IOWA STATE UNIVERSITY
Digital Repository

2012

Cellulose and cellobiose: adventures of a wandering organic chemist in theoretical chemistry

John Ysrael Baluyut

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Physical Chemistry Commons

Cellulose and cellobiose: adventures of a wandering organic chemist in theoretical chemistry

by

John Ysrael Guevarra Baluyut

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Physical Chemistry

Program of Study Committee:

Theresa L. Windus, Major Professor
Thomas A. Holme
Monica H. Lamm

Iowa State University

Ames, Iowa

2012

Copyright © John Ysrael Guevarra Baluyut, 2012. All rights reserved.

TABLE OF CONTENTS

H_3O^+ Complex With B3LYP/6-311+G(d,p)

311+G(d,p)

ACKNOWLEDGEMENT

I owe a deep debt of gratitude to my advisor, Professor Theresa L. Windus, for all the support she has given me in completing this thesis, particularly for sharing with me her vast pool of knowledge in the field of theoretical chemistry. Thanks are also due to Dr. Yana A. Kholod and Dr. Ajitha Devarajan, who have both patiently answered far too many of my questions. I also wish to thank friends from back home in Manila, Philippines like Larry Malijan and Lani Ramos-Mercado, who have always been with me in spirit, if not physically, for constantly and continuously reminding me that we are never too old to pursue our dreams. Finally, thanks to the many friends I have made so far at Iowa State University, who made graduate life the second time around a bit more bearable.

Maraming Salamat Po!

CHAPTER 1

CELLULOSE AND CELLOBIOSE: THE ADVENTURE BEGINS

Introduction

Cellulose is the principal component of biomass, which is generally viewed as a source of renewable fuel.[1] The Energy Information Administration (EIA) has predicted the use of renewable energy sources, such as ethanol, to increase by an average of 3.3% per year.[2] The EIA expects about 40% of the ethanol production in the US will come from cellulosic biomass by 2030. Cellulosic biomass is the most abundant organic compound in nature, but the biggest challenge in ethanol production from biomass is the high cost of the several processing steps.[3]

Cellulose hydrolysis into glucose is an important step for biomass conversion. Glucose is a renewable feedstock molecule, which may be processed into fuels, chemical commodities, food, and medicine.[4] Cellulose breakdown is often accomplished using either enzymatic or acid-catalyzed processes. The former has had some drawbacks such as low thermal stability of the enzymes and difficulties in the separation and recovery of the products from the enzymes. On the other hand, concentrated acid-catalyzed hydrolysis of cellulose has been industrialized for over a century with glucose yields of about 70%, but the chemistry behind this process is poorly understood.[5]

Naturally-occuring enzymes, known as cellulases, break cellulose down into the constituent glucose and cellobiose units. However, enzyme-catalyzed processes come with very slow rates and significant production costs for the enzymes. This is primarily due to a property of plant cell walls known as biomass recalcitrance, which renders cellulosic materials resistant to microbial and enzymatic destruction.[6,7,8]

Knowledge of the structure of native cellulose fibers has important implications in understanding the chemical or enzymatic transformation of cellulosic biomass to valuable chemicals, for the development of new materials, and understanding the physical properties of a wide range of cellulose-based materials. A considerable part of cellulose research has been devoted to cellulosic structure elucidation. Cellulose is synthesized in plants at the cell membrane by an ordered synthase complex[9] by polymerization of D-glucose residues using β(1→4)-glycosidic bonds. Extended parallel polymer chains are then assembled into nanometer thick crystalline microfibrils.[10] These microfibrils have high tensile strength and make up the fundamental structural unit in plant cell walls. A matrix of other polysaccharides cross-link the microfibrils using hydrogen bonding. Use of high-resolution microscopy of a continuous crystal lattice along individual microfibrils, first from highly crystalline algal cellulose[11] and then from ramie cellulose,[12] provided evidence for the presence of single crystal microfibrils.

Most crystalline cellulose, also known as cellulose I, is found naturally in two forms: cellulose Iα and Iβ. Both forms of crystalline cellulose have the same conformation of their heavy atom backbone in the chains, but each possesses a different hydrogen bonding pattern that give different packing between the chains.[13] X-ray (synchrotron source) and neutron diffraction experiments have been used to determine the conformation and packing arrangement in both polymorphs. Cellulose Iα has a triclinic unit cell which contains one chain, belongs to space group $P1$ and has unit cell dimensions: a = 6.717 Å, b= 5.962 Å, c = 10.400 Å, α = 118.08^o, β = 114.80^o, and γ = 80.37^o.[14] The numbering scheme generally used to identify carbon and oxygen atoms in the glucose units of cellulose, as well as the directionality of a cellulose chain with respect to the crystallographic c axis is shown in

Figure 1.[15] All of the axes are chosen to satisfy the right-hand coordinate system, with the upper direction being positive for *z* coordinates. This sets up the unique axis to *c*. In this sense, the cellulose Iα chain is said to be "parallel up" because the *z* coordinate of O5 is greater than that of C5. In this case, the reducing end of the chain, C1, is oriented along the positive side with respect to the *c* axis. Glycosidic linkages between glucose units are all identical to each other, and all of the hydroxymethyl groups are found in the *tg* conformation, where *t* and *g* describe the relative geometries of the –O6H group in the hydroxymethyl group as *trans* and *gauche*, respectively, with respect to the C4 and O5 in the same glucose residue. Finally, adjacent sugar rings are *anti* to each other so that hydroxymethyl groups, for example, are found on opposite sides of the plane defined by the atoms in the glycosidic bond. This leads to the chain having cellobiose as the repeating unit. Atoms in the nonreducing end are generally distinguished from their counterparts in the reducing end with the use of primes, so that the glycosidic bond consists of C4, O4, and C1'.
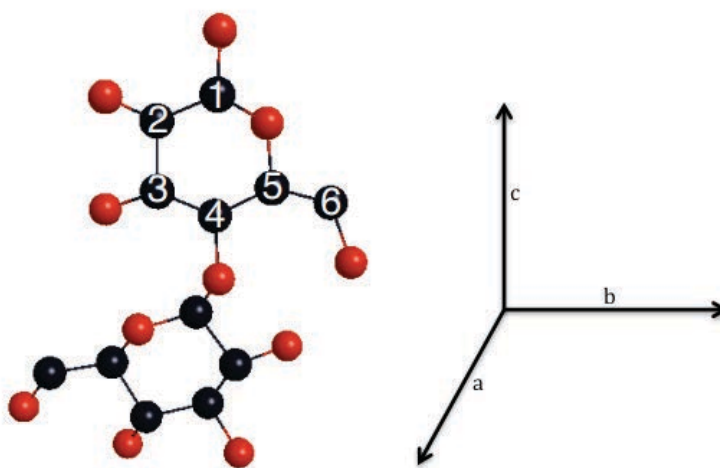
Figure 1. The definition of the directionality with respect to the crystallographic axes as well as the numbering scheme of the cellobiose repeating unit in cellulose. The atoms in the nonreducing unit (unnumbered in this figure) are distinguished from their counterparts in the reducing unit by the use of primes.

On the other hand, cellulose Iβ has a monoclinic $P2_1$ unit cell made up of two different chains and has unit cell dimensions: a = 7.784 Å, b= 8.201 Å, c = 10.380 Å, α = 90°, β = 90°, and γ = 117.1°.[16] As with cellulose Iα, sheets formed by the chains are also in a "parallel-up" fashion, with all hydroxymethyl groups adopting the *tg* conformation.

Both cellulose Iα and Iβ are characterized by having two mutually exclusive networks of hydrogen bonding consisting of O3-H···O5 intrachain, O2-H···O6 intrachain, and O6-H···O3 interchain hydrogen bonds. Here and throughout the thesis, interchain refers to chains in the same sheet; intersheet will be used for interactions between sheets. The chains of glucose residues in each of cellulose Iα and Iβ are linked together by two networks of hydrogen-bonding interactions. Network I of cellulose Iα has two types of O3-H···O5, differentiated from each other in terms of the H···O5 distance and the angle formed by the atoms in the hydrogen bond. These two types of hydrogen bonds alternate within the same chain in cellulose Iα. On the other hand, network I of cellulose Iβ also has two types of O3-H···O5, but these alternate between chains. The intrachain O2-H···O6 bond distances in network I of cellulose Iα are shorter than those present in the corresponding network in cellulose Iβ, while O6-H···O3 interchain hydrogen bonds are longer in cellulose Iα than in cellulose Iβ. In network II, there is one type of intrachain O6-H···O2 bond in Iα while there are two in Iβ, but there are two types of interchain O2-H···O6 bonds in Iα and only one type in Iβ. Network II bond distances are shorter while bond angles are larger in cellulose Iβ than they are in Iα. These subtle differences between cellulose Iα and Iβ still result in densities that are similar to each other for these two cellulose moieties. In either case, weak van der Waals interactions are believed to hold sheets together.[3, 16]

The glucose residues in cellulose have fairly rigid pyranose rings. The torsion angles around the glycosidic bond have some rotational freedom, but like most polysaccharides they are limited to a small region of conformational space owing to steric hindrance.[17] Given the unit-cell parameters, and thus the polymer repeat distance, the values that the glycosidic torsion angles can take is very limited, resulting in a ribbon-shaped flat molecule. High density and tight packing in cellulose crystals also limits the positional and rotational freedoms of the chains with respect to the unit cell to two choices of orientation, pointing in opposite directions, and the possibility of sliding between sheets along the chain axis. Thus, aside from the hydroxyl H atoms, the only rotation torsion angle with a large amount of freedom is around the C5—C6 bond of an exocyclic hydroxymethyl group in each residue.

The positions of H atoms were identified using high-resolution X-ray and neutron fiber diffraction to locate the positions of H atoms involved in hydrogen bonding.[18] For cellulose Iα and Iβ, in addition to the classical interchain hydrogen-bonding schemes O2H⋯O6 and O6H⋯O3, each of these hydrogen bonds can undergo proton exchange so that the hydrogen on O6 is instead found on O2 or O3. This suggests some degree of disorder in the hydrogen bonding network in cellulose I and that individual hydrogen bonding may not be as important a factor in the stabilization of cellulose I as was previously thought.

Cellobiose is the disaccharide that has the same glycosidic linkage as cellulose (basically a subsystem of cellulose) and has been used extensively as a model for cellulose structure.[19, 20, 21, 22] Conformations of cellobiose that have low energies give insight about low-energy conformations of cellulose, although the immediate environment of the cellulose chains play an important role in influencing their shape. In particular, low-energy conformations at the glycosidic linkage of cellobiose indicate regions of the ϕ and φ space allowed for cellulose

chains, where ϕ is the torsion angle in cellobiose formed by atoms H1, C1, O1, and C4', and φ is formed by atoms C1, O1, C4', H4' in the glycosidic linkage. Cellobiose is formed so that the two glucose residues have the same conformation as shown in Figure 1. This conformation results in two hydrogen bonds possibly forming across the glycosidic linkage: the O3'-H···O5 hydrogen bond, which is always present, and the O6'-H···O2 hydrogen bond, which forms only when the hydroxymethyl group on the reducing residue is either *tg* or *gg*. Crystalline cellobiose has a monoclinic unit cell with $P2_1$ symmetry and contains two molecules. The unit cell parameters for cellobiose crystal are: a = 10.972 Å, b = 13.048 Å, c = 5.091 Å, and γ = 90.83°.[23]

The hydrolysis of cellobiose has been extensively studied over the past several years as a model system for cellulose breakdown to investigate glycosidic bond cleavage. Under ordinary conditions, aqueous hydrolysis of cellobiose gives low rates of product formation but acids speed up the process.[24] The reaction proceeds with pseudofirst-order kinetics and an activation energy of about 30-35 kcal/mol. Hydrolysis has been observed to give as much as 70% glucose yields when oxidizing conditions are combined with temperatures of about 150 $^\circ$C and 100 psi pressure.[25]

Detailed investigations of the conformations of cellobiose[26] and cellobiose-water complex[27,28] have been reported. Cellobiose prefers to be in the *anti* conformation (shown in Figure 1) in the gas phase, but is found to be in the *syn* conformation in cellulose chains, crystalline cellobiose, and when cellobiose is dissolved in water.

Car-Parrinello molecular dynamics coupled with metadynamics simulation has been used to determine the energy of activation of acid-catalyzed hydrolysis of cellobiose in water as well as the reaction's overall free energy.[29] Two low-energy paths for the reaction have

been identified. The first one involves protonation at the glycosidic oxygen followed by dissociation, while the second pathway corresponds to a concerted mechanism where protonation and dissociation occur simultaneously.

The Schrodinger Equation

The main interest of quantum chemists is to find solutions to the time-independent Schrodinger equation

$$H|\Phi\rangle = E|\Phi\rangle \tag{1}$$

where H is the Hamiltonian operator, $|\Phi\rangle$ is the wavefunction that describes properties of the system of interest, and E is the system's energy. For a molecule consisting of M nuclei and N electrons, the full Hamiltonian in atomic units is given by

$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{A=1}^{M}\frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{2}$$

where R_{AB} is the distance between nuclei A and B, r_{iA} is the distance between electron i and nucleus A, M_A is the mass of nucleus A relative to that of an electron, and Z_A is the electronic charge on nucleus A. The operators ∇_i and ∇_A pertain to second order differentiation with respect to the coordinates of the ith electron and the Ath nucleus, respectively. In equation (2), the first term yields the kinetic energies of the electrons, the second term determines the kinetic energies of the nuclei, the third term gives the potential energy attraction between nucleus A and electron i, the fourth term represents potential energy repulsion between electrons i and j, and the last term is the potential energy repulsion between nuclei A and B. This Hamiltonian applies to all matter from macromolecules such as carbohydrates to synthetic materials such as plastics. Thus, solving for the eigenvalues and eigenfunctions of this Hamiltonian allows prediction of any measurable property of a given system.

The Born-Oppenheimer Approximation[30]

The first simplification of the solution of the molecular Schrodinger equation uses the fact that nuclei are several times heavier than electrons. This allows treatment of molecular systems as being made up of electrons moving in a field of fixed nuclei. Thus, the kinetic energy of the nuclei can, to a good approximation, be neglected, while the repulsive forces among the nuclei can be treated as a constant (and therefore, has no effect on the eigenfunctions of the Hamiltonian operator). These lead to a simplified Hamiltonian

$$H_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \tag{3}$$

known as the electronic Hamiltonian, which means that one may write an electronic form of the Schrodinger equation

$$H_{elec}|\Phi\rangle_{lec} = E_{elec}|\Phi\rangle_{elec}(\{r_i\}; \{R_A\}) \tag{4}$$

Equation (4) gives a wavefunction that describes electronic motion in terms of the coordinates of electrons and fixed nuclei. The wavefunction depends explicitly on the coordinates of the electrons and parametrically on the nuclear coordinates. The electrons can be seen as particles that move under the influence of the relatively static potential coming from a particular arrangement of the nuclei. In the succeeding sections the symbol H shall refer to the electronic Hamiltonian and the subscript *elec* shall be dropped.

The Hartree-Fock Approximation[31]

The Hartree-Fock (HF) method is the starting point of nearly all approximation methods of computational chemistry. Thus, knowledge of where the HF equations come from as well as their implications is important to quantum chemists. There are several other sources[32] of

the full derivation of these equations and the following is meant to serve as an outline. The notation of Szabo and Ostlund's text is used in this section.

The molecular Hamiltonian H is usually divided into a one-electron term, $h(i)$, and a two-electron term, $g(i,j)$, defined as:

$$h(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} \tag{5a}$$

$$g(i,j) = \frac{1}{r_{ij}} \tag{5b}$$

Substitution of both (5a) and (5b) into equation (3) leads to

$$H = \sum_{i=1}^{N} h(i) + \sum_{i=1}^{N}\sum_{j>i}^{N} g(i,j) \tag{6}$$

The electronic wavefunction is usually approximated using a single Slater determinant

$$\Phi(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix} \tag{7}$$

where $\chi(x)$ is a spin orbital defined as the product of a spatial orbital $\psi(r)$ and a spin function, α or β, to yield

$$\chi(x) = \begin{cases} \psi(r)\alpha \\ \psi(r)\beta \end{cases} \tag{8}$$

and x_i indicates both the space and spin coordinates for electron i. Equation (7) is often abbreviated as

$$\Phi(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}}det|\chi_1\chi_2\ldots\chi_N| \tag{9}$$

assuming that the electron labels are written in the order $x_1, x_2, \ldots, x_N$. Since the spatial orbitals are characterized by orthonormality, so are the spin orbitals

$$\langle \chi_i | \chi_j \rangle = \delta_{ij} \tag{10}$$

where δ_{ij} is the Kronecker delta. On the other hand, the spin functions have the following properties

$$\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1 \tag{11a}$$

$$\langle \alpha | \beta \rangle = \langle \beta | \alpha \rangle = 0 \tag{11b}$$

Slater determinants impose two important properties of multielectron wavefunctions. The first is the antisymmetry property, which yields a sign change for the wavefunction when the positions and spins of any two electrons are switched. This corresponds to switching any two rows or any two columns in the Slater determinant. The second is the Pauli exclusion principle - no two electrons of the same spin can be in exactly the same point in space, since having two identical rows or columns in the Slater determinant gives a value of zero.

The quantum mechanical expression for the total energy of a molecule is written as

$$E = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{12}$$

Therefore, calculation of the energy corresponding to a single Slater determinant wavefunction involves substitution of equation (9) into (12) to obtain

$$E = \frac{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \ldots \chi_N| \left| H \right| \frac{1}{\sqrt{N!}} det|\chi_1 \ldots \chi_N| \right\rangle}{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \ldots \chi_N| \left| \frac{1}{\sqrt{N!}} det|\chi_1 \ldots \chi_N| \right\rangle} \tag{13}$$

Evaluation of (13) requires solving each term of the Hamiltonian separately. For the one-electron terms,

$$E_1 = \sum_{i=1}^{N} \frac{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \left| h(i) \right| \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \right\rangle}{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \left| \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \right\rangle} \tag{14}$$

Expansion of each Slater determinant yields a sum of $N!$ x $N!$ terms for each integral. One then uses the normalization conditions given in equation (10), the fact that $h(i)$ operates only on the ith electron coordinate, and the Slater-Condon rules to reduce to one electron integrals. After fully expanding the Slater determinant, the h matrix element is obtained as

$$h = \frac{1}{N!} (\langle \chi_1(x_i)|h(i)|\chi_1(x_i)\rangle + \cdots + \langle \chi_N(x_i)|h(i)|\chi_N(x_i)\rangle) \tag{15}$$

Since the result is the same for each of the N $h(i)$ matrix elements in the expression for E_1, the one-electron contribution is

$$E_1 = \sum_{i=1}^{N} \langle \chi_i|h(1)|\chi_i\rangle \tag{16}$$

For the two-electron terms,

$$E_2 = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \left| g(i,j) \right| \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \right\rangle}{\left\langle \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \left| \frac{1}{\sqrt{N!}} det|\chi_1 \dots \chi_N| \right\rangle} \tag{17}$$

Following an approach similar to the one taken for the one-electron terms, the expression for the two-electron energy in terms of the spin orbitals is

$$E_2 = \sum_{i=1}^{N} \sum_{j>i}^{N} \left(\begin{array}{c} \int dx_1 dx_2 \chi_i^*(x_1)\chi_i(x_1) \frac{1}{r_{12}} \chi_j^*(x_2)\chi_j(x_2) - \\ \int dx_1 dx_2 \chi_i^*(x_1)\chi_j(x_1) \frac{1}{r_{12}} \chi_j^*(x_2)\chi_i(x_2) \end{array} \right) \tag{18a}$$

Using the following notations:

$$[i|h|j] = \int dx_1 \chi_i^*(x_1)\, h(x_1) \chi_j(x_1) \tag{18b}$$

$$[ij|kl] = \int dx_1 dx_2 \chi_i^*(x_1) \chi_j(x_1) \frac{1}{r_{12}} \chi_k^*(\,_2) \chi_l(x_2) \tag{18c}$$

and combining equations (16) and (18), the HF energy is obtained as

$$E_0[\{\chi_i\}] = \sum_{i=1}^{N} [i|h|i] + \sum_{i=1}^{N}\sum_{j>i}^{N} [ii|jj] - [ij|ji] \tag{19}$$

The next step is to minimize $E_0[\{\chi_i\}]$ with respect to the spin orbitals, provided that the spin orbitals are orthonormal.   The method of Langrange multipliers is then used to determine the optimum functional with the following definition

$$L[\{\chi_i\}] = E_0[\{\psi_i\}] - \sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon_{ji}\big([i|j] - \delta_{ij}\big) \tag{20}$$

where ε_{ij} are the N^2 undetermined multipliers. At this point $\{\chi_i\}$ is determined so that $\delta L = 0$:

$$\delta L[\{\chi_i\}] = \delta E_0[\{\chi_i\}] - \sum_{i=1}^{N}\sum_{j=1}^{N} \varepsilon_{ji}\delta[i|j] = 0 \tag{21}$$

where $\delta[i|j] = [\delta i|j] + [i|\delta j]$. The variations on the right-hand side of (21) are expanded to give

$$\delta E_0[\{\chi_i\}] = \sum_{i=1}^{N} [\delta\chi_i|h|\chi_i] - \sum_{i=1}^{N}\sum_{j=1}^{N}\big([\delta\chi_i\chi_i|\chi_j\chi_j] + [\delta\chi_i\chi_j|\chi_j\chi_i]\big) \\ + complex\ conju\ ate \tag{22}$$

Similarly, one obtains for the second term in (21)

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{ij}\delta[i|j] = \sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon_{ji}[\delta i|j] + complex\ conjugate \tag{23}$$

Two operators are next defined:

$$J_i(1)\chi_j(1) = \left[\iint x_2\chi_i^*(2)\frac{1}{r_{12}}\chi_i(2)\right]\chi_j(1) \tag{24}$$

known as the Coulomb operator, the mean potential at x_1 resulting from all other electrons in

spin orbitals χ_i, where $(1) = x_1,$ $(2) = x_2,$ and so on, and

$$K_i(1)\chi_j(1) = \left[\int dx_2\chi_i^*(2)\frac{1}{r_{12}}\chi_j(2)\right]\chi_i(1) \tag{25}$$

known as the exchange operator, a nonlocal operator whose operation on an orbital depends

on the value of the orbital throughout all space and not just at x_1. Substitution of these

definitions into the variation expansions yields, after some rearrangement,

$$\left[h(1) + \sum_{i=1}^{N}\left(J_j(1) - K_j(1)\right)\right]\chi_i(1) = \sum_{i=1}^{N}\varepsilon_{ij}\chi_j(1), \text{ for i = 1,2,…,N} \tag{26}$$

The term in the square brackets is known as the Fock operator, $f(1)$. Substituting the Fock

operator into (26), one obtains a series of N integral equations known as the HF equations

$$f(1)\chi_i(1) = \sum_{i=1}^{N}\varepsilon_{ij}\chi_j(1), \text{ for i = 1,2,…,N.} \tag{27}$$

The Roothaan-Hall Equations for Closed Shells[33, 34]

The HF equations shown above have no assumptions about the form of the functional for

each of the one electron spin orbitals. For relatively simple systems, these equations can be

solved using methods for numerically solving integral equations. A more useful approach is

to recast the integral equations into matrix form and substitute a basis set expansion for the

spatial part of the orbitals to generate the Roothaan-Hall matrix equations which are central

to HF programs.

For a closed-shell system all the electrons are paired in $N/2$ orbitals. This means that the radial part of the orbital is the same for both α and β spins and the spin orbitals in a closed system can be given as

$\chi_{2i-1} = \psi_i\alpha$, $\chi_{2i} = \psi_i\beta$, ..., $\chi_{N-1} = \psi_{\frac{N}{2}}\alpha$, $\chi_N = \psi_{\frac{N}{2}}\beta$, where i goes from 1 to N.

This allows a set of HF equations to be derived that involves only the radial part of the spin orbitals. The HF restricted determinant is then obtained by pairing the electrons (with α and β spin) in the $N/2$ lowest energy orbitals. Also, diagonalization of the Fock matrix yields the orbital energy of the spin orbital χ_1, which is now written as ε_1. Deriving the spatial HF equations in terms of the α spin-orbitals (the same result is obtained in terms of the β spin) yields

$$f\psi_i(r_1)\alpha(\omega) = \varepsilon_i\psi_i(r_1)\alpha(\omega) \tag{28}$$

Integration over spin results in

$$\left[\int d\omega_1\alpha^*(\omega_1)f(x_1)\,\alpha(\omega_1)\right]\psi_i(r_1) = f(r_1)\psi_i(r_1) = \varepsilon_i\psi_i(r_1). \tag{29}$$

The HF equations can be solved using a finite basis set. These basis sets are of usually either Gaussian[35] or Slater[36] type orbitals, which are used to expand the spatial part of the spin orbitals. In the limit that the basis set approaches completeness, the HF limit is obtained.

The spatial part of the HF molecular orbitals is written as a linear combination of K basis functions, ϕ_μ, to obtain

$$\psi_i(r) = \sum_{\mu=1}^{K} \phi_\mu C_{\mu i} \tag{30}$$

Substitution of this linear combination into (29) yields

$$f(r_1) \sum_{\mu=1}^{K} \phi_\mu C_{\mu i} = \varepsilon_i \sum_{\mu=1}^{K} \phi_\mu C_{\mu i} \tag{31}$$

Multiplying the above equation on the left by a generic basis function, ϕ_ν^*, and integrating

gives

$$\sum_\mu \langle \phi_\nu | f | \phi_\mu \rangle C_{\mu i} = \varepsilon_i \sum_\mu \langle \phi_\nu | \phi_\mu \rangle C_{\mu i} \tag{32}$$

The matrix elements of the Fock operator, f, and the overlap between the basis functions are

redefined as

$$F_{\nu\mu} = \langle \phi_\nu | f | \phi_\mu \rangle = \int \phi_\nu^*(r) f(r) \phi_\mu(r) dr$$

$$\tag{33}$$

$$S_{\nu\mu} = \langle \phi_\nu | \phi_\mu \rangle = \int \phi_\nu^*(r) \phi_\mu(r) dr$$

and the solution to the HF equations is recast to an algebraic eigenvalue problem

$$FC = \varepsilon SC \tag{34}$$

where F is the Fock matrix, S is the overlap matrix, ε is a diagonal matrix of the eigenvalues,

and the expansion coefficients of the molecular orbitals in the finite basis set are arranged

into columns:

$$C = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ \vdots & \vdots & \vdots & \vdots \\ C_{\nu 1} & C_{\nu 2} & \dots & C_{\nu K} \\ \vdots & \vdots & \vdots & \vdots \\ C_{K1} & C_{K2} & \dots & C_{KK} \end{pmatrix} \tag{35}$$

In C, the expansion coefficients of the first molecular orbital are arranged in the first column,

the coefficients of the second molecular orbital in the second column, and so on. The use of

basis functions generates K radial orbitals and $2K$ spin orbitals. Of these, the N spin

orbitals with the lowest energies will be the occupied orbitals in the HF reference state, and

$2K - N$ are the set of virtual orbitals.

The elements of the Fock matrix are next evaluated over the basis functions

$$F_{\mu\nu} = \langle\phi_\mu|f|\phi_\nu\rangle = \int \phi_\mu^*(r)\left[h + \sum_{a=1}^{\frac{N}{2}}(2J_a - K_a)\right]\phi_\nu(r)\,dr \tag{36}$$

where J_a and K_a are defined as in equations (24) and (25). Substituting in the basis set expansion leads to

$$\begin{aligned}
F_{\mu\nu} &= \langle\phi_\mu|\hat{f}|\phi_\nu\rangle \\
&= \langle\phi_\mu|\hat{h}|\phi_\nu\rangle \\
&+ \sum_{a=1}^{\frac{N}{2}}\sum_{\sigma\lambda} C_{\sigma a}^* C_{\lambda a}\left[2\iint \phi_\mu^*(r_1)\phi_\sigma^*(r_2)\frac{1}{r_{12}}\phi_\nu(r_1)\phi_\lambda(r_2)dr_1 dr_2\right. \\
&\left. - \iint \phi_\mu^*(r_1)\phi_\sigma^*(r_2)\frac{1}{r_{12}}\phi_\lambda(r_1)\phi_\nu(r_2)dr_1 dr_2\right]
\end{aligned} \tag{37}$$

The density matrix[37] is then defined as

$$P_{\mu\nu} = 2\sum_{a=1}^{\frac{N}{2}} C_{\nu a}^* C_{\mu a} \tag{38}$$

which allows one to rewrite the expression for the Fock matrix elements as

$$\begin{aligned}
F_{\mu\nu} &= \langle\phi_\mu|\hat{f}|\phi_\nu\rangle \\
&= \langle\phi_\mu|h|\phi_\nu\rangle \\
&+ \sum_{\sigma\lambda} P_{\lambda\sigma}\left[\iint \phi_\mu^*(r_1)\phi_\sigma^*(r_2)\frac{1}{r_{12}}\phi_\nu(r_1)\phi_\lambda(r_2)dr_1 dr_2\right. \\
&\left. - \frac{1}{2}\iint \phi_\mu^*(r_1)\phi_\sigma^*(r_2)\frac{1}{r_{12}}\phi_\lambda(r_1)\phi_\nu(r_2)dr_1 dr_2\right]
\end{aligned} \tag{39}$$

This means that once the density matrix is known, the Fock matrix can be determined. Since the density matrix is defined in terms of the HF molecular orbitals, a solution to the HF equations can only be found through an iterative procedure. Thus, the term self-consistent field (SCF) is used to describe the procedure used to find solutions to the HF equations. The SCF cycle starts with a guess to the density matrix. One possible guess would correspond to a null matrix $P = 0$ and the first SCF cycle thus solves the independent-particle problem.

The electron density operator is a one-electron operator, which in the coordinate representation takes the form

$$\hat{\rho} = \sum_{i=1}^{N} \delta(r - r_i) \tag{40}$$

The HF electron density then is the expectation value of $\hat{\rho}$ with the RHF wavefunction:

$$\hat{\rho} = \langle \Psi_0 | \hat{\rho} | \Psi_0 \rangle = \left\langle \Psi_0 \left| \sum_{i=1}^{N} \delta(r - r_i) \right| \Psi_0 \right\rangle = 2 \sum_{i=1}^{\frac{N}{2}} |\psi_i(r)|^2 \tag{41}$$

where Ψ_0 is the restricted HF solution, ψ_i are the set of occupied HF spatial orbitals, and the factor 2 is due to the spin. The basis set expansion can be inserted into the above equations to yield

$$\rho(r) = 2 \sum_{i=1}^{\frac{N}{2}} \psi_i^*(r)\psi_i(r) = 2 \sum_{i=1}^{\frac{N}{2}} \sum_{\mu} \sum_{\nu} \phi_\nu^* \phi_\mu C_{\nu i}^* C_{\mu i} = \sum_{\mu\nu} P_{\mu\nu} \phi_\nu^* {}_\mu \tag{42}$$

The number of electrons is given by

$$\int \rho(r) = \sum_{\mu\nu} P_{\mu\nu} \int \phi_\nu^* \phi_\mu dr = \sum_{\mu\nu} P_{\mu\nu} S_{\nu\mu} = Tr(PS) \tag{43}$$

where Tr is the trace of the matrix.

RHF theory as constructed using the Roothaan approach[32] suffers from certain chemical and practical limitations. From a practical standpoint, HF theory posed some very challenging technical problems to early computational chemists. One problem was the selection of a basis set. The linear combination of atomic orbitals (LCAO) approach using hydrogenic orbitals remains attractive in principle; however, this basis set requires numerical solution of the four-index integrals appearing in the Fock matrix elements, and that is a very tedious process. Moreover, the number of four-index integrals is daunting. Since each index runs over the total number of basis functions, there are in principle $O(N^4)$ total integrals to be evaluated, although with screening procedures now in place in most calculation packages the total integrals scales down to $O(N^3)$, particularly for large chemical systems.

If the system being studied is small enough, the two-electron integrals can be obtained and then stored in memory. For a molecule with N basis functions and no geometrical symmetry this will mean formally calculating $N^4/8$ two-electron integrals. This becomes rather impractical when the number of basis functions goes beyond even 100. So for the double zeta polarization (DZP) basis set with the second row atoms, consisting of three s-type functions, two p-type functions, and six d (polarization) functions (five when using sphericals) for the first-row elements except for hydrogen, a typical system will be limited to a size of about six heavy atoms for full storage in memory. The d functions allow one to describe some angular bonding functions and are needed to carry out correlated calculations. They also give significant recovery of the correlation energy when used with correlated methods such as those described below.

On the other hand, one may save the two-electron integrals on disk and then read them into the fast memory for each iteration. This technique relies heavily on the capacity and

speed of the I/O system of the machine used to execute the calculation. For gigabyte-sized disk storage, I/O costs start to escalate with CPU time when the size of the basis set reaches 200 functions although with disk farms now on tera-byte scale for large parallel computers, this has become less of an issue.

Still another way to deal with the two-electron integrals is known as the "direct" SCF method.[38,39] This involves recomputing the two-electron integrals at each iteration. One advantage of this approach is the elimination of memory and disk storage requirements. On the other hand, this approach also results in increased CPU time due to integral recomputation. However, if the product between the overlap for functions i and j and that for functions k and l does not exceed some preset threshold, then the integral $(ij|kl)$ is not computed. This results in scaling of about $O(N^2)$, and if the density matrix decreases quickly, scaling down to $O(N \log N)$ is possible.[40]

From a chemical standpoint the one-electron nature of the Fock operators ignores all electron correlation. The effect of correlation on various molecular properties is of interest to many chemists. The manner in which electron correlation is dealt with will be discussed in the following sections in connection with other levels of theory.

Density Functional Theory

Kohn and Sham developed the foundation for the application of density functional theory (DFT) in 1965.[41] The formalism gives a convenient way to solve the Hohenberg-Kohn theorem which states that the ground-state energy of a nondegenerate electronic system as well as its corresponding electronic properties are uniquely determined by the system's electron density.[42] The Hohenberg-Kohn theory further states that there exists a functional that relates electron density to the energy of a system but fails to describe the form of the

functional. The main goal, therefore, of DFT methods is to find functionals that are fully able to connect these two quantities.

The Kohn-Sham formalism starts from a virtual system of noninteracting electrons whose ground-state density corresponds to some real system of chemical interest whose electrons do interact. A challenge for DFT formalisms lies in determining the kinetic energy of the system. Kohn and Sham suggested that the kinetic energy functional of a system may be divided into two parts: one with electrons deemed as noniteracting particles, which can be solved exactly, and another which serves to correct for interactions among the system's electrons. Within this formalism, the electronic energy of the ground state of a system may be written as

$$E[\rho] = \frac{1}{2}\sum_{i=1}^{n} \int \Psi_i^*(r_1)\nabla_i^2 \Psi_i(r_1)dr_1 - \sum_{X=1}^{N} \int \frac{Z_X}{r_{X_i}}\rho(r_1)dr_1 + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{r_{12}}dr_1 dr_2 \qquad (44)$$
$$+ E^{XC}[\rho]$$

where n is the number of electrons, N is the number of nuclei, the Ψ_i's are the Kohn-Sham orbitals, $\rho(r)$ is the electron density at some location r, Z_X is the charge on nucleus X, and E^{XC} is the exchange-correlation energy. The first term of the above equation gives the kinetic energy of the noninteracting electrons, the second term represents interactions between electrons and nuclei, and the third term yields the coulomb repulsions between the total charge distributions at r_1 and r_2. The exchange-correlation energy corrects the kinetic energy due to the actual interacting nature of the electrons, as well as some correction to the repulsion energy among the electrons. It is the description of the exchange-correlation energy that serves as the greatest challenge to DFT.

An important feature of the usual method for solving DFT equations is the solution of the Kohn-Sham orbitals by applying the variational principle to the electronic energy and the charge density from equation (41) so that

$$h_i \Psi_i(r_1) = \varepsilon_i \Psi_i(r_1) \tag{45}$$

where $\hat{h}_i$ is the Kohn-Sham Hamiltonian and ε_i is the orbital energy. The Hamiltonian is in turn given by

$$h_i = -\frac{1}{2}\nabla_1^2 - \sum_{X=1}^{N} \frac{Z_X}{r_{Xi}} + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V^{XC}(r_1) \tag{46}$$

The last term in the above equation is the functional derivative of the exchange-correlation energy and is obtained using

$$V^{XC} = \frac{\delta E^{XC}[\rho]}{\delta \rho} \tag{47}$$

As in the restricted HF method, the solution of the Kohn-Sham equation is done in an iterative manner. One starts with a trial charge density ρ, which can be obtained from a superposition of the atomic densities. A form of the functional that describes E^{XC} is then used to determine V^{XC}. An initial solution to the Kohn-Sham equation is obtained, which yields the first set of orbitals. These orbitals are used to improve the charge density and the process is repeated until convergence criteria are met by both the density and the exchange-correlation energy. Finally the electronic energy is calculated.

The orbitals used in the above process are usually expressed in terms of a set of basis functions. As in HF calculations, solving the Kohn-Sham equations involves determining coefficients for a linear combination of basis functions. Thus, selecting the correct basis set cannot be overemphasized. The number of integrals needed for solving the KS equations also

increases as $O(N^4)$ due to the number of integrals in the coulomb interaction functional, which is the third term in equation (49).

The exchange-correlation term is often viewed as a sum of an exchange term, E^X, and a correlation term, E^C. The exchange term is related to interactions among electrons that possess the same spin while the correlation term relates electrons with opposite spins. The two terms are themselves functionals of the electron density. Functionals vary mainly on how the electron correlation is included. However, because exchange-correlation functionals are mostly empirical, no theoretically founded system for systematically improving DFT exists. On the other hand, Perdew has proposed the "Jacob's ladder" approach, which indicates improvement in the accuracy of a functional the higher it is found on the ladder.[43]

Local density approximation (LDA) assumes that the exchange-correlation energy is a function solely of the electron density and that this density can be described by the behavior of a homogeneous electron gas. The first local density approximation to the exchange energy came from Dirac.[44] This method was combined with the Thomas-Fermi model[45] to give the Thomas-Fermi-Dirac method. The crude nature of the kinetic energy functional from the Thomas-Fermi model resulted in total energies with errors of about 15-50%. Slater then introduced spin dependence into the functional to yield the local spin density approximation (LSDA).[46] LSDA is a more general application of LDA with spin dependence introduced into the functionals, thus, solving several conceptual problems with LDA approaches for chemical systems that involve an external magnetic field, those that are polarized, and those with significant relativistic effects.

An area of concern when dealing with LDA is the difficulty in separating the correlation energy from the exchange energy, if indeed they should be separated. Interpolation formulas

coming from values calculated for different densities of a homogeneous electron gas have been used to address this problem. The Vosko-Wilk-Nusair[47] and the Perdew-Wang[48] formulas used Monte Carlo results from Ceperley[49] to determine the correlation energy of a uniform electron gas.

Like all other types of functionals, LDA fails to sufficiently cancel self-interaction contributions. This results in a tendency to underestimate atomic ground-state energies and ionization energies while overestimating binding energies. LDA performs more poorly with small molecules but does better as system size increases. Since LDA underestimates exchange energies while overestimating correlation energies, E^{XC} values tend to be good because of cancellation of errors.

Generalized gradient approximation (GGA) methods take into consideration the fact that molecular systems are quite different from a homogeneous electron gas, and thus, will have spatially varying densities. GGAs have exchange and correlation energies that rely not only on the density but also on the density gradient. GGAs are either based on numerical fitting procedures using large molecular training sets, such as those first proposed by Becke[50], or those that include exchange-correlation functional based on scaling relations, limits for high and low densities, LSDA limits for slowly varying densities, and exact relations for the exchange and correlation holes, such as those formulated by Perdew.[51] The first group gives accurate atomization energies and reaction barriers for molecules[52] but performs poorly when considering solid-state systems.[53] On the other hand, the second group of functionals has problems with determining atomization energies and reaction barriers for molecules but perform quite well in calculating solid-state properties.[54]

GGA methods generally exhibit significant improvement over LDA methods. GGA usually gives more accurate values for total energies,[55] atomization energies,[56] structural energy differences, and energy barriers.[57] There is also a tendency to stretch and soften bonds. However, GGA methods fail to measure van der Waals interactions[58] and are not much better than LDA when determining ionization potentials and electron affinities.[59]

Meta-GGAs (M-GGA) include semi-local information beyond the first-order density gradient present in GGA. M-GGAs depend explicitly on higher-order density derivatives or on kinetic energy density, which involves derivatives of occupied Kohn-Sham orbitals. This group of functionals gives more accurate values for properties, such as atomization energies, but are often more challenging in terms of numerical stability. Among the more popular M-GGAs are B95[60], TPSS[61], and VSXC.[62]

The most widely used functionals are known as hybrid density functionals (H-GGA). They combine the exchange-correlation of a GGA method with some percent of Hartree-Fock exchange. The weight for each component and the functional mixed is usually optimized based on experimental data such as atomization energies, ionization potentials, proton affinities, and total atomic energies for a set of molecules.[63] Hybrid functionals are widely popular because they provide great improvement over GGAs in calculating molecular properties but they still often fail in solid-state physics. This is due largely to problems in computing the exact-exchange part in a plane-wave basis set. The most popular example of these functionals is B3LYP.[64, 65, 66, 67]

Möller-Plesset Perturbation Theory[68]

Perturbation methods generally assume that there is a mathematical expression for the physical quantity being solved but for which an exact solution is not readily on hand. They

also assume that this physical quantity can be divided into two parts, one of which can be solved exactly and another, which has no analytic solution. It is also usually assumed that the "perturbation" is small compared to the exactly solvable part of the problem. The Rayleigh-Schrodinger perturbation theory [69] (RSPT) provides a prescription for accomplishing this.

In RSPT the equation one wishes to solve may be written as

$$H\Psi_n = E_n\Psi_n \tag{48}$$

with $n = 0,1,2,...$, H is the Hamiltonian, and Ψ_n is a member of an orthonormal basis. RSPT calculates the perturbed form of the state Ψ_0 with energy $E_0^{(0)}$, but this state does not necessarily need to be the ground state of the system. The Hamiltonian of the perturbed system is then written as

$$H = H^{(0)} + \lambda H^{(1)} \tag{49}$$

where λ is a parameter taken to be real with a value between 0 and 1. Towards the end of the calculation, the parameter λ is set to 1. Similarly, the perturbed wavefunction becomes

$$\psi_0 = \psi_0^{(0)} + \lambda\psi_0^{(1)} + \lambda^2\psi_0^{(2)} + \cdots \tag{50}$$

which shows that the unperturbed function is corrected by terms that are in different orders of the perturbation. The energy of the perturbed state also has correction terms of various orders

$$E_0 = E_0^{(0)} + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + \cdots \tag{51}$$

Each of the terms with superscripts (1), (2), … are referred to as the nth-order correction of the energy. This leads to the nth-order energy or wavefunction as being the sum of all terms

leading up to order n. The perturbed wavefunction is usually chosen to be intermediately normalized, that is

$$\left\langle \psi_0^{(0)} \middle| \psi_0 \right\rangle = 1$$

$$\left\langle \psi_0^{(0)} \middle| \psi_0^{(0)} + \lambda \psi_0^{(1)} + \lambda^2 \psi_0^{(2)} + \cdots \right\rangle = 1 \tag{52}$$

$$\left\langle \psi_0^{(0)} \middle| \psi_0^{(0)} \right\rangle + \lambda \left\langle \psi_0^{(0)} \middle| \psi_0^{(1)} \right\rangle + \lambda^2 \left\langle \psi_0^{(0)} \middle| \psi_0^{(2)} \right\rangle = 1$$

Once all the correction terms are determined, normalization of the total wavefunction becomes trivial. For example, the first-order energy correction is obtained as

$$E_0^{(1)} = \left\langle \psi_0^{(0)} \middle| H^{(1)} \middle| \psi_0^{(0)} \right\rangle \tag{53}$$

which means that the first-order correction to the energy is simply the expectation value of the perturbation of the Hamiltonian with the ground-state wavefunction.

Moller-Plesset perturbation theory[67] (MPPT) is a particular form of many-body perturbation theory that takes the Hartree-Fock Hamiltonian, H_{HF}, as a sum of all the one-electron Fock operators, f_i,

$$H_{HF} = \sum_{i=1}^{N} f_i \tag{54}$$

and has electron correlation as a perturbation to the zeroth-order Hamiltonian. A feature of MPPT that characterizes it from other types of theories is size extensivity, which means that the predicted energy for each order of perturbation in MPPT scales with the number of non-interacting particles in the system of interest. This is important when comparing systems with different numbers of electrons and when dealing with infinite systems like crystals and cellulose-like structures. Another distinguishing characteristic of MPPT is that its use of

electron correlation as a perturbation results in a total electronic energy that is not variational in nature.

The HF ground-state wavefunction, Φ_0, is an eigenfunction of H_{HF}, with eigenvalue equal to $E_0^{(0)}$, which is the sum of the orbital energies of all the occupied spin orbitals. The perturbation, $H^{(1)}$, is obtained as

$$H^{(1)} = H - \sum_{i=1}^{N} f_i \tag{55}$$

where H is the electronic Hamiltonian. The energy from H_{HF} is associated with the normalized ground-state HF wavefunction, and can be written as the expectation value using the zero-order wavefunction

$$E_{HF} = \langle \Phi_0 | H | \Phi_0 \rangle \tag{56}$$

which upon substitution of equation (55) and rearrangement yields

$$E_{HF} = \langle \Phi_0 | H_{HF} + H^{(1)} | \Phi_0 \rangle \tag{57}$$

Since Φ_0 is an eigenfunction of H_{HF}, then the zero-order energy is the expectation value

$$E_0^{(0)} = \langle \Phi_0 | H_{HF} | \Phi_0 \rangle \tag{58}$$

and the first-order correction to the energy may be written as

$$E_0^{(1)} = \langle \Phi_0 | H^{(1)} | \Phi_0 \rangle \tag{59}$$

Thus, from equations (57) through (59), one can conclude that the Hartree_Fock energy is simply the sum

$$E_{HF} = E_0^{(0)} + E_0^{(1)} \tag{60}$$

It can also be seen from equation (57) that electron correlation, therefore, begins at order 2. Therefore, the first-order correction to the ground-state energy is obtained from the second-order of Rayleigh-Schrodinger perturbation theory

$$E^{(2)} = \sum_{J>0} \frac{\langle \Phi_J | H^{(1)} | \Phi_0 \rangle \langle \Phi_0 | H^{(1)} | \Phi_J \rangle}{E_0^{(0)} - E_J^{(0)}} \tag{61}$$

where Φ_J is a multiply-excited determinant and an eigenfunction of H_{HF} with eigenvalue $E_J^{(0)}$. The off-diagonal matrix elements $\langle \Phi_J | H^{(1)} | \Phi_0 \rangle$ are then evaluated. However, Brillouin's theorem states that singly excited determinants will not interact directly with a referenced HF determinant. This leads to the conclusion that only double excitations contribute to $E^{(2)}$. Thus,

$$E^{(2)} = \frac{1}{4} \sum_{a,b}^{occ} \sum_{p,q}^{vir} \frac{\langle ab || pq \rangle \langle pq || ab \rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_p - \varepsilon_q} \tag{62}$$

where

$$\langle ab || pq \rangle = \int \phi_a^*(1) \phi_b^*(2) \frac{1}{r_{12}} (1 - \wp_{12}) \phi_p(1) \phi_q(2) dx_1 dx_2 \tag{63}$$

with ϕ_a and ϕ_b being occupied orbitals and ϕ_p and ϕ_q being virtual spin orbitals, and $\wp_{12}$ is the permutation operator, which interchanges the coordinates of electrons one and two.

An MP2 calculation begins by carrying out an SCF calculation in a given basis to obtain Φ_0, E_{HF}, the orbital energy and the molecular orbitals (including the virtual orbitals). The second-order correction to the energy is then calculated by evaluating integrals over spin orbitals in terms of integrals over the basis functions.

MP2 calculations are generally faster than those that involve configuration interaction (CI), a variational method that expresses the exact wavefunction is a linear combination of

N-electron trial functions. CI provides an exact solution to the many-body problem in principle, but is not practical to use in computation except for the smallest of systems or with the smallest of basis sets since it nominally scales as $O(N!)$.[70, 71] The second-order energy correction, on the other hand, can be efficiently evaluated and scales roughly as $O(N^5)$. MP2 calculations consist mainly of the HF calculation, the transformation from atomic orbital to molecular orbital integrals, and the energy calculation. However, only the second part of the calculation formally scales as $O(N^5)$. The other two scale as $O(N^4)$.[72]

Direct MP2 calculations with large molecules are feasible and practical, provided there is reasonable memory available.[73] When the geometry is optimized with polarized functions using the direct SCF method, the MP2 energy is readily obtained at very little extra cost compared to the same calculations done with RHF.

Calculations using small basis sets with MP2 are of little or no value. To be useful, MP calculations generally should use basis sets that are 6-31G(d)[74] or larger. These calculations typically obtain 85 to 95 percent of the correlation energy.[70] With respect to calculations of energies, MP2 is highly efficient in determining energy differences between minima. Errors are reduced by 25-50% in changing the level of theory from HF to MP2.[75]

MP2 is a good method that includes correlation energy for determining the geometry at energy minima.[76] Bond lengths are typically accurate to within 0.015 Å, compared to errors with HF calculations, where errors go up to an average of 0.021 Å for molecules consisting of two to eight atoms. With respect to bond angles, HF calculations are seen to be sufficiently accurate and that calculations at the MP2 level offer little improvement.[77]

Fragment Molecular Orbital Method

The fragment molecular orbital (FMO)[78, 79, 80] method is used to perform high-level *ab initio* quantum chemical calculations for very large molecular systems and can be applied to any *ab initio* method. FMO basically divides a large molecule into fragments, for example, a polysaccharide into its monosaccharide components. The fragmentation is done so that electrons in a covalent bond at the point of fragmentation are intuitively assigned to one of the atoms originally part of the broken bond. Usually, this means that the more electronegative atom in the bond gets assigned the two electrons. In addition, a proton is also assigned from one atom to the other to maintain charge neutrality. As a result the overall charge of the system is maintained throughout the calculation.

The basic algorithm in calculating the energy of a system consisting of N monomers with the FMO method involves the following steps: (1) calculation of the initial electron density distribution for each monomer; (2) construction of the monomer Fock operators using the densities determined in step (1), followed by calculation of the energy of each monomer in the Coulomb bath of the $N-1$ other monomers; (3) iteration of each monomer energy to self-consistency (the monomer SCF) to obtain a converged electrostatic potential (ESP); (4) a single calculation of the energy of each fragment dimer (FMO2) in the converged ESP of the other $N-2$ monomers; and optionally, (5) a single calculation of the energy of each trimer (FMO3) in the converged ESP of the remaining $N-3$ fragments.[81]

For a two-body calculation (FMO2), the total energy of the system may be calculated as

$$E = \sum_{I}^{N} E_I + \sum_{I>J}^{N} \left(E_{IJ} - E_I - E_J \right) \tag{64}$$

where N is the number of fragments, E_I is the energy of monomer I and E_{IJ} is the energy of the dimer made up of fragments I and J. The above equation for the total energy can also be rewritten as:

$$E = \sum_{I>J}^{N} \Delta E_{IJ} - \sum_{I}^{N} E'_I \tag{65}$$

$$\Delta E_{IJ} = \left(E'_{IJ} - E'_I - E'_J \right) + Tr\left(\Delta P^{IJ} V^{IJ} \right) \tag{66}$$

where ΔP^{IJ} is the density difference matrix in the environment of the other fragments, V^{IJ} is the environmental electrostatic potential for dimer IJ, E'_I is the monomer energy without the environmental electrostatic potential, and E'_{IJ} is the dimer energy without the environmental electrostatic potential. ΔE_{IJ} is the pair interaction energy (PIE) between the Ith and Jth fragments. PIEs are obtained from the same FMO calculations, and give useful insight on inter- and intra-molecular interactions.

The basic strategy of FMO is to combine high-order interactions with low-order expansions.[82] Electrostatic interaction is treated at the full N-body order by incorporating the Coulomb field of all N fragments into the self-consistent monomer SCF cycle. Nonelectrostatic interactions (exchange-repulsion, charge transfer, and dispersion), on the other hand, are treated with lower order (2 or 3). The fragments, their pairs (known as FMO2), and triples (FMO3) are explicitly treated quantum-mechanically in the presence of the Coulomb field of the whole system. This results in coupling of the nonelectrostatic interactions, which are generally short-range, with the electrostatics.

The above approximation and fragmentation strategy gives reduced computational cost of calculations with the main limiting factor being the system size in terms of the basis functions. This is true even with expensive algorithms such as MP2 or CCSD(T). Aside

from breaking up the system into much smaller fragments, FMO can take advantage of multilevel parallelization through the use of the generalized distributed data interface (GDDI).[83]

The development of DFT-based FMO2 methods was first proposed by Sugiki et al.[84] and has since been extended to FMO3 by Fedorov and Kitaura.[85] The Moller-Plesset perturbation theory was also first interfaced with FMO by Fedorov and Kitaura.[86] RI-based FMO2-MP2 was suggested by Ishikawa and Kuwata[87] while that based on Cholesky decomposition was started by Okiyama et al.[88] Fedorov and Kitaura[89] developed the FMO3-based MP2 for energy calculations, while Mochizuki et al.[90] extended the same to include gradients.

The analysis of interactions between pairs of fragments is naturally built into the FMO method. In addition to PIEs, Fedorov and Kitaura have developed a pair interaction energy decomposition analysis (PIEDA)[91] which breaks up PIEs into their electrostatic, exchange-repulsion, charge transfer, and dispersion components. The calculation of PIE values allows the determination of the magnitude and nature (whether they are attractive or repulsive) of such interactions.

In the next chapter, the energies arising from the rotation of free hydroxyl groups in the central glucose residue of a cellulose crystalline assembly, calculated using RHF, DFT, and FMO2/MP2 methods, will be presented. In addition, interactions of this central glucose residue with some of the surrounding residues (selected on the basis of the interaction strengths) are analyzed.

The third chapter revisits the mechanism of acid-catalyzed hydrolysis of cellobiose, which is the repeating unit of cellulose. Energies corresponding to the different steps of this

mechanism calculated using RHF and DFT will be compared with those previously reported

using molecular dynamics calculations and with experimental data.

[1] Carroll, A. and Somerville, C. *Annu. Rev. Plant Biol.* **2009**, *60*, 165-182.

[2] Energy Information Administration (2009) Annual energy outlook. Energy Information Administration, Washington, D. C.

[3] Hamelinck, C. n.; Faail, A. P. C.; den Uil, H.; Boerrigter, H. *Energy* **2004**, *29*, 11.

[4] Huber, G. W.; Iborra, S.; Corma, A. *Chem. Rev.* **2006**, *106*, 9.

[5] Mok, W. S.; Antal, M. J.; Varhegyi, G. *Ind. Eng. Chem. Res.* **1992**, *31*, 1.

[6] Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W. and Foust, T. D. *Science*. **2007**, *315*, 804-807.

[7] Zhong, L.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L.; Adney, W. S.; Himmel, M. E. and Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984-1992.

[8] Petersen, L.; Ardevol, A.; Rovira, C. and Relly, P. J. *J. Phys. Chem. B.* **2009**, *113*, 7331-7339.

[9] Itoh, T.; Kimura, S. *J. Plant Res.* **2001**, *114*, 483.

[10] Brown, R. M., Jr. *J. Macromol. Sci. Pure Appl. Chem.* **1996**, *A33*, 1345.

[11] Sugiyama, J.; Harada, H.; Fujiyoshi, Y.; Uyeda, N. *Planta* **1985**, *166*, 161.

[12] Kuga, S.; Brown, R. M., Jr. *Polym. Commun.* **1987**, *28*, 311.

[13] Horii, F.; Hirai, A. and Kitamaru, R. *Macromolecules*. **1987**, *20*, 2117-2120.

[14] Nishiyama, Y.; Langan, P.; Chanzy, H.; *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

[15] Koyama, M.; Helbert, W.; Imai, T.; Sugiyama, J.; Henrissat, B. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9091.

[16] Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300-14306.

[17] French, A. D., Kelterer, A.-M., Johnson, G. P., Dowd, M. K. & Cramer, C. J. (2001). J. Comput. Chem. 22,65

[18] Nishiyama, Y.; Langan, P.; Wadac, M.; V. Forsyth. V. T. *Acta Cryst.* **2010**, *D66*, 1172.

[19] French, A. D.; Johnson, G. P. *Can. J. Chem.* **2006**, *84*, 603.

[20] Strati, G. L..; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833.

[21] Mendonca, S. Johnson, G. P.; French, A. D.; Laine, R. A. *J. Phys. Chem. A* **2002**, *106*, 4115.

[22] French, A. D.; Johnson, G. P. *Cellulose* **2004**, *11*, 449.

[23] Matthews, J. F.; Himmel, M. E.; Brady, J. W. In *Computational modeling in lignocellulosic biofuel production.* **2010**, 17.

[24] Dadach, Z. E.; Kaliaguine, S. *Can. J. Chem. Eng.* **1993**, *71*, 880.

[25] Deng, W.; Lobo, R.; Setthapun, W.; Christensen, S. T.; Elam, J. W.; Marshall, C. L. *Catal. Lett.* **2011**, *141*, 498.

[26] Strati, G. L.; Willett, J. L.; Momany, F. A. *Carbohydr. Res.* **2002**, *337*, 1833.

[27] Bosma, W. B.; Appell, M.; Willett, J. L.; Momany, F. A. *J. Mol. Struct. (THEOCHEM)* **2006**, *776*, 1.

[28] Bosma, W. B.; Appell, M.; Willett, J. L.; Momany, F. A. *J. Mol. Struct. (THEOCHEM)* **2006**, *776*, 21.

[29] Liang, X.; Montoya, A.; Haynes, B. *J. Phys. Chem. B* **2011**, *115*, 10682.

[30] Born, M.; Oppenheimer, J. R. *Ann. Physik* **1927,** *84*, 457.

[31] Szabo, A.; Ostlund, N. S. (1996). *Modern Quantum Chemistry*. Mineola, New York: Dover Publishing.

[32] Jensen, F. *Intro. To Comput. Chem.* 2nd ed. **2007**. John Wiley and Sons, West Sussex, England.

[33] Roothaan, C. C. J. (1951). *Rev. Mod. Phys.* **23**: 69–89.

[34] Hall, G. G. (1951). *Proc. Royal Soc. London A* **205**: 541–552.

[35] Pople, J. A.; Hehre, W. J. (1978). *J. Comp. Phys.* **27** (2): 161–168.

[36] Slater, J.C. (1930). *Phys. Rev.* **36**: 57.

[37] Fano, Ugo (1957). *Rev. Mod. Phys.* **29**: 74–93

[38] Almlof, J.; Faegri, K.; Korsell, K. *J. Comput. Chem.* **1982**, *3*, 385-99.

[39] Cremer, D. and Gauss, J. *J. Comput. Chem.* **1986**, *7*, 274-82.

[40] Head-Gordon, M.; and Pople, J. A. *J. Che. Phys.* **1988**, *89*, 5777-86.

[41] Kohn, W. and Sham, L. J. *Phys. Rev. A.* **1965**, *140*, 1133.

[42] Hohenberg, P. and Kohn, W. *Phys. Rev. B.* **1964**, *136*, 864.

[43] Perdew, J. P.; Ruzsin szky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. I. *J. Chem. Phys.*,

2005, *123*, 062201.

[44] Dirac, P. A. M. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542.

[45] Thomas, L. H. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542.

[46] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.

[47] Vosko, S. H.; Wilk, L. and Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

[48] Perdew, J. P. and Zunger, A. *Phys. Rev. B.* **1981**, *23*, 5048.

[49] Ceperley, D. *Phys. Rev. B.* **1978**, *18*, 3126.

[50] Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524.

[51] Perdew, J. P. and Wang, Y. *Phys. Rev. B.* **1986**, *33*, 8800.

[52] Becke, A. D. *J. Chem. Phys.* **1992**, *96*, 2155.

[53] Kurth, S.; Perdew, J. P. and Blaha, P. *Int. J. Quantum Chem.* **1999**, *75*, 889.

[54] Becke, A. D. *J. Chem. Soc.* **1992**, *97*, 9173.

[55] Langreth, D. C.; Mehl, M. J. *Phys. Rev. B,* **1983**, *28*, 1809.

[56] Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolthais, C.; *Phys. Rev.,* **1982,** *66*, 6671.

[57] Proynov, E. I.; Ruiz, E.; Vela, A.; Salahub, D. R. *Int. J. Quantum Chem.* **1995**, *29*, 61.

[58] Tao, J. M.; Perdew, J. P. *J. Chem. Phys.* **2005**, *122*, 114102.

[59] Kutzler, F. W.; Painter, G. S. *Phys. Rev. B* **1994**, *43*, 6865.

[60] Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

[61] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[62] Van Voorhis, T.; Scuseria, G. E. *J. Chem. Phys.* **1998**, *109*, 400.

[63] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

[64] Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

[65] Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098.

[66] Lee, C.; Yang, W. and Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785.

[67] P.J. Stephens, P. J. ;Devlin, F. J.; Chabalowski, C. F. and Frisch,M. J. *J. Phys. Chem.* **1994**, *98*, 11623.

[68] Moller, C. and Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

[69] Lowdin, P. O. *Phys. Rev. A.* **1965**, *139*, 357.

[70] Sherill, C. D.; Schaefer, H. F. *Adv. In Quantum Chem.* **1999**, *34*, 143.

[71] Bartlett, R. J. *Ann. Rev. Phys. Chem.* **1981**, *32*, 359.

[72] Jensen, F. *Introduction to Computational Chemistry.* 2nd ed. **2007**. John Wiley and Sons, West Sussex, England.

[73] Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1988**, *153*, 503.

[74] (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J.Chem.Phys.*, **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J.Chem.Phys.*, **1972**, *56*, 2257.

[75] Barrows, S. E.; Storer, J. W.; Cramer, C. J.; French, A. D.; Truhlar, D. G. *J. Comput. Chem.* **1998**, *19*, 1111.

[76] Scheiner, A. C.; Baker, J.; Andzelm, J. W. *J. Comput. Chem.* **1997**, *18*, 775.

[77] Feller, D.; Peterson, K. A. *J. Chem. Phys.* **1998**, *108*, 154.

[78] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano,T. and Uebayasi,M. *Chem. Phys. Lett.* **1999**, *313*, 701.

[79] Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M. and Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.

[80] Fedorov, D. G. and Kitaura, K. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, Fragment molecular orbital method: an approximate computational method for large molecules, in Starikov, E. B.; Lewis, J. P. and Tanaka, S. (Eds.). *Modern Methods for Theoretical Physical Chemistry of Biopolymers*. Elsevier. **2006**, 3.

[81] Fedorov, D. G. and Kitaura, K. *J. Phys. Chem. A.* **2007**, *111*, 6904.

[82] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R. and Slipchenko, L. V. *Chem. Rev.* **2011**,

[83] Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S. and Koseki, S. *J. Chem. Comput.* **2004**, *25*, 872.

[84] Sugiki, S.-I.; Kurita, N.; Sengoku, Y.; and Sekino, H. *Chem. Phys. Lett.* **2003**, *382*, 611.

[85] Fedorov, D. G. and Kitaura, K. *Chem. Phys. Lett.* **2004**, *389*, 129.

[86] Fedorov, D. G. and Kitaura, K. *J. Chem. Phys.* **2004**, *121*, 2483.

[87] Ishikawa, T. and Kuwata, K. *Chem. Phys. Lett.* **2009**, *474*, 195.

[88] Okiyama, Y.; Nakano, T.; Yamashita, K.; Mochizuki, Y.; Taguchi, N. and Tanaka, S. *Chem. Phys. Lett.* **2010**, *490*, 84.

[89] Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P. and Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.

[90] Mochizuki, Y.; Nakano, T.; Komeiji, Y.; Yamashita, K.; Okiyama, Y.; Yoshikawa, H. and Yamataka, H. *Chem. Phys. Lett.* **2011**, *504*, 95.

[91] Fedorov, D. G. and Kitaura, K. *J. Comput. Chem.* **2007**, *28*, 222.

CHAPTER 2: As the Hydroxyl Group Turns...

A paper to be submitted to the Journal of Chemical Theory and Computation

John Y. Baluyut, Yana A. Kholod, Ajitha Devarajan, Sergiy Markutsya, Monica H. Lamm,

Mark S. Gordon, and Theresa L. Windus

Abstract

Single-point energies resulting from the rotations of free -OH groups in the central residue of a cellulose Iα fragment consisting of nine cellotriose chains were obtained using restricted Hartree-Fock (RHF) with the 6-31G(d,p) basis set, density functional theory with the B3LYP functional using the 6-31G(d,p), and the fragment molecular orbital (FMO) method at the FMO2 method with second order perturbation theory (MP2) and the 6-31G(d) basis set. Potential energy curves calculated using these three methods are in excellent agreement with each other for the dihedral angles corresponding to energy maxima and minima. The calculated relative energies using the DFT/B3LYP and FMO2/MP2 levels of theory differ from each other by an average of 0.5 kcal/mol, 0.5 kcal/mol, and 1.1 kcal/mol when each of the -OH groups attached to the C2, C3, and C6 atoms, respectively, were rotated. The use of the pair interaction energy decomposition analysis (PIEDA) with the pair interaction energies from the dimer part of the FMO2 calculations also allowed the identification of the glucose residues most significantly involved in contributing to the rotational energy barriers. Intrachain and interchain interactions (those occurring between residues found in the same cellulose sheet) were seen to be stronger than intersheet interactions (occurring between residues found in different cellulose sheets) in contributing to the relative energy changes due to the rotations of free -OH groups in cellulose.

Introduction

Biomass is generally viewed as a source of renewable fuel.[1] Concerns on the amount of CO_2 emissions from transportation systems have also generated widespread interest in the use of biomass to replace fossil fuel for transport purposes since the biomass absorbs CO_2 while it is growing.

Naturally-occuring enzymes, known as cellulases, break cellulose down into the constituent glucose and cellobiose units. However, enzyme-catalyzed processes come with very slow rates and significant production costs for the enzymes. This is primarily due to a property of plant cell walls known as biomass recalcitrance, which renders cellulosic materials resistant to microbial and enzymatic destruction.[2,3,4]

Cellulose consists of D-glucose units linked by $\beta(1\rightarrow4)$-glycosidic bonds and is a linear polymer that goes into an extended and rather stiff rod-like conformation due to the extensive hydrogen bonding among the free hydroxyl groups of the glucose residues. Hydroxyl groups found in a glucose unit from one chain links with units on the same or on a neighboring chain by forming hydrogen bonds which hold the chains firmly together and form microfibrils with relatively high tensile strength. This strength is important in cell walls, where the microfibrils are meshed into a carbohydrate matrix, and gives rigidity to plant cells.

Most crystalline cellulose, also known as cellulose I, is found naturally in two forms: cellulose Iα and Iβ. Both forms of crystalline cellulose have the same conformation of their heavy atom backbone in the chains, but each possesses a different hydrogen bonding pattern that give different packing between the chains.[5] X-ray (synchrotron source) and neutron

diffraction experiments have been used to determine the conformation and packing arrangement in both polymorphs.[6,7] Cellulose Iα has a triclinic unit cell which contains one chain, belongs to space group *P*1 and has unit cell dimensions: a = 6.717 Å, b= 5.962 Å, c = 10.400 Å, α = 118.08°, β = 114.80°, and γ = 80.37°. On the other hand, cellulose Iβ has a monoclinic *P*2$_1$ unit cell made up of two different chains and has unit cell dimensions: a = 7.784 Å, b= 8.201 Å, c = 10.380 Å, α = 90°, β = 90°, and γ = 117.1°. Chains belonging to the same sheet in cellulose are mostly held together by a network of hydrogen bond interactions. Weak van der Waals interactions, on the other hand, are believed to hold sheets together.[3] Hydrogen bonding plays an important role in maintaining structures in native crystalline cellulose Iβ. Through simulation, networks of hydrogen bonding interactions were found in paired cellulose chains, particularly between the interchain and intrachain hydrogen bonds. Increased hydrogen bonding strength was also seen when long cellulose chains are allowed to form sheets, and when sheets were stacked together.[8]

The size of native cellulose has been explained based on competition between hydrogen-bonding energy and electronic strain energy using an *ab initio* pseudopotential method. Interchain interaction, referring to that occurring between glucose residues found in the same cellulose sheet, was found to be largely the domain of hydrogen bonding while weaker "intersheet" interactions dominate between sheets of glucose units in cellulose.[9]

The behavior of hydroxymethyl groups in individual glucose units on the cellulose I surfaces surrounded by water have also been determined using ^{13}C NMR.[10] While these groups are mostly in the *tg* (*trans-gauche*) conformation in the bulk, hydroxymethyl groups at the cellulose-water interface are *gg* in cellulose I and *gt* in cellulose II, where *t* and *g*

describe the relative geometries of the -OH group in the hydroxymethyl group with C4 and O5 in the same glucose residue. The numbers assigned to each atom follow from the numbering system described in the nomenclature section of this paper.

This paper deals with the energies and inter-residue interactions resulting from the rotations of the free hydroxyl groups (*viz.* C2, C3, and C6) of the central residue of a cellulose Iα fragment consisting of nine cellotriose chains arranged in a crystalline fashion as described in Nishiyama *et al*[7,8]. The actual structure was taken from molecular dynamics snapshots representing the crystalline structure of cellulose Iα. The rest of the paper will present computational methods used to obtain the energies and inter-residue interactions, followed by the relationship of these energies to the intra- and interchain hydrogen bonding network and to the repulsive interactions found as a result of the rotations of the dihedral around each of the free –OH groups studied.

Computational Methods

All calculations were performed with the electronic structure code GAMESS,[11] which is freely available from Iowa State University. Structures were visualized with MacMolPlt,[12] a graphical interface to GAMESS. Input files used to calculate the single-point energies using the fragment molecular orbital (FMO) method were prepared with the aid of Facio,[13] a 3-D graphics program for molecular modeling and visualization of quantum chemical calculations, such as molecular orbitals and normal mode vibrations.

The FMO[14,15,16] method is used to decrease the overall cost of high-level *ab initio* quantum chemical calculations for very large molecular systems. FMO basically divides a large molecule into fragments or monomers. For example, a polysaccharide may be broken

into its monosaccharide components. The fragmentation is accomplished so that electrons in a covalent bond at the point of fragmentation are intuitively assigned on one of the atoms originally making up the bond to be broken. Usually, this means that the more electronegative atom in the bond gets assigned the two electrons. As a result the overall charge of the system being fragmented is maintained throughout the calculation.
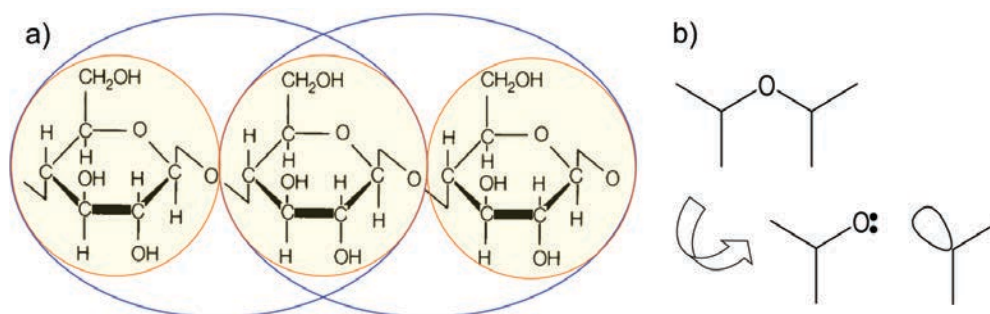
Figure 1. a) The FMO method applied to cellulose. Each glucose residue represents a monomer in FMO (yellow circles), and each pair of glucose molecules forms a dimmer (blue ovals). b) Fractioning of a glycosidic bond in FMO. The oxygen atom is considered a bond acceptor, thus both electrons belong to oxygen; and the carbon atom is considered a bond acceptor, thus carbon has a vacant orbital.

Electronic structure equations for each monomer and pair of fragments (dimers) are solved under the environmental electrostatic field until they converge self-consistently. The results are then used to calculate the total energy of a system:

$$E = \sum_{I>J}^{N} E_{IJ} - (N-2) \sum_{I}^{N} E_I \qquad (1)$$

where N, E_I and E_{IJ} are the number of fragments, total energy of the monomer and that of the dimer, respectively. FMO has been used to calculate energies and inter-fragment interactions for many chemical systems including those involving solvated polypeptides[17], proteins[18], and polysilanes[19]. The above equation for the total energy can also be rewritten as:

$$E = \sum_{I>J}^{N} \Delta E_{IJ} - \sum_{I}^{N} E'_{I} \qquad (2)$$

$$\Delta E_{IJ} = \left(E'_{IJ} - E'_{I} - E'_{J}\right) + Tr(\Delta P^{IJ} V^{IJ}) \qquad (3)$$

where ΔP^{IJ}, V^{IJ}, E'_{I}, and E'_{IJ} are the density difference matrix, the environmental electrostatic potential for dimer IJ, monomer energy without environmental electrostatic potential, and dimer energy without environmental electrostatic potential, respectively. ΔE_{IJ} is the interaction energy between the Ith and Jth fragments, and is defined as the pair interaction energy (PIE), also known in the literature as the inter-fragment interaction energy (IFIE). PIEs are obtained from the same FMO calculations, and give useful insight on inter- and intramolecular interactions. The PIE is divided into electrostatic, exchange-repulsion, charge transfer, and dispersion contributions during the pair interaction energy decomposition analysis (PIEDA) of an FMO run. An FMO analysis with PIEDA calculates PIE by subtracting from each of the calculated dimer energies the converged energy calculated for each of the dimer's component monomers. This allows identification of specific pair interactions that contribute significantly to changes in the total energy of a system made up of several residues such as the one examined in this study.[20]

Single-point energies were calculated for conformers resulting from the rotations of free non-anomeric hydroxyl groups (*viz.* C2, C3, and C6) in the central glucose residue of a fragment consisting of nine cellotriose strands taken from the cellulose Iα crystal as described by Heiner and Teleman.[21] Taken together, these nine strands do not make up a unit cell of the cellulose Iα crystal, but this assembly includes interactions a glucose residue has

with its 26 closest neighbors. Rotations of the free -OH groups were performed in increments of 30° and the energies were calculated using RHF and DFT (B3LYP[22]) with the 6-31G(d,p)[23] basis set, and FMO/MP2[24] with the 6-31G(d) basis set. Reoptimization of the cellulose fragment was not performed to obtain an upper bound of the energy resulting from the rotation of each -OH group and to eliminate any effect the rotations may have on the overall crystal structure.

Single-point energies for conformers resulting from 30-degree rotations of the corresponding hydroxyl groups in free glucose were also calculated using RHF and B3LYP with the 6-31G(d,p) basis set for comparison purposes.

Nomenclature

Each residue in the cellulose Iα fragment is numbered according to the scheme shown in Figure 2. Each circle represents a glucose residue and circles linked by lines refer to glucose units joined by a 1,4-glycosidic bond and are said to form chains. Chains found in the same plane form a sheet of glucose units in the cellulose Iα structure.
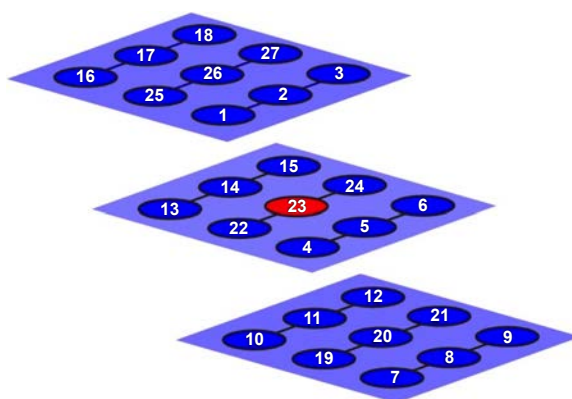
Figure 2. Numbering scheme used to identify glucose residues in the cellulose fragment.

Oxygen atoms in glycosidic linkages between residues are referred to by the residue numbers being connected by such atoms. The carbon atoms in each residue are numbered beginning with the anomeric carbon up to the hydroxymethyl carbon (*viz.* C1 through C6 of each residue). Oxygen atoms which are not part of glycosidic links are numbered based on the number of the carbon to which the oxygen is attached (*viz.* O2 through O6 of each residue). Hydrogen atoms directly attached to carbons also follow the same numbering scheme, including the hydrogen attached to each residue's anomeric carbon (*viz.* H1 through H6 of each residue). Hydrogen atoms that are part of hydroxyl groups are referred to explicitly as such.
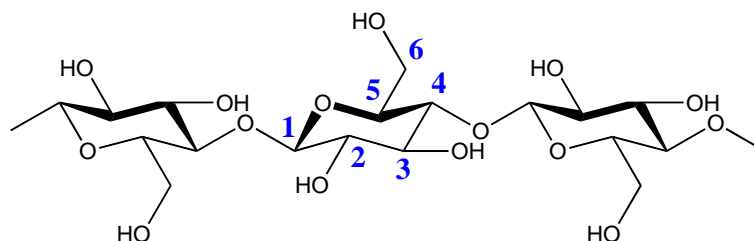
Figure 3. Nomenclature used to distinguish carbon, oxygen, and hydrogen atoms in each of the glucose residues of cellulose Iα.

Results and Discussion

Pair Interaction Energy Decomposition Analysis (PIEDA)

A plot of the absolute values of the pair interaction energies against relative separation between residues in the equilibrium state of cellulose Iα is shown in Figure 3. The relative separation between each pair of residues, *A* and *B*, is the closest distance among all pairs of atoms *i* and *j* in *A* and *B*, respectively, divided by the sum of the van der Waals radii, *W*, of the atoms in the closest pair[19]:

$$R_{AB} = \min_{i \in A, j \in B}[R_{ij}], R_{ij} = \frac{|R_i - R_j|}{W_i + W_j} \tag{4}$$

Since attached dimers (such as those within a chain) are treated differently in PIEDA, the relative separation is given as zero and thus, the intrachain interactions are not included in Figure 4. Only those interactions of Residue 23 with glucose units on each of the other chains are considered in Figure 4. It can be seen that the magnitude of PIE between two residues is a function of the relative separation between them, decreasing by a factor of about $1/R^3$ for every unit increase in the relative separation when fit by one functional parameter.
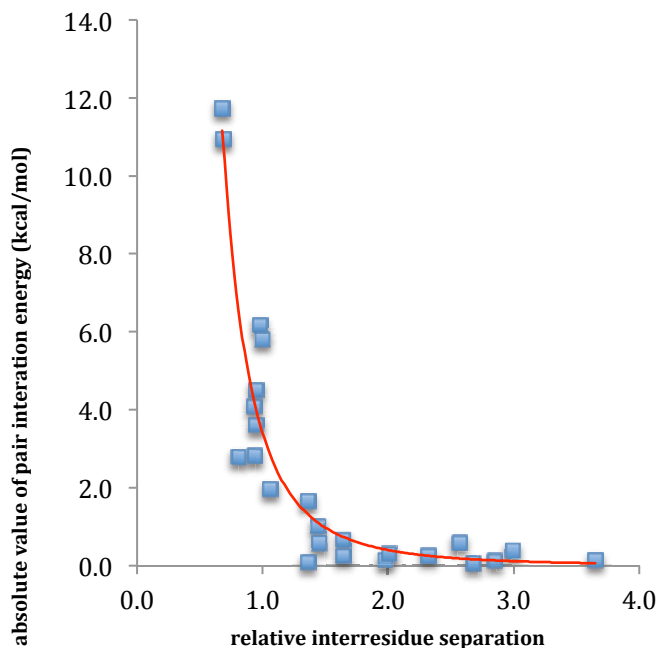
Figure 4. Pair interaction energies in the equilibrium state for cellulose Iα.

The components of the pair interaction energies with the central residue, namely electrostatic, exchange, charge transfer, and dispersion energies, are listed in Table 1. PIE values less than zero imply attractive interactions between the residues involved while positive PIEs correspond to repulsive interactions. Intrachain interactions (those between residues that are adjacent to each other in the same cellulose chain) are predominantly

electrostatic in nature, with other components contributing a total of about 3% of the total PIE. Among residues within the shortest relative separation from Residue 23 (less than or equal to 1.00), electrostatic and exchange energies are among the major components of the interaction energies, although both charge transfer and dispersion energies also contribute significantly to the overall value of PIE and cannot be ignored. Exchange energies vanish when the relative separation between residues approaches a threshold value of 1.50. Interactions between residues for which the relative separation approaches 2.00 become more electrostatic in nature, with other types of interaction contributing less to the overall PIE value. Interactions between residues with relative separation greater than 2.00 are exclusively electrostatic.

Table 1. Pair interaction energy decomposition analysis between Residue 23 (the central residue) and each of the other residues in the cellulose Iα fragment at equilibrium.

Residue Number	Relative interresidue separation	Pair interaction energies (kcal/mol)				
		Total	Electrostatic	Exchange	Charge transfer + mix	Dispersion
1	1.640	0.664	0.854	-0.001	-0.011	-0.179
2	0.950	-4.509	-0.873	2.393	-2.270	-3.760
3	1.060	-1.971	-0.002	0.612	-0.379	-2.202
4	1.440	-1.008	-0.615	0.003	-0.141	-0.255
5	0.690	-10.948	-13.853	11.996	-4.336	-4.755
6	1.450	-0.575	-0.309	0.002	-0.070	-0.197
7	2.990	-0.390	-0.390	0.000	0.000	0.000
8	2.850	-0.122	-0.122	0.000	0.000	0.000
9	2.680	-0.058	-0.058	0.000	0.000	0.000
10	1.980	0.136	0.235	-0.001	-0.005	-0.093

11	0.940	-2.825	-2.543	3.472	-0.074	-3.680
12	0.930	-4.091	-1.374	4.433	-2.968	-4.182
13	1.360	-1.649	-0.950	0.018	-0.268	-0.450
14	0.680	-11.728	-15.314	13.325	-4.819	-4.920
15	1.360	0.093	0.654	0.020	-0.194	-0.386
16	3.650	-0.145	-0.145	0.000	0.000	0.000
17	2.320	0.257	0.257	0.000	0.000	0.000
18	2.570	-0.594	-0.594	0.000	0.000	0.000
19	1.640	-0.261	-0.113	-0.001	-0.008	-0.140
20	0.980	-6.168	-2.088	2.560	-2.317	-4.323
21	1.000	-5.806	-2.610	1.457	-1.135	-3.518
22	0.000	-9581.103	-9305.634	-98.517	-133.912	-43.040
24	0.000	-9518.824	-9247.848	-110.645	-121.895	-38.437
25	2.010	-0.306	-0.306	0.000	0.000	0.000
26	0.950	-3.612	0.143	3.864	-2.622	-4.997
27	0.810	-2.789	-0.053	4.044	-2.475	-4.305

C2 Hydroxyl Group Rotation

The relative energies arising from rotating the C2-OH dihedrals in free glucose and in the central residue (Residue 23) of the cellulose Iα fragment are shown in Figure 5. It can be seen that these curves have similarities with each other as they are both characterized by the presence of two energy maxima and two energy minima. Both sets of energy curves are also characterized by excellent agreement between the relative energies obtained using B3LYP and MP2 levels of theory, with relative energies calculated using each theory differing by about 0.5 kcal/mol at each point. It is also easy to see that relative energy changes for the

cellulose Iα fragment are about three times those of the free glucose molecule. This is to be expected as the rotating hydrogen in cellulose encounters significantly more interactions with other atoms than would affect the rotation of a corresponding hydrogen in free glucose.

The relative energy curves for the cellulose Iα fragment may be explained by examining the relative pair interaction energies (PIEs) that occur between each pair of residues in the fragment as each free hydroxyl group is rotated. In this study, the relative PIE for each pair of residues prior to the rotation of the hydroxyl groups in Residue 23 was taken to be the zero of the interaction between these two residues. The initial value of the PIE is then subtracted from values resulting from the rotation of a hydroxyl group by various angles to give a relative PIE curve corresponding to the rotation of each hydroxyl group.

Figure 5. Relative energy curves arising from the rotation of the C2-OH dihedral in free glucose (top) and Residue 23 in the cellulose fragment (bottom).

A positive change in the relative PIE indicates either an increasingly repulsive or a decreasingly attractive interaction between the residues being considered. The opposite, decreasingly repulsive or increasingly attractive interaction between fragments, is true for negative values of the relative PIE between such residues.

Examples of relative PIE curves of Residue 23 with other residues are shown in Figure 6. Interaction with Residue 22 is an example of an intrachain interaction (between adjacent residues of the same chain) while interaction with Residue 5 is an interchain interaction (between residues coming from adjacent chains of the same sheet). The interaction between Residues 23 and 20 is an intersheet interaction which occurs between residues from different sheets of glucose residues in the cellulose structure. These curves each illustrate the largest change in the relative PIE values obtained for each kind of interaction with Residue 23 as the C2-OH group is rotated.

Figure 6. Important relative pair interaction energy curves arising from the rotation of the C2-OH dihedral in Residue 23 of the cellulose fragment.

The intrachain interaction between Residues 22 and 23 is more important than the one between Residues 23 and 24. This follows from the fact that the C2-OH group in Residue 23 occurs closer to Residue 22. This interaction is attractive in nature due to a hydrogen bond between these two fragments when the C2-OH group in Residue 23 is at its initial position (see Figure 7). At this initial position, the rotating hydrogen is found to be 1.75 Å from the hydroxymethyl oxygen atom in Residue 22. This compares with the same intrachain O2-H···O6 hydrogen bond distance of 1.752 Å in cellulose Iα measured by Nishiyama et al[4]. This intrachain hydrogen bond interaction weakens as the C2-OH rotation proceeds due to an increase in the internuclear distance between the rotating hydrogen and the O6 atom in Residue 22. Consequently, an increase in the relative PIE between Residues 22 and 23 occurs. This internuclear distance goes to as much as 2.94 Å after a rotation of the C2-OH group in Residue 23 is rotated by an angle of 90 degrees. The PIE maximum occurs after a 210-degree rotation when the rotating O-H bond comes into an eclipsing interaction with the C2-H2 bond also in Residue 23. This 210-degree rotation is accompanied by an increase of 24.5 kcal/mol in the relative PIE of the cellulose structure. This is the largest change in the relative PIEs between Residue 23 and each of the other residues as the C2-OH group in Residue 23 is rotated. None of the hydrogen bonds between Residues 23 and 24 is affected by rotation of the C2-OH group in Residue 23, resulting in a maximum PIE change of only 0.5 kcal/mol between these two residues.

The interaction between Residues 23 and 5 is an example of how interchain interactions can also significantly contribute to the total energy changes the cellulose fragment goes through as the C2-OH dihedral in Residue 23 is rotated (Figure 8). Residue 5 is the closest one among the three glucose units in the chain found on the same side of the cellulose sheet

that contains Residue 23 as the rotating hydrogen atom. The relative PIE for this interaction was calculated to decrease by as much as 5.6 kcal/mol following rotation of the C2-OH group in Residue 23 by an angle of 150 degrees from its original position.
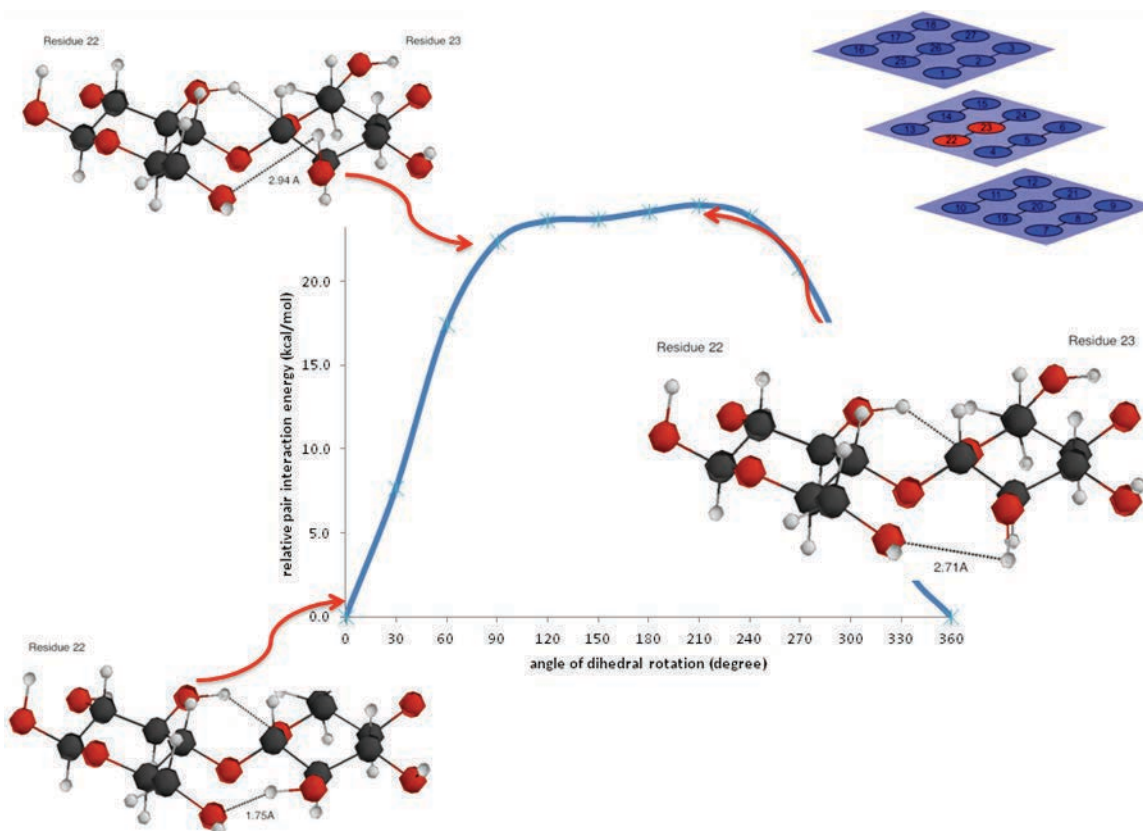
Figure 7. Interaction between Residues 23 and 22 as the C2-OH dihedral in Residue 23 is rotated.

The rotating hydrogen atom is initially found to be 4.17 Å from the O6 atom in Residue 5, which is on a cellotriose chain in the same cellulose sheet as Residue 23. As the hydrogen atom in Residue 23 is moved, it comes close to the O6 atom in Residue 5, giving rise to a hydrogen bonding interaction between these two residues. This internuclear distance is at a minimum of 2.21 Å after the hydrogen in Residue 23 is rotated by an angle of about 150 degrees. Further rotation of the C2-OH group in Residue 23 weakens the attraction with Residue 5 and in fact, the interaction becomes more repulsive as the rotating hydrogen comes

close to another hydrogen atom in the hydroxymethyl group of Residue 5. The smaller change in interaction energy between Residues 23 and 5 is consistent with conclusions from earlier studies that while hydrogen bonding dominates interchain interactions,[10] they are not as strong as intrachain interactions. The other closest glucose unit in the same sheet to Residue 23 is Residue 14, which is found on the side opposite to Residue 5. Interaction with between Residues 23 and 14 is affected by the rotating hydrogen to the extent of less than 0.1 kcal/mol.
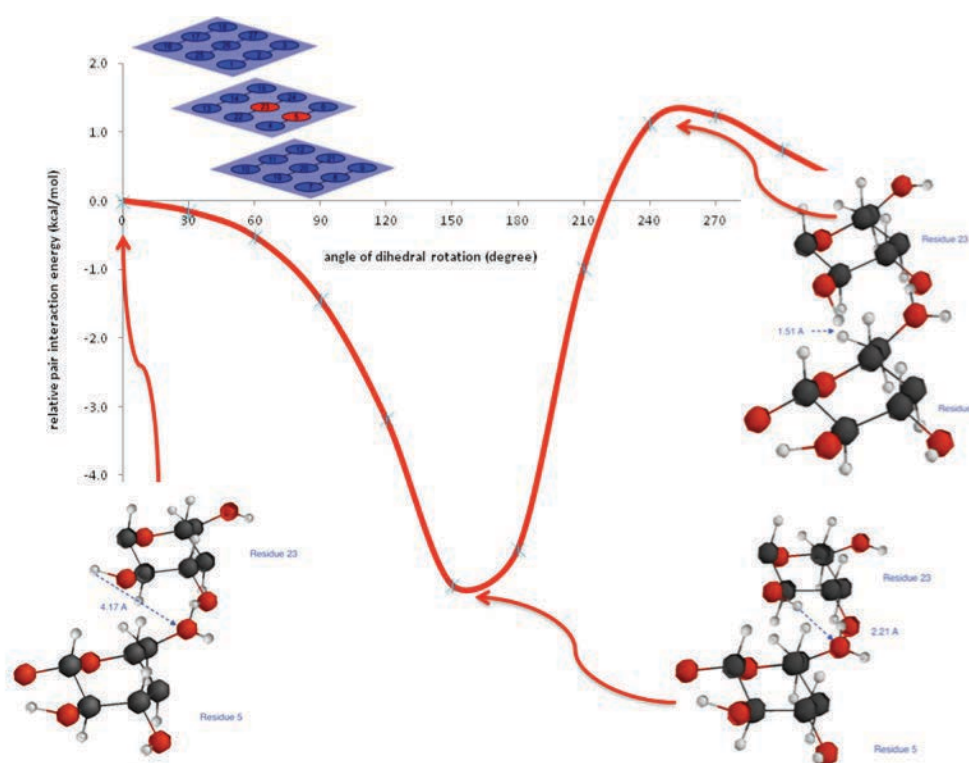
Figure 8. Interaction between Residues 23 and 5 as the C2-OH dihedral in Residue 23 is rotated.

An example of how intersheet interactions are affected by rotating free -OH groups in cellulose is seen with Residue 23's interaction with Residue 20, the central glucose unit in the chain below residue 23 (Figure 9). Rotating the C2-OH group in Residue 23 shortens the distance between the rotating hydrogen and the H5 atom in Residue 20 by about 1.0 Å and

brings about repulsive interactions between the two residues. The repulsion results in an increase in the relative PIE value between Residues 23 and 20 of about 2.6 kcal/mol and is thus seen to be a relatively unimportant contributor to the total energy changes for the cellulose fragment as the C2-OH group in Residue 23 is rotated. This is consistent with earlier studies indicating the absence of intersheet hydrogen bond interactions in cellulose[7, 8] and that intersheet interaction energies are much smaller than interchain interaction energies.[10]
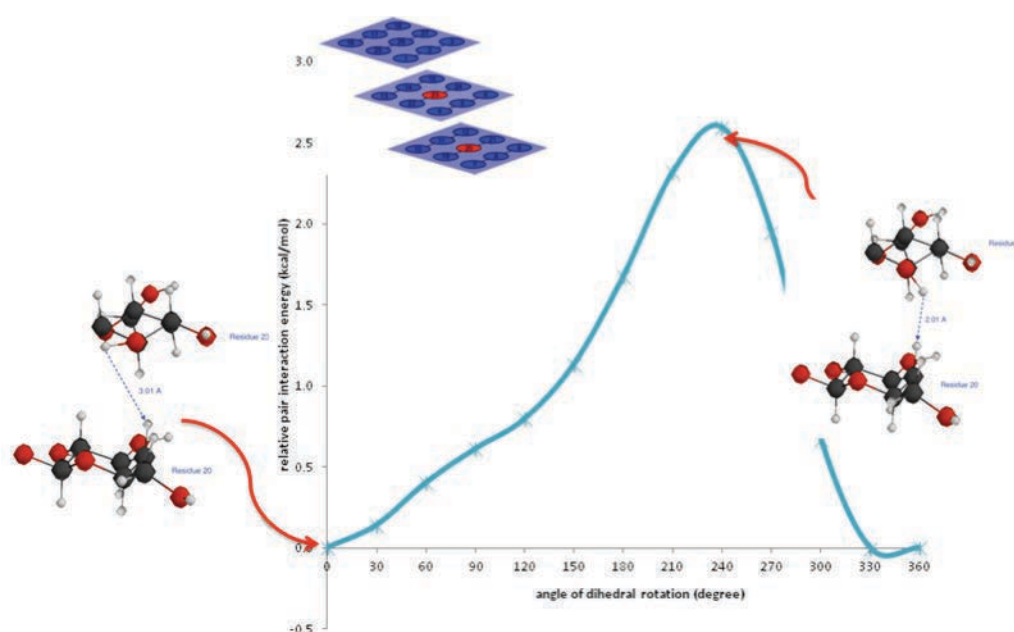
Figure 9. Interaction between Residues 23 and 20 as the C2-OH dihedral in Residue 23 is rotated.

C3 Hydroxyl Group Rotation

As with the rotation of the C2-OH group, rotating the C3-OH dihedral results in relative energy curves for free glucose and cellulose Iα that have characteristics that are similar to each other (Figure 10). Relative energies calculated using B3LYP and MP2 levels of theory differed by an average of 0.5 kcal/mol at each angle of rotation. Relative energy changes in

cellulose as the C3-OH dihedral is rotated are also seen to be approximately three times those that occur as the corresponding dihedral in free glucose is rotated. However, in this case, the relative height of the two maxima change from the gas phase to the crystal.
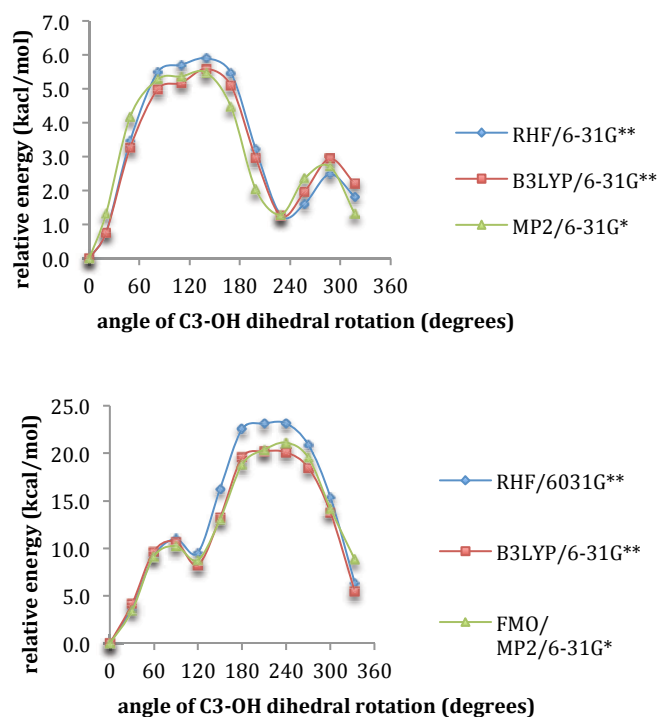
Figure 2. Relative energy curves arising from the rotation of the C3-OH dihedral in free glucose (top) and Residue 23 in the cellulose fragment (bottom).

In terms of changes in relative PIE values, Residues 24 (intrachain) and 5 (interchain in the same sheet) are seen to be the main contributors to changes in the total relative energy of the cellulose Iα fragment as the C3-OH group in Residue 23 is rotated (Figure 11). This again clearly illustrates the greater relative importance of intrachain and interchain interactions over those which are intersheet in determining the energy changes involved as

free -OH groups in cellulose are rotated. This also demonstrates the point that the more important pair interactions among glucose residues in cellulose tend to be with those residues within the immediate vicinity of the -OH dihedral that is being rotated.
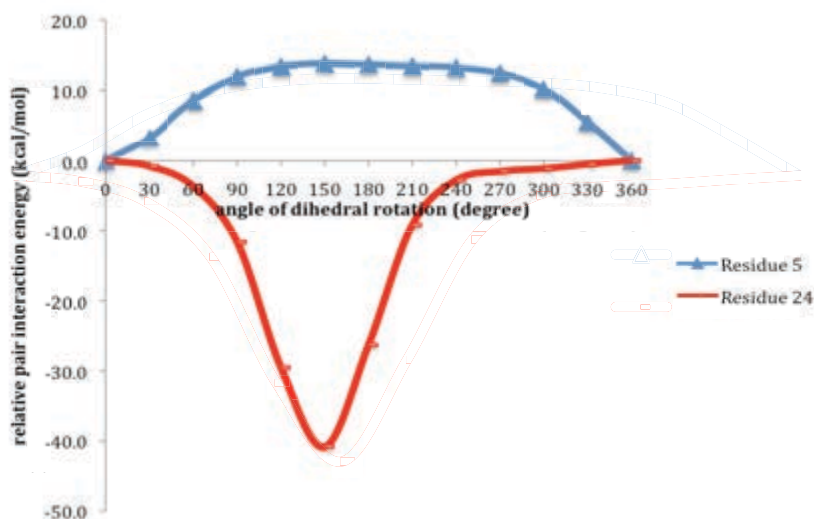
Figure 3. Important relative pair interaction energy curves arising from the rotation of the C3-OH dihedral in Residue 23 of the cellulose fragment.

Residue 24 interacts with Residue 23 in an increasingly attractive manner as the C3-OH group in the latter is rotated. This is due to an O3-H···O5 hydrogen bond interaction that arises as the rotating hydrogen atom is brought closer to O5 in Residue 24. When the rotating hydrogen is 150 degrees away from its initial position, the internuclear distance between the O3-H in Residue 23 and O5 in Residue 24 goes to a minimum of 1.61 Å, which is shorter than the 1.95 Å reported by Nishiyama. This calculated relative PIE value between Residues 23 and 24 decreases by about 40.8 kcal/mol and the intrachain interaction is seen to be the most important one to arise from rotating the C3-OH group in Residue 23. Structures related to the relative PIE critical values between Residues 23 and 24 are shown in Figure 12.

Since Residue 22 is further away from the C3-OH group in Residue 23 than Residue 24 is, intrachain interaction between the first two residues changes by a maximum of 2.5 kcal/mol when the O3-H in Residue 23 is rotated.
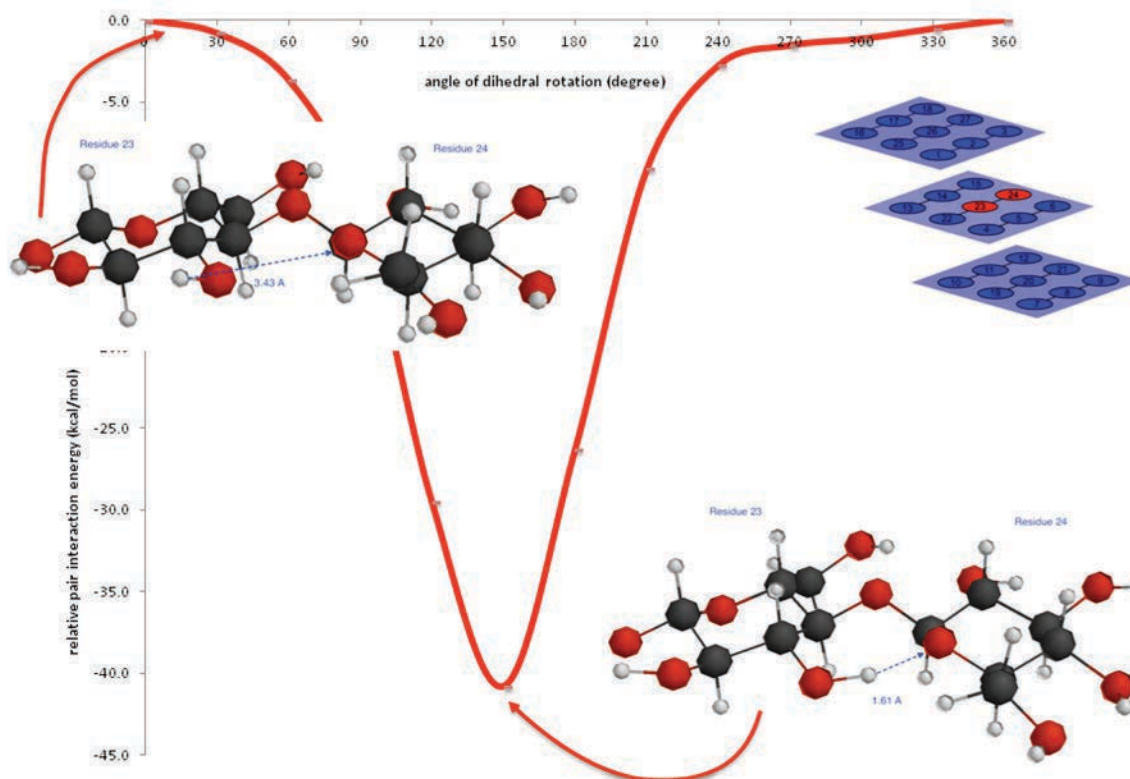
Figure 4. Interaction between Residues 23 and 24 as the C3-OH dihedral in Residue 23 is rotated.

The C3-OH group in Residue 23 also initially forms an interchain hydrogen bonding interaction with the oxygen atom in the hydroxymethyl group in Residue 5 (Figure 13). The internuclear distance between these two atoms is initially about 1.62 Ắ and this distance goes to a maximum of about 3.58 Ắ as the C3-OH dihedral in Residue 23 is rotated by an angle of about 120 degrees. This results in an increase of 13.9 kcal/mol in the relative PIE between Residues 23 and 5. Further rotation gradually brings the rotating hydrogen atom closer to the O6 atom in Residue 5, which brings relative PIE between these two residues back to zero.
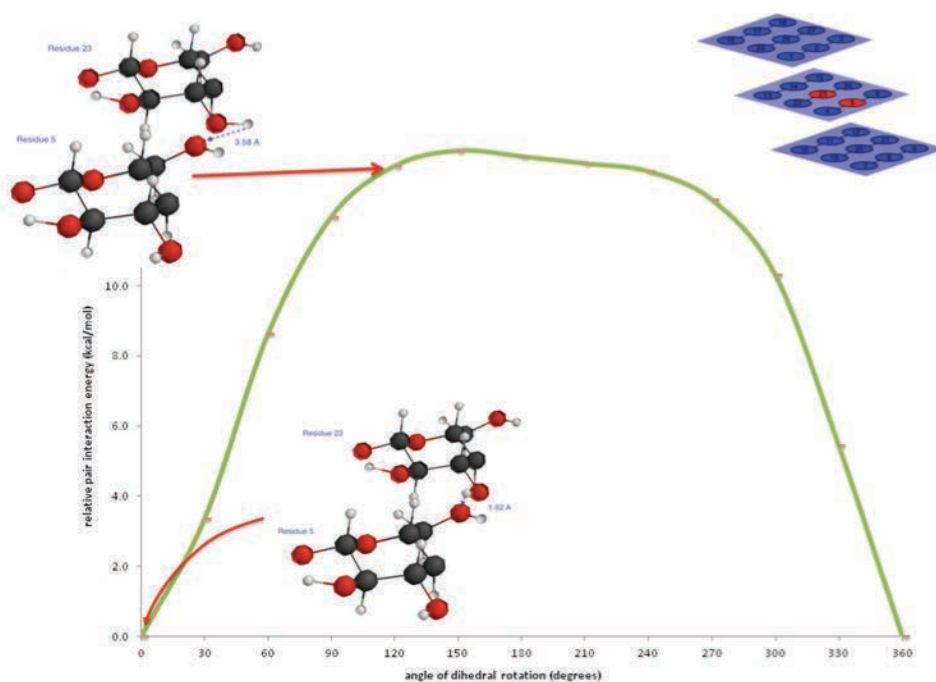
Figure 5. Interaction between Residues 23 and 5 as the C3-OH dihedral in Residue 23 is rotated.

C6 Hydroxyl Group Rotation

Rotations of the C6-OH group in both free glucose and Residue 23 in the cellulose Iα fragment resulted in the largest calculated changes in the relative total energy for any hydroxyl rotation (Figure 14).

Prior to rotation of the hydrogen atom in the hydroxymethyl group of free glucose, an O6-H···O4 intramolecular hydrogen bond interaction is suggested due to the short internuclear distance between the O6-H and O4 atoms, which is about 2.05 Å. Rotating the O6-H atom not only disrupts this attractive force but also introduces repulsions with the O4-H atom as the distance between these two atoms is reduced to about 1.12 Å (Figure 15).
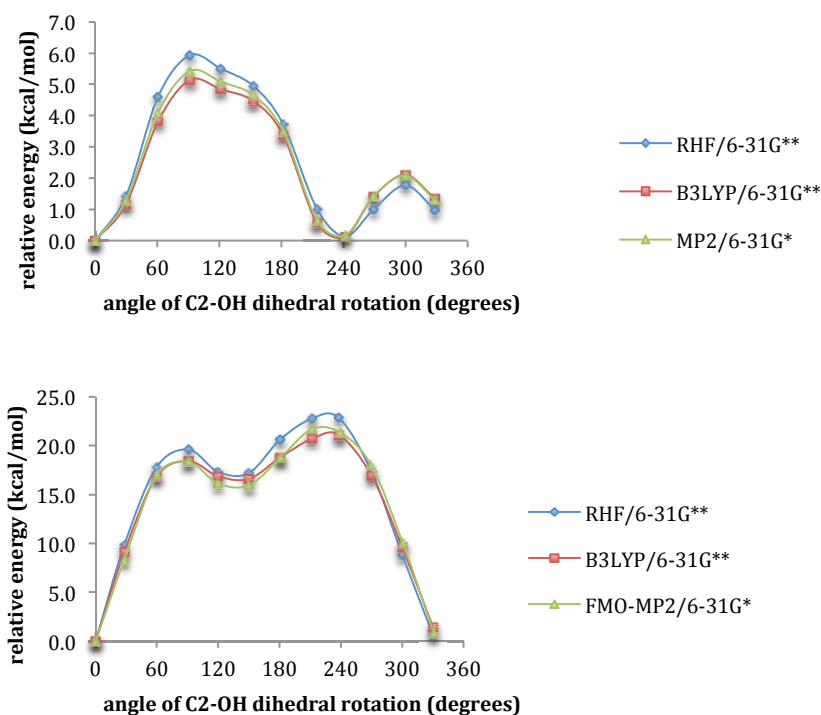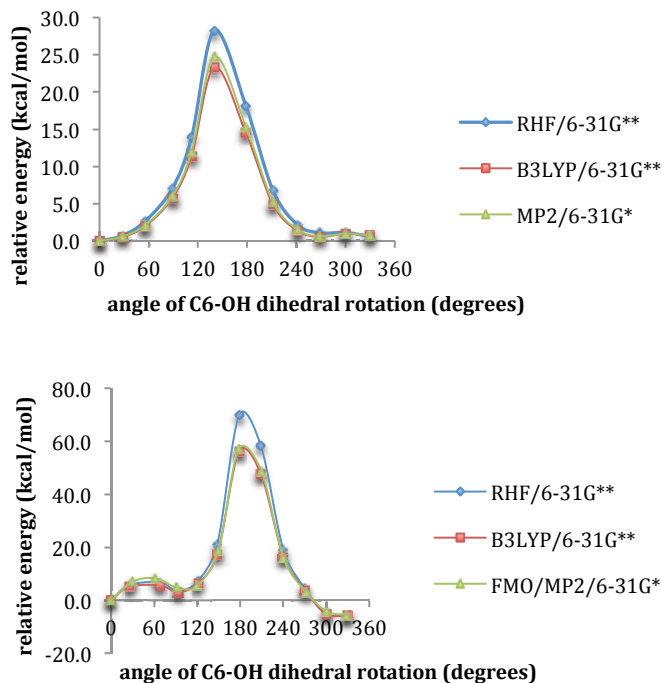
Figure 6. Relative energy curves arising from the rotation of the C6-OH dihedral in free glucose (top) and Residue 23 in the cellulose fragment (bottom).

In the cellulose Iα fragment, the hydroxymethyl group extends out of the general region of space occupied by each glucose unit in the cellulose crystal structure. Thus, the hydroxymethyl group is expected to encounter more interactions with groups of atoms from other glucose units. The rotations of the C6-OH group in cellulose Iα resulted in relative energies that differed from each other by an average of 1.1 kcal/mol at each angle of rotation when calculated with either B3LYP or MP2 level of theory (Figure 14).
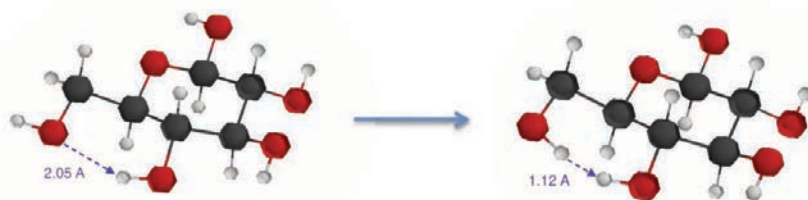
Figure 7. Conformations of the hydroxymethyl group in free glucose at equilibrium (left) and following a 150-degree rotation of the O6-H atom (right).
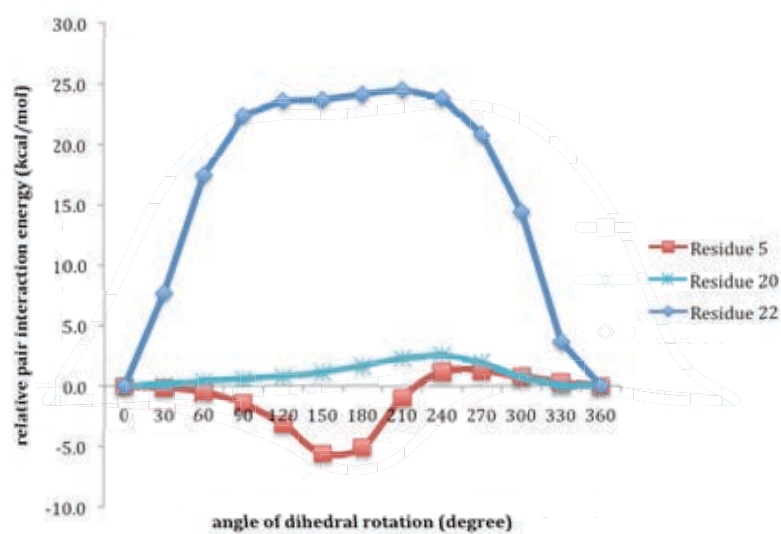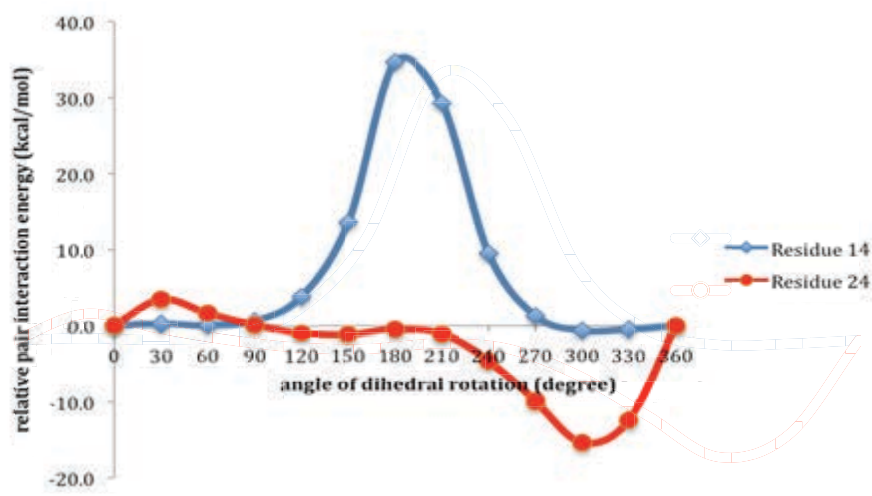
Figure 8. Important relative pair interaction energy curves arising from the rotation of the C6-OH dihedral in Residue 23 of the cellulose fragment.

Changes in the relative energy of the cellulose Iα fragment for the C6-OH rotation are seen to be mainly due to changes in the interaction energies of Residue 23 with Residues 14 (interchain) and 24 (intrachain) as shown in Figure 16. In particular, rotation of the C6-OH group in Residue 23 by an angle of 180 degrees from its original position in cellulose Iα was calculated to bring about an increase of about 34.7 kcal/mol in the relative PIE between Residues 23 and 14. This is explained by a decrease in the internuclear distance between the rotating hydrogen and the hydrogen atom attached to O3 in Residue 14 from 2.64 Ǻ down to 0.75 Ǻ, which brings about significantly increased repulsive interactions between these two residues (Figure 17).

On the other hand, rotation of the C6-OH dihedral in Residue 23 by an angle of 300 degrees from its initial position results in a more attractive interaction with Residue 24, indicated by a decrease in the relative PIE between these two residues by about 15.4

kcal/mol, as the distance between the rotating hydrogen atom and the O2 atom in Residue 24 is decreased from 2.05 Å down to 1.83 Å (Figure 18). This shorter internuclear distance falls well within the range of typical internuclear distances between hydrogen and oxygen atoms that exhibit hydrogen bonding between them.[25]
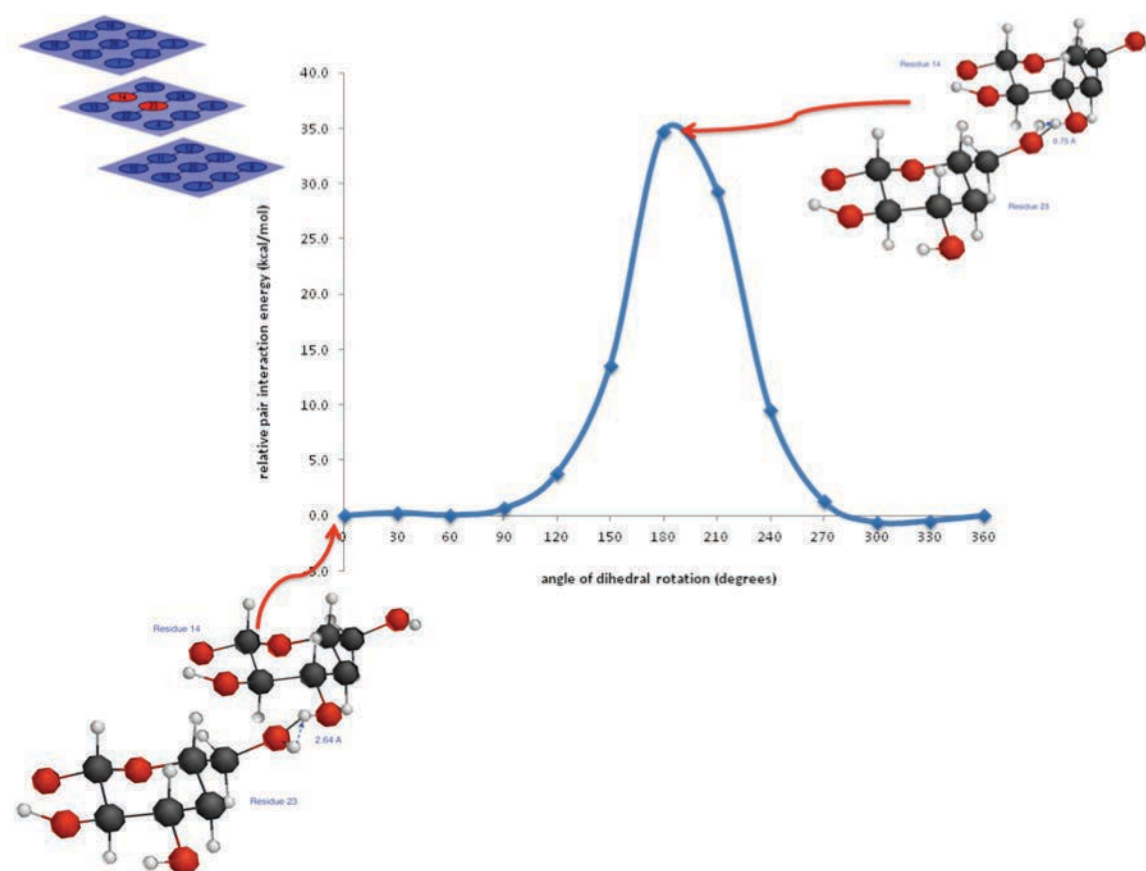
Figure 9. Interaction between Residues 23 and 14 as the C6-OH dihedral in Residue 23 is rotated.
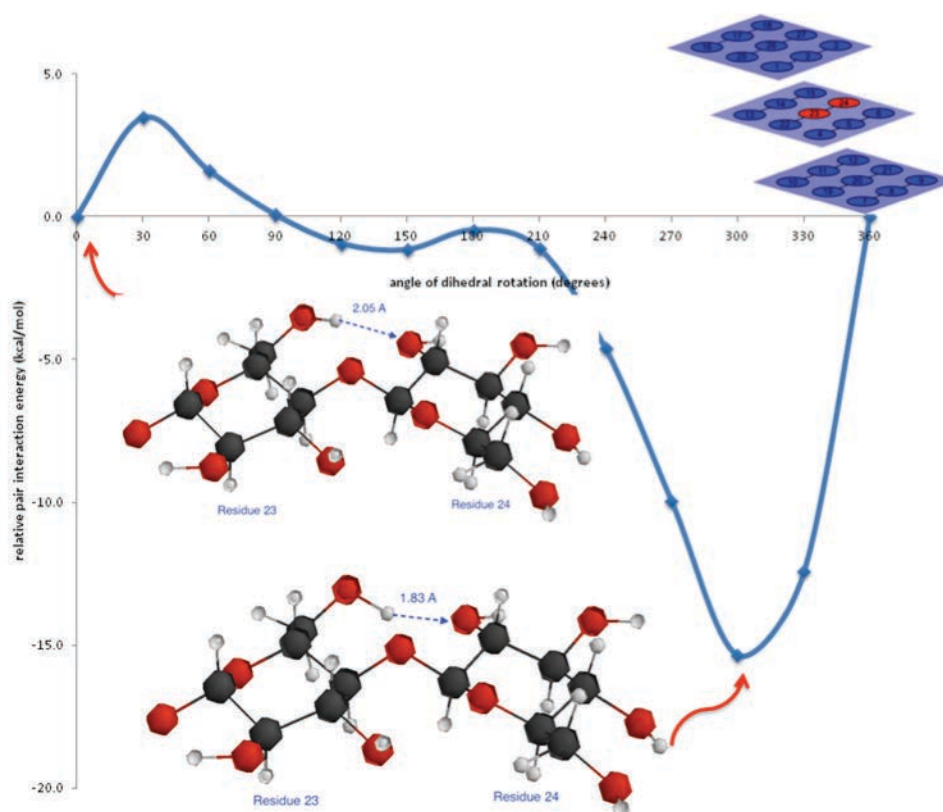
Figure 10. Interaction between Residues 23 and 24 as the C6-OH dihedral in Residue 23 is rotated.

Summary

This study demonstrates that rotations of free –OH groups attached to C2 and C3 atoms in glucose units are freer compared to those bonded to the C6 atom because the latter tends to stick out of the glucose residue and is, therefore, more prone to interactions with atoms belonging to other glucose units in cellulose. Calculations also reflect the relative importance of intrachain and interchain interactions, which occur either to maximize hydrogen bonding interactions or to minimize repulsions among atoms of the various glucose residues, over interactions that occur between sheets of glucose residues. Also seen in this study is the excellent agreement between relative energies calculated using B3LYP and MP2 levels of theory, which are within 1.0 kcal/mol of each other as the various –OH dihedrals are

rotated. This will be very useful for future studies that require more extensive calculations than the already expensive calculations presented here. The use of FMO2/MP2 with PIEDA specifically showed that intrachain interactions are overwhelmingly larger than those that are interchain or intersheet in nature. However, this analysis also showed that all of the interactions must be taken into account to understand the subtle play among the different types of molecular interactions. Interresidue distance is also an important factor both in determining the size of the interaction energy as well as the nature of the energy's contributing components. Electrostatic and exchange energies account for the large parts of interaction energies when the relative separation between residues is small. However, the charge transfer and dispersion contributions cannot be ignored. Exchange energies vanish as the relative separation between glucose units in cellulose approaches 1.50. Interaction energies become purely electrostatic in nature when the relative separation between glucose units in cellulose are greater than 2.00.

[1] Carroll, A. and Somerville, C. *Annu. Rev. Plant Biol.* **2009**, *60*, 165-182.

[2] Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W. and Foust, T. D. *Science*. **2007**, *315*, 804-807.

[3] Zhong, L.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L.; Adney, W. S.; Himmel, M. E. and Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984-1992.

[4] Petersen, L.; Ardevol, A.; Rovira, C. and Relly, P. J. *J. Phys. Chem. B.* **2009**, *113*, 7331-7339.

[5] Horii, F.; Hirai, A. and Kitamaru, R. *Macromolecules*. **1987**, *20*, 2117-2120.

[6] Nishiyama, Y.; Langan, P.; Chanzy, H.; *J. Am. Chem. Soc.* **2002**, *124*, 9074-9082.

[7] Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300-14306.

[8] Qian, X. *Molecular Simulation.* **2008**¸ *34*, 183-191.

[9] Qian, X.; Ding, S.-Y.; Nimlos, M. R.; Johnson, D. K. and Himmel, M. E. *Macromolecules*, **2005**, *38*, 10580-10589.

[10] Newman, R. H. and Davidson, T. C. *Cellullose*, **2004**, *11*, 23-32.

[11](a) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen,J. H.; Koseki, S.; Matsunaga,N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347. (b) Gordon, M. S.; Schmidt, M. W. *Theory and Applications of Computational Chemistry, the first forty years;* Elsevier: Amsterdam, 2005. The GAMESS computational chemistry package may be obtained at http://www.msg.chem.iastate.edu/.

[12] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Modell.* **1998**, *16*, 133.

[13] (a) Suenaga, M. *J. Comput. Chem. Jpn.*, **2008**, *1*, 33. (b) Suenaga, M. *J. Comput. Chem. Jpn.*, **2005**, *1*, 25.

[14] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano,T. and Uebayasi,M. *Chem. Phys. Lett.* **1999**, *313,* 701–706.

[15] Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M. and Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475-480.

[16] Fedorov, D. G. and Kitaura, K. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, Fragment molecular orbital method: an approximate computational method for large molecules, in Starikov, E. B.; Lewis, J. P. and Tanaka, S. (Eds.). *Modern Methods for Theoretical Physical Chemistry of Biopolymers*. Elsevier. **2006**, 3-38.

[17] Nagat, T.; Fedorov, D. G.; Sawada, T.; Kitaura, K. and Gordon, M. S. *J. Chem. Phys.* **2011**, *134*, (034110)-1-12.

[18] Fedorov, D. G.; Ishida, T.; Uebayashi, M. and Kitaura, K. *J. Phys. Chem. A.* **2007**, *111*, 2722-2732.

[19] Ishikawa, T.; Mochizuki, Y.; Imamura, K.; Nakano, T.; Mori, H.; Tokiwa, H.; Tanaka, K.; Miyoshi, E. and Tanaka, S. *Chem. Phys. Lett.* **2006**, *430*, 361-366.

[20] Fedorov, D. G.; Kitaura, K. *J. Comput. Chem.*, **2007**, *28*, 222.

[21] Heiner, A. P.; Teleman, O. *Langmuir*, **1997**, *13*, 511.

[22] (a) Becke, A. D. *J.Chem.Phys.*, **1993**, *98*, 5648. (b) Stephens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *J.Phys.Chem.*, **1994**, *98*, 11623. (c) Hertwig, R. H.; Koch, W. *Chem.Phys.Lett.*, **1997**, *268*, 345.

[23] (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J.Chem.Phys.*, **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J.Chem.Phys.*, **1972**, *56*, 2257.

[24] (a) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A*, **2007**, *111*, 6904. (b) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.*, **2004,** *121*, 2483. (c) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. *J. Comput. Chem.*, **2004**, *25*, 872.

[25] Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, U. K., **1997**.

CHAPTER 3: Cellobiose Hydrolysis: Break It Down!

John Y. Baluyut and Theresa L. Windus

Abstract

Restricted Hartree-Fock (RHF) and density functional theory (DFT) methods were used to determine the energies involved in the acid-catalyzed hydrolysis of cellobiose in the gas phase. A stepwise mechanism for the reaction was used to determine the different species involved. The initial step was protonation of a cellobiose molecule with a hydronium ion, which was followed by removal of a molecule of water to produce protonated cellobiose. Dissociation of the protonated cellobiose followed to produce a β-D-glucose molecule and a glucosyl cation. The cation in turn was hydrated to produce an α-D-glucose molecule. The energy change for the dissociation step was determined to be +36.8 kcal/mol using density functional theory with the B3LYP functional and 6-311+G(d,p) basis set. The calculated value is similar to those obtained from experimental data and from a recent solution phase Car-Parrinello molecular dynamics calculation.

Introduction

Cellulosic biomass is an abundant renewable source of material for producing biofuels. It consists mainly of about 30-50% cellulose, 15-32% hemicelluloses, and 15-25% lignin. Cellulose is a polymer of β-D-glucose, β-D-xylose is the main component of hemicelluloses, and lignin is composed of phenylpropane monomer units. In cellulose, the repeating unit is cellobiose which is formed by glucose units connected to each other by β-1,4'-glycosodic linkages. The biochemical conversion of cellulose to biofuels begins with hydrolysis to monomers (glucose) using chemical and/or biological catalysts in the liquid phase. This is

followed by conversion of the glucose units into ethanol or butanol through fermentation or directly to liquid alkanes using catalytic liquid processes.[1] Since glucose is the most abundant monosaccharide present in biomass, understanding processes with high-glucose yields is crucial to the cost-effective production of biofuels.

Cellobiose serves as a good model for cellulose as cellobiose hydrolysis allows one to understand the cleavage of the glycosidic bond. Experimentally, a maximum glucose yield of 36.8% at 320°C and 5100 psi, below the critical point of water, has been reported.[2] On the other hand, Pt/Al_2O_3 catalysts have been demonstrated to hydrolyze cellobiose completely with about 20% degradation of the glucose product.[3]

Hydrolysis of β-cellobiose is generally known to be an acid-catalyzed reaction involving the formation of an oxacarbenium ion. The commonly accepted reaction mechanism, known as the A-1 mechanism, involves multiple steps.[4] Cellobiose is protonated at the glycosidic oxygen with H_3O^+ to form the oxonium ion. The latter then breaks down into a glucose unit and an oxacarbenium ion. Hydration of the oxocarbenium ion produces another glucose unit and regenerates the hydronium ion. Kinetic studies on cellobiose hydrolysis in dilute sulfuric acid show that the reaction follows pseudo first-order kinetics at temperatures of about 100°C[5] with an activation energy of 32.3 kcal/mol.[6] At temperatures of 300 to 400°C and pressures of 25 to 40 MPa, the activation energy for this reaction decreases to 26.0 kcal/mol.[7] There is evidence, however, that the oxacarbenium ion is too short-lived to support a stepwise mechanism for the hydrolysis reaction. For example, lifetimes of about 10^{-15} s have been estimated for oxacarbenium ions derived from acetophenone acetals in water.[8] Large secondary deuterium isotope effects in the hydrolysis of methoxymethyl derivatives also

indicate differences in the nature of the nucleophilic interactions with the carbon atom bearing the positive charge in the incipient oxacarbenium ion. Such isotope effects are not consistent with unimolecular dissociation mechanisms like the dissociation of protonated glycosidic linkages to yield the oxacarbenium ion.[9] The transition state for acid-catalyzed hydrolysis of cellobiose has, therefore, been postulated to be similar in structure to those produced by S_N2 reactions where the positive charge on the carbon atom is stabilized by loose association with both the attacking and the leaving groups.[10, 11]

This paper deals with calculation of the energies and atomic charges associated with the acid-catalyzed hydrolysis of cellobiose in the gas phase using *ab initio* and density functional levels of theory. The effects of adding diffuse functions to the basis set on the energies and atomic charges were also examined and will be discussed in the succeeding sections.

Computational Methods

Geometry optimized structures for the different species involved in the acid-catalyzed hydrolysis of cellobiose were determined using RHF/6-31G(d,p) [12] and B3LYP [13] /6-311+G(d,p) using the GAMESS[14] electronic structure code, which is freely available from Iowa State University. In some instances, this was achieved by first running constrained optimizations of certain species similar to those thought to be involved in the reaction. The constrained optimizations were accomplished by freezing selected internuclear distances near the reaction center at fixed values while allowing all other degrees of freedom to relax. Further details of these constrained optimizations are described for each step of the reaction in which they were actually performed in the succeeding sections. Full geometry optimizations were then performed at the energy critical point for each series of constrained

optimizations. Single-point energies for each species were also determined using RHF/6-311+G(d,p) and B3LYP/6-31G(d,p) to determine the effects of exclusively using either the extended basis set or the DFT functional. All structures were visualized using MacMolPlt,[15] a graphical interface to GAMESS. Coordinates, absolute energies, and charges associated with all calculations are included in the Appendix.

Results and Discussion

Cellobiose Equilibrium Geometry

The equilibrium geometry of cellobiose (Figure 1) was determined. The numbering of the atoms in the cellobiose structure is given in Figure 1 and remains consistent through all of the other figures. No significant differences of the gross structures were seen between the two levels of theory. Both calculations reveal preference for the *syn* conformation in the gas phase, indicated by the hydrogen atoms (H25 and H30) attached to carbons (C2 and C7) forming the glycosidic linkage being found on the same side of the average plane determined by the cellobiose molecule at equilibrium.
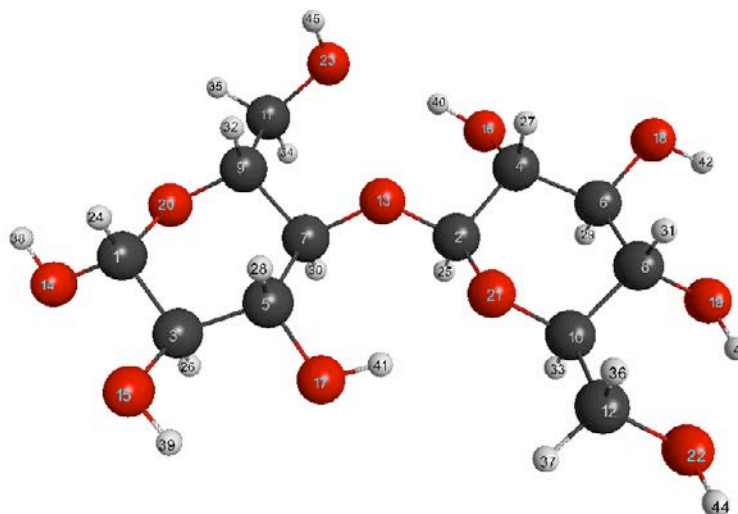
Figure 1. Equilibrium geometry of cellobiose. In all figures, the carbon atoms are indicated by black spheres, oxygen atoms are red, and hydrogen atoms are gray.

The Mulliken[16,17] and Löwdin[18] populations for two of the important oxygens in cellobiose are given in Table 1. Based on the RHF/6-31G(d,p) Mulliken charges, the glycosidic oxygen, O13, is the most negatively charged atom within the vicinity of the reaction center. Using only these charges as indicators, the site of protonation on cellobiose with hydronium ion is predicted to be on the glycosidic oxygen. However, all other levels of theory, basis set and population type assign O13 as the least negative among the oxygen atoms and O23 (on the hydroxymethyl group of the reducing unit of cellobiose) as being the most negative atom within the immediate vicinity of the glycosidic bond. Thus, protonation with hydronium ion is predicted to occur at O23 using this indicator.

Significant changes in the atomic charges are seen between the 6-31G(d,p) basis set and the nominally triple-ζ basis set with diffuse functions added to the heavy atoms, that is, 6-311+G(d,p). When used with atoms having high electron densities, such as oxygen atoms,

the use of diffuse functions on the heavy atoms is important because they possess greater flexibility in allowing the electron density to shift. The magnitude of the changes in atomic charges justifies the use of the expanded basis set for this system and is consistent with basis sets used in earlier studies.[19]

Use of the B3LYP functional tends to result in reduced negative charges on the oxygen atoms in cellobiose. B3LYP has been shown to have excellent agreement with high-level *ab initio* results when predicting partial charges.[20] The calculated values for the Löwdin atomic charges also illustrate a more modest basis set dependence when using B3LYP than with RHF. In addition, the positive charges on O13 in the larger basis set Mulliken charges, suggests that the Mulliken charges are less chemically intuitive than those of the Löwdin charges.

Table 2. Atomic charges on selected oxygen atoms in cellobiose determined using Mulliken (Mull.) and Löwdin (Löw.) formulas.

	RHF				B3LYP			
	6-31G(d,p)		6-311+G(d,p)		6-31G(d,p)		6-311+G(d,p)	
Atomic charges								
	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
O13	-0.743	-0.294	+0.011	-0.220	-0.514	-0.217	+0.060	-0.180
O23	-0.674	-0.354	-0.375	-0.257	-0.522	-0.301	-0.362	-0.246

Cellobiose Protonation

1. Cellobiose-H_3O^+ Complex

As suggested by the charges in the previous section, protonation of cellobiose in the gas phase is predicted to occur at the glycosidic oxygen, O13, using RHF/6-31G(d,p) and at O23 using B3LYP/6-311+G(d,p). As shown in Figures 2 and 3, optimized structures at these levels of theory for the cellobiose-H_3O^+ complex agree with this prediction. At the B3LYP/6-311+G(d,p) level, the O23 is protonated and a water molecule is available near the glycosidic oxygen. As the internuclear distance between O13 and H46 suggest (Table 2), there may very well be a hydrogen bond between the glycosidic oxygen and the water molecule. Another hydrogen bond is believed to be formed between H45 on the hydroxymethyl group on the reducing unit and O16 in the nonreducing residue. These atoms are calculated to be 1.48 Å from each other with B3LYP. The protonation of O23 instead of O13 may explain part of the recalcitrance of cellulose to acid-catalyzed hydrolysis since highly acidic solutions are required for the hydrolysis.

In an attempt to approximate the equilibrium geometry shown in Figure 2 using B3LYP/6-311+G(d,p), a series of constrained optimizations were run by gradually varying the distance between a hydrogen atom in the hydronium ion, H46, and the glycosidic oxygen, O13. This distance ranged from as short as 0.90 Å to as long as 3.00 Å. The O-H distances within the hydronium ion were held constant at 0.96 Å so that the resulting energy became a function solely of the distance between the hydronium and the glycosidic oxygen, as expressed by the O13-H46 distance. All other bonds were allowed to relax. Using this approach, a constrained, local energy minimum was found when the H46-O13 distance was

at 1.62 Å. The energy of this local minimum, however, was determined to be 17.4 kcal/mol higher than the one corresponding to the structure in Figure 3. Once the constraint on the O13-H46 distance was released, however, the structure of the cellobiose-hydronium complex reverted back to the one shown in Figure 3.
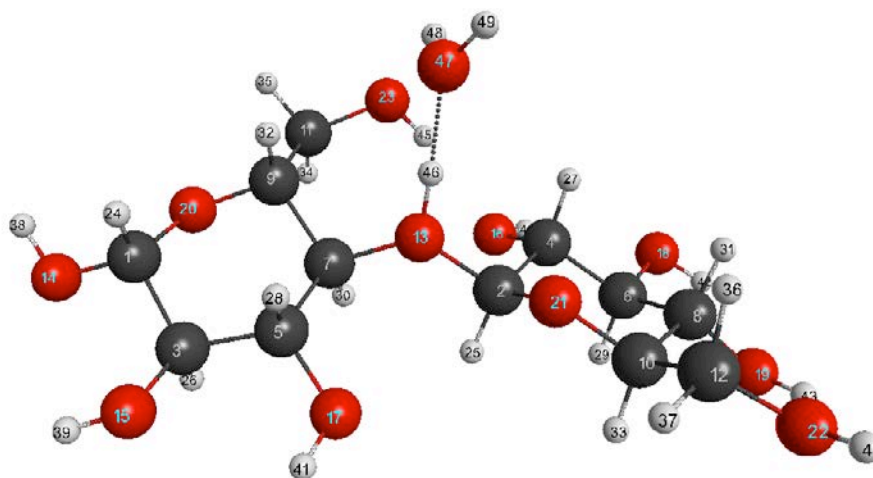
Figure 2. Equilibrium geometry for the cellobiose-hydronium complex predicted using RHF/6-31G(d,p) basis set.
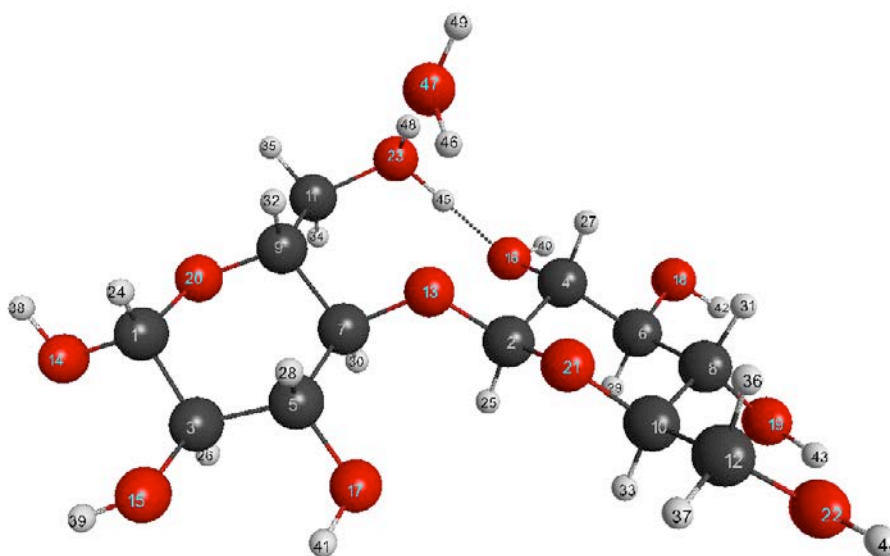
Figure 3. Equilibrium geometry after protonation of cellobiose with hydronium ion at B3LYP/6-311+G(d,p).

Mulliken and Löwdin charges for the cellobiose-H_3O^+ complex are shown in Table 2. At RHF/6-31G(d,p), because the site of protonation occurs almost exclusively at the glycosidic oxygen, O13 is expected to go through a significant increase in atomic charge while the atomic charge on O23 is expected to change minimally.

Table 3. Select internuclear distances (Å) and atomic charges in the cellobiose-H_3O^+ complex.

	RHF		B3LYP	
	6-31G(d,p)	6-311+G(d,p)	6-31G(d,p)	6-311+G(d,p)
Internuclear distances				
H45-O16	2.04			1.48
H46-O13	1.00			2.24
H48-O23	2.11			1.02
H46-O47	1.52			0.96
H48-O47	0.95			1.54
H49-O47	0.94			0.96

Atomic charges

	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
O13	-0.757	-0.106	-0.078	-0.042	-0.501	-0.042	+0.023	-0.068
O23	-0.727	-0.382	-0.376	-0.277	-0.571	-0.327	-0.216	-0.011
O47	-0.740	-0.390	-0.589	-0.228	-0.618	-0.352	-0.563	-0.239

Also, because both O13 and O23 are involved in the protonation process when B3LYP/6-311+G(d,p) is used, one expects both oxygen atoms to have increases in their respective

atomic charges, with O23 having a more significant increase than O13. Only the Löwdin values are consistent with this chemical intuition. Again, the Mulliken charges tend to be much more negative than the Löwdin charges.

Protonation of cellobiose with the hydronium ion is calculated to be exothermic by -58.8 kcal/mol at the RHF/6-31G(d,p) level compared to -69.3 kcal/mol when calculated using B3LYP/6-311+G(d,p).

2. Formation of Cellobiose-H$^+$ Ion

The structure of protonated cellobiose upon dehydration of the cellobiose-H$_3$O$^+$ complexes is shown in Figure 4. At the RHF level of theory, removal of the water molecule from the cellobiose-hydronium complex leaves H46 slightly closer to O13 while still forming a hydrogen bond with O23 on the nearby hydroxymethyl group. With the larger basis and the B3LYP functional, however, electron densities on these two oxygen atoms become more dispersed around the reaction center. The hydrogen atom, H46, is then shared more equally by O13 and O23, as reflected by the smaller difference between the H46-O13 and H46-O23 distances (Table 3).

The charges of oxygen atoms near the hydrolysis reaction center in protonated cellobiose are listed in Table 3. In removing the water molecule from the RHF cellobiose-H$_3$O$^+$ complex, the positive charge from H46 in the cellobiose-H$^+$ ion becomes dispersed over one less oxygen atom than was present in the cellobiose-H$_3$O$^+$ complex.
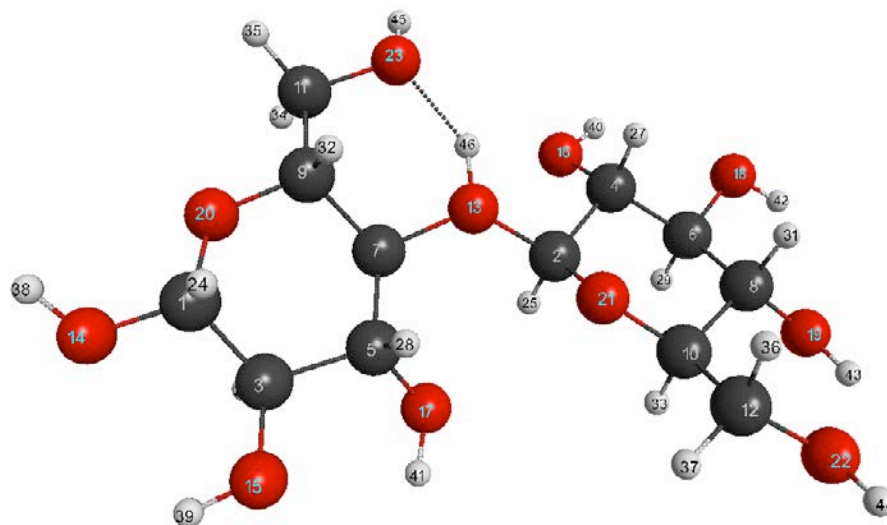
Figure 4. Equilibrium geometry of protonated cellobiose obtained with the use of RHF/6-31G(d,p) and B3LYP/6-311+G(d,p) after removing water from the cellobiose-H_3O^+ complex.

However, changes in the atomic charges on O13 and O23 are minimal because the H46 distance remains nearly the same distances for both the hydrated and dehydrated complexes. With B3LYP/6-311+G(d,p), the charge on O13, surprisingly, does not change much even though the proton has now shifted from O23 to O13. However, the charge on O23 is now more negative, which makes sense given the proton change and the decrease in hydrogen bonding.

Removing the water molecule from the protonated cellobiose involves an endothermicity of +12.0 kcal/mol when calculated using RHF/6-31G(d,p), while an enthalpy change of +18.9 is obtained when B3LYP/6-311+G(d,p) values are used.

Table 4. Select internuclear distances (Å) and atomic charges in the cellobiose-H^+ ion.

	RHF		B3LYP	
	6-31G(d,p)	6-311+G(d,p)	6-31G(d,p)	6-311+G(d,p)
Internuclear distances				
H46-O13	0.98			1.13
H46-O23	1.64			1.30
Atomic charges				

	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
O13	-0.737	-0.099	-0.069	-0.018	-0.498	-0.038	-0.067	-0.054
O23	-0.683	-0.363	-0.466	-0.247	-0.564	-0.299	-0.346	-0.118

This suggests that the water may not truly leave the vicinity of the protonated cellobiose. This is reasonable when one considers the fact that removing this water molecule results in intensifying the negative charge on O23 as electrons near the reaction center become more concentrated over fewer atoms.

Dissociation of Cellobiose-H^+ Ion

Dissociation of the cellobiose-H^+ ion at the glycosidic bond yields a molecule of β-D-glucose and the glucosyl cation (Figure 5). In determining an initial guess for the geometry for the transition state, constrained optimizations were run at various internuclear distances between the O13 and the C2 atoms in the glycosidic bond. The geometry corresponding to the maximum energy among these points was then fully optimized to determine the transition state geometry with RHF/6-31G(d,p). Points on the intrinsic reaction coordinate (IRC) were

determined at the same level of theory using the Gonzalez-Schlegel second-order method[21] with a step size of 0.05 Bohr-sqrt(amu) for each of the first five steps on each side of the transition state and a step size 0.20 Bohr-sqrt(amu) each step thereafter to verify reactants and products.
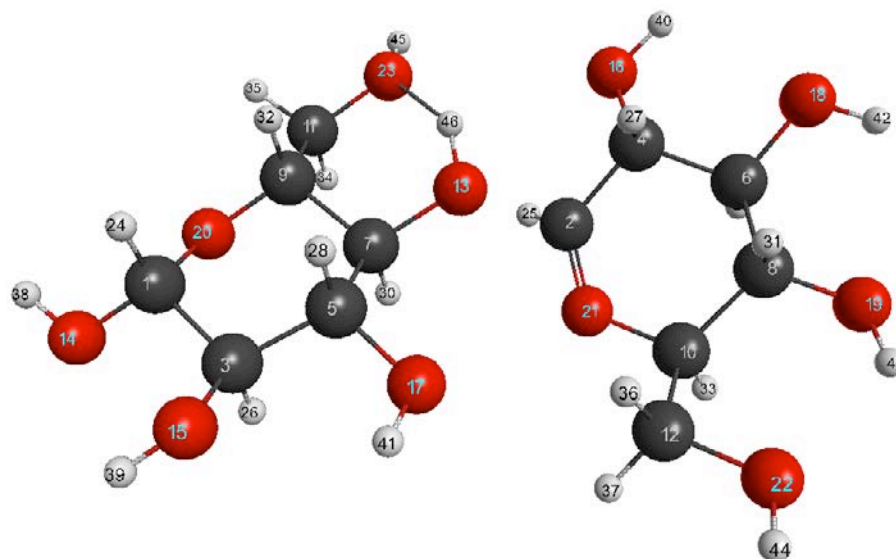
Figure 5. Transition-state geometry arising from the dissociation of the glycosidic bond in cellobiose determined using RHF/6-31G(d,p).

The distance between O13 and C2 changes from 1.46 Å in the cellobiose-H^+ ion to 1.84 Å in the transition state. The latter distance compares well with that reported by Liang and Montoya[22] for the transition state corresponding to the stepwise mechanism for this reaction using Car-Parrinello metadynamics simulation with B3LYP/6-31G(d,p) for the reaction in aqueous solution. Other relevant internuclear distances are listed in Table 4.

The energy barrier for this step was calculated to be 12.1 kcal/mol using RHF/6-31G(d,p) and 14.7 kcal/mol using B3LYP/6-311+G(d,p) single point energy at the RHF transition state.

Table 5. Internuclear distances (Å) between atoms near the reaction center in the transition state for the dissociation of protonated cellobiose.

	C2-O13	C7-O13	C2-O21	O13-H46
Gas-phase calculation using RHF/6-31G(d,p)	1.84	1.45	1.28	0.96
CPMD calculation using B3LYP/6-31G(d,p)[17]	1.85 ± 0.06	1.53 ± 0.06	1.34 ± 0.03	1.00 ± 0.03

Both of these values are slightly less than those reported for the aqueous reaction of Liang and Montoya, which is 15.8 kcal/mol. Starting from the optimized geometry of the cellobiose-H_3O^+ complex, the energy change between these two species is calculated to be +24.1 kcal/mol using RHF/6-31G(d,p).

Glucosyl Cation Equilibrium Geometry

Figure 6 shows the equilibrium geometry of the glucosyl cation, also referred to as the oxacarbenium ion, determined at both RHF and B3LYP levels of theory. Note that in this and the succeeding figures, the atoms involved have been renumbered starting with the carbon atom in the C=O bond as C1. The ion has the usual half-chair geometry[23] typically found in cyclohexenyl cations. This half-chair conformation is also consistent No significant differences in gross structure are seen to arise from adding the use of the B3LYP function and expanding the basis set to include diffuse functions.
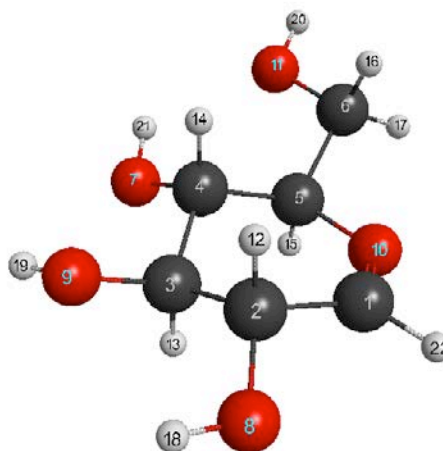
Figure 6. Equilibrium geometry of the glucosyl ion obtained from the dissociation of the glycosidic linkage in cellobiose at both RHF/6-31G(d,p) and B3LYP/6-311+G(d,p).

Examination of the atomic charges reveals that addition of a water molecule will most likely occur on the C1 atom of the glucosyl ion (Table 5). This atom is the most positive in the oxacarbenium ion and should, therefore, attract the oxygen atom from an incoming water molecule the most. This is predicted regardless of the set of atomic charges, level of theory, and basis set used. Also, O10 involved in the C=O bond is predicted to have essentially a neutral charge except for the Mulliken charges with the smallest basis set.

Table 6. Selected internuclear distances (Å) and atomic charges in glucosyl cation.

	RHF				B3LYP			
	6-31G(d,p)		6-311+G(d,p)		6-31G(d,p)		6-311+G(d,p)	
Internuclear distances								
C1-O10	1.23						1.25	
Atomic charges								
	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
C1	+0.470	+0.315	+0.386	+0.282	+0.311	+0.217	+0.250	+0.191
O8	-0.652	-0.322	-0.331	-0.205	-0.511	-0.272	-0.232	-0.176
O10	-0.521	-0.080	+0.006	-0.045	-0.286	+0.023	+0.084	+0.016

Generation of Second Glucose Molecule

1. Addition of Water to Glucosyl Cation

Optimizations show that water approaches the glucosyl cation at C1 from the side opposite to the hydroxymethyl group (Figure 7). The internuclear distance between C1 in the glucosyl cation and O22 in water is, however, longer than the typical 1.38 Å for C-O single bonds. In approaching the glucosyl cation, the water molecule does so in such a way as to not only optimize the attraction between C1 and O22 but to also allow attraction between O8 in the glucosyl cation and H23 in water. This is possible since O8 happens to be the most negatively charged atom within the vicinity of the water molecule as the latter gets close to the C1 atom (Table 5). It should also be noted that the C1-O10 bond becomes only slightly

longer (0.01 – 0.02 Å) as the C1 atom interacts with the oxygen atom from the approaching water molecule.   The B3LYP structure has significantly shorter C1-O22 and O8-H23 distances than the RHF structure as seen in Table 6.
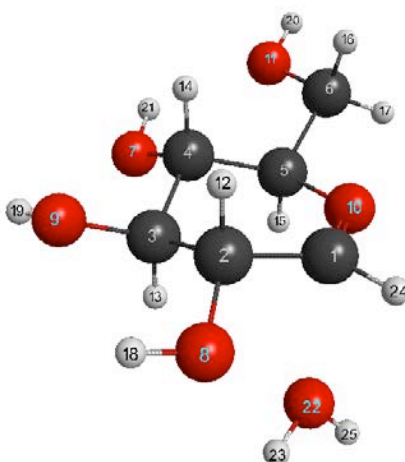
Figure 7. Equilibrium geometry of the assembly formed between the glucosyl cation and water determined using both RHF/6-31G(d,p) and B3LYP/6-311+G(d,p).

Charges calculated using B3LYP/6-311+G(d,p) (Table 6) are seen to be consistent with changes predicted by chemical intuition for each of the atoms directly affected by the approaching water molecule.   The C1 atom, for instance, is expected to have a reduced atomic charge going from the free glucosyl cation earlier described to one attracting the oxygen atom of a nearby water molecule.   As the character of the bond between C1 and O10 starts to change from double to single, as indicated by the slightly longer distance between these two atoms, the O10 atom is also expected to gain some electron density back from the C1 atom.   This results in a corresponding decrease in atomic charge.

Hydration of the glucosyl cation with one molecule of water was determined to be exothermic by -14.6 kcal/mol using RHF/6-31G(d,p) compared to -13.6 kcal/mol with B3LYP/6-311+G(d,p). Weak hydrogen-bonding interaction[24,25] between O8 in the cation and H23 in water as well as between C1 and O22 are believed to contribute to the negative energy values for this step of the reaction.

Table 7. Selected internuclear distances (Å) and atomic charges on atoms near the reaction center as a water molecule approaches the glucosyl cation.

	RHF		B3LYP	
	6-31G(d,p)	6-311+G(d,p)	6-31G(d,p)	6-311+G(d,p)
Internuclear distances				
C1-O10	1.24			1.27
C1-O22	2.32			2.10
O8-H23	2.41			2.09
Atomic charges				

	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
C1	0.519	0.316	0.131	0.269	0.337	0.202	-0.088	0.146
O8	-0.666	-0.328	-0.309	-0.204	-0.525	-0.278	-0.293	-0.207
O10	-0.554	-0.108	0.032	-0.073	-0.319	-0.009	0.084	-0.039
O22	-0.715	-0.405	-0.548	-0.265	-0.598	-0.353	-0.410	-0.193

2. Addition of Second Water Molecule to Glucosyl-H_2O Assembly

The addition of a second water molecule to the assembly in the previous section results in significant molecular changes: decreased internuclear distance between the C1 atom in the

glucosyl cation and the O22 atom from the first water molecule, shorter distance between O8 and H23 also from the first water molecule, and longer distance between C1 and O10 (Table 7). Introduction of the second water molecule, thus, stabilizes the glucose molecule formed by the glucosyl cation and the first water molecule in the assembly described in the previous step. This results in elongation of the bond between C1 and O10 in the glucose moiety to that typical of a C-O single bond, formation of a bond between the same carbon atom and O22, and hydrogen bonding among the atoms coming from the two water molecules (Figure 8).
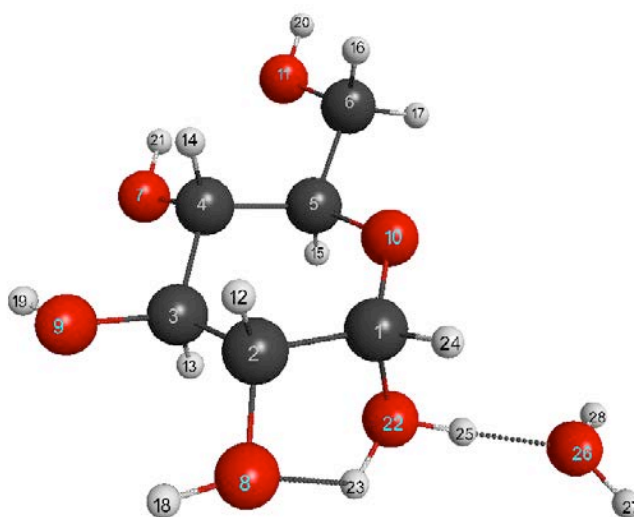
Figure 8. Equilibrium geometry resulting from the addition of a second water molecule to the glucosyl-H₂O assembly as determined using either RHF/6-31G(d,p) or B3LYP/6-311+G(d,p).

As the distance between C1 and O22 becomes shorter, the charge on the carbon atom is expected to decrease, while that of the oxygen is predicted to increase. These are both seen when B3LYP/6-311+G(d,p) is used (Table 7). These changes in atomic charges are also seen in the larger basis set even without switching from RHF to B3LYP. Other atoms expected to

go through changes in atomic charge as a result of the addition of the second water molecule are O10 and O8, both of which becoming more negative.

The addition of the second water molecule to the glucosyl-H_2O complex also results in an enthalpy change, -16.2 kcal/mol, that is slightly greater in magnitude than that associated with the addition of the first water molecule, -17.2 kcal/mol, at B3LYP. Adding the second water molecule introduces new attractive interactions among the species involved and their combined stabilizing effects lead to a further lowering of the overall energy of the system.

Table 8. Selected internuclear distances (Å) and atomic charges resulting from the addition of a second water molecule to the glucosyl-H_2O assembly. Changes with respect to values given in Table 6 for some internuclear distances appear in parentheses.

	RHF		B3LYP	
	6-31G(d,p)	6-311+G(d,p)	6-31G(d,p)	6-311+G(d,p)
Internuclear distances				
C1-O10	1.33 (+0.09)			1.33 (+0.06)
C1-O22	1.52 (-0.80)			1.59 (-0.51)
O8-H23	1.94 (-0.47)			1.83 (-0.26)
O22-H25	0.99			1.02
O26-H25	1.57			1.53
Atomic charges				

	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
C1	0.544	0.230	-0.137	0.097	0.359	0.153	-0.213	0.050
O8	-0.710	-0.358	-0.355	-0.238	-0.558	-0.303	-0.296	-0.223

O10	-0.685	-0.260	-0.053	-0.200	-0.456	-0.177	-0.004	-0.139
O22	-0.646	-0.138	-0.406	0.018	-0.476	-0.098	-0.345	-0.031
O26	-0.725	-0.390	-0.665	-0.238	-0.613	-0.353	-0.625	-0.231

3. Dissociation of H_3O^+ from Glucose

Removing the hydronium ion from the two-water system in the previous section to form the free glucose was done using constrained optimization. In this step, the O-H bonds in the departing hydronium ion were all assumed to have a length of 0.96 Å (to ensure that the hydronium departed intact) and the internuclear distance between the O22 and O26 atoms was fixed at about 4.00 Å. Without the constraints, full optimizations at either level of theory resulted in equilibrium geometries similar to the one depicted in Figure 8 (i.e. the hydronium reverted to the glucosyl-two water complex).
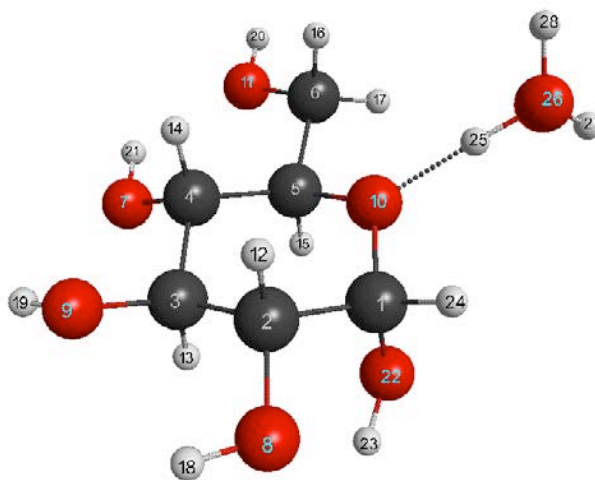
Figure 9. Equilibrium geometry resulting from constrained optimization of a system consisting of a free glucose molecule and a departing hydronium ion using either RHF/6-31G(d,p) or B3LYP/6-311+G(d,p).

On the other hand, a constrained optimization fixing the O22-O26 distance gave a local energy minimum with geometry similar to the one depicted in Figure 9. Here the hydronium ion is seen to move from O22, attached to C1, to O10 in glucose.

This is simply a result of O8 being the closest oxygen atom that also allows the hydronium to satisfy the O22-O26 distance constraint imposed on the system. This shows that the hydronium ion seeks other electron-rich atoms nearby as it is forced to depart from the previous reaction center near C1 in the glucose molecule. This is not surprising in a gas phase environment.

As a result of moving the hydronium ion, O26 in the hydronium ion is expected to have an increase in its atomic charge while the charge on O22 is expected to decrease. These are generally reflected by values calculated and listed in Table 8.

Calculated enthalpy changes for this step are +20.1 kcal/mol using RHF/6-31G(d,p) and +22.4 kcal/mol at the B3LYP level of theory. This suggests that removing the hydronium ion from the glucose-H_3O^+ complex does not proceed in the gas phase and that the presence of solvent molecules may be required to effect this change.

Table 9. Selected internuclear distances (Å) and atomic charges resulting from movement of the hydronium ion from O22 to O10.

	RHF		B3LYP	
	6-31G(d,p)	6-311+G(d,p)	6-31G(d,p)	6-311+G(d,p)
Internuclear distances				
C1-O10	1.42			1.45
C1-O22	1.37			1.39
O8-H23	2.24			2.18
O10-H25	1.50			1.47
O26-H25	0.96			0.96

Atomic charges

	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.	Mull.	Löw.
C1	0.561	0.196	0.039	0.042	0.381	0.130	0.017	-0.002
O10	-0.874	-0.342	-0.404	-0.293	-0.617	-0.254	-0.353	-0.254
O22	-0.667	-0.339	-0.332	-0.200	-0.516	-0.288	-0.264	-0.185
O26	-0.518	-0.033	-0.509	0.166	-0.412	-0.016	-0.404	0.164

Summary

Energies associated with the different steps of acid-catalyzed hydrolysis of cellobiose in the gas phase were determined using RHF/6-31G(d,p) and B3LYP/6-311+G(d,p). The mechanism studied for the reaction involves protonation of cellobiose with hydronium ion,

followed by dissociation of a water molecule from protonated cellobiose, dissociation of the cellobiose-H^+ ion to form a glucose molecule and an oxacarbenium ion, and successive hydration of the oxacarbenium ion to produce a second molecule of glucose. Mulliken and Löwdin atomic charges were compared at both levels of theory in the reaction mechanism. The Löwdin set of charges was found to be more intuitive when used at the higher level of theory.

Acid-catalyzed hydrolysis of cellobiose was found to be an endothermic process using both levels of theory. At RHF, the energy change for the dissociation of protonated cellobiose to form glucose and the glucosyl cation was found to be +24.1 kcal/mol. The oxacarbenium ion obtained from the dissociation of cellobiose-H^+ was observed to have a half-chair conformation. Hydration of the oxacarbenium ion resulted in the formation of α-glucose and regeneration of the hydronium ion in an exoergic process, also consistent with earlier studies.

It should be noted that the use of solvent models with this reaction must still be investigated. Also, concerted types of mechanisms, such as those involving nearly simultaneous nucleophilic attack of water on the C1 atom of the nonreducing unit of cellobiose with dissociation of the glycosidic bond should be examined.

[1] Huber, G. W.; Chheda, J. N.; Barrett, C. J.; Dumesic, J. A. *Science*, **2005**, *308*, 1446.

[2] Kruse, A.; Gawlik, A.; *Ind. Eng. Chem. Res.* **2003**, *42*, 267.

[3] Patrick, T. A.; Abraham, M. A. *Environ. Sci. Technol.* **2000**, *34*, 3480.

[4] Moiseiw, Y. V.; Khalturinskii, N. A.; Zaikov, G. E. *Carbohydr. Res.*, **1976**, *51*, 23.

[5] Dadach, Z. E.; Kaliaguine, S. *Can. J. Chem. Eng.* **1993**, *71*, 880.

[6] Pinto, J. H. Q.; Kaliaguine, S. *AIChE J.* **1991**, *37*, 905.

[7] Kabyemela, B. M.; Takigawa, M.; Adschiri, T.; Malaluan, R. M.; Arai, K. *Ind. Eng. Chem. Res.* **1998**, *37*, 357.

[8] Young, P. R; Jencks, W. P. *J. Am. Chem. Soc.* **1977**, *99*, 8238.

[9] Knier, B. L; Jencks, W. P. *J. Am. Chem. Soc.* **1980**, *102*, 6789.

[10] Sinnott, M. L.; Jencks, W. P. *J. Am. Chem. Soc.* **1980**, *102*, 2026.

[11] Cocker, D.; Sinnott, M. L. *J. Chem. Soc., Perkin Trans.* **1975**, *2*, 1391.

[12] (a) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J.Chem.Phys.*, **1971**, *54*, 724. (b) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J.Chem.Phys.*, **1972**, *56*, 2257.

[13] (a) Becke, A. D. *J.Chem.Phys.*, **1993**, *98*, 5648. (b) Stephens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *J.Phys.Chem.*, **1994**, *98*, 11623. (c) Hertwig, R. H.; Koch, W. *Chem.Phys.Lett.*, **1997**, *268*, 345.

[14] (a) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen,J. H.; Koseki, S.; Matsunaga,N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347. (b) Gordon, M. S.; Schmidt, M. W. *Theory and Applications of Computational Chemistry, the first forty years;* Elsevier: Amsterdam, 2005. The GAMESS computational chemistry package may be obtained at http://www.msg.chem.iastate.edu/.

[15] Bode, B. M.; Gordon, M. S. *J. Mol. Graphics Modell.* **1998**, *16*, 133.

[16] Mulliken, R. S. *J. Chem. Phys.* **1955**, *3*, 564.

[17] Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428.

[18] Lowdin, P.-O. *J. Chem. Phys.* **1950**, *18*, 365.

[19] Csonka, G. I. *J. Mol. Struct. (THEOCHEM)* **2002**, *584*, 1.

[20] Martin, J. M. L.; Proft, F.; Geerlings, P. *Chem. Phys. Lett.* **1996**, *250*, 393.

[21] Gonzalez, C.; Schlegel, H. B. *J. Chem. Phys.* **1989**, *90*, 2154.

[22] Liang, X.; Montoya, A.; Haynes, B. S. *J. Phys. Chem. B* **2011**, *115*, 10682.

[23] Sinnott, M. E. *Biochem. J.* **1984**, *224*, 817.

[24] Jeffrey, G. A. *An Intro. To Hydrogen Bonding.* **1997**, Oxford University Press, Oxford, England.

[25] Desiraju, G. R.; Steiner, T. *The Weak Hydrogen Bond.* **1999**, Oxford University Press, Oxford, England.

CHAPTER 4

Conclusion: The Adventure Continues…

Rotations of free -OH groups attached to C2 and C3 atoms in the glucose residues of cellulose Iα were shown to be freer compared to those bonded to the C6 atom. Hydroxyls found in hydroxymethyl groups extend out of the glucose residues and are, thus, more prone to interactions with atoms belonging to other residues in cellulose. These interactions generally resulted from maximizing hydrogen bonding interactions or minimizing repulsions among atoms of the various glucose residues. Calculations on the relative energy changes arising from these –OH group rotations demonstrated the greater roles of intrachain and interchain interactions over those interactions that occur between sheets of glucose residues. Relative energies calculated using B3LYP and MP2 levels of theory were found to be within 1.0 kcal/mol of each other as the various -OH dihedrals are rotated.

The use of FMO2/MP2 with PIEDA showed that interresidue distance is also an important factor both in determining the size of the interaction energy between each pair of residues in cellulose Iα as well as the nature of each of the interaction energies determined. Electrostatic and exchange energies account for large parts of interaction energies when the relative separation between residues is small. However, the charge transfer and dispersion contributions cannot be ignored. Exchange energies were seen to vanish as the relative separation between glucose units in cellulose approaches 1.50. Interaction energies become purely electrostatic in nature when the relative separation between glucose units in cellulose are greater than 2.00.

It would be interesting to see how these interactions between individual residues would combine to determine interactions among the chains themselves of glucose residues as the cellulose fragment studied is expanded. Interactions arising from small dihedral rotations of entire glucose residues about bonds in the glycosidic linkages may also be investigated, since these residues do not exist as static units in the cellulose crystal at all. The behavior of glucose residues found on the edges or corners of large cellulose fragments might also be differentiated from those found well within the interior of a crystal.

The A-1 mechanism for the acid catalyzed hydrolysis of cellobiose in the gas phase was revisited and energies related to each step in this mechanism using both RHF and B3LYP levels of theory. The reaction was found to be an endoergic process using both levels of theory. With RHF, the transition state for the dissociation of protonated cellobiose to form glucose and the glucosyl cation was found to be 24.1 kcal/mol higher in energy than the cellobiose-H_3O^+ complex. Bond lengths as well as the ring conformation of the oxacarbenium ion from the nonreducing end of cellobiose were also found to generally agree with those reported in previous studies. However, the use of solvent models with this reaction must still be investigated. Also, other concerted types of mechanisms, such as those involving nearly simultaneous nucleophilic attack of water on the C1 atom of the nonreducing unit of cellobiose with dissociation of the glycosidic bond should also be examined.

APPENDIX A

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose

With RHF/6-31G(d,p)

Table A.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.613623882	0.164458797	-3.757778029
C	0.322196612	-0.081647492	1.427354840
C	-1.205026937	-0.911728159	-2.868233984
C	1.011588603	0.826810318	2.437258920
C	-1.226251349	-0.440900278	-1.420893663
C	1.131426118	0.120251938	3.775294039
C	0.141494599	0.063035122	-0.975566256
C	-0.197948280	-0.481519404	4.215685747
C	0.633024362	1.119761908	-1.966243341
C	-0.813116477	-1.329854823	3.102210306
C	2.058960864	1.617689879	-1.739791785
C	-2.241555795	-1.770104747	3.384500127
O	0.065329655	0.658805446	0.296330539
O	-0.417987284	-0.346601181	-5.011621290

O	-2.497628053	-1.192729037	-3.303529132
O	2.286354555	1.133251190	1.982292220
O	-1.674448788	-1.523334737	-0.658400539
O	1.581363865	1.031870444	4.728554507
O	0.084979833	-1.230523556	5.361875160
O	0.634978378	0.560126717	-3.251635164
O	-0.903007311	-0.560110857	1.923988544
O	-2.263549427	-2.490650807	4.590138514
O	2.195477369	2.465412715	-0.631763504
H	-1.278064534	1.030316345	-3.779880652
H	0.969622253	-0.925803210	1.196274963
H	-0.562969256	-1.786674827	-2.947565274
H	0.397232516	1.716891063	2.560605110
H	-1.927800771	0.389763968	-1.340768900
H	1.845782165	-0.695136218	3.661958804
H	0.845478978	-0.765359538	-0.962836028
H	-0.883260087	0.335155569	4.444133484
H	-0.048571755	1.973170856	-1.924841456
H	-0.188998416	-2.207531318	2.938792708
H	2.723154194	0.779774477	-1.573424424
H	2.378914990	2.116682400	-2.649164686

H	-2.871175681	-0.885931383	3.444738292
H	-2.593287508	-2.383440090	2.559390200
H	-0.049851623	0.319180395	-5.569900253
H	-2.882720473	-1.808534159	-2.699797578
H	2.235412925	1.667333023	1.199968006
H	-1.744463456	-1.258953708	0.248311118
H	1.655154639	0.580263611	5.554709308
H	-0.687598446	-1.704173529	5.630119103
H	-3.138303923	-2.784920290	4.780406447
H	1.717771903	3.266909513	-0.768194601

TOTAL ENERGY   =   -1290.7096447592 Hartrees

Table A.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.386670	0.613330	5.810443	0.189557
C	5.385986	0.614014	5.820827	0.179173
C	5.787743	0.212257	5.950614	0.049386
C	5.787315	0.212685	5.949225	0.050775
C	5.801200	0.198800	5.961188	0.038812
C	5.773426	0.226574	5.942382	0.057618
C	5.783611	0.216389	5.951719	0.048281

C	5.775786	0.224214	5.958841	0.041159
C	5.809780	0.190220	5.950777	0.049223
C	5.823388	0.176612	5.949574	0.050426
C	5.872458	0.127542	6.007947	-0.007947
C	5.877661	0.122339	6.012019	-0.012019
O	8.743014	-0.743014	8.293835	-0.293835
O	8.659984	-0.659984	8.320255	-0.320255
O	8.681731	-0.681731	8.353472	-0.353472
O	8.704140	-0.704140	8.356863	-0.356863
O	8.725632	-0.725632	8.373865	-0.373865
O	8.678582	-0.678582	8.349178	-0.349178
O	8.710016	-0.710016	8.370697	-0.370697
O	8.731941	-0.731941	8.327743	-0.327743
O	8.763629	-0.763629	8.327608	-0.327608
O	8.686572	-0.686572	8.369824	-0.369824
O	8.674303	-0.674303	8.353662	-0.353662
H	0.902731	0.097269	0.927031	0.072969
H	0.888569	0.111431	0.918513	0.081487
H	0.863372	0.136628	0.901775	0.098225
H	0.871930	0.128070	0.904996	0.095004
H	0.880973	0.119027	0.911588	0.088412

H	0.877364	0.122636	0.908534	0.091466
H	0.864445	0.135555	0.903441	0.096559
H	0.883076	0.116924	0.912837	0.087163
H	0.893004	0.106996	0.918645	0.081355
H	0.861952	0.138048	0.905599	0.094401
H	0.859833	0.140167	0.890656	0.109344
H	0.871355	0.128645	0.901600	0.098400
H	0.879886	0.120114	0.908709	0.091291
H	0.879508	0.120492	0.904691	0.095309
H	0.655007	0.344993	0.772496	0.227504
H	0.643943	0.356057	0.772741	0.227259
H	0.609807	0.390193	0.781406	0.218594
H	0.618340	0.381660	0.785917	0.214083
H	0.646857	0.353143	0.775422	0.224578
H	0.625430	0.374570	0.779195	0.220805
H	0.644577	0.355423	0.773645	0.226355
H	0.653476	0.346524	0.778006	0.221994

APPENDIX B

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose

With B3LYP/6-311+G(d,p)

Table B.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.622192956	0.154883220	-3.794479421
C	0.328745722	-0.049483356	1.428993896
C	-1.232693516	-0.906031354	-2.883926835
C	0.991104025	0.871019194	2.461020779
C	-1.243224599	-0.412949651	-1.435932912
C	1.114393840	0.151776308	3.799431594
C	0.137839094	0.076832274	-0.991109406
C	-0.209619657	-0.491432474	4.226886697
C	0.664432475	1.117236470	-1.993922222
C	-0.806043995	-1.355335373	3.104507567
C	2.109018330	1.581078918	-1.752994152
C	-2.230245260	-1.829294845	3.391172171
O	0.061123913	0.706206061	0.292827180
O	-0.439965955	-0.385592581	-5.058031119
O	-2.553611884	-1.176451926	-3.321949475

O	2.283524150	1.219730168	2.013507216
O	-1.728758830	-1.492717667	-0.648752377
O	1.545604666	1.080827920	4.780889350
O	0.090689927	-1.251072971	5.392320105
O	0.667639566	0.541324723	-3.297468746
O	-0.922299631	-0.575209297	1.905769844
O	-2.223742158	-2.557303161	4.625312119
O	2.268359274	2.438325150	-0.622504373
H	-1.269374188	1.046969507	-3.821916543
H	1.000125927	-0.888969135	1.192609692
H	-0.605050303	-1.804641168	-2.956676613
H	0.352635513	1.756932202	2.584957261
H	-1.937328440	0.438671030	-1.368078560
H	1.860062880	-0.649566911	3.686488185
H	0.836628116	-0.768112063	-0.948997183
H	-0.922122331	0.313184299	4.463146673
H	-0.002448435	1.995924450	-1.967391509
H	-0.154799011	-2.224208950	2.931206755
H	2.755038201	0.720391491	-1.570969960
H	2.457066156	2.074505375	-2.666456135
H	-2.883614189	-0.951205119	3.457324540

H	-2.574652805	-2.464970655	2.567984102
H	-0.114623822	0.304940539	-5.646285471
H	-2.954192837	-1.752533635	-2.659426910
H	2.211029646	1.714263263	1.179236551
H	-1.716414057	-1.220883632	0.284659323
H	1.595331076	0.602931534	5.617542718
H	-0.693570489	-1.770381943	5.619271336
H	-3.118908809	-2.839482565	4.835481644
H	1.833595509	3.280628746	-0.792215725

TOTAL ENERGY   =   -1297.6626061150 Hartrees

Table B.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.802902	0.197098	6.013424	-0.013424
C	5.892262	0.107738	6.032631	-0.032631
C	6.313506	-0.313506	6.077638	-0.077638
C	6.303485	-0.303485	6.076647	-0.076647
C	6.244536	-0.244536	6.097818	-0.097818
C	6.225924	-0.225924	6.081510	-0.081510
C	5.911679	0.088321	6.093232	-0.093232
C	6.145851	-0.145851	6.095091	-0.095091

C	6.643940	-0.643940	6.090242	-0.090242
C	6.377675	-0.377675	6.085218	-0.085218
C	6.115774	-0.115774	6.074261	-0.074261
C	6.075958	-0.075958	6.076465	-0.076465
O	7.940500	0.059500	8.179707	-0.179707
O	8.232781	-0.232781	8.170619	-0.170619
O	8.233740	-0.233740	8.214654	-0.214654
O	8.328681	-0.328681	8.213412	-0.213412
O	8.325052	-0.325052	8.239954	-0.239954
O	8.219366	-0.219366	8.210022	-0.210022
O	8.280016	-0.280016	8.240625	-0.240625
O	8.133005	-0.133005	8.227382	-0.227382
O	8.172345	-0.172345	8.240231	-0.240231
O	8.370885	-0.370885	8.264762	-0.264762
O	8.361682	-0.361682	8.246446	-0.246446
H	0.850506	0.149494	0.867374	0.132626
H	0.848635	0.151365	0.851003	0.148997
H	0.811996	0.188004	0.853089	0.146911
H	0.823483	0.176517	0.853059	0.146941
H	0.817681	0.182319	0.856890	0.143110
H	0.816597	0.183403	0.855859	0.144141

H	0.844950	0.155050	0.848385	0.151615
H	0.825242	0.174758	0.859520	0.140480
H	0.841808	0.158192	0.863670	0.136330
H	0.818305	0.181695	0.855275	0.144725
H	0.825357	0.174643	0.864095	0.135905
H	0.834672	0.165328	0.875872	0.124128
H	0.831954	0.168046	0.877160	0.122840
H	0.836870	0.163130	0.882226	0.117774
H	0.755494	0.244506	0.824007	0.175993
H	0.744045	0.255955	0.820576	0.179424
H	0.596608	0.403392	0.822463	0.177537
H	0.682608	0.317392	0.825523	0.174477
H	0.755377	0.244623	0.822538	0.177462
H	0.709837	0.290163	0.821705	0.178295
H	0.723611	0.276389	0.829064	0.170936
H	0.752820	0.247180	0.828653	0.171347

APPENDIX C

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose-H₃O⁺ Complex

With RHF/6-31G(d,p)

Table C.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.214074002	0.846237194	-3.910005296
C	0.067602701	0.664342486	1.444109527
C	-0.300080864	-0.431935027	-3.097924730
C	0.996273061	1.520881337	2.278005146
C	-0.685041560	-0.107029968	-1.663052891
C	1.254687563	0.766320306	3.570081043
C	0.248970584	0.946930032	-1.099291835
C	-0.053930757	0.449021290	4.281237094
C	0.359499713	2.172356676	-2.020720275
C	-1.009917311	-0.266278186	3.315781991
C	1.456039616	3.165440264	-1.640997581
C	-2.436098131	-0.372698583	3.830302085
O	-0.280178487	1.347910923	0.218191386
O	0.246423101	0.515799012	-5.150711623
O	-1.276851334	-1.302250278	-3.568980428

O	2.173981020	1.753456916	1.561456322
O	-0.603722684	-1.227083004	-0.841099915
O	2.100381629	1.567509248	4.329996075
O	0.297074290	-0.347043876	5.364154995
O	0.711006143	1.702102371	-3.286564947
O	-1.114895840	0.473795733	2.099548087
O	-2.383387951	-0.993517622	5.083641258
O	1.244903951	3.730410711	-0.375865853
H	-1.184417082	1.341592117	-3.957219985
H	0.527655243	-0.258532676	1.125582402
H	0.686034437	-0.890884828	-3.113927921
H	0.495838816	2.458378887	2.507402872
H	-1.697837856	0.293663084	-1.664827007
H	1.742357921	-0.179575039	3.332749371
H	1.225683739	0.533619117	-0.898357005
H	-0.511040428	1.383752339	4.605947048
H	-0.603005957	2.681806251	-2.049544463
H	-0.620496578	-1.258733929	3.104425357
H	2.416992213	2.666177299	-1.695826666
H	1.451683883	3.956462173	-2.378339981
H	-2.859506903	0.626231685	3.900505508

H	-3.027248274	-0.945619811	3.123247434
H	0.253087403	1.265348869	-5.725693783
H	-1.048735137	-1.608966637	-4.433553328
H	2.870894899	1.945783812	2.175423857
H	-1.116587102	-1.926750129	-1.219442275
H	2.285981177	1.141303312	5.154149291
H	-0.480525728	-0.636957578	5.819448352
H	-3.250046128	-1.151496153	5.420554603
H	1.776106464	3.284755650	0.269727915
H	-0.774982585	2.212321520	0.315817326
O	-1.387268876	3.605969106	0.393707633
H	-0.694489118	4.243444067	0.266031406
H	-2.103946225	3.980539206	0.880543384

Total Energy: -1367.1151225564 Hartrees

Table C.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
H	0.883714	0.116286	0.920226	0.079774
H	0.803474	0.196526	0.875774	0.124226
H	0.842389	0.157611	0.895687	0.104313

H	0.852714	0.147286	0.898739	0.101261
H	0.853793	0.146207	0.900529	0.099471
H	0.856577	0.143423	0.902923	0.097077
H	0.792817	0.207183	0.868876	0.131124
H	0.867613	0.132387	0.907137	0.092863
H	0.853456	0.146544	0.902844	0.097156
H	0.835476	0.164524	0.894159	0.105841
H	0.855008	0.144992	0.893334	0.106666
H	0.835269	0.164731	0.880040	0.119960
H	0.878586	0.121414	0.909052	0.090948
H	0.859626	0.140374	0.894587	0.105413
H	0.635081	0.364919	0.760406	0.239594
H	0.627711	0.372289	0.761622	0.238378
H	0.610324	0.389676	0.753933	0.246067
H	0.622844	0.377156	0.758041	0.241959
H	0.622716	0.377284	0.758264	0.241736
H	0.613020	0.386980	0.768630	0.231370
H	0.631624	0.368376	0.764803	0.235197
H	0.608241	0.391759	0.777043	0.222957
H	0.465985	0.534015	0.734021	0.265979
O	8.739674	-0.739674	8.390466	-0.390466

| H | 0.594776 | 0.405224 | 0.752816 | 0.247184 |
| H | 0.615888 | 0.384112 | 0.749030 | 0.250970 |

APPENDIX D

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose-H_3O^+ Complex

With B3LYP/6-311+G(d,p)

Table D.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.151966314	0.172763240	-3.884527575
C	0.116966096	0.104366689	1.410630295
C	-0.299676201	-1.096152852	-3.047175475
C	0.918596180	1.021586397	2.331287953
C	-0.684656603	-0.751057795	-1.608772759
C	1.295002562	0.262382111	3.586506836
C	0.280118825	0.271937434	-1.014875014
C	0.026598171	-0.240288984	4.279743592
C	0.467552761	1.469999084	-1.963283041
C	-0.864179318	-1.003223408	3.267973112
C	1.625701776	2.405298608	-1.594341013
C	-2.263539219	-1.291376247	3.806273177
O	-0.213913374	0.805188862	0.231385882
O	0.320274614	-0.198514031	-5.129038270
O	-1.327996644	-1.941869056	-3.532197282

O	2.103349176	1.490599268	1.636427772
O	-0.650610418	-1.899330565	-0.778267333
O	2.063060730	1.155613171	4.378992833
O	0.477457626	-1.059046553	5.340435592
O	0.826343820	1.031928936	-3.263227322
O	-1.077389934	-0.217744736	2.078063356
O	-2.109025296	-1.960047952	5.060020232
O	1.495746243	3.028827903	-0.264549405
H	-1.106509769	0.717405113	-3.949550510
H	0.680241150	-0.802494578	1.161589417
H	0.670761541	-1.609807584	-3.052924658
H	0.303158775	1.889284479	2.592027948
H	-1.696064094	-0.318391101	-1.613939879
H	1.901966797	-0.613942585	3.316383093
H	1.247928877	-0.213698856	-0.846123460
H	-0.531301130	0.627406224	4.663467270
H	-0.470974385	2.043690571	-1.994465124
H	-0.368099577	-1.946914058	3.004656879
H	2.574600910	1.871299090	-1.581732333
H	1.682908817	3.226019603	-2.305961109
H	-2.799474151	-0.342389406	3.929172538

H	-2.806451958	-1.912996052	3.087429710
H	0.266354784	0.540545711	-5.746220163
H	-1.072387220	-2.287593704	-4.395725468
H	2.747728812	1.765975052	2.309600094
H	-1.209158686	-2.575136604	-1.183414514
H	2.309091181	0.710336309	5.200208503
H	-0.295499331	-1.481138877	5.744867372
H	-2.970595504	-2.198910000	5.415993049
H	1.789074175	2.404159066	0.518629336
H	-1.370323098	2.717842318	0.436884088
O	-0.958266259	3.557367913	0.183895767
H	0.547569139	3.353004463	-0.078122523
H	-1.467373754	4.290356848	0.548317488

Total Energy: -1374.4685612169 Hartrees

Table D.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.870943	0.129057	6.015385	-0.015385
C	6.163104	-0.163104	6.027534	-0.027534
C	6.430787	-0.430787	6.083264	-0.083264

C	6.063469	-0.063469	6.070807	-0.070807
C	6.012141	-0.012141	6.077189	-0.077189
C	6.659783	-0.659783	6.084239	-0.084239
C	5.871944	0.128056	6.077019	-0.077019
C	6.129773	-0.129773	6.090353	-0.090353
C	6.635525	-0.635525	6.087234	-0.087234
C	6.335422	-0.335422	6.078738	-0.078738
C	6.202204	-0.202204	6.029273	-0.029273
C	5.949809	0.050191	6.067533	-0.067533
O	7.977313	0.022687	8.226623	-0.226623
O	8.250280	-0.250280	8.169179	-0.169179
O	8.244476	-0.244476	8.216269	-0.216269
O	8.313660	-0.313660	8.174310	-0.174310
O	8.240088	-0.240088	8.214027	-0.214027
O	8.249627	-0.249627	8.225433	-0.225433
O	8.254756	-0.254756	8.223500	-0.223500
O	8.129882	-0.129882	8.211665	-0.211665
O	8.072856	-0.072856	8.187449	-0.187449
O	8.367408	-0.367408	8.261867	-0.261867
O	8.216108	-0.216108	8.010992	-0.010992
H	0.834666	0.165334	0.863206	0.136794

H	0.808733	0.191267	0.839262	0.160738
H	0.788991	0.211009	0.850182	0.149818
H	0.783876	0.216124	0.847296	0.152704
H	0.805083	0.194917	0.852982	0.147018
H	0.783029	0.216971	0.851036	0.148964
H	0.819587	0.180413	0.842264	0.157736
H	0.816326	0.183674	0.857218	0.142782
H	0.804998	0.195002	0.853911	0.146089
H	0.801026	0.198974	0.850565	0.149435
H	0.780948	0.219052	0.847363	0.152637
H	0.780635	0.219365	0.850220	0.149780
H	0.833150	0.166850	0.878060	0.121940
H	0.826107	0.173893	0.875813	0.124187
H	0.734540	0.265460	0.813715	0.186285
H	0.734611	0.265389	0.812429	0.187571
H	0.705077	0.294923	0.789470	0.210530
H	0.731867	0.268133	0.809564	0.190436
H	0.726729	0.273271	0.804567	0.195433
H	0.699394	0.300606	0.811798	0.188202
H	0.707123	0.292877	0.821331	0.178669
H	0.527786	0.472214	0.818467	0.181533

H	0.681192	0.318808	0.794525	0.205475
O	8.563036	-0.563036	8.239121	-0.239121
H	0.631317	0.368683	0.810539	0.189461
H	0.648814	0.351186	0.805213	0.194787

APPENDIX E

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose-H[+] Ion

With RHF/6-31G(d,p)

Table E.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.788401809	0.620363378	-3.760016949
C	0.553947898	0.704174884	1.341438926
C	-0.700221356	-0.595564236	-2.857217219
C	1.309878837	1.535280352	2.350457090
C	-0.567383268	-0.187099416	-1.393544310
C	1.257868406	0.778783613	3.666160326
C	0.577242401	0.798399551	-1.251953391
C	-0.185265663	0.491479590	4.065229909
C	0.463590888	1.948686752	-2.267958388
C	-0.935253207	-0.184003188	2.906467549
C	1.689628812	2.849741331	-2.277609670
C	-2.441681472	-0.233144910	3.101911124
O	0.539129622	1.445693185	0.079063491
O	-0.787630158	0.174504796	-5.047548032
O	-1.839805073	-1.389397619	-2.920616959

O	2.602143692	1.717455559	1.868852273
O	-0.291387593	-1.297485444	-0.601049591
O	1.921590234	1.568473213	4.598369385
O	-0.103686263	-0.326317562	5.185752014
O	0.357540243	1.411533651	-3.543288432
O	-0.749295698	0.568996535	1.707071396
O	-2.685555676	-0.869832445	4.324298740
O	1.870209045	3.303236373	-0.949063845
H	-1.675868466	1.211734717	-3.534639087
H	1.035668941	-0.238817081	1.131074841
H	0.188111698	-1.148579635	-3.154096990
H	0.796288526	2.487802670	2.470426614
H	-1.489459555	0.295758136	-1.078956252
H	1.769297455	-0.175692435	3.542444132
H	1.529590923	0.290138124	-1.320711395
H	-0.674711322	1.437057408	4.297421118
H	-0.415776396	2.538526331	-2.010859837
H	-0.546901017	-1.189621345	2.769071280
H	2.550542697	2.285278028	-2.617237899
H	1.522386953	3.680758201	-2.950040943
H	-2.828580974	0.782557566	3.098531203

H	-2.890503086	-0.772629682	2.273610558
H	-0.937730339	0.879429186	-5.658401456
H	-1.926115607	-1.766047741	-3.783494325
H	3.153020598	2.037009470	2.569839490
H	-0.940081507	-1.965137563	-0.772651543
H	1.912651654	1.139304922	5.441398662
H	-0.971282648	-0.584695134	5.461903440
H	-3.611536840	-0.955268693	4.481943657
H	2.695041354	3.746937105	-0.827656645
H	1.158525041	2.209120840	0.051628431

Total Energy: -1291.0703334470 Hartrees

Table E.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.408606	0.591394	5.821658	0.178342
C	5.354502	0.645498	5.778100	0.221900
C	5.820477	0.179523	5.960845	0.039155
C	5.838487	0.161513	5.962517	0.037483
C	5.785155	0.214845	5.942486	0.057514
C	5.799902	0.200098	5.948341	0.051659

C	5.835080	0.164920	5.914138	0.085862
C	5.770792	0.229208	5.955115	0.044885
C	5.822456	0.177544	5.944305	0.055695
C	5.849879	0.150121	5.939925	0.060075
C	5.880176	0.119824	5.999402	0.000598
C	5.870876	0.129124	6.003658	-0.003658
O	8.737256	-0.737256	8.099472	-0.099472
O	8.667169	-0.667169	8.315570	-0.315570
O	8.686588	-0.686588	8.347935	-0.347935
O	8.702355	-0.702355	8.363099	-0.363099
O	8.688800	-0.688800	8.351481	-0.351481
O	8.690812	-0.690812	8.350642	-0.350642
O	8.694551	-0.694551	8.353926	-0.353926
O	8.708771	-0.708771	8.305493	-0.305493
O	8.708631	-0.708631	8.291716	-0.291716
O	8.683124	-0.683124	8.362695	-0.362695
O	8.728312	-0.728312	8.358535	-0.358535
H	0.878342	0.121658	0.915789	0.084211
H	0.806110	0.193890	0.879534	0.120466
H	0.839399	0.160601	0.892588	0.107412
H	0.839720	0.160280	0.894970	0.105030

H	0.834145	0.165855	0.890853	0.109147
H	0.856719	0.143281	0.902487	0.097513
H	0.792068	0.207932	0.870903	0.129097
H	0.863713	0.136287	0.905589	0.094411
H	0.837334	0.162666	0.895026	0.104974
H	0.835750	0.164250	0.894687	0.105313
H	0.842735	0.157265	0.889040	0.110960
H	0.828220	0.171780	0.878944	0.121056
H	0.871550	0.128450	0.906300	0.093700
H	0.867244	0.132756	0.898866	0.101134
H	0.633301	0.366699	0.758717	0.241283
H	0.626190	0.373810	0.760882	0.239118
H	0.616216	0.383784	0.754718	0.245282
H	0.622396	0.377604	0.759027	0.240973
H	0.624576	0.375424	0.759424	0.240576
H	0.614380	0.385620	0.770654	0.229346
H	0.632290	0.367710	0.766501	0.233499
H	0.612558	0.387442	0.752808	0.247192
H	0.492288	0.507712	0.730639	0.269361

APPENDIX F

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of Cellobiose-H$^+$ Ion

With B3LYP/6-311+G(d,p)

Table F.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.648650455	0.956204346	-3.764329942
C	0.028746201	0.844558607	1.573735592
C	-0.525608599	-0.322816119	-2.930264445
C	0.720859185	1.794522354	2.545363538
C	-0.804452209	-0.056410895	-1.443218442
C	1.061067411	0.989198408	3.793753521
C	0.055627374	1.101092113	-0.975713102
C	-0.201267277	0.363247224	4.389686914
C	-0.088371539	2.325505505	-1.906280820
C	-0.954214702	-0.437964800	3.302109787
C	0.850730903	3.463688634	-1.498867222
C	-2.360891775	-0.853430174	3.726538356
O	-0.362748669	1.594544849	0.368095553
O	-0.243225964	0.651908738	-5.046640648
O	-1.460431855	-1.311778684	-3.314458630

O	1.885648064	2.354467084	1.935215049
O	-0.488102103	-1.186002960	-0.651675700
O	1.711208078	1.891293689	4.670175108
O	0.250874336	-0.455985409	5.448204597
O	0.246935232	1.952971459	-3.220608265
O	-1.153092424	0.377013472	2.120662823
O	-2.235119760	-1.550485645	4.966253798
O	0.578616019	3.746355263	-0.084549188
H	-1.673961375	1.355050121	-3.733312230
H	0.686687010	0.036816070	1.238403486
H	0.505188777	-0.683521857	-3.045097148
H	0.019149910	2.594345591	2.818241292
H	-1.864329370	0.211253506	-1.328898877
H	1.745040994	0.170830776	3.522373968
H	1.101990901	0.794351958	-0.895019235
H	-0.855466524	1.165684225	4.761165627
H	-1.129511599	2.681359901	-1.849878699
H	-0.362416816	-1.324602488	3.040688387
H	1.892309334	3.163426228	-1.624564537
H	0.651927764	4.365827256	-2.074688994
H	-2.976803031	0.046764095	3.838597924

H	-2.802202772	-1.489797414	2.952931382
H	-0.469575835	1.365549846	-5.655022640
H	-1.248668071	-1.622266644	-4.203250428
H	2.490183112	2.600519442	2.651874833
H	-0.963028960	-1.945386873	-1.014834300
H	1.940164537	1.423643795	5.483864319
H	-0.509710140	-0.940530796	5.802718723
H	-3.100126405	-1.845012566	5.268847736
H	1.397151012	3.786871349	0.441151455
H	0.008001284	2.656610949	0.347288724

Total Energy: -1298.0118945795 Hartrees

Table F.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.851248	0.148752	6.012369	-0.012369
C	5.807692	0.192308	5.993731	0.006269
C	6.460689	-0.460689	6.078815	-0.078815
C	6.351520	-0.351520	6.079299	-0.079299
C	6.039258	-0.039258	6.071791	-0.071791
C	6.382326	-0.382326	6.084791	-0.084791

C	6.221047	-0.221047	6.043669	-0.043669
C	6.125412	-0.125412	6.090012	-0.090012
C	6.404963	-0.404963	6.080537	-0.080537
C	6.387511	-0.387511	6.071454	-0.071454
C	6.185296	-0.185296	6.029809	-0.029809
C	5.967821	0.032179	6.066148	-0.066148
O	8.067281	-0.067281	8.054284	-0.054284
O	8.259092	-0.259092	8.164664	-0.164664
O	8.242548	-0.242548	8.211382	-0.211382
O	8.253448	-0.253448	8.238883	-0.238883
O	8.229340	-0.229340	8.212276	-0.212276
O	8.241309	-0.241309	8.217377	-0.217377
O	8.256363	-0.256363	8.220434	-0.220434
O	8.114260	-0.114260	8.195964	-0.195964
O	8.079423	-0.079423	8.172676	-0.172676
O	8.364831	-0.364831	8.258575	-0.258575
O	8.345673	-0.345673	8.118359	-0.118359
H	0.827441	0.172559	0.862851	0.137149
H	0.792016	0.207984	0.830394	0.169606
H	0.781972	0.218028	0.849640	0.150360
H	0.774831	0.225169	0.844057	0.155943

H	0.789411	0.210589	0.848252	0.151748
H	0.786665	0.213335	0.851430	0.148570
H	0.781195	0.218805	0.833796	0.166204
H	0.807301	0.192699	0.854577	0.145423
H	0.791161	0.208839	0.852731	0.147269
H	0.796208	0.203792	0.848150	0.151850
H	0.794930	0.205070	0.858304	0.141696
H	0.775788	0.224212	0.857070	0.142930
H	0.825405	0.174595	0.874606	0.125394
H	0.827564	0.172436	0.876691	0.123309
H	0.728580	0.271420	0.810711	0.189289
H	0.729671	0.270329	0.809778	0.190222
H	0.722831	0.277169	0.798325	0.201675
H	0.728382	0.271618	0.805546	0.194454
H	0.723635	0.276365	0.804361	0.195639
H	0.699113	0.300887	0.811133	0.188867
H	0.707438	0.292562	0.820105	0.179895
H	0.691135	0.308865	0.792957	0.207043
H	0.478974	0.521026	0.837232	0.162768

APPENDIX G

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Transition State of Dissociation of the Cellobiose-H$^+$

Glycosidic Bond With RHF/6-31G(d,p)

Table G.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.711134839	0.683418987	-3.876702741
C	1.190115664	0.672944404	1.518021216
C	-1.346128718	-0.184814929	-2.808049688
C	1.113359605	1.813215533	2.502102348
C	-1.205157112	0.457454681	-1.433951225
C	1.083086075	1.222906811	3.901374573
C	0.255751073	0.778473757	-1.169817244
C	-0.171323296	0.376990293	4.010251774
C	0.794187840	1.634703646	-2.332579078
C	-0.162379728	-0.751502689	2.967141689
C	2.270314499	1.977351272	-2.231377819
C	-1.554198338	-1.198422042	2.561963469
O	0.371236891	1.484292692	0.087578957
O	-0.744567705	-0.015399705	-5.048730165
O	-2.711789872	-0.362941702	-3.014095046

O	2.207441247	2.624581204	2.254987318
O	-1.683247547	-0.399541276	-0.446995838
O	1.065012735	2.305022634	4.768424433
O	-0.192497658	-0.157212229	5.290934833
O	0.627018921	0.925650981	-3.519061396
O	0.594141170	-0.433020755	1.754037803
O	-2.172450868	-1.541552278	3.772667939
O	2.424128206	2.702579950	-1.034127975
H	-1.234679667	1.636014035	-3.960268324
H	2.108390822	0.569947347	0.970414058
H	-0.818500178	-1.136286646	-2.812258084
H	0.179662712	2.346345985	2.348100802
H	-1.766995317	1.390665091	-1.426459980
H	1.960483540	0.599639931	4.072269841
H	0.827082080	-0.138844186	-1.088102039
H	-1.031854891	1.022805047	3.848881370
H	0.232080226	2.569470896	-2.356525286
H	0.382898693	-1.584147349	3.385314877
H	2.867202381	1.070436669	-2.225868700
H	2.549160922	2.569808800	-3.094201091
H	-2.073635443	-0.391649594	2.058818883

H	-1.490159075	-2.040795619	1.883240009
H	-0.441226357	0.512296012	-5.771117341
H	-2.858516227	-0.836140018	-3.819124314
H	2.290119412	3.253642989	2.958338496
H	-2.538232009	-0.703291280	-0.717358600
H	0.981795155	2.001384900	5.660792279
H	-1.003729886	-0.625475249	5.428403533
H	-3.035460595	-1.892874520	3.626563668
H	3.284522981	3.078625509	-0.943357068
H	0.937898115	2.248277121	-0.042494062

Total Energy: -1291.0510933461 Hartrees

Table G.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.407811	0.592189	5.820441	0.179559
C	5.436386	0.563614	5.704650	0.295350
C	5.824372	0.175628	5.963119	0.036881
C	5.850979	0.149021	5.958311	0.041689
C	5.763647	0.236353	5.942522	0.057478
C	5.782050	0.217950	5.949690	0.050310

C	5.783198	0.216802	5.937281	0.062719
C	5.805062	0.194938	5.954913	0.045087
C	5.839159	0.160841	5.948217	0.051783
C	5.829851	0.170149	5.923212	0.076788
C	5.876613	0.123387	6.004363	-0.004363
C	5.868156	0.131844	6.009664	-0.009664
O	8.749116	-0.749116	8.259786	-0.259786
O	8.669729	-0.669729	8.319822	-0.319822
O	8.691706	-0.691706	8.353946	-0.353946
O	8.683086	-0.683086	8.341977	-0.341977
O	8.696321	-0.696321	8.352366	-0.352366
O	8.687448	-0.687448	8.342927	-0.342927
O	8.684871	-0.684871	8.345047	-0.345047
O	8.711493	-0.711493	8.310141	-0.310141
O	8.628735	-0.628735	8.188011	-0.188011
O	8.687792	-0.687792	8.368778	-0.368778
O	8.714476	-0.714476	8.374604	-0.374604
H	0.881332	0.118668	0.918638	0.081362
H	0.771522	0.228478	0.854045	0.145955
H	0.846357	0.153643	0.896538	0.103462
H	0.813976	0.186024	0.880730	0.119270

H	0.859397	0.140603	0.902797	0.097203
H	0.848633	0.151367	0.898215	0.101785
H	0.831656	0.168344	0.887397	0.112603
H	0.852322	0.147678	0.899367	0.100633
H	0.854047	0.145953	0.903393	0.096607
H	0.796265	0.203735	0.868009	0.131991
H	0.861263	0.138737	0.899507	0.100493
H	0.840777	0.159223	0.885440	0.114560
H	0.862400	0.137600	0.898888	0.101112
H	0.853252	0.146748	0.892158	0.107842
H	0.635663	0.364337	0.760183	0.239817
H	0.628871	0.371129	0.762037	0.237963
H	0.614859	0.385141	0.753772	0.246228
H	0.628947	0.371053	0.766498	0.233502
H	0.620842	0.379158	0.757837	0.242163
H	0.616307	0.383693	0.764634	0.235366
H	0.630630	0.369370	0.764109	0.235891
H	0.621566	0.378434	0.759201	0.240799
H	0.557058	0.442942	0.752818	0.247182

APPENDIX H

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation

With RHF/6-31G(d,p)

Table H.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.349919367	-0.235562218	-3.862403325
C	-1.089260973	-1.099223408	-2.895105403
C	-1.302518189	-0.364750083	-1.580770510
C	0.024668076	0.204794756	-1.096902882
C	0.611495219	1.174531589	-2.130008853
C	2.114287238	1.391801936	-2.026731442
O	-0.249994180	0.893937361	0.070045547
O	-2.244117475	-1.481709890	-3.539130876
O	-1.811478755	-1.307640396	-0.708268467
O	0.401165668	0.682368183	-3.533805480
O	2.328648719	1.745864216	-0.696625463
H	-0.419455168	-1.946825079	-2.711288403
H	-2.003875123	0.453470435	-1.737609949
H	0.719987576	-0.614519594	-0.923223355
H	0.070719160	2.108296671	-2.113584583

H	2.632572798	0.471944514	-2.287193182
H	2.422037428	2.169952270	-2.716576574
H	-2.755932058	-2.051787884	-2.981789351
H	-2.008217375	-0.912448216	0.129069386
H	3.218434796	2.027742841	-0.552218483
H	0.551734697	1.178462785	0.486200193
H	-0.441964483	-0.415811710	-4.924098116

Total Energy: -607.6590726410 Hartrees

Table H.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.529615	0.470385	5.685065	0.314935
C	5.873691	0.126309	5.961674	0.038326
C	5.774199	0.225801	5.947835	0.052165
C	5.812711	0.187289	5.955717	0.044283
C	5.869707	0.130293	5.910973	0.089027
C	5.863829	0.136171	6.002428	-0.002428
O	8.669084	-0.669084	8.331964	-0.331964
O	8.652145	-0.652145	8.322430	-0.322430
O	8.673517	-0.673517	8.333231	-0.333231

O	8.520634	-0.520634	8.079966	-0.079966
O	8.674688	-0.674688	8.354885	-0.354885
H	0.789302	0.210698	0.859081	0.140919
H	0.831006	0.168994	0.889857	0.110143
H	0.837494	0.162506	0.893549	0.106451
H	0.763723	0.236277	0.854579	0.145421
H	0.863283	0.136717	0.899628	0.100372
H	0.834201	0.165799	0.883336	0.116664
H	0.606180	0.393820	0.743228	0.256772
H	0.616142	0.383858	0.752723	0.247277
H	0.617726	0.382274	0.755195	0.244805
H	0.608543	0.391457	0.759273	0.240727
H	0.718581	0.281419	0.823382	0.176618

APPENDIX I

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation

With B3LYP/6-311+G(d,p)

Table I.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.358327122	-0.245329461	-3.878720287
C	-1.126052418	-1.070762852	-2.916187472
C	-1.312367644	-0.366165429	-1.568446742
C	0.028847294	0.209112212	-1.097878002
C	0.623812373	1.170265245	-2.140179955
C	2.131105725	1.397772392	-2.018168408
O	-0.245427895	0.920565320	0.083605501
O	-2.302310262	-1.464127087	-3.557942077
O	-1.800297267	-1.355780104	-0.699470818
O	0.436809376	0.668348893	-3.567997136
O	2.326219274	1.767203576	-0.662464828
H	-0.450149717	-1.938751556	-2.739892906
H	-2.035208900	0.453058440	-1.687271767
H	0.725528571	-0.620743204	-0.911164054
H	0.072719844	2.112265284	-2.145515707

H	2.664103583	0.469025332	-2.261935020
H	2.448147289	2.183168372	-2.711324996
H	-2.768189772	-2.102693377	-3.001151013
H	-2.015357990	-0.960173797	0.155498132
H	3.237703965	2.043475574	-0.514125400
H	0.588135014	1.241016759	0.459384168
H	-0.450425143	-0.437861401	-4.950671047

Total Energy: -610.9270164099 Hartrees

Table I.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.750039	0.249961	5.808657	0.191343
C	6.135919	-0.135919	6.073545	-0.073545
C	6.193783	-0.193783	6.067879	-0.067879
C	6.041635	-0.041635	6.085662	-0.085662
C	6.621908	-0.621908	6.017868	-0.017868
C	6.043160	-0.043160	6.057947	-0.057947
O	8.257599	-0.257599	8.193482	-0.193482
O	8.232034	-0.232034	8.176112	-0.176112
O	8.265569	-0.265569	8.193262	-0.193262

O	7.915693	0.084307	7.984167	0.015833
O	8.359417	-0.359417	8.245584	-0.245584
H	0.752725	0.247275	0.816030	0.183970
H	0.780330	0.219670	0.844982	0.155018
H	0.785507	0.214493	0.848470	0.151530
H	0.718198	0.281802	0.825975	0.174025
H	0.810511	0.189489	0.870668	0.129332
H	0.800366	0.199634	0.868469	0.131531
H	0.698503	0.301497	0.794830	0.205170
H	0.707074	0.292926	0.800571	0.199429
H	0.689708	0.310292	0.809208	0.190792
H	0.691333	0.308667	0.802975	0.197025
H	0.748990	0.251010	0.813656	0.186344

APPENDIX J

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation-Water

Assembly With RHF/6-31G(d,p)

Table J.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.201762437	-0.211713101	-3.842081720
C	-1.043386135	-1.003737663	-2.897111992
C	-1.243905581	-0.271237237	-1.583476900
C	0.100080254	0.216532524	-1.061200902
C	0.784651395	1.124264271	-2.088849095
C	2.280619844	1.306422925	-1.878706371
O	-0.174050939	0.934614417	0.090245921
O	-2.210067311	-1.356679250	-3.543377673
O	-1.826159936	-1.194965922	-0.730017780
O	0.677938522	0.577232074	-3.464842659
O	2.423114002	1.734008959	-0.559092943
H	-0.421178826	-1.884957508	-2.701160719
H	-1.890714531	0.585817651	-1.746452185
H	0.733268565	-0.644871068	-0.852811562
H	0.271910843	2.072938253	-2.139432548

H	2.784873211	0.358431143	-2.050021543
H	2.663034823	2.032744307	-2.587952891
H	-2.712873348	-1.935899639	-2.987432930
H	-2.007273519	-0.796488609	0.108927174
H	3.317688134	1.967665357	-0.368035109
H	0.629407737	1.214216790	0.506068978
O	-1.922147950	1.247377081	-4.365600870
H	-2.750868459	0.802083941	-4.467253110
H	-0.202233336	-0.450244735	-4.890745991
H	-1.938638852	2.022717638	-4.906059790

Total Energy: -683.7085865161 Hartrees

Table J.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.481235	0.518765	5.684282	0.315718
C	5.860030	0.139970	5.957786	0.042214
C	5.771941	0.228059	5.946934	0.053066
C	5.813516	0.186484	5.955263	0.044737
C	5.857758	0.142242	5.915488	0.084512
C	5.867412	0.132588	6.003825	-0.003825

O	8.675196	-0.675196	8.337263	-0.337263
O	8.666168	-0.666168	8.327655	-0.327655
O	8.681741	-0.681741	8.341112	-0.341112
O	8.554346	-0.554346	8.108336	-0.108336
O	8.677495	-0.677495	8.356834	-0.356834
H	0.797264	0.202736	0.862684	0.137316
H	0.825049	0.174951	0.887085	0.112915
H	0.846037	0.153963	0.897826	0.102174
H	0.777097	0.222903	0.861497	0.138503
H	0.864588	0.135412	0.901187	0.098813
H	0.842005	0.157995	0.886811	0.113189
H	0.609199	0.390801	0.746382	0.253618
H	0.618991	0.381009	0.755226	0.244774
H	0.621277	0.378723	0.757877	0.242123
H	0.609571	0.390429	0.762508	0.237492
O	8.715330	-0.715330	8.405036	-0.405036
H	0.611864	0.388136	0.755664	0.244336
H	0.734805	0.265195	0.830490	0.169510
H	0.620082	0.379918	0.754949	0.245051

APPENDIX K

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation-Water

Assembly With B3LYP/6-311+G(d,p)

Table K.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.342638267	-0.025782391	-3.819956919
C	-1.126228218	-0.896922636	-2.885427944
C	-1.221199396	-0.310062108	-1.480678456
C	0.162360304	0.151292553	-1.016828989
C	0.758296770	1.163362341	-2.009708130
C	2.260751250	1.394534193	-1.847787288
O	-0.036225253	0.760170195	0.237297294
O	-2.371205703	-1.179069642	-3.481992599
O	-1.741442505	-1.347741666	-0.681526013
O	0.619507263	0.707363837	-3.433323875
O	2.447574511	1.749644802	-0.484689209
H	-0.513125572	-1.816779435	-2.826100214
H	-1.892078065	0.558842197	-1.488804224
H	0.825372069	-0.722929938	-0.934473573
H	0.200951110	2.100983239	-1.966737305

H	2.798467968	0.470793994	-2.097604085
H	2.588261130	2.189655922	-2.525266609
H	-2.842585611	-1.813429853	-2.924093743
H	-1.846388613	-1.030935091	0.225100625
H	3.368109131	1.981927252	-0.320412320
H	0.817225350	1.075305390	0.571321338
O	-1.915623970	1.320592918	-4.145912073
H	-2.706009305	0.756307457	-4.195326181
H	-0.317153698	-0.278100746	-4.876918723
H	-1.893135919	1.875205277	-4.938196968

Total Energy: -687.3741022293 Hartrees

Table K. 2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	6.087851	-0.087851	5.853967	0.146033
C	5.937924	0.062076	6.073127	-0.073127
C	6.187983	-0.187983	6.076459	-0.076459
C	6.131740	-0.131740	6.084186	-0.084186
C	6.501131	-0.501131	6.032978	-0.032978
C	6.021375	-0.021375	6.061161	-0.061161

O	8.262050	-0.262050	8.202254	-0.202254
O	8.293113	-0.293113	8.207473	-0.207473
O	8.260640	-0.260640	8.201914	-0.201914
O	7.915585	0.084415	8.039185	-0.039185
O	8.364737	-0.364737	8.248763	-0.248763
H	0.754317	0.245683	0.827038	0.172962
H	0.788085	0.211915	0.845295	0.154705
H	0.796153	0.203847	0.852144	0.147856
H	0.738107	0.261893	0.831066	0.168934
H	0.812111	0.187889	0.871533	0.128467
H	0.810454	0.189546	0.871392	0.128608
H	0.695198	0.304802	0.794162	0.205838
H	0.712327	0.287673	0.802808	0.197192
H	0.696109	0.303891	0.812723	0.187277
H	0.692775	0.307225	0.806544	0.193456
O	8.410109	-0.410109	8.192508	-0.192508
H	0.684922	0.315078	0.787730	0.212270
H	0.764220	0.235780	0.825822	0.174178
H	0.680983	0.319017	0.797767	0.202233

APPENDIX L

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation-$(H_2O)_2$

Assembly With RHF/6-31G(d,p)

Table L.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.479648454	-0.268549629	-3.798004034
C	-1.185948442	-1.098226958	-2.740325200
C	-1.264604950	-0.355115731	-1.421939162
C	0.105149741	0.167469826	-1.022960809
C	0.655603120	1.062453012	-2.137787694
C	2.124530978	1.420967429	-1.969135894
O	-0.100444702	0.878982421	0.151060143
O	-2.451546617	-1.380183698	-3.264415792
O	-1.786049245	-1.259584956	-0.506696453
O	0.610345327	0.390556776	-3.410583981
O	2.261007294	2.025435474	-0.715388215
H	-0.605744604	-2.005787589	-2.604794250
H	-1.926192870	0.504100073	-1.529479401
H	0.780113511	-0.674179369	-0.867458365
H	0.055762884	1.963929958	-2.193213446

H	2.717909299	0.513478071	-2.041915590
H	2.424274392	2.088694479	-2.770780968
H	-2.970621170	-1.845640743	-2.622496289
H	-1.817221679	-0.863753988	0.352385705
H	3.151138056	2.300934675	-0.567122749
H	0.705542121	1.288221649	0.432700225
O	-1.534611395	0.739997490	-4.233296121
H	-2.401035014	0.327403490	-4.174409679
H	-0.259543588	-0.825133771	-4.693831332
H	-1.367496960	1.176870376	-5.107340943
O	-1.050530646	1.817881260	-6.507662648
H	-1.417307159	1.576375187	-7.345306149
H	-0.597777049	2.641439836	-6.614525729

Total Energy: -759.7624546256 Hartrees

Table L.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.455557	0.544443	5.770009	0.229991
C	5.841799	0.158201	5.954390	0.045610
C	5.772161	0.227839	5.954201	0.045799

C	5.788496	0.211504	5.954886	0.045114
C	5.843279	0.156721	5.939773	0.060227
C	5.870588	0.129412	6.005365	-0.005365
O	8.690749	-0.690749	8.350557	-0.350557
O	8.710338	-0.710338	8.358151	-0.358151
O	8.688203	-0.688203	8.347364	-0.347364
O	8.684820	-0.684820	8.261342	-0.261342
O	8.682958	-0.682958	8.361911	-0.361911
H	0.806465	0.193535	0.873155	0.126845
H	0.863970	0.136030	0.904991	0.095009
H	0.857807	0.142193	0.902671	0.097329
H	0.830303	0.169697	0.890620	0.109380
H	0.866357	0.133643	0.902096	0.097904
H	0.862641	0.137359	0.896762	0.103238
H	0.604409	0.395591	0.746143	0.253857
H	0.621104	0.378896	0.757946	0.242054
H	0.628886	0.371114	0.763392	0.236608
H	0.610898	0.389102	0.767609	0.232391
O	8.645529	-0.645529	8.137939	-0.137939
H	0.559064	0.440936	0.722906	0.277094
H	0.786895	0.213105	0.863627	0.136373

H	0.485487	0.514513	0.731965	0.268035
O	8.725065	-0.725065	8.389632	-0.389632
H	0.608578	0.391422	0.745668	0.254332
H	0.607594	0.392406	0.744930	0.255070

APPENDIX M

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry of the Glucosyl Cation-$(H_2O)_2$

Assembly With B3LYP/6-311+G(d,p)

Table M.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-0.417644445	-0.215739246	-3.805756248
C	-1.167676625	-1.049440645	-2.767919566
C	-1.252041778	-0.337773352	-1.426038521
C	0.126625523	0.159863880	-0.996246740
C	0.703359812	1.083474273	-2.082516765
C	2.181715915	1.415595056	-1.881854771
O	-0.087427216	0.853247787	0.214527953
O	-2.455094409	-1.297413810	-3.331747684
O	-1.810391544	-1.276972916	-0.530835230
O	0.669646193	0.445661743	-3.405077962
O	2.303777390	1.970080039	-0.574865667
H	-0.616744202	-1.989075307	-2.643919843
H	-1.908326309	0.541183967	-1.511194079
H	0.797127851	-0.700036860	-0.846613971
H	0.106928930	1.999519129	-2.128406489

H	2.769704606	0.495259587	-1.981315462
H	2.505568734	2.124566844	-2.650911483
H	-3.007944557	-1.739811092	-2.672070333
H	-1.844469339	-0.883936239	0.351366546
H	3.218001413	2.218931101	-0.402959904
H	0.745626413	1.268672899	0.484587909
O	-1.534556944	0.818925035	-4.278430593
H	-2.383994224	0.322775455	-4.171962492
H	-0.186608763	-0.761390460	-4.719358745
H	-1.411402172	1.159535083	-5.233993205
O	-1.200795283	1.716719848	-6.641901250
H	-1.484731994	1.325234979	-7.476830122
H	-0.983181138	2.641378223	-6.812480514

Total Energy: -763.8265383198 Hartrees

Table M.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	6.212638	-0.212638	5.950421	0.049579
C	6.027440	-0.027440	6.066472	-0.066472
C	6.213667	-0.213667	6.084040	-0.084040

C	6.111910	-0.111910	6.084760	-0.084760
C	6.400756	-0.400756	6.057192	-0.057192
C	6.033284	-0.033284	6.064593	-0.064593
O	8.275422	-0.275422	8.215828	-0.215828
O	8.296297	-0.296297	8.222632	-0.222632
O	8.263444	-0.263444	8.209682	-0.209682
O	8.003707	-0.003707	8.139418	-0.139418
O	8.366576	-0.366576	8.254305	-0.254305
H	0.763075	0.236925	0.840560	0.159440
H	0.807191	0.192809	0.852317	0.147683
H	0.805849	0.194151	0.854297	0.145703
H	0.776003	0.223997	0.845055	0.154945
H	0.817851	0.182149	0.873245	0.126755
H	0.823697	0.176303	0.875810	0.124190
H	0.699281	0.300719	0.792382	0.207618
H	0.716278	0.283722	0.804186	0.195814
H	0.705138	0.294862	0.817374	0.182626
H	0.692903	0.307097	0.809809	0.190191
O	8.344738	-0.344738	8.030896	-0.030896
H	0.681862	0.318138	0.775711	0.224289
H	0.772920	0.227080	0.838769	0.161231

H	0.422987	0.577013	0.811069	0.188931
O	8.624932	-0.624932	8.230575	-0.230575
H	0.672500	0.327500	0.799627	0.200373
H	0.667653	0.332347	0.798974	0.201026

APPENDIX N

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry from the Constrained Optimization of

the Glucose-H_3O^+ Assembly With RHF/6-31G(d,p)

Table N.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-1.193573357	-0.081732329	-3.921264654
C	-1.196038485	-1.163029159	-2.853239082
C	-1.055995160	-0.547749925	-1.474508923
C	0.166946268	0.353310454	-1.387024880
C	0.145851399	1.393532539	-2.514234237
C	1.447751599	2.172639126	-2.649118612
O	0.094593943	0.950946613	-0.134496607
O	-2.410845996	-1.827232002	-3.000102060
O	-0.978207343	-1.608227365	-0.575687460
O	-0.025741608	0.718573700	-3.782686536
O	1.716535146	2.777355557	-1.425158161
H	-0.356967719	-1.832472544	-3.030187153
H	-1.930820328	0.064333366	-1.270957823
H	1.063842365	-0.259323939	-1.485077430
H	-0.681887659	2.073163292	-2.368650661

H	2.245724007	1.483261173	-2.929342546
H	1.340128918	2.918339721	-3.436954439
H	-2.525684343	-2.440300394	-2.288440683
H	-0.947450029	-1.275073856	0.308630898
H	2.481141955	3.328390213	-1.469197668
H	0.812845017	1.551191084	-0.004595150
O	-2.287201487	0.743541495	-3.833352818
H	-3.053932291	0.219436703	-3.641419766
H	-1.115531223	-0.529313564	-4.907068394
H	0.680627036	1.039570844	-5.060858447
O	1.102460197	1.180852389	-5.911561868
H	0.788781974	1.929668510	-6.423890271
H	2.051581599	1.044355801	-5.957792255

Total Energy: -759.7285949508 Hartrees

Table N.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.983288	0.016712	6.002061	-0.002061
C	6.174307	-0.174307	6.080446	-0.080446
C	6.190881	-0.190881	6.085039	-0.085039

C	6.071582	-0.071582	6.083929	-0.083929
C	6.417984	-0.417984	6.054379	-0.054379
C	6.098711	-0.098711	6.081938	-0.081938
O	8.272121	-0.272121	8.215148	-0.215148
O	8.273585	-0.273585	8.220403	-0.220403
O	8.276528	-0.276528	8.214151	-0.214151
O	8.353020	-0.353020	8.254009	-0.254009
O	8.352702	-0.352702	8.240519	-0.240519
H	0.807926	0.192074	0.855653	0.144347
H	0.798002	0.201998	0.850367	0.149633
H	0.820153	0.179847	0.858239	0.141761
H	0.747403	0.252597	0.836299	0.163701
H	0.833770	0.166230	0.880408	0.119592
H	0.842663	0.157337	0.887089	0.112911
H	0.717701	0.282299	0.803063	0.196937
H	0.718472	0.281528	0.809182	0.190818
H	0.709955	0.290045	0.817074	0.182926
H	0.701374	0.298626	0.814385	0.185615
O	8.264322	-0.264322	8.185468	-0.185468
H	0.712321	0.287679	0.795057	0.204943
H	0.842582	0.157418	0.870813	0.129187

H	0.358541	0.641459	0.855398	0.144602
O	8.403648	-0.403648	7.835066	0.164934
H	0.631945	0.368055	0.758198	0.241802
H	0.624513	0.375487	0.756221	0.243779

APPENDIX O

Nuclear Coordinates, Energy, and Mulliken and Löwdin Atomic Populations and

Charges Obtained at the Equilibrium Geometry from the Constrained Optimization of

the Glucose-H_3O^+ Assembly With B3LYP/6-311+G(d,p)

Table O.1. Nuclear Coordinates

ATOM	X	Y	Z
C	-1.201556293	-0.083721033	-3.948632743
C	-1.241018454	-1.163020110	-2.863621472
C	-1.063468288	-0.558398378	-1.475942841
C	0.191254300	0.311760303	-1.391420678
C	0.181760584	1.377295875	-2.505446604
C	1.498241477	2.147636936	-2.624690317
O	0.149457435	0.897587556	-0.106179996
O	-2.508476056	-1.779751574	-3.012479480
O	-1.015117745	-1.654782334	-0.580344932
O	-0.002956077	0.715869408	-3.810595813
O	1.752155412	2.744094363	-1.361850865
H	-0.431670563	-1.881414728	-3.047428284
H	-1.925857116	0.080930862	-1.245659697
H	1.081177547	-0.325822937	-1.508888070
H	-0.651383617	2.066342067	-2.351726829

H	2.304705031	1.451311074	-2.896967175
H	1.409113036	2.914872123	-3.405498365
H	-2.644409777	-2.386675787	-2.272634414
H	-0.954198569	-1.316697515	0.321964299
H	2.553373826	3.277682780	-1.396148954
H	0.873630111	1.533456526	-0.019857660
O	-2.302552438	0.765811230	-3.903165562
H	-3.068857303	0.241005509	-3.621827113
H	-1.100231588	-0.548658277	-4.938262389
H	0.676863179	1.117011003	-5.054539027
O	1.133195019	1.355116481	-5.864888973
H	1.613818049	0.642887118	-6.293060491
H	0.611193794	1.868935458	-6.485457089

Total Energy: -763.7893809573 Hartrees

Table O.2. Mulliken (MULL.) and Löwdin (LÖW.) Atomic Populations and Charges

ATOM	MULL.POP.	CHARGE	LÖW.POP.	CHARGE
C	5.983288	0.016712	6.002061	-0.002061
C	6.174307	-0.174307	6.080446	-0.080446
C	6.190881	-0.190881	6.085039	-0.085039

C	6.071582	-0.071582	6.083929	-0.083929
C	6.417984	-0.417984	6.054379	-0.054379
C	6.098711	-0.098711	6.081938	-0.081938
O	8.272121	-0.272121	8.215148	-0.215148
O	8.273585	-0.273585	8.220403	-0.220403
O	8.276528	-0.276528	8.214151	-0.214151
O	8.353020	-0.353020	8.254009	-0.254009
O	8.352702	-0.352702	8.240519	-0.240519
H	0.807926	0.192074	0.855653	0.144347
H	0.798002	0.201998	0.850367	0.149633
H	0.820153	0.179847	0.858239	0.141761
H	0.747403	0.252597	0.836299	0.163701
H	0.833770	0.166230	0.880408	0.119592
H	0.842663	0.157337	0.887089	0.112911
H	0.717701	0.282299	0.803063	0.196937
H	0.718472	0.281528	0.809182	0.190818
H	0.709955	0.290045	0.817074	0.182926
H	0.701374	0.298626	0.814385	0.185615
O	8.264322	-0.264322	8.185468	-0.185468
H	0.712321	0.287679	0.795057	0.204943
H	0.842582	0.157418	0.870813	0.129187

H	0.358541	0.641459	0.855398	0.144602
O	8.403648	-0.403648	7.835066	0.164934
H	0.631945	0.368055	0.758198	0.241802
H	0.624513	0.375487	0.756221	0.243779

APPENDIX P

Initial Coordinates of the Cellulose Iα Fragment Used to Determine Energies Resulting

from the Rotation of Free Hydroxyl Groups in the Central Glucose Residue

ATOM	X	Y	Z
O	-6.036333080	1.445752020	-11.794232370
C	-5.916332720	0.751752020	-10.581229210
C	-7.112332820	-0.322248010	-8.857231140
O	-7.211332800	0.447752090	-10.060231210
C	-5.250332830	1.709752080	-9.616230010
O	-3.980333570	2.140752320	-10.114232060
C	-5.127332690	1.053752180	-8.251231190
O	-4.570333480	1.924752240	-7.271233560
C	-6.476333140	0.553752120	-7.794232850
C	-8.523332600	-0.779248480	-8.555232050
O	-8.569333080	-1.840248580	-7.597232340
H	-5.300020220	-0.198090670	-10.713734630
H	-6.881938460	2.146665570	-11.733567240
H	-6.451449390	-1.235382680	-9.027578350
H	-5.916301730	2.628825660	-9.509511950
H	-3.196169850	1.890203480	-9.384634020
H	-4.432309630	0.156248930	-8.356259350

H	-5.144897460	2.862403870	-7.245045190
H	-7.150511260	1.450803280	-7.593277930
H	-9.003862380	-1.141646620	-9.523405080
H	-9.115878110	0.106132810	-8.149536130
H	-8.568299290	-1.417594910	-6.581672190
O	-6.372333050	-0.229248300	-6.605233190
C	-6.503333090	0.472752180	-5.396232130
C	-5.313332560	1.557752010	-3.672231910
O	-5.209332940	0.800752040	-4.882232190
C	-7.193332670	-0.464248690	-4.428232190
O	-8.481332780	-0.822248520	-4.937232490
C	-7.294332980	0.188751790	-3.058232550
O	-7.856332780	-0.669248340	-2.071231130
C	-5.939332960	0.671752150	-2.610232590
C	-3.914333580	2.020752190	-3.346231700
O	-3.905333280	3.090752120	-2.400232550
H	-7.119487760	1.419799800	-5.548069000
H	-5.980584140	2.467323780	-3.836714270
H	-6.568072800	-1.412418960	-4.330135350
H	-9.077251430	-1.283044340	-4.135641570
H	-7.977369310	1.096066240	-3.157438280

H	-7.823956970	-0.171152560	-1.091000320
H	-5.267289640	-0.229998470	-2.423676490
H	-3.412335400	2.375646350	-4.306256770
H	-3.326948880	1.143623590	-2.915854690
H	-3.920279980	4.052362920	-2.934162620
O	-6.036333080	1.445752020	-1.413231250
C	-5.916332720	0.751752020	-0.200231360
C	-7.112332820	-0.322248010	1.523768190
O	-7.211332800	0.447752090	0.320767490
C	-5.250332830	1.709752080	0.764767230
O	-3.980333570	2.140752320	0.266767590
C	-5.127332690	1.053752180	2.129768130
O	-4.570333480	1.924752240	3.109768630
C	-6.476333140	0.553752120	2.586767200
C	-8.523332600	-0.779248480	1.825766800
O	-8.569333080	-1.840248580	2.783767700
O	-6.372333050	-0.229248300	3.775768760
H	-5.300020220	-0.198090490	-0.332737210
H	-6.451449390	-1.235382680	1.353421210
H	-5.916302200	2.628825660	0.871485050
H	-3.196171280	1.890205030	0.996367220

H	-4.432309630	0.156248930	2.024740460
H	-5.144896510	2.862404580	3.135956290
H	-7.150511260	1.450803520	2.787721630
H	-9.003862380	-1.141647340	0.857594130
H	-9.115879060	0.106132920	2.231462480
H	-8.568299290	-1.417594080	3.799327850
H	-5.924492360	0.377650860	4.576463220
O	-6.930333140	-5.604246620	-9.351230620
C	-7.019332890	-4.942246910	-8.116230960
C	-5.764332770	-3.849246980	-6.413231850
O	-5.719332700	-4.537246700	-7.673232560
C	-7.525332930	-5.980247020	-7.146233560
O	-8.789334300	-6.496245380	-7.579233170
C	-7.614332680	-5.392246720	-5.751232150
O	-8.016333580	-6.360245230	-4.787232880
C	-6.304332730	-4.776247020	-5.340232370
C	-4.346333500	-3.393246650	-6.162232400
O	-4.236333370	-2.418247940	-5.125232220
H	-7.724410530	-4.048603530	-8.178412440
H	-6.024650570	-6.228302960	-9.368309020
H	-6.446252350	-2.939055440	-6.491572860

H	-6.779367450	-6.841966630	-7.122252460
H	-9.316329000	-6.941867830	-6.722671990
H	-8.401782040	-4.568137650	-5.770398140
H	-8.001856800	-5.907437320	-3.784857750
H	-5.548860070	-5.605963710	-5.139157300
H	-3.936345580	-2.942787170	-7.125869270
H	-3.718765740	-4.300653930	-5.875230310
H	-4.391162400	-1.414504170	-5.547756200
O	-6.478333000	-3.989246850	-4.162232880
C	-6.407332900	-4.666246890	-2.934232000
C	-7.688333030	-5.715246680	-1.227231260
O	-7.721333500	-5.016246800	-2.479232070
C	-5.862332820	-3.651247020	-1.966231580
O	-4.602334500	-3.153246880	-2.416232350
C	-5.758332730	-4.252246860	-0.580231960
O	-5.294332500	-3.304246900	0.378767580
C	-7.089333060	-4.816246990	-0.152231980
C	-9.122335430	-6.132246490	-0.963231380
O	-9.227334020	-7.106247430	0.076767180
H	-5.738220690	-5.586496350	-3.005148650
H	-7.040455340	-6.648244380	-1.324028970

H	-6.591661930	-2.776341200	-1.919131280
H	-3.835640430	-3.352243190	-1.652962570
H	-5.005933280	-5.107627390	-0.622881950
H	-5.851921080	-2.363081930	0.263428420
H	-7.805623050	-3.954983230	0.059318210
H	-9.557186130	-6.571972850	-1.920908450
H	-9.723619460	-5.210723400	-0.665151120
H	-9.221291540	-6.601300240	1.054004070
O	-6.930333140	-5.604246620	1.029767750
C	-7.019332890	-4.942246910	2.264768600
C	-5.764332770	-3.849246980	3.967767720
O	-5.719332700	-4.537246700	2.707767490
C	-7.525332930	-5.980247020	3.234769110
O	-8.789334300	-6.496245380	2.801767350
C	-7.614332680	-5.392246720	4.629768370
O	-8.016333580	-6.360245230	5.593767170
C	-6.304332730	-4.776247020	5.040767190
C	-4.346333500	-3.393246650	4.218767170
O	-4.236333370	-2.418247940	5.255767350
O	-6.478333000	-3.989246850	6.218767170
H	-7.724410530	-4.048603060	2.202586650

H	-6.446252350	-2.939055440	3.889426470
H	-6.779367450	-6.841966630	3.258751390
H	-9.316329000	-6.941869260	3.658327580
H	-8.401782040	-4.568137650	4.610602380
H	-8.001856800	-5.907437800	6.596142290
H	-5.548859600	-5.605963710	5.241842270
H	-3.936345580	-2.942787170	3.255130290
H	-3.718765740	-4.300653930	4.505769250
H	-4.391162400	-1.414504170	4.833243370
H	-7.180255890	-3.168224330	6.010857110
O	-6.935332780	-10.847247120	-6.601233480
C	-7.064332960	-10.142247200	-5.393232820
C	-5.859333040	-9.077247620	-3.667231800
O	-5.766333100	-9.826247220	-4.882232190
C	-7.747333530	-11.085247040	-4.425231930
O	-9.030334470	-11.462246890	-4.930232520
C	-7.849332810	-10.429246900	-3.056231740
O	-8.414333340	-11.285246850	-2.067231420
C	-6.491333010	-9.957247730	-2.606231930
C	-4.453333380	-8.637247090	-3.345232250
O	-4.413333890	-7.577248100	-2.389231920

H	-7.678617000	-9.193309780	-5.540790080
H	-6.101336960	-11.560159680	-6.522454260
H	-6.514407160	-8.157596590	-3.824456210
H	-7.111436370	-12.026155470	-4.325605390
H	-9.619670870	-11.925346370	-4.125111100
H	-8.527193070	-9.518215180	-3.156862020
H	-8.381406780	-10.785660740	-1.087777850
H	-5.826973440	-10.864702220	-2.419809100
H	-3.953521010	-8.281087880	-4.305929180
H	-3.876726630	-9.528217320	-2.928943400
H	-4.478533740	-6.610762600	-2.910436870
O	-6.577332970	-9.180247310	-1.410230990
C	-6.464333060	-9.876247410	-0.197232050
C	-7.666332720	-10.941246990	1.527767900
O	-7.761333470	-10.179246900	0.318767640
C	-5.795332910	-8.925247190	0.773768010
O	-4.528333660	-8.486247060	0.275767420
C	-5.668333050	-9.597247120	2.129768130
O	-5.072333340	-8.758247380	3.113768340
C	-7.025332930	-10.067247390	2.589767460
C	-9.076334000	-11.398247720	1.834767580

O	-9.113335610	-12.468247410	2.781768800
H	-5.851245400	-10.828039170	-0.330704450
H	-7.007743360	-11.856605530	1.360502000
H	-6.462604520	-8.008621220	0.892704430
H	-3.742100720	-8.734784130	1.003824470
H	-5.000022890	-10.512186050	2.003878830
H	-5.649207120	-7.825483800	3.198435070
H	-7.685645100	-9.158444400	2.783840890
H	-9.564962390	-11.750272750	0.866820930
H	-9.664563180	-10.516486170	2.254390000
H	-9.112655640	-12.055622100	3.801445010
O	-6.935332780	-10.847247120	3.779768470
C	-7.064332960	-10.142247200	4.987767700
C	-5.859333040	-9.077247620	6.713768010
O	-5.766333100	-9.826247220	5.498767380
C	-7.747333530	-11.085247040	5.955768590
O	-9.030334470	-11.462246890	5.450767040
C	-7.849332810	-10.429246900	7.324769500
O	-8.414333340	-11.285246850	8.313771250
C	-6.491333010	-9.957247730	7.774770260
C	-4.453333380	-8.637247090	7.035769460

O	-4.413333890	-7.577248100	7.991769790
O	-6.577332970	-9.180247310	8.970772740
H	-7.678617000	-9.193309780	4.840210440
H	-6.514407160	-8.157596590	6.556543350
H	-7.111436370	-12.026155470	6.055395130
H	-9.619671820	-11.925347330	6.255887990
H	-8.527193070	-9.518215180	7.224139210
H	-8.381406780	-10.785659790	9.293224330
H	-5.826973440	-10.864702220	7.961193080
H	-3.953520060	-8.281088830	6.075073240
H	-3.876727580	-9.528217320	7.452059270
H	-4.478533740	-6.610762600	7.470564840
H	-7.188256260	-8.284925460	8.783230780
O	0.218666550	6.773753170	-11.798232080
C	0.339666600	6.079752920	-10.585228920
C	-0.855333390	5.014751910	-8.857231140
O	-0.955333410	5.775753020	-10.066231730
C	1.004666810	7.038753030	-9.620231630
O	2.271666290	7.472753050	-10.120230670
C	1.130666730	6.383753300	-8.255230900
O	1.688666340	7.253753190	-7.275233270

C	-0.216333210	5.887752530	-7.797233100
C	-2.262333150	4.570751190	-8.537230490
O	-2.292333130	3.514751910	-7.576232910
H	0.956586540	5.130254270	-10.717373850
H	-0.627617720	7.473818300	-11.737299920
H	-0.200696530	4.096665380	-9.025062560
H	0.336654900	7.956241130	-9.512640950
H	3.058265690	7.221339700	-9.393558500
H	1.824421640	5.485350130	-8.360947610
H	1.113994600	8.191291810	-7.247422220
H	-0.888129060	6.786435600	-7.595592500
H	-2.759208680	4.207392690	-9.496753690
H	-2.844247340	5.460189820	-8.125072480
H	-2.291607620	3.941493030	-6.562383170
O	-0.111333220	5.105751990	-6.607233050
C	-0.248333280	5.805753230	-5.400232790
C	0.941666840	6.889753340	-3.674231770
O	1.044666770	6.136752610	-4.886232850
C	-0.928333340	4.863751410	-4.431231500
O	-2.216333150	4.495751380	-4.928232670
C	-1.033333300	5.519752030	-3.064232110

O	-1.597333430	4.658751490	-2.081231360
C	0.319666620	5.999752520	-2.614232300
C	2.338666440	7.351753230	-3.348232510
O	2.346666340	8.421753880	-2.404232260
H	-0.869819160	6.749514580	-5.550798420
H	0.273163970	7.799135210	-3.834633110
H	-0.295219900	3.920798300	-4.333216190
H	-2.805478810	4.040555950	-4.118476870
H	-1.715948820	6.427184100	-3.165243630
H	-1.565779210	5.153217320	-1.099137780
H	0.992840530	5.098640440	-2.428668020
H	2.841615680	7.705488200	-4.308187480
H	2.925406690	6.474926470	-2.916363720
H	2.333343980	9.382806780	-2.939207790
O	0.218666550	6.773753170	-1.417230960
C	0.339666600	6.079752920	-0.204232020
C	-0.855333390	5.014751910	1.523768190
O	-0.955333410	5.775753020	0.314766970
C	1.004666810	7.038753030	0.760767520
O	2.271666290	7.472753050	0.260767070
C	1.130666730	6.383753300	2.125768420

O	1.688666340	7.253753190	3.105768920
C	-0.216333210	5.887752530	2.583768840
C	-2.262333150	4.570751190	1.843766330
O	-2.292333130	3.514751910	2.804767610
O	-0.111333220	5.105751990	3.773767950
H	0.956586480	5.130254750	-0.336378510
H	-0.200696470	4.096665380	1.355937120
H	0.336654900	7.956241130	0.868358080
H	3.058266640	7.221338750	0.987438320
H	1.824421640	5.485350130	2.020051240
H	1.113995790	8.191292760	3.133579730
H	-0.888128940	6.786435600	2.785409930
H	-2.759206300	4.207390310	0.884242530
H	-2.844249010	5.460190300	2.255921600
H	-2.291607620	3.941494940	3.818616390
H	0.338775900	5.712915420	4.572988510
O	-0.581333400	-0.283248450	-9.340230940
C	-0.726333440	0.406752200	-8.129231450
C	0.492666720	1.507752060	-6.403232570
O	0.554666760	0.841752050	-7.671233650
C	-1.212333440	-0.632248400	-7.155233380

O	-2.441333060	-1.196248530	-7.606232170
C	-1.354333520	-0.032248140	-5.774231910
O	-1.796333070	-0.991248670	-4.818232060
C	-0.043333560	0.563752110	-5.344232080
C	1.909666300	1.953752280	-6.150232320
O	2.020666360	2.927752020	-5.114232540
H	-1.454180480	1.278981210	-8.224389080
H	0.337347150	-0.887504700	-9.310226440
H	-0.190283000	2.418087480	-6.470085620
H	-0.436193410	-1.465606930	-7.103079800
H	-2.960164550	-1.669447660	-6.759534840
H	-2.130090000	0.801501450	-5.825841430
H	-1.781031010	-0.542067410	-3.814239030
H	0.707605840	-0.274985130	-5.164761070
H	2.324941640	2.401317600	-7.112953190
H	2.531158450	1.042930010	-5.860835550
H	1.864326600	3.931448460	-5.536312100
O	-0.204333280	1.324751970	-4.150232790
C	-0.042333510	0.632752120	-2.941231970
C	-1.258333440	-0.481248650	-1.215231300
O	-1.312333460	0.168751810	-2.489232300

C	0.432666900	1.672752020	-1.966231580
O	1.660666470	2.237752440	-2.419232610
C	0.573666810	1.075752140	-0.583232280
O	1.002666710	2.038752320	0.375767320
C	-0.735333380	0.472752180	-0.158231540
C	-2.687333110	-0.901248510	-0.991231260
O	-2.855333570	-1.830248360	0.079767440
H	0.700916530	-0.225557510	-3.043623450
H	-0.583188180	-1.398632760	-1.261977910
H	-0.346624850	2.503392460	-1.917851210
H	2.446713210	2.053527830	-1.672110080
H	1.355465890	0.247382100	-0.630064010
H	0.465253290	2.984376430	0.211506400
H	-1.492895130	1.307583810	0.011345860
H	-3.074200870	-1.383572340	-1.948987480
H	-3.305894140	0.028817770	-0.763286290
H	-2.843020440	-1.290853980	1.038361310
O	-0.581333400	-0.283248450	1.040767430
C	-0.726333440	0.406752200	2.251768110
C	0.492666720	1.507752060	3.977767940
O	0.554666760	0.841752050	2.709768300

C	-1.212333440	-0.632248400	3.225768330
O	-2.441333060	-1.196248530	2.774768830
C	-1.354333520	-0.032248140	4.606767650
O	-1.796333070	-0.991248670	5.562768460
C	-0.043333560	0.563752110	5.036767480
C	1.909666300	1.953752280	4.230768200
O	2.020666360	2.927752020	5.266767020
O	-0.204333280	1.324751970	6.230768200
H	-1.454180360	1.278981330	2.156610250
H	-0.190283180	2.418087240	3.910914900
H	-0.436193410	-1.465606930	3.277922630
H	-2.960164790	-1.669448020	3.621465680
H	-2.130090000	0.801501390	4.555157660
H	-1.781030890	-0.542066930	6.566761020
H	0.707605840	-0.274985130	5.216238020
H	2.324941640	2.401316880	3.268047090
H	2.531158210	1.042930130	4.520165440
H	1.864326480	3.931447980	4.844686510
H	-0.903890730	2.153288360	6.045979500
O	-0.679333330	-5.529246810	-6.598233220
C	-0.811333360	-4.834246640	-5.387232300

C	0.390666720	-3.752246860	-3.667231800
O	0.483666780	-4.496246810	-4.885232450
C	-1.449333670	-5.796247010	-4.406231880
O	-2.708333490	-6.260245800	-4.895232680
C	-1.586333390	-5.115246770	-3.056231740
O	-2.161333320	-5.956246850	-2.062231300
C	-0.234333340	-4.635246750	-2.604232070
C	1.799666400	-3.321246620	-3.351231810
O	1.848666430	-2.272248740	-2.384231810
H	-1.449838640	-3.899451490	-5.521653180
H	0.168692890	-6.226548670	-6.530257230
H	-0.263235570	-2.830746650	-3.818389890
H	-0.761584100	-6.697340010	-4.285268310
H	-3.267870660	-6.734488490	-4.075469490
H	-2.265890360	-4.209105490	-3.185500140
H	-2.125544310	-5.448643210	-1.087009670
H	0.435486560	-5.538846490	-2.418629170
H	2.293905020	-2.956137900	-4.311452390
H	2.376246690	-4.218530650	-2.948693750
H	1.770317670	-1.300301790	-2.893329620
O	-0.320333240	-3.859246730	-1.409231540

C	-0.206333250	-4.554246900	-0.196231650
C	-1.406333450	-5.615246770	1.527767900
O	-1.502333640	-4.857246880	0.317767230
C	0.455666900	-3.601246830	0.777767720
O	1.720666410	-3.156247140	0.282767390
C	0.586666820	-4.275246620	2.131767030
O	1.180666800	-3.438246730	3.118769410
C	-0.768333380	-4.745246890	2.592767720
C	-2.813333270	-6.056246280	1.836766360
O	-2.852333550	-7.100247860	2.808767320
H	0.409172800	-5.504729750	-0.327893500
H	-0.752930280	-6.534678460	1.362513180
H	-0.216493370	-2.688355920	0.897895630
H	2.506645440	-3.401234870	1.012299900
H	1.255755900	-5.189252850	2.003265860
H	0.606103000	-2.503854270	3.201152560
H	-1.428744320	-3.836901900	2.788642410
H	-3.299305440	-6.431315900	0.876165810
H	-3.400720830	-5.162847520	2.232258560
H	-2.851205110	-6.662080760	3.817731140
O	-0.679333330	-5.529246810	3.782768730

C	-0.811333360	-4.834246640	4.993768220
C	0.390666720	-3.752246860	6.713768010
O	0.483666780	-4.496246810	5.495767120
C	-1.449333670	-5.796247010	5.974767680
O	-2.708333490	-6.260245800	5.485766890
C	-1.586333390	-5.115246770	7.324769500
O	-2.161333320	-5.956246850	8.318770410
C	-0.234333340	-4.635246750	7.776769160
C	1.799666400	-3.321246620	7.029768940
O	1.848666430	-2.272248740	7.996770860
O	-0.320333240	-3.859246730	8.971772190
H	-1.449838880	-3.899451730	4.859347340
H	-0.263235570	-2.830746410	6.562609670
H	-0.761584100	-6.697340010	6.095730780
H	-3.267870660	-6.734488490	6.305530070
H	-2.265890360	-4.209105490	7.195501330
H	-2.125544310	-5.448642730	9.293992040
H	0.435486610	-5.538846490	7.962371830
H	2.293905260	-2.956136700	6.069548610
H	2.376246690	-4.218530650	7.432306290
H	1.770317910	-1.300301310	7.487673760

H	-0.935885010	-2.966751580	8.785901070
O	6.467667100	12.113752370	-11.797232630
C	6.586667540	11.423752780	-10.579228400
C	5.395667080	10.348752980	-8.856229780
O	5.292666910	11.112752910	-10.063231470
C	7.250666620	12.386753080	-9.615230560
O	8.516665460	12.824752810	-10.119231220
C	7.376667020	11.734752660	-8.247231480
O	7.920665740	12.611752510	-7.265233040
C	6.032667160	11.221753120	-7.792233940
C	3.989666700	9.899752620	-8.536231040
O	3.960666890	8.844753270	-7.574233060
H	7.207228660	10.476154330	-10.707920070
H	5.617062090	12.809125900	-11.743133540
H	6.053033830	9.433130260	-9.026828770
H	6.580236910	13.302677150	-9.509369850
H	9.306434630	12.571769710	-9.396552090
H	8.081368450	10.844424250	-8.348776820
H	7.350440980	13.552381520	-7.257483480
H	5.353909970	12.114120480	-7.585909370
H	3.494890690	9.533835410	-9.495866780

H	3.404540780	10.787848470	-8.125727650
H	3.961380240	9.272503850	-6.560808660
O	6.141666890	10.435752870	-6.605233190
C	6.004667280	11.133752820	-5.396232130
C	7.190667150	12.228753090	-3.673232320
O	7.296667100	11.464753150	-4.878232480
C	5.323667050	10.191752430	-4.427231790
O	4.035666940	9.824751850	-4.925232410
C	5.219666960	10.849752430	-3.060232400
O	4.657667160	9.987751960	-2.077231650
C	6.571667190	11.337752340	-2.610232590
C	8.592665670	12.698752400	-3.361232040
O	8.618665700	13.775753020	-2.423232320
H	5.383491040	12.077806470	-5.546238420
H	6.517904280	13.134142880	-3.838334320
H	5.956027030	9.248354910	-4.328627590
H	3.445907830	9.369116780	-4.116170880
H	4.533615110	11.754508970	-3.161961560
H	4.689050670	10.482024190	-1.095035310
H	7.248823170	10.440123560	-2.422289370
H	9.084258080	13.046601300	-4.329187870

H	9.184822080	11.824518200	-2.931503770
H	8.574957850	14.733222010	-2.962993620
O	6.467667100	12.113752370	-1.416231510
C	6.586667540	11.423752780	-0.198231500
C	5.395667080	10.348752980	1.524767640
O	5.292666910	11.112752910	0.317767230
C	7.250666620	12.386753080	0.765767630
O	8.516665460	12.824752810	0.261767480
C	7.376667020	11.734752660	2.133767840
O	7.920665740	12.611752510	3.115769150
C	6.032667160	11.221753120	2.588768010
C	3.989666700	9.899752620	1.844767690
O	3.960666890	8.844753270	2.806768420
O	6.141666890	10.435752870	3.775768760
H	7.207228660	10.476154330	-0.326924260
H	6.053033830	9.433130260	1.354167700
H	6.580236910	13.302677150	0.871628640
H	9.306433680	12.571769710	0.984446880
H	8.081367490	10.844424250	2.032222510
H	7.350442410	13.552382470	3.123517990
H	5.353909970	12.114119530	2.795093300

H	3.494890450	9.533834460	0.885132190
H	3.404540780	10.787849430	2.255270240
H	3.961380240	9.272504810	3.820192340
H	6.583579060	11.043806080	4.578878880
O	5.669667240	5.043751720	-9.339231490
C	5.522666930	5.733753200	-8.128232000
C	6.741666790	6.834753040	-6.403232570
O	6.803667070	6.168752670	-7.671233650
C	5.036666870	4.694751260	-7.155233380
O	3.807667020	4.132751460	-7.606232170
C	4.894667150	5.294751640	-5.774231910
O	4.454667090	4.334751610	-4.817232610
C	6.205667020	5.890752790	-5.345232490
C	8.161665920	7.277753350	-6.154232030
O	8.278665540	8.250754360	-5.117232800
H	4.794594760	6.605751510	-8.223778720
H	6.588903900	4.440391540	-9.308218960
H	6.060060020	7.746161940	-6.469157220
H	5.812190060	3.860797170	-7.103427890
H	3.288262840	3.659804820	-6.759744640
H	4.118284700	6.127960680	-5.825154300

H	4.469933510	4.784174440	-3.813347340
H	6.956735130	5.052109240	-5.165863040
H	8.574679370	7.725396160	-7.117889400
H	8.782068250	6.365209580	-5.867936130
H	8.114402770	9.254551890	-5.536050320
O	6.043667320	6.651753430	-4.150232790
C	6.207666870	5.959752560	-2.941231970
C	4.995666980	4.848751070	-1.213231440
O	4.938667300	5.496752740	-2.488231900
C	6.681667800	6.999753000	-1.965231180
O	7.906665800	7.568753240	-2.420232060
C	6.825666430	6.403753280	-0.582231880
O	7.254667280	7.366753100	0.375767320
C	5.517667290	5.801753040	-0.156231690
C	3.569666620	4.430751320	-0.983230890
O	3.410667660	3.494751930	0.081767290
H	6.951090810	5.101655480	-3.044140340
H	5.670788760	3.931376930	-1.260513780
H	5.900202750	7.828264240	-1.915420060
H	8.695453640	7.384056570	-1.676121590
H	7.607604030	5.575545310	-0.629620730

H	6.716731070	8.312095640	0.211595540
H	4.760411260	6.636805060	0.013626520
H	3.174842830	3.955965040	-1.941506510
H	2.954163310	5.359900950	-0.743572000
H	3.421900750	4.028344150	1.043616180
O	5.669667240	5.043751720	1.041767840
C	5.522666930	5.733753200	2.252767560
C	6.741666790	6.834753040	3.977767940
O	6.803667070	6.168752670	2.709768300
C	5.036666870	4.694751260	3.225768330
O	3.807667020	4.132751460	2.774768830
C	4.894667150	5.294751640	4.606767650
O	4.454667090	4.334751610	5.563767910
C	6.205667020	5.890752790	5.035768030
C	8.161665920	7.277753350	4.226768490
O	8.278665540	8.250754360	5.263766770
O	6.043667320	6.651753430	6.230768200
H	4.794594760	6.605751510	2.157220360
H	6.060060020	7.746161940	3.911842820
H	5.812190060	3.860797170	3.277574300
H	3.288262610	3.659804340	3.621255640

H	4.118284700	6.127960210	4.555844780
H	4.469933510	4.784174920	6.567653180
H	6.956735130	5.052109240	5.215137480
H	8.574679370	7.725395680	3.263110640
H	8.782068250	6.365209580	4.513064860
H	8.114402770	9.254551890	4.844948770
H	5.343928810	7.480123520	6.045921330
O	5.561666970	-0.206248310	-6.597232820
C	5.430666920	0.488752100	-5.386232850
C	6.633667470	1.569751980	-3.666232350
O	6.725667000	0.823752050	-4.884232040
C	4.788667200	-0.471248660	-4.406231880
O	3.527666810	-0.932248650	-4.895232680
C	4.653666970	0.209751930	-3.055232290
O	4.074666980	-0.629248380	-2.062231300
C	6.007667060	0.686752140	-2.604232070
C	8.042665480	2.001752140	-3.351231810
O	8.094665530	3.050751920	-2.385232210
H	4.794750210	1.425238970	-5.521149640
H	6.408266070	-0.905267600	-6.529097080
H	5.979691510	2.491111990	-3.817926880

H	5.473621370	-1.374480960	-4.285362720
H	2.967275860	-1.405458450	-4.075455670
H	3.976955180	1.118018510	-3.184509750
H	4.110709670	-0.121766210	-1.086955550
H	6.675887580	-0.218217070	-2.419537310
H	8.535692210	2.366347070	-4.312270640
H	8.619706150	1.104547260	-2.949177500
H	8.011440280	4.022360320	-2.894202230
O	5.923666950	1.462752100	-1.409231540
C	6.037667270	0.767752050	-0.196231650
C	4.836667060	-0.292248040	1.528768300
O	4.741667270	0.462752070	0.316766830
C	6.698667050	1.720752000	0.777767720
O	7.964665410	2.166752100	0.283766840
C	6.829667090	1.047752140	2.132768390
O	7.425666810	1.883752110	3.118769410
C	5.474667070	0.577752050	2.593767170
C	3.429667230	-0.731248380	1.837767720
O	3.388667350	-1.775248890	2.809768680
H	6.654214380	-0.182055530	-0.327894630
H	5.489179130	-1.212657330	1.365445380

H	6.025967120	2.633345840	0.897130430
H	8.749936100	1.922891020	1.014439230
H	7.498026370	0.133223830	2.004183530
H	6.850417140	2.817450050	3.204186680
H	4.814446930	1.486345290	2.789131880
H	2.943222760	-1.105584140	0.877119720
H	2.843550920	0.163014590	2.233193400
H	3.389853480	-1.337098360	3.818739890
O	5.561666970	-0.206248310	3.783768180
C	5.430666920	0.488752100	4.994767670
C	6.633667470	1.569751980	6.714767460
O	6.725667000	0.823752050	5.496768470
C	4.788667200	-0.471248660	5.974767680
O	3.527666810	-0.932248650	5.485766890
C	4.653666970	0.209751930	7.325768950
O	4.074666980	-0.629248380	8.318770410
C	6.007667060	0.686752140	7.776769160
C	8.042665480	2.001752140	7.029768940
O	8.094665530	3.050751920	7.995769500
O	5.923666950	1.462752100	8.971772190
H	4.794750210	1.425238970	4.859850880

H	5.979691510	2.491111990	6.563072680
H	5.473621370	-1.374480960	6.095636370
H	2.967275860	-1.405458450	6.305543900
H	3.976955180	1.118018630	7.196491720
H	4.110709670	-0.121765970	9.294046400
H	6.675887580	-0.218217190	7.961463450
H	8.535692210	2.366347310	6.068730830
H	8.619706150	1.104547020	7.431822780
H	8.011441230	4.022360800	7.486799720
H	5.309601780	2.356350180	8.786283490