Molecular Modeling of Biological Systems: From Chitinase A to Z-DNA

Thesis by

Kenneth A. Brameld

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted August 3, 1998)

To My Parents.

Acknowledgements

The years I have spent at Caltech have been an intense and tremendously rewarding experience. I can state without hesitation that during my tenure here I have enjoyed some of the best times of my life, but then again, there have also been periods when I just haven't wanted to get out of bed in the morning (and not as a result of the poker game the previous evening!). At the heart of Caltech is science, but what I will leave here with is a great fondness for the friends I've made and people I've met. To those people I would like to express my gratitude.

For my advisor Bill Goddard, I have the highest respect. Little did I know when I came to Caltech as an experimentalist that I would leave as (do I dare write it?) a theoretician. Bill is one of the finest and most generous individuals that I have met. I need only to think of my requests for time off for two climbing expeditions to the Himalayas, the trip to McKinley and countless trips to the Sierras, all of which he took in stride. I wonder how many other advisors would have been understanding enough to permit one of their students to disappear for months at a time for such an unacademic pursuit as climbing a mountain. For this, Bill, I am eternally grateful. Bill is also a remarkably good scientist. I can only hope that I leave here having developed some of the same chemical intuition that Bill uses with great success in his fundamental approach to chemistry.

While I don't claim to know everyone in the Goddard group (it would take at least another 5 years), there are some people that truly have made a difference to me. To my office mates Derek and Matt, thanks for all the discussions most of which had nothing to do with science, fortunately. Vaidehi, you are a constant source of optimism and encouragement without which I would have left years ago! Many thanks to Siddharth for being a mentor early on in my grad student career, subsequently an excellent colleague for great science discussions, and most recently, with Barry, for giving me a chance to see the entrepreneurial side of science which has been great. Then there are the behind-the-scenes people without whom life in the Goddard group would have been miserable: Darryl for keeping everything running, Debbie for taming Bill, and Carol for interpreting Bill's hieroglyphics. I also thank Barbara Imperiali and Bill Shrader for starting me down the road of chitinase mechanisms which turned out to be a rich field for computational studies.

It would be impossible to list all my friends at Caltech without missing some. However, some of my best memories involve the following: 201 S. Michigan, the Leslie house, poker nights, Vegas, "my name is Mr. You," Big Wednesday, 556, Ragers, fahrvergnügen, hand and foot, AK-47, shot rule, expedition T-shirts anyone?, big Mak, J-tree, "I smell a rat," and the wall. There are a few people that I will mention by name. Many thanks to the climbing bums I know, Mike, Chris, Sharad, Susan and Peter for moments of panic on rock faces, avalanches, incredible ski runs, big mountains and most of all, keeping me sane. Also thanks to Scott, for being a great roommate and friend of incredible generosity.

To my parents I owe a lot (did I ever pay you back for the truck I sold to go to Dhaulagiri?). I thank you for instilling in me the value of education and for giving me confidence by supporting me in all my endeavors. This thesis is dedicated to you because when all things are said and done, it is you that have provided me with the foundation to succeed and given me a balanced perspective as to what it truly means to be successful in life.

The best thing that happened to me while at Caltech was meeting Natalia. Nata, thank you for your eternal support, for listening to me lament about the latest research disaster and how impossible completing a Ph.D. is. You make me happier than I could possibly have imagined. You are my best friend and the love of my life.

Thesis Abstract

Quantum chemical methods and molecular dynamics simulations are used herein to address interesting problems associated with chemical systems of biological relevance. Two such systems are investigated: The mechanisms of family 18 and family 19 chitinases and the development of a force field for simulations of nucleic acids from first principles calculations.

Chitinases catalyze the hydrolysis of chitin, a β(1,4)-linked *N*-acetyl-glucosamine polymer. Family 18 and family 19 chitinases are glycosyl hydrolases with different structures and mechanisms. Using a combination of quantum chemical and molecular dynamics methods, several interesting and unexpected features of the hydrolysis mechanisms of chitinases were discovered. Family 18 chitinases induce substrate distortion forcing the *N*-acetyl-glucosamine sugar bound at subsite -1 to adopt a boat conformation. Protonation of the β(1,4)-anomeric oxygen leads to spontaneous bond cleavage and the formation of an oxazoline ion intermediate. The oxazoline ion is stabilized through anchimeric assistance from the neighboring *N*-acetyl group. In contrast, family 19 chitinases do not induce substrate distortion and utilize an oxocarbenium ion intermediate. The first of two acidic residues in the active site serves to protonate the β(1,4)-anomeric oxygen while the second acidic residue stabilizes the oxocarbenium ion through a conformational change within a flexible loop of the enzyme. The second acidic residue also coordinates with and activates a water molecule for nucleophilic attack at the C1' anomeric carbon to complete the hydrolysis mechanism.

For molecular dynamics simulations of biomolecules, it is desirable to use accurate potential energy functions (force fields) which are also generic enough to be parameterized for most any conceivable molecule. A hierarchical approach is undertaken herein to achieve this flexibility and accuracy. To begin, a rule based force field (UFF) forms the foundation upon which additional parameters are added so as to reproduce structural and energetic properties important for nucleic acids. The specific substructures within nucleic acids which require additional parameterization are the phosphodiester backbone, sugar ring pseudorotation, glycosidic bond and base pair hydrogen bonding. The potential energy surfaces for each of these substructures are determined from high level quantum mechanical calculations and the force field parameterized to reproduce these results.

Contents

Chapter 4: The Role of Enzyme Distortion in the Single-Displacement Mechanism of Family 19 Chitinases

Chapter 5: Distance Dependent Hydrogen Bond Potentials for Nucleic Acid Base Pairs From *Ab initio* Quantum Mechanical Calculations (LMP2/cc-pVTZ)

List of Figures

List of Tables

Chapter 1

An Introduction to Molecular Modeling Methods and Applications

I. Introduction

Molecular modeling is a suitably generic term to encompass the many computational techniques used in the study of chemical systems. These techniques include and are not limited to the modern methods of quantum mechanics, molecular mechanics and dynamics, and statistical dynamics. The idealized goal of each of these methods is the consistent and accurate prediction of chemical properties using theoretical approaches to alleviate the need for experiments.

Chemistry has long been an experimental field for which hypotheses are generated and tested at the laboratory bench. However, the last several decades have witnessed the birth and rapid development of a new field of chemistry in which experiments are no longer carried out in flasks and beakers, but in the microprocessors of supercomputers. A major factor contributing to growth in this field comes from the tremendous advances in computer technology. As the speed of computers continues to increase while the costs decrease, the question faced by a chemist will change from: "Can I do experiment X?" to "Can I do experiment X more cheaply, easily, accurately and quickly using computational methods rather than traditional bench chemistry?" The reality is that robust computational methods will not replace experiments in the near future. However, the methods discussed in the following sections of Chapter 1 and the

applications presented in the remaining chapters suggest that many useful insights may be gained from theoretical studies of biomolecules.

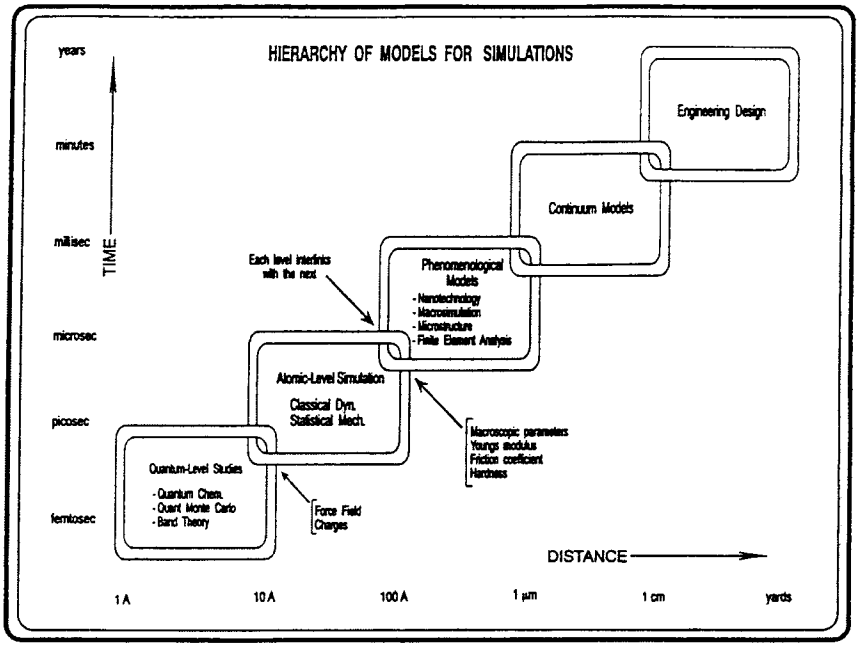


Figure 1.1. A hierarchical view of different theoretical methods which cover a range of time and distance spaces. Biological applications fall within the first two regions with the interface being of particular importance. This figure is courtesy of the MSC.

In the broadest context, the choice of computational method is determined by the nature of the chemical system being studied and the properties of interest. These properties may be characterized in terms of time and distance. Figure 1.1 presents a graphical view of these properties and the related simulation models. At the foundation of the hierarchy lies quantum chemical methods which act on electrons and nuclei in the femtosecond regime typically over a distance of 1-10 Å. The next level is atomistic modeling which includes molecular mechanics (MM) and molecular dynamics (MD) methods. Here the properties of electrons are averaged and the fundamental particles are atoms which obey classical dynamical behavior. At the interface between these two methods, quantum mechanics (QM) and MM/MD, lies the force field. A force field is

simply a collection of mathematical functions which classically describe the potential energy of an atom. The field of chemistry is intimately associated with the time/distance regime of 10^{-13} to 10^{-9} seconds and 1 to 100 Å. Systems with larger dimensions fall into fields such as nanotechnology, material science, and engineering and will not be discussed.

Biological systems are often very complex and can include large biomolecules such as proteins or nucleic acids, medium sized molecules such as cofactors or drugs, and small molecules such as aqueous solvent and counterions. Simulation techniques which are capable of taking into account all these factors explicitly and at the highest level of accuracy are not practical yet. However, judicious use of the correct simulation method for the properties of interest allows one to make considerable progress towards this end. The following sections discuss each of these methods in detail and the types of questions which may be addressed.

II. Quantum Chemical Methods

At the heart of quantum mechanics lies the concept that electrons must be described in terms of a wavefunction Ψ and every observable may be determined by an appropriate operator acting on that wavefunction. Use of the Hamiltonian operator H gives rise to the fundamental Schrödinger equation which states,

$$H\Psi = E\Psi \qquad (1.1)$$

where E is the energy of the system. The solution to this differential equation is a set of allowed wavefunctions or eigenfunctions such that Ψ_n corresponds to an allowed energy E_n. Allowed wavefunctions are continuous functions which satisfy the Pauli principle.

The Schrödinger equation may be solved exactly for the H atom, nothing more (unless the Born-Oppenheimer approximation separating nuclear and electronic motions is made which then allows for the solution of other one electron problems such as H_2^+ and He^+). As a result, all molecular wavefunctions of consequence are approximations. The final goal of most quantum chemistry calculations is a molecular wavefunction Ψ which is sufficiently accurate to calculate the desired observable properties to an acceptable degree of uncertainty.

Using the orbital approximation in which each electron occupies an orbital, the molecular wavefunction may be expanded,

$$\Psi = \phi_1 \phi_2 \phi_3 \ldots \phi_n \tag{1.2}$$

where each function ϕ_i describes the spin-orbital of one electron. The resulting wavefunction must still remain antisymmetric with respect to exchange of electrons due to the Pauli principle. Each of these molecular orbitals ϕ_i is further expanded in terms of atomic orbitals and may be expressed as a linear combination of known atomic orbital functions (solutions to the H atom):

$$\phi_i = \sum_k c_{ik} \chi_k \tag{1.3}$$

where c_{ik} are coefficients and χ_k is an atomic orbital function which itself is a product of a function of r and a spherical harmonic function of θ and ϕ (and a normalization factor). The hydrogen-like atomic orbitals developed by Slater utilize an exponential function of r and are generalizable to any atom, provided the correct exponents appropriate for that particular atom are used. The exact nature of this collection of atomic

orbital functions or "basis set" will be discussed later. Using this basis set of atomic orbitals, the problem of solving for Ψ is thus reduced to a problem of solving for the best coefficients c_{ik} in (1.3).

Just as the Schrödinger equation (1.1) can be applied to the complete molecular wavefunction, it also may be applied to each molecular orbital ϕ_i using the one-electron hamiltonian H,

$$H\phi_i = \varepsilon_i \phi_i \tag{1.4}$$

expanding the linear atomic orbitals of (3),

$$H\sum_k c_{ik}\chi_k = \varepsilon_i \sum_k c_{ik}\chi_k \tag{1.5}$$

multiplying by one basis function χ_i and integrating over all space,

$$\sum_k c_{ik}\left(\int \chi_l H \chi_k dv\right) = \varepsilon_i \sum_k c_{ik}\left(\int \chi_l \chi_k dv\right) \tag{1.6}$$

rewritten as,

$$\sum_k c_{ik}\left(H_{lk} - \varepsilon_i S_{lk}\right) = 0 \tag{1.7}$$

where the following conventional notation is used,

$$H_{lk} = \int \chi_l H \chi_k dv \tag{1.8}$$

and

$$S_{lk} = \int \chi_l \chi_k dv \tag{1.9}$$

The secular equations (1.6) have a solution only when

$$\det\left|H_{lk} - \varepsilon S_{lk}\right| = 0 \tag{1.10}$$

The coefficients c_{ik} are calculated in a "self-consistent" manner by starting with an initial guess, computing all the matrix elements H_{lk} and S_{lk}, solving for the eigenenergies ε_i of (1.10) and back computing the coefficients from (1.6). Because (1.4) depends on ϕ_i which in turn depends on c_{ik} it would seem that one needs to know the coefficients c_{ik} before starting. In practice, after each round of calculations, the new coefficients c_{ik} are substituted into (1.4) and the process is repeated until convergence of c_{ik} is achieved to within a specified limit.

Up to this point, no discussion has been made regarding the exact nature of the Hamiltonian operator H. Many different forms of the Hamiltonian exist which differ in computational expense and rigor. *Ab initio* methods require the calculation of all H_{lk} and S_{lk} integrals explicitly where the Hamiltonian is the Hartree-Fock (HF) self-consistent field operator H^{SCF}.

$$H^{SCF} = H^N + \sum_j J - \sum_j {}' K \tag{1.11}$$

where

$$J_j \phi_i(1) = \left(\int \phi_j^2(2) \frac{1}{r_{12}} dv_2 \right) \phi_i(1) \tag{1.12}$$

and

$$K_j \phi_i(1) = \left(\int \phi_j(1) \phi_i(2) \frac{1}{r_{12}} dv_2 \right) \tag{1.13}$$

The coulombic terms J_j are the interaction of electron (1) in the average field of electron (2). The exchange terms K_j have no classical equivalent and are summed over electrons with the same spin as indicated by the prime notation.

HF calculations are perhaps the most common *ab initio* method used and it is important to understand their limitations. The exact HF solution (assuming it could be achieved with an exhaustive basis set) will not yield an exact solution to the Schrödinger equation, instead reaching the HF limit. The shortcomings of the HF method are manifested in the incorrect dissociation limit for the H_2^+ diatomic molecule. However, the HF method is fairly accurate for predicting properties near the optimum internuclear distance of a diatomic molecule. Therefore, the HF method is capable of accurately predicting optimum geometries and other atomic properties such as electrostatic potentials or dipole moments.

There are two approximations made in the *ab initio* method presented above which lead to an error in the calculated energies. First, the Schrödinger equation does not take into account relativistic effects which play an increasingly important role as the number of core electrons increase. For heavy atoms, the core electrons begin traveling at speeds approaching the speed of light and their masses becomes non-negligible. For biological applications which typically involve first and second row elements only, errors due to relativistic effects are inconsequential. A much more serious error is introduced by the Hartree-Fock operator which calculates the electron-electron repulsion of electron (1) with regards to the average field of electron (2). In reality, the motions of electrons (1) and (2) will be correlated and the probability of finding electron (1) and (2) on the

same side of a nuclei is less than on opposite sides of the nuclei. Thus, the electron correlation energy may be described by,

$$E_{cor} = E_{true} - E_{rel} - E_{HF} \tag{1.14}$$

Electron correlation errors may be further subdivided into dynamical and non-dynamical effects. Dynamical correlation refers to correlation of the motion of electrons and is important for systems near the equilibrium geometry. Non-dynamical correlation accounts for deficiencies in the wavefunction such as the incorrect HF dissociation limit or the inability to include nearly degenerate states.

When considering conformational energies of a molecule, dynamical correlation is of considerable importance and can vary strongly with local changes in bond angles and torsions. One routine method for calculating correlation energies utilizes many-body perturbation theory. The most common perturbation is second-order in which two electrons are excited from occupied orbitals to virtual orbitals. The zeroth-order wavefunction is the HF ground-state wavefunction and the perturbation is typically referred to as Moller-Plesset second-order perturbation (MP2). The second-order term has the form,

$$E_i^{(2)} = \frac{1}{4}\sum_{i}^{occ}\sum_{j}^{occ}\sum_{a}^{virt}\sum_{b}^{virt}\frac{(ij|\ |ab)(ab|\ |ij)}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{1.15}$$

where ϕ_i and ϕ_j are occupied orbitals and ϕ_a and ϕ_b are virtual orbitals. The MP2 calculation is considerably more expensive in terms of computation time and it is still not feasible to carry out a full geometry optimization for large molecules at the MP2 level. It

is much more common to determine the HF optimum geometry with a medium sized basis set, then a single point MP2 energy with a large basis set.

A final discussion of basis sets is in order. The ideal basis set or set of atomic orbitals χ_k that each molecular orbital is expanded into are Slater-type atomic orbitals which have the form,

$$\chi_k = Ce^{-\sigma r}Y_{lm} \tag{1.16}$$

where Y_{lm} is the angular component of the function and ς is the orbital exponent. Integrals over exponential functions are difficult and a more convenient form is to fit several gaussian functions to reproduce each atomic orbital. Using three gaussian functions gives rise to the STO-3G basis set (Slater-type orbital - 3 gaussian). A minimal basis set contains one Slater-type function for each atomic orbital. The double zeta basis set will contain two such functions for each atomic orbital. A further refinement is achieved with a split-valence basis set most commonly of the form 6-31G**. Here the basis set of the core electrons are contracted. In the case of a first-row atom, six gaussian functions represent the core 1s orbital while the valence orbitals are represented by two sets of functions. One function is comprised of three gaussians and the second by a single gaussian, hence the 6-31G notation. A single * indicates polarization functions on heavy atoms and a ** indicates additional polarization functions on both heavy atoms and hydrogen.

There are many more different flavors of quantum chemistry calculations which have not been discussed. However, for biological applications which deal with predominately first-row elements, excellent results may be obtained using HF/6-31G**

for geometry optimization followed by an LMP2/cc-pVTZ single point energy. The calculations presented herein use almost exclusively these methods. The practical limit for modern QM programs with a multi-processor supercomputer is in the 10 to 100 atom range. Therefore, simulations of several hundred or thousand atoms require an alternative approach; molecular mechanics and dynamics.

III. Molecular Mechanics and Dynamics

For quantum chemical methods the fundamental entities of interest are electrons and nuclei. In contrast, molecular mechanics averages over the electrons and nuclei and deals with atoms. Furthermore, the interactions between atoms are treated classically and thus a much larger collection of atoms may be included in a given simulation than for an *ab initio* calculation. The simplest representation of an atom is a sphere with a fixed radius, mass and charge located at the center of the sphere. The forces which act on this atom are determined using a force field (FF) which includes terms for nonbonded and valence interactions. Given energies and forces, it is a simple matter of classical physics to calculate dynamics trajectories or optimum structures.

The FF is of critical importance for MM/MD simulations. Such calculations can only be as accurate as the potential functions which govern atomic interactions. Therefore, it is useful to examine those potentials in greater detail. The total MM energy may be described as,

$$E_{tot} = E_{nonbond} + E_{valence} \tag{1.17}$$

where $E_{nonbond}$ and $E_{valence}$ are the nonbond and valence contributions respectively. The nonbond potential may be further separated into Coulombic and van der Waals

components. Each atom has associated with it an atomic charge. The electrostatic energy is then just the coulombic interaction between each atom.

$$E_{elec} = \sum \frac{q_i q_j}{r_{ij}} \tag{1.18}$$

Similarly, the van der Waals energy is also a pairwise interaction. Two common potential forms are used; a Leonard-Jones 6-12 and a Buckingham exponential-6:

$$E_{LJ} = AR^{-12} - BR^{-6} \tag{1.19}$$

and

$$E_{\exp 6} = Ae^{-CR} - BR^{-6} \tag{1.20}$$

The exponential-6 form has a softer inner wall and is a better representation of the true potential but requires three parameters rather than two.

Valence potentials describe bonding interactions between atoms. The simplest potential form for bond stretching and angle bending are harmonic. However, the harmonic potential only describes those regions close to the minimum and certainly do not allow for correct bond cleavage.

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion} \tag{1.21}$$

where

$$E_{bond} = \frac{1}{2} K_{IJ} (R - R_0)^2 \tag{1.22}$$

$$E_{angle} = \frac{1}{2} K_{IJK} (\theta - \theta_0)^2 \tag{1.23}$$

Torsion potentials, which apply to four bonded atoms (IJKL), are slightly more complex and are periodic in nature. Intuitively, these interactions reproduce the barriers to rotation about a given bond and are represented by a Fourier series of the form

$$E_{torsion} = \frac{1}{2}\sum_{n=1}^{6} K_{\theta,n}(1 - d\cos(n\phi) \qquad (1.24)$$

where $K_{\theta,n}$ is the torsion energy barrier, $d = \pm 1$ and is the phase factor and ϕ is the torsion angle ($\phi = 0$ for cis). Torsion potentials are most critical in predicting conformational energies. Finally, an inversion term (improper torsion) is needed to enforce a planar geometry for sp^3 hybridized atoms, such as aromatic amines or the barrier to inversion for ammonia. For cases where the nonplanar geometry is the minimum:

$$E_{inversion} = \frac{1}{2}C_I(\cos\omega - \cos\omega^0)^2 \qquad (1.25a)$$

where

$$C_I = \frac{K_{inv}}{(\sin\omega^0)^2} \qquad (1.25b)$$

However, for cases where the planar geometry is the minimum:

$$E_{inversion} = K_{inv}(1 - \cos\omega) \qquad (1.26)$$

This completes the list of standard potential forms. There are times when a more specialized potential may be implemented such as for hydrogen bonding or many-body potentials for metals.

Given these various potential forms, the next task is parameterization of the FF with the correct force constants and equilibrium geometries. It is these parameters which

define the quality of a FF. Various approaches have been used to define atom types and associated parameters. In the most general case of the Dreiding FF,[1] all bond and angle force constants are 700 kcal/mol Å (1400 kcal/mol Å for double-bonds) and 100 kcal/mol deg respectively. Equilibrium bond distances are derived from atomic radii and equilibrium bond angles are derived from the atom hydrides. This type of generic FF is most useful for systems which have unusual arrangements of atoms or for a new system for which no experimental or QM data is available. An additional level of complexity may be achieved by allowing for bond order corrections as is done with UFF.[2] Once again, this is a generic FF which has parameters for most any conceivable arrangement of atoms and bonds, but now the force constants are somewhat dependent on the exact atomic constituents which form the bonds.

At the other extreme are highly specialized spectroscopic FFs which include many complex cross-terms such as coupling between bond stretching and angle stretching modes for the purpose of reproducing vibrational spectra. These may be so finely tuned so as to be useful for only a very limited class of molecule. The commonly used biological FFs which include Amber,[3-5] CHARMM[6] and OPLS[7] tend to be closer to the category of specialized FFs and fail to be suitably general. The high degree of parameterization makes them ideal for simulations of systems which contain naturally occurring amino acids or nucleic acids. However, any unusual ligands such as drug molecules, cofactors, substrates or post-translational modifications are difficult to incorporate in a self-consistent manner. Unfortunately, the generic FFs are not accurate enough for simulations of biomolecules for which the exact molecular structure plays such a critical role. Hence there is a need for a generic FF which incorporates just

enough specificity for accurate simulations of biomolecules but still allows the flexibility to model most any organic molecule.

IV. Application to Real Problems

The remainder of this thesis presents the results of studies that apply the modern methods of quantum chemistry and molecular mechanics and dynamics to interesting problems of a biomolecular origin. Two separate cases are discussed that differ substantially from a molecular biology viewpoint but may be reduced to the same principles of molecular modeling. The first case involves the elucidation of the mechanism of family 18 and family 19 chitinases. The second case presents an *ab initio* derived generic FF for the simulation of nucleic acids. Both of these applications rely heavily on a combination of QM and MM/MD methods and demonstrate the feasibility of solving difficult problems without the need for traditional experimental chemistry.

Chitin, a β(1,4)-linked *N*-acetyl-glucosamine (GlcNAc) polysaccharide, is a major structural component of fungal cell walls and the exo-skeletons of invertebrates, including insects and crustaceans. This linear polymer may be degraded through the enzymatic hydrolysis action of chitinases. Chitinases have been found in a wide range of organisms including bacteria,[8,9] plants,[10] fungi,[11] insects,[12] and crustaceans.[13] For those organisms that utilize the structural properties of chitin, chitinases are critical for the normal life cycle functions of molting and cell division.[14,15] Because chitin is not found in vertebrates, it has been suggested that inhibition of chitinases may be used for the treatment of fungal infections and human parasitosis.[16]

Based on amino acid sequence, the glycosyl hydrolases have been classified into 45 families.[17] Using this classification method, the chitinases form families 18 and 19 which are unrelated, differing in structure and mechanism. Sequence analysis shows little homology between these classes of chitinases. Family 19 chitinases (found in plants) share the bilobal $\alpha + \beta$ folding motif of lysozyme, which forms a well defined substrate binding cleft between the lobes.[18,19] In contrast, family 18 chitinases share two short sequence motifs which form the catalytic $(\beta\alpha)_8$-barrel active site.[20] Family 18 chitinases with diverse sequences have been isolated from a wide range of eukaryotes and prokaryotes.

The hydrolysis mechanisms of the chitinases have been investigated by examining the reactivity of the chitin substrate alone and in the presence of the enzyme. The results of these studies are presented in Chapter 2. It was found that anchimeric assistance from the neighboring *N*-acetyl group is critical in stabilizing the resulting oxazoline ion intermediate. MD simulations of the complete enzyme with bound substrate offer further insights to the differing mechanisms of family 18 and 19 chitinases. The oxazoline ion intermediate forms as a result of substrate distortion induced within the active site of family 18 chitinases (Chapter 3). Yet surprisingly, the family 19 chitinases do not utilize an oxazoline ion intermediate and undergo a considerable change in enzyme conformation to stabilize the resulting oxocarbenium ion intermediate (Chapter 4).

The remaining chapters (5-7) address the development of a generic FF for simulations of nucleic acids. The goal of this work is the parameterization of a biological force field (MSCFF) from first principles that maintains the flexibility of a

generic FF. To this end, the MSCFF is built upon the foundation of a generic FF (UFF[2]) with additional terms added which are specific to nucleic acids. These terms are derived from high level *ab initio* QM calculations of small clusters which accurately represent the relevant potential energy surfaces for the complete nucleic acid system.

The nature of these clusters or model systems is defined by the fundamental subunits of nucleic acids: base, sugar and phosphodiester backbone. For the bases, the key interactions involve hydrogen bonding. Chapter 5 presents a detailed examination of the hydrogen bond potentials for GC and AT base pairs. For the 2'-deoxyribose or ribose sugars, the critical issue is pseudorotation of the pentofuranose ring. Interconversion of sugar pucker conformations is critical in defining the overall helical parameters of DNA or RNA. The potential energies of pseudorotation and the conformational preferences of 2'-deoxyribose and ribose sugars are the topic of Chapter 6. Finally, the complete MSCFF that includes additional terms for the phophodiester backbone is presented in Chapter 7. Here too are a series of detailed validation studies that demonstrate the accuracy of the FF for simulations ranging from isolated bases to a complete crystal structure of Z-DNA with explicit solvent molecules and counterions.

V. References

(1) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(2) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.

(3) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

(4) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230-252.

(5) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(6) Mackerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(7) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810-2819.

(8) Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka, H. *J. Biol. Chem.* **1990**, *265*, 15659-15665.

(9) Perrakis, A.; Tews, I.; Dauter, Z.; Oppenheim, A. B.; Chet, I.; Wilson, K. S.; Vorgias, C. E. *Structure* **1994**, *2*, 1169-1180.

(10) Collinge, D. B.; Kragh, K. M.; Mikkelsen, J. D.; Nielsen, K. K.; Rasmussen, U.; Vad, K. *Plant J.* **1993**, *3*, 31-40.

(11) Bartnicki-Gracia, S. *Annu. Rev. Microbiol.* **1968**, *22*, 87-108.

(12) Kramer, K. J.; Dziadik-Turner, C.; Koga, D. *Comprehensive Insect Physiology, Biochemistry, and Pharmacology: Integument, Respiration, and Circulation. Vol. 3*; Pergamon Press: Oxford, 1985.

(13) Koga, D.; Isogai, A.; Sakuda, S.; Matsumoto, S.; Suzuki, A.; Kimura, S.; Ide, A. *Agri. Biol. Chem.* **1987**, *51*, 471-476.

(14) Fukamizo, T.; Kramer, K. J. *Insect Biochem.* **1985**, *15*, 141-145.

(15) Kuranda, M. J.; Robbins, P. W. *J. Biol. Chem.* **1991**, *266*, 19758-19767.

(16) Robertus, J. D.; Hart, P. J.; Monzingo, A. F.; Marcotte, E.; Hollis, T. *Can. J. Bot.* **1995**, *73*, S1142-S1146.

(17) Henrissat, B.; Bairoch, A. *Biochem. J.* **1993**, *293*, 781-788.

(18) Hart, P. J.; Pfluger, H. D.; Monzingo, A. F.; Hollis, T.; Robertus, J. D. *J. Mol. Biol.* **1995**, *248*, 402-413.

(19) Monzingo, A. F.; Marcotte, E. M.; Hart, P. J.; Robertus, J. D. *Nat. Struct. Biol.* **1996**, *3*, 133-140.

(20) Terwisscha van Scheltinga, A. C.; Hennig, M.; Dijkstra, B. W. *J. Mol. Biol.* **1996**, *262*, 243-257.

Chapter 2

Substrate Assistance in the Mechanism of Family 18 Chitinases: Theoretical Studies of Potential Intermediates and Inhibitors

Abstract

Based on first principles and molecular mechanics calculations, the mechanism of hevamine (a family 18 chitinase) is consistent with an oxazoline ion intermediate stabilized by the neighboring C2′ acetamido group. In this intermediate, the acetamido carbonyl oxygen forms a covalent bond to C1′ of *N*-acetyl-glucosamine and has a transferred positive charge from the pyranose ring onto the acetamido nitrogen leading to an anchimeric stabilization of 38.1 kcal/mol when docked with hevamine.

This double displacement mechanism involving an oxazoline intermediate distinguishes the family 18 chitinase (which have one acidic residue near the active site) from family 19 chitinase and from hen egg white lysozyme which have two acidic residues near the active site.

The structural and electronic properties of the oxazoline intermediate are similar to the known chitinase inhibitor allosamidin, suggesting that allosamidins act as transition state analogs of an oxazoline intermediate. Structural and electronic features of the oxazoline ion likely to be important in the design of new chitinase inhibitors are discussed.

I. Introduction

Chitinases catalyze the hydrolysis of chitin, a β(1,4)-linked *N*-acetyl-glucosamine (GlcNAc) polysaccharide (Figure 2.1), and are critical for the normal development of insects[1] and fungi.[2] Structurally diverse chitinases have also been isolated from plants, where they may act as a defense against fungal pathogens.[3,4] The five known classes of chitinases are grouped into two glycosyl hydrolase families.[5] Family 18 consists of class III and class V chitinases found in a wide range of organisms including bacteria, plants, animals and fungi. Family 19 consists of class I, II, and IV chitinases and are found only in plants. The enzyme structures and hydrolysis mechanisms for the two families appear to be quite different. A knowledge of the reaction mechanism is important for the design of new transition state analogs, which may act as chitinase inhibitors and potential insecticides[6,7] or fungicides.[8]

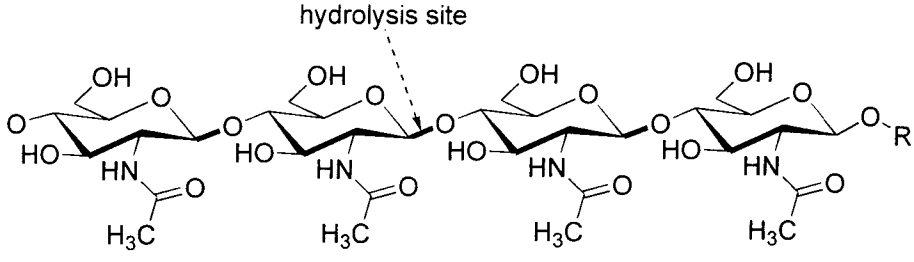


Figure 2.1. An arrow marks the hydrolysis site of chitin, the β(1,4)-*N*-acetyl-glucosamine (GlcNAc) polysaccharide substrate of chitinases.

Reported herein are the use of first principles methods and molecular mechanics modeling to examine the reaction mechanism for hevamine, a family 18 chitinase. First *ab initio* quantum mechanics (QM) was used to study (in both the gas and solution

phases) three possible intermediates formed during the enzymatic hydrolysis of chitin (**1-3**, Figure 2.2). Second a force field (FF) was developed that correctly describes the structure and energy differences between these three intermediates for use in molecular dynamics (MD) studies. This FF was then used to examine the structures of these three intermediates (with a (GlcNAc) dissacharide substitution at O4′ (**4 -6** , Figure 2.2)), bound to the active site of hevamine, a family 18 chitinase. Finally, these structures were used for single point QM energy calculations including the effects of the enzyme active site and solvation.

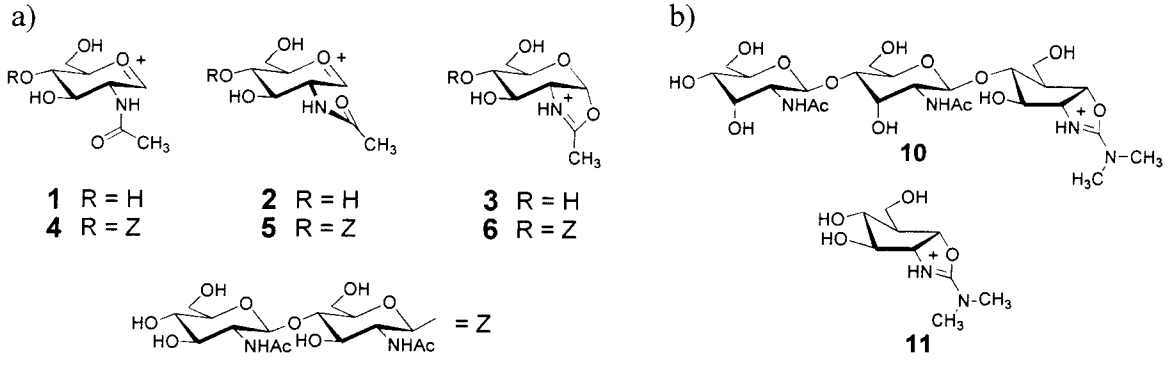


Figure 2.2. (a) Three putative intermediates involved in the chitinase hydrolysis mechanism were examined. **1** and **4** are the oxocarbenium species with an extended *N*-acetyl geometry; **2** and **5** have a rotated *N*-acetyl geometry; **3** and **6** are the global minimum oxazoline ion structures. (b) Allosamizoline (**11**) is the aglycone of allosamidin (**10**), a potent chitinase inhibitor.

These results support the intermediacy of an oxazoline species in the reaction mechanism. Such a mechanism differs substantially from that of the family 19 chitinases and hen egg white lysozyme (the best studied glycosyl hydrolase). The electrostatic potential surfaces of the hevamine bound oxazoline intermediate and allosamidin, a

potent chitinase inhibitor,[6,9] are quite similar, supporting the proposal that the allosamidins act as transition state analogs.[10]

II. Review of Mechanistic Understanding

Extensive studies of the mechanism for hen egg white lysozyme (HEWL) show that glycoside hydrolysis requires two acidic residues (Glu35 and Asp52), one of which is protonated.[11] The consensus view[12] of the mechanism (Figure 2.3a) involves protonation of the β(1,4)-glycosidic oxygen, leading to an oxocarbenium ion intermediate which is stabilized by the second carboxylate (either through covalent or electrostatic interactions). Nucleophilic attack by water yields the hydrolysis product which necessarily retains the initial anomeric configuration. This is commonly referred to as the double displacement mechanism of hydrolysis.[12-15]

Crystallographic analysis of family 18 chitinases isolated from both bacteria[16] and plants[17,18] reveal a common structural motif within two short homologous sequences. Thus, the family 18 chitinases share a similar active site, located at the carboxy-terminus and comprising a $(\beta\alpha)_8$-barrel domain. Hydrolysis has been reported to proceed with retention of the anomeric configuration,[11,19] consistent with the double displacement hydrolysis mechanism, as in HEWL. However, for hevamine and other family 18 chitinases, the three-dimensional structure clearly shows only one acidic residue in the active site.[17,18] Thus, there must be an alternative mechanism which allows a double displacement reaction to take place, yet stabilizes the oxocarbenium ion intermediate without requiring a second carboxylate residue.

The spontaneous acid catalyzed hydrolysis of 2-acetamido substituted polysaccharides in solution has been reported to occur through anchimeric assistance by the acetamido carbonyl group of a GlcNAc residue[20,21] (2.3b).

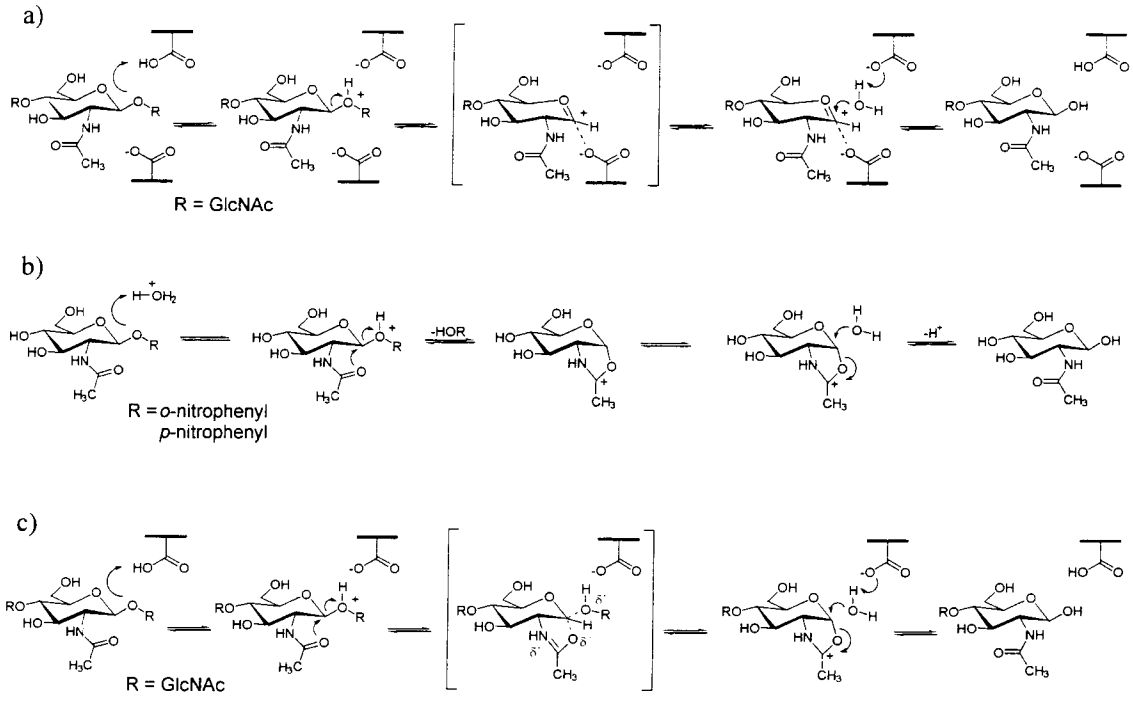


Figure 2.3. (a) The double-displacement mechanism of lysozyme. (b) The proposed anchimeric assisted mechanism of spontaneous hydrolysis of o- and p- nitrophenyl 2-acetamido-2-deoxyglucopyranosides.[20,21] (c) The proposed mechanism for chitinase involving an oxazoline ion intermediate.

It was proposed that anchimeric participation may play a role in the enzymatic mechanism of lysozyme.[22] This proposal was later disproved for HEWL by mutagenesis studies which indicate two acidic residues are essential for enzyme activity.[23] However, such a mechanism could conceivably be operative for the family 18 chitinases,[10,15,24] goose egg white lysozyme (GEWL),[25,26] and soluble lytic transglycosylase.[27,28]

Experimental evidence in support of the anchimeric assistance mechanism is provided by the crystal structure of the inhibitor allosamidin (**10**) bound to the active site of hevamine,[10] where the possibility of "anchimeric assistance by the C2′ *N*-acetyl" group is suggested. Such C2′ acetamido participation would offer a simple mechanistic alternative (Scheme 2.3c) for hydrolysis by the family 18 enzymes, which lack a second active site carboxylate group capable of stabilizing an oxocarbenium ion.

III. Methods

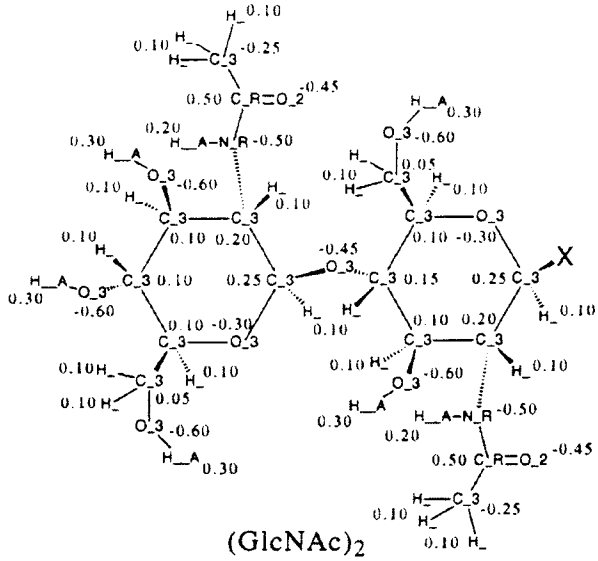
The *ab initio* QM calculations (HF/6-31G**) were carried out with the PS-GVB 2.24 program[29] from Schrodinger, Inc. Initial geometries of **1-3** were obtained using the Dreiding FF[30] with the PolyGraf MD program which is distributed by Molecular Simulations Inc., San Diego. The geometry was then optimized using QM. No constraints were placed on **1** and **3**. However, to determine the energetic effects of anchimeric stabilization, the C--N--C2′--C3′ torsion angle of **2** was constrained to 68.5°. This torsion angle was chosen because it is a local minimum determined from the unconstrained geometry optimization of **3**. Solvent polarization was included self-consistently in the QM wavefunction using the Poisson-Boltzmann continuum description ($\varepsilon = 80.37$) outside the solvent accessible surface (probe radius = 1.2 Å).[31] The geometry optimizations were carried out in the gas phase. For **3**, we also optimized the geometry while simultaneously including the forces due to solvation. However, these calculations found the change in geometry to be insignificant (RMS coordinate difference 0.03 Å).

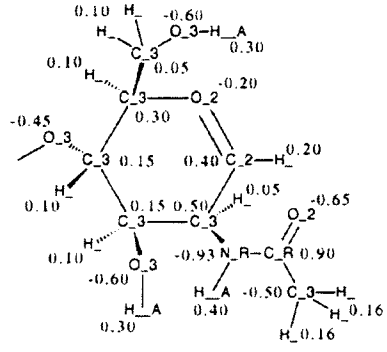
Atomic point charges were determined[32-34] from the electrostatic potential derived from the electron density distribution (constrained to reproduce the molecular monopole and dipole moments) calculated from the converged wavefunctions. Electrostatic potential surfaces were calculated and displayed using the PS-GVB output with the Spartan 4.1.2 visualizer, distributed by Wavefunction Inc., Irvine, CA.

MD studies for the isolated intermediates were carried out with the Dreiding FF[30] using the PolyGraf molecular modeling program. A standard coulomb potential was used with no distance dependent dielectric and all nonbond interactions considered explicitly. The O_2 atom type van der Waals radius was reduced to 3.00 Å, and all other parameters were unchanged.

The starting structures for the (GlcNAc) disaccharide of intermediates **4-6** were taken from the allosamidin/hevamine crystal structure and merged with the aglycone intermediates **1-3**. Docking of these glycosylated intermediates with hevamine was achieved in a three step process using PolyGraf. First, a least squares alignment was carried out matching the oxazoline **6** and constrained oxocarbenium ion **5** intermediates with the allosamidin/hevamine complex. The allosamidin was then deleted from the structure, leaving only the docked intermediates. In the case of the extended *N*-acetyl geometry of **4**, annealing simulations using MD were carried out to determine the optimum binding position.

QM derived charges were used for the intermediates (Figure 2.4) and the default Charmm charges invoked with the Dreiding FF were used for hevamine. The total charge within the active site was zero with a net charge of +1 on each intermediate and -1 on Asp 125. A spline cutoff of 13.5 was used for the nonbond terms.





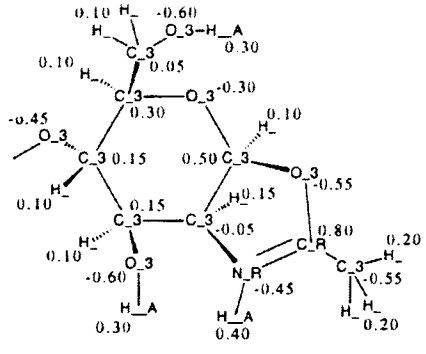


Figure 2.4. Charges and FF atom types used for MD calculations of intermediates **4-6**.

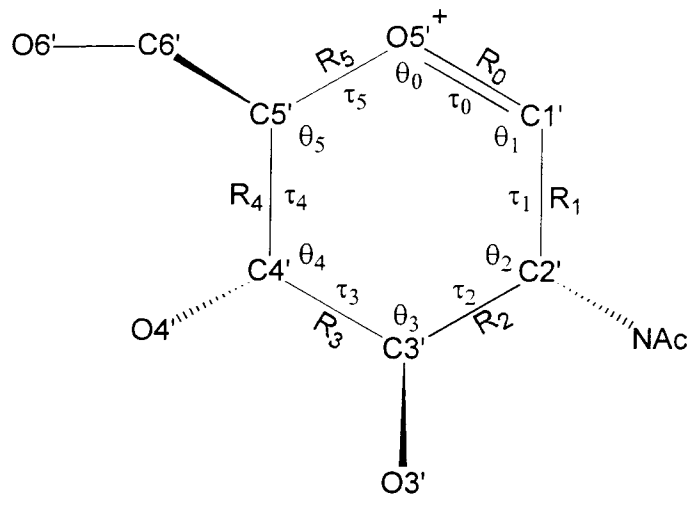


Figure 2.5. Atom numbering and definition of geometric parameters for the oxocarbenium ion.

A series of **10** cycles of MD quenching simulations from 50-300° K for **2** ps each was carried out for each intermediate. For the first **5** cycles, the enzyme was fixed in the crystallographic conformation. For the remaining **5** cycles, the five active site residues (Asp 125, Glu 127, Gln 181, Tyr 183, and Trp 255) were allowed to relax in addition to the intermediates. Energy minimization of the annealed structures yielded the final docked conformations. Coordinates obtained from these structures optimized with the FF were used for the *ab initio* QM calculations (HF/6-31G**). These clusters (**7**-**9**) included the hydrogen terminated amino acid sidechains of residues Asp 125, Glu 127, Gln 181, Tyr 183, and Trp 255.

IV. Results and Discussion

Energetics of Hydrolysis Intermediates

To explore the possibility that an oxocarbenium ion intermediate may be internally stabilized by the acetamido carbonyl group of a GlcNAc residue, QM

calculations on a C1′-deoxy-*N*-acetyl-glucosamine ion were carried out. The global energy minimum leads to the oxazoline species, **3**, in which the acetamido carbonyl group rotates towards C1′, forming a covalent bond. In order to determine the magnitude of this stabilization, two additional conformations (**1** and **2**) were examined which differ in the orientation of this internal carbonyl; either extended or constrained by fixing the C--N--C2′--C3′ torsion angle. Conformation **1** is extended and has the carbonyl group away from the active site while conformation **2** orients the carbonyl group to be parallel to and below the plane of the hexose ring, enabling it to interact with the positive O5′ without forming a covalent bond. Table 1 lists the relative energies for each conformation in the gas phase and solvated in water (using the Poisson-Boltzmann continuum solvent approximation in conjunction with the solvent accessible surface).

Table 2.1. Relative Energies of Hydrolysis Intermediates (kcal/mol).

Conformation	HF/6-31G**	HF/6-31G**(solv)[a]	Dreiding FF[b]
1	21.6	20.0	20.5
2	15.6	16.9	16.5
3	0.0[c]	0.0[d]	0.0[e]

[a]Solvation energy calculated[31] using a Poisson-Boltzmann continuum description (ε = 80.37) outside the solvent accessible surface (probe radius = 1.2 Å). [b]Molecular mechanics energy calculated with the Dreiding FF.[30] [c]Total energy = -739.66680 hartree. [d]Calculated solvation energy = 70.3 kcal/mol. [e]Total energy = 18.06 kcal/mol.

It is found that conformation **2** is 16.9 kcal/mol higher in energy than the oxazoline ion, while the extended conformation is 21.7 kcal/mol higher in energy. These values are with solvation, and the gas phase values are 16.1 and 29.8 kcal/mol,

respectively. Thus, anchimeric stabilization is a significant force in making the oxazoline ion a thermodynamically accessible intermediate in solution. These results are consistent with the experimental observation that anchimeric assistance occurs during spontaneous hydrolysis of 2-acetamido substituted polysaccharides.[20,21]

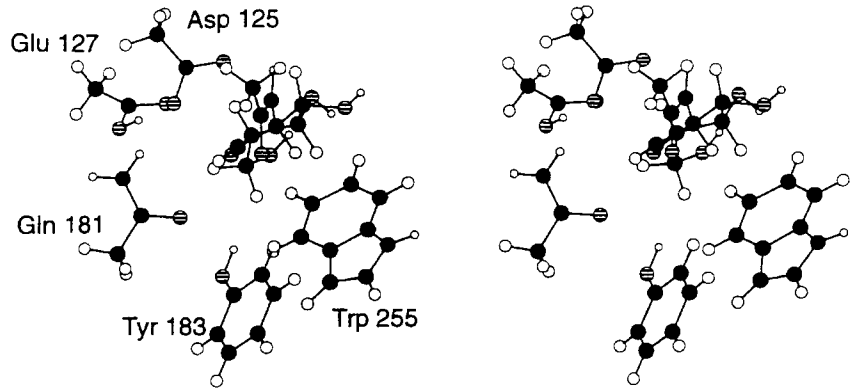
It is conceivable that the enzymatic hydrolysis mechanism of β(1,4)-linked C2′-*N*-acetyl-polysaccharides could still proceed through an oxocarbenium ion if there is sufficient electronic stabilization within the enzyme active site. Theoretical studies have shown that local electrostatic factors are important for lysozyme which has an asymmetric charge distribution.[35] Electrodynamics calculations[36] for the lysozyme active site reveal a large induced electric field which is proposed to stabilize bond cleavage and the resulting charge separation. Hevamine and other family 18 chitinases differ from lysozyme by having only one acidic residue capable of charge stabilization in the active site. Nevertheless, the possibility of electrostatic stabilization of an oxocarbenium ion when bound to a family 18 chitinase (in vacuum and solvated) was explored. As expected, no such stabilization was observed due to the lack of local charged groups capable of stabilizing the positive oxocarbenium ion.

The extent of enzymatic stabilization induced by hevamine was examined using a combination of molecular mechanics and *ab initio* QM calculations (HF/6-31G**). Using the Dreiding FF, it was first established that the gas phase molecular mechanics structures and relative energies for each intermediate (**1-3**) were consistent with those obtained using quantum mechanics. The molecular mechanics energies are listed in Table 1 and show good agreement (within 1 kcal/mol) with the QM results. The RMS difference in coordinates between structures determined with the two methods was only

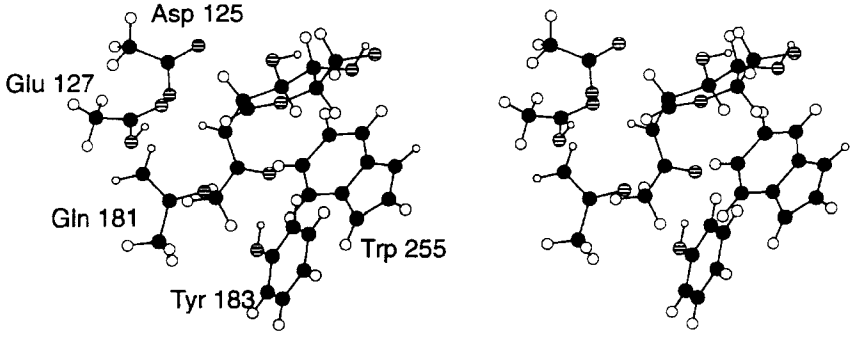
0.16 and 0.10 Å for conformations **1** and **3** respectively. The constrained conformation **2** has a larger RMS coordinate difference (0.31 Å) primarily due to the non-planar C2′ *N*-acetyl nitrogen which was not reproduced by molecular mechanics.

Each of the three possible (GlcNAc) disaccharide substituted intermediates was aligned with the allosamidin/hevamine complex crystal structure. These docked conformations served as starting points for a series of MD and energy minimization cycles. Initially, the entire enzyme was held fixed in the crystal structure geometry. For the final series of quenched dynamics cycles, the five residues (Asp 125, Glu 127, Gln 181, Tyr 183, and Trp 255) forming the active site binding pocket were free to move. The results of these MD simulations are consistent with simulations of the isolated intermediates. The lowest energy structure remains the oxazoline, with the constrained oxocarbenium ion lying 7.5 kcal/mol higher in energy. The extended *N*-acetyl geometry is much less favorable than the oxazoline ion with an energy difference of 31.8 kcal/mol. The optimum structures from these MD simulations were used for more detailed QM calculations. Each amino acid side chain which forms part of the active site binding pocket was then terminated with a hydrogen in place of the C-α carbon. This reduced system served as a model for which single point energy calculations, both in the gas phase and solvated in water, were carried out which included the effects of enzymatic stabilization.

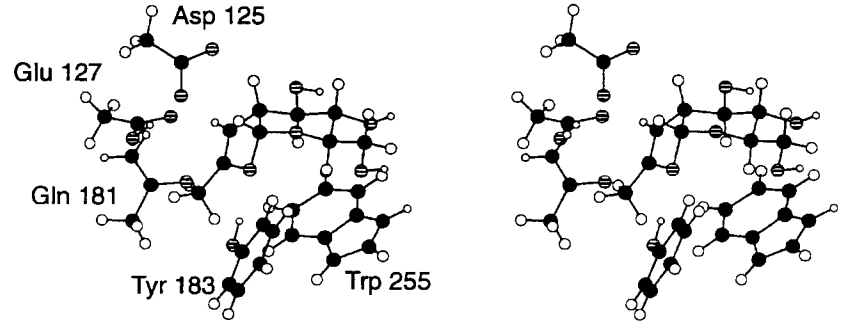
Molecular mechanics and *ab initio* QM (HF/6-31G**) single point energies were calculated for model systems **7-9** (Figure 2.6). Table 2 lists the energies for each cluster. Molecular mechanical methods gave the correct ordering of the three intermediates when compared to the QM results. Both methods clearly indicate that there is no enzymatic



7



8



9

Figure 2.6. Stereo view of the clusters used for quantum mechanical calculations including the hevamine active site residues. Dotted atoms are carbon, hatched are oxygen, solid are nitrogen and white are hydrogens. Here **7**, **8**, and **9** correspond to **1**, **2**, and **3** of Figure 2.2a.

stabilization of oxocarbenium ion **1** over the oxazoline ion intermediate. In all environments examined (gas phase, solvated, gas phase/bound to hevamine, and solvated/bound to hevamine) the oxocarbenium ion intermediate is 21.6, 20.0, 38.1, and 41.3 kcal/mol higher in energy than the oxazoline intermediate. Indeed, in the presence of the active site residues, the oxocarbenium ion with an extended *N*-acetyl geometry is destabilized relative to the other conformations.

Table 2.2. Relative Energies of Hydrolysis Intermediates Explicitly Including Hevamine Active Site Residues (kcal/mol).

Conformation	Dreiding FF[a]	HF/6-31G**	HF/6-31G**(solv)[a]
4	21.4	38.1	41.3
5	12.5	16.2	23.3
6	0.0[b]	0.0[c]	0.0[d]

[a]Molecular mechanics energy calculated with the Dreiding FF.[30] [b]Total energy = -130.4 kcal/mol. [c]Total energy = -2069.829454 hartree. [d]Total solution phase energy = -2069.91545.

In the calculations discussed above, we have not explored the formation of each of the different intermediates from the protonated chitin substrate. Rather, we have started with the assumption that glycosidic bond cleavage (the first step in any glycosidic hydrolysis mechanism) will result in a charged intermediate. The energetics of glycosidic bond cleavage and the barriers to formation of an oxocarbenium ion and oxazoline ion will be addressed in Chapter 3.

Structural and Electronic Properties of Intermediates

Analysis of the structural and electronic properties of the three putative intermediates reveals interesting differences important for the design of transition state analogs which may act as inhibitors. Bond lengths, angles, and torsions for the hexose ring (see Figure 2.5 for nomenclature) of conformations **1** and **3** are listed in Table 3. The complete set of Cartesian coordinates for optimized structures **1** and **3** are listed in Tables 4 and 5.

Conformations **1** and **2** have a C1′ center, forcing the hexose ring to adopt a more planar half-chair geometry. In contrast, the oxazoline ion (**3**) maintains a mildly distorted full chair conformation since all atoms of the hexose ring have sp^3 hybridization. This is evident by comparing the C5′-O5′-C1′-C2′ torsion angle (τ_0) of each conformation. The more planar oxocarbenium ion **1** has a torsion angle of -5.4° while the oxazoline ion is puckered to -32.5°. The O5′-C1′ bond length is 1.23 A for **1** and **2** , indicating a double bond order. This is consistent with the longer O5′-C1′ bond length of 1.34 Å for **3** and the bond length of 1.49 A for the newly formed bond between C1′ and the *N*-acetyl carbonyl oxygen.

The electrostatic potential surfaces for intermediates **1** and **3** are shown in Figure 2.7. For the oxocarbenium ion, **1**, a localized positive charge density is observed on C1′ and O5′, as expected. For the more stable oxazoline ion, **3**, the charge is transferred to the protonated oxazoline, formed upon cyclization of the *N*-acetyl group. This migration alters the electrostatic potential surface of the entire hexose ring, allowing greater delocalization of the positive charge (now distributed between the N, C and O atoms of the *N*-acetyl group).

Table 2.3. Energy Optimized Geometrical Parameters for the Hexose Ring of the Oxocarbenium Ion (**1**) and Oxazoline Ion (**3**).

Parameter[a]	(**1**) Gas Phase	(**3**) Gas Phase	(**3**) Solvated
R_0[b]	1.231	1.338	1.349
R_1	1.502	1.537	1.537
R_2	1.524	1.529	1.530
R_3	1.524	1.522	1.521
R_4	1.528	1.527	1.526
R_5	1.494	1.437	1.431
θ_0[c]	126.4	120.5	120.0
θ_1	123.5	117.8	118.0
θ_2	109.6	114.5	114.4
θ_3	108.9	110.8	110.0
θ_4	110.3	108.1	107.9
θ_5	110.8	107.9	108.2
τ_0[c]	-5.4	-35.2	-33.1
τ_1	24.9	25.5	25.2
τ_2	-52.0	-37.5	-39.1
τ_3	62.2	57.0	59.1
τ_4	-41.3	-63.5	-64.6
τ_5	13.4	54.3	52.7

[a]See Figure 2.5 for atom numbering and definitions of geometric parameters. [b]Bond lengths in Å. [c]Bond angles and torsions in degrees.

Table 2.4. Cartesian Coordinates of the Optimized Geometry for the Oxocarbenium Ion Intermediate **1** (HF/6-31G** gas phase).

Atom	X	Y	Z
C1′	0.00000	0.00000	0.00000
HC1′	-0.54822	-0.92756	-0.00127
C2′	1.50152	0.00000	0.00000
HC2′	1.76657	-0.07910	-1.05746
C3′	2.01247	1.32879	0.54451
HC3′	1.74836	1.41651	1.59878
O3′	3.38349	1.33112	0.37244
HO3′	3.75420	2.12235	0.73588
C4′	1.34168	2.46567	-0.21708
HC4′	1.58284	2.36863	-1.27303
O4′	1.82288	3.66147	0.30724
HO4′	2.01369	4.29725	-0.36357
C5′	-0.17074	2.42893	0.00000
HC5′	-0.41807	2.74304	1.00246
O5′	-0.67980	1.02463	-0.04653
C6′	-0.99802	3.20256	-1.00249
H1C6′	-0.89705	2.74855	-1.98659
H2C6′	-0.58332	4.20745	-1.04683
O6′	-2.30467	3.19297	-0.54509
HO6′	-2.88488	3.62599	-1.15107
N1	2.00580	-1.12385	0.74282
HN1	2.83716	-0.94787	1.26540
C7	1.81582	-2.38284	0.21224
O7	1.01600	-2.52652	-0.66998
C8	2.62144	-3.49359	0.82380
H1C8	2.29899	-4.43224	0.39961
H2C8	2.48965	-3.51031	1.90009
H3C8	3.67602	-3.34643	0.61356

Table 2.5. Cartesian Coordinates of the Optimized Geometry for the Oxazoline Ion Intermediate **3** (HF/6-31G** gas phase).

Atom	X	Y	Z
C1′	0.00000	0.00000	0.00000
HC1′	0.79497	-0.60072	-0.41318
C2′	0.00000	0.00000	1.53721
HC2′	0.99603	-0.15878	1.92462
C3′	-0.59855	1.25642	2.17052
HC3′	-1.68588	1.18692	2.14278
O3′	-0.14500	1.28806	3.47924
HO3′	-0.46809	2.06885	3.90589
C4′	-0.22876	2.49536	1.36839
HC4′	0.85086	2.63339	1.37625
O4′	-0.88567	3.55212	1.99932
HO4′	-0.43052	4.37061	1.88413
C5′	-0.71300	2.29514	-0.06555
HC5′	-1.77602	2.07737	-0.07984
O5′	0.00000	1.17850	-0.62149
C6′	-0.44743	3.46237	-0.99935
H1C6′	-0.94111	4.33951	-0.60217
H2C6′	0.62283	3.65742	-1.03818
O6′	-0.98026	3.23073	-2.25943
HO6′	-0.44492	2.61440	-2.73226
N1	-0.84179	-1.17460	1.78628
HN1	-0.98630	-1.57610	2.68914
C7	-1.54010	-1.47608	0.73727
O7	-1.23072	-0.78346	-0.29153
C8	-2.60780	-2.50136	0.64787
H1C8	-2.76120	-3.00002	1.59391
H2C8	-3.52224	-2.01322	0.33047
H3C8	-2.33063	-3.22003	-0.11437

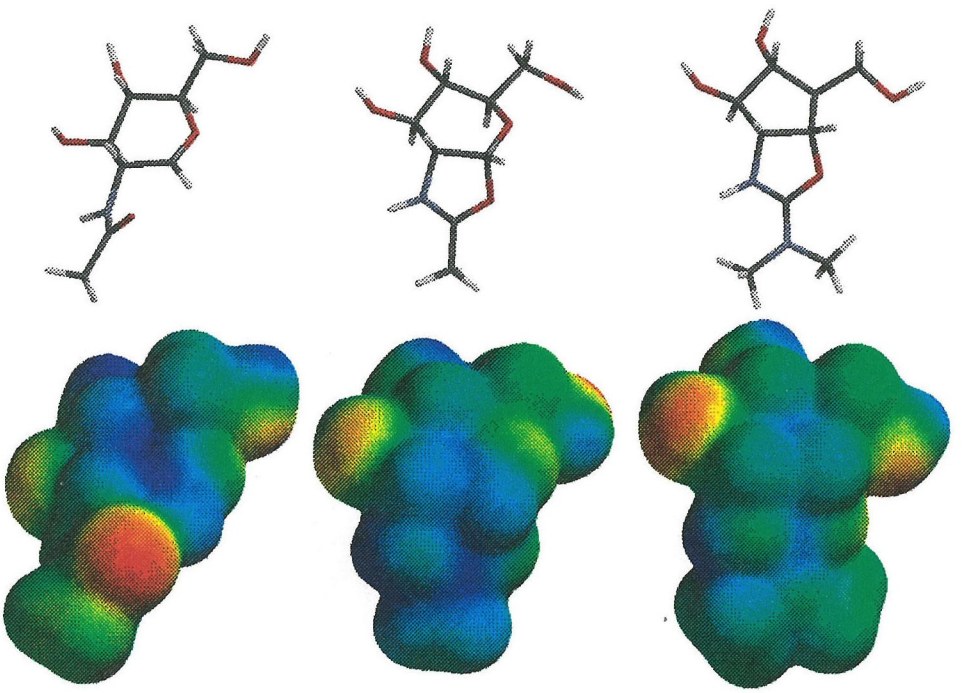


Figure 2.7. Electrostatic potential surfaces of the oxocarbenium (**1**) and oxazoline (**3**) intermediates compared to inhibitors allosamizoline (**11**). Red indicates a negative potential and blue a positive potential.

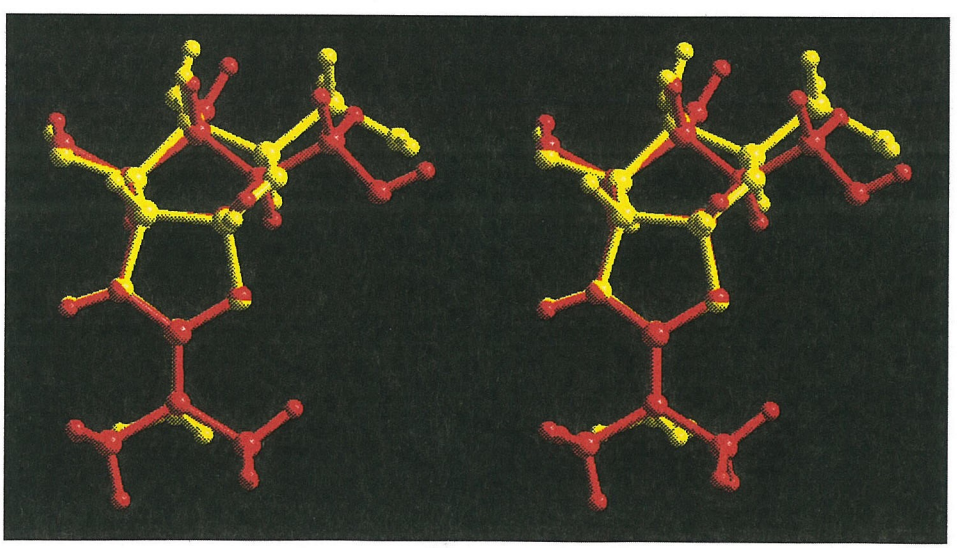


Figure 2.8. A stereo view of the optimized oxazoline ion intermediate **3** (yellow) overlaid with the optimum allosamizoline structure (red).

Transition State Inhibitors

The mechanism of HEWL involves an oxocarbenium ion intermediate, stabilized by a second carboxylate in the enzyme active site.[11-15] Lactones have been reported to act as lysozyme inhibitors by mimicking the electronic and structural characteristics of the oxocarbenium ion intermediate.[37] Unique properties of the oxocarbenium ion (**3**) include a planar region of the hexose ring and a localized positive charge near C1′ and O5′. By analogy, a compound with structural and electronic properties mimicking the oxazoline ion intermediate would be expected to act as an inhibitor of the family 18 chitinases (assuming the oxazoline ion intermediate to be involved in the mechanism).

Allosamidin, **10** (Figure 2.2b), and related natural products are pseudotrisaccharide chitinase inhibitors[6] isolated from *Streptomyces* sp. mycelium. Although detailed inhibition and kinetics studies have not been reported, allosamidin is believed to act as a transition state analog.[10,24,38] Moreover, allosamidin shows no inhibitory activity against lysozyme,[7] which is consistent with the suggestion that the two enzymes have different mechanisms (and therefore unique transition states) for hydrolysis. It was originally proposed that the planar protonated oxazoline ring acts as an oxocarbenium ion transition state analog.[38] However, an X-ray structure of hevamine with allosamidin bound in the active site[10] led to the suggestion that allosamizoline (**11**), the aglycone portion of allosamidin, instead resembles an oxazoline ion reaction intermediate.

QM calculations were carried out, at the same level as for the chitinase reaction intermediates **1-3** and without constraints, to optimize the geometry of allosamizoline

(**11**) (Figure 2.2b). This aglycone allosamidin derivative has a hydroxyl substitution for the *N*-acetyl-glucosamine disaccharide of allosamidin (**10**), thus allowing a comparison with intermediates **1-3**. The optimized geometry and electrostatic potential surface is shown in Figure 2.7. An overlay of the oxazoline ion intermediate and allosamizoline (Figure 2.8) shows the striking match between the two structures, particularly in the charged region of the acetamido group. Indeed, the RMS coordinate difference between equivalent non-hydrogen atoms is 0.21 Å! To carry out this RMS match calculation, the following atoms from oxazoline ion were not included: O5′, C5′, C6′ and O6′; and from allosamizoline: atoms equivalent to C5′, C6′ and the dimethyl portion of the amine.

The electrostatic potential surfaces of allosamizoline and putative chitinase reaction intermediates **1** and **3** (Figure 2.7) indicate clearly that the electronic properties of allosamizoline (**11**) are similar to those of the oxazoline intermediate (**3**) and dissimilar to the oxocarbenium ion (**1**). Both **3** and **11** have a localized positive charge density on the oxazoline. In contrast, oxocarbenium ion **1** has a localized positive charge between O5′ and C1′. Enzymatic stabilization of these two species (**1** and **2**) would require different active site structures and electrostatic properties. For **1**, the active site would need acidic residues adjacent to the oxazoline, whereas for **3** the acidic residues would need to be on the opposite side of the binding cleft to stabilize the oxocarbenium ion.

Indeed, an analysis of the electrostatic potential surface of the enzyme active site model (Figure 2.9) reveals a large negatively charged region opposite the oxazoline moiety. In addition, visualization of the complementary electrostatic potential surface of the oxazoline intermediate shows great similarity between the enzymatic surface and the oxazoline surface. It is evident that the charge distribution within the enzyme active site

cannot stabilize an oxocarbenium ion intermediate and this likely results in the observed destabilization of this species.

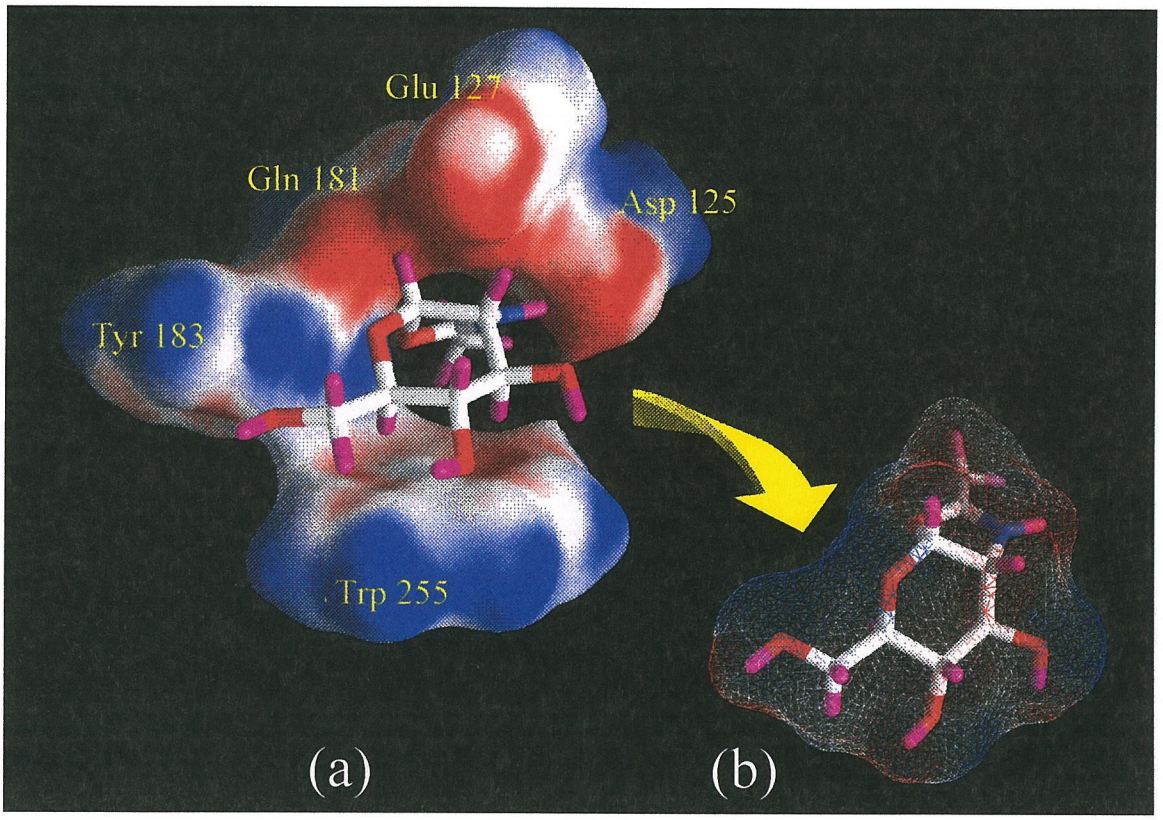


Figure 2.9. (a) Electrostatic potential surface of the hevamine active site model cluster (**9**). (b) Complementary electrostatic surface (shown as a grid) of the oxazoline intermediate. Red denotes a negative potential and blue a positive potential.

V. Conclusions

Using *ab initio* quantum mechanical calculations, the reaction intermediates for the enzymatic hydrolysis of chitin were examined. It is found that the oxazoline ion intermediate is stabilized by the C2′ acetamido group, both in an isolated system and bound to hevamine. Similar calculations on a known inhibitor belonging to the allosamidin family of natural products revealed structural and electronic similarities

between this inhibitor and the oxazoline ion intermediate. These calculations lend strong theoretical support to the increasing body of experimental evidence which suggests that the mechanism of family 18 chitinases involves anchimeric assistance stabilizing an oxazoline ion reaction intermediate.

The double displacement mechanism involving an oxazoline intermediate may not be unique to the chitinases. Other enzymes have been identified which hydrolyze β(1,4)-linked C2′-*N*-acetyl-polysaccharides but are also missing a second acidic residue. These include GEWL[25,26] and the soluble lytic transglycosylase.[27,28] These enzymes may utilize the same oxazoline ion intermediate assisted by anchimeric stabilization.

The oxazoline transition state serves as a target for the rational design of more potent glycosidase inhibitors specific to family 18 chitinases. Simple analogs of allosamidin which incorporate the key features of a delocalized positive charge while maintaining a chair-like sugar conformation may prove to be synthetically more accessible than allosamidin. Such analogs could lead to a new generation of chitinase transition state inhibitors.

Acknowledgments

This research was funded by DOE-BCTR (DE-FG36-93CH105 81, David Boron). The facilities of the MSC are also supported by grants from NSF (CHE 95-22179 and ASC 92-100368), Chevron Petroleum Technology Co., Aramco, Asahi Chemical, Owens-Corning, Chevron Chemical Company, Asahi Glass, Chevron Research and Technology Co., Hercules, BP Chemical, and Beckman Institute. Some calculations were carried out at the NSF San Diego Supercomputer Center (SDSC) and on the JPL Cray. K.A.B. acknowledges an NIH fellowship.

VI. References

(1) Fukamizo, T.; Kramer, K. J. *Insect Biochem.* **1985**, *15*, 141-145.

(2) Bartnicki-Gracia, S. *Annu. Rev. Microbiol.* **1968**, *22*, 87-108.

(3) Boller, T.; Gehri, A.; Mauch, F.; Vogeli, U. *Planta* **1983**, *157*, 22-31.

(4) Collinge, D. B.; Kragh, K. M.; Mikkelsen, J. D.; Nielsen, K. K.; Rasmussen, U.; Vad, K. *Plant J.* **1993**, *3*, 31-40.

(5) Henrissat, B.; Bairoch, A. *Biochem. J.* **1993**, *293*, 781-788.

(6) Sakuda, S.; Isogai, A.; Matsumoto, S.; Suzuki, A. *Tetrahedron Lett.* **1986**, *27*, 2475-2478.

(7) Koga, D.; Isogai, A.; Sakuda, S.; Matsumoto, S.; Suzuki, A.; Kimura, S.; Ide, A. *Agri. Biol. Chem.* **1987**, *51*, 471-476.

(8) Schlumbaum, A.; Mauch, F.; Vogeli, U.; Boller, T. *Nature* **1986**, *324*, 365-367.

(9) Kinoshita, M.; Sakuda, S.; Yamada, Y. *Biosci. Biotech. Biochem.* **1993**, *57*, 1699-1703.

(10) Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W. *Biochemistry* **1995**, *34*, 15619-15623.

(11) Phillips, D. C. *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 484-495.

(12) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171-1202.

(13) Koshland, D. E. *Biol. Rev.* **1953**, *28*, 416-436.

(14) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Bio.* **1994**, *4*, 885-892.

(15) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-859.

(16) Perrakis, A.; Tews, I.; Dauter, Z.; Oppenheim, A. B.; Chet, I.; Wilson, K. S.; Vorgias, C. E. *Structure* **1994**, *2*, 1169-1180.

(17) Terwisscha van Scheltinga, A. C.; Kalk, K. H.; Beintema, J. J.; Dijkstra, B. W. *Structure* **1994**, *2*, 1181-1189.

(18) Terwisscha van Scheltinga, A. C.; Hennig, M.; Dijkstra, B. W. *J. Mol. Biol.* **1996**, *262*, 243-257.

(19) Armand, S.; Tomita, H.; Heyraud, A.; Gey, C.; Watanabe, T.; Henrissat, B. *Febs Letters* **1994**, *343*, 177-180.

(20) Piszkiewiez, D.; Bruice, T. C. *J. Am. Chem. Soc.* **1967**, *89*, 6237-6243.

(21) Piszkiewiez, D.; Bruice, T. C. *J. Am. Chem. Soc.* **1968**, *90*, 2156-2163.

(22) Lowe, G.; Sheppard, G.; Sinnott, M. L.; Williams, A. *Biochem. J.* **1967**, *104*, 893-899.

(23) Parsons, S. M.; Raferty, M. A. *Biochemistry* **1969**, *8*, 4199-4205.

(24) Van Vranken, D. L. *Synthesis of the aminohydroxy-glycosidase inhibitors and enantioselective palladium catalysis.*; Van Vranken, D. L., Ed.; Stanford University, 1992, pp 8-9.

(25) Weaver, L. H.; Grutter, M. G.; Matthews, B. W. *J. Mol. Biol.* **1995**, *245*, 54-68.

(26) Grutter, M. G.; Weaver, L. H.; Mathews, B. W. *Nature* **1983**, *303*, 828-831.

(27) Thunnissen, A. M. W. H.; Dijkstra, A. J.; Kalk, K. H.; Rozeboom, H. J.; Engel, H.; Keck, W.; Dijkstra, B. W. *Nature* **1994**, *367*, 750-754.

(28) Thunnissen, A. M. W. H.; Rozeboom, H. J.; Kalk, K. H.; Dijkstra, B. W. *Biochemistry* **1995**, *34*, 12729-12737.

(29) Ringnalda, M. N.; Langlois, J.-M.; Greeley, B. H.; Murphy, R. B.; Russo, T. V.; Cortis, C.; Muller, R. P.; Marten, B.; Donnelly, R. E.; Mainz, D. T.; Wright, J. R.; Pollard, W. T.; Cao, Y.; Won, Y.; Miller, G. H.; Goddard, W. A., III; Friesner, R. A. Jaguar 3.0 from Schrödinger, Inc., located in Portland, OR.

(30) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(31) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.

(32) Muller, R. P.; Langlois, J. M.; Ringnalda, M. N.; Friesner, R. A.; Goddard, W. A., III *J. Chem. Phys.* **1994**, *100*, 1226-1235.

(33) Murphy, R. B.; Friesner, R. A.; Ringnalda, M. N.; Goddard, W. A., III *J. Chem. Phys.* **1994**, *101*, 2986-2994.

(34) Langlois, J. M.; Yamasaki, T.; Muller, R. P.; Goddard, W. A., III *J. Phys. Chem.* **1994**, *98*, 13498-13505.

(35) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.

(36) Dao-Pin, S.; Liao, D. I.; Remington, S. J. *Proc. Natl. Acad. Sci., USA* **1989**, *86*, 5361-5365.

(37) Ford, L. O.; Johnson, L. N.; Machin, P. A.; Phillips, D. C.; Tjian, R. *J. Mol. Biol.* **1974**, *88*, 349-371.

(38) Robertus, J. D.; Hart, P. J.; Monzingo, A. F.; Marcotte, E.; Hollis, T. *Can. J. Bot.* **1995**, *73*, S1142-S1146.

Chapter 3

Substrate Distortion to a Boat Conformation at Subsite -1 is Critical in the Mechanism of Family 18 Chitinases

Abstract

Using molecular dynamics simulations, the plausible conformations for a hexaNAG substrate bound to the active site of chitinase A were examined. It is found that (i) the hydrolysis mechanism of chitinase A (a family 18 chitinase from *Serratia marcescens*) involves substrate distortion; (ii) the first step of acid catalyzed hydrolysis (protonation of the linking anomeric oxygen between GlcNAc residues -1 and +1) requires a boat conformation for the GlcNAc residue at binding subsite -1; (iii) *ab initio* quantum mechanical (QM) calculations (HF/6-31G**) predict that protonation of a GlcNAc in a boat conformation leads to spontaneous anomeric bond cleavage to yield an oxazoline ion intermediate.

Several conformations of two possible hydrolysis intermediates were also studied: the oxocarbenium ion and the oxazoline ion. Only the oxazoline ion orients in the enzyme active site so as to allow stereo-selective attack by water. This leads to retention of configuration in the anomeric product as observed experimentally.

It is possible that all family 18 chitinases share a common mechanism. Hence, it is suspected that distortion of the substrate into a boat form at subsite -1 is required for any glycosyl hydrolase which has only one acidic residue in the active site. The design of an inhibitor for these systems based on the boat distorted sugar conformation is discussed.

I. Introduction

Chitin, a β(1,4)-linked *N*-acetyl-glucosamine (GlcNAc) polysaccharide, is a major structural component of fungal cell walls and the exo-skeletons of invertebrates, including insects and crustaceans. This linear polymer may be degraded through the enzymatic hydrolysis action of chitinases. Chitinases have been found in a wide range of organisms including bacteria,[1,2] plants,[3] fungi,[4] insects,[5] and crustaceans.[6] For those organisms that utilize the structural properties of chitin, chitinases are critical for the normal life cycle functions of molting and cell division.[7,8] In addition, plants produce chitinases as a defense against fungal pathogens.[3,9] Because chitin is not found in vertebrates, it has been suggested that inhibition of chitinases may be used for the treatment of fungal infections and human parasitosis.[10]

Based on amino acid sequence, the glycosyl hydrolases have been classified into 45 families.[11] Using this classification method, the chitinases form families 18 and 19 which are unrelated, differing in structure and mechanism. Sequence analysis shows little homology between these classes of chitinases. Family 19 chitinases (found in plants) share the bilobal folding motif of lysozyme, which forms a well defined substrate binding cleft between the lobes.[12,13] In contrast, family 18 chitinases share two short sequence motifs which form the catalytic (βα)-barrel active site.[14] Family 18 chitinases with diverse sequences have been isolated from a wide range of eukaryotes and prokaryotes. There are two general mechanistic pathways for acid catalyzed glycosyl hydrolysis which result in:

i. retention of the stereochemistry of the anomeric oxygen at C1′ relative to the initial configuration[15] (see Figure 3.1a) or

ii. inversion of the stereochemistry (see Figure 3.1b).

An example of the retaining mechanism is hen egg white lysozyme (HEWL), which has been shown to require two acidic residues, one of which is protonated.[16] This mechanism is believed to proceed as follows.[17,18] The β-(1,4)-glycosidic oxygen is first protonated (leading to an oxocarbenium ion intermediate) which is stabilized by a second carboxylate (either through covalent or electrostatic interactions). Nucleophilic attack by water yields the hydrolysis product, which necessarily retains the initial anomeric configuration. This is commonly referred to as the double displacement mechanism of hydrolysis (Figure 3.1a).[15-19]

Although the X-ray crystal structure of a family 19 chitinase reveals a lysozyme-like fold (suggesting a double displacement mechanism[12,13]), the hydrolysis products for two family 19 chitinases show inversion of the anomeric configuration.[20,21] This leads to the second commonly discussed hydrolysis mechanism; a concerted single displacement reaction in which a bound water molecule acts as the nucleophile (Figure 3.1b). The crystal structure suggests that the second catalytic carboxylate may be sufficiently close to allow coordination of a water molecule consistent with a single displacement mechanism.

Family 18 chitinases have been reported to yield hydrolysis products which retain the anomeric configuration at C1′.[22] However, the X-ray crystal structure of two family 18 chitinases reveals no second acidic residue in the active site capable of stabilizing the

oxocarbenium ion.[2,23] Thus, neither the single nor double-displacement mechanism is consistent with the observed structure and hydrolysis products. An increasing body of experimental[24] and theoretical evidence points to an oxazoline ion intermediate formed through anchimeric assistance by the neighboring *N*-acetyl group (Figure 3.1c) as being the likely mechanism for family 18 chitinases. Such an intermediate alleviates the need for a second acidic residue. In solution, spontaneous acid catalyzed hydrolysis of 2-acetamido substituted polysaccharides have been reported to occur through anchimeric assistance.[25,26]

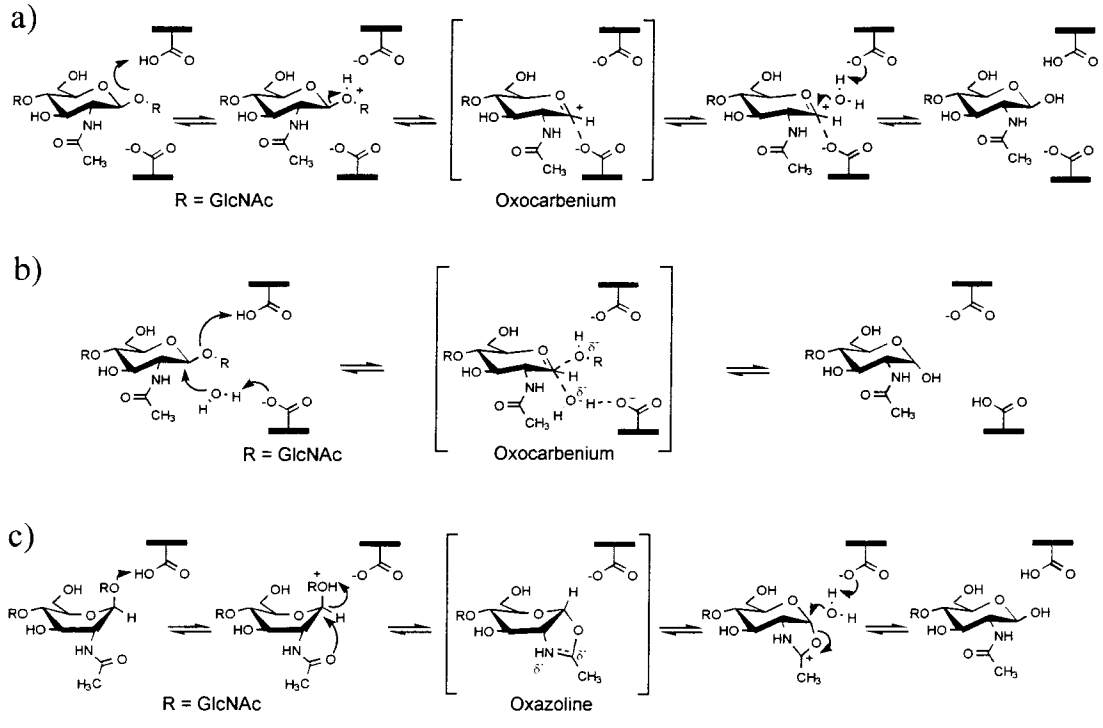


Figure 3.1. (a) The double-displacement hydrolysis mechanism which requires two acidic residues in the active site and leads to retention of the anomeric configuration. (b) The single-displacement mechanism which requires only one acidic residue in the active site and results in inversion of the anomeric configuration. (c) The anchimeric stabilization hydrolysis mechanism of family 18 chitinases. The substrate is distorted to a boat conformation and the oxazoline ion intermediate is stabilized through anchimeric assistance from the neighboring C2′ *N*-acetyl group.

The oxazoline double-displacement mechanism may not be unique to family 18 chitinases. The X-ray structure of a bacterial chitobiase[27] (family 20) complexed with the unhydrolyzed natural substrate reveals that the glycosidic bond to be cleaved is in a "quasi-axial" orientation. In addition, the C2′ *N*-acetyl group is found to be in a position to allow the formation of an oxazoline ion intermediate. A similar substrate distortion was observed for endoglucanase I,[28] a cellulase from family 7, complexed with a nonhydrolyzable thiooligosaccharide substrate analog. As in the chitobiase structure, the substrate occupies subsites spanning the points of enzymatic cleavage and reveals a distortion in the sugar conformation at the cleavage site. Once again, a "quasi-axial" orientation for the glycosidic bond was observed and the nucleophile of endoglucanase I was in a similar position to the C2′ *N*-acetyl group of the chitobiase structure.

Reported herein are molecular dynamics (MD) simulations of a family 18 bacterial chitinase (chitinase A from *Serratia marcescens*). Initial substrate binding and the possible resulting hydrolysis intermediates are investigated. It is found that the hexaNAG substrate is forced to distort to a boat sugar geometry at subsite -1, prior to protonation, which then leads to spontaneous anomeric bond cleavage and subsequent formation of an oxazoline ion.

II. Methods

Simulation Methods

All MD simulations were carried out using the MSC-PolyGraf[29] program with the Dreiding force field (FF).[30] Averaged charge equilibrium (QEq) charges were used for all GlcNAc residues. Charges for the oxocarbenium and oxazoline ions are from *ab initio* quantum mechanics (QM) calculations (HF/6-31G**). Figure 3.3 illustrates the FF atom

types and charges used in the simulation. Charges for the amino acid side chains are from CHARMM. For protonated Glu 315, the net charge was neutralized, with the following assignment: CA 0.05, N -0.40, H 0.25, C 0.60, O -0.55, CB -0.10, HB 0.05, CG -0.25, HG 0.05, CD 0.40, OE1 -0.30, OE2 -0.30, HOE 0.35.

A standard Coulomb potential was used without a distance dependent dielectric constant and all nonbond interactions were considered explicitly. A nonbond cutoff of 9.5 Å was used during MD simulations and extended to 13.5 Å for single point energy calculations. Solvation energies were estimated using the continuum solvent model in the Delphi program.[31]

The *ab initio* QM calculations (HF/6-31G**) were carried out with the PS-GVB program[32-35] from Schrödinger, Inc. During geometry optimization of the two protonated methyl-GlcNAc conformations, the O4′-C1′-O1′-C(methyl) torsion was constrained to be 120°. This constraint was necessary to prevent free rotation about the anomeric bond and proton transfer from O1′ to the *N*-acetyl group carbonyl oxygen. An analogous conformational constraint occurs for the hexaNAG substrate upon binding to Chitinase A.

Starting Structures

Crystal structures have been solved for hevamine and Chitinase A,[2] both family 18 chitinases. There is little structural information known regarding substrate binding to Chitinase A. However, several structures for hevamine have been reported which include complexes with triNAG[23] and allosamidin.[24] We used the detailed knowledge of the hevamine system to aid in the initial docking of the ligands to Chitinase A. A sequence alignment of several members of the family 18 glycosyl hydrolases[14] simplified the task

of structural alignment between Chitinase A and hevamine. The following residues were included in a least squares coordinate match: Tyr 6, Asp 123, Asp 125, Glu 127, Tyr 183, and Trp 255 of the hevamine active site and Tyr 163, Asp 311, Asp 313, Glu 315, Tyr 390, and Trp 539 of Chitinase A.

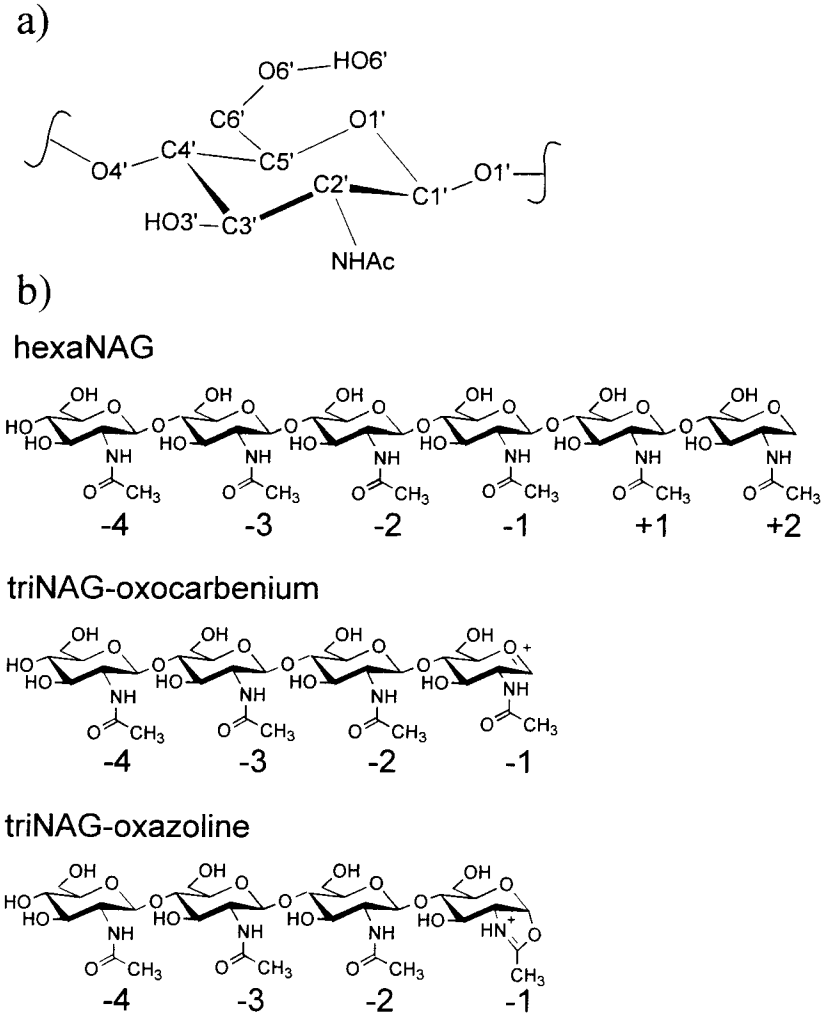


Figure 3.2. Structures and sugar labeling scheme used for the hexaNAG substrate and the triNAG-oxocarbenium ion and triNAG-oxazoline ion intermediates. Sugars are numbered -4 through +2 from the non-reducing end as is the current convention.

The triNAG/hevamine complex structure served as a starting point for the structure of hexaNAG residues -4, -3, and -2.[36] The remaining three GlcNAc residues (-

1, +1, and +2) were built individually and optimized through simulated annealing. The allosamidin/hevamine complex was used as a template for the *N*-acetyl geometry at subsite -1. In total, six different starting conformations for GlcNAc residues -1 through +2 were examined. Each conformation was subjected to 10 annealing cycles during which the temperature was raised from 0 k to 350° K and back to 0° K in increments of 50° K every 100 fs for a total of 2.1 ps. This resulted in only two distinct conformations, which differed only at subsite -1, denoted as -1–chair and -1–boat.

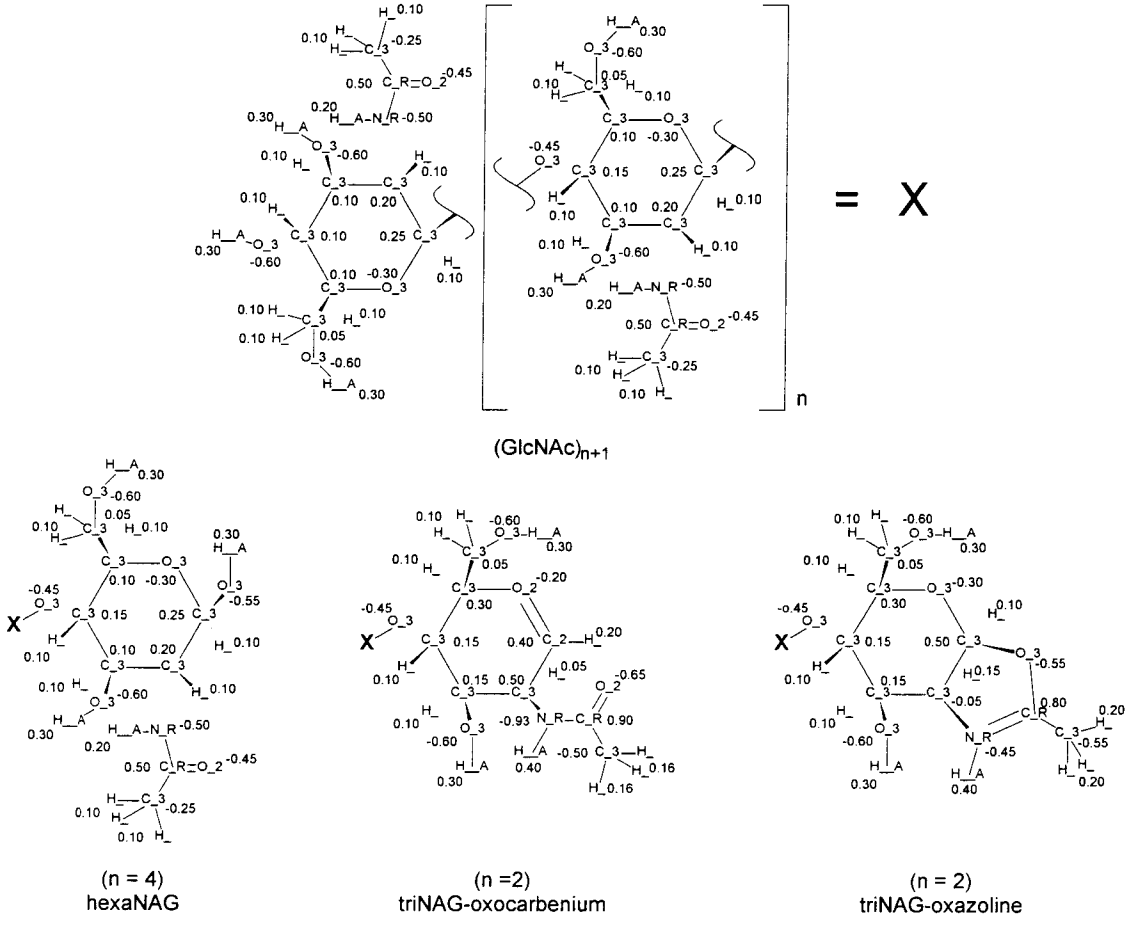


Figure 3.3. Atomic charges and force field atom types used for the simulation. X denotes a GlcNAc residue in the middle of the chain. The chain may be terminated in one of three ways: hexaNAG, a GlcNAc residue; triNAG-oxocarbenium, an oxocarbenium ion; triNAG-oxazoline, an oxazoline ion.

The structures of the intermediates were based on the equilibrated hexaNAG models. GlcNAc residues +1 and +2 were removed and the correct changes in atom hybridization were applied to generate an oxocarbenium ion or oxazoline ion.

Counter ions and crystallographic water molecules were included during all simulations to ensure a net neutral charge. Water molecules in van der Waals (vdW) contact with the docked ligand were moved to avoid high energy starting conformations. Residues 24-150 were deleted from the simulation. These residues form the N-terminal domain which has unknown function. Based on a 6 Å distance cutoff from the bound hexaNAG substrate, the following residues were movable during all simulations: 163-172, 189-191, 204-213, 228-231, 273-277, 311-316, 361-364, 386-391, 442-446, 470-478, 539-543. All other residues were fixed.

III. Results

Simulations of hexaNAG Substrate Binding

The binding of a hexaNAG substrate (Figure 3.2) to chitinase A was studied using MD simulations. Starting from the reported X-ray crystal structures of both the triNAG/hevamine[23] and the alloamidin/hevamine complex,[24] a tetraNAG substrate was docked to chitinase A and two additional NAG residues were built and optimized. Six different binding conformations for hexaNAG were constructed, each differing in the geometry of the sugars bound at subsites -1 through +2 as these sites could not be extrapolated from the hevamine crystal structures. Simulated annealing cycles, carried out for each conformation and the surrounding binding site side chains, resulted in only two globally different binding modes. These differed primarily at the sugars bound in

subsites -2 and -1 and have been labeled -1–boat and -1–chair in reference to the pyranose ring conformation of sugar residue -1.

100 ps of MD was carried out on both optimized hexaNAG conformations. The binding site residues (Figure 3.4), crystallographic waters, counter ions, and hexaNAG substrate were all free to move during the simulation. Both the -1–chair and -1–boat conformations were stable and little deviation from the crystal structure was observed for the binding site residues (Table 3.1). For the -1–boat conformation, analysis of an ensemble of snapshots taken at regular time intervals reveals considerable mobility at the hexaNAG termini (sugars -4 and +2) while the central residues are tightly bound (Table 3.2). The -1–chair conformation shows greater flexibility or instability compared to the -1–boat conformation, greatest at sugar -4 and decreasing to sugar +2.

The -1–boat binding geometry is stabilized through a series of specific hydrogen bonds between each GlcNAc residue and the binding site of the enzyme (Figure 3.4). O6′ of sugar -4 interacts with a counter ion bound to Asp 230. The *N*-acetyl carbonyl of sugar -3 forms a hydrogen bond to the side chain of Thr 276 while the O6′ hydrogen bonds to Asn 474. Hydrophobic contacts are also made with Trp 167. The *N*-acetyl amide of sugar -2 donates a hydrogen bond to Glu 473 while the carbonyl accepts a hydrogen bond from Arg 172. Non-specific contacts are made with the hydrophobic "floor" of the binding cleft. Sugar -1 is tightly bound through a hydrogen bond from Tyr 390 to the *N*-acetyl carbonyl and from the *N*-acetyl amide to Glu 315 (one oxygen is protonated). There are also critical hydrophobic interactions with Tyr 163 and Trp 539 which force the -1 sugar into a boat conformation . Sugar +1 is less tightly bound and forms a hydrogen bond between Arg 446 and the *N*-acetyl carbonyl. Similarly, sugar +2 also makes few

specific contacts except for a hydrogen bond between the *N*-acetyl carbonyl and main chain amide hydrogen of Asp 391.

Table 3.1. RMS Coordinate Difference (Å) for Binding Site Residues.

substrate	av RMS[a]
hexaNAG(-1–chair)	1.59
hexaNAG(-1–boat)	1.43
triNAG-oxocarbenium (-1–chair)	1.54
triNAG-oxocarbenium (-1–boat)	1.50
triNAG-oxazoline	1.49
none	2.13

[a]Coordinate RMS difference was calculated as the difference between the crystal structure coordinates[2] and the average position for all non-hydrogen atoms during the MD interval from 30 to 100 ps.

Table 3.2. RMS Coordinate Fluctuation (Å) from Dynamical Average For HexaNAG Substrate Binding[a].

	binding subsites					
conformation	-4	-3	-2	-1	+1	+2
-1–chair	1.188	0.565	0.568	0.653	0.500	0.346
-1–boat	0.805	0.443	0.340	0.335	0.339	0.512

[a]RMS coordinate fluctuations were calculated as follows: The average structure for the MD interval from 30 to 100 ps was determined. The RMS coordinate differences between this average structure and snapshots taken every 5 ps from 30 to 100 ps were determined. The average of this RMS is reported, separated by sugar residue.

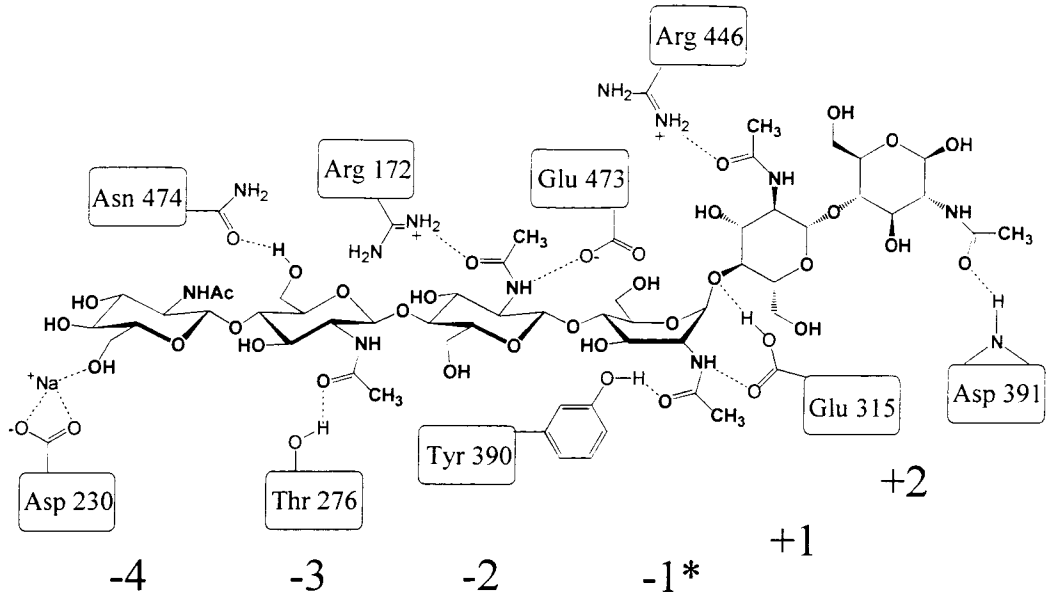


Figure 3.4. A schematic of the hydrogen bonds observed for the -1–boat hexaNAG binding mode. Note: In this schematic representation Tyr 390 appears to the left of the *N*-acetyl carbonyl oxygen of residue -1. However, the true structure has this carbonyl group rotated below the plane of the sugar (see Figure 3.6) with Tyr 390 positioned on the opposite side of the binding cleft. Thus the hydrogen bond to Tyr 390 helps to position the *N*-acetyl group prior to formation of an oxazoline ion and will not slow catalysis.

The -1–chair conformation makes fewer stabilizing contacts with the chitinase A binding site, particularly in the region of sugars -2 and -1 adjacent to the cleavage site. The first two sugars, -4 and -3, are in a similar position to the -1–boat geometry and form the same hydrogen bonds. Sugar -2 is no longer oriented perpendicular to the binding cleft, as observed for the hevamine triNAG complex, but rather has rotated to become parallel to the cleft, as is seen for HEWL. This places the *N*-acetyl group toward the binding cleft interior and allows for the formation of only one hydrogen bond from the *N*-acetyl amide to the Trp 275 backbone carbonyl. Facing towards the solvent, HO6′ forms a hydrogen bond with Glu 206. Similarly, sugar -1 only makes two hydrogen bonds, one between O3′ and Arg 446 and the other between O6′ and the proton of Glu 315. This

greatly limits the possibility of proton transfer to the β(1,4)-glycosidic oxygen which is the first step in the hydrolysis mechanism. The implications of this result are discussed in greater detail shortly. The final two sugars, +1 and +2, share features with the -1–boat conformation and form the same transient hydrogen bonds described above.

Table 3.3. HexaNAG Binding Energies (kcal/mol).

components	-1–chair	-1–boat	$E_{boat}-E_{chair}$
MM energy	-39.7	-15.3	-24.2
solvation energy[a]	139.1	108.4	30.7
total energy[b]	99.4	93.1	6.3

[a]Calculated using a continuum solvation method with the Delphi program.[31] [b]Total energy is the sum of the MM energy and solvation energy.

Relative binding energies for the two hexaNAG conformations is difficult to assess. Many factors influence ligand binding, including solvation energies, entropic consequences of water displacement, and electrostatic interactions. These difficulties are made even more complex as a result of the large number of possible hydrogen bond conformations available to carbohydrates which result in many local minima of similar energy. However, an attempt has been made to determine the energetic differences between the two binding modes (see Table 3.3). These calculations suggest that the -1–boat conformation has an internal energy 24.4 kcal/mol lower than the -1–chair conformation. To estimate the differential solvation energy a continuum solvation approximation[31] is used for each conformation in water (both bound to the enzyme and free in solution). It is calculated that the -1–chair conformation has a solvation energy 30.7 kcal/mol more favorable than the -1–boat conformation. Combining these two

contributions, the -1–chair geometry is preferred by 6.3 kcal/mol. Probably the uncertainty in these estimated energies is comparable to this difference, indicating that both conformations are thermodynamically accessible.

Simulations of Bound Intermediates

Following binding of the chitin substrate, any acid catalyzed hydrolysis mechanism requires a proton transfer from Glu 315 to the β(1,4)-glycosidic oxygen linking sugars -1 and +1. Upon proton transfer and the departure of chitobiose (formed from the +1 and +2 sugars), there remains a positively charged intermediate. Starting from the two hexaNAG conformations discussed above, three intermediates are possible: two oxocarbenium ion intermediates which differ in the geometry of the *N*-acetyl group, and an oxazoline ion. The results of a study of the aglycones of these intermediates using *ab initio* QM methods is discussed in Chapter 2. It was found that the oxazoline ion intermediate was substantially (18 kcal/mol) lower in energy both when isolated and while bound to the active site of hevamine, also a family 18 chitinase. This work is now extended to include the full enzyme and the triNAG substituted intermediates in MD simulations of Chitinase A.

Binding of the -1–chair Derived Oxocarbenium Ion

Starting from the -1–chair hexaNAG structure, sugars +1 and +2 were removed and sugar -1 was converted to an oxocarbenium ion. Following a series of annealing cycles, 100 ps of MD was carried out.

The positively charged C1′-O5′ is stabilized by interactions with Glu 315 and Asp 391. However, in order to attain these favorable interactions, the specific hydrogen bonds observed for the -4 through -2 sugars are disrupted. Indeed, only a few hydrogen bonds

remain intact and considerable mobility of sugars -4 through -2 is observed during the simulation (Table 3.4). HO6′ and HO4′ of sugar -4 form transient hydrogen bonds to Glu 208. Arg 172 now forms a hydrogen bond with O6′ of sugar -3 instead of the *N*-acetyl carbonyl group of sugar -2. Arg 446 replaces Arg 172 and forms a hydrogen bond to the *N*-acetyl carbonyl group of sugar -2.

Table 3.4. RMS Coordinate Fluctuation (Å) from Dynamical Average For TriNAG Intermediates.[a]

	binding subsites			
conformation	-4	-3	-2	-1
triNAG-oxocarbenium-1–chair	0.528	0.478	0.704	0.524
triNAG-oxocarbenium-1–boat	0.414	0.339	0.573	0.257
triNAG-oxazoline	0.445	0.407	1.021	0.267

[a]See comments for Table 3.2

No water molecules were observed near the oxocarbenium ion during the simulation. However, it is readily observable that the positioning of the oxocarbenium ion in the active site cleft, relative to Glu 315 and Asp 391, will result in little stereo-selectivity during nucleophilic attack by water.

Binding of the -1–boat Derived Oxocarbenium and Oxazoline Ion

The simulation results for the -1–boat derived oxocarbenium ion and oxazoline ion intermediates were very similar. The docked structures were generated from the hexaNAG -1–boat model, following removal of sugars +1 and +2. A series of annealing cycles were followed by 100 ps of MD.

The interactions of sugars -4 through -2 with specific residues of the binding site of Chitinase A for both intermediates were analogous to the hexaNAG simulation. Only the hydrogen bond from Asn 474 to O6′ of sugar -3 was missing. This was replaced by Arg 172, leaving the total number of hydrogen bonds unchanged. This is reflected in the greater stability of each of these intermediates over the -1–chair oxocarbenium ion (Table 3.4). Occupation of the active site by the oxazoline ion or -1–boat oxocarbenium ion is stabilized through favorable electrostatic interactions with Glu 315. In addition, hydrogen bonds were observed between O6′ and Arg 446 and from the *N*-acetyl amide to Glu 315.

It is evident from the low RMS coordinate fluctuation that the oxazoline ion binds tightly at subsite -1. This serves to stabilize the oxazoline intermediate and greatly limits the solvent accessible sites. Due to these geometric constraints, nucleophilic attack by water only can lead to retention of the β-anomer, as is observed experimentally.

Simulation of Isolated Chitinase A

As a control study, the same methods were applied to Chitinase A with no substrate or intermediates bound in the active site. The purposes of this simulation were to ensure that the methods used would generate a stable trajectory and to identify any structural changes that may take place upon substrate binding. The results indicate that the enzyme binding site does remain stable and the overall fold is unchanged. The RMS coordinate difference between the average dynamics structure and the crystal structure is 2.13 Å. This is on average 0.5 Å greater than the value observed for simulations which included a ligand (Table 3.1).

While no gross structural changes in the binding site were observed, some conformational changes took place during the early pre-equilibration portion (time = 0-20 ps) of the trajectory. These changes persisted for the remainder of the simulation. The side chains of residues Trp 275 and Trp 167, which form part of the hydrophobic "floor" of the binding site cleft, were very flexible. The largest deviation from the crystal structure was observed for the residues 163-172 and 470-478 which define the walls of the binding site cleft of subsites -4 and -3. These include several charged sidechains (Arg 172, Glu 473, Asp 478) which exhibited a tendency to move towards the solvent and widen the cleft. Little motion was observed in the enzyme active site pocket comprised of residues Asp 313, Glu 315, Met 388, Tyr 390 and Trp 539.

Ab initio QM Calculations for Protonated GlcNAc

Ab initio QM (HF/6-31G**) were used to assess the affects of the -1–boat substrate distortion upon the energetics of hydrolysis. The geometry of a protonated GlcNAc sugar residue was optimized starting from both a chair and boat conformation. When no geometric constraints were used, free rotation about the glycosidic bond allowed the proton to rotate towards the *N*-acetyl carbonyl, eventually transferring completely. As this rotation is not possible in the chitin polymer, a methyl substitution was made at O1′ (in place of a full neighboring GlcNAc residue) and the C(methyl)-O1′-C1′-C2′ torsion was constrained to remain at 120°. This constraint was used for both the chair and boat geometry optimizations.

Starting from a boat conformation of the protonated O1′-GlcNAc **2** (Figure 3.7a), a geometry optimization leads to spontaneous glycosidic bond cleavage with subsequent formation of an oxazoline ion/methanol complex, **3** (Figure 3.7a). In contrast, starting

from a chair conformation leads to a stable oxonium ion, **1** (Figure 3.7a). These interesting results suggest an important role for the -1–boat substrate distortion observed during the MD simulations. The optimum structure for the protonated chair lies 20.8 kcal/mol higher in energy than the oxazoline complex. While an exhaustive transition state search has not been carried out, a scan of different O1′-C1′ distances did not reveal an energy barrier which would hinder the formation of the oxazoline ion intermediate.

IV. Discussion

Proton Donation by Glu 315

It is generally accepted that the first step of the acid catalyzed hydrolysis mechanism of Chitinase A involves a proton transfer from Glu 315.[2] Evidence in support of this includes the observation that Glu 315 is completely conserved in family 18 chitinases. In addition, site-directed mutagenesis of the corresponding Glu residue in the Bacillus circulans chitinase to a Gln was reported to essentially eliminate activity.[37]

For the chitinase system, the likelihood of proton transfer primarily depends on the distance between the proton donor and acceptor. Making the assumption that Glu 315 is the donor, the proposed proton acceptor is the β(1,4)-glycosidic oxygen between sugar residues -1 and +1. A plot of the proton-oxygen distance for the hexaNAG simulations is shown in Figure 3.5. It is evident that the extended *N*-acetyl geometry of the -1–chair conformation places the glycosidic oxygen too far from Glu 315 (7 Å) for efficient proton transfer. In contrast, the -1–boat geometry places the proton between 3-4 Å away from the glycosidic oxygen and occasionally much closer when a direct hydrogen bond is made. From these simulations, it may be concluded that binding of a chitin substrate in

the -1–chair geometry is not compatible with protonation. Such a binding event would not lead to rapid hydrolysis.

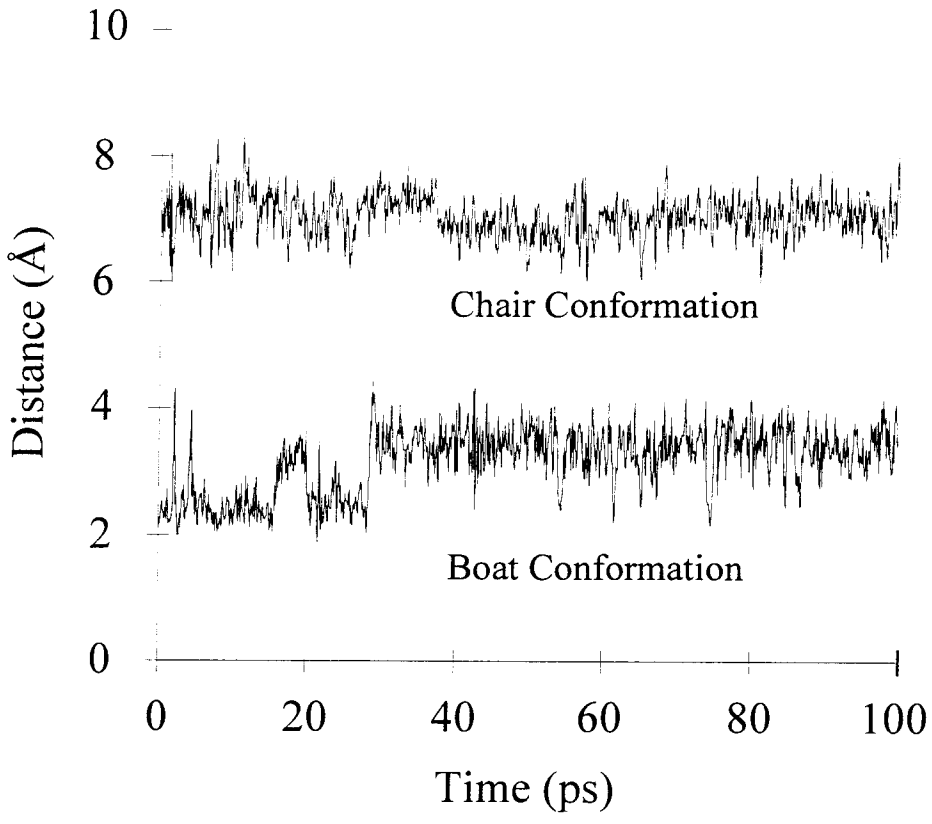


Figure 3.5. Proton– anomeric oxygen distance for the two stable hexaNAG conformations during the dynamics simulation. Proton transfer is likely for the -1–boat geometry only. Note the small fluctuations in these distances which indicates a stable trajectory during the simulation.

Substrate Distortion

A basic tenet of enzyme catalysis states that the catalytic rate enhancement achieved by an enzyme is a result of preferential binding (stabilization) of the transition state relative to the substrate.[38] In some instances, initial binding may induce a geometrical distortion in the glycosyl substrate. Evidence has been reported supporting

the role of substrate distortion for influenza virus neuraminidase,[39,40] endoglucanase I,[28] bacterial chitobiase,[27] 1,3-β-glucanases[41] and suggested for hen lysozyme.[15] Herein, it is reported that tight binding of the -1–boat hexaNAG substrate distorts the sugar residue at subsite -1 to induce a boat conformation (Figure 3.6). This is not observed for any other GlcNAc residues, all of which prefer the lower energy chair conformation.

The boat conformation observed for hexaNAG residue -1 is a consequence of several factors. A complementary fit of the -1–boat *N*-acetyl geometry to the active site pocket is essential and anchors C2′ in place. C4′ and C1′ are prevented from assuming the more favorable chair conformation due to the steric clash between the binding site cleft and hexaNAG residues +1 and +2. Indeed, upon removal of residues +1 and +2, molecular mechanics minimization of the remaining tetraNAG substrate reveals that the sugar at subsite -1 now prefers a chair conformation.

There is considerable debate regarding the significance of such substrate distortion for enzyme catalysis by the glycosyl hydrolases.[42] A recently reported QM study of 2-oxanol[43] suggests that substrate distortion determines the mechanistic path of acid-catalyzed glycosidic cleavage. Protonation of a chair conformation follows a step-wise path to a stable (high-energy) oxonium ion which then dissociates to an oxocarbenium ion. Protonation of all other sugar conformations induces a concerted dissociation leading directly to an oxocarbenium ion.

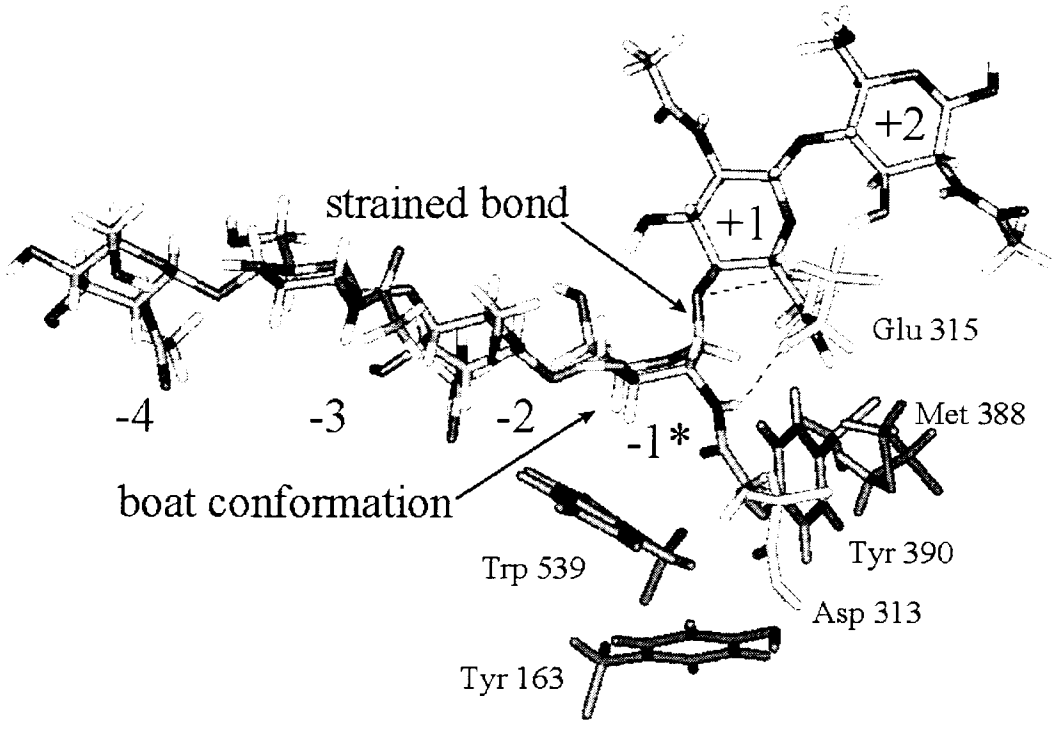


Figure 3.6. The minimum energy structure for the -1–boat hexaNAG conformation. A boat geometry for GlcNAc residue -1 and the twist between residues -1 and +1 strain the linking glycosidic bond. Glu 315 is found to be oriented so as to allow rapid proton transfer to the linking anomeric oxygen and to form a hydrogen bond with the *N*-acetyl amide.

We have extended this work to include O1′-methyl-GlcNAc. Geometry optimization starting from a boat conformation, **2** (Figure 3.7a), leads directly to bond cleavage and the formation of an oxazoline ion/methanol complex, **3**. The analogous optimization starting from a chair geometry does not lead to bond cleavage, but rather a stable oxonium ion (**1**), as was observed for 2-oxanol. Figure 3.7b shows a plot of the relative energy versus the C1′-O1′ distance (monitored during the geometry optimization). The chair conformation remains in a local minimum 20.8 kcal/mol above the oxazoline ion/methanol complex.

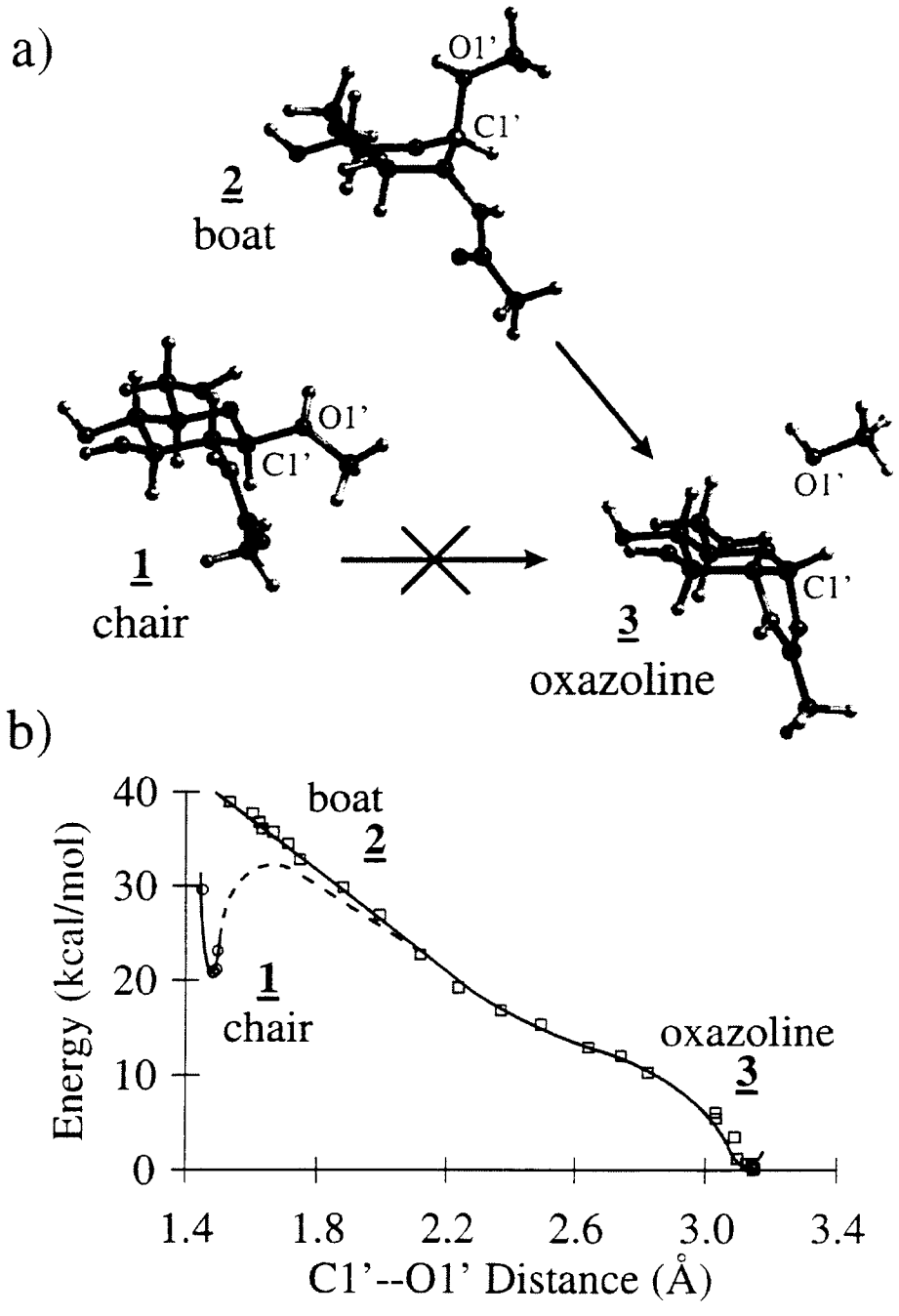


Figure 3.7. (a) (**1**) The optimized methyl substituted GlcNAc oxonium ion in a chair conformation. (**2**) The protonated boat conformation prior to glycosidic bond cleavage and formation of an oxazoline ion. (**3**) The optimum oxazoline ion/methanol complex. Absolute energy of **3** is -854.726 515 hartree. The optimized chair conformation (**1**) is 20.8 kcal/mol higher in energy than the oxazoline ion. (b) A plot of relative energy versus the C1′-O1′ distance during geometry optimization. The local minimum about the chair conformation is visible and no barrier between the protonated boat and the oxazoline ion is observed.

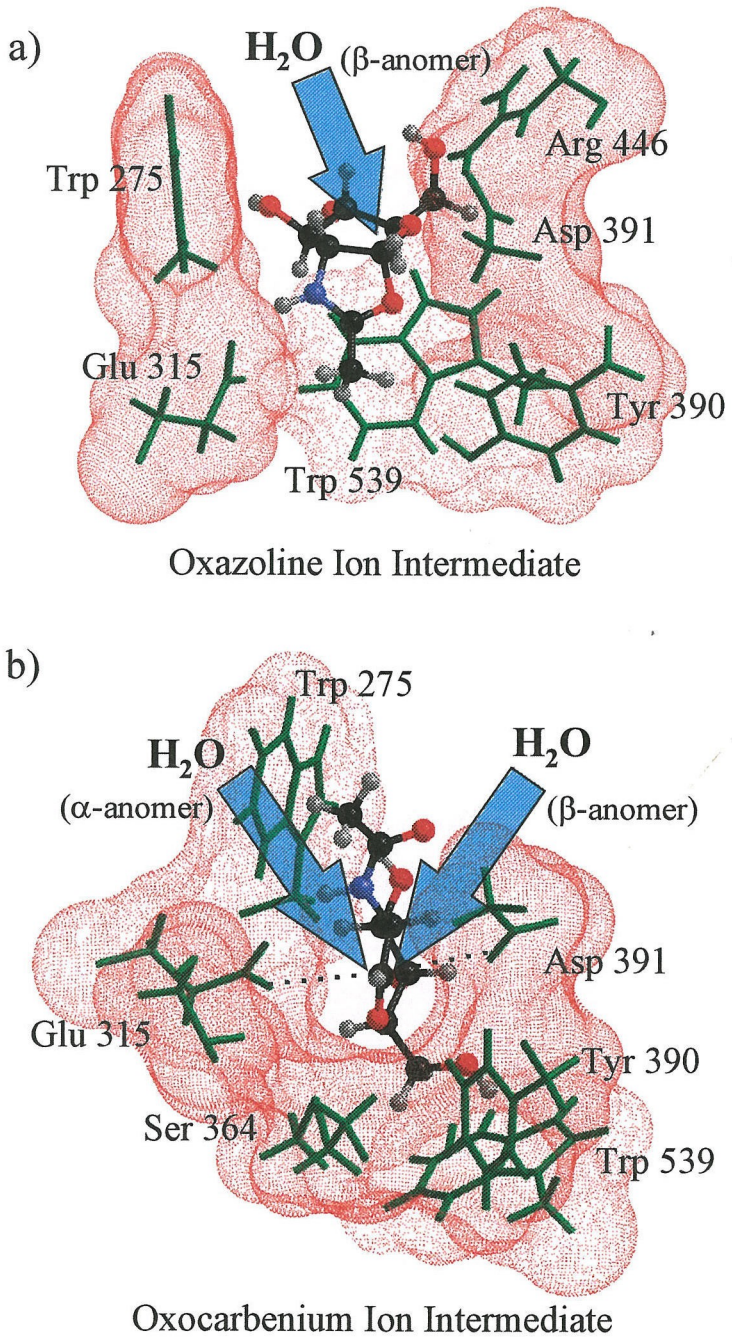


Figure 3.8. (a) The oxazoline ion intermediate bound to the chitinase A active site, shown with a vdW surface. Only one face of the oxazoline is open to attack by water at C1′ (as indicated with an arrow) which will lead to a single anomeric product. (b) The extended oxocarbenium ion intermediate is stabilized by Glu 315 and Asp 391 (dotted line) and the active site cleft is more narrow than for the oxazoline ion. Attack by water at C1′ is hindered equally on both sides of the cleft (as indicated by two arrows) and will likely yield a mix of anomeric products.

In light of this QM data and our simulation results it is reasonable to suggest that substrate distortion is a critical component of the oxazoline double-displacement mechanism. However, in the presence of a C2′-*N*-acetyl group with a favorable orientation, we propose protonation of a boat conformation will result in a concerted dissociation directly to an oxazoline ion with little or no barrier. For the purpose of the molecular dynamics simulations, we have assumed a stepwise mechanism in which proton transfer is followed by bond cleavage and formation of the oxazoline ion intermediate. Although these studies show that such a stepwise mechanism is allowed, they do not rule out the possibility of a completely concerted reaction mechanism with the proton transfer and glycosidic bond cleavage occurring simultaneously.

Active Site Structure

A model of the chitinase active site was proposed based on the X-ray structure of Chitinase A complexed with tetraNAG (N, N′, N″, N‴-tetra-acetylo-chitotetraose).[2] The crystal structure reveals only one sugar residue bound, perhaps because of hydrolysis of the substrate or disorder. Using this limited data, a model was generated based on the assumption that this sugar was bound to subsite -1. An alternative model was suggested in which this sugar was a product of the reaction and hence bound to subsite +1. However, this was subsequently dismissed based on poor electron density for a second bound sugar which would indicate the occupation of two "product" sites leaving the four remaining "substrate" sites unoccupied.

These simulation results are consistent with a model in which the sugar observed by X-ray crystallography does indeed occupy subsite +1, not subsite -1. This is not a

surprising result after considering the recent analysis of the conserved sequence and structure motifs for six family 18 glycosyl hydrolases.[14] This analysis allows an alignment to be made between Chitinase A and hevamine. Detailed structural information is available for hevamine including X-ray structures of a hevamine/triNAG complex and hevamine/allosamidin complex. Five of the six residues forming the active site of hevamine are conserved, therefore extrapolation to Chitinase A is trivial.

The reported electron density of the sugar residue observed by crystallography supports our model for sugars +1 and +2 of the -1–boat hexaNAG system. The most striking conformational feature of these two sugar residues is the twist induced between residues -1 and +1. The binding of residues +1 and +2 is reminiscent of that observed for HEWL, in which the sugars are parallel with the binding cleft. In contrast, the plane defined by the pyranose rings of sugar residues -4 through -1 is perpendicular to the binding cleft. This helps to induce the boat conformation observed for sugar -1 and may further induce stereochemical strain the glycosidic linkage at that point.

Potential Products from Reaction Intermediates

Although there is considerable theoretical evidence in support of an oxazoline ion intermediate, it is still useful to compare the predicted products of each possible intermediate. In these simulations, the oxazoline ion and -1–boat derived oxocarbenium ion both present the same solvent accessible surface. As a result of steric constraints of the active site pocket, nucleophilic attack by water would be predicted to yield exclusively the β-anomer (Figure 3.8a). This agrees with the experimentally observed product.

It is predicted that the -1–chair derived oxocarbenium ion intermediate with an "extended" *N*-acetyl geometry will not occur. However, if such a species were to arise, these simulations indicate the hydrolysis product would be a mix of α- and β-anomers. Both faces of the oxocarbenium ion are equally solvent accessible (Figure 3.8b). In addition, there are no specific side chain interactions which could exert a significant preference for nucleophilic attack by water from one side over the other as Glu 315 and Asp 391 are a similar distance from C1′ (3.19 Å and 3.04 Å, respectively). A slight preference for the β-anomer may be induced by Trp 275 which flanks a portion of the binding site. Therefore, the products predicted for an oxocarbenium ion intermediate are inconsistent with the reported experimental results.

Design of a New Class of Inhibitor

The inhibition of chitinases has been identified as a possible objective in the development of novel anti-fungal therapeutics.[10] When considering the rational design of a family 18 chitinase inhibitor, it is useful to identify unique points along the hydrolysis pathway which may be targeted. Two such points are the distorted boat conformation assumed at subsite -1 and the oxazoline ion intermediate. Indeed, inhibitors which mimic an oxazoline ion intermediate have been identified in the allosamidin family of natural products.[10,24,44] The allosamidins have been reported to be potent chitinase inhibitors and it is shown in Chapter 2, using *ab initio* QM, that allosamizoline shares characteristic structural and electronic properties of the oxazoline ion.

Based on these Chitinase A simulations, an alternative inhibitor may be designed to mimic the substrate bound at subsite -1 prior to protonation. The key structural features to be included in such a design are:

i. a constrained pyranose ring in a boat conformation (possibly through a substituted bicyclo[2.2.2]ocatane),

ii. an *N*-acetyl group in a position equivalent to C2′ of GlcNAc, and

iii. a hydrogen bond acceptor for proton on Glu 315.

Further modifications could be made in an attempt to occupy the "substrate" and "product" binding sites -3 through +2. In the case of the allosamidins, only the "substrate" sites are occupied. This inhibitor has the additional benefit of being neutral whereas the oxazoline ion is positively charged.

V. Conclusion

MD simulation methods have been applied to Chitinase A complexed with different hexaNAG substrate conformations and reaction intermediates. The results of these simulations offer considerable insight to the understanding of the hydrolysis mechanism. Protonation of the hexaNAG substrate by Glu 315 is likely to occur only for the -1–boat binding mode. The -1 sugar residue is distorted to a boat conformation by tight binding of the *N*-acetyl group and steric constraints between sugar residues +1 and +2 and the enzyme binding cleft. In addition, the planes of sugar residues -1 and +1 are twisted relative to one another, placing strain on the glycosidic linkage. QM data (HF/6-31G**) indicates that protonation of the distorted hexaNAG substrate leads to cleavage of the glycosidic bond and formation of an oxazoline ion in a concerted reaction mechanism with little or no barrier.

In Chapter 2, the possibility of an oxazoline ion intermediate is suggested based on the crystal structure of an allosamidin/hevamine complex and QM data. The allosamidins are potent inhibitors of some chitinases and are presumably transition state

analogs of the oxazoline intermediate. Herein, an alternative inhibitor design targeted against the initial hexaNAG binding event is proposed. Such an inhibitor would have preorganized structural features which take advantage of the need for a distorted boat conformation and have a neutral charge.

Acknowledgments

We are grateful to Professor Barbara Imperiali for many helpful discussions. This research was funded by DOE-BCTR (DE-FG36-93CH105 81, David Boron). The facilities of the MSC are also supported by grants from NSF (CHE 95-22179 and ASC 92-100368), Chevron Petroleum Technology Co., Saudi Aramco, Asahi Chemical, Owens-Corning, Exxon, Chevron Chemical Company, Chevron Research and Technology Co., Avery-Dennison, Hercules, BP Chemical, and Beckman Institute.

VI. References

(1) Watanabe, T.; Suzuki, K.; Oyanagi, W.; Ohnishi, K.; Tanaka, H. *J. Biol. Chem.* **1990**, *265*, 15659-15665.

(2) Perrakis, A.; Tews, I.; Dauter, Z.; Oppenheim, A. B.; Chet, I.; Wilson, K. S.; Vorgias, C. E. *Structure* **1994**, *2*, 1169-1180.

(3) Collinge, D. B.; Kragh, K. M.; Mikkelsen, J. D.; Nielsen, K. K.; Rasmussen, U.; Vad, K. *Plant J.* **1993**, *3*, 31-40.

(4) Bartnicki-Gracia, S. *Annu. Rev. Microbiol.* **1968**, *22*, 87-108.

(5) Kramer, K. J.; Dziadik-Turner, C.; Koga, D. *Comprehensive Insect Physiology, Biochemistry, and Pharmacology: Integument, Respiration, and Circulation. Vol. 3*; Pergamon Press: Oxford, 1985.

(6) Koga, D.; Isogai, A.; Sakuda, S.; Matsumoto, S.; Suzuki, A.; Kimura, S.; Ide, A. *Agri. Biol. Chem.* **1987**, *51*, 471-476.

(7) Fukamizo, T.; Kramer, K. J. *Insect Biochem.* **1985**, *15*, 141-145.

(8) Kuranda, M. J.; Robbins, P. W. *J. Biol. Chem.* **1991**, *266*, 19758-19767.

(9) Boller, T.; Gehri, A.; Mauch, F.; Vogeli, U. *Planta* **1983**, *157*, 22-31.

(10) Robertus, J. D.; Hart, P. J.; Monzingo, A. F.; Marcotte, E.; Hollis, T. *Can. J. Bot.* **1995**, *73*, S1142-S1146.

(11) Henrissat, B.; Bairoch, A. *Biochem. J.* **1993**, *293*, 781-788.

(12) Hart, P. J.; Pfluger, H. D.; Monzingo, A. F.; Hollis, T.; Robertus, J. D. *J. Mol. Biol.* **1995**, *248*, 402-413.

(13) Monzingo, A. F.; Marcotte, E. M.; Hart, P. J.; Robertus, J. D. *Nat. Struct. Biol.* **1996**, *3*, 133-140.

(14) Terwisscha van Scheltinga, A. C.; Hennig, M.; Dijkstra, B. W. *J. Mol. Biol.* **1996**, *262*, 243-257.

(15) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-859.

(16) Phillips, D. C. *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 484-495.

(17) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Bio.* **1994**, *4*, 885-892.

(18) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171-1202.

(19) Koshland, D. E. *Biol. Rev.* **1953**, *28*, 416-436.

(20) Fukamizo, T.; Koga, D.; Goto, S. *Biosci. Biotech. Biochem.* **1995**, *59*, 311-313.

(21) Iseli, B.; Armand, S.; Boller, T.; Neuhaus, J. M.; Henrissat, B. *Febs Letters* **1996**, *382*, 186-188.

(22) Armand, S.; Tomita, H.; Heyraud, A.; Gey, C.; Watanabe, T.; Henrissat, B. *Febs Letters* **1994**, *343*, 177-180.

(23) Terwisscha van Scheltinga, A. C.; Kalk, K. H.; Beintema, J. J.; Dijkstra, B. W. *Structure* **1994**, *2*, 1181-1189.

(24) Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W. *Biochemistry* **1995**, *34*, 15619-15623.

(25) Piszkiewiez, D.; Bruice, T. C. *J. Am. Chem. Soc.* **1967**, *89*, 6237-6243.

(26) Piszkiewiez, D.; Bruice, T. C. *J. Am. Chem. Soc.* **1968**, *90*, 2156-2163.

(27) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. *Nat. Struct. Biol.* **1996**, *3*, 638-648.

(28) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schulein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280-15287.

(29) POLYGRAF MSC Modified Version from Molecular Simulations, Inc., located in San Diego, CA.

(30) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(31) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144-1149.

(32) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J. M.; Goddard, W. A., III; Donnelly, R. E.; Ringnalda, M. N. *J. Chem. Phys.* **1994**, *101*, 4028-4041.

(33) Langlois, J. M.; Yamasaki, T.; Muller, R. P.; Goddard, W. A., III *J. Phys. Chem.* **1994**, *98*, 13498-13505.

(34) Muller, R. P.; Langlois, J. M.; Ringnalda, M. N.; Friesner, R. A.; Goddard, W. A., III *J. Chem. Phys.* **1994**, *100*, 1226-1235.

(35) Murphy, R. B.; Friesner, R. A.; Ringnalda, M. N.; Goddard, W. A., III *J. Chem. Phys.* **1994**, *101*, 2986-2994.

(36) Davies, G. J.; Wilson, K. S.; Henrissat, B. *Biochem. J.* **1997**, *321*, 557-559.

(37) Watanabe, T.; Kobori, K.; Miyashita, K.; Fujii, T.; Sakai, H.; Uchida, M.; Tanaka, H. *J. Biol. Chem.* **1993**, *268*, 18567-18572.

(38) Wolfenden, R. *Nature* **1969**, *223*, 704.

(39) Varghese, J. N.; McKimmin-Breschkin, J. L.; Caldwell, J. B.; Kortt, A. A.; Colman, P. M. *Proteins* **1992**, *14*, 327-332.

(40) Burmeister, W. P.; Ruigrok, R. W. H.; Cusack, S. *EMBO* **1992**, *11*, 49-56.

(41) Hrmova, M.; Garrett, T. P. J.; Fincher, G. B. *J. Biol. Chem.* **1995**, *270*, 14556-14563.

(42) Fresht, A. *Enzyme Structure and Mechanism*; W. H. Freeman and Co.: New York, 1985.

(43) Smith, B. J. *J. Am. Chem. Soc.* **1997**, *119*, 2699-2706.

(44) Sakuda, S.; Isogai, A.; Matsumoto, S.; Suzuki, A. *Tetrahedron Lett.* **1986**, *27*, 2475-2478.

Chapter 4

The Role of Enzyme Distortion in the Single-Displacement Mechanism of Family 19 Chitinases

Abstract

Using molecular dynamics simulations, the binding of a hexaNAG substrate and two potential hydrolysis intermediates (an oxazoline ion and an oxocarbenium ion) to a family 19 barley chitinase has been studied. It is found that the hexaNAG substrate binds with all sugars in a chair conformation, unlike the family 18 chitinase which causes substrate distortion. Glu 67 is in a position to protonate the anomeric oxygen linking sugar residues D and E while Asn 199 serves to hydrogen bond with the C2' *N*-acetyl group of sugar D, thus preventing the formation of an oxazoline ion intermediate. In addition, Glu 89 is part of a flexible loop region allowing a conformational change to occur within the active site to bring the oxocarbenium ion intermediate and Glu 89 closer by 4-5 Å. A hydrolysis product with inversion of the anomeric configuration occurs due to nucleophilic attack by a water molecule which is coordinated by Glu 89 and Ser 120. Issues important for the design of inhibitors specific to family 19 chitinases over family 18 chitinases are also discussed.

I. Introduction

Plants respond to pathogenic attack by producing defense proteins such as chitinases. Chitinases catalyze the hydrolysis of chitin, a β(1,4)-linked *N*-acetyl-glucosamine (GlcNAc) polysaccharide. Chitin, a fibrous and insoluble polymer, is a major structural component of many organisms including fungi, insects, and crustaceans. Fungal growth is limited by the degradation of fungal cell walls by the hydrolytic action of plant chitinases.[1,2] In addition, transgenic tobacco plants expressing a bean endochitinase gene have been reported to resist fungal infection better than non-transformed plants[3] demonstrating the importance of chitinases in the defense mechanism of higher plants.

Chitinases are included in the broad classification of glycosyl hydrolases and have been isolated from many different organisms including plants, insects, and bacteria. Based on amino acid sequence, chitinases are subdivided into two families (families 18 and 19) which differ in structure and mechanism.[4-6] Plant chitinases are also divided into classes I-V in which classes III and V belong to family 18 and classes I, II and IV comprise family 19. Recently, the first prokaryotic family 19 chitinase was isolated from *Streptomyces griseus* HUT 6037.[7] This is an important addition to the family 19 chitinases, previously found only in higher plants.

Before discussing the hydrolysis mechanism of family 19 chitinases, the focus of this paper, it is useful to review the general understanding of the mechanism for glycosyl hydrolases. Acid catalyzed glycosidic hydrolysis may proceed to yield a hydrolyzed product with either retention or inversion of the anomeric configuration (at C1') relative to the starting conformation. All family 18 chitinases reported to date[6,8,9] yield a β-

anomer hydrolysis product (retaining mechanism) while family 19 chitinase result in the α-anomer (inverting mechanism).[10]

Based on crystallographic structural data[9] and theoretical models (see Chapters 2 and 3), the retaining mechanism for family 18 chitinases has been proposed to involve anchimeric assistance by the C2' *N*-acetyl group and substrate distortion of sugar residue D (-1) to a boat conformation. Figure 4.1a shows how this mechanism requires only one active site carboxylic acid which acts as both a general acid and general base in addition to stabilizing the oxazoline ion intermediate.

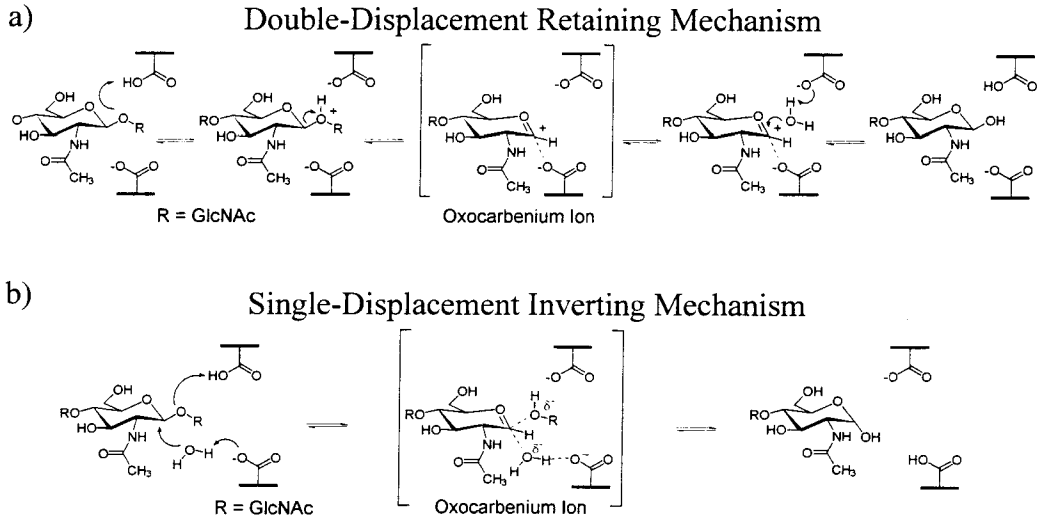


Figure 4.1. (a) The double-displacement hydrolysis mechanism proposed for family 18 chitinases. Protonation of a GlcNAc residue in a boat conformation leads to an oxazoline intermediate which may be hydrolyzed to form a product with retention of the anomeric configuration. (b) The single-displacement hydrolysis mechanism proposed for family 19 chitinases. Two acidic residues are required in the active site and the hydrolysis product has inversion of the anomeric configuration.

The crystal structure of a family 19 plant endochitinase isolated from barley (*Hordeum vulare* L.) seeds[11] reveals two acidic residues (Glu 67 and Glu 89) in the active site separated by 9.3 Å. A single displacement mechanism for family 19 chitinases is

often cited to account for inversion of the anomeric product and the need for two largely separated acidic residues within the active site (see Figure 4.1b). The single displacement mechanism requires one acidic residue to act as a general acid (Glu 67) and the other as a general base (Glu 89); activating water for a concerted nucleophilic attack at C1'.

The single displacement mechanism proposed for family 19 chitinases necessarily involves an intermediate with considerable oxocarbenium ion character. A similar intermediate arises for hen egg white lysozyme (HEWL) with the important exception that the second acidic residue is within a few angstroms of the oxocarbenium ion and stabilizes the charge by forming either a covalent bond to C1' or an ion pair.[12,13] Theoretical studies of HEWL indicate that this charge stabilization is a critical component of the enzymatic rate enhancement.[14,15] The proximity of the acidic residue also has the affect of forcing retention of the anomeric configuration. In solution, without the presence of enzymatic stabilization, *ab initio* quantum mechanical (QM) calculations predict the oxocarbenium ion to be 20 kcal/mol higher in energy than an oxazoline ion. Indeed, this provides much of the driving force for the oxazoline ion intermediate proposed for family 18 chitinases. Unlike the HEWL system, the second acidic residue of family 19 chitinases is considerably further from the forming oxocarbenium ion intermediate. Thus the problem of oxocarbenium ion charge stabilization is not well addressed in the current consensus view of the single displacement mechanism.

In order to investigate the detailed mechanism of family 19 chitinases, we have carried out molecular dynamics (MD) simulations on the barley seed endochitinase complexed with a hexaNAG substrate and two possible intermediates (triNAG-oxocarbenium ion and triNAG-oxazoline). The following conclusions are reached:

i. The hexaNAG substrate binds with all sugars in a chair conformation such as to favor protonation by Glu 67 of the anomeric oxygen linking sugar residues D and E.

ii. Asn 199 serves to hydrogen bond with the C2' *N*-acetyl group of sugar D, preventing the formation of an oxazoline ion intermediate.

iii. Glu 89 is situated in a flexible loop region allowing a conformational change to occur within the active site to bring the oxocarbenium ion intermediate 4-5 Å closer to Glu 89.

iv. Glu 89 and Ser 120 coordinate with a water molecule which may be activated for nucleophilic attack; this would yield a product with an inverted anomeric configuration (as observed).

v. Inhibitors designed to be transition state analogs of the more planar oxocarbenium ion are predicted to be selective against family 19 chitinases over family 18 chitinases.

II. Methods

Details regarding the simulation methods, charges, solvation calculations, and simulation software were described in Chapters 2 and 3.

The starting structure for the hexaNAG/chitinase simulation was based on the crystal structure of NAG-NAM-NAG/HEWL.[16] In predicting the structure for barley chitinase, we matched residues 56-68, 112-120, and 148-160 of barley chitinase to residues 24-36, 51-59, and 89-101 of HEWL. This provided a model for sugar residues B-D (after conversion to all GlcNAc sugar). Sugar residues A, E-F were built individually, starting from a range of different conformations and optimized through simulated annealing. Each conformation was subjected to 10 annealing cycles during

which the temperature was raised from 0° K to 350° K and lowered back to 0° K in increments of 50° K every 100 fs for a total of 2.1 ps. This resulted in only one family of hexaNAG structures. The lowest energy conformation in this family was then energy minimized prior to dynamics simulations.

Starting structures for the bound oxocarbenium ion and oxazoline ion intermediates were based on the optimized hexaNAG model. GlcNAc residues E and F were removed and the correct changes in atom hybridization were applied. These conformations were then subjected to annealing dynamics and energy optimization using a protocol analogous to that described above for the hexaNAG substrate.

Only residues comprising the binding site of barley chitinase were allowed to move during the optimizations and MD simulations. This chitinase binding site included residues: 66-70, 86-96, 110-130, 154-165, 197-203, and 211-214 (based on a 6 Å distance cutoff from the bound hexaNAG substrate). The simulations included crystallographic water molecules plus counterions for solvent exposed residues (leading to a net neutral charge). Water molecules in van der Waals (vdW) contact with the docked ligand were displaced sufficiently to avoid high energy starting conformations.

III. Results

Simulations of HexaNAG Substrate Binding

MD simulations were used to study the binding of a hexaNAG substrate to the family 19 barley chitinase. The conformation of the hexaNAG substrate bound to barley chitinase was based on the reported crystal structure of HEWL complexed with a NAG-NAM-NAG trisaccharide. Sugar residues A, E-F (labeled A-F from the non-reducing end, following the convention for lysozyme) were built and optimized to yield the final

hexaNAG substrate. A range of conformations for the added sugar residues A, E, and F were examined but only one low energy global conformation was found to be stable during simulated annealing simulations from 0° K to 350° K.

A 100 ps MD simulation was carried out in which the enzyme binding site residues, crystallographic waters, counterions and hexaNAG substrate were all free too move. During this simulation, the hexaNAG substrate was stable as was the enzyme binding site. The average coordinate RMS for all movable residues was 1.31 Å (see Table 4.1). All six GlcNAc residues remained in a chair conformation. Table 4.2 shows the average coordinate RMS fluctuation of each sugar residue. This result suggests that the D sugar residue is bound most tightly while residue A is least strongly bound.

Table 4.1. RMS Coordinate Difference (Å) for the Binding Site Residues of Barley Chitinase.[a]

substrate	av RMS[b]
hexaNAG	1.31
triNAG-oxocarbenium	1.14
triNAG-oxazoline	1.51
none	1.23

[a]Comparing molecular dynamics with the crystal structure obtained without substrate.
[b]Coordinate RMS difference was calculated as the difference between the crystal structure coordinates and the average position for all nonhydrogen atoms during the MD interval from 30 to 100 ps.

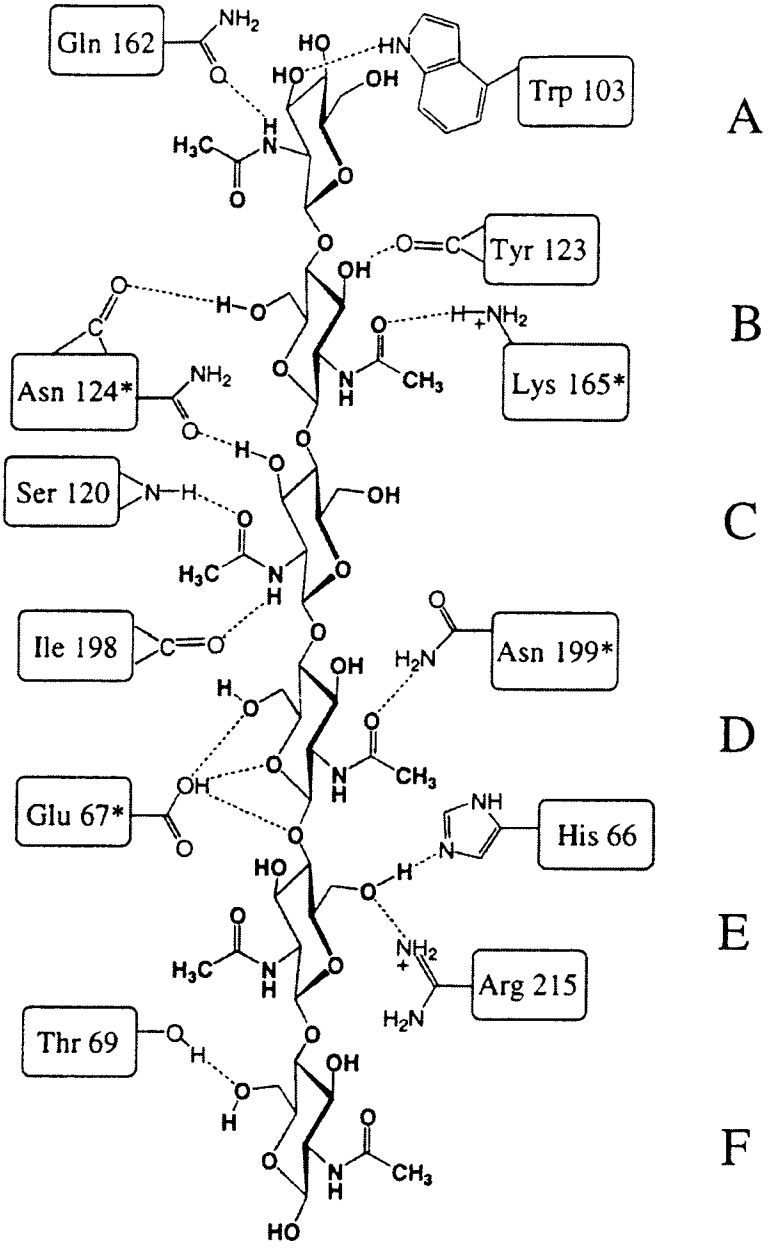


Figure 4.2. A schematic of the hydrogen bonds observed for a hexaNAG substrate bound to barley chitinase. Residues marked with an (*) are strongly conserved in family 19 chitinases.

Table 4.2. RMS Coordinate Fluctuations (Å) from Molecular Dynamics of Sugars Bound to Barley Chitinase.[a]

	binding subsites					
conformation	A	B	C	D	E	F
hexaNAG	0.908	0.498	0.458	0.405	0.531	0.486
triNAG-oxocarbenium	0.501	0.447	0.384	0.411	–	–
triNAG-oxazoline	0.411	0.340	0.381	0.450	–	–

[a]The hexaNAG substrate and two triNAG intermediates were considered. [b]RMS coordinate fluctuations were calculated as follows: The average structure for the MD interval from 30 to 100 ps was determined. The RMS coordinate difference between this average structure and snapshots taken every 5 ps from 30 to 100 ps were determined. The average of this RMS is reported separated by sugar residue.

Tight binding of the hexaNAG substrate is achieved through well defined hydrogen bonds which are constant during the simulations. Figure 4.2 shows a schematic of these interactions and highlights residues which are strongly conserved in family 19 chitinases. The *N*-acetyl amide of sugar A donates a hydrogen bond to the side chain of Gln 162 and O3' accepts a hydrogen bond from Trp 103. HO3' and HO6' of sugar B donate hydrogen bonds to the backbone carbonyls of Tyr 123 and Asn 124, respectively. In addition, the *N*-acetyl carbonyl forms a hydrogen bond with the charged side chain of Lys 165. The *N*-acetyl amide of sugar C donates a hydrogen bond to the backbone carbonyl of Ile 198 and the carbonyl of the *N*-acetyl group accepts a hydrogen bond from the backbone amide of Ser 120. Asn 124 also forms a hydrogen bond with O3' of sugar C. Two critical hydrogen bonding interactions are observed for sugar D. The first is between Asn 199 and the *N*-acetyl carbonyl which serves to constrain the *N*-acetyl geometry. The second hydrogen bond is transient forming between the protonated Glu 67

and O5', O6' or O1' (see Figure 4.3). Sugar residues E and F make the fewest specific contacts. HO6' of sugar E donates a hydrogen bond to His 66 while O6' accepts a hydrogen bond from Arg 215. Sugar F forms only one hydrogen bond between O6' and the side chain of Thr 69.

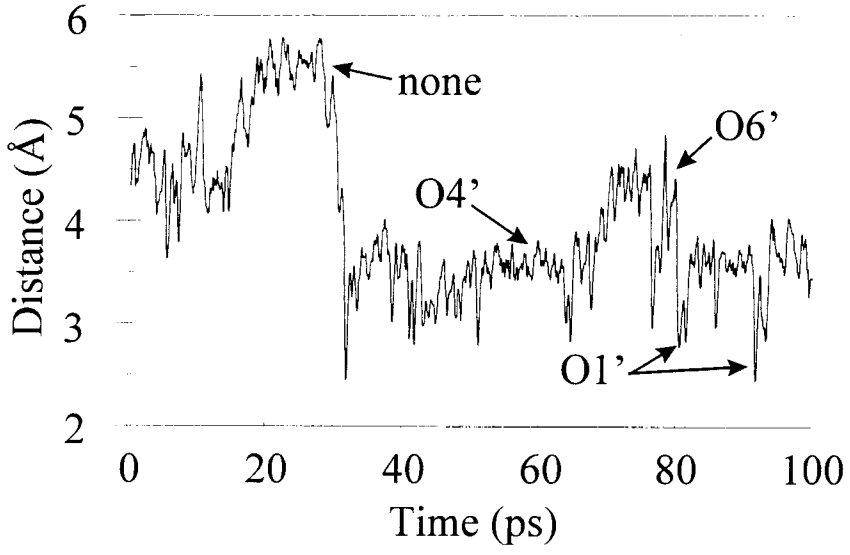


Figure 4.3. The distance between the proton of Glu 67 and the C1' anomeric oxygen linking sugar residues D and E. At times hydrogen bonds are made with O5' and O6' as indicated.

Figure 4.3 shows the average distance during the MD simulation between the proton of Glu 67 and the β(1,4)-glycosidic oxygen linking sugar residues D and E. During the early part of the simulation (0-30 ps), only brief hydrogen bonds are formed. However, for the remainder of the trajectory, transient bonds are seen oscillating between O5', O6' and O1' (with O5' most frequent and 01' least frequent).

Simulations of Hydrolysis Reaction Intermediates

Protonation of the anomeric oxygen linking sugar residues D and E followed by subsequent bond cleavage results in a charged intermediate. Since the experimentally

observed hydrolysis product has an inverted stereochemistry at C1', an oxocarbenium ion intermediate is probable. However, both triNAG-oxocarbenium and triNAG-oxazoline intermediates bound to barley chitinase were simulated for 100 ps. Starting structures were generated from the hexaNAG/chitinase complex by deletion of sugar residues E and F followed by conversion of sugar D to the appropriate intermediate.

Binding of the Oxocarbenium Ion

The MD simulation of the triNAG-oxocarbenium ion bound to barley chitinase yielded a stable trajectory with an average enzyme coordinate RMS of 1.14 Å. The sugar residues are slightly more stable than observed for the hexaNAG substrate as is evident by the smaller coordinate RMS fluctuations reported in Table 4.2. Most surprising was the observed change in conformation for residues 89-92 (the "Glu 89 loop"). These changes do not result in a gross structural change of the enzyme (as indicated by the low total coordinate RMS). However, the conformation of the terminus of the acidic sidechain of Glu 89 shifts 2 Å towards C1' of the oxocarbenium ion. Simultaneously, the departure of sugar residues E and F allows the oxocarbenium ion to also move 2-3 Å towards Glu 89. This results in a reduction of the Glu 89 - C1' distance by 4-5 Å (see Figure 4.4) and dramatic charge stabilization.

In addition to stabilizing the oxocarbenium ion, part of Glu 89 is well solvated and a specific water molecule is observed to coordinate between Ser 120, Glu 89 and the positive charge of the oxocarbenium ion. Figure 4.5 shows a snapshot from the simulation which highlights this bridged water molecule. This result indicates that the conformational change of Glu 89 and of the oxocarbenium ion intermediate place the Glu

89-H O-C1' atoms appropriately to complete the hydrolysis reaction and inversion of the D sugar.

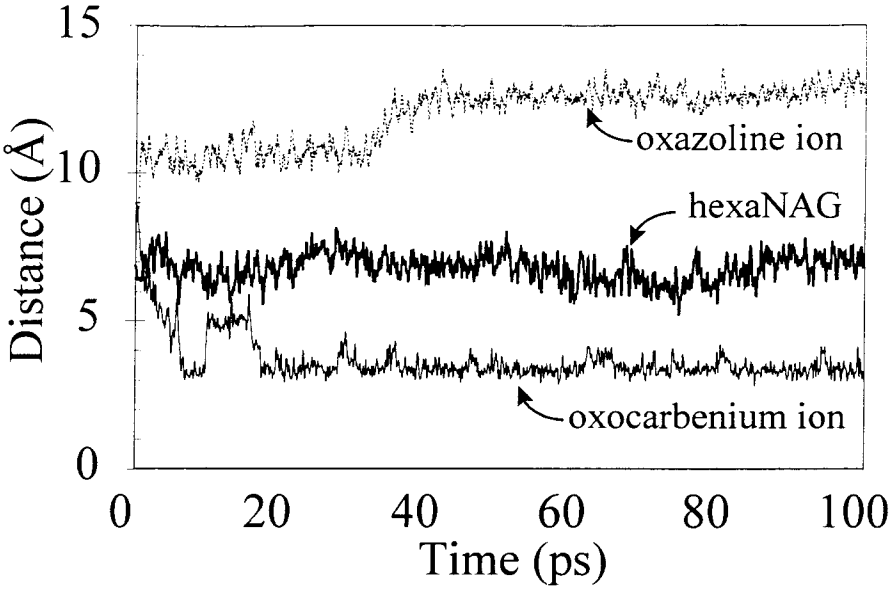


Figure 4.4. The distance between the Glu 89 carbonyl oxygen and C1' of sugar D (or N for the oxazoline case).

No changes in the hydrogen bonds for sugar residues A-C took place as a result of the shift in conformation at residue D. However, Asn 199 no longer forms a hydrogen bond with the *N*-acetyl carbonyl of sugar D which now becomes solvent accessible. A new hydrogen bond is formed between the deprotonated Glu 67 and HO6'. The proximity of Glu 89 and the bridging water molecule to C1' prevents the formation of an oxazoline ion because of steric constraints between the bound water molecule and the *N*-acetyl group. Such an intermediate could possibly occur if the formation of the oxazoline ion was sufficiently favorable enough to displace the bound water molecule.

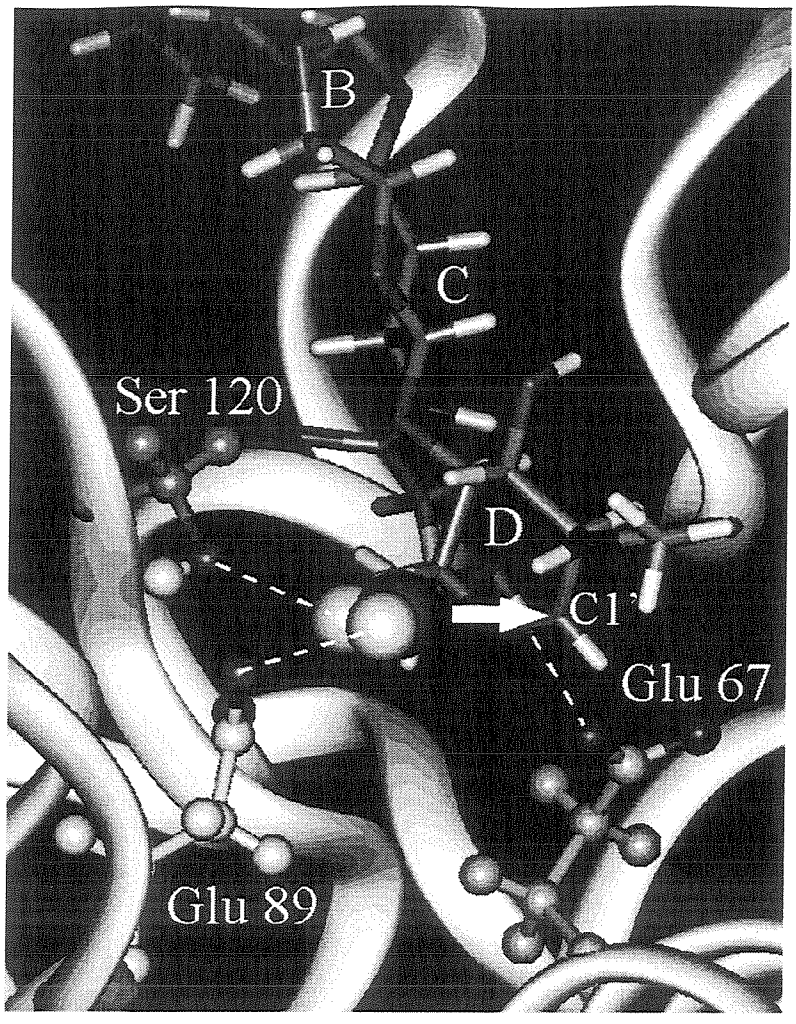


Figure 4.5. A snapshot from the dynamics simulations of a tri-NAG oxocarbenium ion intermediate bound to barley chitinase. An arrow marks the α-face of the oxocarbenium ion which will result in inversion of the anomeric configuration upon nucleophilic attack by water. Ser 120 and Glu 89 are positioned to coordinate with this water molecule as shown with dashed lines. Glu 67 forms a hydrogen bond with HO6'.

Binding of the Oxazoline Ion

The MD simulations of triNAG-oxazoline bound to barley chitinase indicate that an oxazoline reaction intermediate is extremely unlikely. Although the net charge is the same for the oxazoline ion and oxocarbenium ion intermediates, these intermediates complexed with barley chitinase behave very differently. The oxazoline ion is not stabilized by Glu 89, in contrast to the oxocarbenium ion. Indeed, throughout the

simulation, the average distance between the positive oxazoline nitrogen and Glu 89 *increases*, eventually converging to a distance of 12-13 Å (see Figure 4.4). This distance is >8 Å greater than for the oxocarbenium ion, resulting in a significant loss of charge stabilization. The oxazoline ion is not charge stabilized by other binding site residues and is predominately exposed to solvent.

The calculation of relative binding energies for small ligands to proteins is fraught with peril due to complications resulting from solvation, polarization, and entropic affects. However, such calculations may reveal trends useful in obtaining mechanistic insights. Consequently, the relative binding energies for the triNAG-oxazoline are compared to the triNAG-oxocarbenium ion. Due to the charge stabilization by Glu 89, the molecular mechanics ligand binding energy (internal energy) favors the oxocarbenium ion by 81 kcal/mol. The differential solvation energy for each ligand bound and unbound was calculated using the Poisson-Boltzmann continuum solvation approximation.[17] Because of the greater solvent accessibility, the oxazoline ion is expected to lose less solvation energy upon binding. Indeed, the oxazoline ion was found to have a more favorable solvation energy by 51 kcal/mol. Combining these contributions, we estimate the oxocarbenium ion to be more stable by 30 kcal/mol. The validity of these methods for determining binding energies has not been established and the 30 kcal/mol difference should be considered as a qualitative estimate. These results do indicate a preference for the oxocarbenium ion intermediate.

IV. Discussion

These MD simulations on a family 19 barley chitinase complexed with a hexaNAG substrate and two possible hydrolysis intermediates lead to a single

displacement mechanism with an inverted anomeric configuration for the hydrolysis product. These simulation results are in excellent agreement with experimental results, offering considerable insight to the specific interactions which stabilize the oxocarbenium ion intermediate. I propose a modified single displacement reaction mechanism in which Glu 89 serves to charge stabilize the oxocarbenium ion in addition to recruiting and activating a nucleophilic water. Next I will discuss the roles of other important enzyme residues within the active site and indicate some biochemical tests of these predictions.

Protonation of the HexaNAG Substrate

The presented model for binding of the hexaNAG substrate to barley chitinase is in general agreement with the model proposed by Hart et al.[11] Minor differences were observed in the positioning of sugar residues A and F (which make few hydrogen bonds and are not tightly bound). It is likely that Glu 67 donates a proton to the β(1,4) glycosidic oxygen linking sugar residues D and E. Figure 4.3 shows the distance between the Glu 67 proton and this anomeric oxygen. Following an equilibration period of 30 ps, the proton remains within 4 Å of O1'. At times a hydrogen bond is made directly with the anomeric oxygen (O1') although much of the time the proton is involved in a hydrogen bond with either O5' or O6'.

Ab initio QM calculations predict that protonation at O1' of 2-oxanol in the chair conformation leads to an oxonium ion.[18] There is a significant energy barrier (0.82 kcal/mol) preventing formation of an oxocarbenium ion directly from the oxonium ion (this is the barrier to glycosidic bond cleavage). The magnitude of this barrier will depend on the local enzyme structure. Our simulations assume a stepwise mechanism in which protonation is followed by bond cleavage and formation of a charged intermediate

(oxocarbenium or oxazoline ion). Certainly there will be oxocarbenium character in the intermediate but it is not certain that the lifetime for this will be sufficiently long to be a discrete observable intermediate.

Charge Stabilization by Glu 89

Simulations of the oxocarbenium ion intermediate bound to barley chitinase yields the surprising result that Glu 89 serves to charge stabilize the oxocarbenium ion through a conformational change in both substrate and enzyme. Commonly the role of the second acidic residue in a single displacement reaction mechanism (the first having donated a proton as discussed above) is referred to in general terms as accommodating a water molecule which can participate in nucleophilic attack at C1'.[19,20] Although this is also observed in the simulations, charge stabilization is likely to be of comparable importance for this mechanistic pathway.

The distance of Glu 89 from C1' (or the protonated oxazoline nitrogen) of sugar residue D depends upon the exact nature of the substrate. Figure 4.4 shows a plot of this distance during the MD simulations. There are three distinct responses observed:

i. HexaNAG substrate binding has little affect on the Glu 89 loop region and the average distance of 8 Å remains fairly constant throughout the simulation.

ii. Upon departure of sugar residues E and F, the oxocarbenium ion is allowed to move towards Glu 89. A simultaneous change in the conformation of Glu 89 results in a 4 Å reduction in the separation distance and considerable charge stabilization.

iii. In contrast, for the oxazoline ion (which has the same net charge) the Glu 89 moves away from the charge center eventually reaching a distance of 12 Å. This indicates that the oxazoline gains no stabilization from Glu 89.

This lack of stabilization by Glu 89 provides additional evidence against a role for the oxazoline ion intermediate. In these calculations, the free energy change associated with this change in enzyme structure has not be determined explicitly and the assumed stabilization is a qualitative estimate.

Crystallographic Data Supports a Flexible "Glu 89 Loop"

Our model of a flexible "Glu 89 loop" is supported by crystallographic data. Residues 89-95 have the highest temperature factors (25-30 $Å^2$) within the entire enzyme (excluding end effects).[11] This is in agreement with the average RMS coordinate differences between the simulation and crystal structure as shown in Figure 4.6. In addition, this indicates considerable thermal instability of the Glu 89 loop and the lack of a well defined packed structure.

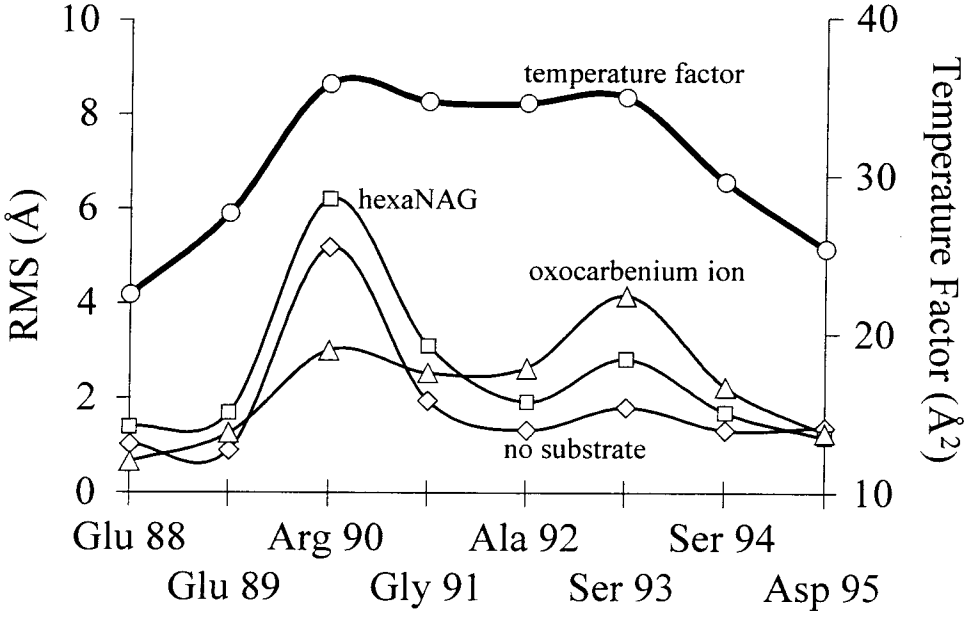


Figure 4.6. Comparison of the RMS coordinate difference of various residues from MD (30-100ps) with the experimental temperature factors. This data supports the flexibility of the Glu 89 loop region. The average temperature factors for all other residues (except the amino and carboxy-termini) are less than 22 $Å^2$.

Soaking of monoclinic barley chitinase crystals with tetraNAG caused the crystals to crack and dissolve, only to reform 24 hours later. Based on these observations, it was suggested that "tetraNAG substrate binding causes the molecule to undergo a conformational change incompatible with maintenance of crystal contacts."[11] The crystals could reform in the native conformation upon cleavage of tetraNAG to the disaccharide, which remains unbound. Our simulations indicate that binding of hexaNAG does not cause a change in conformation. However, during the hydrolysis mechanism, a conformational change occurs to stabilize the oxocarbenium ion. Such a change could cause the crystals to dissolve.

Predicted Mechanistic Details

The MD simulations also offer considerable insight about the role of specific active site residues during substrate binding and hydrolysis. Some of these residues are strictly conserved in all family 19 chitinases. Figure 4.2 shows the contacts made during hexaNAG binding and highlights conserved residues. Particularly critical are the interactions with Lys 165, Asn 124, Asn 199 and Glu 67, all of which are strongly conserved in family 19 chitinases, including the recently isolated bacteria chitinase. It is found that Asn 199 serves a unique role of hydrogen bonding to the *N*-acetyl group of sugar D. This forces an extended geometry for the *N*-acetyl group and prevents the formation of an oxazoline ion intermediate. As a test of these results, mutations of any of these residues would be predicted to reduce hydrolysis activity.

Indeed, some of these mutagenesis studies have been recently reported following the development of a heterologus expression system for a barley endochitinase.[21] The mutations of Glu 67 and Glu 89 to Gln both resulted in complete loss of catalytic activity

as would be expected for a single displacement reaction mechanism. Furthermore, the mutation of Asn 124 to Ala reduced the enzyme activity to 0.82 of the wild-type. These findings are consistent with our model for substrate binding in which Asn 124 forms two hydrogen bonds to the hexaNAG substrate.

In addition to stabilizing the positively charged oxocarbenium ion, Glu 89 coordinates with several water molecules. One of these water molecules becomes bridged between Ser 120, Glu 89, and the oxocarbenium ion (Figure 4.5). The positioning of this water ensures nucleophilic attack from the α-face of the oxocarbenium ion to yield only an inverted stereochemistry at C1'. A mutation which would position Glu 89 close enough to the oxocarbenium ion to form a covalent bond to C1' would force an alternative mechanism to occur (the double displacement mechanism of HEWL) leading to retention of the anomeric configuration.

An analysis of the HEWL structure matched to barley chitinase suggests a Gly 113 to Asp mutation (and the complementary Glu 89 to Gly mutation) would result in an active site architecture similar to HEWL and the associated change in mechanism. However, simulations of this mutation show that there is still sufficient space to accommodate a water molecule between the oxocarbenium ion and Asp 113. A mutation to Glu 113 would be predicted to occupy a large enough space to prevent coordination of a water molecule and force a double-displacement retaining mechanism.

Design of Family Specific Chitinase Inhibitors

Family 18 and 19 chitinases utilize different hydrolysis mechanisms: family 18 involving an oxazoline intermediate while family 19 involves an oxocarbenium ion intermediate. As a result, specific transition state analogs could be designed for each

family of chitinase. Indeed, the allosamidins mimic an oxazoline ion intermediate,[9] making them potent family 18 chitinase inhibitors.[22-24] Family 19 inhibitors could be designed to mimic the more planar oxocarbenium ion geometry with the delocalized positive charge. Some such inhibitors have been synthesized including amidines,[25] amidrazones,[25] and nojiritetrazoles.[26] It is expected that selective inhibition of family 19 chitinases can be achieved by condensation of one of these transition state analogs with the NAG-NAG-NAG glycone specificity of chitinase.

V. Conclusion

The methods of MD simulations have been applied to examine the hydrolysis mechanism of a family 19 barley chitinase. These simulations are consistent with experimental results which indicate a single displacement reaction mechanism. In addition, the oxocarbenium ion intermediate may be stabilized through electrostatic interactions with Glu 89 following a conformational change both in the binding geometry of the tri-NAG oxocarbenium ion intermediate and the enzyme active site. A similar conformational change for the oxazoline ion intermediate was not observed and the oxazoline ion is not likely to be part of the hydrolysis mechanism.

Acknowledgments

This research was funded by DOE-BCTR (DE-FG36-93CH105 81, David Boron). The facilities of the MSC are also supported by grants from NSF (CHE 95-22179 and ASC 92-100368), Chevron Petroleum Technology Co., Saudi Aramco, Asahi Chemical, Owens-Corning, Exxon, Chevron Chemical Company, Chevron Research and Technology Co., Avery-Dennison, Hercules, BP Chemical, and Beckman Institute.

References

(1) Boller, T.; Gehri, A.; Mauch, F.; Vogeli, U. *Planta* **1983**, *157*, 22-31.

(2) Schlumbaum, A.; Mauch, F.; Vogeli, U.; Boller, T. *Nature* **1986**, *324*, 365-367.

(3) Broglie, K.; Chet, I.; Holliday, M.; Cressman, R.; Biddle, P.; Knowlton, S.; Mauvais, C. J.; Broglie, R. *Science* **1991**, *254*, 1194-1197.

(4) Henrissat, B. *Biochem. J.* **1991**, *280*, 309-316.

(5) Henrissat, B.; Bairoch, A. *Biochem. J.* **1993**, *293*, 781-788.

(6) Iseli, B.; Armand, S.; Boller, T.; Neuhaus, J. M.; Henrissat, B. *Febs Letters* **1996**, *382*, 186-188.

(7) Ohno, T.; Armand, S.; Hata, T.; Nikaidou, N.; Henrissat, B.; Mitsutomi, M.; Watanabe, T. *J. Bacter.* **1996**, *178*, 5065-5070.

(8) Armand, S.; Tomita, H.; Heyraud, A.; Gey, C.; Watanabe, T.; Henrissat, B. *Febs Letters* **1994**, *343*, 177-180.

(9) Terwisscha van Scheltinga, A. C.; Armand, S.; Kalk, K. H.; Isogai, A.; Henrissat, B.; Dijkstra, B. W. *Biochemistry* **1995**, *34*, 15619-15623.

(10) Fukamizo, T.; Koga, D.; Goto, S. *Biosci. Biotech. Biochem.* **1995**, *59*, 311-313.

(11) Hart, P. J.; Pfluger, H. D.; Monzingo, A. F.; Hollis, T.; Robertus, J. D. *J. Mol. Biol.* **1995**, *248*, 402-413.

(12) Phillips, D. C. *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 484-495.

(13) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171-1202.

(14) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227-249.

(15) Dao-Pin, S.; Liao, D. I.; Remington, S. J. *Proc. Natl. Acad. Sci., USA* **1989**, *86*, 5361-5365.

(16) Kelly, J. A.; Sielecki, A. R.; Sykes, B. D.; James, M. N. G.; Phillips, D. C. *Nature* **1979**, *282*, 875-878.

(17) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144-1149.

(18) Smith, B. J. *J. Am. Chem. Soc.* **1997**, *119*, 2699-2706.

(19) McCarter, J. D.; Withers, S. G. *Curr. Opin. Struct. Bio.* **1994**, *4*, 885-892.

(20) Davies, G.; Henrissat, B. *Structure* **1995**, *3*, 853-859.

(21) Andersen, M. D.; Jensen, A.; Robertus, J. D.; Leah, R.; Skriver, J. K. *Biochem. J.* **1997**, *322*, 815-822.

(22) Sakuda, S.; Isogai, A.; Matsumoto, S.; Suzuki, A. *Tetrahedron Lett.* **1986**, *27*, 2475-2478.

(23) Koga, D.; Isogai, A.; Sakuda, S.; Matsumoto, S.; Suzuki, A.; Kimura, S.; Ide, A. *Agri. Biol. Chem.* **1987**, *51*, 471-476.

(24) Kinoshita, M.; Sakuda, S.; Yamada, Y. *Biosci. Biotech. Biochem.* **1993**, *57*, 1699-1703.

(25) Papandreou, G.; Tong, M. K.; Ganem, B. *J. Am. Chem. Soc.* **1993**, *115*, 11682-11690.

(26) Ermert, P.; Vasella, A.; Weber, M.; Rupitz, K.; Withers, S. G. *Carbo. Res.* **1993**, *250*, 113-128.

Chapter 5

Distance Dependent Hydrogen Bond Potentials for Nucleic Acid Base Pairs From *Ab initio* Quantum Mechanical Calculations (LMP2/cc-pVTZ)

Abstract

Hydrogen bonding between base pairs in nucleic acids is a key determinant of their structures. The distance dependence of the hydrogen bonding of AT-WC (Watson-Crick), GC-WC and AT-H (Hoogsteen) base pairs have been examined using *ab initio* quantum mechanics, LMP2/cc-pVTZ(-f) energies at HF/cc-pVTZ(-f) optimized geometries. From these curves, Morse potentials have been extracted between the H atoms and the acceptor atoms that accurately reproduces the quantum mechanical energies for a range of geometries. Using these parameters, the complexation energies of the remaining 26 possible pairwise combinations have been calculated and the agreement with previously reported *ab initio* calculations is excellent. Lennard-Jones 12-6 parameters to be used with the popular AMBER95 and CHARMM95 FFs have also been extracted. These off-diagonal terms significantly improve their descriptions of the base pairing energy and optimum geometry.

I. Introduction

Hydrogen bonds play a key role in maintaining structure and specificity of biological systems. In particular, the base pairing of nucleic acids (stemming from the specific formation of hydrogen bonds between Watson-Crick base pairs) is essential for the transfer of genetic information. We examine here the three of greatest biological relevance; the Watson-Crick base pairs, in which adenine hydrogen bonds to thymine (AT-WC) or guanine hydrogen bonds to cytosine (GC-WC), and the Hoogsteen adenine-thymine pair (AT-H).

Given the importance of hydrogen bonding in biological systems, considerable theoretical attention has been focused on exploring the nature and strength of these interactions. Extensive calculations have been carried out on the nucleic acid base pairs using semi-empirical or *ab initio* quantum mechanical (QM) methods. For a recent review covering the current state of the field and discussing the extensive progress from semi-empirical to *ab initio* QM methods, see reference 1. It is now established that the determination of accurate hydrogen bond energies and geometries requires a large diffuse basis set and the inclusion of electron correlation.[1-11] Many of the earlier studies of the nucleic acid base pairs did not include electron correlation effects, or estimated correlation energies with empirical methods. Recently, a series of studies have been reported in which dispersion energies have been evaluated using second order Moller-Plesset perturbation theory (MP2)[12-18] or density functional theory (DFT).[19]

Experimentally, nucleic acid base pairing has been difficult to study and there is little data with which theoretical results may be compared. Gas phase association

energies have been reported for some systems[20] and solution studies have been undertaken in nonpolar solvents.[21] The limited experimental data leads to uncertainty in the accuracy of empirical potentials.

Herein are report *ab initio* QM calculations (LMP2/cc-pVTZ(-f)) on the complexation energies of the AT-WC, GC-WC and AT-H nucleic acid base pairs. In addition to determining the minimum energy and geometry of these complexes, QM was used to calculate the strength of these interactions as a function of intermolecular distance. These potential energy curves were used to determine distance dependent hydrogen bond functions (described as Morse potentials). We find that force fields (FF) incorporating these hydrogen bond potentials in conjunction with standard van der Waals and coulombic terms accurately reproduce the full QM potential curve. This differs from the approach of current standard FF (AMBER95,[22] CHARMM,[23] and OPLS[24]) which have been parameterized to reproduce only the energy and geometry at the bottom of the potential curve.

II. Methods

Calculations involving the nucleic acid base pairs were carried out with the PS-GVB software package.[25] Optimum geometries (HF/6-31G**) of the free bases 1-hydroxymethyl-cytosine, 9-hydroxymethyl-guanine, 1-hydroxymethyl-thymine, 9-hydroxymethyl-adenine, 1-methyl-thymine, and 9-methyl-adenine were determined while maintaining planar C_s symmetry. This constraint was also imposed for the optimization of the AT-WC, GC-WC and AT-H base pairs. For the unpaired bases containing an exocyclic amino group, the optimum geometry has been reported to be nonplanar.[26]

However, it has been shown that for the three base paired structures we examine, the optimum structure is planar,[12] and hence the errors due to the symmetry constraint are negligible.

Initially we obtained the optimum base pair geometries by gradient optimization (HF/6-31G**) while keeping internal base coordinates fixed in the optimum monomer conformation. These HF/6-31G** results proved to be inadequate and yielded geometries which did not match experimental structures. Full optimization with a more diffuse cc-pVTZ(-f) basis set greatly improved the results. For these calculations no symmetry constraints were used.

Energetics were obtained using local second order Moller-Plesset perturbation theory (LMP2)[27-30] with frozen core orbitals for geometries determined at the HF/cc-pVTZ(-f) level. The LMP2 method differs from canonical MP2 by limiting each occupied orbital's space of correlating virtual orbitals to be localized on the atoms of the occupied HF orbital. A Boys localization[31] of the HF reference wavefunction is used to determine the local MP2 reference wavefunction. This local correlation method has been reported to be essentially free from correlation basis set super-position error (BSSE) and offers significant computational savings.[30]

Starting from the optimum geometry, energies were calculated at the LMP2/cc-pVTZ(-f) level as the central N–H ::: N hydrogen bond length was systematically increased. This yielded a potential energy curve as a function of distance for each base pair. The HF BSSE correction was calculated for each geometry, using the counterpoise method of Boys and Bernardi.[32]

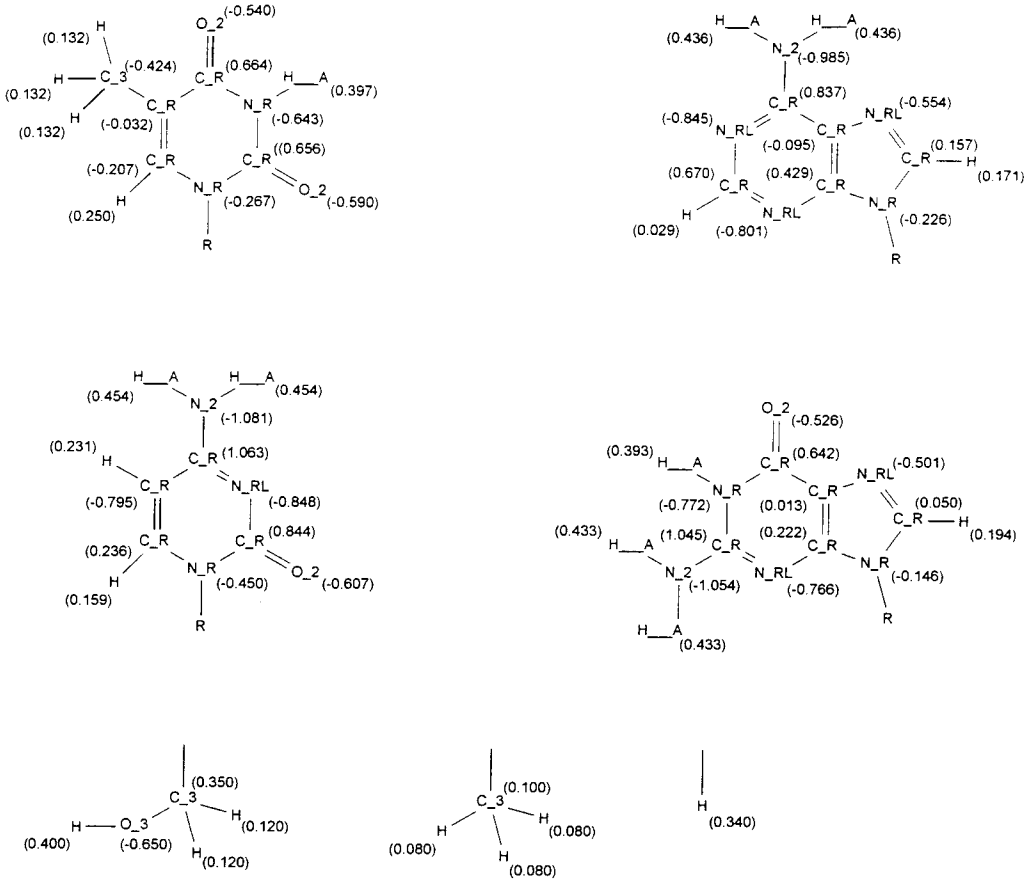


Figure 5.1. Diagram of the nucleic acid bases; thymine, adenine, cytosine, and guanine including force field atom types and partial charges. Charges were determined from LMP2/cc-pVTZ(-f) quantum mechanical calculations. The methyl and methoxy groups on the last row are common to all bases and are abbreviated as R in the base structures.

To simplify the calculations, we want to replace the deoxyribose sugar with either a methyl or hydroxymethyl substitution on the base. To establish which is best we carried out a calculation on 2'-deoxyadenosine, including the sugar explicitly. Comparing the potential derived charges for 2'-deoxyadenosine case with those of 9-methyl-adenine and 9-hydroxymethyl-adenine, we found that the hydroxymethyl substitution gave results closer to that with the complete sugar (see Figure 5.1). Since these hydrogen bond potentials are intended for parameterization of a force field (FF) which will include the sugar, the hydroxymethyl substituted bases were used for all subsequent calculations.

This hydroxymethyl substitution is expected to lead to a complexation energy similar to that of the full sugar because the charges for all atoms in close contact are quite similar. Experimental hydrogen bond energies have been measured for the methyl substituted base pairs. In order to compare our calculations directly with these experiments, we also examined the methyl substituted AT Hoogsteen base pair.

The complexation enthalpies for base pair dimerization have been measured experimentally using mass spectrometry.[20] To compare between these experimental values and theoretical calculations, several corrections must be made. The free bases undergo a small conformational change upon forming base paired complexes, resulting in a positive deformation energy associated with base pairing. Thus, the QM energy may be defined as:

$$\Delta E_{QM} = \Delta E_{HF} - BSSE_{HF} + E_{cor} + E_{def} \qquad (5.1)$$

where ΔE_{HF} is the HF energy, $BSSE_{HF}$ is the basis set superposition error, E_{cor} is the LMP2 correlation energy, and is E_{def} the deformation energy. Finally, ΔH_{300} may be determined from:

$$\Delta H_{300} = \Delta E_{QM} + \Delta ZPE + \Delta H_{0\rightarrow 300} \qquad (5.2)$$

in which ΔZPE is the change in zero-point energy between the free bases and the dimer, while $\Delta H_{0\rightarrow 300}$ is the temperature dependence of ΔH from 0 to 300 K. ΔZPE and $\Delta H_{0\rightarrow 300}$ were calculated from the vibrational frequencies obtained from the FF. The two terms have opposite sign with ΔZPE decreasing the value of ΔE while $\Delta H_{0\rightarrow 300}$ increases ΔE.

Atomic point charges were determined from the electrostatic potential (ESP) derived from the electron density distribution (constrained to reproduce the molecular monopole and dipole moments) calculated from the converged LMP2/cc-pVTZ(-f) wavefunctions (Figure 5.1).[33] MM calculations were performed with the POLYGRAF[34] software package using the Dreiding FF[35] with an exponential-6 van der Waals potential (5.3).

$$E_{\exp 6} = Ae^{-CR} - BR^{-6} \tag{5.3}$$

The procedure for determining the *ab initio* derived hydrogen bond potentials was as follows:

1. The total nonbond potential energy for the complex was determined for each geometry *excluding* the specific van der Waals contribution between hydrogen bonding atoms. This steeply attractive potential matches the hydrogen bond energy at large distances.

2. The difference between the "electrostatic" potential and the hydrogen bond curve determined from QM is fit with a simple Morse function (5.4).

$$E_{morse} = D_0\left(\chi^2 - 2\chi\right) \tag{5.4a}$$

$$\chi = \exp\left[-\frac{\xi}{2}\left(1-\frac{R}{R_0}\right)\right] \tag{5.4b}$$

This off-diagonal Morse function is specific for each pair of hydrogen bond donor and acceptor atoms and enables the MM calculations to reproduce energies and geometries obtained from QM over the full range of hydrogen bond distances.

Thus, the nonbond interactions involve superposing two-body van der Waals (E_{vdw}), hydrogen bond (E_{morse}) and electrostatic terms:

$$E_{nonbond} = E_{vdw} + E_{morse} + (332.0637)\frac{q_i q_j}{R_{ij}} \qquad (5.5)$$

Hydrogen bond curves for the FF were obtained analogous to the QM methods. The hydrogen bonded complexes were optimized while keeping the internal degrees of freedom fixed at the HF/6-31G** optimum geometry. The intermolecular distance was then increased and single point energies calculated. We chose to keep the internal base geometry fixed so as to compare to the nonbond potentials of the AMBER95,[22] CHARMM,[23] and OPLS[24] FF without correcting for valence distortions.

III. Results and Discussion

3.1. Quantum Mechanical Calculations for Nucleic Acid Base Pairs.

QM calculations on systems as large nucleic acid base pairs are still computationally intensive. Our initial approach was to carry out all calculations at the HF level with a 6-31G** basis set. Earlier optimizations reported for the 6-31G* basis set yielded hydrogen bond lengths which were an average of 0.1 Å too long.[18] We obtained similar results with the 6-31G** basis set. Using the cc-pVTZ(-f) basis set gave much better geometries when compared to experiment. HF/6-31G** energies were determined for all points and corrected for BSSE. These energies were found to underestimate the ΔH of dimerization. To improve upon both the optimum geometries and energies obtained from QM, we used a larger basis set and included electron correlation for the single-point energy calculations (LMP2/cc-pVTZ(-f)).

The optimum hydrogen bond lengths for the AT-WC, GC-WC and AT-H base pairs are reported in Table 5.1. The HF/6-31G** geometries lead to hydrogen bond lengths that are consistently too long by 0.15 Å. Increasing the basis set size [cc-pVTZ(-f)] gave much better results with an average difference of only 0.05 Å from the experimental geometry (some differences may be expected between the gas phase geometries of isolated bases and the crystal structure of a two base-pair dimer).

As discussed in Section 2, hydroxymethyl was used to replace the full ribose or deoxyribose sugar. For the AT-H base pair, both the methyl and hydroxymethyl substitutions were examined. The effect this substitution has on the HF/6-31G** geometry is worth noting. The optimum distances for N7:::(H)N3 and N6:::(H) O4 hydrogen bonds in the hydroxymethyl case are 3.09 Å and 2.98 Å respectively whereas in the methyl case they are 2.96 Å and 3.12 Å respectively. This marks a reversal in the relative lengths for both hydrogen bonds, yet this effect is not observed for the HF/cc-pVTZ optimized structures. These observations further support the need for a large basis set to correctly describe nonbonded interactions.

Table 5.1. Hydrogen Bond Lengths of the DNA Base Pairs (Å).

Atom Pair	HF/3-21G[a]	HF/6-31G**[b]	HF/cc-pVTZ[c]	X-ray[d]
		GC Watson-Crick		
O6–N4	2.77	2.91	2.83	2.91
N1–N3	2.91	3.08	2.95	2.95
N2-O2	2.86	3.11	2.92	2.86
		AT Watson-Crick		
N6–O4	2.96	3.17	3.06	2.95
N1–N3	2.78	3.02	2.92	2.82
		AT Hoogsteen		
N6–O4	2.99	2.98	2.91	2.86
N7–N3	2.75	3.09	2.95	2.93

[a]As reported by Gould et al.[18] [b]Internal geometry held fixed. [c]Full geometry optimization with no constraints. [d]From experimental X-ray crystallography data.[36]

The individual bases were also optimized (HF/cc-pVTZ(-f)) so as to determine internal strain energies associated with base pairing. These geometries were found to compare favorably with crystal structures of the free bases. Table 5.2 lists the RMS difference in coordinates, bond lengths, and angles for each base. The average RMS difference in coordinates for the pyrimidine bases was 0.02 Å and for the purine bases, 0.05 Å.

Table 5.2. RMS Comparsion Between *Ab initio* (HF/cc-pVTZ(-f)) and Experimental Nucleic Acid Base Structures.

Base	Coordinates (Å)	Bonds (Å)	Angles (deg)
1-methylcytosine	0.022	0.024	1.448
1-methylthymine	0.020	0.200	1.433
9-methyladenine	0.049	0.040	4.121
9-methylguanine	0.049	0.032	3.573

Table 5.3. DNA Base Pair Energies for HF Optimized Geometries (kcal/mol).

Method	GC Watson-Crick	AT Watson-Crick	AT Hoogsteen
HF/6-31G**	-23.8	-12.0	-11.5
HF/6-31G**(BSSE)	-21.4	-9.8	-9.4
HF/cc-pVTZ(-f)	-24.2	-10.3	-10.5
HF/cc-pVTZ(-f)(BSSE)	-22.8	-9.2	-9.4
LMP2/cc-pVTZ(-f)	-24.9	-12.0	-12.6
LMP2/cc-pVTZ(-f)(BSSE)	-23.8	-11.1	-11.7
ΔE_{QM}	-22.4	-10.8	-11.0
ΔZPE	1.42	0.84	0.77
$\Delta H_{1\rightarrow 300}$	-0.17	-0.23	-0.25
ΔH_{300}	-21.2	-10.2	-10.5
$\Delta H_{\exp}$	-21.0	–	-12.1

HF/6-31G** complexation energies were calculated for the optimized base pairs and corrected for BSSE. The base pair energies are listed in Table 5.3 and found to be substantially weaker than the reported experimental ΔH, even before taking ΔZPE corrections into account. The BSSE was found to be 2.1 and 2.4 kcal/mol for the AT-WC and GC-WC base pairs respectively. This compares favorably with the reported results of Gould et al.[18] in which the effects of BSSE corrections are discussed in detail for various methods and basis sets.

Using a cc-pVTZ(-f) basis set improves upon these results. However the inclusion of correlation energies is necessary to adequately describe the system. The HF/cc-pVTZ(-f) BSSE is reduced to 1.1 and 1.3 kcal/mol for the AT-WC and GC-WC base pairs respectively. With the LMP2 method, there is no additional correction to the

dispersion energy due to BSSE.[30] This offers a distinct advantage over standard MP2 methods for which BSSE corrections are difficult to estimate and computationally intensive.

As discussed earlier, LMP2/cc-pVTZ(-f) energies can be compared to experimental complexation enthalpies only after a number of factors have been taken into account. Table 5.3 lists ΔE_{QM}, ΔZPE, and $\Delta H_{1\rightarrow 300}$, the sum of which ($\Delta H_{300}$) may be compared directly to experimentally determined enthalpies ($\Delta H_{\exp}$). For the GC-WC base pair, the calculated value of -21.2 kcal/mol is in excellent agreement with the experimental value of -21.0 kcal/mol. The calculated hydrogen bond energies of the AT-H base pair is considerably lower than the experimental results, -10.5 versus -13.0 kcal/mol. These energies differ from those of Gould et al.[18] who found good agreement with the AT-H base pair but reported a GC-WC base pair energy 4.4 kcal/mol too negative when compared with experiment. The energies reported by Sponer et al.[17] are within 1 kcal/mol of the energies we report.

We examined the AT-H complexation enthalpies in greater detail to understand the discrepancy between the theoretical and experimental energies. A methyl versus hydroxymethyl substitution for the full sugar makes no difference in the complexation energy. The HF/cc-pVTZ(-f) energy for methyl-AT-H is -9.45 kcal/mol compared to -9.42 kcal/mol for hydroxymethyl substitution. This is also in agreement with the experimental report that ΔH is independent of methylation.[20]

The AT base pair may assume one of four conformations; Watson-Crick, reverse Watson-Crick, Hoogsteen or reverse Hoogsteen, all of which have been reported to have

energies within 1 kcal/mol of each other.[17] Experimental gas phase complexation enthalpies were not corrected for contributions from all four possible orientations. At 60° C, this leads to an overestimation of ΔH by 0.9 kcal/mol. Thus the experimental AT complexation enthalpy should be 12.1 kcal/mol. While this correction reduces the difference between our theoretical energies and experiments, the discrepancy of 1.6 kcal/mol is still significant.

Although stabilization between base pairs is dominated by electrostatics, dispersion energies contribute significantly more to the AT-H and AT-WC base pairs versus GC-WC. For the GC-WC base pair dispersion energies account for only 3% of the total complexation energy whereas for the AT-WC and AT-H pairs this contribution is increased to 15% and 18% respectively. It may be the case that an LMP2 geometry optimization of the AT and GC base pairs would be found to preferentially stabilize the AT pair.

Given an optimum geometry for each of the Watson-Crick and Hoogsteen base pairs, it is possible to determine the complexation energy as a function of distance. For each base pair, the distance between atoms of the central (N:::H–N) hydrogen bond was increased and single-point energy calculations performed followed by BSSE corrections. Complete HF/6-31G** curves were obtained for all the base pairs. The AT-WC and GC-WC curves are shown in Figure 5.2. As discussed above, these hydrogen bond lengths are slightly too long and the energies too weak to be used to fit a FF.

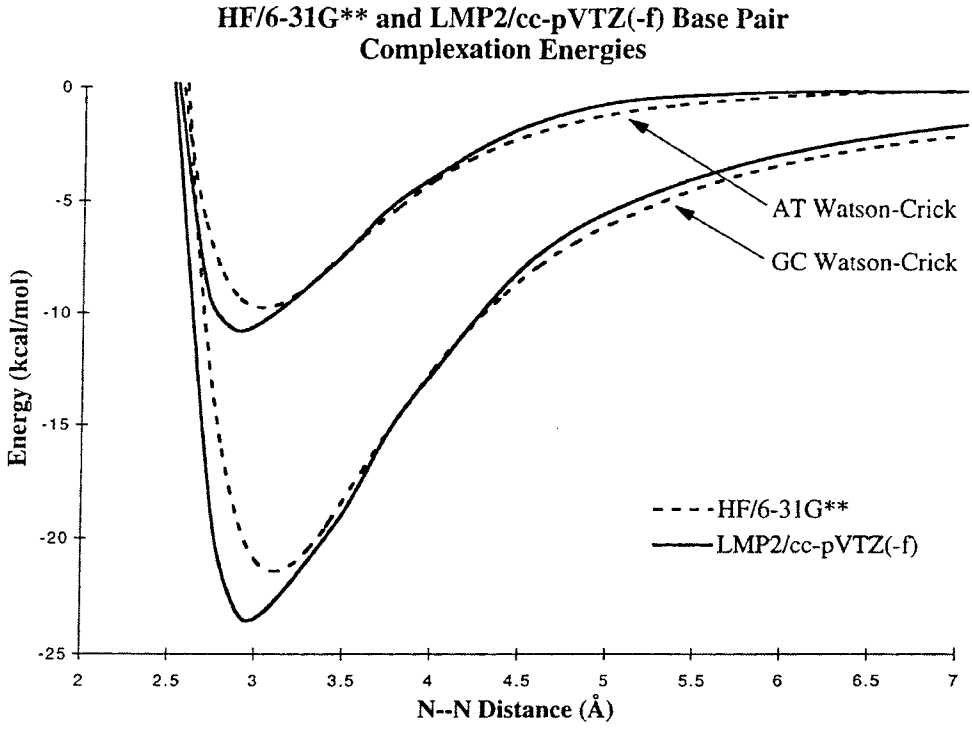


Figure 5.2. Potential energy curves for the AT-WC and GC-WC base pairs. The base pair separation is defined as the distance between heteroatoms of the central hydrogen bond, atoms N1 of the pyrimidines and N3 of the purines. Note the overestimation of the complexation by the Hartree-Fock calculation at distances greater than 4 Å.

For each base pair, six single-point LMP2/cc-pVTZ energies were calculated: One at the HF/cc-pVTZ optimum geometry and five additional points up to a hydrogen bond length of 5 Å. A final distance of 5 Å between heteroatoms was chosen as the dispersion energies are nearly zero and the total potential energy largely determined by electrostatics. For each geometry, the component and total energies are listed in Tables 5.4-5.6 for the GC-WC, AT-WC and AT-H base pairs respectively. To compare with the HF/6-31G** results, the points for the AT-WC and GC-WC pair are plotted in Figure 5.2.

Table 5.4. LMP2/cc-pVTZ(-f) Total and Component Energies (kcal/mol) for the GC-WC Base Pair at Variable N1(H):::N3 Distances (Å).

N1(H):::N3	E_{HF}	$BSSE_{HF}$	E_{cor}	E_{LMP2}
2.50	+5.053	1.660	-4.884	+1.829
2.75	-18.473	1.226	-2.328	-19.575
2.95	-24.172	1.347	-0.737	-23.562
3.45	-20.611	0.718	+0.530	-19.363
3.95	-14.516	0.493	+0.957	-13.066
5.00	-6.840	0.227	+1.176	-5.437

Table 5.5. LMP2/cc-pVTZ(-f) Total and Component Energies (kcal/mol) for the AT-WC Base Pair at Variable N1(H):::N3 Distances (Å).

N1(H):::N3	E_{HF}	$BSSE_{HF}$	E_{cor}	E_{LMP2}
2.50	+1.862	1.539	-4.577	-1.176
2.72	-7.420	1.316	-2.876	-8.980
2.92	-10.347	1.157	-1.630	-10.820
3.42	-8.054	0.947	-0.863	-7.970
3.92	-4.872	0.705	-0.212	-4.379
5.00	-1.360	0.304	+0.339	-0.717

Table 5.6. LMP2/cc-pVTZ(-f) Total and Component Energies (kcal/mol) for the AT-H Base Pair at Variable N7:::(H)N3 Distances (Å).

N7:::(H)N3	E_{HF}	$BSSE_{HF}$	E_{cor}	E_{LMP2}
2.50	+3.119	1.469	-6.789	-2.201
2.75	-6.975	1.287	-5.111	-10.799
2.95	-10.750	1.327	-1.884	-11.307
3.45	-8.741	0.958	-0.775	-8.558
3.95	-5.345	0.675	-0.059	-4.729
5.00	-1.561	0.332	+0.469	-0.760

Examination of the correlation energies reveals a positive contribution to the total energy at intermediate distances. Calculations on the CC and GG dimers, the only other DNA systems for which distance dependent hydrogen bond potentials have been reported, show the same tendencies.[17] The LMP2 electron correlation energies may be decomposed as in (5.6)[37,38]

$$E_{cor}^{LMP2} = E_{disp} + \Delta E_{rep} \qquad (5.6)$$

where E_{disp} is the dispersion energy and is always negative, while ΔE_{rep} represents the remaining repulsive terms which include the intrasystem correlation correction to the electrostatic effect, the induction correlation, and exchange terms. MP2 calculations have been reported for simple hydride dimers such as $(HF)_2$.[37] These calculations indicate that ΔE_{rep} is slightly smaller in magnitude than E_{disp}, with the two terms largely canceling each other leading to a small total correlation correction. A fine balance is maintained between these components which individually are sensitive to the intermolecular separation, yet vary slowly once summed. At intermediate to long hydrogen bond distances, where E_{disp} becomes negligible, ΔE_{rep} still contributes slightly to the total energy resulting in a positive ΔE_{cor}. Thus the HF calculation overestimates the electrostatic energies and has a lower relative energy for these points. Similarly, a FF which incorporates charges determined from isolated monomers will overestimate the electrostatic energy at points where there is a positive change in electrostatic energy, with decreasing distance, due to electron correlation.

To examine whether such positive correlation energies observed using MP2 or LMP2 methods are general, we carried out extensive calculations on a water dimer

system. The optimum dimer geometry used in these calculations was the same as that determined by Saebo et al.[28] The intermolecular distance was then varied and for each geometry the LMP2/cc-pVTZ(-f) energies compared to MP2, MP3 and (SDTQ)MP4/cc-pVTZ energies after full BSSE corrections. The correlation energies as a function of distance are listed in Table 5.7 and the total energies plotted in Figure 5.3. All levels of theory show the same distance dependent behavior with a positive correlation energy at intermediate intermolecular distances. The complexation energy determined with LMP2 (4.67 kcal/mol) is within the range of experiment 5.4 ±0.7 kcal/mol[39] (MP2 with augmented cc-pVTZ basis set leads to 5.58 kcal/mol and 4.65 kcal/mol when corrected for BSSE and monomer relaxation).[40]

Table 5.7. Dependence of Correlation Energies (kcal/mol) for the H_2O Dimer on O:::O Distances (Å) and Level of Theory, Using the cc-pVTZ Basis Set.

O:::O	MP2	MP3	MP4	LMP2[a]
2.50	-2.018	-1.834	-1.794	-1.917
2.66	-1.481	-1.363	-1.337	-1.659
2.97	-0.765	-0.712	-0.695	-1.055
4.00	-0.036	-0.021	+0.001	-0.119
5.00	+0.036	+0.037	+0.056	+0.005
6.00	+0.031	+0.030	+0.042	+0.030
7.00	+0.021	+0.020	+0.028	+0.023

[a]All calculations use a frozen core approximation. The optimum geometry has an O:::O distance of 2.97 Å. LMP2 calculations were carried out with the cc-pVTZ(-f) basis set, as discussed in the Methods section.

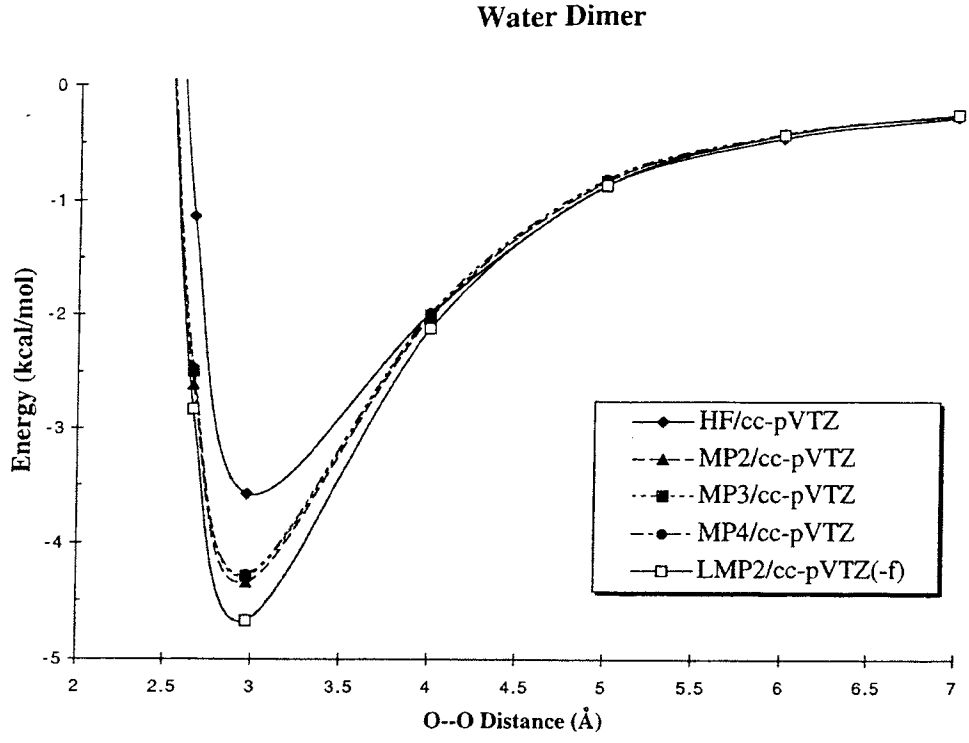


Figure 5.3. Potential energy curves for the H_2O dimer at various levels of theory. The LMP2/cc-pVTZ(-f) minimum energy is closest to the experimental association energy of 5.4 ±0.7 kcal/mol. In addition, all calculations which include electron correlation effects show the same behavior at intermediate distances, with association energies which are weaker than the Hartree-Fock calculation predicts.

3.2 Force Field Parameterization.

Using the curves calculated from QM, we sought a simple FF description for use in molecular mechanics (MM) calculations which could correctly describe both the inner wall of the minimum energy well and the curvature as the hydrogen bond length is increased. Examination of the hydrogen bond potentials determined from QM reveals that at large distances the energies are dominated by electrostatics. We find that using charges derived from QM for all atoms, it is only necessary to define a repulsive Morse potential between the atom pairs directly involved in hydrogen bonding to reproduce the

QM curves. This method has been used successfully in the parameterization of a specialized FF for nylon.[41]

Table 5.8. Morse Potential Parameters for the DNA Base Pairs.

Atom Pair	Atom Types	R_0	D_0	ζ
MSC1 FF GC:				
O6-N4/N2-O2	O_2–H___A	2.48	0.987	7.55
N1-N3	N_RL–H___A	2.32	0.711	8.83
MSC1 FF AT:				
N6-O4	O_2–H___A	2.58	0.241	9.92
N1-N3	N_RL–H___A	2.62	0.192	10.46
MSC2 FF (all):				
	O_2–H___A	2.55	0.200	9.00
	N_RL–H___A	2.70	0.200	9.40

Parameterization of the Morse potentials is closely related to the van der Waals description of the FF used. For these calculations we have used the Dreiding FF with the exponential-6 (Exp-6) van der Waals (vdW) option. All vdW interactions are included explicitly except the specific vdW interaction between two atoms which form a hydrogen bond. This energy is set to zero. For each geometry, a MM energy is calculated with nonbond energies comprised of an Exp-6 vdW term and a standard coulombic function. The difference between the QM energy and the FF Exp-6 + Coulomb energy may be fit with a Morse potential. The Morse potential is specific for the selected pair of atoms forming the hydrogen bond and replaces the normal Exp-6 vdW energy calculated from combination rules.

We report two sets of Morse parameters for the MSC FF which differ in accuracy and generality. The best parameters (MSC1 FF) require different potentials between the AT (-WC and -H) base pairs and the GC-WC base pair. While this gives an excellent fit to the QM data, it is not as useful for studies of uncommon base pair geometries or non-natural bases. A satisfactory fit using a global set of parameters (MSC2 FF) for all the base pairs was also obtained. The MSC2 FF leads to systematic differences from the QM data, probably because of cooperative effects in hydrogen bonding. The AT-WC and AT-H base pair energies are too strong by 0.6 and 0.4 kcal/mol, respectively, and the GC-WC energy is underestimated by the 1.1 kcal/mol. The parameters for both MSC1 FF and MSC2 FF are listed in Table 5.8. The optimum energies and geometries obtained using these new FF are listed with the QM results in Tables 5.9 and 5.10.

Table 5.9. Optimum Hydrogen Bond Lengths of the DNA Base Pairs (Å) as Determined by Molecular Mechanics with the AMBER95.1*, CHARMM*, and MSC Force Fields.

Atom Pair	AMBER95.1*	CHARMM*	MSC1	MSC2	X-ray[d]
			GC Watson-Crick		
O6–N4	2.91	2.88	2.94	2.84	2.91
N1–N3	2.97	2.95	3.02	2.91	2.95
N2-O2	2.89	2.88	2.95	2.84	2.86
			AT Watson-Crick		
N6–O4	2.99	2.97	3.02	2.87	2.95
N1–N3	2.89	3.01	2.93	2.88	2.82
			AT Hoogsteen		
N6–O4	3.08	3.03	3.07	2.91	2.86
N7–N3	2.92	2.93	2.91	2.86	2.93

Table 5.10. Optimum Complexation Energies for the DNA Base Pairs as Determined by Molecular Mechanics with the AMBER95.1*, CHARMM*, and MSC Force Fields. Internal base geometries were defined by the optimum HF/6-31G** structures and held fixed during optimization by each force field.

Method	GC Watson-Crick	AT Watson-Crick	AT Hoogsteen
AMBER95.1*	-23.5	-11.0	-11.4
CHARMM*	-23.1	-11.3	-11.4
MSC1	-23.7	-10.9	-12.0
MSC2	-22.7	-11.7	-12.7
LMP2/cc-pVTZ(-f)	-23.8	-11.1	-11.7

3.3. Molecular Mechanics Force Field Evaluations.

Hydrogen bonding has been addressed in a number of ways in various FF. Earlier versions of AMBER,[42] CHARMM, and Dreiding used a Lennard-Jones (LJ) 12-10 potential for hydrogen bonding atom pairs. The current releases of AMBER (AMBER95[22]), CHARMM[23] and OPLS[24] all use a LJ 12-6 potential, but differ in the derivation of atomic charges and the specific LJ parameters. AMBER95 and CHARMM both use QM derived charges. AMBER95 uses the restrained electrostatic potential (RESP)[43] method, and alters the LJ parameters of the hydrogen bonding atoms to fit experimental energies and geometries. The OPLS FF uses general LJ parameters based on element type only and uses a charge fitting scheme which reproduces the energy and geometries of water-base complexes as determined from *ab initio* QM. I find that these standard MM FF lead to an inaccurate description of the curvature of the hydrogen bond potential and usually do not predict correct complexation energies and geometries.

For each FF (including MSC1 FF) I have evaluated the base pair complexation energies as a function of distance. The internal geometry of each base was held fixed in the HF/6-31G** optimum conformation. Intermolecular geometries were determined by gradient optimization of the nonbond energies, followed by single point energy calculations at fixed intermolecular distances. Two sets of charges are available for the AMBER95 FF: (i) AMBER95.0 denotes calculations using the original RESP charge scheme (listed in the supplementary materials of reference [44]), for which the complexation energies are reported in the FF paper, reference [22], and (ii) AMBER95.1

denotes the charge scheme released with the full FF, which do not match the energies in reference 22.

The results of this analysis, plotted in Figures 5.4a-c, reveal that FF which employ *ab initio* QM derived charges are more successful at reproducing the QM energies. For all three cases examined, AMBER95.0 and AMBER95.1 overestimate the hydrogen bond energies by 3-5 kcal/mol. CHARMM describes GC-WC to within 1 kcal/mol, but overestimates the AT-WC and AT-H energies by 3 kcal/mol and 2.5 kcal/mol respectively. The OPLS FF underestimates the complexation energies by 1-2 kcal/mol and is considerably worse at reproducing the QM energies away from the optimum geometry. All FF incorporating a LJ6-12 description suffer from the same shortcoming intrinsic to the form of the potential. Specifically, the attractive portion of the potential is too steep, creating a minimum energy well which is too narrow. In comparison, the Morse potential used with the MSC FF leads to a softer inner wall and a broader minimum energy well which better describes the true curve.

In spite of this improvement, the MSC1 Morse potential does not reproduce the exact QM energies for the AT-H and AT-WC pairs at longer hydrogen bond distances (see Figures 5.5b-c). As discussed earlier, inclusion of electron correlation decreases the electrostatic energy leading to a weaker bonding at the longer distances. An electrostatic model based on independent monomers correctly describes the system around the minimum, but overestimates the stabilization energy at longer distances. For the GC-WC pair, where dispersion contributes less to the total bonding, this effect is negligible, but for AT complexes, which depend more strongly upon dispersion, it is significant.

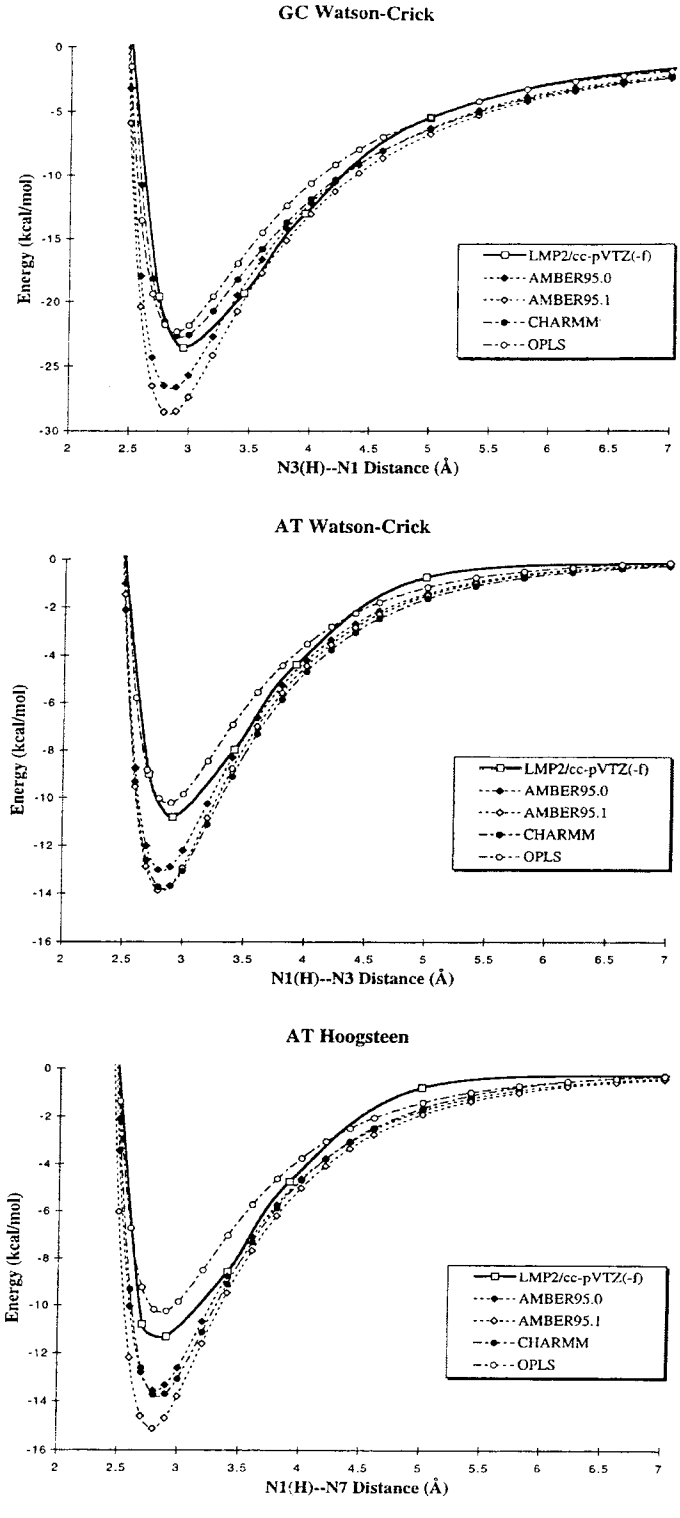


Figure 5.4. Comparison between potential energy curves for the DNA base pairs (a) GC-WC, (b) AT-WC and (c) AT-H as determined with different molecular mechanics force fields and quantum mechanical results.

In addition to determining optimum off-diagonal nonbond parameters for the MSC FF, we carried out similar optimizations for CHARMM and AMBER95.1. For both FF we report the optimum LJ12-6 parameters needed to define the vdW potentials between the heteroatoms involved in hydrogen bonding. These off-diagonal van der Waals terms replace the standard R_e and D_e determined using combination rules. No charges or valence terms have been altered. The parameters are listed in Table 5.11. We denote calculations using these modified terms as CHARMM* and AMBER95*. The new optimum geometries and energies are compared in Tables 5.9 and 5.10, respectively. Full potential curves are plotted in Figures 5.5a-c.

Table 5.11. LJ12-6 Parameters for the DNA Base Pairs Used to Correct the AMBER95.1 and CHARMM Force Fields.

Atom Types	R_0	D_0
	AMBER95.1	
NA-NC	4.395	0.020
N2-O	4.410	0.017
NA-NB	4.500	0.020
	CHARMM	
NN3-NN2G	3.38	0.550
NN3A-NN2U	4.01	0.130
NN4-NN2U	4.15	0.060
NN1-ON1	3.59	0.180

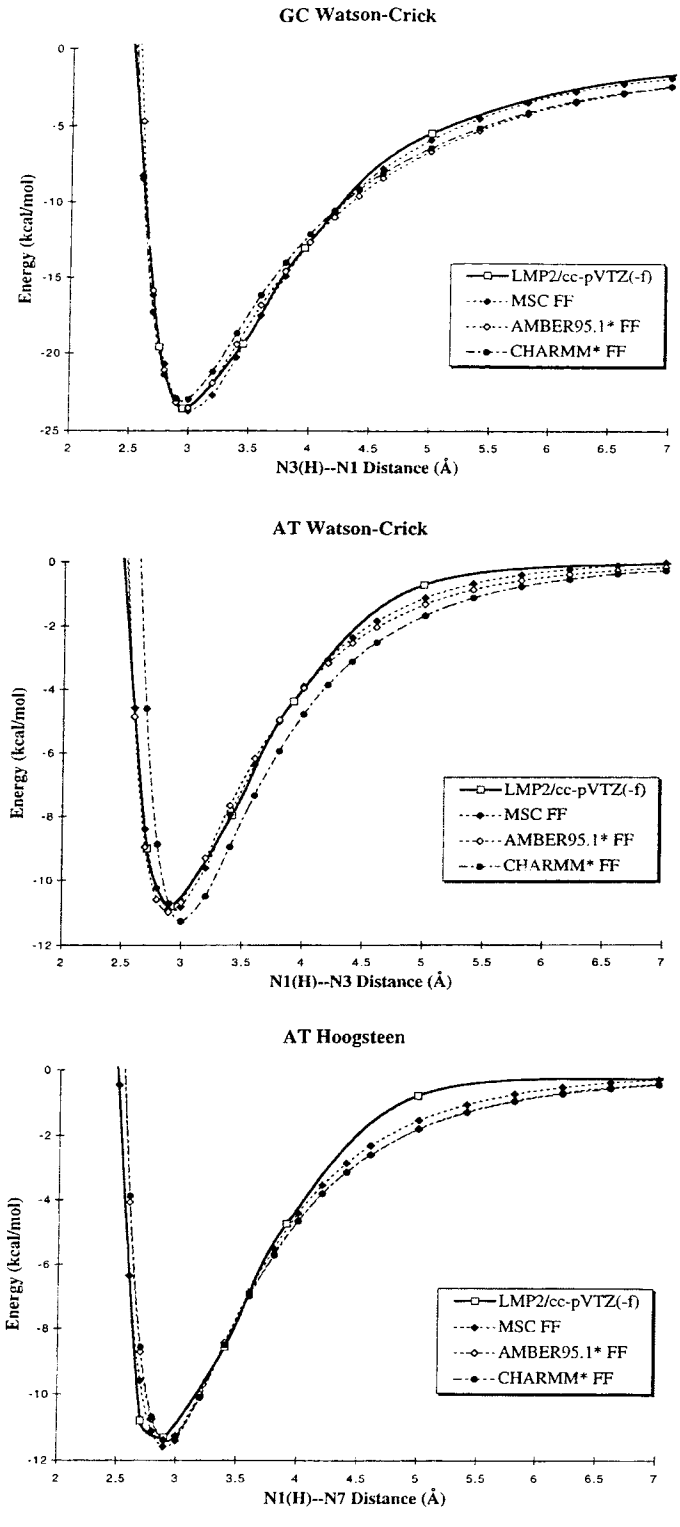


Figure 5.5. Comparison between potential energy curves for the DNA base pairs (a) GC-WC, (b) AT-WC and (c) AT-H as determined by *ab initio* quantum mechanics, the MSC FF and modified versions of AMBER95 and CHARMM.

Incorporating these special off-diagonal terms greatly improves both FF. For AMBER95.1* (which uses the charge scheme distributed with AMBER95), the minimum geometry and energy for all three base pairs could be adjusted to reproduce the QM results with quite satisfactory distance dependence. Due to the additional number of atom types used by the CHARMM FF, all three base pair complexes could be fit independently; however, this was not sufficient to attain a satisfactory fit for AT-WC. Here the final optimum hydrogen bond lengths are too long by 0.1 and the energies too strong by 0.4 kcal/mol. As with AMBER95.1*, both the GC-WC and AT-H complexes could be fit and yielded similar distance dependent curves shown in Figures 5.5a-c. These off-diagonal van der Waals parameters are simple to include during MM and dynamics simulations and greatly improve the accuracy of these two popular FF.

The stabilization energies of 26 possible geometries for the DNA base pairs, in which a minimum of two hydrogen bonds are formed, has been evaluated using *ab initio* QM (MP2/6-31G**).[17] These energies have been compared to those obtained from empirical potentials such as AMBER95, CHARMM and OPLS FF.[45] A linear regression analysis of each of these potentials compared to the MP2 results simplifies the task of assessing which FF best reproduce the QM energies. Such an analysis has been previously reported[45] for the potentials listed above and we have now extended this work to include the general MSC2 FF.

The MP2 and MSC2 FF energies of each complex are plotted in Figure 5.6a and the linear regression plot is shown in Figure 5.6b. The MSC2 FF energies correlate very well with the QM results. The specific linear regression parameters for each of the FF are listed in Table 5.12. The correlation coefficient is an overall measure of how well the

MM potentials and QM methods are correlated, with a value closest to 1.0 representing the highest possible correlation. In addition, the standard deviation reflects the difference in energies between the two methods for each geometry. The MSC2 FF has the highest correlation coefficient (0.98) of the FF examined and the smallest standard deviation (0.83 kcal/mol). Other parameters may also be examined such as the intercept (A) or slope (B) of the linear fit which should be close to 0.0 and 1.0 respectively. Finally, MCS2 FF leads to the smallest average absolute error.

Table 5.12. Linear Regression Analysis of Complexation Energies for 26 Hydrogen Bonded DNA Base Pair Geometries Using the AMBER95.1, CHARMM, and MSC Force Fields Compared to *Ab Initio* Energies at the MP2/6-31G** Level.[a]

	MSC2 FF	AMBER95	CHARMM	OPLS
R[c]	0.98	0.96	0.93	0.96
SD[d]	0.83	1.06	1.60	1.11
A	0.60	0.35	0.66	-1.34
B	0.98	0.98	0.96	0.92
AAE[e]	0.7	1.0	1.0	2.4

[a]MP2/6-31G** energies reported by Sponer et al.[17] Linear regression analysis reported by Hobza et al.[45] [c]Correlation coefficient. [d]Standard deviation (kcal/mol). [e]Average absolute error (kcal/mol).

$$AAE = \frac{1}{26}\sum_{i}^{26}\left|\Delta E^{MP2} - \Delta E^{FF}\right|$$

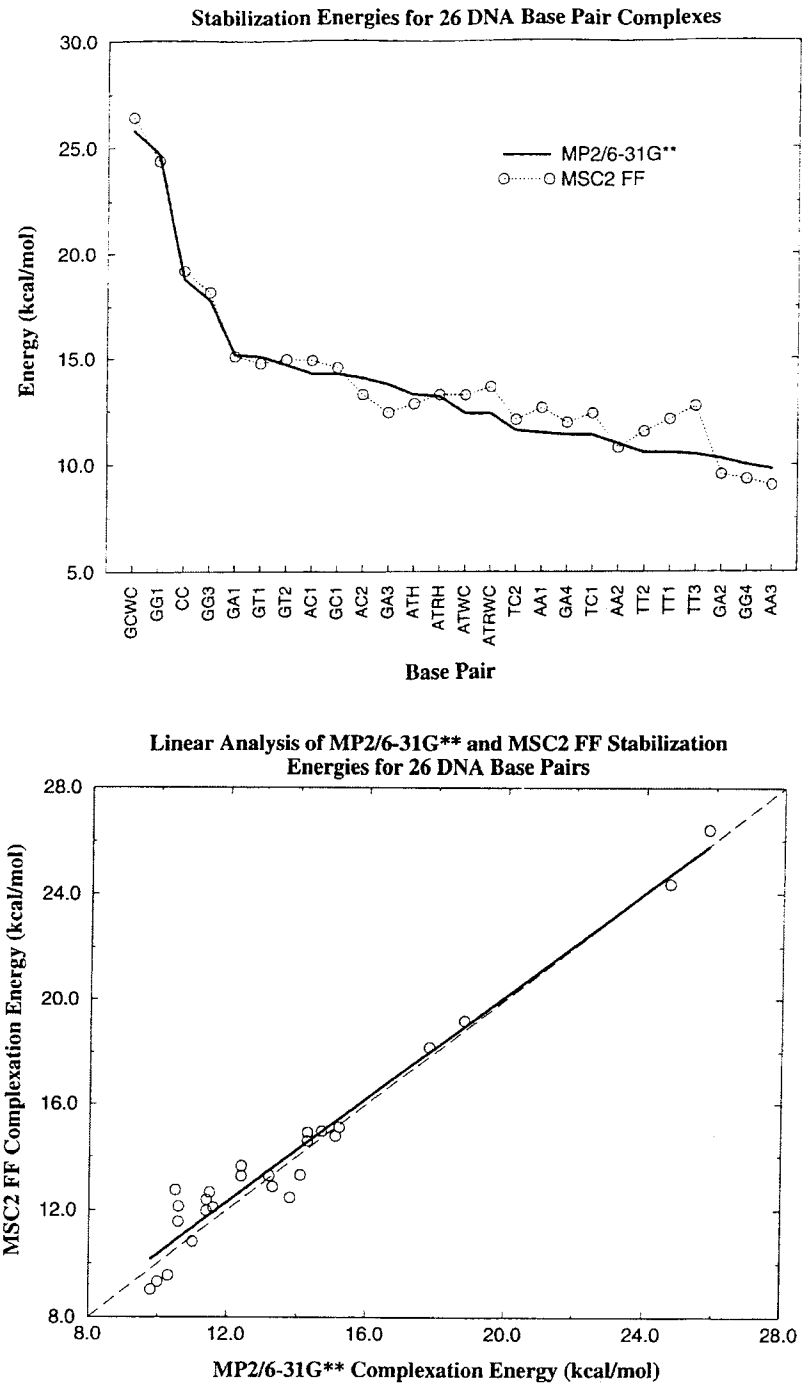


Figure 5.6. (a) Complexation energies for 26 possible base pair conformations determined with *ab initio* quantum mechanics (MP2/6-31G**) and the general MSC2 FF. See reference [17] for specific geometries. (b) For each complex, the MSC2 FF energies are plotted against the MP2 energies. The dashed line represents the ideal case in which the two energies are equal. The solid line is the least squares best fit through all points. See Table 5.12 for linear regression analysis.

Since none of the MP2 energies were used to parameterize the MSC2 FF, this represents an independent test of both the method of parameterization and the quality of LMP2 complexation energy calculations compared to the MP2 method. It is evident that a generalized hydrogen bond potential derived from the full LMP2 potential energy surfaces of three base pair geometries is sufficient to accurately describe all possible base pair interactions.

IV. Conclusion

The distance dependent hydrogen bond energies of the AT-WC, GC-WC and AT-H base pairs were examined using *ab initio* QM (LMP2/cc-pVTZ(-f)//HF/cc-pVTZ(-f)). Using these energies as a standard, an alternative hydrogen bond potential for use in MM and dynamics simulations is presented. This new FF incorporates a Morse potential to describe the nonbond energies between hydrogen bonding atoms and reproduces the QM energies for a range of geometries. In addition to parameterizing the new FF, we examined the accuracy of two commonly used FF, AMBER95 and CHARMM95. Neither FF was found to correctly describe the optimum geometry or energy of the base paired complexes. The accuracy of these FF was greatly improved by replacing the LJ12-6 potential between the heteroatoms directly involved in hydrogen bonding with an off-diagonal LJ12-6 potential tuned for the base pairs. The new parameters and the resulting complex energies and geometries are reported.

These *ab initio* calculations and analytical hydrogen bond potential energy functions are part of an ongoing effort to develop general first principles FF for biological systems. The goal of this work is to provide a relatively simple set of functions which

have been fit to reproduce *ab initio* calculations on small model compounds. Parameters thus defined are general and may be easily extended to new systems including non-natural amino acids and nucleic acid bases. Such a new generation FF may be extended using QM theory and is thus independent of experimental parameters.

Acknowledgments. This research was funded by DOE-BCTR (DE-FG36-93CH105 81). The facilities of the MSC are also supported by grants from NSF (CHE 95-22179 and ASC 92-100368), Chevron Petroleum Technology Co., Aramco, Asahi Chemical, Owens-Corning, Chevron Chemical Company, Asahi Glass, Chevron Research and Technology Co., Hercules, BP Chemical, and Beckman Institute. Some calculations were carried out at the NSF San Diego Supercomputer Center (SDSC).

V. References

(1) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117-135.

(2) Latajka, Z.; Scheiner, S. *J. Chem. Phys.* **1984**, *81*, 2713-2716.

(3) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comp. Chem.* **1983**, *4*, 294-301.

(4) Spitznagel, G. W.; Clark, T.; Chandrasekhar, J.; Schleyer, P. V. R. *J. Comp. Chem.* **1982**, *3*, 363-371.

(5) Delbene, J. E.; Shavitt, I. *J. Mol. Struct. (Theochem)* **1994**, *113*, 27-34.

(6) Delbene, J. E. *J. Phys. Chem.* **1993**, *97*, 107-110.

(7) Delbene, J. E.; Shavitt, I. *Int. J. Quantum Chem.* **1989**, *97*, 445-452.

(8) Delbene, J. E. *Int. J. Quantum Chem.* **1987**, *97*, 27-35.

(9) Delbene, J. E. *J. Comp. Chem.* **1987**, *8*, 810-815.

(10) Delbene, J. E. *J. Chem. Phys.* **1987**, *86*, 2110-2113.

(11) Szczesniak, M. M.; Latajka, Z.; Scheiner, S. *J. Mol. Struct. (Theochem)* **1986**, *28*, 179-188.

(12) Sponer, J.; Florian, J.; Hobza, P.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827-833.

(13) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590-5596.

(14) Sponer, J.; Hobza, P. *Chemical Physics* **1996**, *204*, 365-372.

(15) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Comp. Chem.* **1996**, *17*, 841-850.

(16) Sponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695-706.

(17) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965-1974.

(18) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493-2499.

(19) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, *118*, 7574-7577.

(20) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149-1170.

(21) Mayevsky, A. A.; Sukhorukov, B. I. *Nuc. Ac. Res.* **1980**, *8*, 3029.

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(23) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(24) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810-2819.

(25) Ringnalda, M. N.; Langlois, J.-M.; Greeley, B. H.; Murphy, R. B.; Russo, T. V.; Cortis, C.; Muller, R. P.; Marten, B.; Donnelly, R. E.; Mainz, D. T.; Wright, J. R.; Pollard, W. T.; Cao, Y.; Won, Y.; Miller, G. H.; Goddard, W. A., III; Friesner, R. A. Jaguar 3.0 from Schrödinger, Inc. located in Portland, OR.

(26) Sponer, J.; Burcl, R.; Hobza, P. *J. Biomol. Struct. Dyn.* **1994**, *11*, 1357-1376.

(27) Saebo, S.; Pulay, P. *Chem. Phys. Lett.* **1986**, *131*, 384-388.

(28) Saebo, S.; Pulay, P. *J. Chem. Phys.* **1988**, *88*, 1884-1890.

(29) Saebo, S.; Pulay, P. *Ann. Rev. Phys. Chem.* **1993**, *44*, 213-236.

(30) Saebo, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170-2175.

(31) Foster, J. M.; Boys, S. F. *Rev. Mod. Phys.* **1960**, *32*, 300.

(32) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *25*, 553.

(33) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.

(34) POLYGRAF MSC Modified Version from Molecular Simulations, Inc., located in San Diego, CA.

(35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(36) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(37) Latajka, Z. *J. Mol. Struct. (Theochem)* **1991**, *83*, 245-260.

(38) Latajka, Z. *J. Mol. Struct. (Theochem)* **1992**, *85*, 225-241.

(39) Curtis, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703.

(40) Xantheas, S. S.; Dunning, T. H. *J. Chem. Phys.* **1993**, *99*, 8774-8792.

(41) Dasgupta, S.; Hammond, W. B.; Goddard, W. A. *J. Am. Chem. Soc.* **1996**, *118*, 12291-12301.

(42) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230-252.

(43) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comp. Chem.* **1995**, *16*, 1357-1377.

(44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631.

(45) Hobza, P.; Hubalek, F.; Kabelac, M.; Mejzlik, P.; Sponer, J.; Vondrasek, J. *Chem. Phys. Lett.* **1996**, *257*, 31-35.

Chapter 6

An *Ab Initio* Quantum Mechanical Study of the Structures and Energies for the Pseudorotation of 2'-Deoxyribose and Ribose Sugars

Abstract

Ab initio quantum mechanical (QM) methods are used to determine the potential energy of pseudorotation for 2'-deoxyribose and ribose sugars. A pyrrole is used as a substitute for the naturally occurring nucleic acid bases; adenine, thymine, guanine, cytosine, and uracil. At the highest calculation level (HF/6-31G**//LMP2/cc-pVTZ(-f)) for 2'-deoxyribose, the C2'-endo conformation is the global minimum. The C3'-endo conformation is a local minimum 0.6 kcal/mol higher in energy and an eastern barrier of 1.6 kcal/mol separates the two minima. Pseudorotation energies of ribose are quite complex and are strongly affected by local orientations of the 2' and 3' hydroxyl groups. When the hydroxyl groups are allowed to assume any conformation, the global minimum remains the C2'-endo conformation. The eastern barrier increases slightly to 1.8 kcal/mol and the C3'-endo local minimum lies 0.6 kcal/mol above the global minimum. Constraining the torsion angle of the C3' hydroxyl group to -146°, as is found in RNA polymers, results in the C3'-endo conformation becoming the only energy minimum with the C2'-endo conformation 1.9 kcal/mol higher in energy. Bond angles within the pentofuranose ring are correlated to the pseudorotational phase, as is observed by X-ray crystallography and is predicted by pseudorotation theory. Finally, a force field for use in molecular mechanics and dynamics simulations is presented which reproduces the QM potential energies for the 2'-deoxyribose and ribose sugar.

I. Introduction

Ribose and 2'-dexoyribose sugars are the basic subunits which differentiate RNA and DNA respectively. The conformations of these substituted pentofuranose sugars is of critical importance to the global structure of nucleic acids and has been the subject of considerable theoretical and experimental study. A fundamental concept governing the conformations of 5-membered rings was first formulated to describe the interconversion of conformations of cyclopentane through "pseudorotation."[1] For cyclopentane, the lowest energy conformation has one atom positioned out of the plane which is defined by the remaining four atoms of the ring. Through pseudorotation, ring conformations having different out-of-plane atoms may interconvert without traversing a high energy planar structure.

For cyclopentane, this process of pseudorotation has little or no energy barrier resulting in essentially "free" ring pucker motion.[2-4] In the case of tetrahydrofuran (THF) and the more complex 2'-deoxyribose and ribose sugars, there are barriers to interconversion stemming from the O4' oxygen in the ring and the exocyclic hydroxyl groups.[3,5] NMR experiments offer a straightforward approach to determine the relative energies of low energy pucker states.[6-21] However, the magnitude of the energy barrier between these stable conformations has escaped experimental measurement, except for purine ribosides.[6] The complete energy landscape for pentofuranose sugars is critical both for the understanding of the forces which govern these conformational preferences and for the development and evaluation of force field potentials used for molecular mechanics and dynamics simulations.

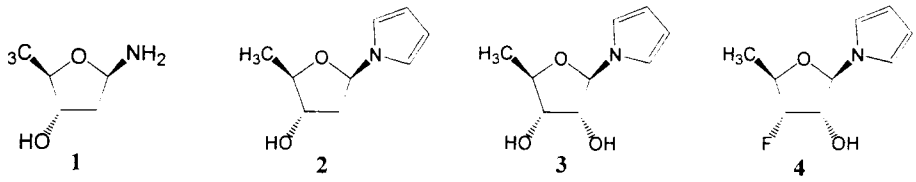


Figure 6.1. Model systems used for all *ab initio* quantum mechanical calculations. The 2'-deoxyribose models (**1** and **2**) differ by having an amino group or pyrrole group as a substitution for the base at C1'. The ribose (**3**) model and 3'-fluoro-ribose (**4**) both have a pyrrole substitution for the base at C1'.

We have undertaken a systematic study of substituted pentofuranose sugars (Figure 6.1) to examine the geometries and potential energies of pseudorotation. Initially it was necessary to establish a suitable substitution at C1' which was computationally feasible and still retained the properties of the full nucleic acid base. Using a pyrrole substitution at C1', we then identified the level of QM calculation and basis set necessary for accurate geometries (HF/6-31G**) and energies (LMP2/cc-pVTZ), when compared to experimental data. This model cluster and computational method allowed us to determine:

1. for 2'-deoxyribose, the C3'-endo conformation is 0.6 kcal/mol higher in energy than the global minimum C2'-endo conformation and is separated by the 1.6 kcal/mol eastern barrier.
2. for ribose, the C2'-endo conformation remains the global minimum because of the formation of a hydrogen bond between the 3'-OH and 2'-OH, as has been observed by X-ray crystallography.[22,23]

3. constraining the rotation of the 3'-OH to prevent hydrogen bond formation alters the global minimum to C3'-endo, consistent with the observed sugar pucker preferences of RNA polymers.

In addition, a force field capable of reproducing these *ab initio* QM results is presented. These new sugar parameters, embedded in the framework of a generic force field, such as UFF[24], and modified to include our accurate hydrogen bond potential,[25] are ideal for molecular mechanics and dynamics simulations of nucleic acids interacting with a diverse collection of organic and inorganic molecules. To begin, it is useful to review some important concepts regarding 2' and 3' substituted pentofuranose sugars including pseudorotation, structural nomenclature, anomeric and gauche effects, and previous theoretical work in the field.

Pseudorotation and Structural Nomenclature

Pseudorotation involves the conversion of one ring conformation to another without passing through a high energy planar state (see Figure 6.2). Altona and Sundaralingam[26] have derived a simple mathematical relationship relating the five ring torsion angles, $j = 0$ - 4 (see Figure 6.3), of pentofuranose to two parameters: τ_m and P.

$$\tau_j = \tau_m \cos(P + 0.8\pi(j-2)) \qquad (6.1)$$

τ_m is the maximum pucker amplitude and P is the pseudorotation phase that may range from 0 to 2π. However, in this work P will be in units of degrees and in the range between 0° and 360°. Other more complex and rigorous expressions for pseudorotation have been derived.[27] Some require artificial constructs to define the plane $z = 0$ and are useful for systems with unequal bond lengths and angles.[28] A statistical survey of over

100 ribose and deoxyribose crystal structures indicates that τ_m is dependant on both pucker geometry and nature of the sugar.[29] The average τ_m for 2'-deoxyribose and ribose was found to be 35.2° and 37.1°, respectively.

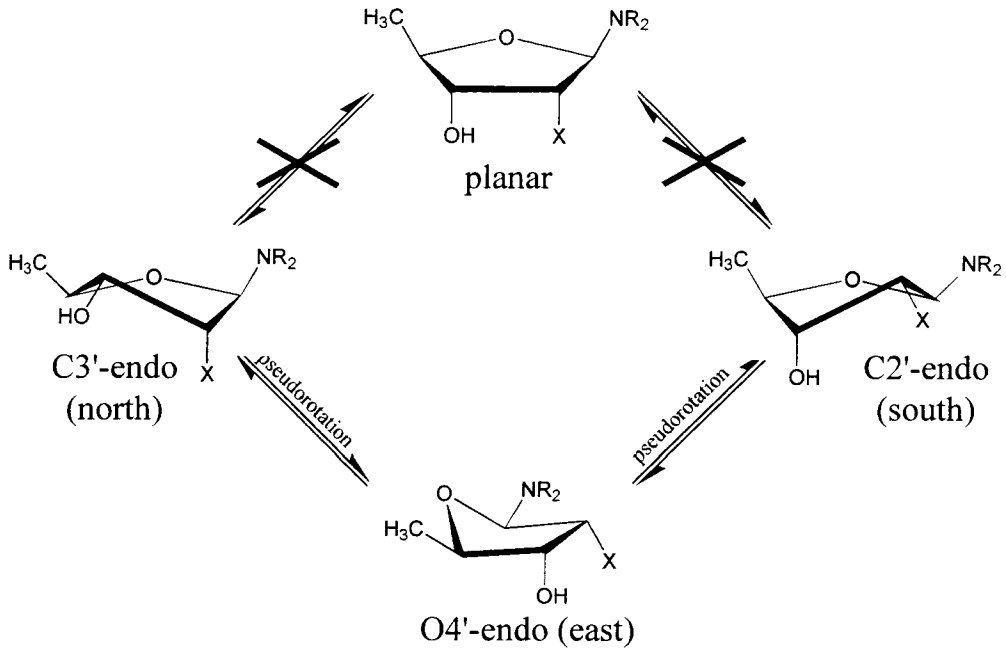


Figure 6.2. A schematic of the pseudorotation process in which low energy conformations (C3'-endo and C2'-endo) inter-convert without passing through a planar intermediate.

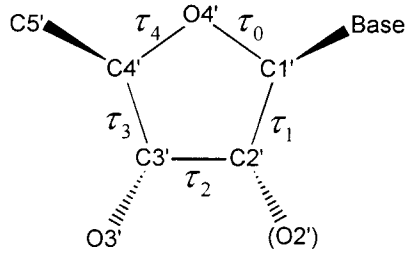


Figure 6.3. Atom names and torsion definitions for substituted pentofuranose rings. In order to enforce the pseudorotation pathway, sugar torsion angles τ_0 and τ_4 were constrained as described in the Methods section.

In general, there are only two major conformations observed for all ribose and 2'-deoxyribose sugars; C2'-endo and C3'-endo. The 2' or 3' designation refers to the atom which is out of the plane of the remaining atoms forming the ring. This atom may lie above the ring plane on the same side as C5' or on the opposite side, resulting in an "endo" or "exo" conformation, respectively (see Figure 6.2). Using the pseudorotation phase coordinate *P*, C3'-endo lies at $P = 18°$ while the C2'-endo conformation is at $P = 162°$. The full pseudorotation cycle also may be divided into quadrants in which conformations clustered around C2'-endo are known as southern or S-type sugars. Similarly, the C3'-endo pucker is a northern or N-type sugar conformation. This nomenclature system is used extensively in the interpretation of NMR experiments in which the measurement of vicinal coupling constants can be used to determine the ratio of N/S sugar conformations.[8-12,15-17,20]

The two energy barriers between the C3'-endo and C2'-endo conformations depend upon the direction of pseudorotation: $P = 18° \rightarrow 162°$ via 90° (east) or via 270° (west). Several crystal structures have been reported in which the ring geometry is E-type (O4'-endo), but none have been reported for structures with a W-type (O4'-exo) conformation.[29] Furthermore, NMR experiments indicate a rapid interconversion between the C2'-endo and C3'-endo conformations. Based on these observations, it is widely accepted that the western barrier is higher than the eastern barrier, which must be sufficiently low to allow rapid interconversion at room temperature.[5,30-38] Using ^{13}C-NMR relaxation experiments, the eastern barrier for purine ribonucleosides is reported to be 4.7 ± 0.5 kcal/mol.[6]

Anomeric and Gauche Effects

The anomeric effect and the gauche effect are two principles that have been invoked to explain the existence and nature of the barriers to pseudorotation for ribose and 2'-deoxyribose. The anomeric effect was first discovered in sugars with the observation that a methoxy group attached to the anomeric carbon prefers the sterically unfavorable axial position.[39] In chemical systems R-X-C-Y where X has at least one set of lone pair electrons and Y is an electronegative atom, the anomeric effect will favor a gauche torsion angle.[14,17-19,40-44] In that geometry, the lone pair of X may donate into the σ^* antibonding orbital of the C-Y bond. Maximum overlap with the σ^* antibonding orbital occurs when the lone pair is trans to the C-Y bond. For nucleic acids, the anomeric effect arises only for the O4'-C1'-N1(N9)-R and the phosphodiester torsions. It has been suggested that the anomeric effect is critical in defining the eastern barrier during pseudorotation.[40]

Conceptually, the gauche effect is quite similar to the anomeric effect but affects chemical systems of the type X-C-C-Y. Here too electronegative substitutions at X and Y will stabilize the gauche conformation relative to the trans conformation.[14,18,19,42,45,46] The gauche effect is important in the ribose and 2'-deoxyribose systems because there are several X-C-C-Y type torsions. For 2'-deoxyribose, these include O4'-C4'-C3'-O3' and O4'-C4'-C5'-O5', while ribose has three additional torsions, O4'-C1'-C2'-O2', N1(N9)-C1'-C2'-O2', and O2'-C2'-C3'-O3'. These gauche effects result in C2'-endo being the preferred structure for the 2'-deoxyribose sugars found in DNA polymers.

Theoretical Approaches to Pentofuranose Pseudorotation

Early theoretical work using empirically derived potential energy functions gave conflicting results and led to discrepancies between theoretical predictions and experimental observations.[30,47] It was not until the gauche effect was explicitly taken into account that Olson was able to derive an empirical energy function that correlated well with known crystallographic and solution data.[5] Indeed, the potential energies of pseudorotation derived with these functions have been the benchmark against which several popular molecular mechanics/dynamics force fields have been parameterized.[48-52]

Several quantum mechanical studies have been carried out that examine the conformational energies of pentofuranose sugars. Semi-empirical and *ab initio* methods have been used in the pseudorotational analysis of THF.[2,37,53] Substituted pentofuranose sugars also have been studied but in those cases either the nucleic acid base was not included[14,54] or only the minimum energy conformations were calculated.[55] Consequently, the potential energy for pseudorotation of 2'-deoxyribose and ribose systems comparable to those found in natural nucleic acids heretofore has not been determined using *ab initio* QM methods. The results of a systematic *ab initio* QM study of substituted pentofuranose sugars are presented herein which addresses this shortcoming.

II. Methods

Ab Initio Quantum Mechanical Calculations

All *ab initio* QM calculations were carried out using the Jaguar 3.0 software package.[56] In order to determine the potential energy of pseudorotation, sugar torsion

angles τ_0 and τ_4 (Figure 6.3) were constrained. Values for τ_0 and τ_4 were calculated using equation (6.1) and $\tau_m = 40$. For each point along the pseudorotation pathway, full geometry optimization (HF/6-31G**) was carried out while maintaining the desired τ_0 and τ_4 angles. Single point energies were determined using local second order Möller-Plesset perturbation theory (LMP2)[57-60] with the cc-pVTZ(-f) basis set[61] and frozen core orbitals. In order to ensure the points along this constrained pseudorotation pathway were a reasonable approximation of the true pseudorotation pathway, the C3'-endo and C2'-endo conformations were also optimized (HF/6-31G** and LMP2/cc-pVTZ(-f)) with no constraints. In addition, a transition state optimization of O4'-endo conformation (eastern barrier) at the same level with no constraints was also carried out. For some ribose calculations, the C4'–C3'–O3'–H torsion angle was constrained to be -146°, as is found in RNA polymers.[29] This constraint enabled the pseudorotation potential for ribose to be calculated for conformations which were not compromised by an intramolecular hydrogen bond.

Rotamer Definitions

Exocyclic substitutions of an amino group at C1' or hydroxyl groups at C2' and C3' complicates the potential energy surfaces of these sugars. In many cases, the torsion angle about C–N or C–O bond has a considerable impact on the energy of a particular sugar pucker. Figure 6.4 illustrates the nomenclature which we have adopted to describe each of these rotamers. For the C1'-NH_2 bond, general torsion angles are defined relative to the O4'-C1' bond and the NH_2 lone pair vector. Thus the *+synclinal* (*+sc*), *antiperiplanar* (*ap*) and *-synclinal* (*-sc*) geometries correspond to (lone-pair)–N–C1'–O4'

torsion angles of 60°, 180° and -60° respectively. Similarly, for hydroxyl group substitutions at C2' and C3', torsion angles are measured relative to the H–O and C1'–C2' bonds for substitution at C2' or the H–O and C3'–C4' bonds for substitution at C3'. In describing ribose conformational rotamers, the C3' substitution will be listed first, then the C2' substitution. For example, a ribose conformation C2'-endo(*ap*/+*sc*) would have an *ap* C3' hydroxyl group and a +*sc* C2' hydroxyl group.

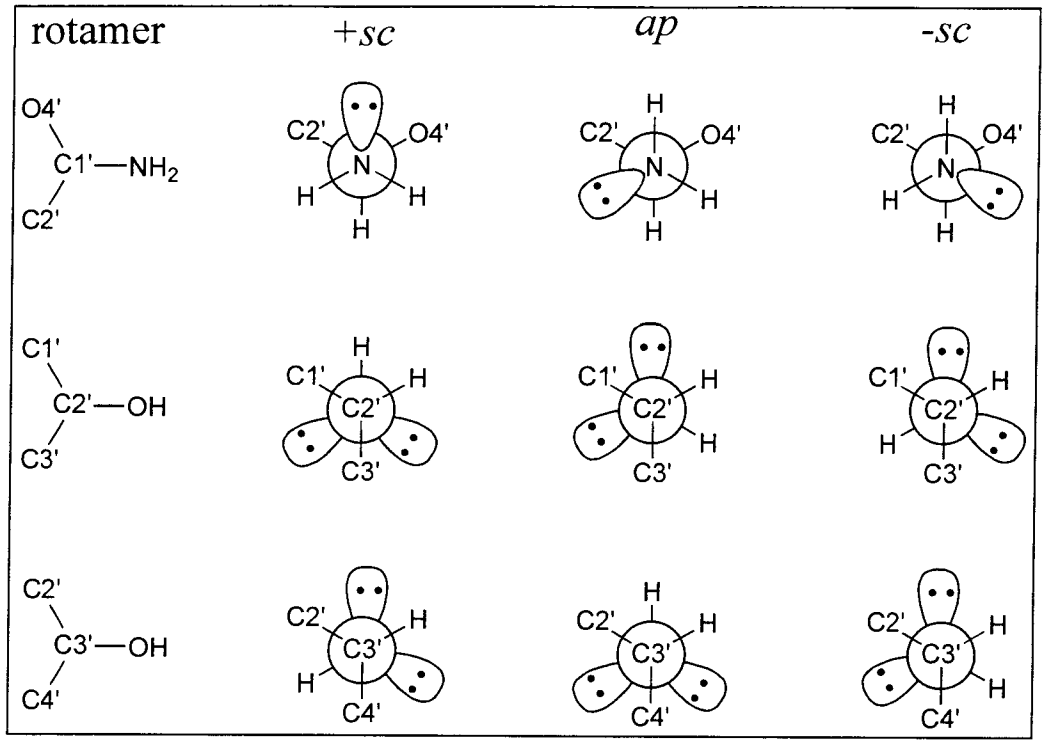


Figure 6.4. The three possible rotamer conformations for substitutions at C1', C2' and C3'.

Force Field Calculations

The Cerius^2 3.5 software package[62] was used for all molecular mechanics and dynamics simulations. The strategy we employ for force field development begins with a simple generic force field (UFF[24]) which is then tuned to reproduce QM potential energies for small model systems, as has been discussed elsewhere.[63] All standard

valence terms including bonds, angles, torsions and inversions are taken from UFF.[24] The Dreiding FF[64] exponential-6 parameters are used for all van der Waals interactions and a standard coulombic potential completes the non-bond terms. For pseudorotation of ribose and deoxyribose sugars, two new atom types were required: O_S for O4' and C_S for C3'. These new atom types allow for the parameterization of several new torsion potentials which take into account anomeric and gauche effects. A standard torsion potential is used and can be represented by a Fourier series of the form

$$E_{torsion} = \frac{1}{2}\sum_{n=1}^{6} K_{\theta,n}(1 - d\cos(n\phi)) \qquad (6.2)$$

where $K_{\theta,n}$ is the torsion energy barrier, $d = \pm 1$ and is the phase factor and ϕ is the torsion angle ($\phi = 0$ for *cis*). A complete description of the force field atom types and point charges is shown in Figures 6.5 and 6.6. Atomic point charges were determined from the electrostatic potential derived from the electron density distribution (constrained to reproduce the molecular monopole and dipole moments) calculated from the converged LMP2/cc-pVTZ(-f) wavefunction.[65] Calculations using Amber 4.1[51] and CFF95[66-68] were also carried out in Cerius^2 3.5 and used an $\varepsilon = 1$ dielectric constant.

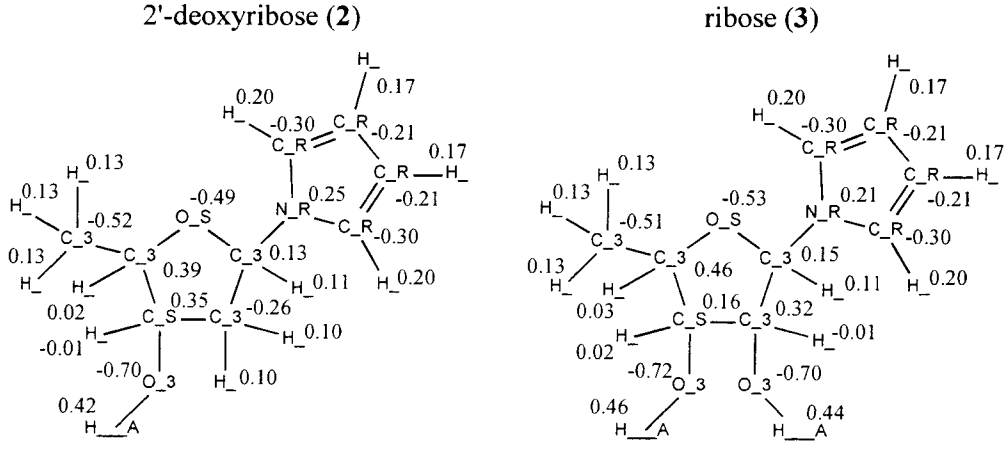


Figure 6.5. Atomic point charges and MSCFF force field atom types for 2'-deoxyribose and ribose sugars.

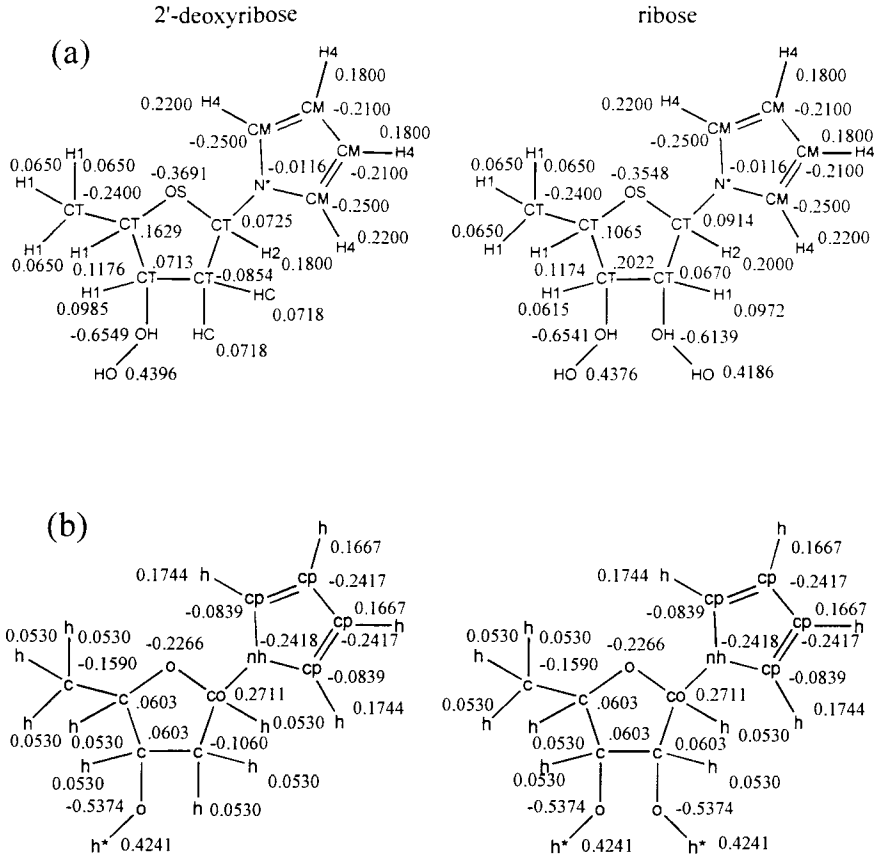


Figure 6.6. (a) Atomic point charges and Amber 4.1 force field atom types for 2'-deoxyribose and ribose sugars. (b) Atomic point charges and CFF95 force field atom types for 2'-deoxyribose and ribose sugars.

III. Results and Discussion

Model Systems and Level of Theory

Complete potential energy calculations for nucleosides of all five natural bases (adenine, thymine, guanine, cytosine and uracil) is computationally intensive and an alternative sugar model is necessary which still retains the key properties of the full bases. For the first 2'-deoxyribose model system, the full base was substituted with a simple amine (**1**). Figure 6.7 shows a plot of the HF/6-31G** relative energies for pseudorotation and the dramatic dependency upon the amine rotamer conformation. The lowest energy pseudorotation path is almost exclusively the *ap* rotamer in which the

nitrogen lone pair electrons are *trans* to the C1'–O4' bond (see the Methods section and Figure 6.4 for rotamer nomenclature).

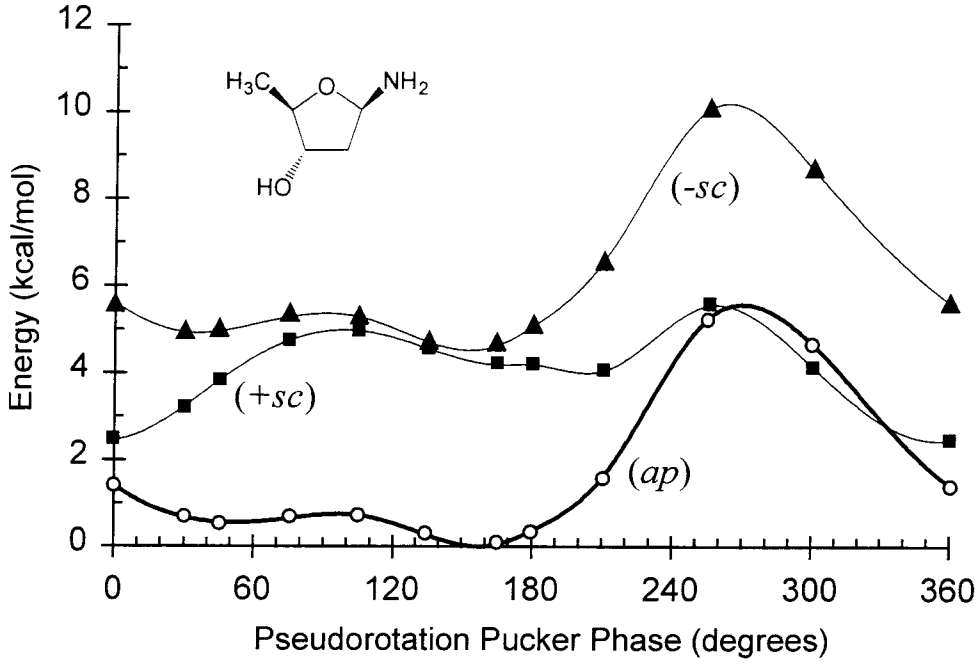


Figure 6.7. The HF/6-31G** potential energy (kcal/mol) for 2'-deoxyribose (**1**) with an amine substitution for the naturally occurring base as a function of pseudorotation angle P. The three possible rotamer conformations for the amino group have very different energies. All energies are reported relative to C2'-endo(*ap*) conformation (-399.927569 hartree).

Several features of the amine substitution make it unsatisfactory as a model for the aromatic base. In addition to complicating the potential energy surface due to several possible rotamers, the amino group nitrogen has sp^3 hybridization thereby allowing the lone pair electrons to participate in a strong anomeric effect with O4'. This anomeric effect likely stabilizes the *ap* rotamer relative to ±*sc* rotamers and would not be present for the full base. The lack of an eastern barrier between the C3'-endo and C2'-endo conformations is also problematic and is not consistent with NMR experiments. For

these reasons, the amine model was abandoned and a pyrrole substitution, which maintains the aromatic nitrogen, was pursued.

A pyrrole substitution for the complete nucleic acid base is a satisfactory model. The symmetry of the pyrrole ring alleviates complications due to rotamers while simultaneously enforcing sp^2 character on the nitrogen atom. The pseudorotation potential energy curves for 2'-deoxyribose (**2**) are plotted in Figure 6.8. The potential energy curve of **2** has a well defined eastern barrier consistent with NMR data and the expected minima at C2'-endo and C3'-endo.

Table 6.1. Relative Energies (kcal/mol) for Fully Optimized 2'-deoxyribose (**2**).

Method	C3'-endo	O4'-endo	C2'-endo
HF/6-31G**//HF/6-31G**	0.44	1.16	0.00[a]
HF/6-31G**//LMP2/cc-pVTZ(-f)	0.40	1.31	0.00[b]
LMP2/cc-pVTZ(-f)//LMP2/cc-pVTZ(-f)	0.61	1.19	0.00[c]

[a]Absolute energy = -552.548294 hartree. [b]Absolute energy = -554.634699 hartree.
[c]Absolute energy = -554.636953 hartree.

Several calculations were carried out for the 2'-deoxyribose (**2**) model system to establish the optimum level of computational rigor which yields the best accuracy while still remaining computationally feasible. Table 6.1 reports the relative energies for complete geometry optimization with no constraints of **2** in the C3'-endo, O4'-endo (eastern barrier), and C2'-endo conformations. The HF/6-31G** and LMP2/cc-pVTZ(-f) optimized geometries are very similar. The only systematic difference observed between the two levels of theory is the C4'–O4'–C1' angle which averages 2.5° less for the LMP2/cc-pVTZ(-f) optimized structures. Given the small structural difference between

these two methods, all subsequent geometry optimizations were carried out at the HF/6-31G** level, followed by LMP2/cc-pVTZ(-f) single point energies.

2'-deoxyribose

The HF/6-31G** and LMP2/cc-pVTZ(-f) pseudorotation potential energies are plotted in Figure 6.8a. Both curves have the same general trend with a global minimum C2'-endo conformation. The C3'-endo conformation is 0.6 kcal/mol and 0.4 kcal/mol higher in energy for the LMP2/cc-pVTZ(-f) and HF/6-31G** curves respectively. The O4'-endo barrier is well defined with a maximum energy of 1.6 kcal/mol and 1.1 kcal/mol for the LMP2/cc-pVTZ(-f) and HF/6-31G** curves respectively. There is excellent agreement between the fully optimized HF/6-31G** points listed in Table 6.1 and the constrained pseudorotation curve in Figure 6.8a. This indicates that the torsion constraints used for the pseudorotation calculations are a reasonable representation of the true pseudorotation pathway.

An analysis of the endocyclic bond angles of the furanose ring reveals a strong correlation between pseudorotational phase and bond angles (Figure 6.9). This correlation is observed experimentally[29] and is predicted from pseudorotation theory.[1,7,27] Endocyclic bond angles for which the central atom is C1', C2', C3' or C4' all oscillate in the range of 101° to 106° and are fairly symmetrical. In contrast, the C4'–O4'–C1' bond angle oscillates in the range of 108° to 112° and is noticeable unsymmetrical. For the O4'-exo conformations found near P = 270, steric interactions between the C4' and C1' substituents force the C4'–O4'–C1' bond angle to increase and that portion of the curve to be unsymmetrical when compared to the O4'-endo conformation. The resulting angle strain, in combination with van der Waals interactions, leads to a high western barrier.

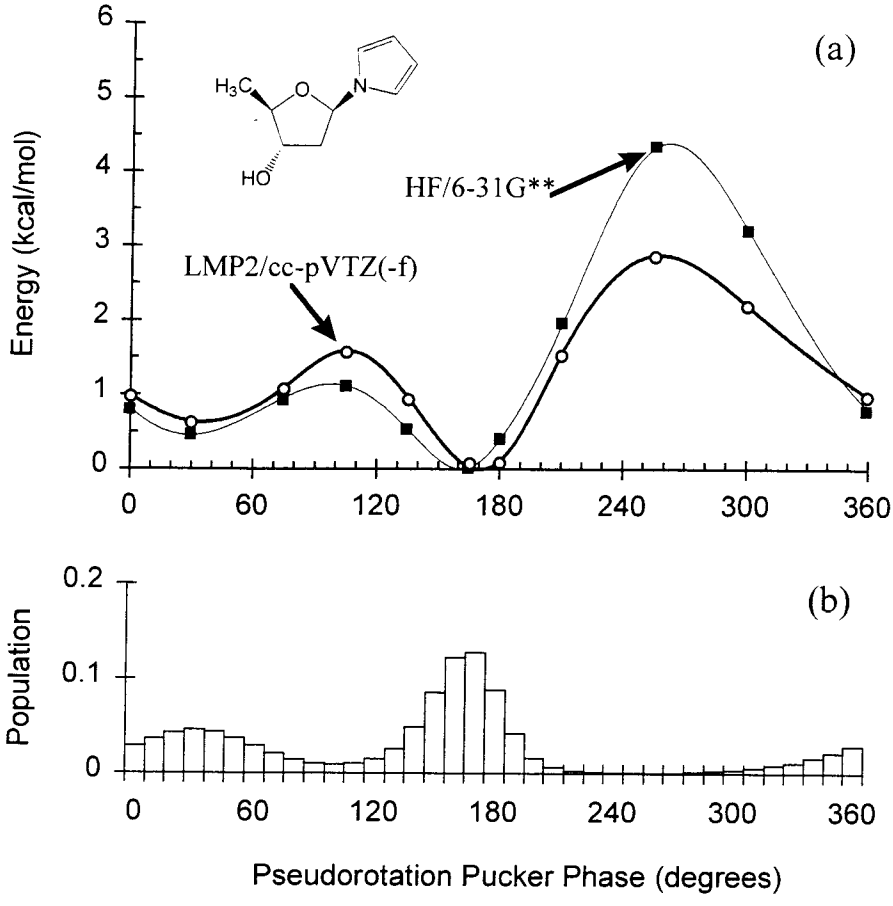


Figure 6.8. (a) The HF/6-31G** and LMP2/cc-pVTZ(-f) potential energies (kcal/mol) as a function of pseudorotation angle P for 2'-deoxyribose (**2**) with a pyrrole substitution for the naturally occurring base. All energies are reported relative to the C2'-endo conformation (HF/6-31G** = -552.548378; LMP2//cc-pVTZ(-f) = -554.634699 hartree). (b) Normalized statistical weights σ at T = 298° K using the LMP2/cc-pVTZ(-f) energies.

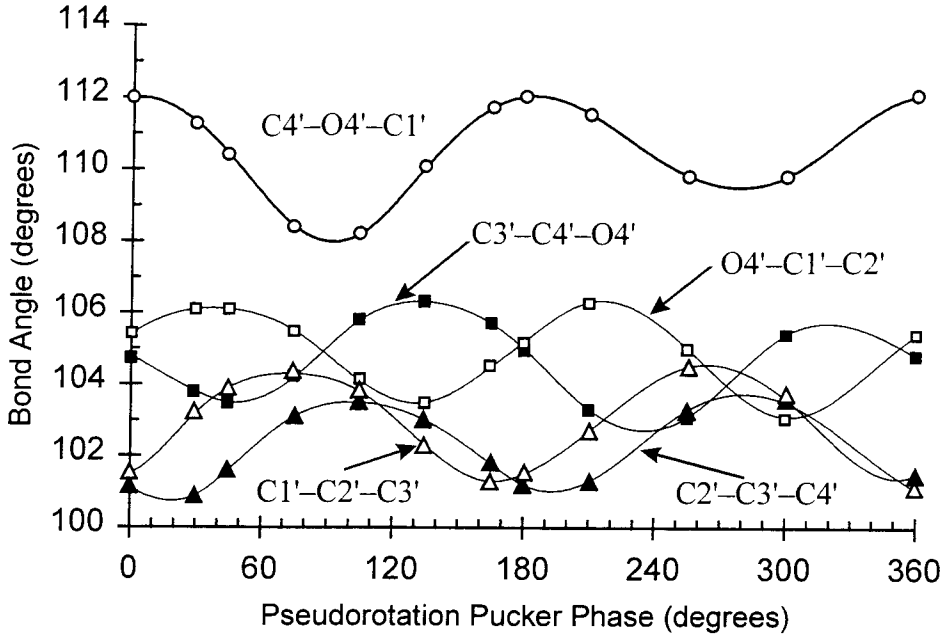


Figure 6.9. Endocyclic bond angles for 2'-deoxyribose (**2**) as a function of pseudorotation angle P. There is a strong coupling between the pseudorotational phase and all the bond angles.

The statistical populations (T=298° K) of (**2**) are plotted versus pseudorotation phase in Figure 6.8b. These populations are derived from the LMP2/cc-pVTZ(-f) single point energies and are normalized by the total energy of the system. Summing each of these contributions into quadrant populations gives the following: $\sigma_n = 0.26$, $\sigma_e = 0.18$, $\sigma_s = 0.54$, $\sigma_w = 0.02$. As expected from crystallographic studies, the σ_s quadrant is heavily favored. These populations are different from those reported by Olson using an empirically derived potential energy function.[5] The empirical potential energy function overestimates σ_s ($\sigma_s = 0.74$) and predicts σ_n and σ_e to be almost equal. A study of 127 mononucleoside and mononucelotide crystal structures[29] reported 78 C2'-endo conformations ($\sigma_s = 0.61$), 31 C3'-endo conformations ($\sigma_n = 0.24$) and 18 conformations

(σ_e = 0.14) which were neither C2'-endo nor C3'-endo and presumably in the O4'-endo range. Thus the LMP2/cc-pVTZ(-f) potential energies are in excellent agreement with available crystallographic data.

Using NMR, it is difficult to quantify the energy barrier between low energy conformations. However, NMR experiments which measure vicinal coupling constants can be used to determine σ_n and σ_s. A study measuring σ_s of five 2'-deoxyribose nucleosides (dA, dG, dC, dT and dU) at 278° K and 358° K has been reported.[19] An average over the five different bases gives a mean experimental value for σ_s of 0.66 and 0.62 at the two temperatures respectively. Summing the QM energy derived populations into hemispheres (NMR experiments resolve the sugar conformations only into N-type or S-type) at 278° K and 358° K gives σ_s = 0.64 and σ_s = 0.59, consistent with the NMR experiments. The remarkable correlation of *ab initio* QM results with both crystallographic and NMR data suggests that both our model system containing a pyrrole and the HF/6-31G**//LMP2/cc-pVTZ(-f) potential energies give an accurate representation of 2'-deoxyribose pseudorotation.

Ribose

Applying the same computational methods established for 2'-deoxyribose (**2**) to determine the pseudorotation potential of ribose (**3**) is considerably more complex due to the multiple possible conformations of the 2'- and 3'-hydroxyl groups. Each hydroxyl group may assume one of three possible orientations, *+sc*, *-sc* and *ap*, giving a total of 9 distinct rotamer combinations for each pucker state. Some of these rotamer combinations are sterically forbidden and most are energetically unfavorable. In order to identify those rotamers which are most likely sampled during pseudorotation, all 9 possible rotamers in

both the C3'-endo and C2'-endo sugar conformations were geometry optimized (HF/6-31G**). No constraints were used to enforce the rotamer geometry; therefore, some conformations did not yield minimized structures. Table 6.2 reports the energies of those conformations which did constitute local minima. Interestingly, the lowest energy conformations are C2'-endo(*ap*/+*sc*) and C3'-endo(+*sc*/+*sc*), both sharing the +*sc* rotamer at the C2' hydroxyl group.

Pseudorotation potential energies were calculated for ribose (**3**) in both the (*ap*/+*sc*) and (+*sc*/+*sc*) rotamer conformations. As expected, both rotamers are required to map the lowest energy pseudorotation pathway for ribose as shown in Figure 6.10a. The interconversion between the two rotamer conformations occurs at the eastern (O4'-endo) and western (O4'-exo) barriers. The (*ap*/+*sc*) conformation is favorable in the northern hemisphere (P = 270-360; 0-90) while the (+*sc*/+*sc*) conformation is preferred in the southern hemisphere (P = 90-270). Single point energies (LMP2/cc-pVTZ(-f)) were calculated only for the lowest energy rotamer at each point. Perhaps the most surprising result is that the C2'-endo conformation remains the global minimum for ribose. The C3'-endo conformation lies 0.6 kcal/mol higher in energy with an eastern barrier of 1.8 kcal/mol separating the two minima. These energies are unexpected given the known preference for RNA polymers to assume an A-form helix which requires the C3'-endo sugar conformation.

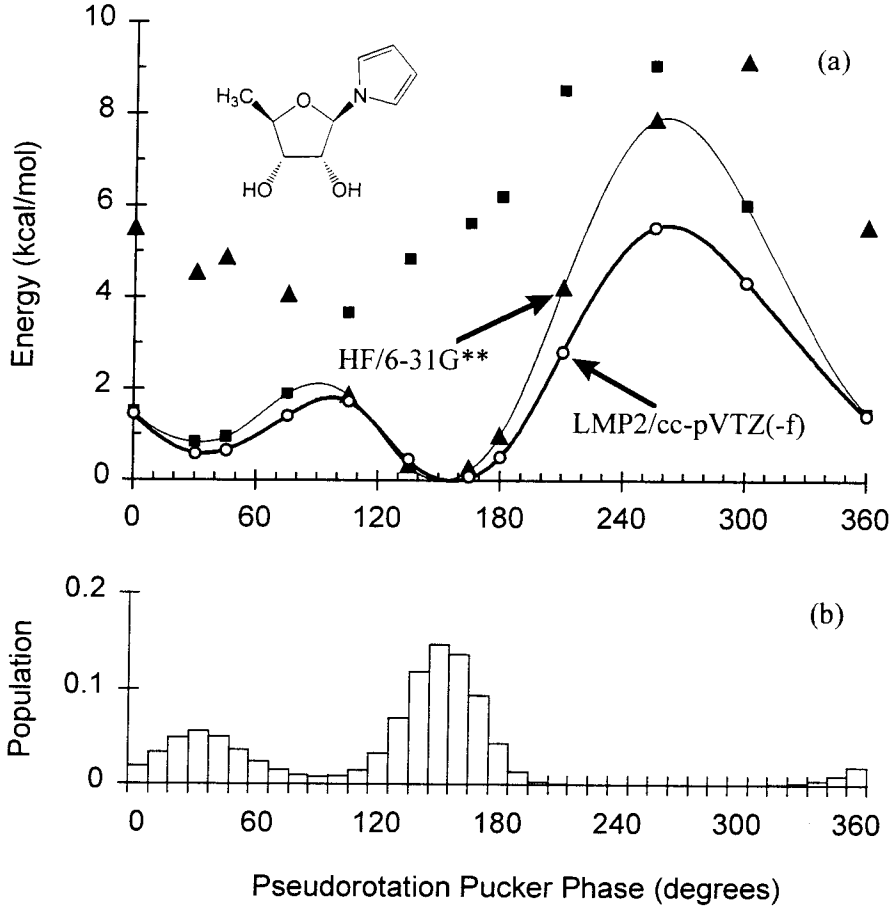


Figure 6.10. (a) The HF/6-31G** and LMP2/cc-pVTZ(-f) potential energies (kcal/mol) as a function of pseudorotation angle P for ribose (**3**) with a pyrrole substitution for the naturally occurring base. For the HF/6-31G** points, two possible C3' rotamers were examined: (▲) = *ap*/+*sc*; (■) = +*sc*/+*sc*. It is clear that both rotamers are needed to trace out the lowest energy path shown. LMP2/cc-pVTZ(-f) energies are calculated only for the most favorable rotamer. All energies are reported relative to the C2'-endo(*ap*/+*sc*) conformation (HF/6-31G** = -627.407468; LMP2//cc-pVTZ(-f) = -629.725164 hartree). (b) Normalized statistical weights σ at T = 298° K using the LMP2/cc-pVTZ(-f) energies.

Table 6.2. HF/6-31G** Relative Energies for All 9 Possible Rotamers for Ribose (**3**) in Both the C3'-endo and C2'-endo Optimum Conformations.

Rotamer	*-sc* (C3)	*ap* (C3)	*+sc* (C3)
		C3'-endo	
-sc (C2)	n/a	n/a	3.14
ap (C2)	1.14	0.98	1.38
+sc (C2)	6.48	n/a	**0.92**
		C2'-endo	
-sc (C2)	2.03	2.60	3.88
ap (C2)	n/a	n/a	n/a
+sc (C2)	5.23	**0.00**[a]	n/a

[a]Absolute energy = -627.407965 hartree.

NMR studies of ribonucleosides reveal that the nature of the base has a marked influence upon the preferred sugar pucker. Purine bases A and G have σ_s populations of 0.70 and 0.67 respectively at T = 278° K. However, pyrimidine bases C, T and U have a greater preference for the C3'-endo pucker with σ_s populations of 0.35, 0.49 and 0.46.[19] Using the LMP2/cc-pVTZ(-f) single point energies, the predicted populations of **3** are shown in Figure 6.10b. Due to the slight broadening of the C2'-endo potential energy minimum and an increase in the eastern barrier height when compared to 2'-deoxyribose, σ_s for ribose increases to 0.70, in agreement with the purine ribonucleoside NMR data. The unique behavior of the pyrimidine ribonucleosides is not predicted by the pyrrole model. Perhaps these compounds assume non-standard base geometries which exert an altered electronic influence on the furanose ring.

A closer examination of the global minimum C2'-endo(*ap*/+*sc*) conformation of ribose (**3**) suggests that a likely explanation for the unusual stability of this conformation stems from an intramolecular hydrogen bond between the C3' hydroxyl hydrogen and C2' hydroxyl oxygen atoms. Experimentally, an intramolecular hydrogen bond has been observed for several ribose crystal structures and does lead to a C2'-endo sugar conformation.[22,23] The presence of this intramolecular hydrogen bond compromises the ribose sugar (**3**) as a model for a ribose polymer or for comparison to experimental studies in aqueous solution. In the first case of a ribose polymer, the C3' hydroxyl group hydrogen atom is not present, having been replaced by a phosphate group, and thus cannot form the hydrogen bond in question. Secondly, in solution, such intramolecular hydrogen bonds are largely replaced by solvent interactions and are overestimated in gas phase calculations.[69,70] To better understand ribose sugar behavior in RNA polymers and for the purpose of force field development, it is desirable to design a model system which is not encumbered by an intramolecular hydrogen bond.

3'-fluoro-ribose

Substitution of the C3' hydroxyl group of ribose with a fluorine atom as in 3'-fluoro-ribose (**4**) may at first seem like reasonable model for a ribose sugar in an RNA polymer. Two advantages of the fluoro system over native ribose (**3**) are removal of rotamer complications at the C3' hydroxyl group and elimination of possible hydrogen bond formation in which the C3' hydroxyl group acts as the hydrogen bond donor. As was observed for the ribose system, two rotamers are required to cover the lowest energy pseudorotation pathway. Figure 6.11 shows the potential energy plot for (**4**) and it is clear that the C3'-endo conformation is now the global minimum and the C2'-endo

conformation is 0.5 kcal/mol higher in energy. The barrier between the C3'-endo and C2'-endo minima is 1.7 kcal/mol, similar to that calculated for 2'-deoxyribose (1.6 kcal/mol) and ribose (1.8 kcal/mol).

Interconversion between the most stable hydroxyl rotamers of 3'-fluoro-ribose occurs at the two barriers, O4'-endo and O4'-exo, as is observed for ribose. However, unlike ribose, 3'-fluoro-ribose prefers the *ap* rotamer for N-type conformations and the *-sc* rotamer for S-type conformations. The preference for a given rotamer is remarkably susceptible to subtle changes in the local environment. For ribose (**3**), the lowest energy conformations along the pseudorotation pathway all maintain the *+sc* rotamer at C2' since this allows for the best hydrogen bonding. When hydrogen bonding is precluded, the *ap* and *-sc* rotamers are preferred, such as observed for **4**. However, it is known that fluoro substitutions can participate in stronger anomeric and gauche effects than their hydroxyl counterparts; therefore, the 3'-fluoro-ribose systems remains an imperfect model.

Constrained 3'-OH Ribose

One method to eliminate the hydrogen bond donor properties of the C3' hydroxyl group is by constraining the C4'–C3'–O3'–H torsion angle to -146°, the same average position assumed by a phosphate atom in dinucleoside monophosphate and trinucleoside diphosphate crystals.[29] The resulting pseudorotation potential for ribose with a constrained C3' hydroxyl group is shown in Figure 6.12a. Now, the C3'-endo conformation is the global minimum. Indeed, there is little or no local minimum corresponding to C2'-endo, and it is clear why RNA polymers assume a C3'-endo sugar pucker. An analysis of the relative populations of N-type and S-type sugars indicates an overwhelming preference for the northern hemisphere ($\sigma_n = 0.90$).

There is experimental data which supports a strong correlation between the C3'–O3' torsion angle and ribose sugar pucker. NMR studies of ribodinucleoside monophosphates indicate a C3'-endo ↔ C2'-endo equilibrium with a bias towards the C3'-endo pucker. Furthermore, the C3'-endo pucker is associated with an *ap* torsion angle while the C2'-endo pucker is present with a *-sc* torsion angle.[71] This NMR data is consistent with the constrained pseudorotation potential for ribose which has a C3'-endo global minimum for the *ap* conformation.

While a DNA duplex may undergo a transition from a B-form helix (C2'-endo) to an A-form (C3'-endo) helix with changes in the aqueous environment, RNA polymers do not experience an analogous transition and remain in an A-form conformation. Various explanations based upon the presence of the C2'-hydroxyl group have been put forth to account for the structural stability of RNA polymers.[3] Intramolecular stabilization through hydrogen bonding between O2'-H and O4' of the neighboring sugar[72] or a water mediated hydrogen bond between O2'-H and 3'-phosphate[73,74] have been proposed based on NMR data. Alternative rationalizations include electronic effects due to the preference for a C3'-endo pucker with increasing electronegativity of the C2' substituent[75] or unfavorable steric interactions which preclude other sugar conformations. Our calculations indicate that no intramolecular hydrogen bond with a neighboring nucleotide or with a bridging water molecule is required to induce a preference for a C3'-endo pucker. The combined electronic effects of the 2'-hydroxyl group and the constrained orientation of the C3'–O3' torsion angle are sufficient to drive the pseudorotation potential of ribose sugars in RNA polymers completely to the A-form or C3'-endo pucker.

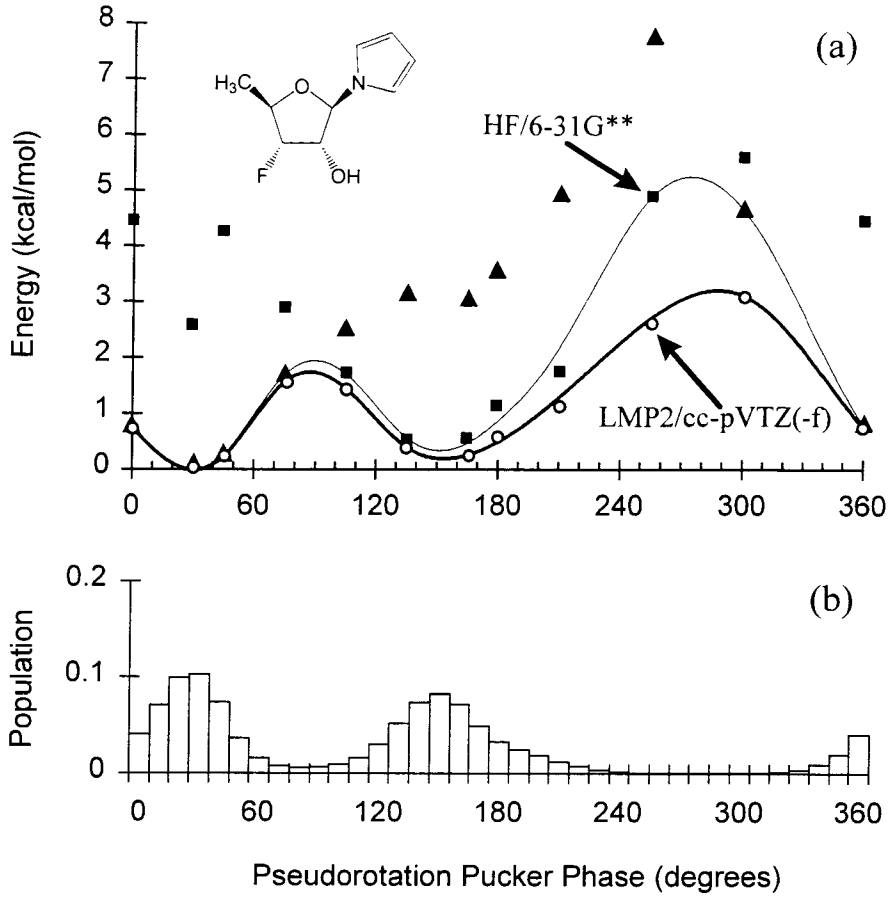


Figure 6.11. (a) The HF/6-31G** and LMP2/cc-pVTZ(-f) potential energies (kcal/mol) as a function of pseudorotation angle P for 2'-fluoro-ribose (**4**). All energies are reported relative to the C3'-endo conformation (HF/6-31G** = -651.399237 hartree; LMP2//cc-pVTZ(-f) = -653.732892 hartree). (b) Normalized statistical weights σ at T = 298° K using the LMP2/cc-pVTZ(-f) energies.

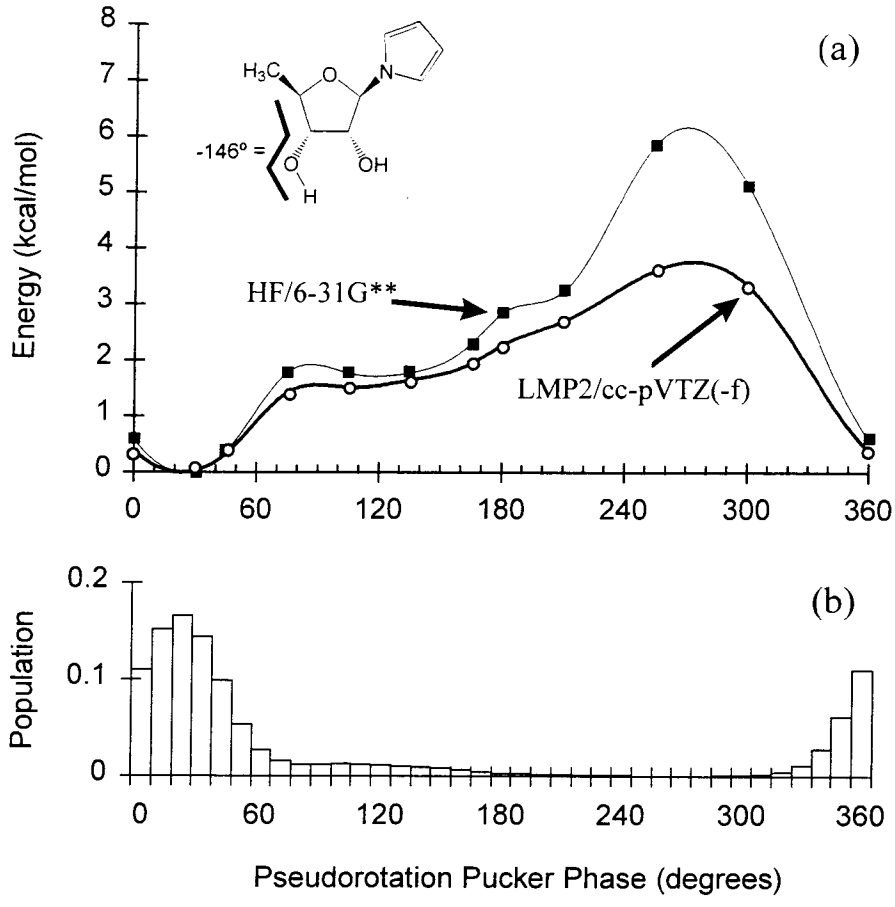


Figure 6.12. (a) The HF/6-31G** and LMP2/cc-pVTZ(-f) potential energies (kcal/mol) as a function of pseudorotation angle P for ribose (**3**) but with the C4'–C3'–O3'–H torsion angle constrained to -146°. This geometric constraint mimics the properties of RNA polymers. All energies are reported relative to the C3'-endo conformation (HF/6-31G** = -627.406345; LMP2//cc-pVTZ(-f) = -629.724598). (b) Normalized statistical weights σ at T = 298° K using the LMP2/cc-pVTZ(-f) energies.

Force Field Development

Using the potential energies for pseudorotation of 2'-deoxyribose and ribose sugars presented above, a set of force field (FF) parameters have been developed for accurate molecular mechanics and dynamics simulations of 2'-deoxyribose or ribose containing systems. All standard valence terms including bonds, angles, torsions and inversions are taken from UFF.[24] The Dreiding FF[64] exponential-6 parameters are used for all van der Waals interactions and a standard coulombic potential completes the non-bond terms. Upon the framework of this generic FF, two atom types unique to pentofuranose sugars are defined; C_S and O_S for C3' and O4' respectively. These atom types allow for the explicit parameterization of anomeric and gauche effects necessary to reproduce QM potential energies. A complete listing of FF parameters is included in Appendix A.

Only two additional torsion types are needed to reproduce the pseudorotation potentials of 2'-deoxyribose and ribose (unconstrained and constrained 3'-hydroxyl group) sugars. The first torsion potential accounts for the gauche preference of O_S–C_3–C_S–C_3 by adding a two-fold term to the existing C–C three-fold potential. Table 6.3 reports the QM and FF pseudorotation energies for tetrahydrofuran which serves as the model system for this generic X–C–C–C torsion. The second new potential applies to all O_x–C_x–C_x–(O_x or N_R) torsions where x = 3 or S types. In practice, this accounts for all torsions of the type X–C–C–Y in which X and Y are electronegative atoms. It is common for these types of torsions to be parameterized with a combination of two-fold and three-fold potentials.[5,48,50-52] Surprisingly, we find it is not possible to reproduce the QM potential energies using only two-fold and three-fold potentials for X–C–C–Y type

torsions. This type of potential results in local minima located at P = 0° and 180°, not the desired P = 30° and 165°. The use of a two-fold, three-fold and four-fold potential is capable of reproducing the QM potential energies. The same torsion barriers are used for all atom types and the resulting FF accurately reproduces the pseudorotation potentials of both ribose and 2'-deoxyribose sugars.

Table 6.3. Relative energies (kcal/mol) for pseudorotation of tetrahydrofuran (THF).

P	HF/6-31G**	MSCFF
0	0.00[a]	0.00[b]
30	0.15	0.14
60	0.37	0.32
90	0.44	0.48

[a]Absolute energy = -230.988771 hartree. [b]Absolute energy = 15.62 kcal/mol.

Force Field Evaluation

Figure 6.13 plots the pseudorotation potential energies calculated with the MSCFF for 2'-deoxyribose and ribose sugars. The position and energies for the two minima at C3'-endo and C2'-endo are well described, as is the eastern O4'-endo barrier. In some cases, the western barrier is overestimated as a result of steric interactions between C5' and the pyrrole base. The FF has a higher penalty for deformation of the pyrrole ring to relieve these steric clashes than is calculated by QM methods. These regions of the potential energy surface will not be populated during typical MD simulations and the resulting errors will be negligible.

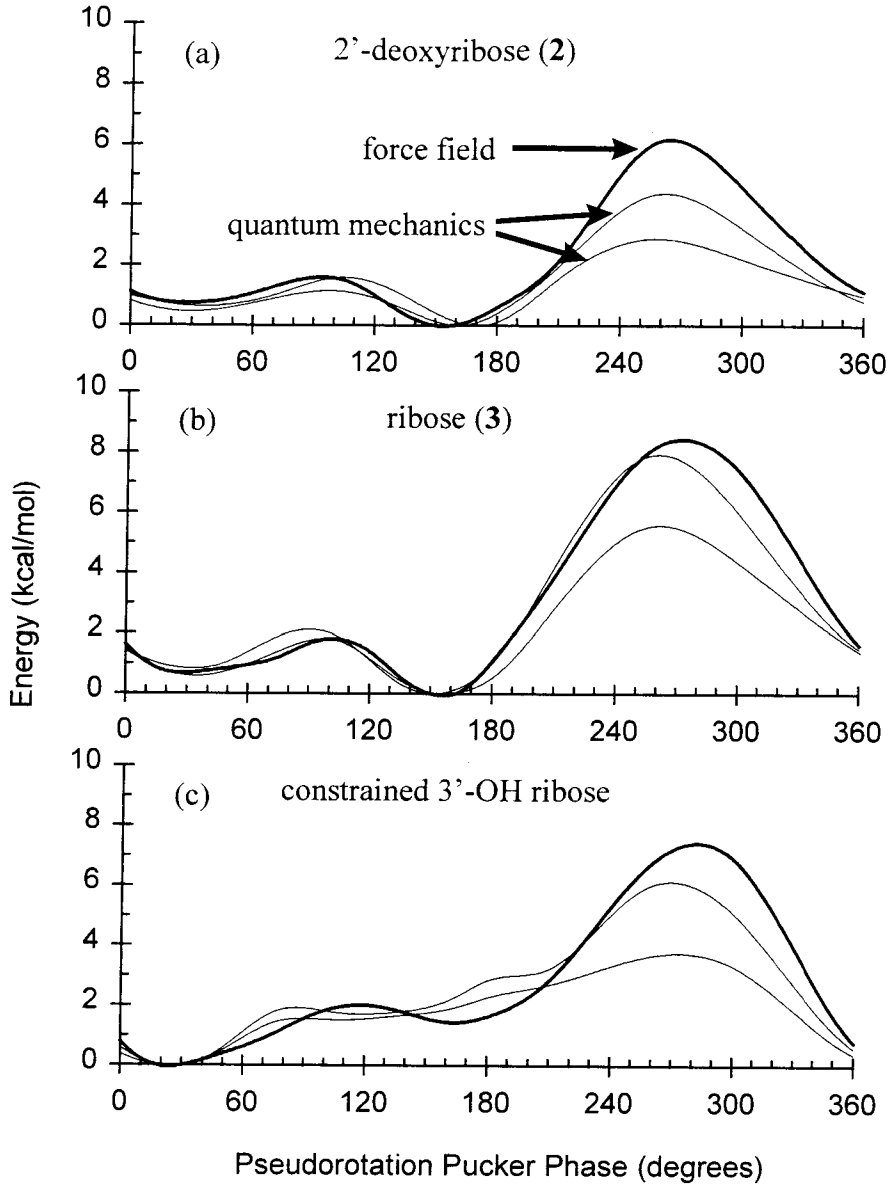


Figure 6.13. MSCFF potential energies for (a) 2'-deoxyribose, (b) ribose and (c) ribose with a constrained C3'-hydroxyl group. The heavy line is the force field energies while the lighter lines are HF/6-31G** and LMP2/cc-pVTZ(-f) energies (plotted in detail in Figures 6.8a, 6.10a and 6.12a).

Table 6.4. Statistical Weights for the North, East, South and West Quadrants of Pseudorotation Phase as Determined From Potential Energy Calculations.

Method	σ_n	σ_e	σ_s	σ_w
	2'-deoxyribose (**2**)			
HF/6-31G**//				
LMP2/cc-pVTZ(-f)	**0.26**	**0.18**	**0.54**	**0.02**
MSCFF	0.18	0.19	0.63	0.00
Amber 4.1	0.13	0.21	0.66	0.00
CFF95	0.76	0.16	0.08	0.00
	Ribose (**3**)			
HF/6-31G**//				
LMP2/cc-pVTZ(-f)	**0.22**	**0.21**	**0.58**	**0.00**
MSCFF	0.18	0.19	0.63	0.00
Amber 4.1	0.59	0.24	0.17	0.01
CFF95	0.05	0.06	0.89	0.00
	Constrained 3'-OH Ribose			
HF/6-31G**//				
LMP2/cc-pVTZ(-f)	**0.79**	**0.17**	**0.04**	**0.01**
MSCFF	0.66	0.27	0.07	0.00
Amber 4.1	0.83	0.16	0.00	0.01
CFF95	0.96	0.03	0.02	0.00

Table 6.4 compares the quadrant populations at T = 298° K for the LMP2/cc-pVTZ(-f) energies and several force fields including MSCFF, Amber 4.1[51] and CFF95.[68] The MSCFF populations are in good agreement with the QM predictions for all three cases examined. Amber 4.1 is biased towards C2'-endo conformation for 2'-deoxyribose and C3'-endo for ribose sugars. This bias is expected as the Amber 4.1 FF is intended for simulations of canonical DNA or RNA polymers. However, the resulting simulations may reflect an unrealistic stability for B-form DNA and A-form RNA. CFF95 does not

reproduce the populations of 2'-deoxyribose and ribose sugars correctly. Indeed, C3'-endo is the preferred conformation for 2'-deoxyribose while C2'-endo is preferred for ribose, in complete contradiction to the QM predictions.

IV. Conclusion

The pseudorotation potential energy calculations presented here provide an updated estimate of the local minima and energy barriers associated with the conformations of 2'-deoxyribose and ribose sugars. The energies predicted using *ab initio* QM methods are in good agreement with available NMR and crystallographic data. There is also a qualitative agreement with the benchmark empirical potential derived by Olson based on experimental data.[5,7] However, the Olson potential overestimates the ribose eastern barrier and does not correctly describe the positions of the local minima; a shortcoming consistent with the use of only a two-fold and three-fold torsion potential to represent the gauche effect. Because the parameterization of many force fields for nucleic acids are based upon the Olson potential, these new QM potential energies are significant for the evaluation of current force fields and the development of second-generation potentials.

Here it is demonstrated that the complex energy potential of pseudorotation may be accurately determined using high level quantum mechanical calculations. Furthermore, as detailed simulations of nucleic acids, including explicit solvent, become more common, it is increasingly critical that the force fields used reflect an accurate potential energy surface, not just the global minimum. These *ab initio* calculations and the MSCFF are part of a continuing effort to parameterize biological force fields from

first principles. The final goal is a new generation force field containing relatively simple functions which are parameterized independently from experimental observations.

Acknowledgements

This research was funded by DOE-BCTR (DE-FG36-93CH105 81, David Boron). The facilities of the MSC are also supported by grants from NSF (CHE 95-22179 and ASC 92-100368), Chevron Petroleum Technology Co., Saudi Aramco, Asahi Chemical, Owens-Corning, Exxon, Chevron Chemical Company, Chevron Research and Technology Co., Avery-Dennison, Hercules, BP Chemical, and Beckman Institute. K. A. B. acknowledges an NIH Predoctoral Biotechnology Training Grant.

V. References

(1) Kilpatrick, J. E.; Pitzer, K. S.; Spitzer, R. *J. Am. Chem. Soc.* **1947**, *69*, 2483-2488.

(2) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1358.

(3) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(4) Cui, W. L.; Li, F. B.; Allinger, N. L. *J. Am. Chem. Soc.* **1993**, *115*, 2943-2951.

(5) Olson, W. K. *J. Am. Chem. Soc.* **1982**, *104*, 278-286.

(6) Roder, O.; Ludemann, H.-D.; von Goldammer, E. *European J. Biochem.* **1975**, *53*, 517-525.

(7) Olson, W. K.; Sussman, J. L. *J. Am. Chem. Soc.* **1982**, *104*, 270-278.

(8) de Leeuw, F. A. A. M.; Vanbeuzekom, A. A.; Altona, C. *J. Comp. Chem.* **1983**, *4*, 438-448.

(9) de Leeuw, F. A. A. M.; Altona, C. *J. Comp. Chem.* **1983**, *4*, 428-437.

(10) de Leeuw, F. A. A. M.; Haasnoot, C. A. G.; Altona, C. *J. Am. Chem. Soc.* **1984**, *106*, 2299-2306.

(11) de Leeuw, F. A. A. M.; Vankampen, P. N.; Altona, C.; Diez, E.; Esteban, A. L. *J. Mol. Struct. (Theochem)* **1984**, *125*, 67-88.

(12) Diez, E.; Esteban, A. L.; Bermejo, F. J.; Altona, C.; de Leeuw, F. A. A. M. *J. Mol. Struct. (Theochem)* **1984**, *125*, 49-65.

(13) Chuprina, V. P.; Nerdal, W.; Sletten, E.; Poltev, V. I.; Fedoroff, O. Y. *J. Biomol. Struct. Dyn.* **1993**, *11*, 671-683.

(14) Plavec, J.; Tong, W. M.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1993**, *115*, 9734-9746.

(15) Emsley, L.; Dwyer, T. J.; Spielmann, H. P.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 7765-7771.

(16) Carmichael, I.; Chipman, D. M.; Podlasek, C. A.; Serianni, A. S. *J. Am. Chem. Soc.* **1993**, *115*, 10863-10870.

(17) Oleary, D. J.; Kishi, Y. *J. Org. Chem.* **1994**, *59*, 6629-6636.

(18) Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **1994**, *116*, 8033-8037.

(19) Plavec, J.; Thibaudeau, C.; Chattopadhyaya, J. *Pure Appl. Chem.* **1996**, *68*, 2137-2144.

(20) Lam, S. L.; Auyeung, S. C. F. *J. Biomol. Struct. Dyn.* **1996**, *13*, 803-814.

(21) Luyten, I.; Thibaudeau, C.; Sandstrom, A.; J., C. *Tetrahedron* **1997**, *53*, 6433-6464.

(22) Murai, Y.; Shiroto, H.; Ishizaki, T.; Iimori, T.; Kodama, Y.; Ohtsuka, Y.; Oishi, T. *Heterocycles* **1992**, *33*, 391-404.

(23) Thewalt, U.; Bugg, C. E.; Marsh, R. E. *Acta Cryst., Sect. B* **1970**, *26*, 1089.

(24) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.

(25) Brameld, K. A.; Dasgupta, S.; Goddard, W. A., III *J. Phys. Chem. B* **1997**, *101*, 4851-4859.

(26) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205-8221.

(27) Tomimoto, M.; Go, N. *J. Phys. Chem.* **1995**, *99*, 563-577.

(28) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354-1358.

(29) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S. H.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 519-529.

(30) Broyde, S.; Wartell, R. M.; Stellman, S. D.; Hingerty, B. *Biopolymers* **1978**, *17*, 1485-1506.

(31) Pearlman, D. A.; Kim, S. H. *J. Biomol. Struct. Dyn.* **1985**, *3*, 99-125.

(32) Harvey, S. C.; Prabhakaran, M. *J. Am. Chem. Soc.* **1986**, *108*, 6128-6136.

(33) Schlick, T.; Peskin, C.; Broyde, S.; Overton, M. *J. Comp. Chem.* **1987**, *8*, 1199-1224.

(34) Gorin, A. A.; Ulyanov, N. B.; Zhurkin, V. B. *Mol. Biol.* **1990**, *24*, 1036-1047.

(35) Gabb, H. A.; Harvey, S. C. *J. Am. Chem. Soc.* **1993**, *115*, 4218-4227.

(36) Rudnicki, W. R.; Lesyng, B.; Harvey, S. C. *Biopolymers* **1994**, *34*, 383-392.

(37) Dobado, J. A.; Molina, J. M.; Espinosa, M. R. *J. Mol. Struct. (Theochem)* **1994**, *109*, 205-212.

(38) Pechenaya, V. I.; Rudnitsky, W.; Gritsuk, T.; Lesyng, B. *Mol. Biol.* **1995**, *29*, 616-627.

(39) Edward, J. T. *Chem. Ind. (London)* **1955**, 1102-1104.

(40) Thatcher, G. R. J. *The Anomeric Effect and Associated Stereoelectronic Effects*; American Chemical Society: Washington, DC, 1993.

(41) Kneisler, J. R.; Allinger, N. L. *J. Comp. Chem.* **1996**, *17*, 757-766.

(42) Murcko, M. A.; Dipaola, R. A. *J. Am. Chem. Soc.* **1992**, *114*, 10010-10018.

(43) Jaffe, R. L.; Smith, G. D.; Yoon, D. Y. *J. Phys. Chem.* **1993**, *97*, 12745-12751.

(44) Juaristi, E.; Cuevas, G. *The Anomeric Effect*; CRC Press: Boca Raton, 1995.

(45) Phillips, L.; Wray, V. *J. Chem. Soc. Chem. Comm.* **1973**, 90-91.

(46) Wiberg, K. B.; Murcko, M. A.; Laidig, K. E.; Macdougall, P. J. *J. Phys. Chem.* **1990**, *94*, 6956-6959.

(47) Warshel, A.; Levitt, M. *J. Am. Chem. Soc.* **1978**, *100*, 2607-2613.

(48) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

(49) Nilsson, L.; Karplus, M. *J. Comp. Chem.* **1986**, *7*, 591-616.

(50) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230-252.

(51) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(52) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(53) Cadioli, B.; Gallinella, E.; Coulombeau, C.; Jobic, H.; Berthier, G. *J. Phys. Chem.* **1993**, *97*, 7844-7856.

(54) Serianni, A. S.; Chipman, D. *J. Am. Chem. Soc.* **1987**, *109*, 5297-5303.

(55) Grana, A. M.; Pereiras, A. J.; Rios, M. A. *J. Mol. Stuct. (Theochem)* **1993**, *280*, 211-222.

(56) Ringnalda, M. N.; Langlois, J.-M.; Greeley, B. H.; Murphy, R. B.; Russo, T. V.; Cortis, C.; Muller, R. P.; Marten, B.; Donnelly, R. E.; Mainz, D. T.; Wright, J. R.; Pollard, W. T.; Cao, Y.; Won, Y.; Miller, G. H.; Goddard, W. A., III; Friesner, R. A. Jaguar 3.0 from Schrödinger, Inc., located in Portland, OR.

(57) Saebo, S.; Pulay, P. *Chem. Phys. Lett.* **1986**, *131*, 384-388.

(58) Saebo, S.; Pulay, P. *J. Chem. Phys.* **1988**, *88*, 1884-1890.

(59) Saebo, S.; Pulay, P. *Ann. Rev. Phys. Chem.* **1993**, *44*, 213-236.

(60) Saebo, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170-2175.

(61) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007-1023.

(62) Cerius^2 3.5 from Molecular Simulations, Inc., located in San Diego, CA.

(63) Dasgupta, S.; Brameld, K. A.; Fan, C. F.; Goddard, W. A., III *Spect. Chim. Acta. A* **1997**, *53*, 1347-1363.

(64) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(65) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.

(66) Maple, J. R.; Dinur, U.; Hagler, A. T. *Proc. Natl. Acad. Sci., USA* **1988**, *85*, 5350-5354.

(67) Dinur, U.; Hagler, A. T. *J. Comp. Chem.* **1990**, *11*, 1234-1246.

(68) Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comp. Chem.* **1994**, *15*, 162-182.

(69) Nagy, P. I.; Durant, G. J.; Smith, D. A. *Modeling the Hydrogen Bond* **1994**, *569*, 60-79.

(70) Shan, S. O.; Herschlag, D. *Proc. Natl. Acad. Sci., USA* **1996**, *93*, 14474-14479.

(71) Sarma, R. H. *Nucleic Acid Geometry and Dynamics*; Pergamon Press: New York, 1980.

(72) Young, P. R.; Kallenbach, N. R. *J. Mol. Biol.* **1978**, *126*, 467-479.

(73) Bolton, P. H.; Kearns, D. R. *Biochim. Biophys. Acta* **1978**, *517*, 329-337.

(74) Bolton, P. H.; Kearns, D. R. *J. Am. Chem. Soc.* **1979**, *101*, 479-484.

(75) Uesugi, S.; Miki, H.; Ikehara, M.; Iwahashi, H.; Kyogoku, Y. *Tetrahedron Lett.* **1979**, *42*, 4073-4076.

Chapter 7

MSCFF – A Complete Force Field for Nucleic Acids: Parameterization and Validation

Abstract

A new force field, MSCFF, is presented for the simulation of nucleic acids. Based upon the rule based generic force field UFF, the MSCFF has additional parameters for accurate simulations of nucleic acids. These additional parameters are fit from *ab initio* quantum mechanical calculations carried out on small model systems. Two of these model systems, base pair hydrogen bonding and sugar pseudorotation, are discussed in detail in Chapters 5 and 6 respectively. The remaining important substructures for nucleic acids include the phosphodiester backbone and glycosidic bond. Model systems are used to determine these parameters which complete the MSCFF. Extensive simulations have been carried out to test structural and energetic properties of the MSCFF. These validation studies demonstrate that the MSCFF is accurate and provide a benchmark against which other force fields may be compared.

I. Introduction

The development of force fields (FFs) for the simulation of biological macromolecules such as nucleic acids or proteins poses several challenges. Because the functions of these molecules are directly associated with their structures, it is of paramount importance that a FF accurately predict structures and energies. Furthermore, for nucleic acids in particular, these systems are quite complex with substructures including aromatic bases, ribose or 2'-deoxyribose sugars and a poly-anionic phosphodiester backbone. Current FFs such as Amber[1] and CHARMM[2] solve this problem by including specialized parameters for each of these substructures. The resulting force field then may be capable of accurately predicting the preferred gauche/gauche conformation of dimethylphosphate or the preferred sugar puckers of 2'-deoxyribose, but no properties for other chemical system which lie outside the realm of biological macromolecules. Therefore, a problem which arises with these highly parameterized FFs is a loss of generality.

Rarely do simulations involve pure nucleic acids or proteins. Typically, systems of interest include additional modifications such as glycosylation or phosphorylation, small ligands (drug molecules or cofactors), metals, or polymer scaffolds. It is desirable for a force field to be general enough to handle such additional molecules self-consistently with those detailed parameters developed specifically for nucleic acids or proteins. The final goal is parameterization of a biological force field (MSCFF) from first principles that maintains the flexibility of a generic FF. To this end, MSCFF is built upon the foundation of a generic FF (UFF[3]) with additional terms added which are specific to biomolecules. The first part of this FF which covers nucleic acids is presented

herein. These terms are derived from high level *ab initio* quantum mechanical (QM) calculations of small clusters which accurately represent the relevant potential energy surfaces present in a complete nucleic acid system.

The nature of these clusters or model systems is defined by the fundamental subunits of nucleic acids: base, sugar and phosphodiester backbone. For the bases, the key interactions involve hydrogen bonding. Chapter 5 presents a detailed examination of the hydrogen bond potentials for GC and AT base pairs. For the 2'-deoxyribose or ribose sugars, the critical issue is pseudorotation of the pentofuranose ring. Interconversion of sugar pucker conformations is critical in defining the overall helical parameters of DNA or RNA. The potential energies of pseudorotation and the conformational preferences of 2'-deoxyribose and ribose sugars are the subject of Chapter 6. Finally, this chapter presents the complete MSCFF that includes additional terms for the phosphodiester backbone and glycosidic bond. Here too are a series of detailed validation studies that demonstrate the accuracy of the FF for simulations ranging from isolated bases to a complete crystal structure of Z-DNA with explicit solvent molecules and counterions.

II. Methods

All *ab initio* QM calculations were carried out using the Jaguar 3.0 software package (or earlier versions).[4] The Cerius[2] 3.5 software package[5] was used for all molecular mechanics and dynamics simulations. A complete listing of force field parameters is provided in Appendix B. Atom types and charges are listed in the form of a conversion table for DNA and RNA in Appendices C and D respectively. Crystal simulations are carried out using full periodic boundary conditions and the Ewald summation method for all nonbond energies. All phosphate containing systems are made

neutral by the addition of Na^+ counterions. Additional solvent molecules are not included unless specified explicitly.

The MSCFF contains the following valence energy terms:

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion} \tag{7.1}$$

where

$$E_{bond} = \frac{1}{2} K_{IJ} (R - R_0)^2 \tag{7.2}$$

$$E_{angle} = \frac{1}{2} K_{IJK} (\theta - \theta_0)^2 \tag{7.3}$$

Periodic torsion potentials, which apply to four bonded atoms (IJKL), are represented by a Fourier series of the form

$$E_{torsion} = \frac{1}{2} \sum_{n=1}^{6} K_{\theta,n} (1 - d \cos(n\phi) \tag{7.4}$$

where $K_{\theta,n}$ is the torsion energy barrier, $d = \pm 1$ and is the phase factor and ϕ is the torsion angle ($\phi = 0$ for cis). For a given bond vector, every possible torsion angle is calculated and the resulting torsion energy is scaled by the total number of torsions. Thus for a torsion explicitly defined as A1–B–C–D1 in which there are 9 different torsions possible for the B–C bond, such as in the case of ethane (A1-A3 and D1-D3 are all hydrogen atoms), the contribution from the A1–B–C–D1 torsion is only 1/9 of the total for the B–C bond. Generic torsions include the X atom type which may be any atom. Finally, an inversion term is needed to enforce a planar geometry for sp^3 hybridized atoms, such as aromatic amines or the barrier to inversion for ammonia. For cases where the nonplanar geometry is the minimum:

$$E_{inversion} = \frac{1}{2} C_1 (\cos \omega - \cos \omega^0)^2 \qquad (7.5a)$$

where

$$C_1 = \frac{K_{inv}}{(\sin \omega^0)^2} \qquad (7.5b)$$

However, for cases where the planar geometry is the minimum:

$$E_{inversion} = K_{inv} (1 - \cos \omega) \qquad (7.6)$$

All valence parameters for MSCFF are originally derived from UFF, with additional torsion parameters added specifically for nucleic acids. The following new atom types are added: O_S (sugar O4' atom); C_S (sugar C3' atom); C_B (C8 or C6 for purines and pyrimidines); O_P (O1 and O2 phosphate oxygen atoms). These are necessary to reproduce the correct potential energies for pseudorotation, glycosidic bond rotation, and phosphate bond angles respectively.

Because UFF is a rule based FF with valence force constants which vary as a result of bond orders determined from electronegativities, it is possible for two bonds between the same pair of atom types to have slightly different force constants. In order to allow users who may not have access to the UFF force field generator to still utilize the MSCFF, the force constants for common atom types are averaged and reported in Appendix B. This is the parameter set for which all benchmarks are carried out. Use of the UFF generator is expected to yield similar, but not identical, results.

The Dreiding FF[6] exponential-6 parameters are used for all van der Waals (vdW) interactions and a standard coulombic potential completes the non-bond terms. A Morse potential is used for hydrogen bonding, as is described in Chapter 5. Fixed atomic

charges are determined from the electrostatic potential derived from the electron density distribution (constrained to reproduce the molecular monopole and dipole moments) calculated from the converged wavefunction of small model systems.[7]

Validation studies of the MSCFF provide a check on the accuracy of nonbond and valence terms. Heats of sublimation for free bases are calculated by

$$H_s = \Delta E_0 + \Delta H_{0\to 300} + \Delta ZPE \qquad (7.7)$$

where

$$\Delta E_0 = E_{gas} - E_{xtl} \qquad (7.8)$$

$\Delta H_{0\to 300}$ and ΔZPE are calculated from the vibrational frequencies calculated with the MSCFF. E_{xtl} is the crystal total energy determined using the Ewald nonbond method. E_{gas} is the gas phase energy with all nonbond terms included explicitly.

Simulations of all nucleic acid crystals start from a minimized structure for which the counterions and solvent molecules have been equilibrated for 10 ps (where applicable). The NPT simulations allow all lattice parameters to be movable. For smaller cases including A, U, T, AT, GC, ApU and GpC, 50 ps of dynamics was carried out. For the Z-DNA simulation, 200 ps of dynamics was carried out. Nucleic acid internal structural properties such as torsion angles are calculated with the Curves[8] program and visualized with Dials and Windows.[9]

III. Parameterization

The strategy employed herein for force field development begins with a simple generic force field (UFF[3]) which is then tuned to reproduce QM potential energies for small model systems relevant to nucleic acids. Torsion potentials are perhaps the most

important parameters for a biological force field which is not concerned with reproducing spectroscopic properties. All global changes in structure occur as local changes in torsion angles and, for this reason, these are the potentials which are tuned carefully in the MSCFF. Figure 7.1 shows the torsion angles important in nucleic acids. For each of these groups of torsions, a model system is used to determine the potential energy surface which in turn is fit by the FF. The hydrogen bond potential and parameterization of 2'-deoxyribose and ribose sugars are described in Chapters 5 and 6 respectively.

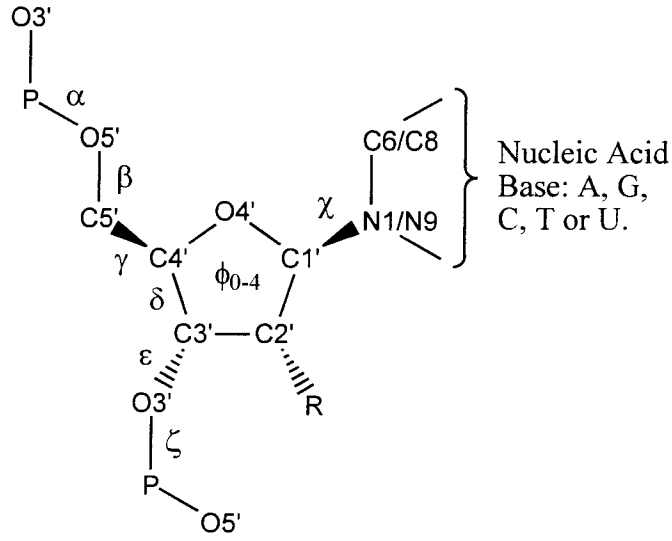


Figure 7.1. Definition of torsion angles important for nucleic acids. **R** is H– for 2'-deoxyribose and HO– for ribose sugars.

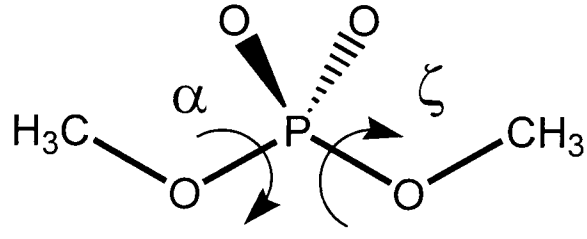


Figure 7.2. Dimethylphosphate is a model system for the phosphodiester backbone. Torsions α and ζ prefer a *gauche* conformation due to the anomeric effect.

Dimethylphosphate

Dimethylphosphate (DMP) is an excellent model system for the phosphodiester backbone torsions α and ζ (see Figure 7.2). As a result of the anomeric effect, the preferred conformation for these two torsion angles is not *trans/trans* but instead *gauche/gauche*. *Ab initio* QM calculations (MP2/6-31+G*//HF/6-31+G*)[10] of the DMP anion indicate that the *trans/trans* conformation is 3.66 kcal/mol higher in energy. UFF, with electrostatic potential derived charges, predicts the *trans/trans* conformation to be the global minimum and the *gauche/gauche* conformation to lie 2.20 kcal/mol higher in energy (Table 7.1). Clearly, this is an example of a case where an additional set of torsion parameters are required to obtain the correct conformational energies.

Table 7.1. Conformational Energies (kcal/mol) for Dimethylphosphate.

Method	*g/g*	*g/t*	*t/t*	*TS*[b]
QM (MP2/6-31+G*)[a]	0.00	1.45	3.66	2.26
UFF	2.20	1.10	0.00	3.30
MSCFF	0.00	1.51	3.15	2.31

[a]From reference 10. [b]Transition state between g/g and g/t.

Accompanying the standard X–P_3+3–O_3–X three-fold torsion term in UFF, a specific O_3– P_3+3–O_3–C_3 torsion with single-fold and two-fold terms is added. Both of these terms have a maximum at zero (*cis*) and help to disfavor the *trans* conformation. The resulting 2-D potential energy surface is shown in Figure 7.3. In

addition to fitting the local minimum geometries correctly, the barriers separating these local minima were also fit. For MSCFF, the global minimum is now the *gauche/gauche* conformation and the *trans/trans* conformation is 3.15 kcal/mol higher in energy.

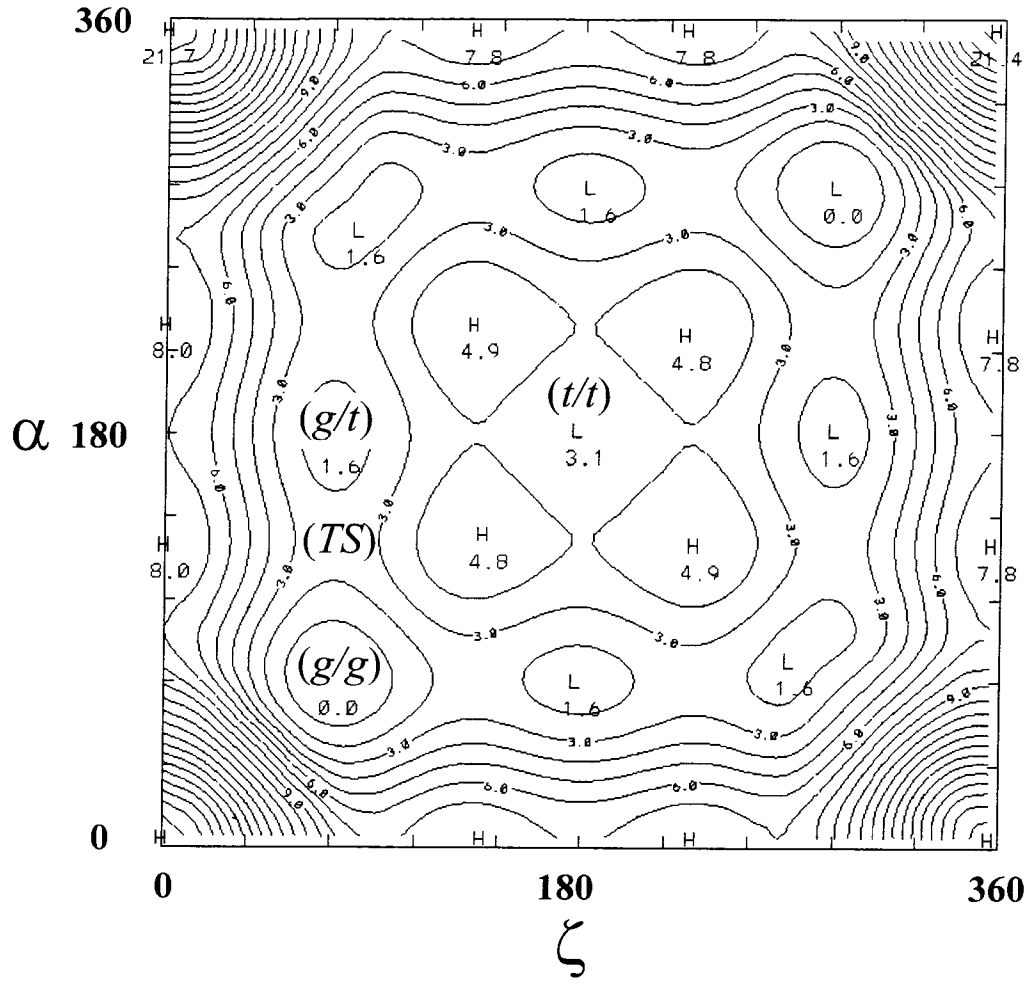


Figure 7.3. A contour map of the MSCFF predicted conformational energies for dimethylphosphate upon rotation of α and ζ torsion angles. *Gauche/gauche* (*g*/*g*) is the lowest energy conformation.

Glycosidic Bond

The 2'-deoxribose or ribose sugar and nucleic acid base are connected by a glycosidic bond. The conformation of the glycosidic bond is defined by the χ torsion

angle: O4'–C1'–N9–C4 for purine nucleosides and O4'–C1'–N1–C2 for pyrimidine nucleosides (Figure 7.1). The potential surface for rotation about χ is quite complex due to electronic and steric effects and is coupled to the pseudorotation pucker phase of the sugar.[11] In both helical A- and B-forms, χ is in an *anti* conformation (-150° to -160°). Whereas the free nucleosides may assume either an *anti* or *syn* conformation, with *anti* preferred for the pyrimidine bases and either conformation for the purine bases.

For the purposes of FF parameterization, conformational energies of different χ angles for the C2'-endo sugar conformation of dT and dA were determined by *ab initio* QM. Both the *syn* and *anti* conformations were completely geometry optimized (HF/6-31G**) followed by a single point energy (LMP2/cc-pVTZ(-f)) calculation. Other points along the potential surface were determined by constraining the χ angle and optimizing all other coordinates. The MSCFF is fit to reproduce the QM energies through the addition of a new torsion parameter: O_S–C_3–N_R–C_B where the C_B atom type is defined for atom C4 or C2 for purine and pyrimidine bases respectively. Figure 7.4 shows the complete MSCFF χ potential energy curve for dT and dA and the QM single point energies. There is good agreement between the FF and QM energies, including correct positioning of the local minima and relative energies which place the purine *syn* conformation lower in energy than the pyrimidine *syn* conformation.

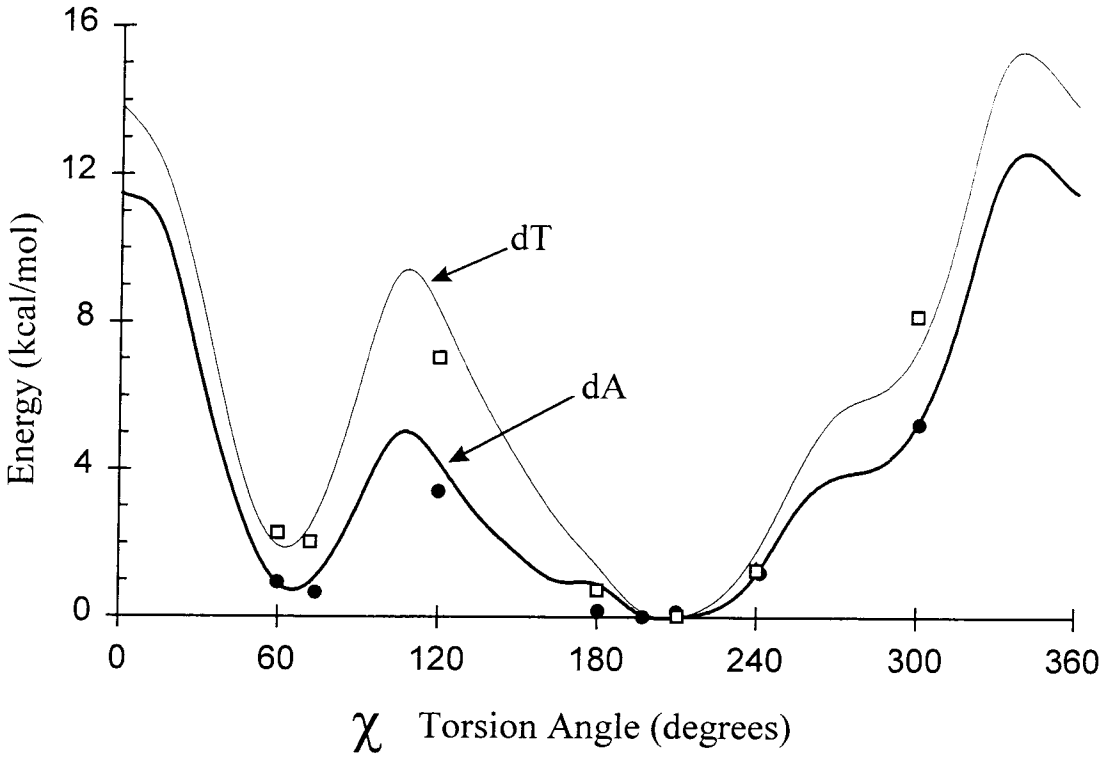


Figure 7.4. Conformational energies as determined by *ab initio* QM ((HF/6-31G**//LMP2/cc-pVTZ)) for the glycosidic bond of dA (●) and dT (□) compared to the MSCFF potential shown as solid lines.

Sugar Pseudorotation

Pseudorotation of the 2'-deoxyribose or ribose sugars involves the coupled rotations of torsions $\phi_{0\text{-}4}$ (Figure 7.1), the details of which are presented in Chapter 6. For 2'-deoxyribose, the global minimum conformation is C2'-endo, with C3'-endo 0.6 kcal/mol higher in energy. Ribose sugars, which have a C3' hydroxyl group, also prefer the C2'-endo conformation. In an RNA polymer that has no C3' hydroxyl group (instead having been replaced by a phosphate group), the preferred conformation is C3'-endo. The MSCFF reproduces these properties and correctly predicts the energy barriers between these local minima.

IV. Validation Studies

Validation of the MSCFF for nucleic acids is subdivided into two parts: nonbond parameters and valence parameters. Nonbond parameters which include hydrogen bonding, van der Waals (vdW), and electrostatic components are best tested with small crystals of free nucleic acid bases. Lattice dimensions and heats of sublimation offer rigorous tests of these nonbond components. Valence parameters are tested better using larger crystals that include a complete phosphodiester backbone. For these cases, lattice dimensions, total coordinate RMS difference and specific torsion angles are good measures of force field accuracy.

Nonbond Parameters

Five cases are used to test the nonbond parameters of the nucleic acid bases and include the following crystal structures: uracil,[12] 9-methyladenine,[13] 1-methylthymine,[14] 9-methyladenine:::1-methylthymine base pair,[15] and 9-ethylguanine:::1-methylcytosine base pair.[16] Three of these cases are isolated bases and two are hydrogen bonded base pairs. The AT base pair is in a Hoogsteen geometry while the GC base pair is in a Watson-Crick geometry. Molecular mechanics minimization, allowing all cell parameters to relax, results in small contraction of unit cell dimensions (Table 7.2). The overall agreement between the predicted and experimental structures is good with the average reduction of volume being only a few percent.

Table 7.2. Experimental and Predicted Lattice Properties for Nucleic Acid Bases.

	A	B	C	α	β	γ	Volume
9-methyladenine ::: 1-methylthymine							
Crystal	**8.304**	**6.552**	**12.837**	**90.00**	**106.83**	**90.00**	668.52
Minimization	8.231	6.572	12.810	90.00	108.59	90.00	656.80
Dynamics	8.389	7.460	12.723	84.60	106.95	86.50	756.93
Percent Error							
Minimization	-0.9	0.3	-0.2	0.0	1.6	0.0	-1.8
Dynamics	1.0	13.9	-0.9	-6.0	0.1	-3.9	13.2
9-ethylguanine ::: 1-methylcytosine							
Crystal	**8.838**	**11.106**	**7.391**	**107.82**	**87.05**	**91.45**	689.52
Minimization	8.635	10.834	7.485	107.82	87.87	91.45	666.10
Dynamics	10.524	11.451	9.058	118.14	106.56	75.53	893.38
Percent Error							
Minimization	-2.3	-2.4	1.3	0.0	0.9	0.0	-3.4
Dynamics	19.1	3.1	22.5	9.6	22.4	-17.4	29.6
Adenine							
Crystal	**7.670**	**12.240**	**8.470**	**90.00**	**123.43**	**90.00**	663.62
Minimization	8.018	11.922	8.244	90.00	125.77	90.00	639.40
Dynamics	8.015	12.298	8.524	89.97	124.34	90.53	693.69
Percent Error							
Minimization	4.5	-2.6	-2.7	0.0	1.9	0.0	-3.6
Dynamics	4.5	0.5	0.6	0.0	0.7	0.6	4.5
Thymine							
Crystal	**7.351**	**12.091**	**7.602**	**90.00**	**89.97**	**90.00**	675.67
Minimization	7.324	11.856	7.711	90.00	89.31	90.00	669.60
Dynamics	7.677	12.056	8.008	89.47	88.84	89.61	741.01
Percent Error							
Minimization	-0.4	-1.9	1.4	0.0	-0.7	0.0	-0.9
Dynamics	4.4	-0.3	5.3	-0.6	-1.3	-0.4	9.7
Uracil							
Crystal	**11.938**	**12.376**	**3.655**	**90.00**	**120.90**	**90.00**	463.39
Minimization	11.489	12.522	3.633	90.00	121.81	90.00	444.19
Dynamics	12.982	13.407	3.625	71.92	119.19	97.30	519.38
Percent Error							
Minimization	-3.8	1.2	-0.6	0.0	0.8	0.0	-4.1
Dynamics	8.7	8.3	-0.8	-20.1	-1.4	8.1	12.1

Molecular dynamics simulations for 50 ps at 300° K are a considerably more rigorous test of these crystal structure predictions than are simple minimization. Once again, all cell parameters are free to change (NPT dynamics). For these simulations, the average cell volume increases by approximately 10%. The average dynamical structures for two cases, GC and U, differ quite substantially from the experimental structure. In both of these systems, the bases are all parallel and the discrepancy arises from a shearing motion between the base planes. The experimental stacking distances are maintained, but the unit cell dimensions change considerably. Presumably, the energy barrier to shearing is quite low, particularly for uracil which has no exocylic amino or methyl group and is completely planar. This result may also be an artifact of the size of the unit cell which contains only 4 molecules in each case. Nevertheless, these results are comparable to other more specialized and narrowly parameterized force fields.[2]

Crystal unit cell parameters offer structural information but little energetic data. Nonbond energies may be tested by comparing experimental heats of sublimation to predicted values. The heat of sublimation is the change in enthalpy during the transition from the solid phase (crystal) to the gas phase. In addition to the change in total energy as determined from molecular mechanics, corrections need to be made for the change in vibrational energies during the phase transition (see Methods). Table 7.3 lists the experimental[17] and predicted heats of sublimation for 9-methyladenine, 1-methylthymine, and uracil. MSCFF gives excellent agreement with the experimental values. For comparison, heats of sublimation as determined with the Dreiding FF,[6] a generic FF, are also included. The improvements observed for MSCFF, which uses Dreiding vdW

parameters, is a result of a more accurate hydrogen bond potential and electrostatic potential derived charges.

Table 7.3. Experimental and Calculated Heats of Sublimation (kcal/mol) for Nucleic Acid Bases.

Base	Experimental[a]	MSCFF	Dreiding 2.21
9-methyladenine	**32.0**	32.3 (+0.3)	26.3 (-5.7)
1-methylthymine	**29.3**	29.1 (-0.2)	35.7 (+6.4)
uracil	**28.8**	27.5 (-1.3)	37.7 (+8.9)

[a]Experimental heats of sublimation from reference 17. Quantities in parentheses reflect the error between predicted and experimental values.

Valence Parameters

Two high resolution crystal structures of helical duplex dimers (ApU[18] and GpC[19]) are good systems to further test the MSCFF nonbond parameters and provide a check of the valence parameters, specifically the torsion potentials. The ApU and GpC crystals are the smallest cases that have a complete set of backbone torsion angles. In addition, these structures include water molecules and counterions. Average lattice parameters from a 50 ps NPT dynamics simulation at 300° K of ApU and GpC are in good agreement with the experimental structure and differ by just a few percent (Table 7.4). Average base pair hydrogen bond lengths measured during the simulation are reported in Table 7.5. The MSCFF hydrogen bond potential is fit from *ab initio* QM calculations[20] and the predicted hydrogen bond lengths are in very good agreement with the experimental structure.

Table 7.4. Experimental and Predicted Lattice Properties for ApU and GpC Crystals.

	A	B	C	α	β	γ	Volume
			ApU Crystal				
Crystal	**18.025**	**17.501**	**9.677**	**90.00**	**99.45**	**90.00**	**3011.2**
Dynamics	18.319	17.401	9.704	87.02	99.15	93.00	3045.5
			Percent Error				
Dynamics	1.6	-0.6	0.3	-3.3	-0.3	3.3	1.1
			GpC Crystal				
Crystal	**21.460**	**16.927**	**9.332**	**90.00**	**90.54**	**90.00**	**3389.7**
Dynamics	21.279	16.660	9.738	89.60	91.67	90.47	3450.4
			Percent Error				
Dynamics	-0.8	-1.6	4.4	-0.4	1.2	0.5	1.8

The ability of the valence parameters of the MSCFF to reproduce an experimental crystal structure may be quantified with a global measure, the coordinate RMS difference, or with internal coordinates, such as the average torsion angles during a simulation. Figure 7.5 plots the coordinate RMS as a function of time. The simulations converge within 5 to 10 ps to stable structures, which have an average coordinate RMS of 0.8 Å and 0.5 Å for ApU and GpC respectively. There is no significant difference between the NVT simulations, which have fixed cell parameters, and the NPT simulations with all cell parameters movable.

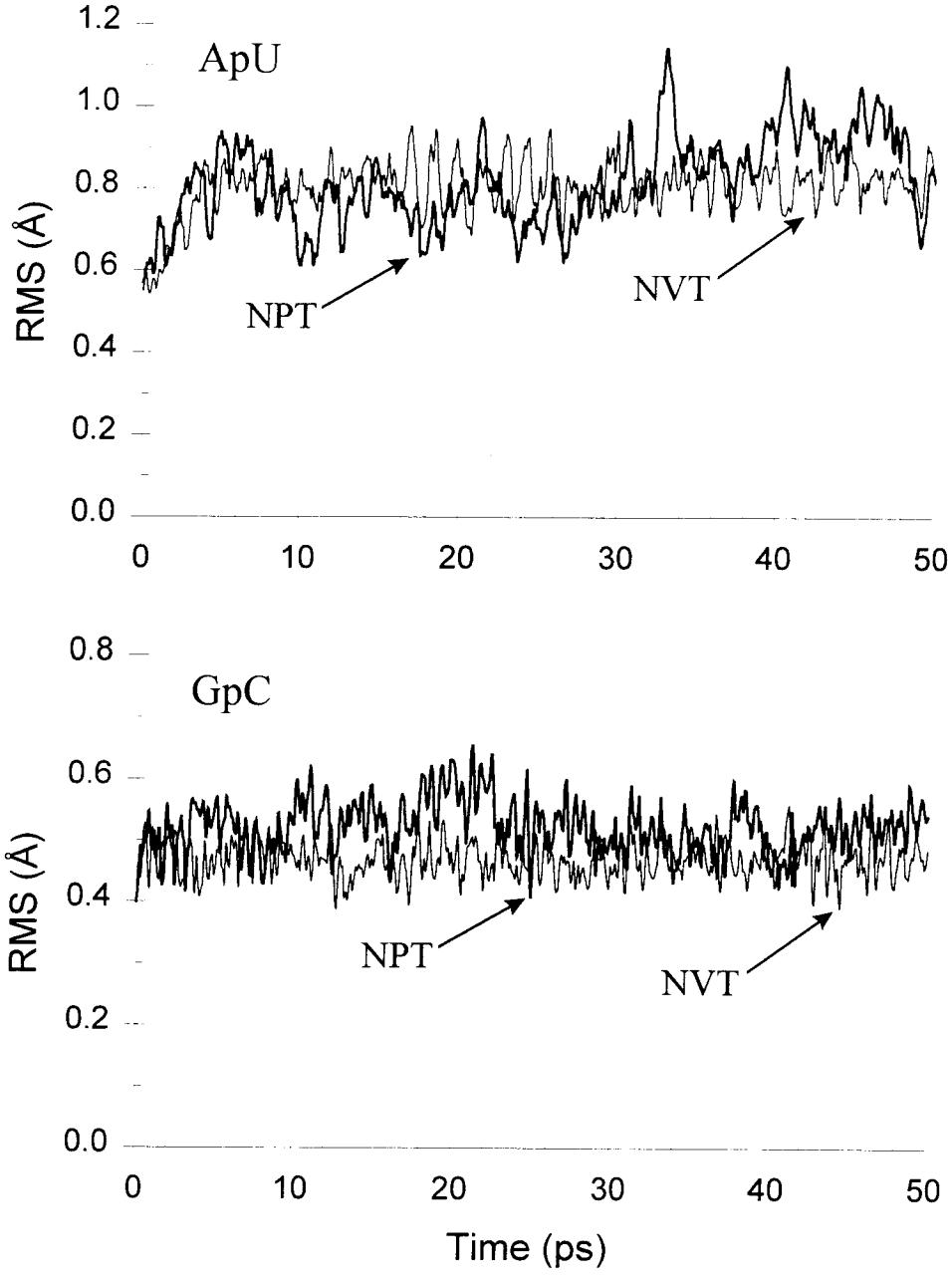


Figure 7.5. Coordinate RMS difference between the experimental crystal structure and molecular dynamics structure of ApU and GpC crystals.

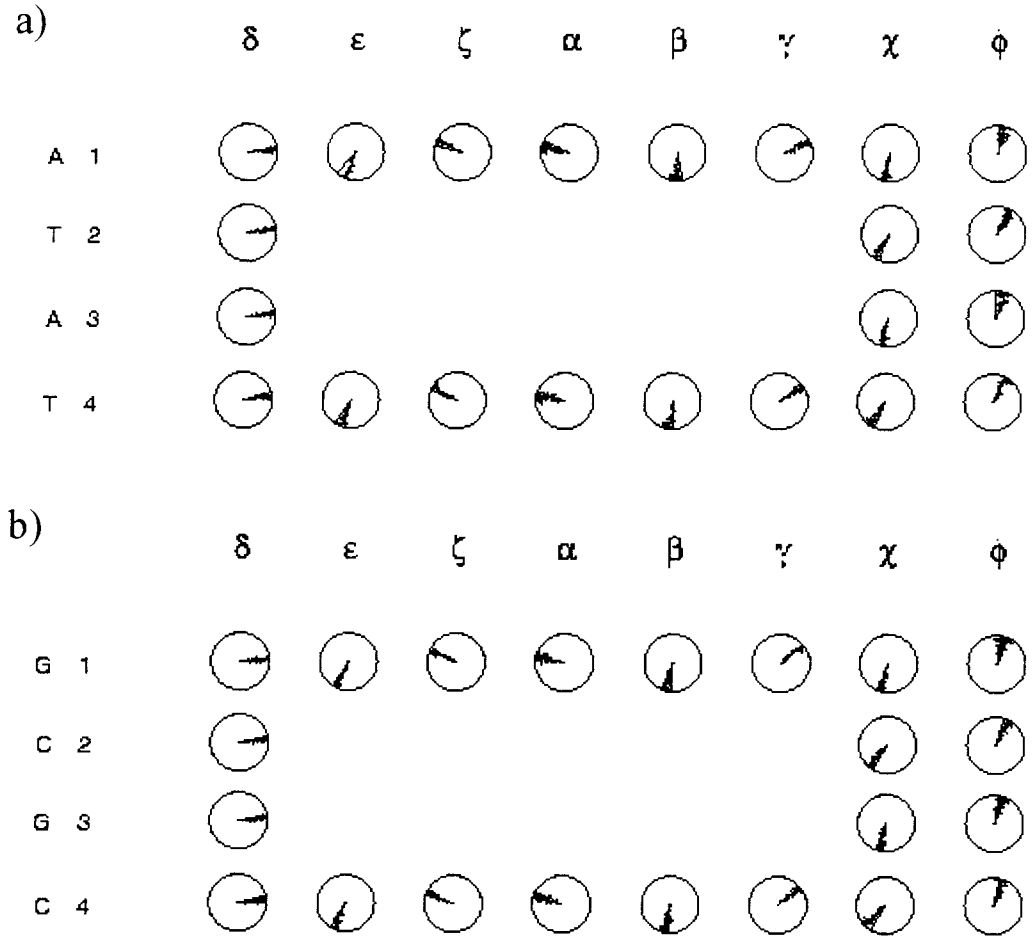


Figure 7.6. Torsion dials from molecular dynamics simulations of (a) ApU and (b) GpC crystals. The dials represent torsion angles from 0° to 360° (0° at the top of the dial). A solid line marks the crystal structure geometry while the molecular dynamics trajectory is plotted as a function of time, starting at the center and moving outward. In many instances, the molecular dynamics plot lies directly on the experimental crystal structure.

An ideal set of internal coordinates for nucleic acids are the backbone torsion angles (α, β, γ, δ, ε, and ζ), the glycosidic bond χ, and the sugar pseudorotation phase ϕ (Figure 7.1). These may be monitored during a dynamics simulation and plotted as a function of time using "Curves, Dials and Windows."[8,9] The MSCFF is fit to QM conformational energies for each of these torsion angles making these coordinates

particularly suitable for validation studies. Figure 7.6 presents results in the form of torsion "dials" from the ApU and GpC 50 ps NPT dynamics simulation. The experimental structure is shown as a solid line in the dial while each snapshot from the dynamics trajectory is plotted starting from T = 0 ps at the center and ending with T = 50 ps at the outer edge. In most cases, the superimposed plots are indistinguishable as the dynamics trajectory lies directly on the line that marks the experimental structure. The remarkable success of the MSCFF in reproducing the experimental crystal structures suggests that the valence parameters are accurately parameterized.

A considerably more challenging simulation is that of a non-standard DNA form with significant structural deviations from canonical A- and B-form duplexes. A Z-DNA system (5'-dCGCGCG) was chosen because a high resolution crystal structure has been reported[21] and the structure contains multiple sugar conformations (both C2'-endo and C3'-endo), a distorted backbone geometry and many solvent/counterion molecules. As was done for earlier Z-DNA simulations,[22] 32 water molecules and 8 Na counterions, which were undetermined in the crystal structure, were added to fill voids, neutralize the charge, and bring the total cell density to 1.42 g/cm^3.

A molecular dynamics simulation for 200 ps (NVT; 300° K) results in excellent agreement with the experimental structure. Figure 7.7 plots the coordinate RMS for all atoms and for DNA atoms alone. The all atom plot reveals considerable solvent diffusion as the RMS continues to increase over the entire 200 ps trajectory. In contrast, the DNA coordinate RMS stabilizes at 1.0 Å after only 15 ps. Furthermore, an analysis of the structurally important torsion angles (Figure 7.8) suggests little deviation from the experimental structure. For several bases, the sugar pucker oscillates between the

C3'-endo and C2'-endo local minima as can be expected given the low barrier and small energy difference between these conformations (see Chapter 6). Only two systematic discrepancies from the experimental structure are observed. The 3'-terminal guanine bases (G6 and G12) prefer a C3'-endo sugar conformation, while the C2'-endo conformation is observed experimentally. In addition, the highly unusual α and β torsion angles for bases G2 and C11 are not maintained. It is interesting to note that a survey of reported crystal structures turned up no other cases which share these α and β torsion values.[23] Indeed, all other reported crystal structures are similar to the values observed for the remaining bases and are correctly predicted by this simulation.

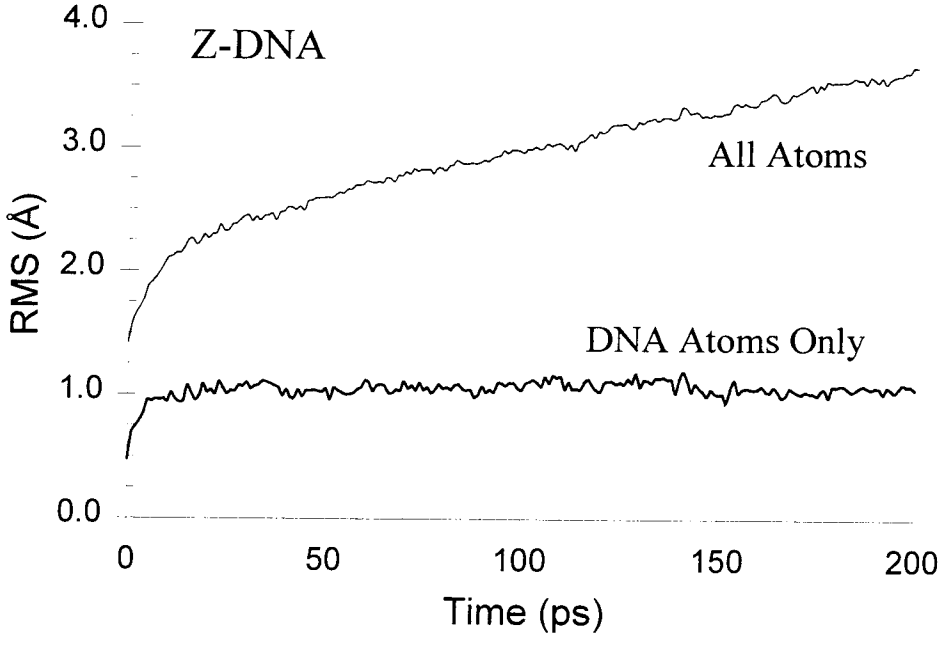


Figure 7.7. Coordinate RMS difference between the experimental crystal structure and molecular dynamics structure of Z-DNA. The all atom coordinate RMS does not reach a constant value due to solvent diffusion.

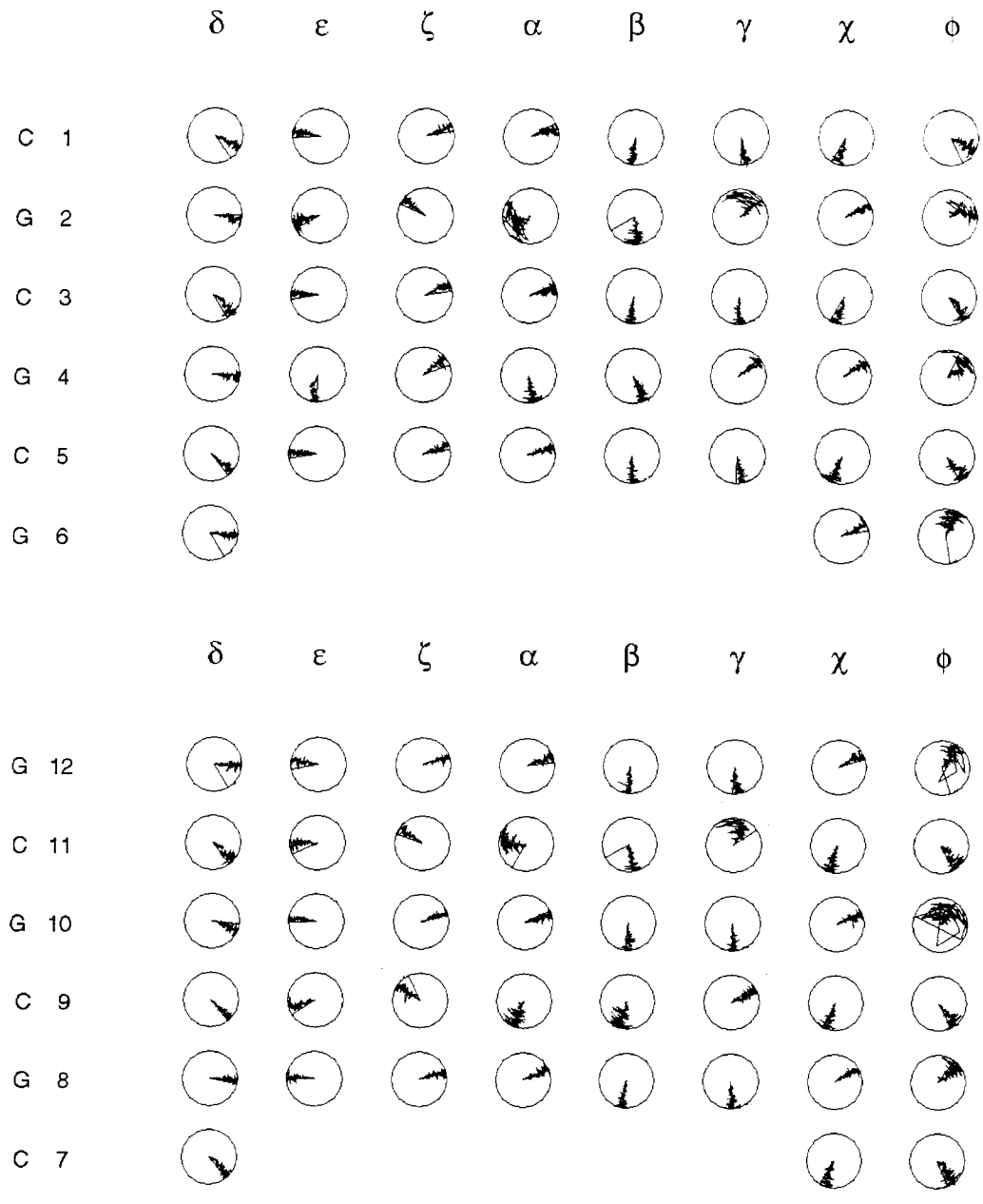


Figure 7.8. Torsion dials from molecular dynamics simulations of a Z-DNA crystal. See Figure 7.6 for comments regarding the dials representation.

V. Conclusion

A generic FF, with additional parameters fit from *ab initio* QM calculations, is presented for simulations of nucleic acids. Examples of the fitting procedure are given for dimethylphosphate, which is a model system for the α and ζ backbone torsion angles, and for the glycosidic bond. The hydrogen bond potential and sugar pseudorotation are presented elsewhere. Rigorous validation studies including simulations of free bases, base pair dimers and Z-DNA all suggest that the MSCFF is of equal if not superior quality to other highly parameterized FFs. The advantage of a generic FF, such as MSCFF, is the wide parameterization that encompasses a diverse range of possible chemistries in addition to nucleic acids. This work is part of an ongoing effort to parameterize a general FF for biological simulations.

VI. References

(1) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

(2) Mackerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946-11975.

(3) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024-10035.

(4) Ringnalda, M. N.; Langlois, J.-M.; Greeley, B. H.; Murphy, R. B.; Russo, T. V.; Cortis, C.; Muller, R. P.; Marten, B.; Donnelly, R. E.; Mainz, D. T.; Wright, J. R.; Pollard, W. T.; Cao, Y.; Won, Y.; Miller, G. H.; Goddard, W. A., III; Friesner, R. A. Jaguar 3.0 from Schrödinger, Inc., located in Portland, OR.

(5) Cerius^2 3.5 from Molecular Simulations, Inc., located in San Diego, CA.

(6) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897-8909.

(7) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875-11882.

(8) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63-91.

(9) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669-699.

(10) Liang, C.; Ewig, C. S.; Stouch, T. R.; Hagler, A. T. *J. Am. Chem. Soc.* **1993**, *115*, 1537-1545.

(11) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(12) Stewart, R. F.; Jensen, L. H. *Acta Crystallogr.* **1967**, *23*, 1102-1105.

(13) Stewart, R. F.; Jensen, L. H. *J. Chem. Phys.* **1964**, *40*, 2071-2075.

(14) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 28-38.

(15) Frey, M. N.; Koetzle, T. F.; Lehmann, M. S.; Hamilton, W. C. *J. Chem. Phys.* **1973**, *59*, 915-924.

(16) O'Brien, E. J. *Acta Crystallogr.* **1967**, *23*, 92-106.

(17) Glukhova, O. T. *Stu. Biophys.* **1984**, *101*, 25-26.

(18) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145-167.

(19) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Chem. Phys.* **1976**, *59*, 915-924.

(20) Brameld, K. A.; Dasgupta, S.; Goddard, W. A., III *J. Phys. Chem. B* **1997**, *101*, 4851-4859.

(21) Gessner, R. V.; Frederick, C. A.; Quigley, G. J.; Rich, A.; Wang, A. H. J. *J. Biol. Chem.* **1989**, *264*, 7921-7935.

(22) Lee, H.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *102*, 3830-3834.

(23) Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S. H.; Olson, W. K.; Berman, H. M. *J. Am. Chem. Soc.* **1996**, *118*, 519-529.

Appendix A

Sugar Force Field Parameters

Tables A1.1-A1.5 present the MSCFF parameters used for all sugar calculations in Chapter 6. The potential forms are described in the Methods section of Chapter 6.

Table A1.1. MSCFF Exponential-6 Van Der Waals Parameters.

Atom Type	R_0 (Å)	D_0 (kcal/mol)	ζ
H_	3.1950	0.0152	12.382
H___A	3.1950	0.0001	12.000
C_3	3.8983	0.0951	14.034
C_S	3.8983	0.0951	14.034
C_R	3.8983	0.0951	14.034
N_R	3.6621	0.0774	13.843
O_3	3.4046	0.0957	13.483
O_S	3.4046	0.0957	13.483
O_3::::H___A[a]	2.7500	0.2000	9.000

[a]A specific off-diagonal non-bond Morse potential for hydrogen bonding.[63]

Table A1.2. MSCFF Harmonic Bond Parameters.

Bond (*ij*)		K (kcal/mol)	R_0 (Å)
C_3	-H_	659.8	1.111
C_3	-C_3	699.6	1.514
C_S	-H_	659.8	1.111
C_S	-C_3	699.6	1.514
C_R	-H_	712.3	1.083
C_R	-C_R	951.8	1.407
N_R	-C_3	1046.5	1.456
N_R	-C_R	1359.3	1.373
O_3	-H___A	1128.4	0.988
O_3	-C_3	1030.8	1.415
O_3	-C_S	1030.8	1.415
O_S	-C_3	1030.8	1.415

Table A1.3. MSCFF Angle Parameters.

Angle (*ijk*)			K (kcal/mol)	θ_0 (°)
H_	-C_3	-H_	75.6	109.5
C_3	-C_3	-H_	117.4	109.5
C_S	-C_3	-H_	117.4	109.5
C_S	-C_3	-C_3	214.2	109.5
N_R	-C_3	-H_	170.1	109.5
N_R	-C_3	-C_3	303.6	109.5
O_3	-C_3	-H_	165.9	109.5
O_3	-C_3	-C_3	290.0	109.5
O_3	-C_3	-C_S	290.0	109.5
O_S	-C_3	-H_	165.9	109.5
O_S	-C_3	-C_3	290.0	109.5
O_S	-C_3	-C_S	290.0	109.5
O_S	-C_3	-N_R	413.5	109.5
C_3	-C_S	-H_	117.4	109.5
C_3	-C_S	-C_3	214.2	109.5
O_3	-C_S	-H_	165.9	109.5
O_3	-C_S	-C_3	290.0	109.5
C_R	-C_R	-H_	117.1	120.0
C_R	-C_R	-C_R	229.3	120.0
N_R	-C_R	-H_	162.4	120.0
N_R	-C_R	-C_R	314.5	120.0
C_R	-N_R	-C_3	216.8	120.0
C_R	-N_R	-C_R	247.2	120.0
C_3	-O_3	-H___A	174.3	109.5
C_S	-O_3	-H___A	174.3	109.5
C_3	-O_S	-C_3	297.8	112.0

Table A1.4. MSCFF Torsion Parameters.

Torsion (*ijkl*)				v0	v2	v3	v4	v6
X	-C_3	-C_3	-X	1.059		1.059		
X	-C_S	-C_3	-X	1.059		1.059		
X	-O_3	-C_3	-X	1.000		1.000		
X	-O_S	-C_3	-X	1.000		1.000		
X	-O_3	-C_S	-X	1.000		1.000		
X	-C_R	-C_R	-X	15.000	-15.000			
X	-N_R	-C_R	-X	15.000	-15.000			
X	-N_R	-C_3	-X	0.500				-0.500
H___A	-O_3	-C_3	-H_	4.000	3.000	1.000		
H___A	-O_3	-C_S	-H_	4.000	3.000	1.000		
C_3	-C_S	-C_3	-O_S	9.159	8.100	1.059		
O_3	-C_S	-C_3	-O_3	11.559	1.500	1.059	-9.000	
O_3	-C_S	-C_3	-O_S	11.559	1.500	1.059	-9.000	
O_3	-C_3	-C_3	-N_R	11.559	1.500	1.059	-9.000	
O_S	-C_3	-C_3	-O_3	11.559	1.500	1.059	-9.000	

Table A1.5. MSCFF Inversion Parameters.

Inversion (*cijk*)				K (kcal/mol)
N_R	-X	-X	-X	6.000
C_R	-X	-X	-X	6.000

Appendix B

Full Cerius2 Force Field Parameter Set for MSCFF

```
VERSION
 CERIUS2      1
END
#
HEADER
END
#
PREFERENCES
 BONDS                                  T
 ANGLES                                 T
 COULOMB                                T
 INVERSIONS                             T
 TORSIONS                               T
 UREY_BRADLEY                           F
 STRETCH_STRETCH                        F
 SEPARATED_STRETCH_STRETCH              F
 STRETCH_BEND_STRETCH                   F
 BEND_BEND                              F
 TORSION_STRETCH                        F
 TORSION_BEND_BEND                      F
 BEND_TORSION_BEND                      F
 STRETCH_TORSION_STRETCH                F
 HYDROGEN_BONDS                         F
 DIAGONAL_VAN_DER_WAALS                 T
 OFF_DIAGONAL_VAN_DER_WAALS             T
 GENERATE_UNDEFINED_TERMS               F
 IGNORE_UNDEFINED_TERMS                 F
 NON-BONDED_3-BODY                      F
 SHRINK_CH_BONDS                        F
 SHRINK_CH_H_ATOM                       H__C
 SHRINK_CH_FACTOR                          0.91500
 SINGLE_TORSION                         F
 SCALE_TORSIONS_ABOUT_COMMON_BOND       T
 SCALE_BY_N_DEFINED_TORSIONS            T
 EXOCYCLIC_TORSIONS_SCALE_FACTOR           1.00000
 SINGLE_INVERSION                       F
 H-BOND_METHOD                          SPLINE
 H-BOND_LIST                            T
 H-BOND_DIRECT_RCUT                        4.00000
 H-BOND_SPLINE_DISTANCE_ON                 4.00000
 H-BOND_SPLINE_DISTANCE_OFF                4.50000
 H-BOND_SPLINE_ANGLE_ON                   65.00000
 H-BOND_SPLINE_ANGLE_OFF                  75.00000
 H-BOND_LIST_DISTANCE_OFF                  6.50000
 H-BOND_LIST_ANGLE_OFF                    90.00000
 NON_BOND_LIST                          T
 NON_BOND_BUFFER_DISTANCE                  2.00000
```

```
  H-BOND_BUFFER_DISTANCE                    2.00000
  COU_DIELETRIC_CONSTANT                      1.00000000000000000
  COU_INTER_CUT_OFF                       100.00000
  COU_SPLINE_OFF                            8.50000
  COU_SPLINE_ON                             8.00000
  EWALD_SUM_COU_ACCURACY                    0.00100
  EWALD_SUM_COU_ETA                         2.50000
  EWALD_SUM_COU_KCUT                        0.50000
  EWALD_SUM_COU_RCUT                        6.00000
  EWALD_SUM_COU_AUTO_OPT                  T
  COU_EXCLUDE_1-2                         T
  COU_EXCLUDE_1-3                         T
  COU_EXCLUDE_1-4                         F
  COU_1-4_SCALE_FACTOR                      1.00000
  COU_METHOD                              DIRECT
  COU_DIRECT_CUT-OFF                        8.00000
  VDW_COMBINATION_RULE                    ARITHMETIC
  VDW_INTER_CUT_OFF                        100.00000
  VDW_EXCLUDE_1-2                         T
  VDW_EXCLUDE_1-3                         T
  VDW_EXCLUDE_1-4                         F
  VDW_1-4_SCALE_FACTOR                      1.00000
  VDW_METHOD                              DIRECT
  VDW_SPLINE_ON                             8.00000
  VDW_SPLINE_OFF                            8.50000
  EWALD_SUM_VDW_AUTO_OPT                  T
  EWALD_SUM_VDW_ACCURACY                    0.00100
  EWALD_SUM_VDW_ETA                         2.50000
  EWALD_SUM_VDW_KCUT                        0.50000
  EWALD_SUM_VDW_RCUT                        6.00000
  EWALD_SUM_VDW_REP_CUT                     6.00000
  FAST_EWALD_SUM_RATIO                     10.00000
  SLOW_EWALD_SUM_RATIO                      5.00000
  MINIMUM_IMAGE                           F
  ASSIGN_MASS                             T
  ASSIGN_CHARGE                           F
  ASSIGN_HYBRIDIZATION                    F
  ATOM_TYPE                               F
  ATOM_TYPE_ALL                           T
  CALCULATE_BOND_ORDER                    F
END
#
ATOMTYPES
  H_          H        1.00790   0.0000    0    0    0
  H___A       H        1.00790   0.0000    0    0    0
  C_3         C       12.01100   0.0000    3    0    0
  C_S         C       12.01100   0.0000    3    0    0
  C_2         C       12.01100   0.0000    2    0    0
  C_R         C       12.01100   0.0000    2    0    0
  C_B         C       12.01100   0.0000    2    0    0
  C_1         C       12.01100   0.0000    1    0    0
  N_3         N       14.00670   0.0000    3    0    1
  N_2         N       14.00670   0.0000    2    0    1
  N_R         N       14.00670   0.0000    2    0    1
```

```
 N_RL        N       14.00670   0.0000    2    0    1
 N_1         N       14.00670   0.0000    1    0    1
 O_3         O       15.99940   0.0000    3    0    2
 O_S         O       15.99940   0.0000    3    0    2
 O_2         O       15.99940   0.0000    2    0    2
 O_P         O       15.99940   0.0000    2    0    2
 O_R         O       15.99940   0.0000    2    0    1
 O_1         O       15.99940   0.0000    1    0    1
 P_3         P       30.97376   0.0000    3    0    1
 P_3+3       P       30.97376   0.0000    3    0    0
 S_31        S       33.06790   0.0000    3    1    2
 S_3         S       32.06000   0.0000    3    0    2
 Cl          Cl      35.45300  -1.0000    0    0    3
 Na          Na      22.98977   1.0000    0    0    0
 Ca          Ca      40.08000   2.0000    0    0    0
 NaH         Na      22.98977   1.0000    0    0    0
 ClH         Cl      35.45300  -1.0000    0    0    3
 H__OJ       H        1.00800   0.0000    0    0    0
 O_3_J       O       15.99940   0.0000    3    0    2
 Mg          Mg      24.30500   2.0000    0    0    0
 F_          F       18.99840   0.0000    0    0    0
END
#
DIAGONAL_VDW
 H_          EXPO_6          3.1950   0.1520E-01   12.3820
 H___A       EXPO_6          3.1950   0.1000E-03   12.0000
 C_3         EXPO_6          3.8983   0.9510E-01   14.0340
 C_S         EXPO_6          3.8983   0.9510E-01   14.0340
 C_2         EXPO_6          3.8983   0.9510E-01   14.0340
 C_R         EXPO_6          3.8983   0.9510E-01   14.0340
 C_B         EXPO_6          3.8983   0.9510E-01   14.0340
 C_1         EXPO_6          3.8983   0.9510E-01   14.0340
 N_3         EXPO_6          3.6621   0.7740E-01   13.8430
 N_2         EXPO_6          3.6621   0.7740E-01   13.8430
 N_R         EXPO_6          3.6621   0.7740E-01   13.8430
 N_RL        EXPO_6          3.6621   0.7740E-01   13.8430
 N_1         EXPO_6          3.6621   0.7740E-01   13.8430
 O_3         EXPO_6          3.4046   0.9570E-01   13.4830
 O_S         EXPO_6          3.4046   0.9570E-01   13.4830
 O_2         EXPO_6          3.4046   0.9570E-01   13.4830
 O_P         EXPO_6          3.4046   0.9570E-01   13.4830
 O_R         EXPO_6          3.4046   0.9570E-01   13.4830
 O_1         EXPO_6          3.4046   0.9570E-01   13.4830
 P_3         EXPO_6          4.1500   0.3200E+00   12.0000
 P_3+3       EXPO_6          4.1500   0.3200E+00   12.0000
 S_31        EXPO_6          4.0300   0.3440E+00   12.0000
 S_3         EXPO_6          4.0300   0.3440E+00   12.0000
 Cl          EXPO_6          3.9503   0.2833E+00   13.8610
 Na          EXPO_6          2.8500   0.5000E+00   12.0000
 Ca          EXPO_6          3.4720   0.5000E-01   12.0000
 NaH         EXPO_6          5.0000   0.1000E+00   12.0000
 ClH         EXPO_6          5.0000   0.1000E+00   12.0000
 H__OJ       EXPO_6          3.1950   0.1000E-03   12.3800
 O_3_J       EXPO_6          3.5364   0.1521E+00   13.4830
```

```
 Mg          EXPO_6          2.5549  0.3649E+00    12.0000
 F_          EXPO_6          3.3640  0.5000E-01    13.4830
END
#
ATOM_TYPING_RULES
END
#
#
OFF_DIAGONAL_VDW
 N_RL      H___A       VDW_MORSE       2.7000  0.2000E+00     9.4000
 O_3       H___A       VDW_MORSE       2.7500  0.2000E+00     9.0000
 O_2       H___A       VDW_MORSE       2.5500  0.2000E+00     9.0000
END
#
BOND_STRETCH
 C_3       H_          HARMONIC      659.7507    1.1110
 C_3       C_3         HARMONIC      699.5920    1.5140
 C_S       H_          HARMONIC      659.7507    1.1110
 C_S       C_3         HARMONIC      699.5920    1.5140
 C_R       H_          HARMONIC      712.2570    1.0830
 C_R       C_3         HARMONIC      739.8881    1.4690
 C_R       C_R         HARMONIC      951.8002    1.4074
 C_B       H_          HARMONIC      712.2570    1.0830
 C_B       C_R         HARMONIC      951.8002    1.3664
 N_3       C_3         HARMONIC      700.0000    1.4620
 N_2       H___A       HARMONIC     1030.9469    1.0830
 N_2       C_R         HARMONIC     1359.2566    1.3726
 N_R       H_          HARMONIC     1030.9469    1.0830
 N_R       H___A       HARMONIC     1030.9469    1.0830
 N_R       C_3         HARMONIC     1046.4963    1.4560
 N_R       C_R         HARMONIC     1359.2566    1.3726
 N_R       C_B         HARMONIC     1359.2566    1.3726
 N_RL      C_3         HARMONIC     1046.4963    1.4560
 N_RL      C_R         HARMONIC     1359.2566    1.3726
 N_RL      C_B         HARMONIC     1359.2566    1.3726
 O_3       H_          HARMONIC     1050.0039    1.0120
 O_3       H___A       HARMONIC     1128.3962    0.9880
 O_3       C_3         HARMONIC     1030.7742    1.4150
 O_3       C_S         HARMONIC     1030.7742    1.4150
 O_S       H_          HARMONIC     1050.0039    1.0120
 O_S       C_3         HARMONIC     1030.7742    1.4150
 O_2       C_3         HARMONIC     1085.0542    1.3910
 O_2       C_R         HARMONIC     1621.0470    1.2440
 P_3+3     O_3         HARMONIC      803.3712    1.7590
 P_3+3     O_P         HARMONIC     1050.0000    1.4800
 P_3+3     O_R         HARMONIC     1082.4950    1.5930
 O_3_J     H__OJ       HARMONIC      700.0000    0.9800
 F_        C_3         HARMONIC      761.3979    1.4250
 F_        C_S         HARMONIC      761.3979    1.4250
END
#
ANGLE_BEND
 H_        C_3       H_          COS_HARMON    75.6040  109.4710
 C_3       C_3       H_          COS_HARMON   117.3990  109.4710
```

C_3	C_3	C_3	COS_HARMON	214.2065	109.4710
C_S	C_3	H_	COS_HARMON	117.3990	109.4710
C_S	C_3	C_3	COS_HARMON	214.2065	109.4710
C_R	C_3	H_	COS_HARMON	121.8610	109.4710
C_B	C_3	H_	COS_HARMON	121.8610	109.4710
N_3	C_3	H_	COS_HARMON	170.1120	109.4710
N_3	C_3	C_3	COS_HARMON	214.2065	109.4710
N_R	C_3	H_	COS_HARMON	170.1120	109.4710
N_R	C_3	C_3	COS_HARMON	303.5660	109.4710
N_RL	C_3	H_	COS_HARMON	170.1120	109.4710
N_RL	C_3	C_3	COS_HARMON	303.5660	109.4710
O_3	C_3	H_	COS_HARMON	165.8800	109.4710
O_3	C_3	C_3	COS_HARMON	290.0060	109.4710
O_3	C_3	C_S	COS_HARMON	290.0060	109.4710
O_3	C_3	N_R	COS_HARMON	413.4660	109.4710
O_3	C_3	N_RL	COS_HARMON	413.4660	109.4710
O_S	C_3	H_	COS_HARMON	165.8800	109.4710
O_S	C_3	C_3	COS_HARMON	290.0060	109.4710
O_S	C_3	C_S	COS_HARMON	290.0060	109.4710
O_S	C_3	N_3	COS_HARMON	413.4660	109.4710
O_S	C_3	N_R	COS_HARMON	413.4660	109.4710
O_S	C_3	N_RL	COS_HARMON	413.4660	109.4710
F_	C_3	H_	COS_HARMON	119.8158	109.4710
C_3	C_S	H_	COS_HARMON	117.3990	109.4710
C_3	C_S	C_3	COS_HARMON	214.2065	109.4710
O_3	C_S	H_	COS_HARMON	165.8800	109.4710
O_3	C_S	C_3	COS_HARMON	290.0060	109.4710
F_	C_S	H_	COS_HARMON	119.8158	109.4710
F_	C_S	C_3	COS_HARMON	212.2212	109.4710
H_	C_R	H_	COS_HARMON	117.0535	120.0000
C_R	C_R	H_	COS_HARMON	117.0535	120.0000
C_R	C_R	C_3	COS_HARMON	202.4230	120.0000
C_R	C_R	C_R	COS_HARMON	229.2697	120.0000
C_B	C_R	H_	COS_HARMON	117.0535	120.0000
C_B	C_R	C_3	COS_HARMON	202.4230	120.0000
C_B	C_R	C_R	COS_HARMON	229.2697	120.0000
N_2	C_R	C_R	COS_HARMON	314.5356	120.0000
N_R	C_R	H_	COS_HARMON	162.4060	120.0000
N_R	C_R	C_R	COS_HARMON	314.5356	120.0000
N_R	C_R	N_2	COS_HARMON	433.7220	120.0000
N_R	C_R	N_R	COS_HARMON	433.7220	120.0000
N_RL	C_R	H_	COS_HARMON	162.4060	120.0000
N_RL	C_R	C_R	COS_HARMON	314.5356	120.0000
N_RL	C_R	N_2	COS_HARMON	433.7220	120.0000
N_RL	C_R	N_R	COS_HARMON	433.7220	120.0000
N_RL	C_R	N_RL	COS_HARMON	433.7220	120.0000
O_2	C_R	C_R	COS_HARMON	326.3435	120.0000
O_2	C_R	N_R	COS_HARMON	451.2690	120.0000
O_2	C_R	N_RL	COS_HARMON	451.2690	120.0000
H_	C_B	H_	COS_HARMON	117.0535	120.0000
C_R	C_B	H_	COS_HARMON	117.0535	120.0000
N_R	C_B	H_	COS_HARMON	162.4060	120.0000
N_R	C_B	C_R	COS_HARMON	314.5356	120.0000
N_R	C_B	N_R	COS_HARMON	433.7220	120.0000

```
 N_RL      C_B       H_          COS_HARMON   162.4060   120.0000
 N_RL      C_B       C_R         COS_HARMON   314.5356   120.0000
 N_RL      C_B       N_R         COS_HARMON   433.7220   120.0000
 H___A     N_3       H___A       COS_HARMON    71.3950   109.4710
 H___A     N_2       H___A       COS_HARMON    71.3950   120.0000
 C_3       N_2       H___A       COS_HARMON   126.4483   109.4710
 C_R       N_2       H___A       COS_HARMON   126.4483   120.0000
 H___A     N_R       H___A       COS_HARMON    71.3950   120.0000
 C_3       N_R       H___A       COS_HARMON   126.4483   120.0000
 C_3       N_R       C_3         COS_HARMON   126.4483   120.0000
 C_R       N_R       H_          COS_HARMON   126.4483   120.0000
 C_R       N_R       H___A       COS_HARMON   126.4483   120.0000
 C_R       N_R       C_3         COS_HARMON   216.7528   120.0000
 C_R       N_R       C_R         COS_HARMON   247.1814   120.0000
 C_B       N_R       H___A       COS_HARMON   126.4483   120.0000
 C_B       N_R       C_3         COS_HARMON   216.7528   120.0000
 C_B       N_R       C_R         COS_HARMON   247.1814   120.0000
 C_R       N_RL      C_R         COS_HARMON   247.1814   120.0000
 C_B       N_RL      C_R         COS_HARMON   247.1814   120.0000
 H___A     O_3       H___A       COS_HARMON   175.0000   109.5000
 C_3       O_3       H_          COS_HARMON   174.2860   109.5000
 C_3       O_3       H___A       COS_HARMON   174.2860   109.5000
 C_3       O_3       C_3         COS_HARMON   311.6010   109.5000
 C_S       O_3       H_          COS_HARMON   174.2860   109.5000
 C_S       O_3       H___A       COS_HARMON   174.2860   109.5000
 P_3+3     O_3       C_3         COS_HARMON   321.2860   120.0000
 P_3+3     O_3       C_S         COS_HARMON   321.2860   120.0000
 C_3       O_S       H_          COS_HARMON   165.6001   104.5100
 C_3       O_S       C_3         COS_HARMON   297.8331   112.0000
 O_3       P_3+3     O_3         COS_HARMON   200.0000   104.5100
 O_P       P_3+3     O_3         COS_HARMON   200.0000   108.2000
 O_P       P_3+3     O_P         COS_HARMON   280.0000   119.9000
 H__OJ     O_3_J     H__OJ       COS_HARMON   320.0000   104.5100
END
#
TORSIONS
 X       C_3   C_3   X       DIHEDRAL     2.1190    3.0000   -1.0000
 O_3     C_3   C_3   N_R     DIHEDRAL     3.0000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
                             18.0000    4.0000    1.0000
 O_S     C_3   C_3   C_3     DIHEDRAL    16.2000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
 O_S     C_3   C_3   O_3     DIHEDRAL     3.0000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
                             18.0000    4.0000    1.0000
 X       C_S   C_3   X       DIHEDRAL     2.1190    3.0000   -1.0000
 C_3     C_S   C_3   O_S     DIHEDRAL    16.2000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
 O_3     C_S   C_3   O_3     DIHEDRAL     3.0000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
                             18.0000    4.0000    1.0000
 O_3     C_S   C_3   O_S     DIHEDRAL     3.0000    2.0000   -1.0000
                              2.1190    3.0000   -1.0000
                             18.0000    4.0000    1.0000
```

```
 C_B     N_R   C_3   O_S      DIHEDRAL    10.0000    1.0000   -1.0000
                                          30.0000    2.0000   -1.0000
                                           7.0000    3.0000    1.0000
                                          10.0000    5.0000   -1.0000
 H___A   O_3   C_3   H_       DIHEDRAL     6.0000    2.0000   -1.0000
                                           2.0000    3.0000   -1.0000
 H___A   O_3   C_S   H_       DIHEDRAL     6.0000    2.0000   -1.0000
                                           2.0000    3.0000   -1.0000
 X       C_R   C_3   X        DIHEDRAL     1.0000    6.0000    1.0000
 X       C_R   C_R   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       C_B   C_3   X        DIHEDRAL     1.0000    6.0000    1.0000
 X       C_B   C_R   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       N_3   C_3   X        DIHEDRAL     2.1190    3.0000   -1.0000
 X       N_2   C_R   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       N_R   C_3   X        DIHEDRAL     1.0000    6.0000    1.0000
 X       N_R   C_R   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       N_R   C_B   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       N_RL  C_R   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       N_RL  C_B   X        DIHEDRAL    30.0000    2.0000    1.0000
 X       O_3   C_3   X        DIHEDRAL     2.0000    3.0000   -1.0000
 X       O_3   C_S   X        DIHEDRAL     2.0000    3.0000   -1.0000
 X       O_S   C_3   X        DIHEDRAL     2.0000    3.0000   -1.0000
 X       P_3+3 O_3   X        DIHEDRAL     3.6000    3.0000   -1.0000
 O_3     P_3+3 O_3   C_3      DIHEDRAL     2.7022    1.0000   -1.0000
                                           8.6752    2.0000   -1.0000
END
#
INVERSIONS
 C_R     X       X       X          UMBRELLA      6.0000    0.0000
 C_R     O_2     X       X          UMBRELLA     50.0000    0.0000
 C_R     X       X       O_2        UMBRELLA     50.0000    0.0000
 C_B     X       X       X          UMBRELLA      6.0000    0.0000
 N_2     X       X       X          UMBRELLA      6.0000    0.0000
 N_R     X       X       X          UMBRELLA      6.0000    0.0000
 N_RL    X       X       X          UMBRELLA      6.0000    0.0000
END
#
COULOMBIC
 X       X          CONST-EPS
END
```

Appendix C

MSCFF Conversion Table for DNA Atom Types and Charges

```
C*************************************************************************
C  MSCFF
C  Conversion table for DNA.
C  12/10/97
C
C  Record format is 1X,A3, 1X,A7, 1X,A7, 1X,I2, 1X,I2, 1X,F8.4 and
C  corresponds to:
C
C  RES OLDTYP NEWTYP MB NL CHARGE
C
C  where RES     = residue name
C        OLDTYP = old atom type (type on input data file)
C        NEWTYP = new atom type (type used in calculations)
C        MB     = maximum number of covalent connections (for building)
C        NL     = number of lone pairs at MB (for H-bond acceptors)
C        CHARGE = charge on atom (used in electrostatic calculation)
C
C*************************************************************************
*  BASES
   A  N9     N_R      3  0  -0.2260
   A N9      N_R      3  0  -0.2260
   A  C8     C_B      3  0   0.1570
   A C8      C_B      3  0   0.1570
   A HC8     H_       1  0   0.1710
   A  N7     N_RL     2  1  -0.5540
   A N7      N_RL     2  1  -0.5540
   A  C5     C_R      3  0  -0.0950
   A C5      C_R      3  0  -0.0950
   A  C6     C_R      3  0   0.8370
   A C6      C_R      3  0   0.8370
   A  N6     N_2      3  0  -0.9850
   A N6      N_2      3  0  -0.9850
   A HN6     H___A    1  0   0.4360
   A  N1     N_RL     2  1  -0.8450
   A N1      N_RL     2  1  -0.8450
   A  C2     C_R      3  0   0.6700
   A C2      C_R      3  0   0.6700
   A HC2     H_       1  0   0.0290
   A  N3     N_RL     2  1  -0.8000
   A N3      N_RL     2  1  -0.8000
   A  C4     C_R      3  0   0.4290
   A C4      C_R      3  0   0.4290
C  Sugar and Phosphate conversions for deoxy!
   A P       P_3+3    3  1   1.3250
   A  P      P_3+3    3  1   1.3250
   A O1P     O_P      1  2  -0.7770
   A  O1P    O_P      1  2  -0.7770
```

```
   A O2P      O_P       1  2  -0.7770
   A  O2P     O_P       1  2  -0.7770
   A O5*      O_3       2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   A  O5*     O_3       2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   A C5*      C_3       4  0   0.0800
   A  C5*     C_3       4  0   0.0800
   A HC5*     H_        1  0   0.1000
   A C4*      C_3       4  0   0.3900
   A  C4*     C_3       4  0   0.3900
   A HC4*     H_        1  0   0.0200
   A O4*      O_S       2  2  -0.4900
   A  O4*     O_S       2  2  -0.4900
   A C3*      C_S       4  0   0.3500
   A  C3*     C_S       4  0   0.3500
   A HC3*     H_        1  0  -0.0100
   A O3*      O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   A  O3*     O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   A C2*      C_3       4  0  -0.2600
   A  C2*     C_3       4  0  -0.2600
   A HC2*     H_        1  0   0.1000
   A  C1*     C_3       4  0   0.1300
   A C1*      C_3       4  0   0.1300
   A HC1*     H_        1  0   0.1100
   A HO3*     H_        1  0   0.4200
   A HO5*     H_        1  0   0.3600
*
   C N1       N_R       3  0  -0.4500
   C  N1      N_R       3  0  -0.4500
   C C2       C_R       3  0   0.8440
   C  C2      C_R       3  0   0.8440
   C O2       O_2       1  2  -0.6070
   C  O2      O_2       1  2  -0.6070
   C N3       N_RL      2  1  -0.8480
   C  N3      N_RL      2  1  -0.8480
   C C4       C_R       3  0   1.0630
   C  C4      C_R       3  0   1.0630
   C N4       N_2       3  0  -1.0810
   C  N4      N_2       3  0  -1.0810
   C HN4      H___A     1  0   0.4540
   C C5       C_R       3  0  -0.7950
   C  C5      C_R       3  0  -0.7950
   C HC5      H_        1  0   0.2310
   C C6       C_B       3  0   0.2360
   C  C6      C_B       3  0   0.2360
   C HC6      H_        1  0   0.1590
C  Sugar and Phosphate conversions for deoxy!
   C P        P_3+3     3  1   1.3250
   C  P       P_3+3     3  1   1.3250
   C O1P      O_P       1  2  -0.7770
   C  O1P     O_P       1  2  -0.7770
   C O2P      O_P       1  2  -0.7770
   C  O2P     O_P       1  2  -0.7770
   C O5*      O_3       2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   C  O5*     O_3       2  2  -0.4850  !change to OH (-0.4600) 5' terminal
```

```
    C C5*      C_3       4  0   0.0800
    C  C5*     C_3       4  0   0.0800
    C HC5*     H_        1  0   0.1000
    C C4*      C_3       4  0   0.3900
    C  C4*     C_3       4  0   0.3900
    C HC4*     H_        1  0   0.0200
    C O4*      O_S       2  2  -0.4900
    C  O4*     O_S       2  2  -0.4900
    C C3*      C_S       4  0   0.3500
    C  C3*     C_S       4  0   0.3500
    C HC3*     H_        1  0  -0.0100
    C O3*      O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
    C  O3*     O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
    C C2*      C_3       4  0  -0.2600
    C  C2*     C_3       4  0  -0.2600
    C HC2*     H_        1  0   0.1000
    C  C1*     C_3       4  0   0.1300
    C C1*      C_3       4  0   0.1300
    C HC1*     H_        1  0   0.1100
    C HO3*     H_        1  0   0.4200
    C HO5*     H_        1  0   0.3600
*
    G N9       N_R       3  0  -0.1460
    G  N9      N_R       3  0  -0.1460
    G C8       C_B       3  0   0.0500
    G  C8      C_B       3  0   0.0500
    G HC8      H_        1  0   0.1940
    G N7       N_RL      2  1  -0.5010
    G  N7      N_RL      2  1  -0.5010
    G C5       C_R       3  0   0.0130
    G  C5      C_R       3  0   0.0130
    G C6       C_R       3  0   0.6420
    G  C6      C_R       3  0   0.6420
    G O6       O_2       1  2  -0.5260
    G  O6      O_2       1  2  -0.5260
    G N1       N_R       3  0  -0.7720
    G  N1      N_R       3  0  -0.7720
    G HN1      H___A     1  0   0.3930
    G C2       C_R       3  0   1.0450
    G  C2      C_R       3  0   1.0450
    G N2       N_2       3  0  -1.0540
    G  N2      N_2       3  0  -1.0540
    G HN2      H___A     1  0   0.4330
    G N3       N_RL      2  1  -0.7660
    G  N3      N_RL      2  1  -0.7660
    G C4       C_R       3  0   0.2220
    G  C4      C_R       3  0   0.2220
C   Sugar and Phosphate conversions for deoxy!
    G P        P_3+3     3  1   1.3250
    G  P       P_3+3     3  1   1.3250
    G O1P      O_P       1  2  -0.7770
    G  O1P     O_P       1  2  -0.7770
    G O2P      O_P       1  2  -0.7770
    G  O2P     O_P       1  2  -0.7770
```

```
   G O5*     O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   G  O5*    O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   G C5*     C_3      4  0   0.0800
   G  C5*    C_3      4  0   0.0800
   G HC5*    H_       1  0   0.1000
   G C4*     C_3      4  0   0.3900
   G  C4*    C_3      4  0   0.3900
   G HC4*    H_       1  0   0.0200
   G O4*     O_S      2  2  -0.4900
   G  O4*    O_S      2  2  -0.4900
   G C3*     C_S      4  0   0.3500
   G  C3*    C_S      4  0   0.3500
   G HC3*    H_       1  0  -0.0100
   G O3*     O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   G  O3*    O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   G C2*     C_3      4  0  -0.2600
   G  C2*    C_3      4  0  -0.2600
   G HC2*    H_       1  0   0.1000
   G  C1*    C_3      4  0   0.1300
   G C1*     C_3      4  0   0.1300
   G HC1*    H_       1  0   0.1100
   G HO3*    H_       1  0   0.4200
   G HO5*    H_       1  0   0.3600
*
   T N1      N_R      3  0  -0.2670
   T  N1     N_R      3  0  -0.2670
   T C2      C_R      3  0   0.6560
   T  C2     C_R      3  0   0.6560
   T O2      O_2      1  2  -0.5900
   T  O2     O_2      1  2  -0.5900
   T N3      N_R      3  0  -0.6430
   T  N3     N_R      3  0  -0.6430
   T HN3     H___A    1  0   0.3970
   T C4      C_R      3  0   0.6640
   T  C4     C_R      3  0   0.6640
   T O4      O_2      1  2  -0.5400
   T  O4     O_2      1  2  -0.5400
   T C5      C_R      3  0  -0.0320
   T  C5     C_R      3  0  -0.0320
   T C5A     C_3      4  0  -0.4240
   T  C5A    C_3      4  0  -0.4240
   T HC5A    H_       1  0   0.1320
   T  C6     C_B      3  0  -0.2070
   T C6      C_B      3  0  -0.2070
   T HC6     H_       1  0   0.2500
C  Sugar and Phosphate conversions for deoxy!
   T P       P_3+3    3  1   1.3250
   T  P      P_3+3    3  1   1.3250
   T O1P     O_P      1  2  -0.7770
   T  O1P    O_P      1  2  -0.7770
   T O2P     O_P      1  2  -0.7770
   T  O2P    O_P      1  2  -0.7770
   T O5*     O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
   T  O5*    O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
```

```
   T C5*      C_3      4  0   0.0800
   T  C5*     C_3      4  0   0.0800
   T HC5*     H_       1  0   0.1000
   T C4*      C_3      4  0   0.3900
   T  C4*     C_3      4  0   0.3900
   T HC4*     H_       1  0   0.0200
   T O4*      O_S      2  2  -0.4900
   T  O4*     O_S      2  2  -0.4900
   T C3*      C_S      4  0   0.3500
   T  C3*     C_S      4  0   0.3500
   T HC3*     H_       1  0  -0.0100
   T O3*      O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   T  O3*     O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   T C2*      C_3      4  0  -0.2600
   T  C2*     C_3      4  0  -0.2600
   T HC2*     H_       1  0   0.1000
   T  C1*     C_3      4  0   0.1300
   T C1*      C_3      4  0   0.1300
   T HC1*     H_       1  0   0.1100
   T HO3*     H_       1  0   0.4200
   T HO5*     H_       1  0   0.3600
*
 5MU N1       N_R      3  0  -0.2670
 5MU  N1      N_R      3  0  -0.2670
 5MU C2       C_R      3  0   0.6560
 5MU  C2      C_R      3  0   0.6560
 5MU O2       O_2      1  2  -0.5900
 5MU  O2      O_2      1  2  -0.5900
 5MU N3       N_R      3  0  -0.6430
 5MU  N3      N_R      3  0  -0.6430
 5MU HN3      H___A    1  0   0.3970
 5MU C4       C_R      3  0   0.6640
 5MU  C4      C_R      3  0   0.6640
 5MU O4       O_2      1  2  -0.5400
 5MU  O4      O_2      1  2  -0.5400
 5MU C5       C_R      3  0  -0.0320
 5MU  C5      C_R      3  0  -0.0320
 5MU C5A      C_3      4  0  -0.4240
 5MU  C5A     C_3      4  0  -0.4240
 5MU HC5A     H_       1  0   0.1320
 5MU  C6      C_B      3  0  -0.2070
 5MU C6       C_B      3  0  -0.2070
 5MU HC6      H_       1  0   0.2500
C  Sugar and Phosphate conversions for deoxy!
 5MU P        P_3+3    3  1   1.3250
 5MU  P       P_3+3    3  1   1.3250
 5MU O1P      O_P      1  2  -0.7770
 5MU  O1P     O_P      1  2  -0.7770
 5MU O2P      O_P      1  2  -0.7770
 5MU  O2P     O_P      1  2  -0.7770
 5MU O5*      O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
 5MU  O5*     O_3      2  2  -0.4850  !change to OH (-0.4600) 5' terminal
 5MU C5*      C_3      4  0   0.0800
 5MU  C5*     C_3      4  0   0.0800
```

```
 5MU HC5*    H_       1  0   0.1000
 5MU C4*     C_3      4  0   0.3900
 5MU  C4*    C_3      4  0   0.3900
 5MU HC4*    H_       1  0   0.0200
 5MU O4*     O_S      2  2  -0.4900
 5MU  O4*    O_S      2  2  -0.4900
 5MU C3*     C_S      4  0   0.3500
 5MU  C3*    C_S      4  0   0.3500
 5MU HC3*    H_       1  0  -0.0100
 5MU O3*     O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
 5MU  O3*    O_3      2  2  -0.6660  !change to OH (-0.7000) 3' terminal
 5MU C2*     C_3      4  0  -0.2600
 5MU  C2*    C_3      4  0  -0.2600
 5MU HC2*    H_       1  0   0.1000
 5MU  C1*    C_3      4  0   0.1300
 5MU C1*     C_3      4  0   0.1300
 5MU HC1*    H_       1  0   0.1100
 5MU HO3*    H_       1  0   0.4200
 5MU HO5*    H_       1  0   0.3600
*
   U N1      N_R      3  0  -0.2830
   U  N1     N_R      3  0  -0.2830
   U C2      C_R      3  0   0.6670
   U  C2     C_R      3  0   0.6670
   U O2      O_2      1  2  -0.5880
   U  O2     O_2      1  2  -0.5880
   U N3      N_R      3  0  -0.6350
   U  N3     N_R      3  0  -0.6350
   U HN3     H___A    1  0   0.3860
   U C4      C_R      3  0   0.7920
   U  C4     C_R      3  0   0.7920
   U O4      O_2      1  2  -0.5650
   U  O4     O_2      1  2  -0.5650
   U C5      C_R      3  0  -0.5400
   U  C5     C_R      3  0  -0.5400
   U HC5     H_       1  0   0.2180
   U C6      C_B      3  0   0.0160
   U  C6     C_B      3  0   0.0160
   U HC6     H_       1  0   0.1920
C  Sugar and Phosphate conversions for deoxy!
   U P       P_3+3    3  1   1.3250
   U  P      P_3+3    3  1   1.3250
   U O1P     O_P      1  2  -0.7770
   U  O1P    O_P      1  2  -0.7770
   U O2P     O_P      1  2  -0.7770
   U  O2P    O_P      1  2  -0.7770
   U O5*     O_3      2  2  -0.4850  !change to OH (-0.6500) 5' terminal
   U  O5*    O_3      2  2  -0.4850  !change to OH (-0.6500) 5' terminal
   U C5*     C_3      4  0   0.0800
   U  C5*    C_3      4  0   0.0800
   U HC5*    H_       1  0   0.1000
   U C4*     C_3      4  0   0.3900
   U  C4*    C_3      4  0   0.3900
   U HC4*    H_       1  0   0.0200
```

```
   U O4*      O_S       2  2  -0.4900
   U  O4*     O_S       2  2  -0.4900
   U C3*      C_S       4  0   0.3500
   U  C3*     C_S       4  0   0.3500
   U HC3*     H_        1  0  -0.0100
   U O3*      O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   U  O3*     O_3       2  2  -0.6660  !change to OH (-0.7000) 3' terminal
   U C2*      C_3       4  0  -0.2600
   U  C2*     C_3       4  0  -0.2600
   U HC2*     H_        1  0   0.1000
   U  C1*     C_3       4  0   0.1300
   U C1*      C_3       4  0   0.1300
   U HC1*     H_        1  0   0.1100
   U HO3*     H_        1  0   0.4200
   U HO5*     H_        1  0   0.3600
*
 HOH OH       O_3_J     2  2  -0.8340
 HOH OW       O_3_J     2  2  -0.8340
 HOH O        O_3_J     2  2  -0.8340
 HOH  OH      O_3_J     2  2  -0.8340
 HOH  OW      O_3_J     2  2  -0.8340
 HOH  O       O_3_J     2  2  -0.8340
 HOH HO       H__OJ     1  0   0.4170
 HOH HOW      H__OJ     1  0   0.4170
 HOH H        H__OJ     1  0   0.4170
 HOH  HO      H__OJ     1  0   0.4170
 HOH  HOW     H__OJ     1  0   0.4170
 HOH  H       H__OJ     1  0   0.4170
```

Appendix D

MSCFF Conversion Table for RNA Atom Types and Charges

```
C*********************************************************************
C  MSCFF
C  Conversion table for RNA.
C  12/10/97
C
C  Record format is 1X,A3, 1X,A7, 1X,A7, 1X,I2, 1X,I2, 1X,F8.4 and
C  corresponds to:
C
C  RES OLDTYP NEWTYP MB NL CHARGE
C
C  where RES    = residue name
C        OLDTYP = old atom type (type on input data file)
C        NEWTYP = new atom type (type used in calculations)
C        MB     = maximum number of covalent connections (for building)
C        NL     = number of lone pairs at MB (for H-bond acceptors)
C        CHARGE = charge on atom (used in electrostatic calculation)
C
C*********************************************************************
*  BASES
   A  N9     N_R      3  0  -0.2260
   A N9      N_R      3  0  -0.2260
   A  C8     C_B      3  0   0.1570
   A C8      C_B      3  0   0.1570
   A HC8     H_       1  0   0.1710
   A  N7     N_RL     2  1  -0.5540
   A N7      N_RL     2  1  -0.5540
   A  C5     C_R      3  0  -0.0950
   A C5      C_R      3  0  -0.0950
   A  C6     C_R      3  0   0.8370
   A C6      C_R      3  0   0.8370
   A  N6     N_2      3  0  -0.9850
   A N6      N_2      3  0  -0.9850
   A HN6     H___A    1  0   0.4360
   A  N1     N_RL     2  1  -0.8450
   A N1      N_RL     2  1  -0.8450
   A  C2     C_R      3  0   0.6700
   A C2      C_R      3  0   0.6700
   A HC2     H_       1  0   0.0290
   A  N3     N_RL     2  1  -0.8000
   A N3      N_RL     2  1  -0.8000
   A  C4     C_R      3  0   0.4290
   A C4      C_R      3  0   0.4290
C  Sugar and Phosphate conversions for deoxy!
   A P       P_3+3    3  1   1.3250
   A  P      P_3+3    3  1   1.3250
   A O1P     O_P      1  2  -0.7770
   A  O1P    O_P      1  2  -0.7770
```

```
    A O2P      O_P       1  2  -0.7770
    A  O2P     O_P       1  2  -0.7770
    A O5*      O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
    A  O5*     O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
    A C5*      C_3       4  0   0.0800
    A  C5*     C_3       4  0   0.0800
    A HC5*     H_        1  0   0.1000
    A C4*      C_3       4  0   0.4600
    A  C4*     C_3       4  0   0.4600
    A HC4*     H_        1  0   0.0200
    A O4*      O_S       2  2  -0.5300
    A  O4*     O_S       2  2  -0.5300
    A C3*      C_S       4  0   0.1600
    A  C3*     C_S       4  0   0.1600
    A HC3*     H_        1  0   0.0200
    A O3*      O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
    A  O3*     O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
    A C2*      C_3       4  0   0.3200
    A  C2*     C_3       4  0   0.3200
    A HC2*     H_        1  0  -0.0100
    A O2*      O_3       2  2  -0.7000
    A  O2*     O_3       2  2  -0.7000
    A HO2*     H___A     1  0   0.4400
    A  C1*     C_3       4  0   0.1500
    A C1*      C_3       4  0   0.1500
    A HC1*     H_        1  0   0.1100
    A HO3*     H_        1  0   0.4600
    A HO5*     H_        1  0   0.3920
*
    C N1       N_R       3  0  -0.4500
    C  N1      N_R       3  0  -0.4500
    C C2       C_R       3  0   0.8440
    C  C2      C_R       3  0   0.8440
    C O2       O_2       1  2  -0.6070
    C  O2      O_2       1  2  -0.6070
    C N3       N_RL      2  1  -0.8480
    C  N3      N_RL      2  1  -0.8480
    C C4       C_R       3  0   1.0630
    C  C4      C_R       3  0   1.0630
    C N4       N_2       3  0  -1.0810
    C  N4      N_2       3  0  -1.0810
    C HN4      H___A     1  0   0.4540
    C C5       C_R       3  0  -0.7950
    C  C5      C_R       3  0  -0.7950
    C HC5      H_        1  0   0.2310
    C C6       C_B       3  0   0.2360
    C  C6      C_B       3  0   0.2360
    C HC6      H_        1  0   0.1590
C   Sugar and Phosphate conversions for deoxy!
    C P        P_3+3     3  1   1.3250
    C  P       P_3+3     3  1   1.3250
    C O1P      O_P       1  2  -0.7770
    C  O1P     O_P       1  2  -0.7770
    C O2P      O_P       1  2  -0.7770
```

```
   C  O2P    O_P      1  2  -0.7770
   C O5*     O_3      2  2  -0.4850  !change to OH (-0.5120) 5' terminal
   C  O5*    O_3      2  2  -0.4850  !change to OH (-0.5120) 5' terminal
   C C5*     C_3      4  0   0.0800
   C  C5*    C_3      4  0   0.0800
   C HC5*    H_       1  0   0.1000
   C C4*     C_3      4  0   0.4600
   C  C4*    C_3      4  0   0.4600
   C HC4*    H_       1  0   0.0200
   C O4*     O_S      2  2  -0.5300
   C  O4*    O_S      2  2  -0.5300
   C C3*     C_S      4  0   0.1600
   C  C3*    C_S      4  0   0.1600
   C HC3*    H_       1  0   0.0200
   C O3*     O_3      2  2  -0.6660  !change to OH (-0.7200) 3' terminal
   C  O3*    O_3      2  2  -0.6660  !change to OH (-0.7200) 3' terminal
   C C2*     C_3      4  0   0.3200
   C  C2*    C_3      4  0   0.3200
   C HC2*    H_       1  0  -0.0100
   C O2*     O_3      2  2  -0.7000
   C  O2*    O_3      2  2  -0.7000
   C HO2*    H___A    1  0   0.4400
   C  C1*    C_3      4  0   0.1500
   C C1*     C_3      4  0   0.1500
   C HC1*    H_       1  0   0.1100
   C HO3*    H_       1  0   0.4600
   C HO5*    H_       1  0   0.3920
*
   G N9      N_R      3  0  -0.1460
   G  N9     N_R      3  0  -0.1460
   G C8      C_B      3  0   0.0500
   G  C8     C_B      3  0   0.0500
   G HC8     H_       1  0   0.1940
   G N7      N_RL     2  1  -0.5010
   G  N7     N_RL     2  1  -0.5010
   G C5      C_R      3  0   0.0130
   G  C5     C_R      3  0   0.0130
   G C6      C_R      3  0   0.6420
   G  C6     C_R      3  0   0.6420
   G O6      O_2      1  2  -0.5260
   G  O6     O_2      1  2  -0.5260
   G N1      N_R      3  0  -0.7720
   G  N1     N_R      3  0  -0.7720
   G HN1     H___A    1  0   0.3930
   G C2      C_R      3  0   1.0450
   G  C2     C_R      3  0   1.0450
   G N2      N_2      3  0  -1.0540
   G  N2     N_2      3  0  -1.0540
   G HN2     H___A    1  0   0.4330
   G N3      N_RL     2  1  -0.7660
   G  N3     N_RL     2  1  -0.7660
   G C4      C_R      3  0   0.2220
   G  C4     C_R      3  0   0.2220
C  Sugar and Phosphate conversions for deoxy!
```

```
   G P        P_3+3     3  1    1.3250
   G  P       P_3+3     3  1    1.3250
   G O1P      O_P       1  2   -0.7770
   G  O1P     O_P       1  2   -0.7770
   G O2P      O_P       1  2   -0.7770
   G  O2P     O_P       1  2   -0.7770
   G O5*      O_3       2  2   -0.4850  !change to OH (-0.5120) 5' terminal
   G  O5*     O_3       2  2   -0.4850  !change to OH (-0.5120) 5' terminal
   G C5*      C_3       4  0    0.0800
   G  C5*     C_3       4  0    0.0800
   G HC5*     H_        1  0    0.1000
   G C4*      C_3       4  0    0.4600
   G  C4*     C_3       4  0    0.4600
   G HC4*     H_        1  0    0.0200
   G O4*      O_S       2  2   -0.5300
   G  O4*     O_S       2  2   -0.5300
   G C3*      C_S       4  0    0.1600
   G  C3*     C_S       4  0    0.1600
   G HC3*     H_        1  0    0.0200
   G O3*      O_3       2  2   -0.6660  !change to OH (-0.7200) 3' terminal
   G  O3*     O_3       2  2   -0.6660  !change to OH (-0.7200) 3' terminal
   G C2*      C_3       4  0    0.3200
   G  C2*     C_3       4  0    0.3200
   G HC2*     H_        1  0   -0.0100
   G O2*      O_3       2  2   -0.7000
   G  O2*     O_3       2  2   -0.7000
   G HO2*     H___A     1  0    0.4400
   G  C1*     C_3       4  0    0.1500
   G C1*      C_3       4  0    0.1500
   G HC1*     H_        1  0    0.1100
   G HO3*     H_        1  0    0.4600
   G HO5*     H_        1  0    0.3920
*
   T N1       N_R       3  0   -0.2670
   T  N1      N_R       3  0   -0.2670
   T C2       C_R       3  0    0.6560
   T  C2      C_R       3  0    0.6560
   T O2       O_2       1  2   -0.5900
   T  O2      O_2       1  2   -0.5900
   T N3       N_R       3  0   -0.6430
   T  N3      N_R       3  0   -0.6430
   T HN3      H___A     1  0    0.3970
   T C4       C_R       3  0    0.6640
   T  C4      C_R       3  0    0.6640
   T O4       O_2       1  2   -0.5400
   T  O4      O_2       1  2   -0.5400
   T C5       C_R       3  0   -0.0320
   T  C5      C_R       3  0   -0.0320
   T C5A      C_3       4  0   -0.4240
   T  C5A     C_3       4  0   -0.4240
   T HC5A     H_        1  0    0.1320
   T  C6      C_B       3  0   -0.2070
   T C6       C_B       3  0   -0.2070
   T HC6      H_        1  0    0.2500
```

```
C  Sugar and Phosphate conversions for deoxy!
   T P       P_3+3     3  1   1.3250
   T  P      P_3+3     3  1   1.3250
   T O1P     O_P       1  2  -0.7770
   T  O1P    O_P       1  2  -0.7770
   T O2P     O_P       1  2  -0.7770
   T  O2P    O_P       1  2  -0.7770
   T O5*     O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
   T  O5*    O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
   T C5*     C_3       4  0   0.0800
   T  C5*    C_3       4  0   0.0800
   T HC5*    H_        1  0   0.1000
   T C4*     C_3       4  0   0.4600
   T  C4*    C_3       4  0   0.4600
   T HC4*    H_        1  0   0.0200
   T O4*     O_S       2  2  -0.5300
   T  O4*    O_S       2  2  -0.5300
   T C3*     C_S       4  0   0.1600
   T  C3*    C_S       4  0   0.1600
   T HC3*    H_        1  0   0.0200
   T O3*     O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
   T  O3*    O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
   T C2*     C_3       4  0   0.3200
   T  C2*    C_3       4  0   0.3200
   T HC2*    H_        1  0  -0.0100
   T O2*     O_3       2  2  -0.7000
   T  O2*    O_3       2  2  -0.7000
   T HO2*    H___A     1  0   0.4400
   T  C1*    C_3       4  0   0.1500
   T C1*     C_3       4  0   0.1500
   T HC1*    H_        1  0   0.1100
   T HO3*    H_        1  0   0.4600
   T HO5*    H_        1  0   0.3920
*
 5MU N1      N_R       3  0  -0.2670
 5MU  N1     N_R       3  0  -0.2670
 5MU C2      C_R       3  0   0.6560
 5MU  C2     C_R       3  0   0.6560
 5MU O2      O_2       1  2  -0.5900
 5MU  O2     O_2       1  2  -0.5900
 5MU N3      N_R       3  0  -0.6430
 5MU  N3     N_R       3  0  -0.6430
 5MU HN3     H___A     1  0   0.3970
 5MU C4      C_R       3  0   0.6640
 5MU  C4     C_R       3  0   0.6640
 5MU O4      O_2       1  2  -0.5400
 5MU  O4     O_2       1  2  -0.5400
 5MU C5      C_R       3  0  -0.0320
 5MU  C5     C_R       3  0  -0.0320
 5MU C5A     C_3       4  0  -0.4240
 5MU  C5A    C_3       4  0  -0.4240
 5MU HC5A    H_        1  0   0.1320
 5MU  C6     C_B       3  0  -0.2070
 5MU C6      C_B       3  0  -0.2070
```

```
 5MU HC6     H_        1  0   0.2500
C  Sugar and Phosphate conversions for deoxy!
 5MU P       P_3+3     3  1   1.3250
 5MU  P      P_3+3     3  1   1.3250
 5MU O1P     O_P       1  2  -0.7770
 5MU  O1P    O_P       1  2  -0.7770
 5MU O2P     O_P       1  2  -0.7770
 5MU  O2P    O_P       1  2  -0.7770
 5MU O5*     O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
 5MU  O5*    O_3       2  2  -0.4850  !change to OH (-0.5120) 5' terminal
 5MU C5*     C_3       4  0   0.0800
 5MU  C5*    C_3       4  0   0.0800
 5MU HC5*    H_        1  0   0.1000
 5MU C4*     C_3       4  0   0.4600
 5MU  C4*    C_3       4  0   0.4600
 5MU HC4*    H_        1  0   0.0200
 5MU O4*     O_S       2  2  -0.5300
 5MU  O4*    O_S       2  2  -0.5300
 5MU C3*     C_S       4  0   0.1600
 5MU  C3*    C_S       4  0   0.1600
 5MU HC3*    H_        1  0   0.0200
 5MU O3*     O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
 5MU  O3*    O_3       2  2  -0.6660  !change to OH (-0.7200) 3' terminal
 5MU C2*     C_3       4  0   0.3200
 5MU  C2*    C_3       4  0   0.3200
 5MU HC2*    H_        1  0  -0.0100
 5MU O2*     O_3       2  2  -0.7000
 5MU  O2*    O_3       2  2  -0.7000
 5MU HO2*    H___A     1  0   0.4400
 5MU  C1*    C_3       4  0   0.1500
 5MU C1*     C_3       4  0   0.1500
 5MU HC1*    H_        1  0   0.1100
 5MU HO3*    H_        1  0   0.4600
 5MU HO5*    H_        1  0   0.3920
*
   U N1      N_R       3  0  -0.2830
   U  N1     N_R       3  0  -0.2830
   U C2      C_R       3  0   0.6670
   U  C2     C_R       3  0   0.6670
   U O2      O_2       1  2  -0.5880
   U  O2     O_2       1  2  -0.5880
   U N3      N_R       3  0  -0.6350
   U  N3     N_R       3  0  -0.6350
   U HN3     H___A     1  0   0.3860
   U C4      C_R       3  0   0.7920
   U  C4     C_R       3  0   0.7920
   U O4      O_2       1  2  -0.5650
   U  O4     O_2       1  2  -0.5650
   U C5      C_R       3  0  -0.5400
   U  C5     C_R       3  0  -0.5400
   U HC5     H_        1  0   0.2180
   U C6      C_B       3  0   0.0160
   U  C6     C_B       3  0   0.0160
   U HC6     H_        1  0   0.1920
```

```
C  Sugar and Phosphate conversions for deoxy!
   U P        P_3+3     3  1    1.3250
   U  P       P_3+3     3  1    1.3250
   U O1P      O_P       1  2   -0.7770
   U  O1P     O_P       1  2   -0.7770
   U O2P      O_P       1  2   -0.7770
   U  O2P     O_P       1  2   -0.7770
   U O5*      O_3       2  2   -0.4850  !change to OH (-0.5120) 5' terminal
   U  O5*     O_3       2  2   -0.4850  !change to OH (-0.5120) 5' terminal
   U C5*      C_3       4  0    0.0800
   U  C5*     C_3       4  0    0.0800
   U HC5*     H_        1  0    0.1000
   U C4*      C_3       4  0    0.4600
   U  C4*     C_3       4  0    0.4600
   U HC4*     H_        1  0    0.0200
   U O4*      O_S       2  2   -0.5300
   U  O4*     O_S       2  2   -0.5300
   U C3*      C_S       4  0    0.1600
   U  C3*     C_S       4  0    0.1600
   U HC3*     H_        1  0    0.0200
   U O3*      O_3       2  2   -0.6660  !change to OH (-0.7200) 3' terminal
   U  O3*     O_3       2  2   -0.6660  !change to OH (-0.7200) 3' terminal
   U C2*      C_3       4  0    0.3200
   U  C2*     C_3       4  0    0.3200
   U HC2*     H_        1  0   -0.0100
   U O2*      O_3       2  2   -0.7000
   U  O2*     O_3       2  2   -0.7000
   U HO2*     H___A     1  0    0.4400
   U  C1*     C_3       4  0    0.1500
   U C1*      C_3       4  0    0.1500
   U HC1*     H_        1  0    0.1100
   U HO3*     H_        1  0    0.4600
   U HO5*     H_        1  0    0.3920
*
 HOH OH       O_3_J     2  2   -0.8340
 HOH OW       O_3_J     2  2   -0.8340
 HOH O        O_3_J     2  2   -0.8340
 HOH  OH      O_3_J     2  2   -0.8340
 HOH  OW      O_3_J     2  2   -0.8340
 HOH  O       O_3_J     2  2   -0.8340
 HOH HO       H__OJ     1  0  0.41700
 HOH HOW      H__OJ     1  0  0.41700
 HOH H        H__OJ     1  0  0.41700
 HOH  HO      H__OJ     1  0  0.41700
 HOH  HOW     H__OJ     1  0  0.41700
 HOH  H       H__OJ     1  0  0.41700
*
```