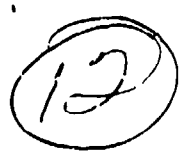

EDGEWOOD RESEARCH, DEVELOPMENT & ENGINEERING CENTER

ERDEC-TR-029

APPLICATION OF MOLECULAR MODELING TO BIOLOGICAL PROCESSES

Alfred H. Lowrey

NAVAL RESEARCH LABORATORY
Washington, DC 20375

George R. Famini
Charles Wick

RESEARCH AND TECHNOLOGY DIRECTORATE

July 1993

U.S. ARMY
CHEMICAL
AND BIOLOGICAL
DEFENSE AGENCY

Aberdeen Proving Ground, Maryland 21010-5423

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1993 July	Final, 92 Jan - 92 Apr

4. TITLE AND SUBTITLE

Application of Molecular Modeling to Biological Processes

5. FUNDING NUMBERS

PR-10162622A553L

6. AUTHOR(S)

Lowrey, Alfred H. (NRL), Famini, George R., and
Wick, Charles (ERDEC)*

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

DIR, ERDEC, ATTN: SCBRD-RTB/SCBRD-RTC,
APG, MD 21010-5423
DIR, NRL, Laboratory for the Structure of Matter,
Washington, DC 20375

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDEC-TR-029

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

*When this work was performed, ERDEC was known as the U.S. Army Chemical Research, Development and Engineering Center, and the ERDEC authors were assigned to the Research Directorate.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Detailed understanding of the molecular basis for biological processes is now available through computational modeling techniques. Advances in computational algorithms and technology allow applications to large biological macromolecules and permits the study of such problems as binding mechanisms, chemical reactivity, structural and conformational effects, and simulations of molecular motions. Recent crystallographic data provides access to detailed structural information that allows analysis and comparison of various computational techniques. Preliminary semi-empirical studies on N-acetylneuraminic acid are presented as an example of computational studies on binding mechanisms. N-acetylneuraminic acid is a substituted carbohydrate, which is a recognition site for binding of proteins (i.e., cholera toxin). These calculations provide some insight into electronic effects on binding in a crystal complex and the effect of the molecular charge on hydrogen bonding in the crystal complex.

14. SUBJECT TERMS

Computational chemistry
Computational biology
Molecular modeling

15. NUMBER OF PAGES

30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Blank

PREFACE

The work described in this report was authorized under Project No. 10162622A553L, CB Defense Assessment Technology. This work was started in January 1992 and completed in April 1992.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for release to the public. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

Blank

CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPLICATION OF MOLECULAR MODELING TO BIOLOGICAL PROCESSES

1. INTRODUCTION

This study concerns molecular modeling of processes that are known to occur on cell surfaces. Detailed structural information is now becoming available at the molecular level.[1] This study provides information about molecular structure and interactions known to affect fundamental biochemistry at cell surfaces.[2] The currently provided computational techniques increased the capacity of molecular modeling techniques to include large-scale assemblies of molecules that simulate biological materials and structures.[3] Understanding this structural information with modeling techniques is the first step toward understanding the effects of surface chemistry on membrane performance. Knowledge of these interactions is a step toward the goal of a useful model for the cell surface membrane.

Computational modeling is the primary tool used in integrating detailed information on the nature of molecular structure and interactions into knowledge of complex biological processes.[4] This technique is based on developments in theoretical chemistry,[5] whose applications have been made possible by the corresponding developments in computational science.[6] It has been possible to use the early success in calculating atomic and molecular structure[7] as building blocks for macromolecules of biological interest[8] and then to further simulate assemblies of these macromolecules as representations of biological materials.[9] Molecular modeling is a systematic approach for understanding complex molecular organization. Computational techniques organize large amounts of chemical information into tools that can be used to explain molecular phenomena (e.g., self-assembly) that are the basis for the functioning of biological structures (e.g., cell walls and membranes).

Modeling these membranes is important for understanding basic cellular processes. The membrane is a complex structure composed of proteins, lipids, complex carbohydrates, and distinct chemical species (e.g., glycolipids and glycoproteins), that combine the biochemistry of these basic building blocks.[10] Understanding the biochemistry of these complex materials at the molecular level will develop fundamental knowledge for building a more complete model that would encompass fundamental questions about the effect of environmental influences on cellular behavior. Specific phenomena (e.g., blood type identification, action of toxins, and cell fertilization and division) are related to specific molecular processes. Broad areas (e.g., cell recognition, adhesion, biochemical transport, and cell movement) are also recognized as being related to particular types of chemical processes. Successful detailed molecular modeling of membranes and cell surfaces would be a breakthrough in understanding such

general concerns as identification, survivability, detection, and other features important to biodefense issues.

The scope of this study is molecular modeling of sialic acid to study its binding to wheat germ agglutinin isolectin 1 (WGA1). This biochemical process has important implications for studies on binding of toxins,[11] on structural effects on cell walls,[12] and other biological phenomena. This report will present results of semi-empirical molecular orbital calculations on N-acetylneuraminic acid (NeuNAc). These results yield information about molecular structure, charge, and electronic distribution. This information will be related to a crystallographic study of WGA1.[13]

2. MOLECULAR MODELING

Computational molecular modeling is based on the observation that complex phenomena in the physical world can be described by theoretical and mathematical methods. "The more progress physical sciences make, the more they tend to enter the domain of mathematics, which is a kind of center to which they all converge. We may even judge the degree of perfection to which a science has arrived by the facility with which it may be submitted to calculation."[14] Current capabilities allow quantitative calculation of molecular properties with precision comparable to experimental methods. Recent advances in computational algorithms and large increases in available computing power make computational methods a unique tool for dealing with the large systems that form the molecular basis for biological materials and structures. Understanding the chemistry of large molecules like proteins, and intricate mechanisms like receptor docking, requires facilities such as modern workstations and computational chemistry systems for computational speed and graphics visualization.

There are many techniques for computational modeling; some important methods are discussed below. The applicability of these methods depends on the capacity of the computational algorithms and the available computational resources. There is a direct relationship between accuracy and the number of atoms in the molecule; larger numbers of atoms necessitate more approximate treatment. Table 1 summarizes the limit on the size of molecules and relative requirements for computational resources.

Ab initio quantum chemistry is a computational technique that depends only on solving the Schrödinger equation using physical constants and approximations for the wavefunction involving linear combinations of atomic orbitals (LCAO).[7] Current capabilities allow routine calculations on molecules with up to 25 atoms.[15] These methods are most readily available using the commercial series of GAUSSIAN programs.[16] *Ab initio* methods

provide intricate details about molecular electronic structure that are needed to understand the complex reactions in molecules like enzymes. These methods allow the calculation of molecular properties, (e.g., electron density and electrostatic potential surfaces) that give information about the structures and reactive properties of small molecules and active sites in biological macromolecules. These methods are the principal techniques for studying the fine details of bond breaking. *Ab initio* methods give fine details about intermolecular interactions, such as hydrogen bonding, which is a central phenomenon in biological processes.[17] *Ab initio* methods have been used to understand spectroscopy of biological molecules in water.[18] These methods are useful in understanding the complex structural transformations in biological macromolecules that are central mechanisms for their activity.

Table 1. Comparison of Various Modeling Techniques

Ab Initio Methods	Most accurate, most detail; resource intensive, small molecules.
Semi-Empirical Methods	Approximate electronic properties; less accurate molecular properties (e.g., vibrational frequencies), rapid calculations, molecules up to 100 atoms.
Molecular Mechanics	Rapid approximate structures; macromolecules with thousands of atoms; no electronic information.
Molecular Dynamics	Based on molecular mechanics, computationally intensive; long calculations for averages over many structures and trajectories; approximate information on molecular motions; molecule size similar to molecular mechanics but more severely limited by available computational resources.
Theoretical Linear Solvation Energy Relationships	Rapid calculations for specialized molecular properties; requires experimental information for series of related molecules; molecule size limited by availability of data and capability of modeling properties desired.

Semi-empirical methods use experimentally derived information to approximate the more exact solutions of the molecular wavefunction. These methods are characterized by rapid calculations and economical use of computing resources. Thus, semi-empirical methods can be used on very large molecules. Semi-empirical computations are primarily available in MOPAC.[19] This computational package has facilities for approximate calculations of most molecular structure and property features available with *ab initio* methods. The semi-empirical approximations allow economical calculation of the changing molecular structure along a reaction path, thus allowing detailed analysis of the fundamental chemistry. Semi-empirical methods have found wide application in molecular interactions and large scale systems such as polymers.

Molecular mechanics is a method that calculates molecular structure, energy, and vibrational properties using classical force field theory and empirically derived force fields.[20] Current techniques allow rapid and accurate structure calculations for small molecules. The use of molecular mechanics techniques is essential for understanding the structure of large scale biomolecules with thousands of atoms. Recent breakthroughs in advanced computational graphics techniques have been coupled with molecular mechanics to perform visual chemistry (e.g., docking of a drug with a receptor site).

Molecular dynamics is a computational tool used for understanding the motions of molecules. It is based on molecular mechanics and Newtonian equations of motion.[3] This method allows simulation of changes in molecular structure associated with chemical reactions and biological activity. The introduction of free energy perturbation techniques uses molecular dynamics techniques for calculating thermodynamic quantities associated with structural and chemical changes. Molecular dynamics is used in understanding the role of water in biological processes. Concerted use of molecular dynamics with experimental data from x-ray crystallography is a new technique for determining the structure of large macromolecules.[21]

Theoretical Linear Solvation Energy Relationships is a method for correlating and predicting physical and biological activities with microscopic thermodynamic parameters.[22,23] This is based on the Quantitative Structure Activity Relationship (QSAR) methods, which have been widely used in medicinal chemistry. The development of modern semi-empirical methods allows systematic integration of experimental data with computational results. The efficiency of these techniques allows the application of these methods to large molecules associated with enzyme activity and transport across cell membranes.[24]

3. MOLECULAR BASIS FOR BIOLOGICAL PROCESSES

Molecules are the fundamental building blocks for biological materials and structures. The chemical basis for many central biological processes is now understood. Common molecular patterns and principles underlie the diverse expressions of life.[10] Molecular structure and function is the fundamental information for understanding biochemical processes. The activity of biological macromolecules is directly related to their conformation, which is determined by the molecular composition. The three-dimensional structure of proteins is uniquely determined by the sequence of the amino acid residues. This controls such processes as the unique specificity of enzymes. Macromolecule biosynthesis is governed by the structure and interaction of the molecular building blocks. For example, this governs the rates at which biological materials are made. Information transmission and storage in genetic materials is determined by the sequence of base pairs. The generation and storage of metabolic energy often depends on small rearrangements of molecular conformation. Computational modeling provides tools for integrating information about molecules for the purpose of understanding the complex chemistry of biological processes. Examining some basic types of molecules in the cell makes it possible to consider the kinds of information computational modeling provides for problems in biological defense.

3.1 Genetic Materials.

DNA and RNA are polymers of nucleotides. They contain sugars and phosphate groups that form the structural elements and the purine and pyrimidine bases that convey the genetic information. DNA is the storehouse for the information that ultimately governs the synthesis of all biomolecules. It is generally found in the Watson-Crick double helix. Breakthroughs in computer graphics and simulations have discovered that intercalation of flat aromatic rings is a major mechanism for mutations.[10] Using molecular mechanics, docking studies show that molecules (e.g., acridine) can insert through the edge of the DNA helix. This effectively displaces the stacking of the bases and leads to insertion or deletion. The effect of such mutations is to alter the reading frame in transcription. Computational modeling is used to understand the binding of drugs to groove sites in the DNA helices. Recent advances suggest that twists in the conformation of the base pairs strongly influence hydrogen bonding.[17]

3.2 Proteins.

Proteins are the basic catalysts for biochemical reactions. Detailed information about their structure is available from x-ray crystallography and molecular modeling. Computational techniques provide information about molecular properties that is useful for interpreting the mechanisms of the

catalytic reactions. Recent advances have used calculations of electrostatic fields to interpret the interactions and docking mechanisms for electron transfer in metalloproteins. Matching the magnitudes and directions of calculated electrostatic vectors for plastocyanin and cytochrome c provided the relative orientation of the two proteins and enabled detailed analysis of the binding between them.[25]

3.3 Lipids.

Lipids are relatively small biological macromolecules that are characterized by their insolubility in water and generally contain long fatty acid chains from 12-24 carbon atoms in length. They are characterized by a high degree of conformational disorder and the absence of subunit connectivity. For these reasons, techniques such as x-ray crystallography yield less information about the molecular structure in the natural state. Computational modeling of lipids has primarily been focused on packed layers because of the flexibility of the individual molecule and the interest in the self-assembled materials that constitute components of the plasma membrane. Recent advances in large-scale molecular mechanics and graphics visualization techniques have shown that a 12-molecule layer of cholic acid, which is a lipid simulant, shows spontaneous formation of irregular channels in an approximately helical structure.[26] From the dimensions of the channels, it is apparent that guest molecules can be incorporated into a cholic acid micell through more than one channel.

3.4 Carbohydrates.

Complex carbohydrates are macromolecules built from simple sugars. The many hydroxyl groups on the sugars give rise to different stereochemical linkages between the monomer units, often with branching to form multiple chains in the single polymer. The sequence of sugars in complex carbohydrates can serve as code words in the molecular language of life. This code is the basis for much of the regulatory and recognition functions on the surfaces of cells. Because the three-dimensional structure of complex carbohydrates is often ill-defined, these molecules are a challenging class of problems for computational techniques.[27] Recent work using semi-empirical methods, calculated structure and charge distributions for some glucosidase inhibitors. Differences in charge density explained inhibition properties while structural changes were correlated with lack of binding to the enzyme.[28]

3.5 Biological Materials.

The application of molecular modeling techniques to basic biological macromolecules provides the foundation

for constructing models of more complicated systems (e.g., membranes). Membranes consist mainly of lipids and proteins. They also contain carbohydrates that are linked to the proteins and lipids. A goal is to assemble computational models of various components to simulate the function of the cell wall membrane. Some preliminary results show the feasibility of this idea. Molecular mechanics and molecular dynamics, coupled with new mathematical techniques, have been used to model lipid bilayers[29] and micelles.[26] Carbohydrate interactions with a lipid bilayer have been modeled.[9] Molecular modeling results were used in understanding experimental data for a lipopolysaccharide membrane.[30] The electrostatic properties of a bacterial phorin were used to understand ion transport across the outer membrane of a gram negative bacteria.[31]

4. CRYSTALLOGRAPHIC STRUCTURE

For those proteins that crystallize, x-ray diffraction techniques provide three-dimensional atomic coordinates. This information is the basis for many computational modeling techniques because of the ability to visualize chemical sites and active processes at specific locations in the protein. The reliability of this information with respect to the structure, and thus function, of the proteins in water (naturally active situation) is always at question because of possible changes in conformations, influences of crystal lattice contacts, and preferential orientation of hydrogen bonded waters in the crystal. It has been concluded that the static x-ray coordinates contribute to understanding enzyme mechanisms for the following reasons:

- The structures of many proteins have repeatedly been determined under many conditions, giving identical results.
- Related proteins invariably have the same polypeptide chain folding.
- Crystal lattice contacts are too weak and too limited in area to disturb the protein folding to any significant extent.
- The properties of molecules in solution can usually be explained by the crystal structure.
- Active molecules (e.g., enzymes) often retain full catalytic activity in the crystal; especially if no large conformational changes occur.

The usefulness of the crystallographic data for studying enzyme sites is important to the whole range of modeling problems for biological materials. It is essential to recognize that

well-defined active sites and the tightly bound complexes found in the crystal structures of enzymes are representative models of reactions even in the solution bound continual motion associated with biological materials.[32]

4.1 WGA1.

This study is based on x-ray crystallographic coordinates for a complex of WGA1 and N-acetylneuraminyl-lactose [(NeuNAc-α(2-3)-Gal-ß-(1-4)-Glc)(13)]. The WGA1 is a plant lectin; a class of molecules that are quantitated by their ability to agglutinate erythrocytes. Oligiomeric plant lectins, which display stringent sugar specificity, constitute a special class of these proteins. As a representative of the highly conserved group of lectins in the Gramineae family, WGA1 possesses several properties distinct from other plant lectins. The most important is the specificity for two different types of saccharides, N-acetyl-D-glucosamine (GlcNAc) and NeuNAc (α anomer). In addition, there is a requirement of two isostructural domains for sugar binding and the existence of two independent noncooperative binding sites. These features are important for linkages in surface binding.*

The WGA1 exhibits specificity only for acetylated sialic acid sugars, which has been explained based on structure.[33] The compound (NeuNAc) satisfies the stereochemical requirement for an equatorial N-acetyl group and an adjacent equatorial OH group. The N-acetylneuraminyl-lactose is an important receptor analog. In the complex, only one binding site is occupied and is determined to be the primary site. The secondary site is unoccupied because of the charge on the NeuNAc moiety. In the binding site, the N-acetyl group makes the largest number of contacts, contributing both hydrogen bonds and an important nonpolar interaction between the methyl group and the phenyl ring of Tyr 73.[13] Molecular modeling studies on NeuNAc will yield information about the nature of these interactions and about molecular specificity.

4.2 Sialic Acids.

Interest in sialic acids has increased as a result of recognition of their involvement in regulating a large number of biological phenomena. Sialic acids play a strong protective role in living cells and organisms. This appears to be a result of their peripheral position in glycoconjugates and, correspondingly, their frequent external location in cell membranes.[12]

---

*Wright, C.S., J. Biol. Chem. (in press).

Figure 1 shows the basic structure diagram for neuraminic acid. This is a nine-carbon sugar characterized by a carboxylic acid group at position 2 and a glycerol tail originating from position 6. The name sialic acid refers to a class of all N- and O-acyl derivatives of neuraminic acid. The most common is NeuNAc with an N-acetyl substitutent at the 5 position. This occurs on both the α and β anomeric form, with respect to the carboxylate group. This molecule is associated with the biological phenomena discussed below. There are many other substituted sialic acids; the 5 glycolyl substituted molecule (Neu5Gc) is associated with the Hanganutziu-Deicher antibodies but not NeuNAc.[12] There is a large literature on other forms, in particular the recent synthesis work of Hartmann and Zbiral at the University of Vienna.[34] Unsubstituted neuraminic acid is not known to exist in as a free molecule.[12]

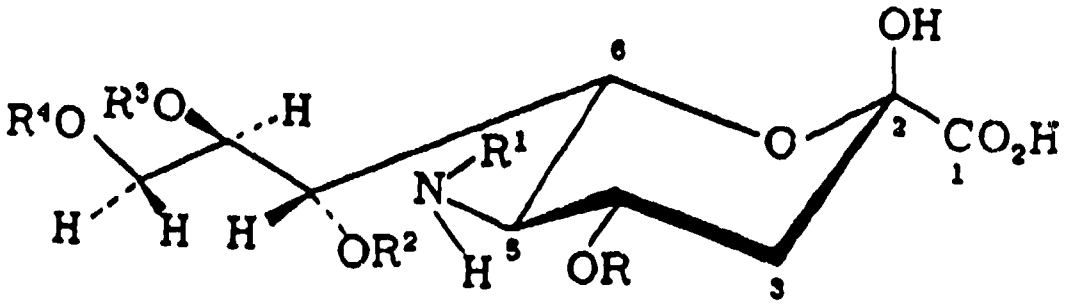

Figure 1. Substitution Diagram for Natural Sialic Acids. N- and O-Substituents occur at the corresponding positions R-R_4, where the most common molecule is NeuNAc, with an acetyl group at position R_1 and hydrogens elsewhere.[12]

Sialic acid residues exist as negatively charged carboxylate ions in the biological environment. Based on the accumulation of the negatively charged sialic acid residues on cell membranes, it may be expected that these compounds strongly influence the behavior of the cells. For example, this prevents aggregation due to electrostatic shielding for blood platelets and erythrocytes where, in others (e.g., chick embryo muscle cells), it appears to facilitate aggregation possibly due to Ca^{++} bridges.[12] The repulsive electrostatic forces of sialic acids contribute to the rigidity of the cell surface in studies on sarcoma cells. Sialic acid residues influence the viscosity of glycoprotein.[12]

Another important function is known as the anti-recognition effect. Ashwell and Morell[35] discovered that sialic acids masks the D-glactosyl residues of various serum-glycoproteins and thus protects their survival in the blood. With enzymatic removal of the sialic acids, these molecules are rapidly degraded by the liver function and removed from circulation.[10]

Another anti-recognition use of sialic acid is exhibited by the K1 strain of *Escherichia coli*. The cell makes long polymers of sialic acids known as colominic acid, which cover the surface to mask its presence from the complement path.[36]

Sialic acid is also known as the receptor molecule for the influenza virus and for the cholera toxin. The cholera toxin consists of two main parts: a catalytic unit A chain and a membrane penetration unit that consists of five B chains.[10] The B chains bind to the monosialoganglioside GM1 and are responsible for penetrating the membrane.[11]

Sialic acid is associated with an important transmembrane protein, glycophorin A. This is a single polypeptide chain with 16 attached oligosaccharide units. It is known that all molecules integral to the membrane point in the same direction.[37] Sialic acid is part of the attached oligiosaccharides and is responsible for giving red blood cells a very hydrophilic, anionic coat.[10] Figure 2 is a schematic diagram illustrating the complex structural interactions between lipids, proteins, and carbohydrates that influence the molecular characteristics of cell surfaces. The carbohydrates are known to exist on the outside of the cell; this is presumed to represent the structural elements for glycosolated transmembrane proteins such as glycophorin A.

5. METHODOLOGY

N-acetylneuraminic acid is a large molecule with 11 carbon atoms, 9 oxygen atoms, 1 nitrogen atom, and 18 hydrogen atoms. Semi-empirical methods are best suited for molecular orbital studies on this molecule. The geometries of NeuNAc and $NeuNAc^-$ ion were optimized with the PM3 algorithm, as contained within MOPAC 5.06.[19] Because the molecule is large and floppy, the PRECISE option was used to insure convergence. Initial geometries were generated using the in-house Chemview visualization package on an Ardent Titan. Final geometrical information was obtained using routines incorporated into the Molecular Modeling, Analysis and Display System (MMADS) molecular modeling package.[38] Visualization of the molecular orbitals and electrostatic potentials was obtained using Chemview and the SPARTAN system, WAVEFUNCTION, Incorporated (Irvine, CA).

6. RESULTS AND DISCUSSION

6.1 Geometry.

Preliminary results are shown in Table 2, which compares bond lengths for the PM3 optimized geometry with those from a crystallographic study.[39] The optimized structure and atom numbering for NeuNAc are displayed in Figure 3.

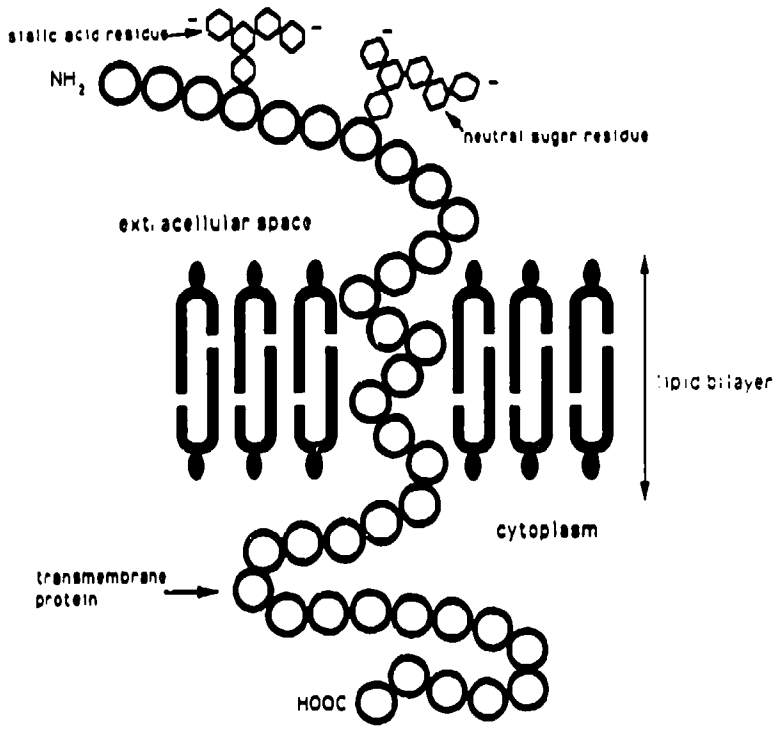

Figure 2. Schematic Representation for a Glycosolated Transmembrane Protein. In Glycophorin A, the Carbohydrate Residues Contain Terminal Sialic Acid Groups as Represented in the Diagram[10]

Table 2. NeuNAc Comparison of Bond Lengths Between PM3 and Experimental Results

Distance	PM3 Geometry Å	Experimental Å
O1 - C1	1.4066	1.420
C1 - C2	1.5410	1.535
C2 - C3	1.5368	1.519
C3 - C4	1.5545	1.517
C4 - C5	1.508	1.532
C5 - O1	1.4290	1.440
C5 - C6	1.5646	1.517
C6 - C7	1.5582	1.532
C7 - C8	1.5506	1.512
C1 - C9	1.5551	1.531
C10 - C11	1.4996	1.494
C1 - O2	1.3981	1.400
C9 - O3	1.3526	1.290
C9 - O4	1.2128	1.198
C3 - O5	1.4170	1.434
C10 - O6	1.2228	1.237
C6 - O7	1.4080	1.428
C7 - C8	1.4165	1.432
C8 - O9	1.4006	1.428
C4 - N1	1.4895	1.452
N1 - C10	1.4327	1.330

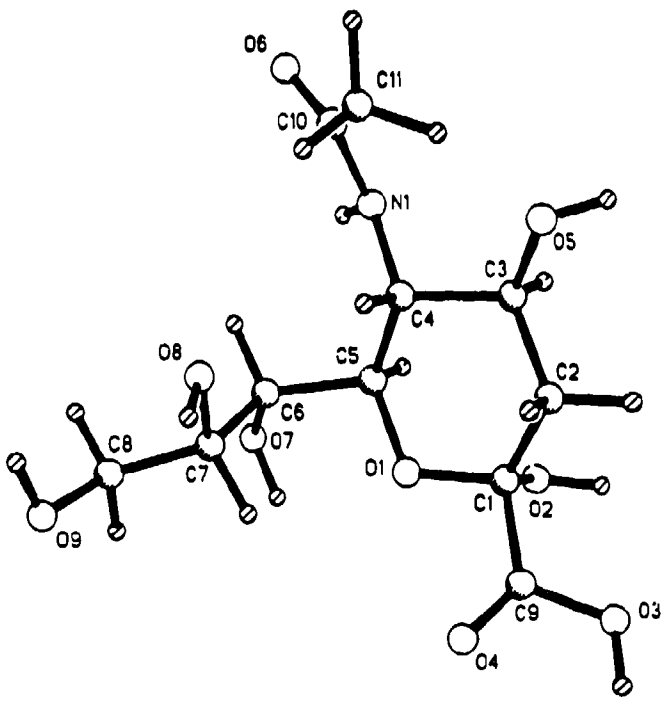

Figure 3. Optimized Structure for NeuNAc Using Semi-Empirical Methods and the PM3 Hamiltonian

Because the parameterization within the PM3 Hamiltonian is derived from averages over all types of experimental data and a large variety of molecules, the rough agreement between the PM3 optimized geometry and the x-ray crystallographic structure is what is expected.[19] As with any computational technique, the differences tend to be systematic in respect to atom type. Thus, the calculated C-C distances are too long, the C-O distances are too short, and the C-O and C-N distances appear more dependent on the local chemical environment. This is a result of the effects of nonspherical atoms and polarization of the electron distribution: problems faced by all computational techniques.[7,20] Within this error limit, the optimized structure of the NeuNAc$^-$ ion is essentially identical to that of the protonated form, with the exception of the geometry for the carboxylate group. The C1-C9 distance increases to 1.612 Å, and the C-O distances become 1.239 and 1.259 Å, respectively. For these studies on binding, the conformation of the molecule becomes the important result. The PM3 calculations produce good agreement between the optimized structure and the experimentally observed molecular framework. The calculated chair form is the same as that observed in the crystal structure, although the positions of the ring atoms do not deviate from the ring plane to the extent observed experimentally. The ring torsion angles involving only carbon atoms are calculated to be very similar (within 2°), while the ring torsions involving oxygen differ to a greater extent (up to 11°), such that the deviations from experiment are greatest for atoms with polarized electron distributions. The optimized structure retains the axial and equatorial substituents observed experimentally; the equatorial positions of the N-acetyl group at C4 and the adjacent equatorial hydroxyl at C3 are essential for binding.[33] The conformations of the side chains are not consistent with those observed in the crystal structure. The positions of O3 and O4 in respect to torsion around C1-C9 bond, and the positions of O6 and C11, in respect to torsion around the N1-C10 bond, are opposite to those observed in the crystal structure. The torsion angle for the conformation of the glycerol tail (C6-C7-C8) is of opposite sign compared to that observed in the crystal structure. Investigating these details of structure with other computational methods and other semi-empirical approximations will be important steps in computational studies of binding of sialic acid.

6.2 Electronic Properties: Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO).

The semi-empirical calculations allow rapid estimation of electronic properties known to be important in binding. Figures 4 through 7 display the important HOMO and LUMO for NeuNAc and NeuNAc$^-$, respectively. They indicate the localized availability of electrons for donation and sites for electron acceptance, which are indicative of reactivity of various

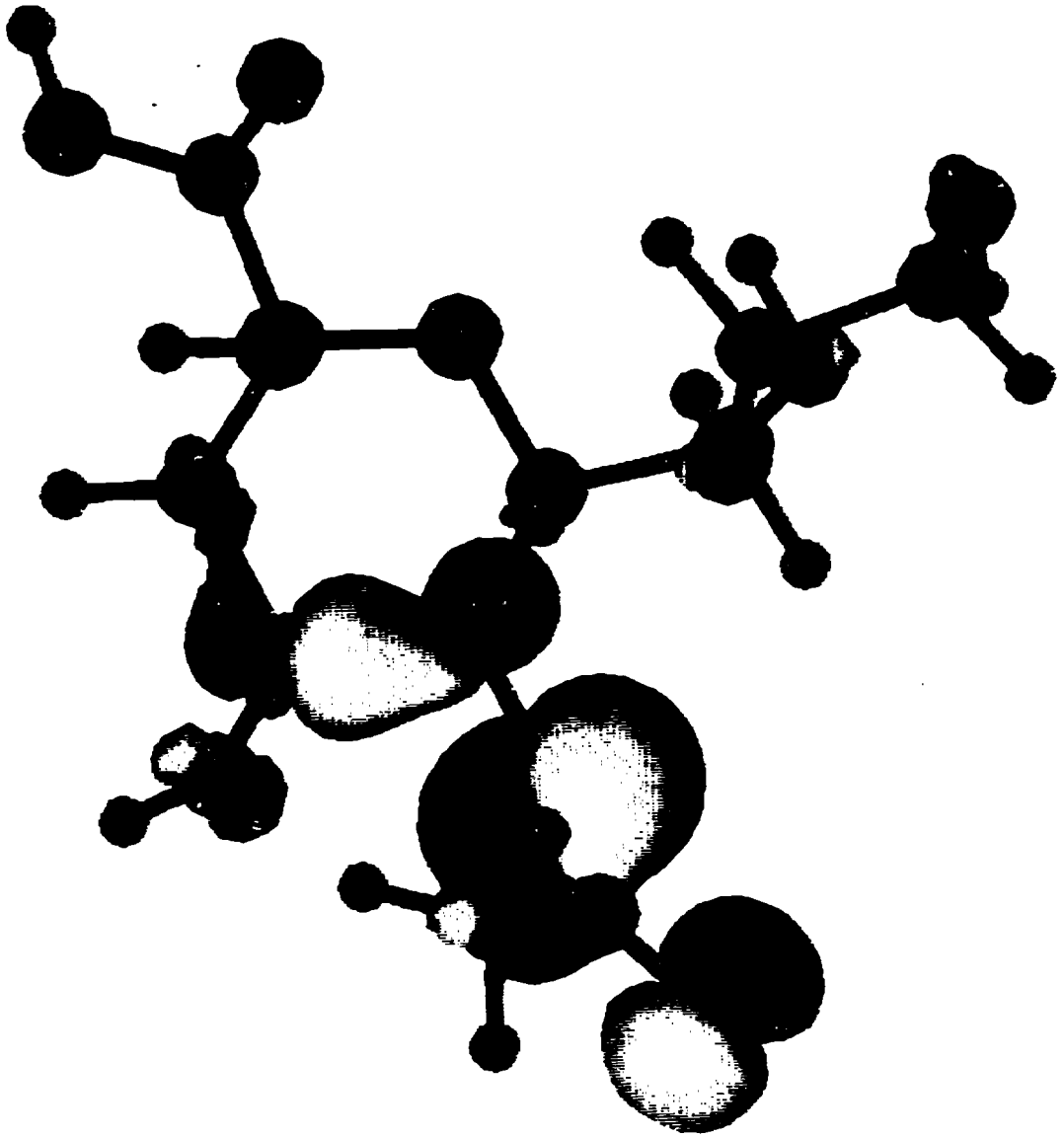

Figure 4. Plot of the HOMO Surface, at ±0.1 Density Value, for Protonated Sialic Acid

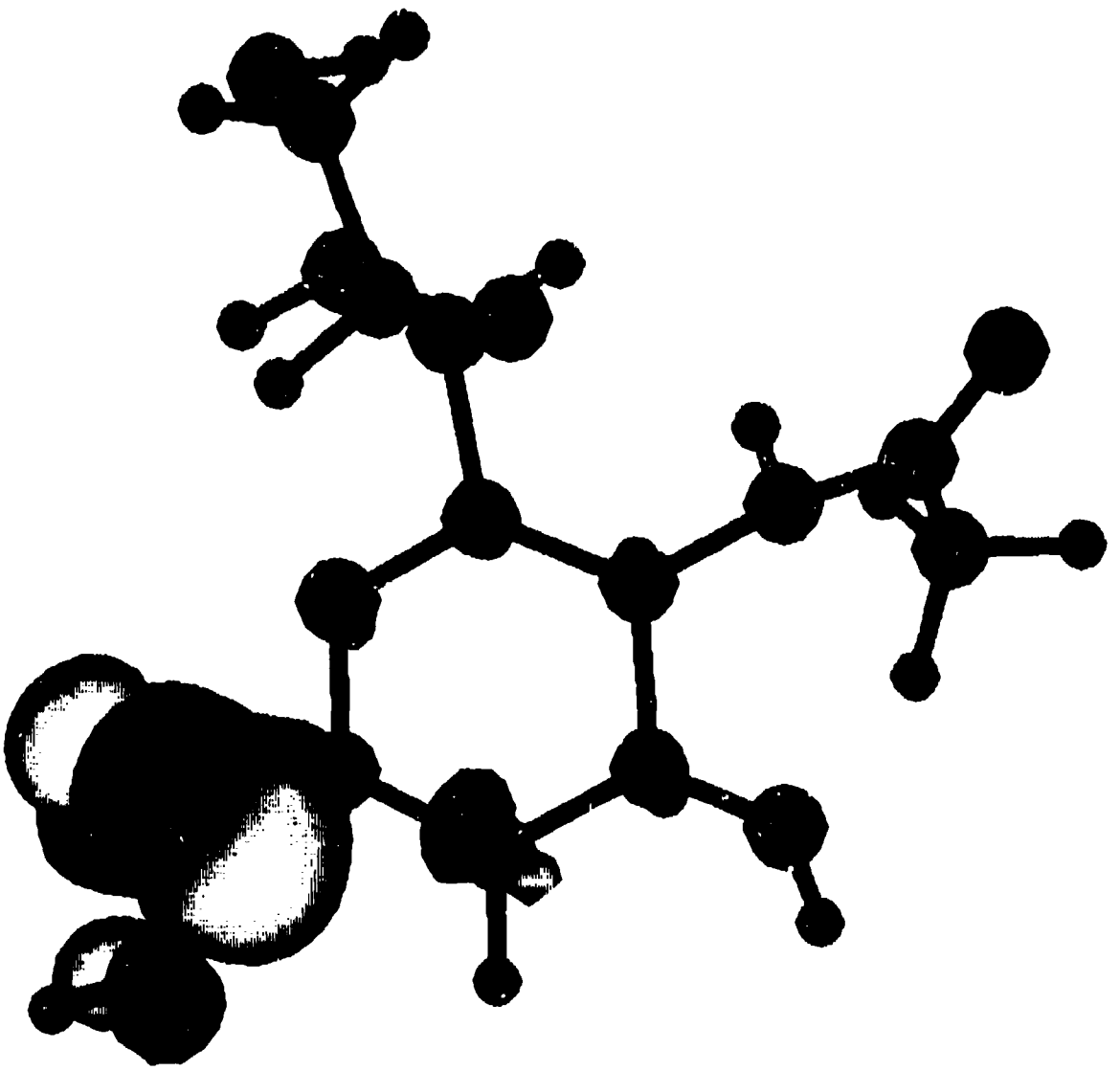

Figure 5. Plot of the LUMO Surface, at ±0.1 Density Value, for Protonated Sialic Acid

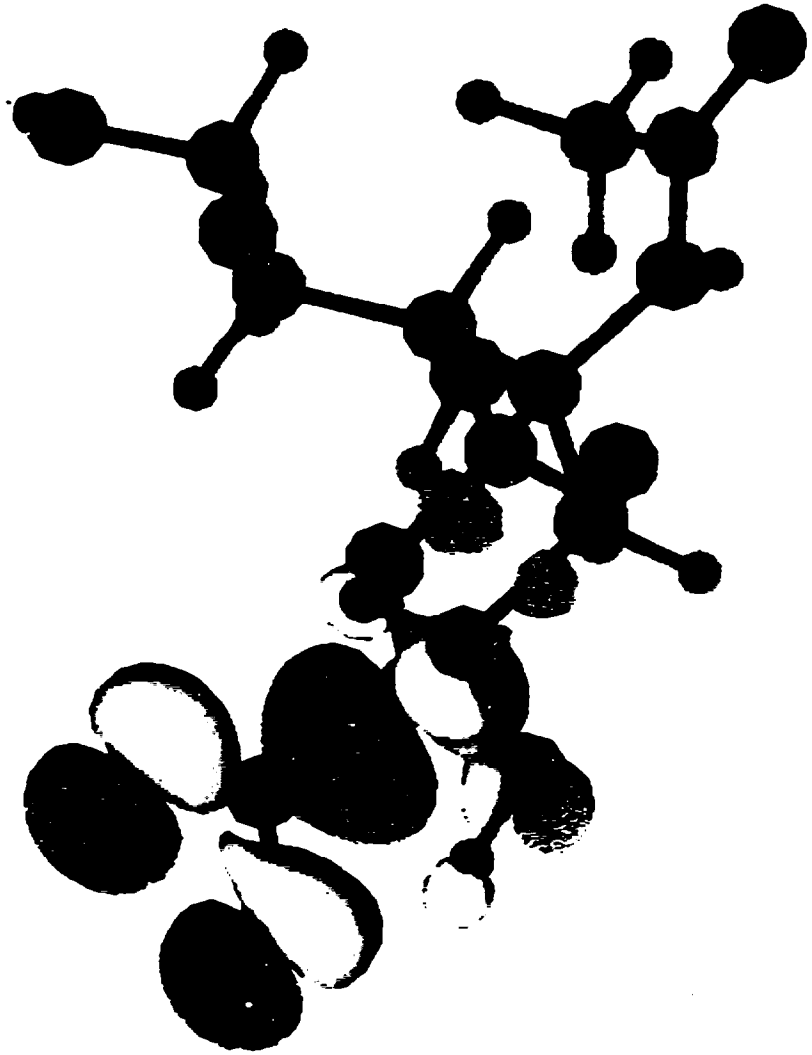

Figure 6. Plot of the HOMO Surface, at ±0.1 Density Value, for Sialic Acid Ion

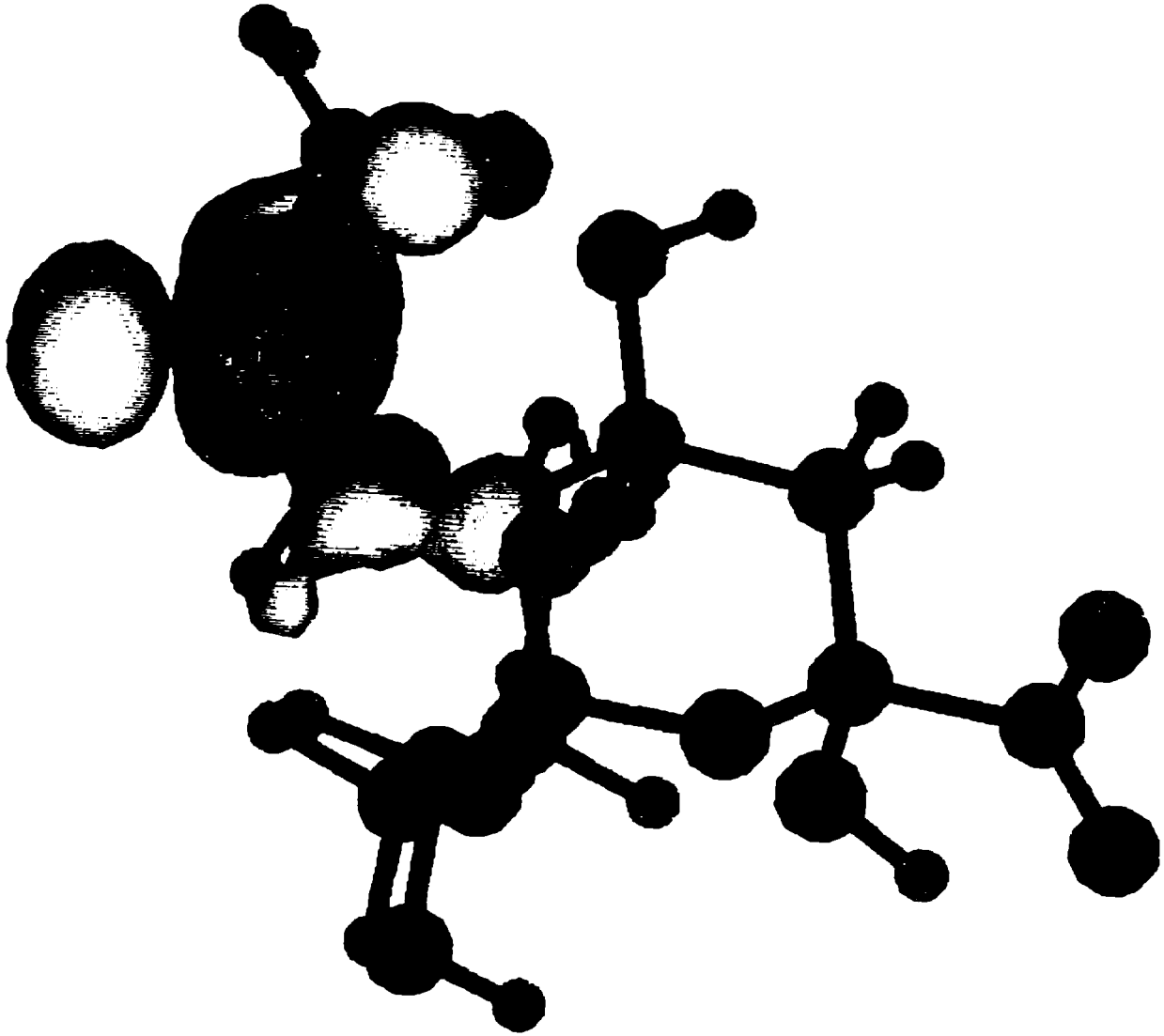

Figure 7. Plot of the LUMO Surface, at ±0.1 Density Value, for Sialic Acid Ion

portions of the molecule.[40-41] These figures show that the distribution of available electrons undergoes dramatic change with the removal of the acidic proton from the carboxylate group. For the protonated molecule, the HOMO electron density clusters in the region of the N-acetyl group; whereas, the LUMO density clusters around the carboxylic acid group. With the removal of the proton, these density clusters shift position, with the HOMO density clustering around the carboxylate ion, and the LUMO density clustering around the N-acetyl group. This is relevant as an indicator of available electrons for hydrogen bonding. These two chemical groups are most involved in hydrogen binding in the WGA1 site.[13] The molecule exists as the negative ion at physiological pH, so the hydrogen bonding at the carboxylate group is expected. The complete localization of the LUMO at the N-acetyl group would seem contrary to its involvement with the hydrogen bond network. However, the carboxylate group binds very tightly to the OH hydrogen of Ser 114 such that "this contact is very close in monomer II of NLI1 where a strong electron density connection is present."[13] The dramatic shifts between the ion and acidic forms indicate that the partial binding of the close contact would induce a mixed distribution of HOMO and LUMO density in both the carboxylate and N-acetyl group.

6.3 Electrostatic Potentials.

The electrostatic potential V (r) that is created by electrons and nuclei in a molecule is a well established tool for interpreting and predicting molecular reactive behavior towards electrophiles.[42-43] It has been extensively used for the study of biological recognition interactions.[44] The value for the electrostatic potential at any point is given rigorously by equation (1).

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\sigma(r')\, d(r')}{|r' - r|} \qquad (1)$$

where

Z_A = nuclear charge

$\sigma(r')$ = electron density

The calculation of the electrostatic potential surface for NeuNAc indicates charge distributions suitable for hydrogen bonding. Based on the crystal structure of the isolated molecule, this electrostatic potential surface suggests that NeuNAc is a highly amphipathic molecule with most of the attractive negative potential distributed as a pocket on one side of the molecule. The preliminary results of the semi-empirical calculations raise important questions about the mechanisms of binding for NeuNAc. From the crystal structure for the WGA complex,[13] clear indications about conformation of the molecule are derived from intermolecular distances between NeuNAc and the oxygen atoms in the protein residues of the binding site. The hydrogen bond contacts indicate that the *trans* conformation for the peptide linkage for the N-acetyl group, as observed in the crystal structure,[39] is important and not well represented by the semi-empirical results, which yield a *cis* conformation for the PM3, AM1, and MNDO Hamiltonians. This surface also shows the clearly distinct potential around the methyl group in the N-acetyl group whose nonpolar contacts with a tyrosine, **Tyr 73**, are known to be important for binding. The potential surface also shows that the attractive potential pocket is located on the side of the molecule associated with nonpolar contacts with the side chains of **Tyr 64** and **Tyr 66**. This raises questions about the role of electrostatic charge in binding, since this is a negatively charged species at natural pH. The electrostatic potential surface for the negative ion shows strong variations between polar and nonpolar regions of the molecule despite its overall charge. The glycerol tail is known to be flexible and in different orientations in the binding site.[13] Changes in the electrostatic potential that depend on the conformation of this group will answer questions about its structural importance for binding.

7. CONCLUSIONS

Molecular modeling techniques are shown to be suitable for a large number of biological processes. Fundamental components of biochemical systems can be simulated with computational methods. These simulations can be combined to provide qualitative visual representations of biological materials and structures and quantitative evaluations of mechanisms and energetics. New technologies will allow simultaneous simulation of biomolecular systems and evaluation of their interactions using computational techniques.[45] These methods will provide new information about the fundamental physics and chemistry of living systems that is not readily available with current experimental techniques.

Preliminary computational studies on sialic acid illustrate the power of molecular modeling and visualization in studying fundamental chemistry of biological processes.

Semi-empirical methods appear suitable for large molecules such as sialic acid. Electrostatic potential energy surfaces give good correlation with proposed binding processes with proteins. The representations of electronic structure give new insight into the important role of molecular charge in the interactions of sialic acid with proteins such as wheat germ agglutinin. Molecular modeling is a useful and important tool for obtaining information that is difficult to obtain experimentally and for developing new lines of inquiry for further investigation.

Further applications of modeling methods can yield important information about biological materials (e.g., lipid membranes, complex carbohydrates, and protein complexes). New structural information and computational technology will dramatically increase the size and scope of molecular modeling applications and provide a new basis for interactions with experimental techniques.

LITERATURE CITED

1. Quiocho, F.A., Current Topics in Microbiology and Immunology Vol. 139, pp 135-148 (1988).

2. Albersheim, P. et al., Acc. Chem. Res. Vol. 25, pp 77-83 (1992).

3. Gunsteren, W.F.V., and Weiner, P.K., Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989.

4. Goddard, W.A. III, , Chem. Eng. News Vol. 70, p 24 (1992).

5. Ostland, A.S.N., Modern Quantum Chemistry, Macmillan Publishing, New York, NY, 1982.

6. Corcoran, E., Scientific American, pp 100-109 (1991).

7. Hehre, W.J., Radom, L., Schlyer, P.V.R., and Pople, J.A., Ab Initio Molecular Orbital Theory, Wiley Interscience, New York, NY, 1986.

8. Kollman, P., In Crystallographic and Modeling Methods in Molecular Design, pp 114-122, C.E. Bugg, and S.E. Ealick, Eds., Springer-Verlag, New York, Berlin, Heidelberg, 1990.

9. Rudolph, B.R., Chandrasekhar, I., Gaber, B.P., and Nagumo, M., Chem. Phys. Lipids Vol. 53, pp 243-61 (1990).

10. Stryer, L., Biochemistry, W.H. Freeman and Company, New York, NY, 1988.

11. Whalley, C.E., Toxins of Biological Origin, CRDEC-SP-021, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, June 1990, UNCLASSIFIED Report (AD B145 632L).

12. Schauer, R., Advances in Carbohydrate Chemistry and Biochemistry Vol. 40, pp 131-234 (1982).

13. Wright, C.S., J. Mol. Biol. Vol. 215, pp 635-651 (1990).

14. Quetelet, A., Instructions Populaires sur le Calcul de Probabilities, Tarlier, Brussels 1828.

15. Boggs, J.E., In Theoretical Models of Chemical Bonding, pp 185-212, Z.B. Maksic, Ed., Springer-Verlag, Heidelberg, 1990.

16. Frisch, M.J. et al., Gaussian, Incorporated, 1992.

17. Wang, A.H.J., Teng, M.-K., In Crystallographic and Modeling Methods in Molecular Design, C.E. Bugg, and S.E. Ealick, Eds., Springer-Verlag, New York, NY, 1990.

18. Williams, R.W., Lowrey, A.H., J. Comp. Chem. Vol. 12, p 761 (1991).

19. Stewart, J.J.P., J. Comp. Mol. Design Vol. 4 (1990).

20. Burkert, U., and Allinger, N.L., Molecular Mechanics, American Chemical Society, Washington, DC, 1977.

21. Brunger, A.T., Kuriyan, J., and Karplus, M., Science Vol. 235, p 458 (1987).

22. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships. I. Molecular Volume, CRDEC-TR-88031, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1988, UNCLASSIFIED Report (AD A191 522).

23. Famini, G.R., Using Theoretical Descriptors in Structural Activity Relationships. II. Polarizability Index, CRDEC-TR-88137, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, September 1988, UNCLASSIFIED Report (AD A199 594).

24. Famini, G.R., and Wilson, L.Y., In Annual International Conference of IEEE Engineering in Medicine and Biology Society, p 1608, IEEE, Eds., 1990.

25. Freeman, H.C., Getzoff, E.D., and Tainer, J.A., In Proceedings of the Division of Coordination and Metal-Organic Chemistry of the Royal Australian Chemical Institute, Royal Australian Chemical Institute, 1984.

26. Lawrence, S.M., Lawrence, M.J., and Barlow, D.J., J. Mol. Graphics Vol. 9, pp 218-225 (1991).

27. French, A.D., and Brady, J.W., In Computer Modeling of Carbohydrate Molecules, M.J. Comstock, Ed., American Chemical Society, Washington, DC, 1990.

28. Kajimoto, T. et al., J. Am. Chem. Soc. Vol. 113, pp 6187-6193 (1991).

29. Lawrence, S.M., Lawrence, M.J., and Barlow, J., Biochem. Soc. Trans. Vol. 18, p 943 (1990).

30. Naumann, D., Schultz, C., Sabisch, A., Kastowsky, M., and Labischinski, H., J. Mol. Struct. Vol. 214, pp 213-246 (1989).

31. Weiss, M.S. et al., Science Vol. 254, pp 1627-1630 (1991).

32. Jansonius, J.N., In Crystallography in Molecular Biology, pp 229-240, D. Moras, J. Drenth, B. Strandberg, D. Suck, and K. Wilson, Eds., Plenum Press, New York, NY, 1987.

33. Wright, C.S., J. Mol. Biol. Vol. 141, pp 267-291 (1980).

34. Hartmann, M., and Zbiral, E., Liebigs Ann. Chem., pp 795-801 (1991).

35. Ashwell, G., and Morell, A.G., Adv. Enzymol. Vol. 41, pp 99-128 (1974).

36. Goodenough, U.W., Am. Sci. Vol. 79, pp 344-355 (1991).

37. Marchesi, V.T., Ann. Rev. Cell. Biol. Vol. 1, pp 531-561 (1985).

38. Leonard, J.M., and Famini, G.R., A User's Guide to the Molecular Modeling, Analysis and Display System (MMADS), CRDEC-TR-030, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, January 1989, UNCLASSIFIED Report (AD A205 674).

39. Flippen, J.L., Acta. Cryst. Vol. B29, pp 1881-1886 (1973).

40. Pearson, R.G., Acc. Chem. Res. Vol. 4, pp 152-160 (1971).

41. Kopman, G., J. Am. Chem. Soc. Vol. 90, pp 223-234 (1968).

42. Scrocco, E., and Tomasi, J., Adv. Quantum Chem. Vol. 11, p 115 (1978).

43. Politzer, P., and Truhlar, D.G., Plenum Press, New York, NY (1981).

44. Politzer, P., and Murry, J.S., In Structure and Reactivity (Molecular Structure and Energetics), J.F. Liebman, and A. Greenberg, Eds., VCH Publishers, New York, NY, 1988.

45. Heller, H., Schaefer, M., and Schulten, K., In High Performance Computing and Challenges in Structural Biology, P.A. Bash and H. Meirovitch, Eds., Florida State University, Tallahassee, FL, 1992.