.5175150      5      109.4511127      7      125.0769692      0
H35     6      1.1305615      5      107.4995124      7     -118.4626334      0
H36    34      1.1165350      6      110.3875216     35     -58.2311531      0
H37    34      1.1166715      6      110.2612248     35      61.6826044      0
H38    34      1.1169039      6      110.7913258     35     -178.3456396      0
C39    10      1.5177071      9      109.1039315     11      125.4817532      0

```



H40	10	1.1304286	9	107.9128335	11	-118.5840927	0
H41	39	1.1165161	10	110.2577988	40	-55.7535957	0
H42	39	1.1168027	10	110.2559675	40	64.2543400	0
H43	39	1.1170215	10	110.8211017	40	-175.7133378	0
C44	14	1.5175765	13	109.0639366	15	125.5606589	0
H45	14	1.1309054	13	107.8367450	15	-118.4472210	0
H46	44	1.1168107	14	110.1868179	45	-57.9756106	0
H47	44	1.1167870	14	110.3709624	45	61.8613010	0
H48	44	1.1170118	14	110.7690381	45	-177.7539054	0
H49	1	1.1169449	2	110.6710496	3	59.2854771	0
H50	1	1.1168520	2	110.2347466	3	179.2179277	0
H51	1	1.1168884	2	110.9143413	3	-60.7867373	0
H52	2	1.1229516	3	108.7157864	4	-161.3647800	0
H53	2	1.1224838	3	109.6067277	4	-44.6211081	0
H54	3	1.1231455	4	107.7662415	2	121.2913732	0
H55	3	1.1206480	4	109.9541558	2	-122.6287295	0
H56	5	1.1242743	6	107.3089975	4	121.0492418	0
H57	5	1.1232186	6	109.6236665	4	-123.0469960	0
H58	7	1.1244422	8	108.2491480	6	121.7906304	0
H59	7	1.1210955	8	110.4312354	6	-121.6730035	0
H60	9	1.1248651	10	106.9061305	8	121.3231306	0
H61	9	1.1242467	10	109.1503447	8	-123.0982038	0
H62	11	1.1253911	12	108.2233956	10	121.2541842	0
H63	11	1.1210173	12	110.4801350	10	-122.1999775	0
H64	13	1.1250692	14	106.8383704	12	121.2746178	0
H65	13	1.1240894	14	109.2529401	12	-123.2262097	0
H66	15	1.1250455	16	108.6113639	14	121.6002517	0
H67	15	1.1242058	16	109.0348289	14	-122.1122590	0
H68	17	1.1220689	18	109.1627619	16	121.0690385	0
H69	17	1.1221061	18	108.8219289	16	-120.7358187	0

§END

```

!
! 01-jan-2006
! This is the initial part of a gamess input -
! just need to add coordinates and $END record
!
! internal coordinates for ANTI-fatty acid (R,S,R,S,R,S,R)

```

```

$CONTRL RUNTYP=OPTIMIZE COORD=ZMT NZVAR=201 $END
$$SYSTEM TIMLIM=100 MEMORY=1000000 $END
$BASIS GBASIS=AM1 $END
$STATPT NSTEP=90 $END
$SCF DIRSCF=.TRUE. $END

```

```
$DATA
```

```
ANTI-fatty acid (R,S,R,S,R,S,R)
```

```

C1
C2      1      1.5069217
C3      2      1.5131260      1      111.3775942
C4      3      1.5227779      2      111.8648099      1      -177.3243230      0
C5      4      1.5273086      3      116.3231679      2      70.5591717      0
C6      5      1.5218732      4      121.5744891      3      47.8704667      0
C7      6      1.5178205      5      115.2993143      4      42.7724435      0
C8      7      1.5266023      6      112.7442567      5      65.7060566      0
C9      8      1.5249952      7      115.4974793      6      75.2992890      0
C10     9      1.5200853      8      117.9213535      7      37.0767442      0
C11    10     1.5213548      9      114.9294100      8      59.1771766      0
C12    11     1.5270224     10     114.7349076      9      61.3500538      0
C13    12     1.5263031     11     115.7517714     10     107.6842174      0
C14    13     1.5210123     12     119.1117802     11     39.0664562      0
C15    14     1.5187831     13     114.9827980     12     50.4365256      0
C16    15     1.5270422     14     114.8749915     13     61.7373310      0
C17    16     1.5294343     15     112.3497018     14     99.4619288      0
C18    17     1.4981950     16     111.8193586     15     73.8898455      0
O19     4     1.4189325      3     110.8567083      5     128.9873037      0
H20     4     1.1346111      3     108.0435809      5     -119.6225214      0
H21    19     0.9656950      4     108.4778194     20     -174.5653523      0
O22     8     1.4192060      7     109.5043387      9     128.1892529      0
H23     8     1.1351277      7     108.3807981      9     -120.4756752      0
H24    22     0.9664550      8     108.5971648     23     161.6937696      0
O25    12     1.4212511     11     110.2943990     13     127.8853503      0
H26    12     1.1334977     11     109.2992557     13     -120.6936134      0
H27    25     0.9653857     12     108.6355262     26     173.3004548      0
O28    16     1.4219659     15     111.1274592     17     123.8925133      0
H29    16     1.1295922     15     110.6217296     17     -121.6865419      0
H30    28     0.9672825     16     108.0786677     29     -165.1689244      0
O31    18     1.3667948     17     118.8765734     16     70.1271946      0
H32    31     0.9701763     18     109.9933050     17     2.6277490      0
O33    18     1.2303370     17     127.2045016     31     -179.0880262      0
C34     6     1.5165256      5     112.1154794      7     -127.0160542      0
H35     6     1.1334075      5     105.5278512      7     117.3799649      0
H36    34     1.1173207      6     110.0352596     35     -55.2310938      0
H37    34     1.1162722      6     111.4576498     35     -175.3350793      0
H38    34     1.1170302      6     110.3660738     35     64.2879411      0
C39    10     1.5174704      9     109.3720949     11     125.1331706      0

```

H40	10	1.1306648	9	107.5380770	11	-118.5061202	0
H41	39	1.1165227	10	110.3716528	40	-58.1560844	0
H42	39	1.1170112	10	110.7700344	40	-178.2218817	0
H43	39	1.1166531	10	110.2701784	40	61.7913701	0
C44	14	1.5176075	13	109.0513567	15	125.6660080	0
H45	14	1.1307350	13	107.9237534	15	-118.4262874	0
H46	44	1.1167908	14	110.1641417	45	-57.7502933	0
H47	44	1.1168220	14	110.3569556	45	62.0572555	0
H48	44	1.1169719	14	110.8016953	45	-177.5448265	0
H49	1	1.1171771	2	110.5675815	3	61.4781835	0
H50	1	1.1166538	2	110.3405277	3	-178.5795124	0
H51	1	1.1168615	2	110.7117673	3	-58.5297882	0
H52	2	1.1232846	3	109.4481252	4	-55.9971723	0
H53	2	1.1221733	3	109.4763826	4	61.2185874	0
H54	3	1.1234690	4	107.9768725	2	120.4665489	0
H55	3	1.1200055	4	110.5223365	2	-122.9776074	0
H56	5	1.1245165	6	107.3080582	4	123.2017877	0
H57	5	1.1260853	6	107.5032688	4	-121.8854742	0
H58	7	1.1223951	8	108.2658665	6	122.3661615	0
H59	7	1.1215118	8	110.1733679	6	-121.9827651	0
H60	9	1.1246361	10	107.1599359	8	120.9904788	0
H61	9	1.1234176	10	109.5955576	8	-123.2325136	0
H62	11	1.1246083	12	108.2314359	10	121.6883412	0
H63	11	1.1211282	12	110.4397841	10	-121.7846961	0
H64	13	1.1250128	14	106.8746537	12	121.2747802	0
H65	13	1.1242249	14	109.1571773	12	-123.2010376	0
H66	15	1.1252962	16	108.6732202	14	121.4247429	0
H67	15	1.1240125	16	109.0458103	14	-122.1817599	0
H68	17	1.1220092	18	109.1779291	16	121.1232728	0
H69	17	1.1221716	18	108.7875629	16	-120.7011922	0

SEND

**MISSING  
PAGE  
NO.**

121, 122

**AT THE TIME OF  
MICROFILMING**

```

        read(iw2_local,*)
        read(iw2_local,*)
        read(iw2_local,*)

!       do i=1,natom
!           read(iw2_local,*)char5_temp, float_temp,
!           . geom(1,i), geom(2,i), geom(3,i)
!       enddo

        do i=1,natom
            read(iw2_local,'(a)')line
        enddo

cjdhd 12/12/05
c  add in read of internal coordinates
c  these appear to be directly after the cartesian coordinates
cjdhd 12/12/05
c
c  find final zmatrix
c
        do j=1,190
            read(iw2_local,'(a42)',end=3223)line
            if{line(1:42).eq.
$               ' THE CURRENT FULLY SUBSTITUTED Z-MATRIX IS')}then
c
c  read in zmatrix
c
            write(iw0,*)' z-matrix coordinates'
1223          format(a6,i5,f12.7,i5,f14.7,i5,f14.7)
            read(iw2_local,1223)at_char(1)
            read(iw2_local,1223)at_char(2),icor(1),zcor(1)
            read(iw2_local,1223)at_char(3),icor(2),zcor(2),
$              icor(3),zcor(3)
            ii=3
            do i=4,natom
                read(iw2_local,1223)at_char(i),
$              (icor(ii+iii),zcor(ii+iii),iii=1,3)
                ii=ii+3
            enddo
c
c  write out internal coords
c
            write(iw0,1223)at_char(1)
            write(iw0,1223)at_char(2),icor(1),zcor(1)
            write(iw0,1223)at_char(3),icor(2),zcor(2),
$              icor(3),zcor(3)
            ii=3
            do i=4,natom
                write(iw0,1223)at_char(i),
$              (icor(ii+iii),zcor(ii+iii),iii=1,3)
                ii=ii+3
            enddo

```

**MISSING  
PAGE  
NO.**

124-131

**AT THE TIME OF  
MICROFILMING**

19. L. Ferrighi, D. Marchesan, K. Ruud, *J. Chem. Phys.* **2005**, *123*, 204104.
20. K. B. Wiberg, Y. Wang, P. H. Vaccaro, J. R. Cheeseman, G. Trucks, M. J. Frisch *J. Phys. Chem. A* **2004**, *108*, 32.
21. R. K. Kondru, P. Wipf, and D. N. Beratan *J. Am. Chem. Soc.* **1998**, *120*, 2204.
22. G. Zuber, M. Goldsmith, T. D. Hopkins, D. N. Beratan, P. Wipf *Org. Lett.* **2005**, *7*, 5269.
23. A. D. Becke *Phys. Rev. A* **1988**, *38*, 3098.
24. J. P. Perdew *Phys. Rev. B* **1986**, *33*, 8822.
25. A. Schäfer, H. Horn, R. Ahlrichs. *J. Chem. Phys.* **1992**, *97*, 2571.
26. J. H. van't Hoff, *Die Lagerung der Atome in Raume*; Vieweg: Braunschweig, Germany, **1908**.
27. N. Weinberg, S. Wolfe *J. Am. Chem. Soc.* **1994**, *116*, 9680.
28. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart *J. Am. Chem. Soc.* **1985**, *107*, 3902.
29. J. J. P. Stewart *J. Comp. Chem.* **1989**, *10*, 209.
30. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.* **1993**, *14*, 1347.
31. 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, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.
32. R. K. Kondru, S. Lim, P. Wipf, D. N. Beratan *Chirality* 1997, 9, 469.
33. N. Metropolis, A. W. Rosenbluth, M. N. Rosebluth, A. M. Teller, E. Teller *J. Chem. Phys.* 1953, 21, 1087.
34. D. P. Landau, K. Binder, *A Guide to do monte Carlo Simulations in Statistical Physics;* Cambridge Univ. Press: UK, 2000.