

THE COOPER UNION FOR THE ADVANCEMENT OF SCIENCE AND ART
ALBERT NERKEN SCHOOL OF ENGINEERING

Structural Characterization of N-Acetyl-2-Aminofluorene and its Guanine and
Deoxyguanosine Adducts via a Molecular Mechanics, Semi-empirical, and Density
Functional Theory Cascade Method

by

Timothy A. Isgro

A thesis submitted in partial fulfillment of the requirements for the degree of Master of
Engineering

5/5/03

Professor Robert Q. Topper, Advisor

THE COOPER UNION FOR THE ADVANCEMENT OF SCIENCE AND ART
ALBERT NERKEN SCHOOL OF ENGINEERING

This thesis was prepared under the direction of the Candidate's Thesis Advisor and has received approval. It was submitted to the Dean of the School of Engineering and the full Faculty, and was approved as partial fulfillment of the requirements for the degree of
Master of Engineering.

Dean, School of Engineering 5/5/03

Prof. Robert Q. Topper 5/5/03
Candidate's Thesis Advisor

To my Grandparents,

who started the family that helped me achieve everything I have, and whose love has been immeasurable in guiding me toward becoming the student and the man that I am.

Acknowledgements

I would like to thank first and foremost Dr. Robert Topper for his invaluable advice concerning both the present work and past projects on which I have participated under his tutelage. Furthermore, I am grateful for his endless patience in the seeing this work completed. I would like to thank my parents and brother for providing a lifetime of love and support which enabled me to achieve that which I have today. I would also like to thank Josephine Scalici, not only for some useful conversations relating to the biochemistry of the current work, but for her unending support as I completed this work.

I am grateful to The Cooper Union for the Advancement of Science and Art for not only providing the Enders' Fellowship in Chemical Engineering, which helped finance the current work, but also for the intense graduate and undergraduate educations I have received here. I would specifically like to thank the faculty of the Chemistry and Chemical Engineering Department for providing years of instruction which I will carry for the rest of my life.

Abstract

N-acetyl-2-aminofluorene (AAF) is a potent carcinogen which forms two adducts with DNA upon reaction both *in vivo* and *in vitro*. It is believed that the minor adduct may be important in the formation of cancer due to its persistence *in vivo*. Four varieties of modified AAF have been studied computationally; AAF bound to C8 of guanine (G) or deoxyguanosine (dG), and AAF bonded to N² of guanine or deoxyguanosine. A computational cascade method has been employed to locate and predict all of the low energy conformers of each molecule. Each was subjected to a systematic molecular mechanics conformer search with the MMFF94 force field, and each resulting conformer was then geometry optimized via the semi-empirical AM1 method. A subset of the resulting conformers was then submitted to density-functional theory (DFT) optimization. Confidence in the method was established via comparison of the structure of the major adduct bound to guanine with x-ray crystallographic data. The major adduct structures reveal a preference for adoption of one of two orientations, with guanine or deoxyguanosine oriented either perpendicular to the fluorenyl ring moiety or approximately 45° with respect to the fluorenyl structure. Repulsion between the ribose sugar and acetyl group in the deoxyguanosine case disrupts the “ideal” dihedral angles observed in the guanine adduct. Results for the minor adduct similarly point toward two general conformer classes, one of which is characterized by hydrogen bonding between the acetyl O and the H bonded to N² of guanine, and the other of which is characterized by minimization of repulsion between the guanine structure and the acetyl group. In the deoxyguanosine case, results differ somewhat due to an attractive interaction between fluorene and guanine hydrogen atoms and ribose group oxygen atoms.

Table of Contents

Section	Page
1. Introduction and Background.....	1
1.1. DNA Background.....	1
1.2. Base Pairing and Carcinogen Biochemistry.....	8
1.3. The Carcinogenic Activity of N-acetyl-2-aminofluorene (AAF)	13
1.4. AAF Adduct Structure.....	18
2. Methodology.....	22
2.1. SPARTAN Program and Its Applications.....	22
2.2. Molecular Mechanics.....	24
2.2.1. Theory.....	24
2.2.2. Force Fields and MMFF94.....	30
2.3. Quantum Mechanics.....	35
2.3.1. Semi-empirical Methods and Theory.....	35
2.3.2. Density Functional Theory.....	38
2.4. Strategy and Progression of Study.....	47
2.4.1. Approach to Calculating Structures.....	47
2.4.2. AAF Alone and Choosing Force Fields.....	52
3. Results.....	55
3.1. G-C8-AAF.....	55
3.2. G-N ² -AAF.....	61
3.3. dG-C8-AAF.....	66
3.4. dG-N ² -AAF.....	74

4. Discussion.....	81
4.1. Confidence.....	81
4.2. Significance of the Study and Carcinogenicity.....	83
5. Future Direction for Study.....	86
6. Conclusions.....	87
7. Appendix - Conformer Energy and Dihedral Angle Tables.....	89
8. References.....	94

List of Figures

1. Figure 1.1. Purine, pyrimidine, and nucleic acid bases.
2. Figure 1.2. A 4-nucleotide single-stranded representation of DNA.
3. Figure 1.3 Hydrogen bonding in DNA.
4. Figure 1.4. Differences between keto and enol forms of nucleic acids.
5. Figure 1.5. Metabolic activation pathways for dG-C8-AF, dG-C8-AAF, and dG-N²-AAF.
6. Figure 1.6. Structures of AAF, G-C8-AAF, G-N²-AAF, dG-C8-AAF and dG-N²-AAF.
7. Figure 1.7. Adduct structures and pertinent dihedral angles for dG-C8-AAF and dG-N²-AAF.
8. Figure 2.1. Representation of torsional twisting.
9. Figure 2.2. A hypothetical 1-D potential energy surface.
10. Figure 2.3. Lowest energy conformers from a 15x15x15 conformer search using a)AM1 or b)PM3.
11. Figure 3.1. Conformers 2 and 6 of G-C8-AAF at the BP/DN** level.
12. Figure 3.2. Conformer 7 of G-N²-AAF at the BP/DN** level.
13. Figure 3.3. Conformer 4 of G-N²-AAF at the BP/DN** level.
14. Figure 3.4. Conformer 1 of dG-C8-AAF at the BP/DN* level.
15. Figure 3.5. Conformers a)4 and b)9 of dG-C8-AAF at the BP/DN* level.
16. Figure 3.6. Conformers a)1 and b)2 of dG-C8-AAF at the BP/DN* level.
17. Figure 3.7. Conformer 6 of dG-N²-AAF at the BP/DN* level.
18. Figure 3.8. Conformers a)1 and b)9 of dG-N²-AAF at the BP/DN* level.

19. Figure 3.9. Conformer 5 of dG-N²-AAF at the BP/DN* level.
20. Figure 3.10. Conformer 7 of dG-N²-AAF at the BP/DN* level.
21. Figure 3.11. Conformer 6 of dG-N²-AAF at the BP/DN* level.

List of Tables

1. Table 1.1. Geometries of DNA.
2. Table 2.1. Conformer energies and dihedral angles for cascade A for DFT BP/DN.
3. Table 2.2. Conformer energies and dihedral angles for cascade B for DFT BP/DN.
4. Table 7.1. Dihedral angles, conformer search parameters, and number of energy-minimized conformers used in MMFF94 conformer searches.
5. Table 7.2. G-C8-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.
6. Table 7.3. G-N²-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.
7. Table 7.4. dG-C8-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.
8. Table 7.5. dG-N²-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.

1. Introduction

1.1. DNA background

The determination of the structure of deoxyribonucleic acid, DNA, by James Watson and Francis Crick in 1953 did much more than paint a picture (or build a model) for the housing of genetic information. Their proposal fittingly sparked life to the fields of molecular biology and biochemistry, which have contributed in countless ways to the improvement of human health. Their revolutionary hypothesis was an imaginative amalgamation of various pieces of information from several sources over prior years. Most important among these pieces were Chargaff's rules, the identification of DNA as a helical molecule, and the correct tautomeric forms of the purine and pyrimidine bases which comprise the interior of the famed double-helix.

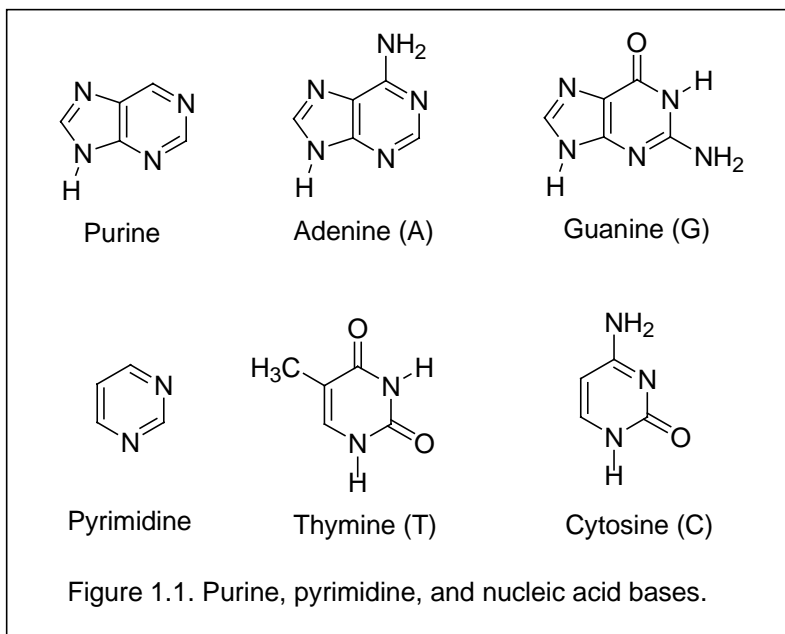
Erwin Chargaff's quantitative analysis of the makeup of DNA in the late 1940's revealed an interesting trend in the amounts of nitrogenous bases in the molecule. Namely, Chargaff discovered that there existed an equal amount of adenine (A) and thymine (T), as well as an equal amount of guanine (G) and cytosine (C) in every strand of DNA. Although the specific amounts of adenine and thymine versus guanine and cytosine may vary widely from organism to organism, the amounts generally remained constant among related species, with the amount of guanine and cytosine ranging from 39 to 46% in mammals. Chargaff's rules led Watson and Crick to propose base pairing of these molecules as the qualitative explanation for this intriguing piece of data.

The x-ray crystallography work of Rosalind Franklin went far in revealing the structure and general shape of the DNA molecule. Franklin's diffraction photographs allowed Crick to deduce the shape of the molecule as helical and also led him to conclude

that the interior of the helix was filled by aromatic bases stacked in such a manner that their aryl rings were perpendicular to the helical axis. Without this important crystallography data, it would have been an incredibly difficult task to determine the nature of a strand of DNA.

The forms of those bases which filled the inside of the DNA helix were firmly established by a great deal of NMR, x-ray crystallography, and spectroscopic research which had taken place years before. The dominant tautomeric forms of the four bases were not fully appreciated until Watson and Crick developed their model and identified base pairing through hydrogen bonding as vital in allowing a molecule of DNA to store genetic information.

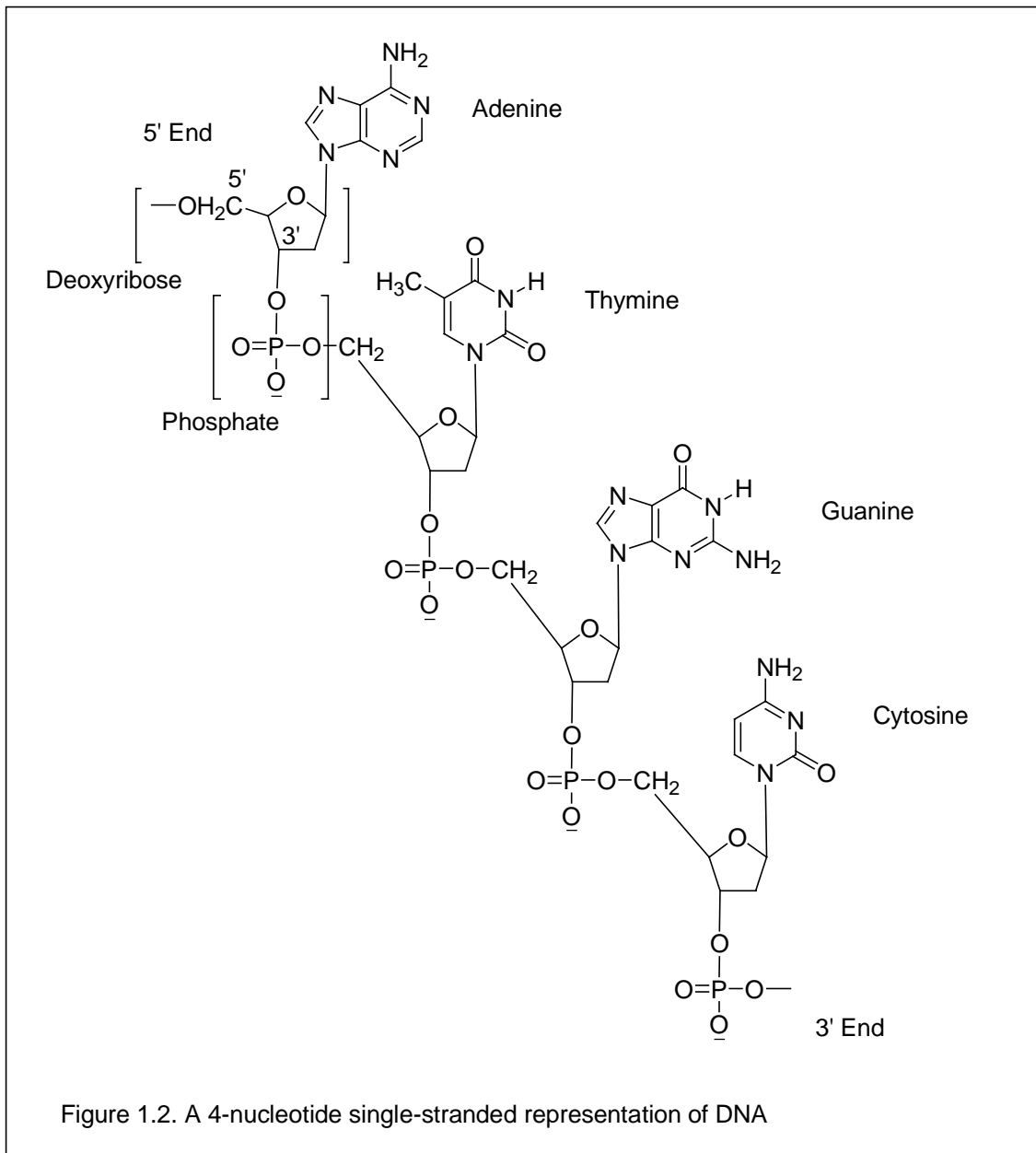
Armed with the previous information, Watson and Crick used model building to compose the correct structure of DNA, which is made up of four nitrogenous bases



adenine, thymine, guanine, and cytosine (Figure 1.1) in polynucleotide strands with deoxyribose-phosphate, or phosphodiester, linkages (Figure 1.2). Two strands link to one another

through hydrogen bonding of the bases (Figure 1.3) and wind around each other to form a right-handed double-helix. In this manner, the bases fill the interior of the helix and stack

themselves in base pairs of adenine bonded to thymine and guanine bonded to cytosine, consistent with Chargaff's rules. The data mentioned on major tautomeric forms of the nitrogenous bases becomes important when one considers the manner in which the aromatic bases hydrogen bond to one another. The dominant keto form (Figure 1.4) among bases enables the Watson-Crick base pairing of A to T and G to C as shown in Figure 1.3. Other hydrogen bonding arrangements are possible between bases; however, the Watson-Crick forms are the most stable under the double-helix conformation. Other base pairing modes, whether or not they have been experimentally observed, can lead to sequence errors but may only take place by inducing abnormal strain on the helix. This safeguard is an important one in allowing only correct interactions to take place and in preventing incorrect ones (Hoogsteen pairs). In this manner, DNA protects itself from a misrepresentation of information should a mismatch between strands happen to occur. The phosphodiester linkages, which hold the polynucleotide chain together, run along the outside of the double-helix and provide a negative charge along the exterior of the molecule. Furthermore, complementary strands run anti-parallel with one 5' to 3' strand pairing with another 3' to 5' strand, as shown in Figures 1.2 and 1.3. In minimizing interaction between negatively charged phosphate groups, the DNA molecule adopts a conformation such that its double-helix contains a major and minor groove, both of which wind their way along and around the axis of the helix.



It should be noted that the work of Watson and Crick was performed on the most biologically common B form of DNA. Also known to exist are two other structural varieties, A-DNA and Z-DNA, which may exist in appropriate solvent and base-sequence conditions. DNA has been known to undergo conformational rearrangements from one form to another, provided the existence of unusual environments which are not thought to exist *in vivo*. Perhaps the most striking features of the two other forms are as follows.

The base pairs of A-DNA are tilted 20° with respect to the helical axis, which allows the conformer to have a very deep major groove. Z-DNA, surprisingly, exists as a left-handed double-helix. A summary of the geometries of all three types is given in Table 1.1.

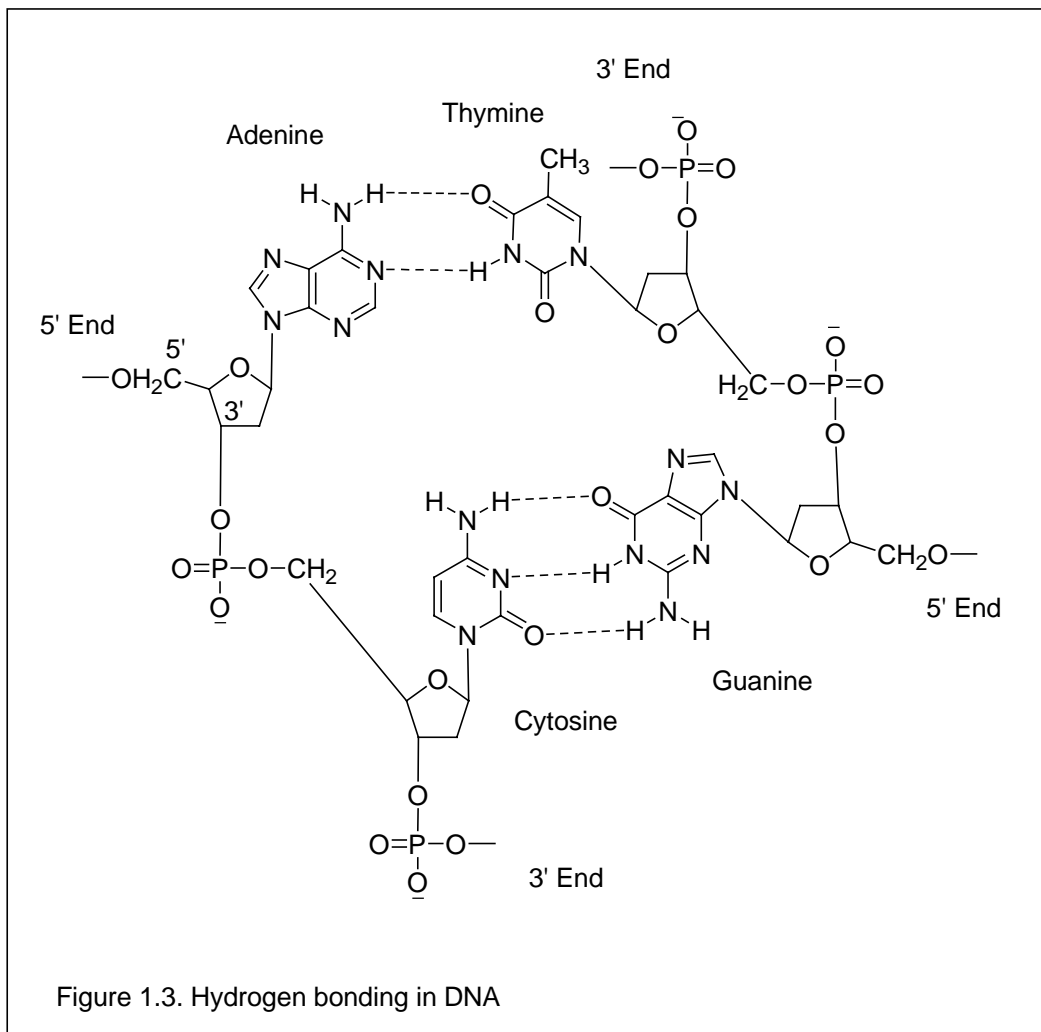
Feature	A-DNA	B-DNA	Z-DNA
Helical direction	Right-handed	Right-handed	Left-handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10	12 (6 dimers)
Helical twist per base pair	33°	36°	60° (per dimer)
Helical pitch (rise per turn)	28 Å	34 Å	45 Å
Helical rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to helical axis	20°	6°	7°
Major groove	Narrow, deep	Wide, deep	Flat
Minor groove	Wide, shallow	Narrow, deep	Narrow, deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines, C3'-endo for purines
Glycosidic bond	<i>Anti</i>	<i>Anti</i>	<i>Anti</i> for pyrimidines, <i>syn</i> for purines

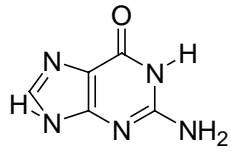
Table 1.1. Geometries of DNA (1).

The four nitrogenous bases which make up the interior of the DNA molecule encode genetic information by the sequence in which they appear along the strand. Upon replication, a molecule of DNA first divides so as to allow the sequence of bases in one chain to be read and copied as a complementary strand of messenger RNA (mRNA) in what is known as transcription. The single-stranded mRNA may then bind to a molecule of transfer RNA (tRNA) three bases at a time. It is by this recognition of three-base sequences, called codons, that translation occurs, in which the molecule of tRNA determines the correct amino acid residue in the formation of a protein. Since three bases code for one amino acid, a total of 64 amino acids may be accommodated by the genetic

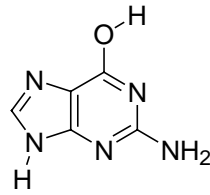
code. However, since 20 standard amino acids are known to exist, the genetic code is degenerate, in that one amino acid may correspond to different and, in many cases, similar codons. In this respect more protection is provided against transcription or translation errors, since a misread actually has a chance for representing the correct amino acid.

With such naturally incorporated protection, *in vivo* DNA replication can occur flawlessly over and over again, but the process is not foolproof, and errors in DNA replication are believed to be of primary importance in the formation of cancer.

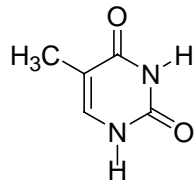




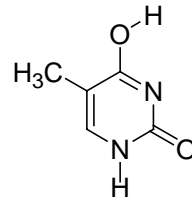
Guanine in
keto form



Guanine in
enol form



Thymine in
keto form



Thymine in
enol form

Figure 1.4. Differences between keto and enol forms of nucleic acids.

1.2. Base pairing and carcinogen biochemistry

The behavior of chemical mutagens which have a propensity to bind to DNA is of great interest to biologists or biochemists working to understand the genes of cancer. Alteration of DNA structure is believed to be a primary mechanism by which carcinogens may initiate cancer formation. In binding to DNA, a molecule may alter the chemical and electronic properties of the DNA strand in such a way that its former identity is unrecognizable to the molecules involved in the replication process. Thus, the base sequence of the strand may be interpreted incorrectly. Such errors are often harmless, however, occurring in non-coding regions of DNA, and the results of a point mutation in the strand may be inconsequential. On the other hand, should a nucleic acid base in a strand of DNA or RNA be interpreted incorrectly and the wrong codon formed, an amino acid may be misplaced in the formation of a protein, and the operation of the protein could be destroyed, particularly if the misrepresentation occurs in a functional part of the protein.

A chemical mutagen which infiltrates a macromolecule of DNA may produce a mutation, which will generally fall into one of two classes, that of a point mutation or of an insertion/deletion mutation. Point mutations describe alterations which occur when one base replaces another. In certain environments, a nitrogenous base may become converted such that its original base pairing partner may be unable to form hydrogen bonds with it, whereas an otherwise false partner will be able to. In this case, the original partner may be cleaved from the DNA strand in favor of the false partner, which may ultimately lead to incorrect replication. Mismatches in which a purine or pyrimidine is replaced by a purine or pyrimidine, respectively, are known as transitions, while

mismatches in which a purine is replaced by a pyrimidine, or vice versa, are known as transversions.

Insertion/deletion mutations are usually the result of intercalating agents infiltrating a strand of DNA and resting between base pairs. Replication is altered to the point that one or more bases are deleted entirely. In this respect, a frameshift mutation may result, whereby the remainder of the coding sequence is incorrectly interpreted via an effective “domino effect” due to the insertion or deletion of a single base pair.

Preventing mutations such as these is vital in halting the formation of faulty DNA, which may lead to tumor formation. In regard to this, the study of classes of molecules which have a penchant for DNA binding is key in discovering and thus reducing mutagenic alterations. One of these classes, perhaps the most notorious among carcinogenic molecules, is that of the polycyclic aromatic hydrocarbons (PAHs). Numerous reviews have been written on the tumorigenic potency of PAHs (2-6), which are found in a variety of sources including tobacco smoke and barbecued meats and fish. The existence of PAHs in the environment can be mainly attributed to the incomplete combustion of fossil fuels, and, as a result, low-level PAH contamination may also be found in the air, oceans, soil, as well as on the surface of fruits and vegetables and in drinking water (2). Within the classification of PAHs exists a subset of molecules whose carcinogenic activity is also prominent.

The aromatic amines were some of the most hazardous substances in an early study by Case et al. of tumorigenic chemicals and their connection to cancer in industrial dye manufacturing (7). The British review could hardly have been a surprise, however, since a German clinician, Ludwig Rehn, first reported the trend of bladder cancer

development among aniline dye workers in 1895 (8). By 1942, as Kriek describes, more than 300 cases of bladder cancer in German chemical workers had been documented (9). Also among a 1988 review of the carcinogenic activity of 506 chemicals, the aromatic amines were the most well represented class with 70 molecules (10). Weisburger's review of the same year provides a historical perspective of the carcinogenic activity of aryl amines and amides, as well as their metabolic effects (11). The works of Clayson (12), Hueper (13), and Scott (14) provide a rather complete listing of the aromatic amines and related compounds known to induce cancer in animals. The carcinogenic activity of aryl amines and amides is indeed well-known and accepted, and studies over the years have established certain specific structural characteristics of these molecules which are associated with carcinogenic properties.

Many early so-called structure-activity studies done on aromatic amines yielded interesting data concerning the position of the amino group and carcinogenic activity (12, 15-19). Carcinogenic potency was connected to the substitution of the amino group at the beta or 2-positions in polynuclear aromatic hydrocarbons and at the *para* position in polyphenyl compounds. Furthermore, absence of carcinogenic activity was generally the case for molecules whose amino group was in the alpha or 1-position or the *ortho* position, respectively. Miller and Miller (16) and Arcos and Argus (17) have studied the effect of adding methyl groups or halogens, which do not always have the effect of increasing the carcinogenicity of a molecule. Ring system size polycyclic aryl amines and amides generally obey the same trend for smaller sized molecules. In general, a larger ring system translates to increased carcinogenicity, as the works of Lotlikar (20) and Weisburger and Fiala (21) demonstrate.

The binding of aromatic amines to DNA *in vivo* was the subject of a 1973 review by Irving (22). Evidence suggests that upon incorporation into a cell, most aromatic amines undergo an activation mechanism, which shall be described in the next section, whereby they retain their basic identity, but become altered such that they may react mainly with the purines in DNA, particularly at the C8 position of guanine (23-25). It should be noted that this is no simple task, however, as the C8 position is not readily accessible in an undistorted strand of B-DNA. Some degree of structural rearrangement must first occur before the infiltration of a mutagen, as suggested by Duane and co-workers (26). Most of the structure-activity correlations mentioned can be explained as their relationship to metabolic reactions which cause the activation of the molecule to reactive mutagenic form (24, 27-29). N-hydroxylation is an important step in this process for aromatic amines and amides, and the degree to which this activation is hindered or helped may depend significantly on the molecular geometry.

Kriek (9) mentions the fact that the active forms of most carcinogenic molecules are strong electrophiles and Miller (23) has demonstrated this relationship between the carcinogenic potency and electrophilic properties. In some cases activation of the source carcinogen makes DNA invasion possible by increasing the electrophilic nature of the chemical.

Although the invasion of a DNA molecule by a single mutagen molecule may seem sufficient to initiate tumorigenesis, it is generally accepted that several point mutations are necessary for the formation of cancer, as has been discussed by Bishop (30). Furthermore, after attack and intrusion has taken place, many defense mechanisms exist for the eradication of the lesion. An insightful review of DNA lesion repair

mechanisms is provided by Braithwaite et al. (31) in which four main types of DNA repair mechanisms in humans are identified. They are base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), and recombination repair. BER and NER are involved in the cleaving of error sequences from a strand of DNA, while MMR corrects mismatched bases during replication. Recombination repair is mainly involved in fixing breakages of a DNA double strand. Countless studies have been performed on the ability of certain DNA enzyme polymerases and nucleases to effectively detect and repair mutations and the rate of repair of such enzymes. The reader is referred to some (32-36) related to the compound of interest in this study, N-acetyl-2-aminofluorene, whose relationship to the study of carcinogenesis is expounded upon in the next section.

1.3. The Carcinogenic Activity of N-acetyl-2-aminofluorene (AAF)

In 1941, the US Department of Agriculture had been involved in testing a variety of organic compounds for use as insecticides. One insecticide which had been claimed to work contained 2-aminofluorene (AF) as its essential ingredient (37). Before it or its more stable counterpart, N-acetyl-2-aminofluorene (AAF) could be used, one criterion needed to be proven: the nontoxicity of the compounds toward man or animal. As a result, in the same year, Wilson et al. tested for and discovered the carcinogenic activity of AAF in rats (38), and the compound was never used as an insecticide. Since then, AAF has been examined extensively as a prototype for the tumorigenic effects of polycyclic aromatic hydrocarbons. In fact, the study of AAF as a cancer causing agent has been so widespread that several reviews of the abundant documented work on the aryl amide have been written (37, 39-41).

AAF alone will not bind to the nucleic acids of DNA, and, like many other carcinogenic molecules, must undergo some degree of activation before bonding is possible. Cramer et al. discovered the initial activation reaction for aromatic amines and amides to be N-oxidation, whereby atomic oxygen reacts at the amine or amide nitrogen site (42). During the same year, Herlinglake et al. discovered the N-oxidation reaction for 2-naphthylamine (43). Evidence for the activation is found in the increased carcinogenic activity of N-hydroxy-AAF (N-OH-AAF) as compared to AAF in a variety induced tumors in several mammalian species (44). Studies of the oxidation of AAF to N-OH-AAF and its importance in the activation of AAF to a potent mutagen, as well as its increased carcinogenicity as compared to the parent compound, resulted in the identification of N-OH-AAF as the first example of a *proximate* carcinogen. The

oxidation is catalyzed by cytochrome P-450, and a review of isozymes capable of causing the oxidation has been given by Kadlubar and Hammons (45).

After oxidation, three nucleic acid adducts are predominantly formed, with the fractional composition depending on whether the intermediate activation takes place *in vivo* or *in vitro*. *In vivo*, the largest fraction of adduct is the deacylated N-(deoxyguanosin-8-yl)-2-aminofluorene (dG-C8-AF) (46-59), with the balance of product being composed of the acylated N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (dG-C8-AAF) and 3-(deoxyguanosin-N²-yl)-2-acetylaminofluorene (dG-N²-AAF). The deacylated adduct may be formed via several routes depending on which metabolic pathways dominate from species to species. In rats, the adduct appears to be formed due to the N,O-acyltransferase-catalyzed reaction in which N-OH-AAF becomes N-acetoxy-2-aminofluorene (N-OAc-AF), which then reacts directly with DNA (56, 60-64). Mice have low N,O-acyltransferase activity (62), and, as a result, another pathway seems to dominate, whereby N-OH-AAF is deacylated to N-hydroxy-2-aminofluorene (N-OH-AF), which is converted into a guanosine reactive N-sulfonyl ester (58, 59).

In vitro, there exists a greater fraction the acylated adducts, dG-C8-AAF and dG-N²-AAF with 60-80% of the total product existing in dG-C8-AAF form. Both adducts are formed via sulfotransferase reaction whereby N-OH-AAF forms N-sulfonyl-2-acetylaminofluorene (49, 58, 59) which then reacts with DNA to form either the major (dG-C8-AAF) or minor (dG-N²-AAF) adduct. The formation of the major and minor adducts may also result from reaction of N-acetoxy-2-acetylaminofluorene (N-OAc-AAF), which is formed from the reaction of N-OH-AAF and acetyl coenzyme A (65).

The difference between *in vivo* and *in vitro* cases may arise from a number of factors, as discussed by Howard et al. (66). *In vitro* conditions may favor the metabolic pathways which lead to an increased formation of the acylated adducts. Acyltransferase, for instance, has been shown to be very dependent on O₂ concentration (67), and as a result, the formation of N-OAc-AF may be inhibited due to high oxygen concentrations present during cell incubation. The various metabolic activation pathways which may lead to the formation of dG-C8-AF, dG-C8-AAF, and dG-N²-AAF are shown in Figure 1.5, adapted from the review by Beland and Kadlubar (39).

The major and minor adducts behave quite differently when incorporated into DNA. Whereas the major adduct is excised quite rapidly from DNA (66), with a half life of approximately seven days in rat liver, the minor adduct appears to be relatively persistent (50, 51, 53-55, 57, 68, 69). Numerous experimental and theoretical studies indicate that the reason for the poor ability of the major adduct to avoid being cleaved from DNA lies in the geometry it adopts upon infiltrating the DNA molecule. In B-DNA, deoxyguanosine normally adopts an *anti* configuration about its glycosyl bond. This bond and its *anti/syn* domains will be explained in the next section. When the major adduct becomes incorporated into a molecule of B-DNA, it preferentially adopts a *syn* conformation which distorts the B-DNA molecule somewhat (70-75). The distortion makes the geometry resemble that of a Z-DNA molecule, and transitions between B and Z form have been shown to take place (76-78). The *anti-syn* distortion causes the AAF to be inserted into the interior of the DNA helix, which results in a denaturation of ~ 12 base pairs (79). The process has been termed “base displacement” (80) or “insertion-denaturation” (81), and the large amount of distortion present in the modified DNA is

believed to be the primary reason for the short half life of the major adduct mutation. Repair enzymes easily recognize the helical deformation as foreign and cleave the DNA. Experimental structural determination of the minor adduct is not available as of yet due to low yields. The study by Kriek (68) was the first to propose a structure for the minor adduct. A theoretical investigation by Beland (82) suggested that the minor adduct, which is much more persistent than the major form, is able to reside in the minor groove of the DNA molecule, producing little distortion and allowing it to evade repair enzymes. The ability of the minor adduct to evade repair mechanisms in comparison to the major adduct is fundamental to the carcinogenic behavior of the molecule. Thus, elucidation of both adduct structures and behaviors when bound to DNA or deoxyguanosine is a primary goal in understanding how AAF causes cancer.

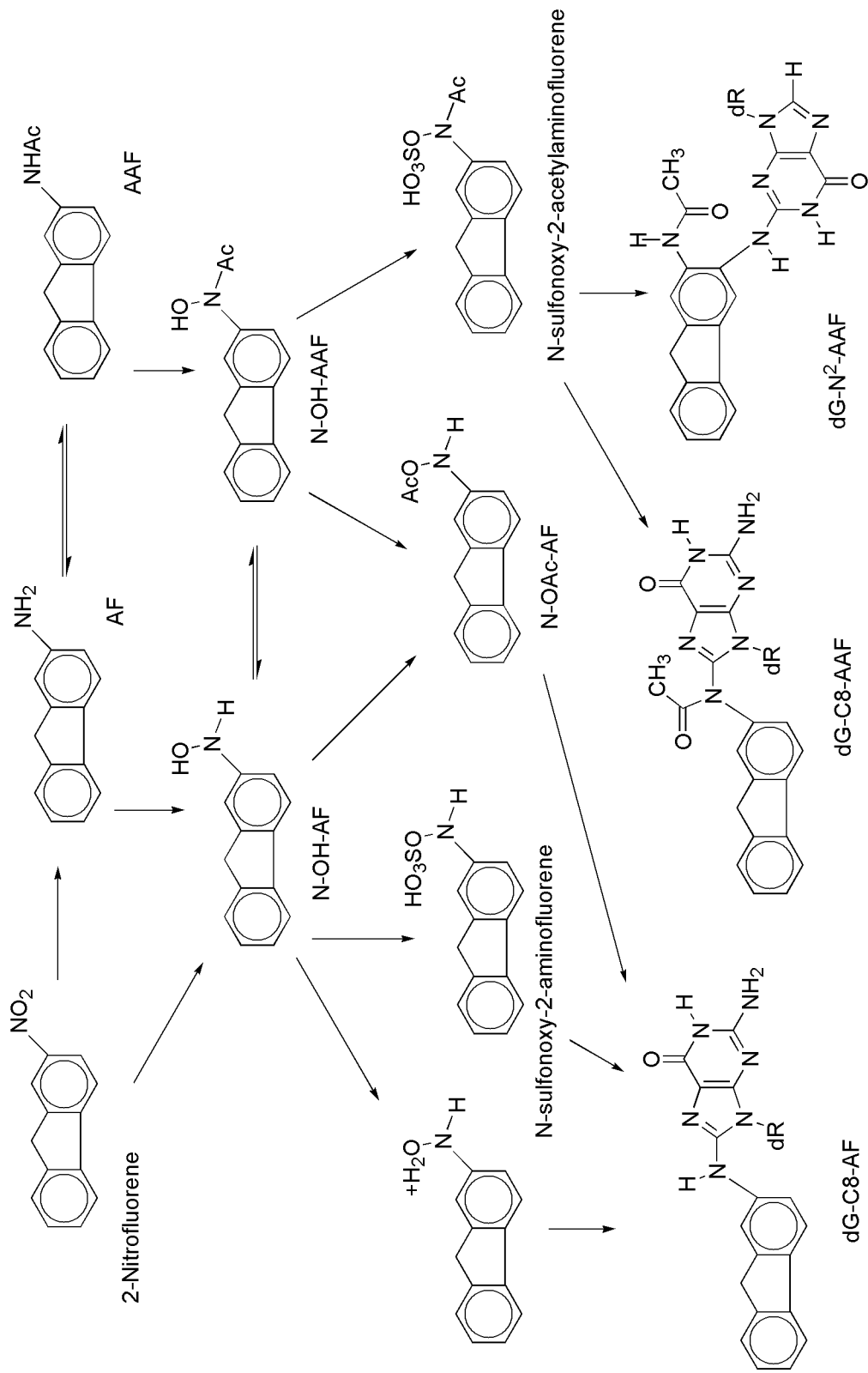


Figure 1.5. Metabolic activation pathways for dG-C8-AF, dG-C8-AAF, and dG-N²-AAF. Ac represents the acetyl group, and dR represents the deoxyribose group.

1.4. AAF Adduct Structure

Experimental determination of the structure of AAF bound to guanine in the major adduct configuration (G-C8-AAF) using x-ray crystallography and NMR techniques has been carried out by Neidle et al. (73). The major conformers observed indicate an approximate perpendicularity of the planar ring structures of the guanine and fluorenyl moieties. The conformation serves to ease the steric interference present between the fluorene and guanine ring systems.

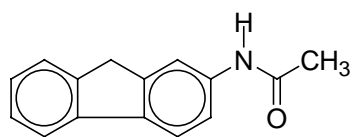
Computational efforts to elucidate the structure of the major adduct in deoxyguanosine bound form have been undertaken using molecular mechanics and dynamics (83) and the semi-empirical MNDO method (84). A mechanics study by Fritsch and Westhof (83) used two sets of force field parameters under AMBER 3.0 and a distance-modified dielectric constant to obtain 20 minimized domains for the major adduct, while a study of the MNDO energy surface displayed two energy minima (84) which only vaguely resemble domains 2 and 24 of the Fritsch and Westhof study. Based on a benchmark computational study of AAF by Topper et al. (85), however, it is believed that the MNDO method is unsuitable for geometry calculations on AAF and its adducts. More explanation is provided in the Methodology section.

Minimized potential energy calculations on single- and double-stranded oligonucleotides in alternating CG sequences modified with AAF have also been carried out (72, 74, 86, 87). The *anti-syn* transition observed in these calculations when the major adduct binds to the oligomer has already been mentioned. Shapiro et al. (88) extended these studies to single-stranded dimers containing each of the four nucleotides adjacent to the major adduct modified guanine. The results confirmed the preference for

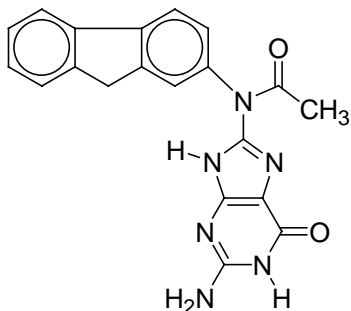
a *syn* glycosidic bond and indicated base stacking disruption, replaced by a degree of carcinogen-base stacking in certain cases. O'Handley et al. (89) studied a major adduct-modified DNA oligomer (9-mer) via NMR, energy minimization, and molecular dynamics, providing more evidence for the above.

Molecular mechanics and dynamics calculations of the minor adduct of AAF incorporated as a lesion into a heptamer double-helix of DNA by Grad et al. (90) showed relaxation of AAF in the minor groove of the helix confirming the initial suggestion of Beland (82). This paper also stressed the importance of sequence dependence, as suggested by Shibutani and Grollman (91), based on the differences in orientation direction adopted by the fluorenyl moiety. Its results will be discussed with respect to the geometries calculated for the present study. These are the only data for the structure of the minor adduct in the literature to date.

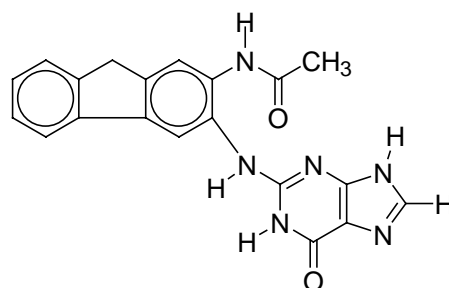
The present study seeks to obtain structural information via computation for two adducts of AAF each in two forms; AAF in major and minor adduct form bound to guanine, G-C8-AAF and G-N²-AAF, and AAF in major and minor adduct form bound to deoxyguanosine, dG-C8-AAF and dG-N²-AAF. The structures of each of the adducts are shown in Figure 1.6, and the dihedral angles pertinent to the present study are shown in Figure 1.7 for the adducts bound to deoxyguanosine. For the guanosine forms, the angles χ and χ^* are omitted. χ is known as the glycosidic dihedral angle. It governs the orientation of the nucleic acid base relative to the backbone ribose group in a DNA molecule. It is described as being in either the *syn* ($-180^\circ < \chi < -90^\circ$ and $90^\circ < \chi < 180^\circ$) or *anti* ($-90^\circ < \chi < 90^\circ$) domain, with the preference for the three forms of DNA shown in Table 1.1.



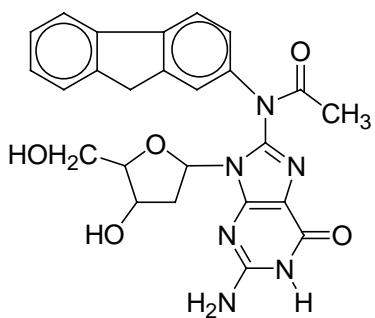
N-acetyl-2-aminofluorene
AAF



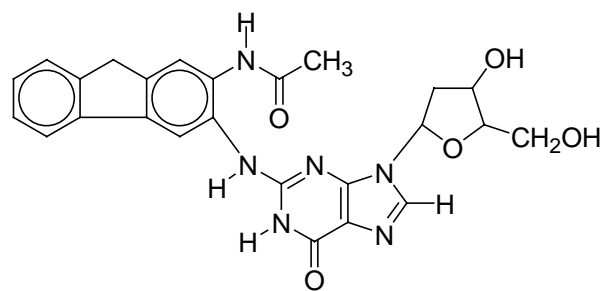
Major Adduct (guanine form)
G-C8-AAF



Minor Adduct (guanine form)
G-N²-AAF



Major Adduct (deoxyguanosine form)
N-(deoxyguanosin-8-yl)-AAF
dG-C8-AAF



Minor Adduct (deoxyguanosine form)
3-(deoxyguanosin-N²-yl)-AAF
dG-N²-AAF

Figure 1.6. Structures of AAF, G-C8-AAF, G-N²-AAF, dG-C8-AAF and dG-N²-AAF.

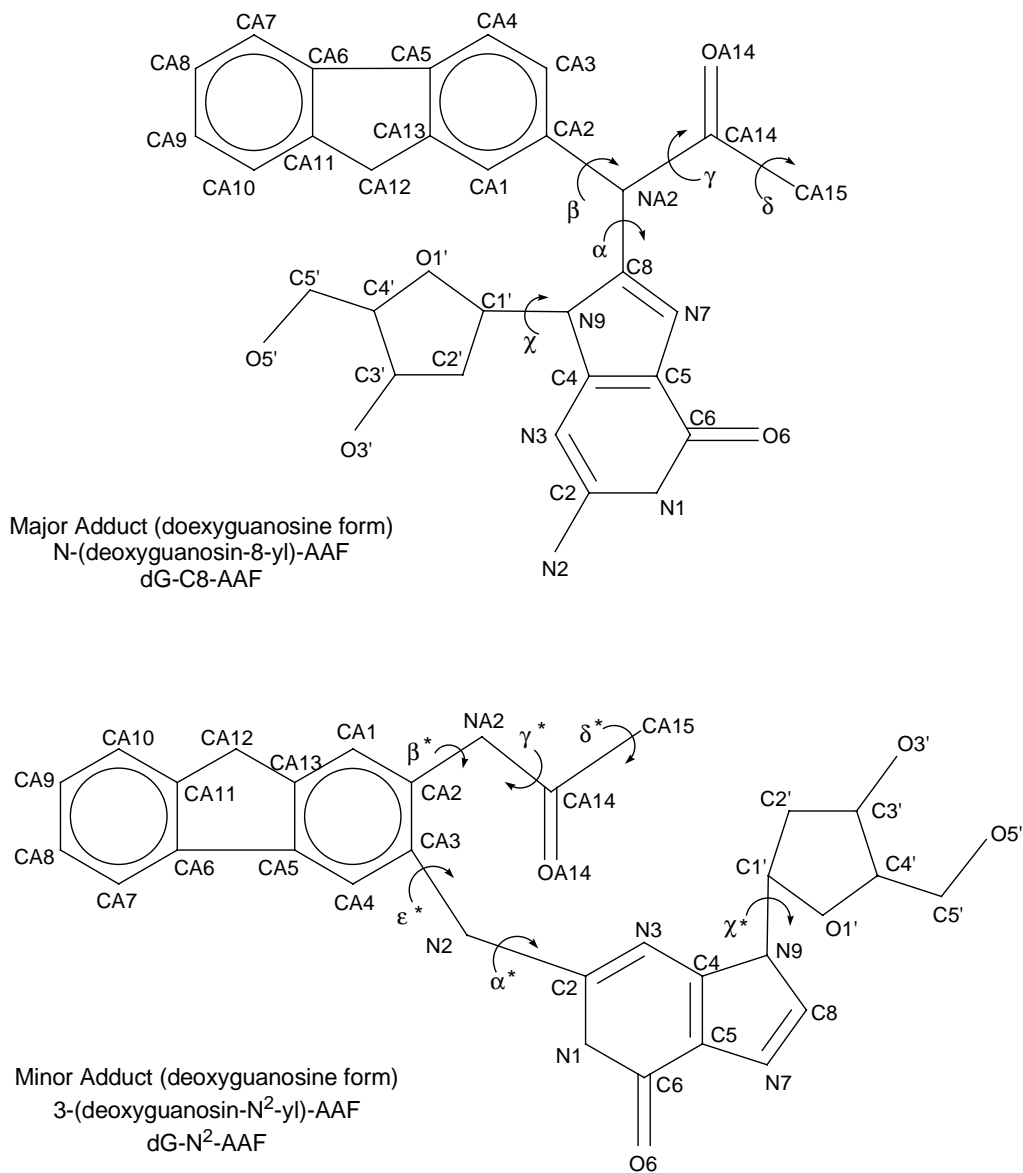


Figure 1.7. Adduct structures and pertinent dihedral angles for dG-C8-AAF and dG-N²-AAF. G-C8-AAF and G-N²-AAF have a H in place of the ribose sugar. Dihedral angles are α : N9-C8-NA2-CA2, β : C8-NA2-CA2-CA1, γ : C8-NA2-CA14-CA15, χ : O1'-C1'-N9-C8, and α^* : N1-C2-N2-CA3, β^* : CA3-CA2-NA2-CA14, γ^* : CA2-NA2-CA14-CA15, ϵ^* : C2-N2-CA3-CA2, χ^* : O1'-C1'-N9-C8. δ is given as NA2-CA14-CA15-HA15, and was only used in a preliminary study of AAF. Changes in its value correspond to rotations of the methyl group of CA15.

Methodology

2.1. SPARTAN Program and its Applications

SPARTAN is a molecular modeling program with a wide variety of uses for both experimental and theoretical chemists wishing to elucidate the properties of any class of compounds. SPARTAN utilizes both classic and quantum mechanical numerical calculations to determine the physical, electronic, and thermodynamic properties of an input molecule under investigation. Information for the following section was taken from the *SPARTAN Version 5.0 User's Guide* (92).

SPARTAN's architecture is divided into seven independent program modules, each of which may reside on the same or different computers. They are the graphical user interface, *ab initio*, density functional, semi-empirical, mechanics, properties, and graphics modules.

The graphical user interface is intended for convenience in creating and inputting any calculations performed with the *ab initio*, density functional, semi-empirical, and mechanics modules. The interface is also responsible for output displays in both text and graphical form, including molecular geometries, electron densities, dipole moments, and other structures calculated. Volumes, as well as isosurfaces and 2-D slices of surfaces from volume data may be displayed for individual molecules or as a comparison between multiple molecules which have been aligned. Animation of vibrational modes is also handled by the graphical user interface, along with the importing and exporting of molecular structures.

The *ab initio*, density functional, and semi-empirical modules each perform their designated quantum mechanical calculations which shall be described later. Each module

is capable of energy and wavefunction calculations for a single geometry, as well as the calculation of structures for both equilibrium and transition-state geometries. Calculation of the Hessian, which shall be described, and the evaluation of normal-mode vibrational frequencies and thermodynamic properties are among the capabilities of these modules, along with molecular conformer searching.

The mechanics module is responsible for molecular mechanics calculations of chemical properties. Among its capabilities are the enumeration of strain energies for a molecular geometry, evaluation of equilibrium geometries, the Hessian matrix, vibrational frequencies, thermodynamic properties, and conformer searching.

The properties module and the graphics module serve in the preparation of text and graphical output, respectively. The properties module calculates vibrational modes, the dipole moment, and solvation energies, along with the evaluation of thermodynamic properties. The graphics module controls the calculations for the mapping of properties to volumes and surfaces based on the wavefunction calculations of the *ab initio*, density functional, and semi-empirical modules.

The present study takes advantage of SPARTAN'S ability to calculate equilibrium geometries for each class of the four computational classes provided. Geometry calculations have been utilized in three of these four classes (no *ab initio* calculations were run) to examine the structure of AAF and its guanine bound adducts, and to study their effect in the carcinogenicity of the molecule.

A variety of publications have also been created for those needing instruction in a particular aspect of SPARTAN's capabilities (92-97).

2.2. Molecular Mechanics

2.2.1. Theory

Molecular mechanics (MM) theory and applications are the simplest and quickest of computational methods for determining molecular structure and properties. The theory is based purely in Newtonian mechanics, and is akin to regarding atoms and the bonds that binds them in molecules as an array of charged particles connected by springs. The theory disregards the often computationally expensive laws of quantum mechanics, favoring instead a representation of forces between atoms, in pursuit of stable molecular configurations. This facet of molecular mechanics methods provides a great advantage over quantum mechanical methods in computational time for identical systems. A benchmark study by Reynolds (98) illustrates typical relative computational times, such as 0.3 seconds for MM versus 2×10^5 seconds for a higher order quantum mechanical method in the optimization of a 29-atom molecule. MM methods also allow one to examine very large ($\sim 10,000$) systems of atoms, varying in size depending on computational capability.

Among the drawbacks of MM are the need for parameterization, or experimental data, to complete the model and the lack of accuracy which may arise due to the merely implicit consideration of quantum-mechanical phenomena. However, the ability to examine many-atom systems along with its capacity for speedy convergence makes MM an indispensable scientific tool.

MM methods rely on the definition of the potential energy as a function of configuration space for the molecule under consideration, based on the many interactions

which may occur between various atom types. A typical survey of these interactions may be given by the representation of the potential energy by the following equation (99):

$$V = V_{str} + V_{bend} + V_{oop} + V_{tors} + V_{vdW} + V_{es} + V_{cross} \quad (2.1)$$

in which stretching, bending, out-of-plane bending, torsional, cross term, van der Waals, and electrostatic influences on the energy of the molecule are taken into account, respectively, and linearly combined to give a total picture of the potential energy of the molecule. There are a number of mathematical representations for each interaction displayed in (2.1), and there are various parameters available to use with each. A set of expressions relating to those above, as well as a set of parameters for use with those expressions defines the “force field” for molecular mechanics. Many force fields have been created and studied, and some are mentioned in the next section, with focus on the Merck Molecular Force Field (MMFF94) which was used in this study. An explanation of each term in the above equation is necessary before specific force fields may be examined.

V_{str} , the first terms of the equation, represents the potential energy of stretching a bond, like a spring, from its minimum energy equilibrium length. In the simplest force fields, the potential energy of stretching may be approximated by the harmonic oscillator potential.

$$V_{str,ij} = \frac{1}{2}k_{IJ}(r_{ij} - r_{IJ})^2 \quad (2.2)$$

Here, r_{ij} represents the distance between atoms i and j , k_{IJ} represents the force constant between the atoms, and r_{IJ} represents a reference bond length between the atoms, the latter two of which must be determined via experimental methods. (2.2) may be written for each pair of atoms in the system under consideration, and the total potential energy of

stretching, as shown in (2.1) may be obtained via a summation over all stretching interactions between atoms bonded directly to one another.

$$V_{str} = \sum_{bonds} V_{str,ij} \quad (2.3)$$

The second term of (2.1) represents the potential energy associated with bending the angle between three atoms. Similar to the potential energy of stretching, the simplest representation of this interaction is a quadratic.

$$V_{bend,ijk} = \frac{1}{2} k_{IJK} (\theta_{ijk} - \theta_{IJK})^2 \quad (2.4)$$

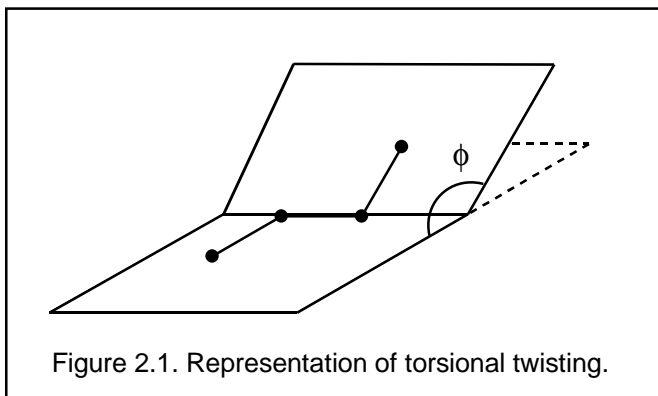
θ_{ijk} , k_{IJK} , and θ_{IJK} each hold analogous definitions as in (2.2), except they represent an interaction between three immediately adjacent bonded atoms and are referenced and measured by the angle formed by these atoms and not by any distances. The value for V_{bend} in (2.1) is obtained via a similar summation as (2.3), this time over all three-atom bonds.

V_{oop} characterizes out-of-plane bending, which may alter the potential energy in systems of four or more atoms. An sp^2 -hybridized carbon atom, for instance, may impose the influence that the three atoms bonded to it lay in the same plane. Angle bending potential energy interactions, however, may not recognize that conformation as relaxed, and when at a potential minimum, the molecule may reach some medium between being planar and at the angle desired by bending interactions. Out-of-plane bending, sometimes called improper torsion, simply gives quantitative value to deviation from planarity, and may also, in simplest form, be written as a quadratic

$$V_{oop,ijkl} = \frac{1}{2} k_{IJKL} \chi_{ijkl}^2 \quad (2.5)$$

where χ_{ijkl} represents the angle by which an atom deviates from planar. Again, the value for V_{oop} in (2.1) is obtained by a summation as in (2.3), this time over all out-of-plane bonded atoms.

Potential energy changes due to torsional bending of four-atom groups bonded in a chain-like configuration are represented by the term V_{tors} in (2.1). For instance, ethene is made up of two sp^2 carbon atoms bonded together and prefers to adopt a planar



geometry. Any twisting of the central double bond results in an increase in energy of the molecule. Similarly, in many circumstances sp^3 atoms prefer a staggered conformation (as opposed to

eclipsed). The expression for torsional potential energy is usually given in the form of a small summation over cosines (100), taking the form

$$V_{tors,ijkl} = K_{IJKL} \sum_{n=0}^m C_n \cos n\phi \quad (2.6)$$

where K_{IJKL} and C_n are parameters depending on angle type and ϕ is as indicated in Figure 2.1. Again, V_{tors} is obtain via summation of all $V_{tors,ijkl}$.

Variations in the potential energy of the system due to van der Waals interactions among atoms are accounted for by the V_{vdW} term of (2.1). Van der Waals forces, or London dispersion forces, exist between non-bonded atoms, and are typically regarded as influential among atoms separated by roughly three bond lengths or more. The Lennard-Jones 6-12 potential (101) is particularly useful in this case.

$$V_{vdW,ij} = \epsilon_{IJ} \left[\left(\frac{R_{IJ}^*}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{IJ}^*}{r_{ij}} \right)^6 \right] \quad (2.7)$$

Here, r_{ij} is the distance between atoms i and j , ϵ_{IJ} is called the well-depth parameter, and R_{IJ}^* is the distance between atoms i and j when the van der Waals potential energy is at a minimum. It should be noted that the first term of (2.7) represents a repulsive force which decreases much more quickly with distance than does the second term, an attractive force representation of potential energy. Here too, V_{vdW} is gotten via a summation over all atoms to which van der Waals forces apply.

The V_{es} term of (2.1) represents electrostatic interactions between atoms and, like van der Waals forces, they are usually taken into account for non-bonded atoms separated by three bond lengths or more. In its simplest incarnation, the electrostatic term appears as Coulombic.

$$V_{es,ij} = \frac{Q_i Q_j}{\epsilon_r r_{ij}} \quad (2.8)$$

Atoms are assigned partial charges in determining the values of Q_i and Q_j . ϵ_r is a dielectric constant, and r_{ij} is the distance between atoms i and j . A summation is applied to obtain V_{es} .

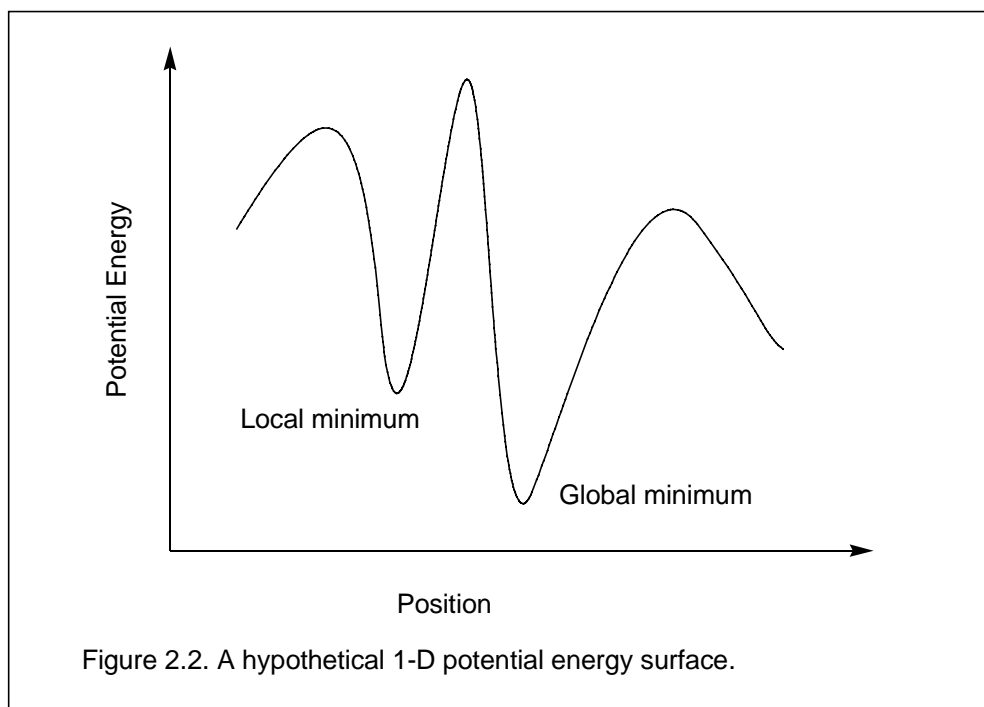
V_{cross} is a quantification of the interactions which may exist between different modes of conformation shifting mentioned above. For example, stretching bonds will change the van der Waals force between atoms, which in turn will change the force needed to bend an angle in question. Most common among cross terms are stretch-bend and stretch-stretch for two bonds to the same atom and stretch-torsion, bend-torsion, and

bend-bend for two angles with the same central atom. A typical stretch-bend cross term may appear as

$$V_{cross,ij} = \frac{1}{2}k_{IJ}(\Delta r_1 - \Delta r_2)\Delta\theta \quad (2.9)$$

where the Δr 's and $\Delta\theta$ are deviations from bond lengths and angle, respectively.

Once sufficient form has been provided for the terms in (2.1) and accurate parameters provided for these expressions, the potential energy surface may be mapped completely, and conformations may be affixed an energy value. Exploring the surface is actually a matter of determining the magnitude of the forces on each atom, and thus the configurations which most reduces those forces. Figuring the magnitude of the forces is achieved by taking the spatial derivative of the potential energy surface. The molecule will exist in a mechanical equilibrium when a conformation resides at a flat point on the potential energy surface, i.e. in a “valley.” (Other configurations of interest rest at “hilltops” or saddle points, but these positions are unstable equilibria.) In most cases, there is more than one minimum on the surface. The point of lowest overall potential energy is termed the “global” minimum, while all other valleys are dubbed “local” minima. Figure 2.2 shows a typical potential energy surface.



Starting from a given molecular conformation, or point on the potential energy surface, various mathematical techniques may be employed to search the surface in pursuit of the most stable conformations at the minima, such as the Newton-Raphson method or that of successive substitution. An explanation of these is not attempted here, but a fine review is provided by Rappé and Casewit (102).

2.2.2. Force Fields and MMFF94

A number of molecular mechanics force fields have been created and parameterized, each for the purpose of studying of a particular subset of compounds. The following represents a list of the more commonly used ones. MM2 (103) was developed by Allinger and co-workers for use with small organic compounds, and it has been modified to create MM3 (104) and MM4 (105 - 109). Both are also parameterized for use with small organics, but MM3 is suited for calculations on proteins and polypeptides,

while MM4 is specialized for hydrocarbons. AMBER (assisted model building with energy refinement) (110-112), which was developed by Kollman and co-workers, CHARMM (chemistry at Harvard molecular modeling) (113-115), which is the force field of Karplus and co-workers, designed at Harvard University, and OPLS (optimized potential for liquid simulations) (116-118), which was developed by Jorgensen and co-workers, are all optimized for calculations involving polypeptides, nucleic acids, and proteins. CFF93 (119) and CFF95 (120) (consistent force field, versions 1993 and 1995) are force fields which were developed for use with small organics, nucleic acids, and proteins. The UFF (121) (universal force field) is appropriately named, due to its parameterization for compounds of all elements. SYBYL (122), which is also referred to as the Tripos force field, is designed for calculations on small organic molecules and proteins. The SYBYL force field was examined for use in the present study of AAF and its adducts, but ultimately was discounted in favor of the MMFF94 force field, due to its inability to predict correct geometries of AAF. While some of the present work covers the force field study, a more complete analysis was performed by co-workers in this laboratory (85).

The MMFF94 force field is parameterized toward performing calculations on small organic compounds, especially nucleic acids and proteins, and as such, is well-suited for the present study. Among the strengths of MMFF94 is the large amount of *ab initio* quantum mechanical data used to parameterize the force field. Over 2,800 molecular structures optimized with high level *ab initio* techniques were utilized in the creation of MMFF94. The force field, while primarily designed to be a “small molecule” and “protein” force field, may also be accurately used in a wide variety of chemical

systems and is well defined for over 20 chemical families. Furthermore, its unique form for the van der Waals interaction and a systematic correlation of van der Waals parameters with well-characterized small molecule interactions and noble gas atoms, solidify MMFF94 as an accurate force field. The following is an outline of the MMFF94 force field and how each of the potential terms expounded upon in the previous section are represented under it. A complete description of the force field has been written by Halgren (123-127), its primary developer, and parameters are currently available via the Internet (128).

MMFF94 represents bond stretching interactions via the following quartic function

$$V_{str,ij} = 143.9325 \frac{kb_{IJ}}{2} \Delta r_{ij}^2 \left(1 + cs \Delta r_{ij} + \frac{7}{12} cs^2 \Delta r_{ij}^2 \right) \quad (2.10)$$

where $\Delta r_{ij} = r_{ij} - r_{IJ}^0$, the difference between the distance separating atoms i and j and a reference bond length between the atoms. kb_{IJ} represents the force constant between the atoms, and $cs = -2 \text{ \AA}^{-1}$ is the “cubic-stretch” constant. The function is obtained via a fourth order expansion of a Morse function with “alpha” = cs .

Bending interactions are normally represented by MMFF94 as

$$V_{bend,ijk} = 0.043844 \frac{ka_{IJK}}{2} \Delta \vartheta_{ijk}^2 (1 + cb \Delta \vartheta_{ijk}) \quad (2.11)$$

where $\Delta \vartheta_{ijk} = \vartheta_{ijk} - \vartheta_{IJK}^0$, the difference between the angle separating atoms i and j and a reference bond angle between the atoms. ka_{IJK} represents the force constant between the atoms, and $cb = -0.4 \text{ rad}^{-1}$ is the “cubic-bend” constant. For angles occurring in delocalized single bonds or small rings, MMFF94 uses the same form found in UFF

$$V_{bend,ijk} = 143.9325 ka_{IJK} (1 + \cos \vartheta_{ijk}) \quad (2.12)$$

For out-of-plane bending interactions, MMFF94 uses the potential energy representation

$$V_{oop,ijkl} = 0.043844 \frac{koop_{IJKL}}{2} \chi_{ijkl}^2 \quad (2.13)$$

where χ_{ijk} is the Wilson angle between the $j-l$ bond and the $i-j-k$ plane and $koop_{IJK}$ represents the force constant between the atoms.

MMFF94 utilizes the following representation for torsional bending interactions

$$V_{tors,ijkl} = \frac{1}{2} [V_1(1 + \cos \phi) + V_2(1 - \cos 2\phi) + V_3(1 + \cos 3\phi)] \quad (2.14)$$

where V_1 , V_2 , and V_3 are constants which depend on the atom types, and ϕ is the $i-j-k-l$ torsion angle. In the above case, there exist bonds between atoms i and j , j and k , and k and l .

Van der Waals interactions are represented by MMFF94 in the following “Buffered-14-7” form

$$V_{vdW,ij} = \epsilon_{IJ} \left[\left(\frac{1.07R_{IJ}^*}{R_{ij} + 0.07R_{ij}^*} \right)^7 \left(\frac{1.12R_{IJ}^{*7}}{R_{ij}^7 + 0.12R_{IJ}^{*7}} - 2 \right) \right] \quad (2.15).$$

Here, the minimum-energy separation R_{IJ}^* is given by the expression

$$R_{IJ}^* = \frac{1}{2} (R_{II}^* + R_{JJ}^*) \left[1 + 0.2 \left(1 - e^{-12\gamma_{IJ}^2} \right) \right] \quad (2.16)$$

where

$$R_{II}^* = A_I \alpha_I^{0.25} \quad (2.17)$$

and

$$\gamma_{IJ} = \frac{R_{II}^* - R_{JJ}^*}{R_{II}^* + R_{JJ}^*} \quad (2.18).$$

α_I in (2.17) is the atomic polarizability of atom I . The well-depth is given by the Slater-Kirkwood expression

$$\epsilon_{IJ} = \frac{181.16G_I G_J \alpha_I \alpha_J}{R_{IJ}^*{}^6 \left[\left(\frac{\alpha_I}{N_I} \right)^{\frac{1}{2}} + \left(\frac{\alpha_J}{N_J} \right)^{\frac{1}{2}} \right]} \quad (2.19)$$

The buffered Coulombic form is utilized by MMFF94 to represent electrostatic interactions where

$$V_{es,ij} = 332.07 \frac{Q_i Q_j}{D(R_{ij} + \delta)^n} \quad (2.20).$$

R_{ij} is the internuclear separation between atoms i and j , D is the dielectric constant, and δ is the electrostatic buffering constant and assumes a value of 0.05\AA . Normally, $n = 1$, however, MMFF94 also supports an $n = 2$ distance dependence. MMFF9 includes van der Waals and electrostatic interactions when atoms i and j are separated by three or more bonds.

MMFF94 utilizes the following cross term form to relate bond stretch-bending interactions

$$V_{cross,ijk} = 2.51210(kba_{iJK} \Delta r_{ij} + kba_{kJI} \Delta r_{kj}) \Delta \vartheta_{ijk} \quad (2.21).$$

Here, kba_{iJK} and kba_{kJI} are force constants which couple the i - j and k - j stretches to the i - j - k bend, respectively, and Δr and $\Delta \vartheta$ are defined in the stretching and bending interactions. The form is omitted when (2.12) is employed.

2.3 Quantum Mechanics

Quantum mechanical approaches to calculating molecular properties are inherently different from mechanics methods in that they attempt to derive information from a species via its quantum properties; in most cases, its electronic wavefunction. Quantum mechanics, as the most successful theory of modern physics, is invaluable in its ability to predict the physical behavior of small species, from particle momentum to magnetic properties. As a result, quantum computational methods have been developed to determine the energies, structures, and thermodynamic properties of molecules, among other things. These methods are generally regarded to be more reliable in the determination of physical properties than molecular mechanics methods.

2.3.1. *Semi-Empirical Methods and Theory*

The semi-empirical quantum mechanical methods were developed in response to so-called *ab initio*, or “first principles”, quantum mechanical methods, which attempt to calculate atomic or molecular properties directly from the atomic or molecular wavefunction. Semi-empirical methods use the same method of attack, but use parameterized values in many calculations instead of performing those calculations directly. They may be described as “approximate Hartree-Fock methods.” Such parameterization increases reliance on experimental data, but practically speaking, means a reduction in computational time when the methods are implemented. Information in the following section was adapted from Levine (129).

The earliest semi-empirical methods were developed to treat only the π -electrons of conjugated molecules. Justification for the separation rests on the greater

polarizability of the π -electrons and thus their increased susceptibility to electronic interactions, such as those which happen during chemical reactions.

The most celebrated π -electron semi-empirical theory is the Hückel molecular orbital theory, whereby the π -electron Hamiltonian is cast in the form

$$\hat{H}_\pi = \sum_{i=1}^{n_\pi} \hat{H}^{eff}(i) \quad (2.22)$$

where $\hat{H}^{eff}(i)$, whose form is not actually given by the theory, incorporates π -electron repulsions. Since the Hamiltonian involves single particles only, we can separate variables to obtain

$$\hat{H}^{eff}(i)\phi_i = e_i\phi_i \quad (2.23)$$

where ϕ_i are the Hückel molecular orbitals and e_i is the energy associated with each orbital. Instead of solving (2.23) directly, the assumption is made that the Hückel molecular orbitals may be written as a linear combination of atomic orbitals. That is,

$$\phi_i = \sum_{r=1}^n c_{ri} f_r \quad (2.24)$$

where the sum here is over atoms, as opposed to π -electrons in (2.22), and f_r are the chosen atomic orbitals. In order to determine the coefficients of (2.24), we note that their optimum values obey the equation

$$\sum_{s=1}^n [(H_{rs}^{eff} - S_{rs} e_i) c_{si}] = 0 \quad (2.25)$$

where orbital energies are determined by the secular equation

$$\det(H_{rs}^{eff} - S_{rs} e_i) = 0 \quad (2.26)$$

The distinguishing features of π -electron Hückel theory arise in the assumptions made for the integrals appearing in (2.26). Considering carbon atoms, the integrals are

$$H_{rr}^{eff} = \int f_r^*(i) \hat{H}^{eff}(i) f_r(i) d\mathbf{v}_i \equiv \alpha \quad (2.27)$$

$$H_{rs}^{eff} = \int f_r^*(i) \hat{H}^{eff}(i) f_s(i) d\mathbf{v}_i \equiv \beta \quad \text{for bonded } C_r \text{ and } C_s \quad (2.28)$$

$$H_{rs}^{eff} = 0 \quad \text{for non-bonded } C_r \text{ and } C_s \quad (2.29)$$

$$S_{rs} = \int f_r^*(i) f_s(i) d\mathbf{v}_i = \delta_{rs} \quad (2.30)$$

$$f_r = C_r 2p\pi \quad (2.31)$$

where it has been assumed that H_{rr}^{eff} has the same value for every carbon atom in the molecule, H_{rs}^{eff} has the same value for all bonded carbon atoms and equals zero for non-bonded atoms, and S_{rs} is assumed to be equal to 1 for the same orbitals and zero for different orbitals. α is known as the Coulomb integral, β is called the bond integral, and S_{rs} is known as the overlap integral. The basis atomic orbitals are shown to be carbon 2p π orbitals by (2.31). Assuming a value of zero for H_{rs}^{eff} for non-bonded carbon atoms is a reasonable assumption, since any non-bonded carbons are reasonably well separated in space, however, the choice of zero for the overlap integral for different atoms turns out to be a poor assumption. Values for S_{rs} can be around 0.3 for adjacent carbon atoms whose atomic orbitals are approximated by Slater type orbitals, which are of the form

$$\frac{(2\zeta/a_0)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\zeta r/a_0} Y_l^m(\theta, \phi) \quad (2.32)$$

where ζ is known as the orbital exponent. Some semi-empirical methods have arisen out of an attempt to remedy any inaccuracies arising from the zero overlap assumption.

The NDDO method (neglect of diatomic differential overlap), for example, neglects differential overlap only between atomic orbitals centered on different atoms. That is, $\int f_r^*(i)f_s(i)dV_i = 0$ only when orbitals r and s are on different atoms. Parameterization of the NDDO method proved rather unsuccessful (130), and the method was eventually modified to give the MNDO (modified neglect of diatomic overlap) method (131-136). The AM1 (Austin Method 1, which was named for the University of Texas at Austin) (137) is an improved version of the MNDO method, which has been parameterized for H, B, Al, C, Si, Ge, Sn, N, P, O, S, F, Cl, Br, I, Zn, and Hg. AM1 differs from MNDO in allowing the valence orbital exponents on the same atom to differ and AM1 represents core repulsion between valence electrons and the nucleus and inner shell electrons with a modified set of basis functions. It should be noted that NDDO, MNDO, and AM1 are not merely π -electron semi-empirical methods, which are applicable to only planar conjugated molecules. These methods take into account all valence electron interaction and are applicable to all molecules.

A 1999 search showed the AM1 method to be the most widely used semi-empirical method, followed by PM3 (parametric method 3) (138-141), which is a reparameterization of AM1.

A comparison of MNDO, AM1, and PM3 for geometry optimization of AAF will be developed.

2.3.2. Density Functional Theory (DFT)

DFT differs from other quantum methods in that it attempts to calculate molecular properties not via the electronic wavefunction of the molecule under study, but rather via

its electron density. Utilizing the electron density reduces the number of independent variables to 3, one for each spatial dimension in the electron density, from $3N$ spatial and N spin coordinates when the wavefunction is considered, where N is the number of electrons in the system. The information for the following section is adapted from Levine (142).

The Hohenberg-Kohn Theorem, which was developed in 1964 by Pierre Hohenberg and Peter Kohn, proved that the ground state electron probability density $\rho_0(x,y,z)$ uniquely determines molecular properties such as the ground state energy and ground state wavefunction for molecules with a non-degenerate ground state (143). The ground state energy is termed a *functional* of the electron density, since it depends on the density which is a function itself. The functional dependence of the ground state energy on the ground state electron density is indicated with square brackets, and is written $E[\rho_0]$.

For an n -electron molecule, the electronic Hamiltonian in atomic units may be represented as follows:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 + \sum_{i=1}^n v(\vec{r}_i) + \sum_j \sum_{i>j} \frac{1}{r_{ij}} \quad (2.33)$$

with the “external potential” v given by

$$v(\vec{r}_i) = -\sum_a \frac{Z_a}{r_{ia}} \quad (2.34).$$

The three terms in (2.33) represent electron kinetic energy, Coulomb interaction with a fixed nucleus, and interelectron Coulomb repulsion, respectively. Hohenberg and Kohn proved that for a nondegenerate ground state, the ground state electron density uniquely determines the external potential up to a constant. The proof is as follows. Suppose two

different external potentials, v_a and v_b , give rise to the same ground state electron density, ρ_0 , and are not necessarily given by (2.34). Let $\psi_{a,0}$ and $\psi_{b,0}$ and $E_{a,0}$ and $E_{b,0}$ be the normalized ground state wavefunctions and energies associated with the respective Hamiltonians H_a and H_b , corresponding to v_a and v_b . The variational theorem states that any normalized trial function, ϕ , not equal to the ground state wavefunction will yield an energy higher than the ground state energy when evaluated in the variational integral. That is,

$$\langle \phi | \hat{H} | \phi \rangle > E_0 \quad (2.35).$$

Inserting $\psi_{b,0}$ into the variational integral with H_a , one obtains

$$E_{a,o} < \langle \psi_{b,0} | \hat{H}_a | \psi_{b,0} \rangle = \langle \psi_{b,0} | \hat{H}_b - \hat{H}_a - \hat{H}_b | \psi_{b,0} \rangle = \langle \psi_{b,0} | \hat{H}_b - \hat{H}_a | \psi_{b,0} \rangle + \langle \psi_{b,0} | \hat{H}_b | \psi_{b,0} \rangle \quad (2.36)$$

which yields

$$E_{a,o} < \langle \psi_{b,0} | \sum_{i=1}^n [v_a(\vec{r}_i) - v_b(\vec{r}_i)] | \psi_{b,0} \rangle + E_{b,0} \quad (2.37)$$

since the Hamiltonians differ only by the external potentials. This leads to

$$E_{a,o} < \int \rho_{b,0}(\vec{r}) [v_a(\vec{r}) - v_b(\vec{r})] d\vec{r} + E_{b,0} \quad (2.38).$$

Since our initial assumption was that the different external potentials give rise to the same ground state electron density, $\rho_{b,0} = \rho_{a,0} = \rho$, and (2.38) must be incorrect since the same result could have been obtained (with a minus sign in front of the integral) by simply exchanging systems a and b in the argument. Adding the two results yields

$$E_{a,o} + E_{b,0} < E_{b,0} + E_{a,0} \quad (2.39)$$

which is nonsense, so our initial assumption must be false and the ground state electron density must uniquely determine the external potential. From this, we can also deduce

that the ground state electron density uniquely determines the electronic Hamiltonian and thus the ground state energy, wavefunction, and other properties.

In order to determine the ground state energy, we take the average of (2.33) over the ground state of the system under consideration. This yields

$$E = \bar{T} + \bar{V}_{Ne} + \bar{V}_{ee} \quad (2.40)$$

where each term on the right respectively corresponds to the terms in (2.33), and overbars are used to denote average values. Note that each term is a functional of the ground state electron density, and the second term in (2.40) is known.

$$\bar{V}_{Ne} = \langle \psi_0 | \sum_{i=1}^n v(\vec{r}_i) | \psi_0 \rangle = \int \rho_0(\vec{r}) v(\vec{r}) d\vec{r} \quad (2.41)$$

However, the first and last terms on the right hand side of (2.40) are unknown. Finding a representation for the two terms is key toward implementing density functional theory. Furthermore, the theory in principle tells us only that the ground state electron density determines the ground state properties, but not how to find these properties once the electron density is known. This is the second challenge to be overcome.

The Kohn-Sham method (144) solves the second problem, but still leaves an unknown functional to be approximated. The method considers a noninteracting system of electrons which each experience the same external potential, $v_s(\vec{r}_i)$, different from the actual external potential of the system, and chosen such that the electron density of the noninteracting system, ρ_s is equal to that of the actual system, ρ_0 . The Hamiltonian for the noninteracting system is written

$$\hat{H}_s = \sum_{i=1}^n \left[-\frac{1}{2} \nabla_i^2 + v_s(\vec{r}_i) \right] \equiv \sum_{i=1}^n \hat{h}_i^{KS} \quad (2.42).$$

If desired, the noninteracting system may be related to the actual system by considering interelectron repulsion to be a perturbative effect and representing the Hamiltonian as

$$\hat{H}_\lambda = \hat{T} + \sum_{i=1}^n v_\lambda(\vec{r}_i) + \lambda \hat{V}_{ee} \quad (2.43).$$

Kohn and Sham chose a noninteracting system because it leads to the separable Hamiltonian (2.42) for each electron in the system, which enables separation of the ground state wavefunctions. This fact, along with the Pauli exclusion principle, show that the ground state wavefunction of the noninteracting system, $\psi_{s,0}$ is the antisymmetrized product, or Slater determinant, of lowest energy Kohn-Sham spin orbitals, u_i^{KS} , of the system, with the spatial part of each spin orbital, $\theta_i^{KS}(\vec{r}_i)$ an eigenfunction of the single electron Hamiltonian, \hat{h}_i^{KS} . That is

$$\psi_{s,0} = |u_1 u_2 \dots u_n| \quad (2.44)$$

where

$$u_i = \theta_i^{KS}(\vec{r}_i) \sigma_i \quad (2.45)$$

and

$$\hat{h}_i^{KS} \theta_i^{KS} = \varepsilon_i^{KS} \theta_i^{KS} \quad (2.46).$$

Here, σ_i are spin functions and ε_i are the energies of the Kohn-Sham spatial spin orbitals. Kohn and Sham then rewrote the energy expression (2.40) for convenience. They defined an average kinetic energy difference between the actual ground state and the noninteracting system,

$$\Delta \bar{T}[\rho] = \bar{T}[\rho] - \bar{T}_s[\rho] \quad (2.47).$$

Note that there is no need to distinguish between electron densities for the two system based on the original requirement. They also defined an average interelectron repulsion difference term as

$$\Delta V_{ee}[\rho] = V_{ee}[\rho] - \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (2.48)$$

where the second term on the right hand side is the classical expression for Coulomb repulsion among two charge densities. r_{12} is the distance between elements $d\vec{r}_1 d\vec{r}_2$ and the $\frac{1}{2}$ is included to prevent counting the same elements twice. The expression (2.40) for E then becomes

$$E_0 = E_v[\rho] = \int \rho(\vec{r})v(\vec{r})d\vec{r} + \bar{T}_s[\rho] + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{xc}[\rho] \quad (2.49)$$

where the exchange-correlation energy functional has been defined

$$E_{xc}[\rho] = \Delta \bar{T}[\rho] + \Delta \bar{V}_{ee}[\rho] \quad (2.50).$$

(2.49) simply serves to rewrite the unknown terms in (2.40) in the different fashion of (2.50). Most density functional theories are formulated with approximations to E_{xc} , and accurate calculations rely on good approximations to the functional.

In order to evaluate (2.49), we need to know the ground state electron density, which is identical to that of the noninteracting system by definition. It may be proven that the electron probability density of a system described by the wavefunction (2.45) may be written as

$$\rho = \sum_{i=1}^n |\theta_i^{KS}|^2 \quad (2.51).$$

We are able to find E_0 if we are able to find the Kohn-Sham spatial spin orbitals and a suitable form for E_{xc} .

There have been numerous approximations to the exchange-correlation energy functional E_{xc} , the simplest of which assumes a near constant electron density, one which is slowly varying in space. Then Hohenberg and Kohn showed that E_{xc} is given quite accurately by the integral over all space,

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) \epsilon_{xc}(\rho) d\vec{r} \quad (2.52)$$

where ϵ_{xc} is the exchange and correlation energy per electron in a homogenous electron gas. This form for E_{xc} is known as the local density approximation.

Various other forms for E_{xc} may be employed. SPARTAN 5.1.3 utilizes the Becke-Perdew gradient corrected approximation to E_{xc} . Gradient corrected methods seek to rectify any inaccuracies which arise in E_{xc} due to a uniform medium assumption by including changes low order changes in the electron density via its gradient. They assume the form

$$E_{xc}^{GGA}[\rho^\alpha, \rho^\beta] = \int f(\rho^\alpha(\vec{r}), \rho^\beta(\vec{r}), \nabla\rho^\alpha(\vec{r}), \nabla\rho^\beta(\vec{r})) d\vec{r} \quad (2.53)$$

and also often split the exchange and correlation parts of E_{xc}^{GGA} as

$$E_{xc}^{GGA} = E_x^{GGA} + E_c^{GGA} \quad (2.54)$$

where ρ^α and ρ^β are separate densities for α and β spin electrons, and f is a function of the spin densities and their gradients.

The Becke-Perdew gradient corrected approximation is derived from Perdew's 1986 approximation (145) to the correlation energy functional and Becke's 1988 approximation (146) to the exchange energy functional. Perdew's approximation is given as

$$E_c(\rho^\alpha, \rho^\beta) = \int d\vec{r} \rho \epsilon_c(\rho^\alpha, \rho^\beta) + \int d\vec{r} \frac{C(\rho) |\nabla\rho|^2}{de^\Phi \rho^{4/3}} \quad (2.55)$$

where

$$\Phi = 1.745 \tilde{f} \left[\frac{C(\infty)}{C(\rho)} \right] \frac{|\nabla \rho|}{\rho^{7/6}} \quad (2.56),$$

$$d = 2^{1/3} \left[\left(\frac{1+\zeta}{2} \right)^{5/3} + \left(\frac{1-\zeta}{2} \right)^{5/3} \right]^{1/2} \quad (2.57),$$

$$C(n) = 0.001667 + \frac{(0.002568 + \alpha r_s + \beta r_s^2)}{(1 + \gamma r_s + \delta r_s^2 + 10^4 \beta r_s^3)} \quad (2.58),$$

$$\rho = \rho^\alpha + \rho^\beta \quad (2.59),$$

$$\zeta = \frac{(\rho^\alpha - \rho^\beta)}{\rho} \quad (2.60),$$

$\tilde{f} = 0.11$, $n = 3/4\pi r_s^3$, $\alpha = 0.023266$, $\beta = 7.389 \cdot 10^{-6}$, $\gamma = 8.723$, $\delta = 0.472$, and ε is the correlation energy per particle taken from parameterization of other results. Becke approximated the exchange energy functional as

$$E_x = E_x^{LSDA} - b \sum_{\sigma=\alpha,\beta} \int d\vec{r} \frac{(\rho^\sigma)^{4/3} \chi_\sigma^2}{1 + 6b \chi_\sigma \sinh^{-1} \chi_\sigma} \quad (2.61)$$

where

$$\chi_\sigma = \frac{|\nabla \rho^\sigma|}{(\rho^\sigma)^{4/3}} \quad (2.62),$$

$$\sinh^{-1} x = \ln \left(x + (x^2 + 1)^{1/2} \right) \quad (2.63),$$

$$E_x^{LSDA} = -\frac{3}{4} \left(\frac{6}{\pi} \right)^{1/3} \int d\vec{r} \left[(\rho^\alpha)^{4/3} + (\rho^\beta)^{4/3} \right] \quad (2.64),$$

and $b = 0.0042$ is an empirical parameter.

Becke-Perdew density functional theory, as utilized by SPARTAN makes use of an expansion of the spatial part of the Kohn-Sham spin orbitals, θ_i^{KS} , in terms of a set of orthogonal basis functions, a common approach to expressing the orbitals. The basis sets, BP/DN (147), BP/DN* (148), and BP/DN** (149) each form a representation in which the valence shell of all atoms is split into inner and outer parts. BP/DN* includes atomic polarization functions to all atoms except hydrogen, and BP/DN** includes polarization functions on hydrogens.

It is noted that a weakness of density functional theory is its inability to directly provide an expression for the ground state electron density, leading to the need for a basis set representation. Another weakness of the theory lies in the fact that it is only useful for calculating ground state properties, as no one has yet been able to prove its ability to predict accurate excited state properties. Despite these, the theory has been largely successful in calculating molecular ground state properties.

2.4. Strategy and Progression of Study

2.4.1. Approach to Calculating Structures

Computational studies provide an invaluable scientific tool for supplementing experimental work, and may provide clues into the physical nature of the system under study in cases where experiment has been unsuccessful. Such is the case with the present study. Experimental structural identification of the major adduct has been performed, but there no experimental work has been possible on the minor adduct to date because of the low yields which result from isolation from tissue samples. The situation makes a computational study of the adducts of AAF quite attractive.

The present work seeks to utilize the experimental data on the major adduct as a means of verification that the computational method being employed is appropriate for the study of adducts of AAF. By calculating major adduct geometries computationally and comparing to experiment, convincing evidence may be developed that the methods being applied are suitable for the present system.

Calculations were performed using the molecular modeling program SPARTAN 5.1.3 on a dual-processor SGI Octane under IRIX 6.4.

The computational method invoked for calculating geometries in the present case is termed a “cascade method” because of its use of molecular mechanics and semi-empirical methods as precursors for more accurate DFT methods. The method has been described elsewhere by Hehre (97). The attractiveness of the method lies in its ability to make calculations less computationally taxing by relegating initial geometry calculations to less computationally intensive (and possibly more inaccurate) methods. The initial calculations, which may be initialized in a geometry far from that of equilibrium, are

performed by those methods requiring less computational effort, allowing equilibrium geometries to be “honed in on” in later stages, leaving the refining to the more accurate and computationally intensive theories.

The cascade chosen here employed the initial step of a systematic molecular mechanics conformer search with the MMFF94 force field. A molecular mechanics precursor is necessary at this stage since subjecting many conformers to DFT optimization is not presently feasible. Systematic conformer searching of the pertinent dihedral angles for AAF and its adducts mentioned in the Introduction involves dividing the 360° dihedral angle into a given number of smaller angles and calculating molecular energies for the configuration. For example, a 4-fold conformer search divides the dihedral angle into 90° segments. The conformer search parameters were obtained for each adduct by systematically increasing their values until convergence was obtained, as described by Topper et al. (85). Once the geometries are calculated, they are energy minimized with the same force field. Obviously, some equilibrated conformers may end up with quite similar geometries. SPARTAN 5.1.3 utilizes a threshold cutoff distance of 0.25Å between conformer atomic positions below which they may be considered identical. Conformer search parameters are displayed in Table 7.1. Conformers calculated from the MMFF94 conformer search were then submitted to a more intensive semi-empirical AM1 geometry optimization, where SPARTAN’s NOSYMTRY keyword was used to prevent any accidental bias. Conformers within 10 kcal/mol of the lowest energy conformer from the MMFF94 conformer search were submitted to AM1 optimization. The reason for choosing the AM1 as part of the cascade will be discussed. Following the semi-empirical optimization, the conformers were visually inspected to

determine which would be submitted to Becke-Perdew DFT optimization. A sub-cascade was utilized for DFT optimizations whereby the conformers were first submitted to optimization with the BP/DN basis set, followed by optimization with the BP/DN* basis, and, in the cases of G-C8-AAF and G-N²-AAF, the BP/DN** basis set. All DFT optimizations were performed using the NOSYMTRY keyword and the NOPGN keyword (except G-C8-AAF at the BP/DN level), which eliminates the prescreening of the numerical integrations for efficiency. In some cases prescreening can reduce the likelihood of differential buildup errors in the density matrix; however, no such problems were encountered in any of the systems considered here. The “restart using Hessian” option, which uses the Hessian from the previous calculation as an initial guess, was also utilized in all DFT calculations, with BP/DN geometry optimizations using the AM1 Hessian.

Starting configurations for the cascade were generated via construction of the molecules in SPARTAN 5.1.3 via the program’s molecular viewer. The constructed geometries were then energy minimized with the MMFF94 force field. In order to determine if the cascade method is an appropriate one for the study of adducts of AAF, a preliminary study was performed on G-C8-AAF in which the molecule was subjected to an MMFF94 conformer search followed by optimization at the semi-empirical AM1 level followed by optimization at the DFT BP/DN level. A second cascade was performed whereby the secondary semi-empirical optimization was omitted. These shall be referred to as cascades A and B, respectively. The experimental work of Neidle et al. (73) provided the test against which the chosen computational method was cast. The x-ray crystallography study shows a preference for perpendicular fluorene and guanine groups,

reporting values for dihedral angles as $\beta = 108^\circ$, $\gamma = 174^\circ$, and $\alpha = 28^\circ$. The work stresses the importance of the configuration for stability, as it seems to reduce steric interaction among the fluorene and guanine groups, both of which contain a certain degree of π character.

The MMFF94 conformer search returned 8 stable conformers within 10 kcal/mol of the lowest energy conformer. These were submitted to further optimization. Results for the cascade method A are in much better agreement with the experimental work of Neidle et al., displaying the perpendicular arrangement of fluorene and guanine ring structures already mentioned. At the BP/DN level of calculation, cascade A provides much better agreement with experiment than cascade B. For one conformer, cascade A calculates $\beta = 110^\circ$, a 2° difference from the 108° report of Neidle et al. Furthermore, two other conformers, 4 and 2 on Table 2.1, which contains results for cascade A, show values for β within 12° of 90° . No conformers calculated by cascade B have β values residing around 90° , and the best case scenario is a 19° difference in the case of conformer 3B, which shows $\beta = 127^\circ$. Dihedral angle values and conformer relative energies for cascade B at the BP/DN level are available in Table 2.2. In addition, conformers 4 and 2 show very close agreement with the value of $\gamma = 178^\circ$ predicted by Neidle et al. This value for γ indicates that the acetyl group should be oriented parallel to the fluorene ring to which it is attached. Cascade B lacks agreement with this prediction as well.

Cascade A Geometry Optimization Results				
Conformer	Relative Energy, kcal/mol	β	γ	α
1	0.0000	-63	167	-167
3	0.2366	110	169	-166
4	0.3062	-98	-178	174
2	0.6150	78	-176	166
8	2.2459	-37	-4	-76
5	2.7391	-141	5	66
7	2.9713	42	3	70
6	3.1476	139	-7	-60

Table 2.1. Conformer energies and dihedral angles for cascade A for DFT BP/DN.

Cascade B Geometry Optimization Results				
Conformer	Relative Energy, kcal/mol	β	γ	α
1B	0.0000	-52	169	-176
4B	0.2805	-127	-167	-179
2B	0.4744	52	-168	-180
3B	0.6357	127	168	-176
7B	4.8023	44	12	30
5B	5.7618	-137	12	30
6B	5.9118	137	-15	-26
8B	7.1530	-42	-13	-27

Table 2.2. Conformer energies and dihedral angles for cascade B for DFT BP/DN.

The value of having the AM1 method between the molecular mechanics and DFT stages can be seen upon viewing the effect on geometry induced by the AM1 method. As a precursor for DFT optimization, AM1 "pushes" dihedral angles toward values in closer agreement with higher level DFT methods. The method serves as a computationally affordable way to tune geometries before submitting the conformers to the more rigorous and time-demanding DFT methods. AM1 is known to perform well for geometry calculation on molecule of the class and size of AAF (85, 137). It is for these reasons that cascade A was chosen for implementation in the current study.

Further calculations for other conformers and at higher levels of DFT optimization, which are described in the Results section, confirm the ability of AM1 to tune geometries well. Admittedly, AM1 is also noted to be a poor predictor of relative energies (137), but it is believed that the value of the method as a geometry precursor outweighs the method's poor energy predictions.

2.4.2. AAF Alone and Choosing Force Fields

In utilizing the present cascade method, or performing any structural determination for that matter, among the most important considerations is the choice of force field or quantum methods to be used. The present study chooses the MMFF94 force field and the semi-empirical AM1 method based on a previous benchmark study performed in our laboratory on AAF alone (85). The study compared the molecular mechanics force fields MMFF94 and SYBYL as well as the semi-empirical methods AM1, PM3, and MNDO for their ability to predict AAF geometries in comparison to quantum mechanical Hartree-Fock (HF) and density functional theories. With different force fields or methods optimized to predict accurate structures for a given class of molecule, it is necessary to perform such benchmark studies as preliminary to further studies of a given molecule.

Conformer searches of the dihedral angles β , γ , and δ were performed for AAF using each of the mechanics and semi-empirical methods above in order and compared to DFT searches using the BP/DN and basis sets and to HF searches using the 3-21G* and 6-31G* basis sets. The molecule and dihedral angles are shown in Figures 1.6 and 1.7. DFT and HF calculations predict a planar structure overall between the fluorene and

acetyl groups, favoring $\beta \sim 180^\circ$ and $\gamma \sim 0^\circ$. This is in very good agreement with conformers calculated using the MMFF94 and AM1 methods, in contrast to those predicted with SYBYL, PM3, and MNDO. These methods favor significant non-planarity of the amide group preserving the nitrogen NA2 pyramidal structure. HF molecular orbital calculations, however, explain the planarity. Determination of the highest occupied molecular orbital in a simpler, but very similar compound, acetamidobenzene, reveal an aromatic character for the electronic orbital between NA2 and CA14, which serves to stabilize the planarity of the molecule. A typical difference in structural prediction between MMFF94 and AM1 and SYBYL, PM3, and MNDO is shown in Figure 2.3. The figure illustrates the difference between the lowest energy conformers returned in a 15x15x15 conformer search using either AM1 or PM3.

The benchmark study (85) also discusses coordinate driving studies which were performed to create a dihedral angle energy landscape. The β dihedral angle was rotated by 360° in with step sizes in the range of 20 to 40 increments. These calculations again show the MMFF94 and AM1 predictions to be in agreement with higher level HF and DFT results. Each method displays qualitative agreement, producing two energy minima at $\beta \sim 0^\circ$ and 180° . Landscapes produced by SYBYL and by PM3 and MNDO are in stark disagreement with MMFF94 and AM1 calculations, respectively, producing minima near angles which AM1 has predicted are energy maxima.

The ability of MMFF94 and AM1 to predict AAF structures more accurately than other methods led to the choosing of these methods for the present cascade calculations. The success of the methods for calculating geometries of AAF alone make them logical

candidates for calculations on the molecule modified by guanine or deoxyguanosine in either major or minor adduct form.

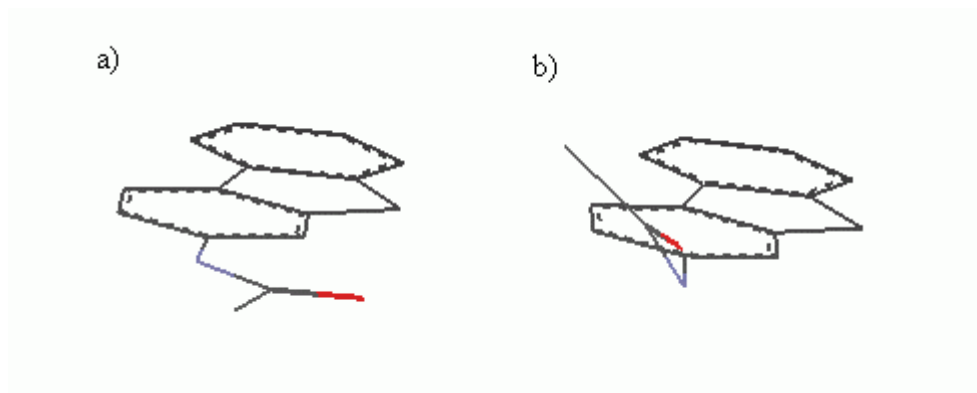


Figure 2.3. Lowest energy conformers from a 15x15x15 conformer search using a)AM1 or b)PM3.

3. Results

3.1. G-C8-AAF

Determination of the structure of G-C8-AAF was useful as a study to provide confidence in the cascade method utilized. Comparison of the present computationally calculated G-C8-AAF structures to the x-ray crystallographic and NMR experimentally determined structures of Neidle (73) displayed similarities which indicated that the cascade method was appropriate for the AAF-guanine and -deoxyguanosine structures under study. Results for all G-C8-AAF conformers are displayed in Table 7.2 and are available on the CD provided. Atomic coordinates for all calculations on G-C8-AAF and figures for the 8 conformers which were selected for DFT optimization are also available on the CD provided.

The 12-fold MMFF94 conformer search of α , β , and γ for G-C8-AAF, in which 1,728 systematically generated conformers were energy-minimized, returned only eight distinct minimum-energy conformers below the 10 kcal/mol energy cutoff, four of which had relative energies ranging from 0.00 - 0.33 kcal/mol, while the four others had relative energies in the range of 5.32 - 5.52 kcal/mol. A distinct symmetry can be seen in the results returned by the molecular mechanics search. β values within 5° of + or - 45° and 135° , γ values near + and - 165° and 11° , and α values near + and - 177° and 23° characterize the 8 returned conformers. Upon investigation of the γ and α values among the four lowest energy conformers, a correlation between positive and negative values is seen. Positive γ values correspond to negative α values in each of the four conformers. The consequence of this positive-negative matching is an intramolecular hydrogen bond which is formed between the acetyl oxygen OA14 and the hydrogen atom bonded to the nitrogen N7 of guanine. This hydrogen bonding appears to somewhat stabilize the four

lower energy conformers relative to the four higher energy ones. The hydrogen bonding is preserved through the BP/DN** level of calculation and hydrogen bond distances will be discussed at that level, along with the significance of the distinct values of dihedral angles mentioned. Figures of both types of conformer are represented by conformers 2 and 6 at the BP/DN** level in Figure 3.1.

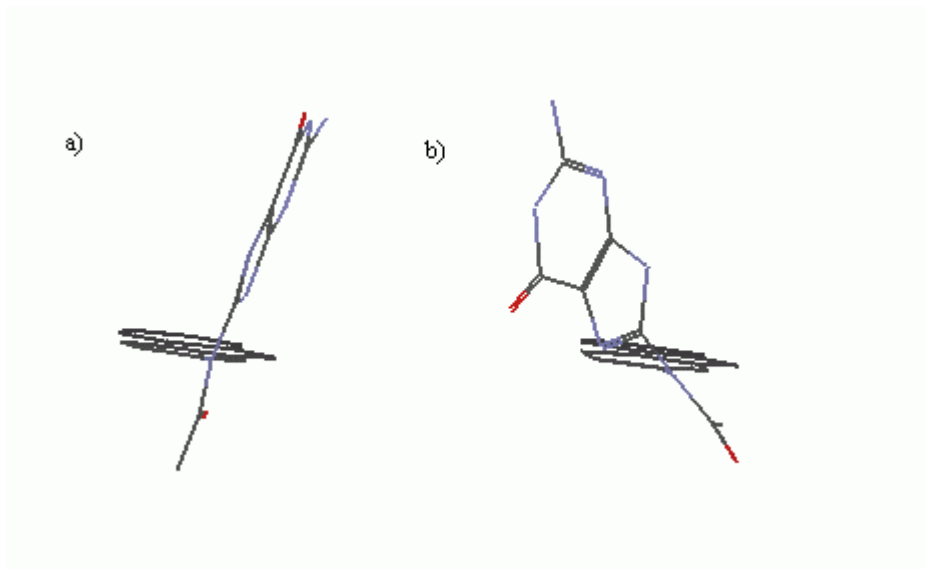


Figure 3.1. Conformers 2 and 6 of G-C8-AAF at the BP/DN** level.

Submission of the eight conformers to AM1 geometry optimization switched the energy ordering of the lowest four conformers, as well as the ordering of the four highest energy conformers, but preserved the dual grouping seen in the MMFF94 conformer search. The lowest energy conformers now displayed a relative energy spread of 0.00 - 0.15 kcal/mol, while the four higher energy conformers showed a spread of 1.31 - 1.35 kcal/mol. As mentioned, the AM1 method is not recognized as a very good predictor of absolute energies, however the method preserves the dual grouping from the MMFF94 conformer search. Furthermore, AM1 optimization altered β values in the four lowest

energy conformers by $\sim 21^\circ$ from 47° to 68° in the two lowest energy conformers and from $\sim 132^\circ$ to 111° in the next two lowest. The method also altered α values by $\sim 46^\circ$ from $\sim 24^\circ$ to $\sim 70^\circ$ in the four highest energy conformers. These changes more closely represent results for the DFT level of calculation, and underlie the importance of the AM1 method in the cascade.

Upon submission to DFT calculations, the dual grouping was still apparent for data at the BP/DN level, and is present, but less obvious among BP/DN* and BP/DN** results, wherein the relative energy varies somewhat smoothly from 0.00 - 4.98 kcal/mol. BP/DN level calculations preserve a distinct dual grouping with energy values ranging from 0.00 - 0.61 kcal/mol for the four low energy conformers and from 2.25 - 3.15 kcal/mol for the four highest energy conformers. Ordering is exchanged among the groups when compared to AM1 results. BP/DN level calculations differ from the AM1 results with an average absolute change in β of $\sim 8^\circ$ and a maximum change of 13° for conformer 4. Values for γ do not change significantly, except for changes of 8° and 6° for conformers 4 and 2, respectively. Perhaps surprisingly, the average absolute value of α for the four lowest energy conformers is $\sim 168^\circ$, which is closer to the average absolute value of $\sim 178^\circ$ given by MMFF94 results, compared to the average absolute value of 147° reported by the AM1 method.

BP/DN* calculations display a more uniform energy spread than do BP/DN level calculations, with a relative energy difference of 3.84 kcal/mol between lowest and highest energy conformers. The four lowest energy conformers saw average absolute value changes of $\sim 5^\circ$ in β , 4° in γ , and $\sim 5^\circ$ in α , with maximum changes of 10° , 9° , and

7°, for conformers 3, 3, and 1 and 2, respectively. The changes indicate the importance of including polarization functions in the basis set used for DFT calculations.

The only significant change from the BP/DN* to BP/DN** level was the change in energy ordering among conformers 1 and 2. BP/DN** results show that conformer 1 is the lowest in energy and conformer 2 is the second lowest with a 0.96 kcal/mol difference between the two. The overall energy difference between highest and lowest energy conformers at the BP/DN** level is 4.97 kcal/mol. Upon examination of the BP/DN** results, it can be seen that the four lowest energy conformers place the guanine and fluorene ring structures approximately perpendicular to one another, with two conformers displaying values for β near 90° while the other two show β residing near -90°, as well as values for α around either 170° or -170°. The benefit of orthogonal ring structures arises as repulsion of the acetyl group is minimized. The four highest energy conformers are unique with respect to one another, but very similar, placing the two planar structures approximately 45° apart in each case. Furthermore, the four lowest energy conformers display an intramolecular hydrogen bond between the acetyl oxygen OA14 and the hydrogen atom bonded to the nitrogen N9 of guanine, as mentioned. The bond distances are 1.93 Å, 1.94 Å, 1.94 Å, and 1.94 Å, for conformers 1 - 4, respectively. This explains the planarity of the acetyl and guanine groups with respect to one another. It is also interesting to note that a similar electrostatic interaction serves to stabilize the four highest energy conformers, with the attraction being between the acetyl oxygen OA14 and the hydrogen atom bonded to the carbon CA1 in the case of conformers 6 and 5, or the hydrogen atom bonded to the carbon CA3 in case of conformers 7 and 8. Both

interaction are similar with distances between OA14 and either CA1 or CA3 being 2.52 Å, 2.55 Å, 2.54 Å, and 2.52 Å for conformers 6, 5, 7, and 8, respectively.

X-ray crystallography and NMR measurements for the structure of G-C8-AAF were reported by Neidle et al. (73). The group's x-ray data indicate $\beta = 108^\circ$, $\gamma = 178^\circ$, and $\alpha = -28^\circ$, which closely resembles the third lowest energy conformer at the BP/DN** level obtained in the current study. The value of α is noted to differ significantly ($\sim 142^\circ$). This may be due to crystal packing effects. For example, unmodified AAF experiences considerable strain in the crystal phase, with β residing around 44° as determined by Van Meersche et al. (150) and Haisa et al. (151), whereas higher level quantum calculations indicate planar structures ($\beta \sim 0^\circ, 180^\circ$) (85). Haisa et al. (151) attributed the loss of planarity to hydrogen bonding between the N-H of AAF and the O of its nearest neighbor in the crystal.

In order to support these findings, the G-C8-AAF adduct was built with SPARTAN's molecular builder and energy minimized under the MMFF94 force field. Then, each of the three dihedral angles was set to the corresponding value reported by Neidle et al. (73) for crystalline G-C8-AAF. This structure was then submitted directly to DFT geometry optimization in the same BP/DN to BP/DN* to BP/DN** cascade used for the guanine adducts. The resulting BP/DN** conformer displayed dihedral angle values of $\beta = 107^\circ$, $\gamma = 177^\circ$, and $\alpha = -173^\circ$, which resembles the conformer third highest in energy from our G-C8-AAF BP/DN** calculation, with each angle differing by 8° , 2° , and 2° , respectively. It should be noted that the value of α changed by 145° from its original value of -28° .

NMR measurements of G-C8-AAF in methanol from Neidle et al. (73) indicate that the fluorene ring system and guanine are approximately perpendicular when $\gamma \sim 180^\circ$. That is, when γ resides around 180° or -180° , $\beta \sim 90^\circ$ or -90° . This is similar to the four lowest energy gas-phase conformers from the present BP/DN** calculation. Furthermore, NMR data from Neidle et al. (73) indicate that the orthogonal characteristic is not present when $\gamma \sim 0^\circ$ due to π conjugation from the C8 nitrogen to the fluorene ring. This effect is also supported by the present BP/DN** calculations, with $\beta = 137^\circ, -137^\circ, 41^\circ,$ and -40° , and $\alpha = -60^\circ, 65^\circ, 67^\circ,$ and -76° in order of relative energy, respectively, when γ resides around 0° .

Such agreement with experimental study serves to support the belief that the present cascade method is a valid one for use in the study of AAF.

3.2. G-N²-AAF

The G-N²-AAF adduct complex was submitted to an 8x8x8x8 conformer search of dihedral angles α^* , β^* , γ^* , and ϵ^* . Energy minimization of all 4,096 initial structures returned 16 distinct stable conformers with relative energies less than the 10 kcal/mol cutoff. Results for the G-N²-AAF conformers which were selected for DFT optimization are displayed in Table 7.3. Results for all G-N²-AAF conformers through every stage of study, atomic coordinates for all calculations on G-N²-AAF, and figures for the 8 conformers which were selected for DFT optimization are available on the CD provided.

The 16 conformers were returned from MMFF94 optimization with a spread in relative energy of 9.49 kcal/mol, and the mechanics method predicted excellent values for β^* and γ^* when compared to higher level calculations. The consistency of values for the two dihedral angles from the MMFF94 stage of cascade through AM1 and DFT optimization is quite remarkable. Of the 16 conformers, none showed any change in either β^* or γ^* from MMFF94 to AM1 optimization. Furthermore, of the eight conformers selected for DFT optimization, none showed any change in the value of β^* or γ^* through the BP/DN** level of optimization. α^* and ϵ^* values change among the 16 conformers from MMFF94 to AM1 optimization, with average differences of $\sim 10^\circ$ and $\sim 27^\circ$ for each angle, respectively.

AM1 geometry optimization on the 16 conformers resulted in some change in energy ordering, with a spread of 3.93 kcal/mol. Visual inspection of the conformers led to the selection of the eight lowest energy conformers from this group for BP/DN, BP/DN*, and BP/DN** optimization. The three levels of optimization resulted in energy spreads of 4.72 kcal/mol, 3.72 kcal/mol, and 4.04 kcal/mol, respectively. Through each

stage of DFT optimization, α^* and ϵ^* dihedral angle values remain consistent with moderate changes from stage to stage.

Seven of the eight conformers studied at the BP/DN** level fall into one of two distinct conformer groups. Four of the eight structures display co-planarity of the N-acetyl group with the fluorenyl moiety to within 2° of β^* and γ^* . Three of the remaining four conformers, numbers 1, 7, and 6, place $\gamma^* \sim -178^\circ$ and $\beta^* = 9^\circ, 28^\circ,$ and 25° , respectively. This notable difference in β^* may be attributed to hydrogen bonding, as shall be discussed. The remaining conformer, number 8, is similar to one of the groups, but has some uniqueness in its values of γ^* , β^* , and ϵ^* .

Conformers 1, 7, and 6 appear to be structurally similar with the exception of a roughly 161° rotation about α^* in the case of conformer 1, compared to the other two. The average value for β^* in this group is $\sim 21^\circ$, which is quite different from the β^* values around 180° for all of the other conformers. The difference in β^* between this group and the other conformers may be attributed to intramolecular hydrogen bonding between the acetyl O and the H bonded to guanine N^2 , with distances of 1.75 Å, 1.83 Å and 1.82 Å, respectively, between the two atoms. The effect is shown in Figure 3.2 for conformer 7. The roughly 161° difference in α^* for conformer 1 corresponds to a flipping of guanine and does not appear to have a great effect on the energy of the conformer.

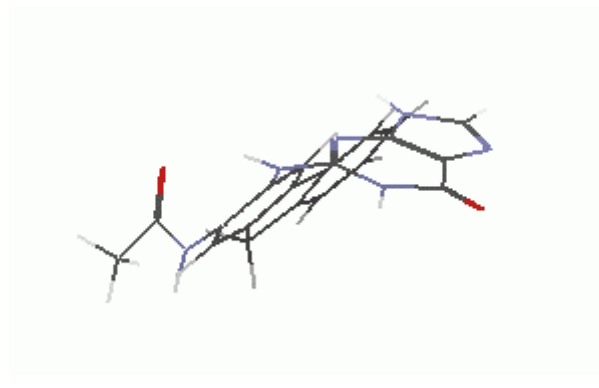


Figure 3.2. Conformer 7 of G-N²-AAF at the BP/DN** level.

Conformers 4, 2, 3, and 5 appear to have structural similarity, with some difference arising in the value of $\alpha^* = 99^\circ$ for conformer 4, compared to 120° and 127° for conformers 2 and 3. The difference does not significantly alter the general structural features of the conformer when compared to the other three. All four conformers in the group are classified by values for β^* and $\gamma^* \sim 180^\circ$, leading to planarity of the acetyl group and the fluorene moiety. The planarity is stabilized by intramolecular hydrogen bonding between the acetyl oxygen OA14 and the hydrogen attached to the fluorenyl carbon CA1, as well as by a favorable electrostatic attraction between the hydrogen attached to AAF nitrogen NA2 and the guanine nitrogen N². For conformers 4, 2, 3, and 5 the bond distances between OA14 and the CA1 hydrogen are 2.19 Å, 2.19 Å, 2.21 Å, and 2.19 Å, respectively, and the distances between the NA2 hydrogen and N² are 2.24 Å, 2.22 Å, 2.22 Å, and 2.22 Å, respectively. The effect is shown in Figure 3.3 for conformer 4. Conformer 5 differs from the other three in its values of $\alpha^* = 19^\circ$ and $\epsilon^* = -116^\circ$. Conformers 4, 2, and 3 each have opposite the sign for both angles. The difference places guanine on the other side of the acetyl-fluorene plane in a reflective symmetric configuration.

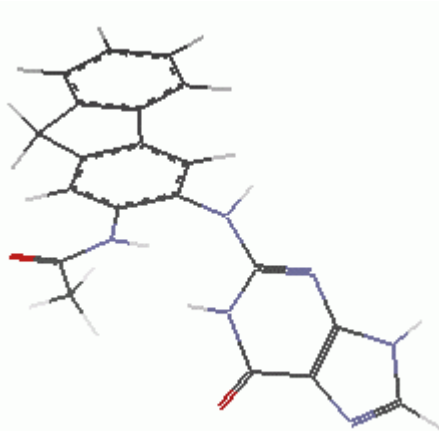


Figure 3.3. Conformer 4 of G-N²-AAF at the BP/DN** level.

Conformer 8 is very similar to conformers 4, 2, and 3, but displays a slightly higher value for $\epsilon^* = 140^\circ$, and a slightly lower value of $\beta^* = 172^\circ$. Most notably, however, the conformer differs from every other one with a γ^* value of -6° , nowhere near the 180° value favored by the rest. The γ^* value destroys the hydrogen bonding interaction between OA14 and the CA1 hydrogen, which existed for conformers 4, 2, 3, and 5. The ϵ^* value of 140° tilts the guanine group a bit more so than in the case of those conformers, but allows the preservation of some electrostatic interaction between the NA2 hydrogen and N², with a distance of 2.42 Å between the two atoms.

Although a grouping has been chosen to highlight the differences in β^* among the conformer groups, it should be noted that there is a maximum difference of 21° in the absolute value of α^* (except in the case of conformer 1 as mentioned) and 25° in ϵ^* (excluding conformer 4) between both groups, indicating that the position of guanine relative to the fluorene ring structure is similar in each case. This position is characterized by $\alpha^* \sim -25^\circ$ (except for conformer 1) and $\epsilon^* \sim 135^\circ$. The two angles

combine to produce a geometry characterized by the guanine and fluorene ring structure being approximately perpendicular to one another, with the guanine structure minimizing repulsion from the acetyl group. The result reinforces the observation that energy stabilization arises from perpendicular guanine and fluorene ring structures, an effect reflected in the G-C8-AAF low energy conformers. Pyramidization of the nitrogen N² is also barely evident in all conformers, perhaps a result of the two electronic interactions which the nitrogen is involved in from conformer to conformer. An average CA3-N²-H bond angle of ~116° characterizes the weak pyramidization.

3.3. dG-C8-AAF

The 12x12x12x12 conformer search of dihedral angles α , β , γ , and χ for dG-C8-AAF generated 20,736 initial conformers. After molecular mechanics energy minimization, 67 stable conformers were obtained with a relative energy less than the 10 kcal/mol cutoff. After submission to AM1 geometry optimization, visual inspection led the nine lowest energy conformers to be selected for DFT optimization. Those nine have an energy ranking among the 67 MMFF94 conformers as follows: 1, 9, 18, 21, 39, 41, 44, 50, 53. Note that the conformers are labeled 1-9, respectively, in Table 7.4, such that relative energy ranking may be more easily tracked through the cascade. The ability of AM1 to rearrange geometries toward much better agreement with DFT calculations than MMFF94 reinforces the value of having such an intermediate method in the cascade from molecular mechanics to DFT calculations. While MMFF94 seems to possess better ability to predict relative energy rank, the AM1 semi-empirical method is an invaluable tool for geometries. Results for the dG-C8-AAF conformers which were selected for DFT optimization are displayed in Table 7.4. Results for all dG-C8-AAF conformers through every stage of study, atomic coordinates for all calculations on dG-C8-AAF, and figures for the 9 conformers which were selected for DFT optimization are available on the CD provided.

At the MMFF94 geometry optimization level, the 9 conformers have a relative energy spread of 6.15 kcal/mol and differ somewhat from the 8 which were returned for G-C8-AAF, in average absolute values for β differing by $\sim 13^\circ$ for the 5 conformers of dG-C8-AAF with large β values when compared to 4 comparable conformers of G-C8-AAF. An average absolute difference of 12° is also noted between dG-C8-AAF and G-

C8-AAF among the conformers with low absolute values for α . Similar differences in absolute values between molecules are $\sim 8^\circ$ for low γ values and $\sim 27^\circ$ for low α values. Differences are effectively resolved at the DFT level, when values for β and γ for dG-C8-AAF agree with those for G-C8-AAF, but α values do not agree. The difference will be discussed.

It may be interesting to note that all of the nine conformers eventually submitted to DFT optimization were returned by MMFF94 optimization with a C2'-*endo* ribose ring pucker, as seen in B-DNA, rather than the C3'-*endo* pucker, which is characteristic of A-DNA (152). BP/DN* calculations, which favor quite different values for the glycosidic torsion angle, alter the pucker geometry, as will be discussed.

Submission of the 67 conformers to AM1 geometry optimization yielded conformers 1 - 9 as having the lowest energies of the 67 list, with a spread of 1.37 kcal/mol. Conformer geometries are very much altered from their MMFF94 configurations, and are seen to be in much better agreement with higher level DFT calculations. Inspection of conformer relative energies shows a somewhat drastic reordering from the MMFF94 to the AM1 level, signaled by conformer 1 going from lowest to highest relative energy. DFT calculations place conformer 1 back in the lowest energy position, agreeing with the molecular mechanics assessment. This again demonstrates the poor ability of AM1 as an energy predictor. However, the agreement of dihedral angles between AM1 and DFT methods again demonstrates its value as a precursor to DFT geometry optimization.

DFT optimization at the BP/DN level yields a 4.48 kcal/mol relative energy spread among conformers. Geometry characteristics present at the BP/DN level are

preserved in general through the BP/DN* level with minor changes. Energy ordering somewhat varies between the different levels of calculation, with the notable change of conformer 7 becoming the second lowest energy structure at the BP/DN* level, as opposed to the second highest at the BP/DN level.

BP/DN* calculations show a relative energy difference of 3.16 kcal/mol between lowest and highest energy conformers, with conformer 1, the original lowest energy conformer at the MMFF94 level retaining its lowest energy status throughout DFT calculations. At this level, eight of the nine conformers, show the glycosidic bond, χ , to be within 14° of the borderline between *anti* ($-90^\circ < \chi < 90^\circ$) and *syn* ($-180^\circ < \chi < -90^\circ$ and $90^\circ < \chi < 180^\circ$) configurations. Three display χ values of -85° , -82° , and -78° in the *anti* domain, while four conformers show glycosidic bond values just outside the *anti* domain, two at -94° and two at -104° and -96° . The only conformer which shows a distinct χ *syn* configuration is the lowest energy conformer, which has a value of $\chi = -141^\circ$. Furthermore, every conformer displays a value for γ within 12° of planarity.

When compared to the BP/DN** results for G-C8-AAF, the dG-C8-AAF BP/DN* level results display a somewhat different energy ordering. The two conformers which display β values close to $\pm 90^\circ$ and γ values close to -180° correspond to the four perpendicular ring structures described for G-C8-AAF, which are lower in energy than the other "45°" conformers on the G-C8-AAF list. However, for dG-C8-AAF these conformers assume positions 2 and 9 on the relative energy scale, second-lowest and highest, respectively. The relative energy difference between the two is due to the fact that the ribose sugar breaks an effective symmetry of the molecule which existed in the G-C8-AAF form. Positive and negative values of the same β are no longer symmetric for

dG-C8-AAF as they were for G-C8-AAF. For the BP/DN* calculation on dG-C8-AAF, the lower energy conformer with $\beta = -84^\circ$ is able to assume a configuration with less steric hindrance between the sugar and the guanine and fluorene moieties. The higher energy conformer, with $\beta = 84^\circ$, on the other hand, assumes a configuration wherein the ribose sugar is in a somewhat more crowded configuration with respect to the guanine and fluorene groups.

Furthermore, the majority of β values are not near 90° for dG-C8-AAF, as was the case for the lowest energy conformers of G-C8-AAF. The presence of the ribose group is the reason. Every conformer of dG-C8-AAF places the three ringed groups, fluorene, guanine, and ribose, approximately perpendicular to one another (the definition not being rigorous for the puckered ribose ring) in an attempt to minimize steric repulsion among electrons. Conformers of dG-C8-AAF adopt combined values of β and α in such a way as to minimize steric repulsion by allowing each group to occupy a large volume and so as to have the groups oriented perpendicular to one another.

Of the remaining seven conformers, all are quite similar and characterized into one of two groups except for the lowest energy one, conformer 1. The conformer displays a unique ribose-guanine interaction not observed in any other conformer. A hydrogen bond between guanine nitrogen N3 and the hydrogen attached to ribose oxygen O5' is observed for the conformer, with bond distance of 1.93\AA . The bond also allows for another intramolecular attraction to exist, that between the O5' itself and one hydrogen attached to the guanine nitrogen N². The interatomic distance for this interaction is calculated to be 2.45\AA . It is this interaction which leads to the different value of χ compared to the other conformers. Other than the double interaction, the

molecule is characterized by the perpendicular orientation for each ring system described above, with little steric interaction among the ringed groups. The interaction is shown in Figure 3.4.

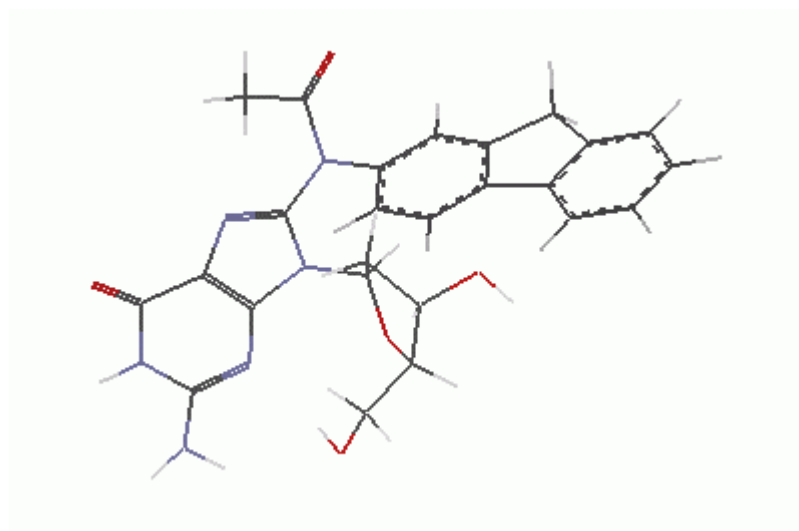


Figure 3.4. Conformer 1 of dG-C8-AAF at the BP/DN* level.

Conformers 9, 3, and 5 show dihedral angle values very close to one another with the largest difference among all angles being a 9° difference in χ values for conformers 9 and 5. All have $\beta = \sim 140^\circ$. The conformers very much resemble the four high energy conformers of G-C8-AAF, particularly conformer 6. The dG-C8-AAF conformers differ somewhat with α values averaging -80° , as opposed to -60° for conformer 6 of G-C8-AAF. Each of the dG-C8-AAF conformers adopts the modified geometry in response to the presence of the ribose group. Steric repulsion between the ribose and fluorene rings forces α to adopt the 20° difference.

Conformers 4, 2, and 6 are nearly identical to 9, 3, and 5, again corresponding to the four highest conformers of G-C8-AAF, with the closest partner being conformer 8 of G-C8-AAF. The conformers differ from 9, 3, and 5 in values for β averaging -35° . This

corresponds to a rotation of $\sim 180^\circ$ around the CA2-NA2 bond, and since the fluorene moiety is nearly symmetric (with exception of its middle ring) in 180° rotations about β , the conformers are nearly identical to 9, 3, and 5. Comparing the six conformers with the four G-C8-AAF, it is recognized that no comparable conformers of dG-C8-AAF exist with $\beta \sim -140^\circ$ or $\sim 35^\circ$. This is due to the fact that the ribose ring sterically interferes with the fluorene ring at these angles. With the ribose attached, the molecule no longer possesses the effective 4-fold rotational symmetry it had for G-C8-AAF. The two groups are represented in Figure 3.5 with conformers 9 and 4.

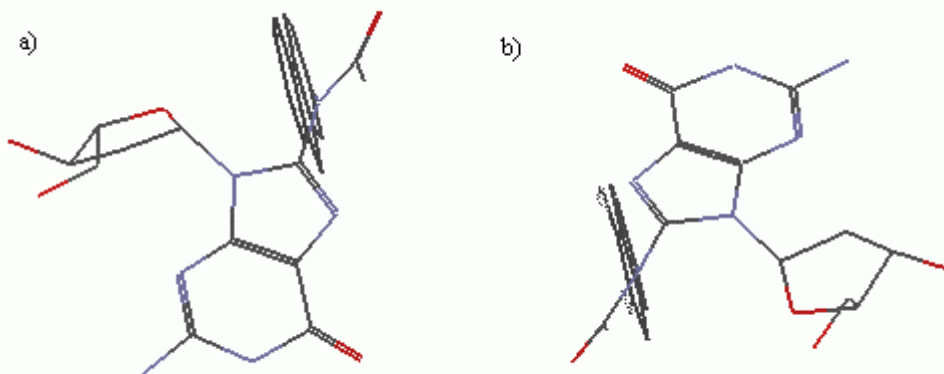


Figure 3.5. Conformers a)4 and b)9 of dG-C8-AAF at the BP/DN* level.

Inspection of the ribose ring pucker shows BP/DN* optimization reveals a preference for either a more planar or a C3'-*endo* ring configuration for all conformers but the lowest energy one. The glycosidic torsion angles, χ , for these conformers have values within 14° of -90° , which puts the guanine constituent somewhat above the ribose ring, almost directly above the C2' carbon, hindering the atom from adopting an *endo* configuration, and causing the ring to adopt either a near planar geometry or, depending

on exact χ value, one in which the C3' may adopt the out-of-plane pucker. Conformer 1, with $\chi = 141^\circ$, is the only conformer to adopt the C2'-*endo* configuration at the BP/DN* level of calculation. The effect may be seen for conformers 1 and 2 in Figure 3.6.

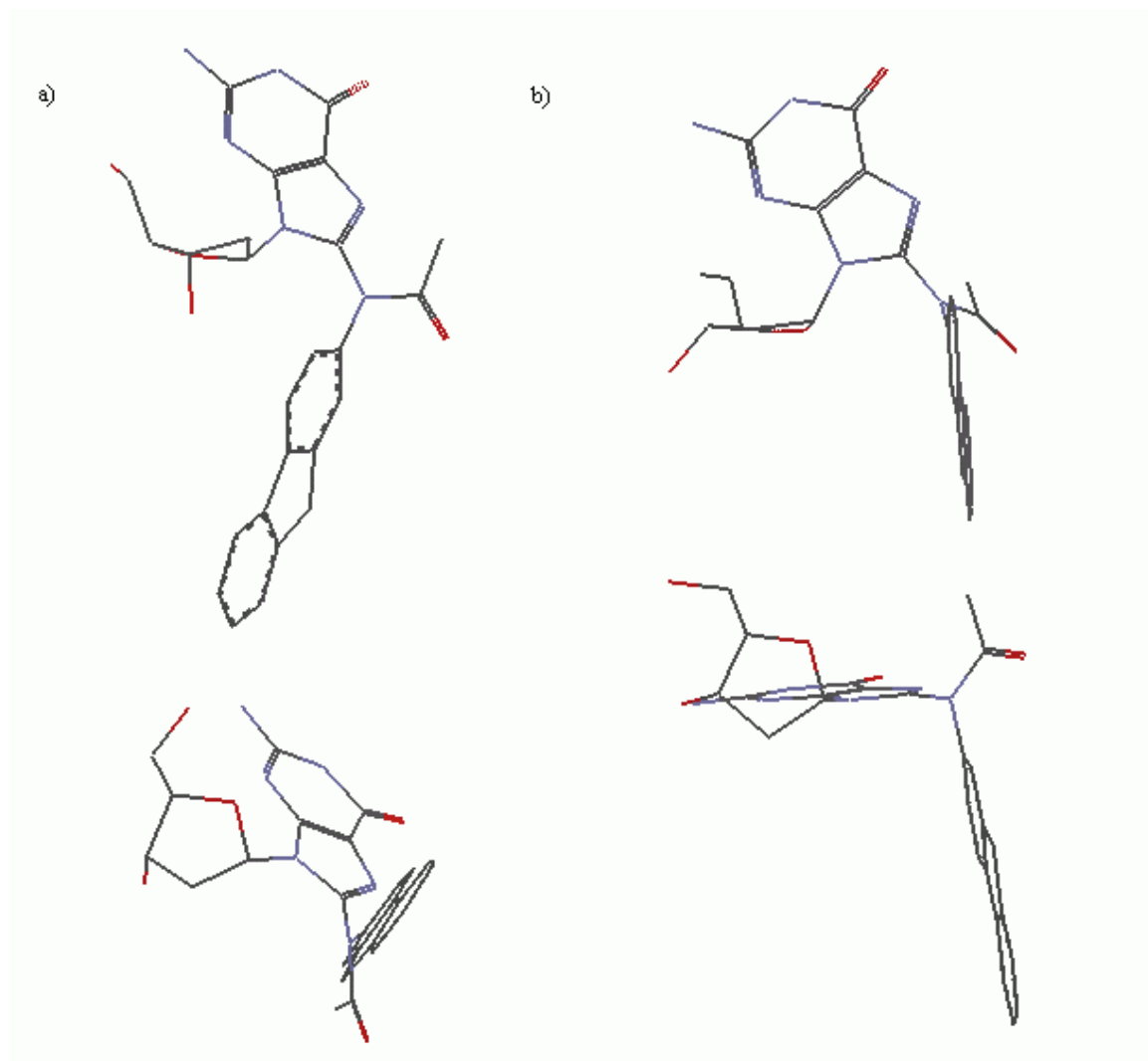


Figure 3.6. Conformers a)1 and b)2 of dG-C8-AAF at the BP/DN* level.

The BP/DN* results differ qualitatively from previous computational studies on dG-C8-AAF using the semi-empirical MNDO method (84) and molecular mechanics implementation using AMBER 3.0 (153), both of which studied the constrained α - χ

energy landscapes which result from keeping various dihedral angles fixed. The MNDO study, for instance held $\beta = -11.88^\circ$ and $\gamma = 180^\circ$, values quite different than those predicted by the present study for any single conformer. Previous work in this laboratory has shown that MNDO incorrectly predicts non-planar geometries for unmodified AAF (85).

3.4. dG-N²-AAF

The 8x8x8x8x8 MMFF94 conformer search of angles α^* , β^* , γ^* , ϵ^* , and χ^* for dG-N²-AAF generated 32,768 initial conformers. After subsequent molecular mechanics geometry optimization, 96 conformers were returned with relative energy less than the 10 kcal/mol cutoff. All were submitted to semi-empirical AM1 geometry optimization whereby a spread of 6.53 kcal/mol in relative energy was observed. Visual inspection again led to the ten lowest energy conformers from this group to be selected for geometry optimization at the BP/DN and BP/DN* level. The ten conformers showed an energy ranking at the MMFF94 level as follows: 22, 25, 53, 54, 55, 64, 65, 79, 84, and 97. Note that the conformers are labeled 1-10, respectively, in Table 7.5, such that relative energy ranking may be more easily tracked through the cascade. At the BP/DN and BP/DN* levels, the relative energy spreads among conformers changed to 8.14 kcal/mol and 6.95 kcal/mol respectively. Results for the dG-N²-AAF conformers which were selected for DFT optimization are displayed in Table 7.5. Results for all dG-N²-AAF conformers through every stage of study, atomic coordinates for all calculations on dG-N²-AAF, and figures for the 10 conformers which were selected for DFT optimization are available on the CD provided.

At the MMFF94 level of calculation, dG-N²-AAF conformers bear a fair resemblance to conformers returned at the same level for G-N²-AAF. The four dihedral angles shared among the molecules all show values which are relatively consistent at the molecular mechanics level. ϵ^* values do differ significantly, with lower values $\sim 90^\circ$ favored by dG-N²-AAF, as opposed to values $\sim 155^\circ$ for G-N²-AAF. Between MMFF94 and the BP/DN* level of calculation for dG-N²-AAF, the notable dihedral angle trends

are an overall decrease in the value of α^* and an increase in β^* for 3 conformers. Also, ϵ^* and χ^* values tend to cluster around given values more so at the higher level of calculation than at the MMFF94 level.

AM1 optimization again shows its effectiveness for geometry prediction by starting the trends mentioned above. The method is in much better agreement with the BP/DN* level than MMFF94 in terms of geometry prediction. Energy ordering is somewhat rearranged at the higher levels compared to the assignments given by AM1.

The BP/DN level of optimization shows very good geometry agreement with BP/DN*, with modest changes occurring at the BP/DN* level. Furthermore, energy ordering is relatively consistent between the two levels, with some changes made, the most extreme being the reordering of conformer 2 from lowest energy to fourth highest at the BP/DN* level.

At the BP/DN* level, trends observed among conformers allow their classification into groupings as will be discussed. All conformers but two have glycosidic dihedral angle χ^* values within 21° of the *syn-anti* borderline, two of those (1 and 9) having *anti* χ^* values of 83° and 84° , which is quite similar to the value of 85° found in B-DNA (84). The two other conformers have distinctly *anti* values of $\chi^* = 7^\circ$ and -2° . In the case of every conformer, γ^* resides around 180° , with a deviation of 6° being the most extreme.

Conformers 6, 8, and 2 display similarities which are typified by χ^* values very close to the $\sim 109^\circ$ average for all three conformers. *Syn* χ^* values are realized for these conformers due to an electrostatic attraction between hydrogen atoms in the acetyl group and the oxygen O5' in ribose. The attraction occurs between O5' and both the hydrogen

attached to NA2 and a methyl hydrogen. Distances between O5' and NA2 average 2.46 Å, and distances between O5' and the nearest methyl hydrogen average 2.69 Å for the three conformers. The attraction causes the glycosidic bond to be twisted further from the *anti* domain than is viewed for other conformers. The effect is displayed in Figure 3.7 for conformer 6.

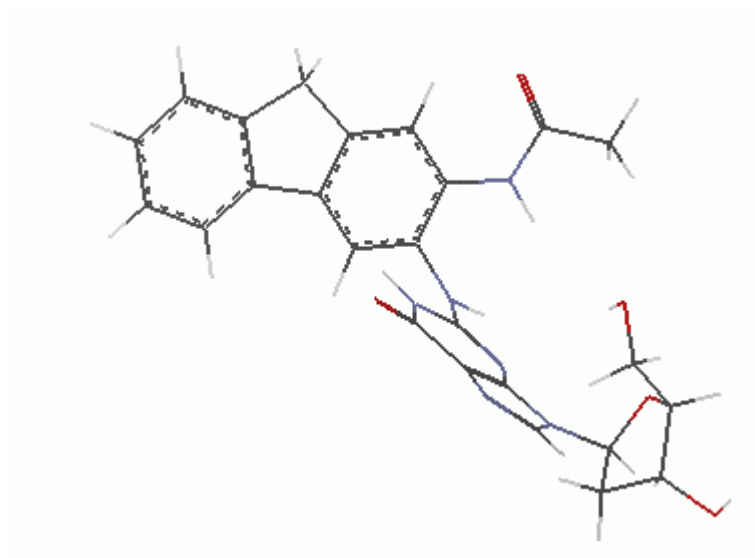


Figure 3.7. Conformer 6 of dG-N²-AAF at the BP/DN* level.

Conformers 1 and 9 appear to quite structurally similar, having nearly identical values of β^* , γ^* , and χ^* . α^* and ϵ^* are similar in number, although ϵ^* differs by almost 30° in the two cases, but opposite in sign. The opposing signs serve to flip the fluorene group about the plane of guanine. The flipping among conformers can be seen in Figure 3.8. The two conformers appear to contain the least intramolecular interaction among the ten under consideration. Their adopted geometry favors a spreading out of the acetyl, fluorene, guanine, and ribose groups, minimizing steric interaction.

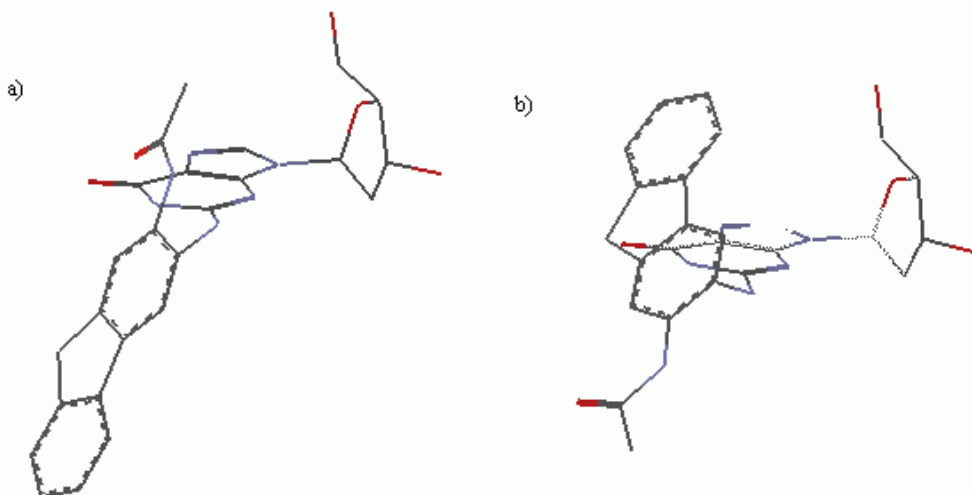


Figure 3.8. Conformers a)1 and b)9 of dG-N²-AAF at the BP/DN* level.

Conformers 5 and 4, the two highest energy conformers, appear to be quite similar to conformers 1 and 9, respectively, with corresponding values between every angle except χ^* . Conformers 5 and 4 display χ^* values of 7° and -2° , in contrast to values of 83° and 84° for conformers 1 and 9. The change by $\sim 80^\circ$ in χ^* appears to be stabilized for 5 and 4 by some degree of electrostatic interaction between the hydrogen attached to guanine carbon C8 and both oxygen atoms O1' and O5' of the ribose sugar. The average distance between atoms for both conformers is 2.61 Å for H-O1' and 2.89 Å for H-O5'. The effect is shown in Figure 3.9 for conformer 5.

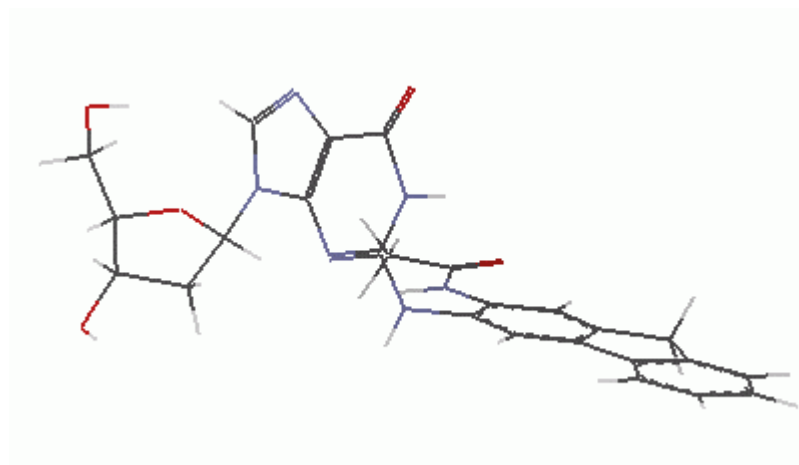


Figure 3.9. Conformer 5 of dG-N²-AAF at the BP/DN* level.

Each of the conformers listed above contains a common feature in the values of β^* and γ^* . The two angles are both close to values of 180° in every case, with β^* deviating by 20° in the most extreme cases. The deviation does little to disrupt the effect which is common amongst the conformers. The same intramolecular attractions which typified conformers 4, 2, 3, and 5 in the case of G-N²-AAF exist here. The observed effect is an attraction between acetyl oxygen OA14 and the hydrogen attached to CA1, as well as an attraction between the hydrogen bonded to amide nitrogen NA2 and guanine nitrogen N². Average values for all conformers listed above are 2.22 Å for the OA14-H interaction and 2.34 Å for the N²-H interaction. The ribose interaction which was described for conformers 6, 8, and 2 appears to slightly disrupt the N²-H interaction.

Conformers 7, 3, and 10 are typified by β^* absolute values close to 50° , markedly different from the other conformers. The reason for the difference is an intramolecular hydrogen bond which is formed between the acetyl oxygen OA14 and the hydrogen attached to guanine nitrogen N². Bond distances between the two atoms are 1.86 Å, 1.90 Å, and 1.92 Å for conformers 7, 3, and 10, respectively. The interaction is the same one

which exists for conformers 1, 7, and 6 in the case of G-N²-AAF. Interestingly, dihedral angle values for conformer 10 are very similar in value to the other two conformers but opposite in sign in every case. The opposite signs for β^* , γ^* , α^* , and ϵ^* may be thought of as a reflection through the fluorene ring of the molecule. The opposite sign for χ^* simply rotates the group by $\sim 180^\circ$ in this case, keeping the bulky oxygen arms of the sugar as far from steric interaction with the guanine ring as possible. The reflection is not totally symmetric, however, as conformer 10 lacks an interaction between a methyl hydrogen of the acetyl group and ribose oxygen O5', which is seen for conformers 7 and 3. Interatomic distances are 2.43 Å and 2.32 Å for these conformers, respectively. The group is represented by conformer 7 in Figure 3.10.

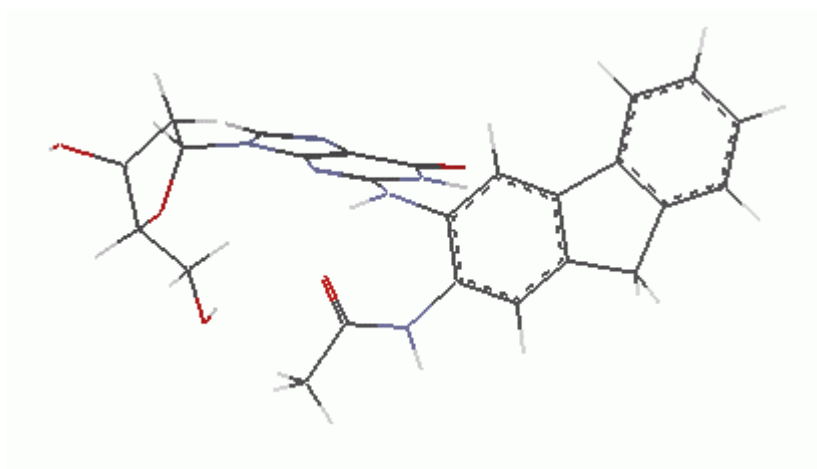


Figure 3.10. Conformer 7 of dG-N²-AAF at the BP/DN* level.

It is quite interesting to note that each conformer puts the three rings in a near perpendicular arrangement. This is, of course, an inexact characterization for the puckered ribose ring, but a general perpendicularity is apparent for every conformer. The perpendicularity is necessitated by steric repulsion from hydrogens on each group,

whereby the three ringed groups have adopted the geometry to effectively maximize the space which the hydrogen atoms may occupy. This perpendicularity is shown in Figure 3.11 for conformer 6.

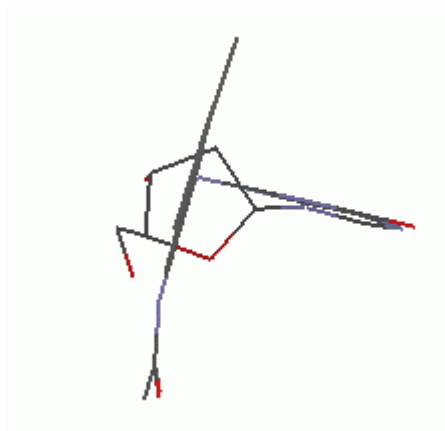


Figure 3.11. Conformer 6 of dG-N²-AAF at the BP/DN* level.

Ribose ring puckers at the BP/DN* optimization level show a preference for a nearly planar conformation with no distinct pucker, or a C3'-*exo* configuration, whereby the C3' carbon lies on the opposite side from C5' of a plane formed by the other ring atoms. This pucker is not common in any form of DNA, but attractive interactions here between the acetyl group and the ribose ring oxygen atoms, as described above, makes it possible.

4. Discussion

4.1. Confidence

The cascade method utilized was chosen because of its ability to save computational effort using lower level structural calculation methods to determine initial structures and use higher level, more computationally intense methods to simply refine these initial calculations. The method appears to be quite successful in the present case of the adducts of AAF. The method provides quite good agreement with experimental data in the case of the guanine bound major adduct form, G-C8-AAF. The method is able to predict the perpendicular nature of the fluorene and guanine ring systems and agrees reasonably well with the work of Neidle et al. (74) on G-C8-AAF. This is in contrast to the results predicted when a simpler MMFF94 conformer search was fed directly to DFT optimization (method B). This method was unable to predict the perpendicular character very well, instead favoring a more 45° arrangement of the groups (which was also seen by the cascade, but at higher energy than the 90° arrangement).

The benefit of using the cascade method also lay in the substantial reduction of computational effort at the DFT level when using the cascade. The number of calculations cycles performed at the DFT level until convergence on an equilibrium geometry was reduced for every conformer using the cascade method in comparison with method B. Method B utilized 33.4 cycles on average for the eight conformers, while the cascade method utilized 20.6 cycles on average for DFT BP/DN optimization. The computational effort at the DFT is therefore greatly reduced. Although the total computational time for the cascade method may be greater for small systems considering the extra step of AM1 optimization, the effectiveness of the cascade method to reduce

computational effort becomes more apparent as the size of the system increases, when time spent performing DFT optimization dominates the calculation more and more. For larger molecules, the value of the cascade method is fully realized.

Admittedly, a weakness of the method lies in the inability of the semi-empirical AM1 method to correctly predict energy ordering compared to DFT methods. As discussed before, the method is quite useful and important to the cascade method for its ability to predict geometries in agreement with higher level DFT methods. However, poor energy predictions made by the method may lead to the inadvertent omitting of some low energy geometries upon selection of AM1 low energy conformers for DFT optimization. The risk here is believed acceptable due to the ability of AM1 to hone in on correct geometries relative to DFT, as discussed before.

4.2. Significance of the Study and Carcinogenicity

The present study provides the most computationally intensive results to date for the major and minor adduct in guanine or deoxyguanosine form. Although perhaps not immediately useful in determining the reasons for the carcinogenicity of the adducts, due to interactions with the DNA helix itself when then carcinogen binds to guanine, the calculations provide accurate results useful for determining the more general characteristics of the molecule which lead to its ability to cause mutations. The value lies in determining some of the more general features of the molecules. Furthermore, the present calculations are of value to future studies on adducts of AAF. This study suggests that the greater conformational freedom of the minor adduct with respect to the major may play an important role in the ability of the minor adduct to avoid distorting the DNA helix upon becoming incorporated. Perhaps future studies will investigate this claim with respect to other molecules. Moreover, the present study is of value to future studies of the adducts of AAF in recognizing the preferred orientations of the molecule alone. While full relation to carcinogenicity must utilize the adducts bound to DNA, and the most preferred DNA-bound orientations will depend heavily on interactions with of the DNA chain itself, a knowledge of the preferred orientations of the adducts alone is of great value in determining strain on the adducts bound to DNA.

The present study also serves as a testament to the value of the cascade method for accurately producing high level quantum mechanical equilibrium geometries. The method has been shown to be quite effective in reproducing experimental structure, and is thus a valuable technique for investigating those structures which cannot be deduced experimentally. Its true value lies in its ability to reduce computational effort at the upper

levels of calculation while maintaining accuracy. Finally, our calculations may be useful to the development of improved force fields for accurate characterization of molecules of the type and class of AAF.

As mentioned, to get a complete picture as to any differences in carcinogenicity amongst the major and minor adducts, interaction between the adducts and the DNA to which they bind must be taken into account. Again, in this respect, the present study is not designed to confirm or deny the ability of the minor adduct to evade repair mechanisms, but perhaps it is interesting to compare the present study with a past computational study involving the minor adduct bound to DNA.

Grad et al. (90) performed a molecular mechanics and dynamics study of the minor adduct bound to three different sequences of DNA oligomer. The primary result of the work was showing that the minor adduct rests in the minor groove of DNA for the majority of conformers calculated with no major distortions of the double helix. The adduct is stabilized in the helix through interaction of the ribose groups on the DNA backbone with the fluorene ring system of AAF. The ring system seems to fit quite comfortably in the minor groove, with hydrogen bonding assisting in locking the molecule into place. Compared to the present study (which is admittedly completely in the gas phase), these conformations are somewhat unfeasible. Present results suggest values for α^* which indicate a spreading out of the fluorene, guanine, and ribose ring groups, favoring a decrease in steric interaction, whereas the results reported by Grad et al. show markedly different values for α^* , placing the fluorene ring structure in closer proximity to the ribose group. These conformations are feasible in DNA because of the more limited orientations available to the adduct in DNA than in the gas phase.

However, the present study seems to suggest that these conformers should be rather higher in energy than the structures observed in the molecular dynamics study.

5. Future Direction for Study

The present study is ripe for continuation in several areas. Investigation of larger molecular forms, that is, of the adducts bound to DNA dimers or oligomers, is of great interest. A computational study of the major and minor adducts bound to single- or double-chain DNA oligomers at higher levels than molecular mechanics or dynamics studies has not been performed to date. Furthermore, data concerning variation of the sequence of base pairs of DNA in the study would be of great interest. While performing semi-empirical or DFT geometry optimizations for such large systems would be quite computationally taxing, the results discovered would certainly be useful. Simulations involving a strand of DNA would need to be performed in proper solvent conditions to obtain accurate results and avoid any denaturing of the helix, with a price of increased computational effort. If high level calculations on an oligomer are deemed too computationally expensive, perhaps confining the 3' and 5' atoms of the ribose in a short chain would be useful so as to "simulate" the backbone of the DNA molecule without using a large number of atoms in the system.

Investigation of the applicability of the cascade method utilized in the present study might also be continued. The success of the cascade method in reducing computational effort and providing accurate structural data could be investigated for other aromatic molecules similar in size and class to AAF. A benchmark study of the method for other classes of molecules (such as polyaromatic hydrocarbons) would also prove quite interesting. Finally, it would be interesting to carry out an exhaustive molecular mechanics conformer search of dG-N²-AAF using exactly the same force field as that used by Grad et al. (90) in their study of the minor adduct.

6. Conclusion

Four varieties of modified AAF have been studied using SPARTAN 5.1.3 via a computational cascade method which utilized molecular mechanics conformer searching using the MMFF94 force field and subsequent geometry optimizations at the semi-empirical level with the AM1 method and at DFT levels using the Becke-Perdew functional. Results for G-C8-AAF at the BP/DN** level reveal the striking preference for adoption of one of two general orientations: that of guanine perpendicular to the fluorenyl ring moiety and that of guanine at approximately 45° to the fluorenyl structure. Confidence in the cascade method applied was established via comparison of these results with the experimental work of Neidle et al. (74). Results for dG-C8-AAF at the BP/DN* level display similarity to those for G-C8-AAF, with a difference in energy ordering when compared to G-C8-AAF calculations. Two of the nine conformers display the same perpendicularity of the guanine and fluorene rings, while six other seem to favor the 45° orientation. Steric interaction and repulsion between the acetyl group and ribose sugar play a key role in determining these structures. Results for G-N²-AAF at the BP/DN** level point toward the establishment of two general conformer classes, one of which is characterized by hydrogen bonding between the acetyl O and the H bonded to guanine N², and the other of which is characterized by a minimizing of repulsion between the guanine structure and the acetyl group. dG-N²-AAF results at the BP/DN* level indicate similarity to those for G-N²-AAF only to some degree with a variety of attractive electrostatic interactions appearing for various conformers among guanine and fluorene groups, and most notably between oxygen atoms on the ribose group and hydrogen atoms in the rest of the molecule. The differences which arise due to the addition of the ribose

group are obviously important should one wish to compare present gas-phase results with conformations adopted by modified AAF *in vivo*. Further higher level calculations on the species presently studied in a DNA oligomer would be quite useful.

The cascade method utilized in the present study proved quite useful in providing reasonable structural data, while reducing the computational effort at the DFT level through utilization of the MMFF94 force field and AM1 semi-empirical method as precursor methods.

7. Appendix - Conformer Energy and Dihedral Angle Tables

Molecule	Dihedral Angles	Conformer Search Parameters	Number of Conformers
G-C8-AAF	α, β, γ	12x12x12	1728
G-N ² -AAF	$\alpha^*, \beta^*, \gamma^*, \epsilon^*$	8x8x8x8	4096
dG-C8-AAF	$\alpha, \beta, \gamma, \chi$	12x12x12x12	20736
dG-N ² -AAF	$\alpha^*, \beta^*, \gamma^*, \epsilon^*, \chi^*$	8x8x8x8x8	32768

Table 7.1. Dihedral angles, conformer search parameters, and number of energy-minimized conformers used in MMFF94 conformer searches.

Conformer	Relative Energy, kcal/mol	β	γ	α
MMFF94 12x12x12 Conformer Search Results				
1	0.0000	-47	165	-178
2	0.1205	47	-165	177
3	0.2330	133	165	-178
4	0.3341	-132	-166	177
5	5.3216	-140	11	23
6	5.3406	141	-11	-24
7	5.4658	41	12	24
8	5.5200	-41	-12	-24
AM1 Geometry Optimization Results				
2	0.0000	68	-170	146
1	0.0321	-68	170	-147
4	0.1038	-111	-170	147
3	0.1430	111	170	-148
6	1.3221	148	-6	-70
8	1.3364	-31	-6	-70
7	1.3476	31	6	69
5	1.3512	-148	7	69
DFT DN Geometry Optimization Results				
1	0.0000	-63	167	-167
3	0.2366	110	169	-166
4	0.3062	-98	-178	174
2	0.6150	78	-176	166
8	2.2459	-37	-4	-76
5	2.7391	-141	5	66
7	2.9713	42	3	70
6	3.1476	139	-7	-60
DFT DN* Geometry Optimization Results				
2	0.0000	80	-177	173
1	0.3376	-66	172	-174
3	1.3416	100	178	-170
4	2.1775	-95	-179	174
6	2.3475	137	-7	-60
5	2.7937	-137	8	65
7	3.0321	40	6	68
8	3.8366	-39	-9	-76
DFT DN** Geometry Optimization Results				
1	0.0000	-70	177	-172
2	0.9626	79	-177	174
3	1.3630	99	178	-170
4	2.1762	-94	-179	174
6	2.5935	137	-6	-60
5	3.7688	-137	8	65
7	4.4541	41	5	67
8	4.9718	-40	-7	-76

Table 7.2. G-C8-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.

Conformer	Relative Energy, kcal/mol	β^*	γ^*	α^*	ϵ^*
MMFF94 8x8x8x8 Conformer Search Results					
1	0.0000	9	-179	176	161
2	1.3084	178	178	-45	152
3	1.3237	179	178	-45	152
4	2.2303	-178	-179	-33	88
5	2.5235	-180	-178	-54	-58
6	2.5771	25	-178	-43	163
7	3.2739	28	-177	37	78
8	6.3368	172	-6	-47	157
AM1 Geometry Optimization Results					
4	0.0000	-178	-179	-33	93
3	0.2907	179	178	-35	131
2	0.3009	178	178	-34	126
7	0.5783	28	-177	32	84
6	0.8278	25	-178	-38	147
5	1.4258	-180	-178	-42	-67
8	1.8143	172	-6	-39	144
1	1.8322	9	-179	-164	80
DFT DN Geometry Optimization Results					
1	0.0000	9	-179	177	127
5	1.1276	-180	-178	-22	-79
3	1.3284	179	178	-16	120
7	1.4157	28	-177	-12	137
2	1.4834	178	178	-22	121
4	1.7244	-178	-179	-20	88
6	2.4473	25	-178	-27	142
8	4.7201	172	-6	-25	134
DFT DN* Geometry Optimization Results					
2	0.0000	178	178	-25	121
4	0.3621	-178	-179	-27	95
1	0.6250	9	-179	178	126
5	0.9224	-180	-178	17	-114
7	0.9369	28	-177	-12	137
3	1.0831	179	178	-24	127
6	2.9355	25	-178	-28	144
8	4.3405	172	-6	-32	139
DFT DN** Geometry Optimization Results					
4	0.0000	-178	-179	-29	99
2	0.1964	178	178	-25	120
1	0.7373	9	-179	179	126
7	0.9758	28	-177	-12	137
3	1.1534	179	178	-24	127
5	1.2400	-180	-178	19	-116
6	2.7843	25	-178	-27	145
8	4.7760	172	-6	-33	140

Table 7.3. G-N²-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.

Conformer	Relative Energy, kcal/mol	β	γ	α	χ
MMFF94 12x12x12x12 Conformer Search Results					
1	0.0000	-155	21	53	-156
2	2.5555	-28	-19	-50	-34
3	3.2534	151	-19	-51	-113
4	3.3237	-28	-20	-50	-113
5	4.9329	153	-18	-50	-35
6	5.0594	-28	-19	-50	-35
7	5.1732	-44	156	159	-27
8	5.7997	136	156	159	-23
9	6.1526	152	-20	-50	-33
AM1 Geometry Optimization Results					
9	0.0000	145	-7	-85	-82
2	0.0951	-35	-4	-76	-88
8	0.1375	98	-171	114	-79
5	0.7115	146	-5	-78	-90
6	0.7130	-35	-4	-78	-90
7	1.2918	-84	-170	115	-83
4	1.2957	-35	-7	-86	-85
3	1.3047	146	-7	-87	-85
1	1.3675	-151	9	78	-158
DFT DN Geometry Optimization Results					
1	0.0000	-146	10	70	-150
3	1.7081	139	-3	-81	-84
6	1.7890	-35	-8	-91	-95
5	2.1467	145	-9	-82	-87
9	2.2289	144	-8	-86	-83
4	2.6600	-36	-5	-83	-96
2	3.0045	-38	-4	-80	-81
7	3.4231	-85	-174	114	-83
8	4.4792	95	-174	112	-80
DFT DN* Geometry Optimization Results					
1	0.0000	-145	12	70	-141
7	0.3263	-84	-173	116	-94
9	0.8020	141	-10	-80	-85
3	0.9155	139	-9	-77	-90
4	1.1138	-36	-9	-84	-104
5	1.9045	143	-12	-83	-94
2	2.0639	-35	-10	-78	-82
6	2.6657	-34	-11	-89	-96
8	3.1639	84	-174	112	-78

Table 7.4. dG-C8-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.

Conformer	Relative Energy, kcal/mol	β^*	γ^*	α^*	ε^*	χ^*
MMFF94 8x8x8x8x8 Conformer Search Results						
1	0.0000	-178	-179	44	-159	145
2	0.3856	-179	-179	-33	-96	132
3	2.2108	-25	-179	-23	-98	101
4	2.4638	-178	-179	-33	87	-18
5	2.5149	178	179	33	-88	-17
6	3.0604	179	-176	-55	-62	139
7	3.1569	-23	179	-20	-98	107
8	4.2587	-168	-173	-74	-139	152
9	4.5531	-178	-179	-34	88	68
10	5.8459	28	-177	-28	155	-179
AM1 Geometry Optimization Results						
3	0.0000	-62	-169	-26	-78	96
7	0.0163	-62	-169	-27	-80	96
2	0.3628	-156	-177	-42	-92	114
8	0.3658	-157	-177	-41	-92	113
10	0.7087	61	173	35	80	-82
6	0.7602	161	-179	-41	-74	118
9	0.7818	-170	-180	-32	93	86
4	0.9071	-168	-180	-32	92	-1
5	0.9283	169	180	33	-93	-1
1	0.9770	-162	179	31	-123	86
DFT DN Geometry Optimization Results						
2	0.0000	-162	-180	-10	-113	103
8	1.1521	-160	-178	-19	-107	106
7	1.3266	-46	-179	-7	-101	92
6	1.7194	166	-179	-16	-84	104
1	2.1247	-174	-177	21	-114	86
3	2.5515	-50	-173	-14	-89	94
9	3.0252	-171	-180	-25	86	84
5	5.6055	172	-180	28	-88	4
10	7.7956	52	178	19	91	-82
4	8.1463	-170	178	-25	89	-3
DFT DN* Geometry Optimization Results						
7	0.0000	-49	-177	-12	-100	92
6	0.7681	164	-177	-18	-84	110
8	0.9413	-160	-179	-20	-108	107
1	2.2741	-173	-178	24	-118	83
9	2.2804	-170	-180	-29	90	84
3	3.3747	-53	-174	-16	-91	95
2	4.0267	-160	177	-3	-126	111
10	5.1807	53	175	19	96	-86
5	6.2305	172	180	34	-96	7
4	6.9547	-169	178	-30	95	-2

Table 7.5. dG-N²-AAF conformer searching/geometry optimization cascade results. See Table 7.1 for conformer search parameters used in this study.

8. References

1. Voet, D., Voet, J. G., and Pratt, C. W. (1999) *Fundamentals of Biochemistry*. p. 728, John Wiley and Sons, Inc., New York.
2. Dipple, A., Cheng, S. C., and Bigger, C. A. H. (1990) Polycyclic Aromatic Hydrocarbon Carcinogens. In *Mutagens and Carcinogens in the Diet*, Ed. Pariza, M. W., Felton, J. S., Aeschbacher, H.-U., and Sato, S., pp. 109-127, Wiley-Liss Inc., New York.
3. Dipple, A., Moschel, R. C., and Bigger, C. A. H. (1984) Polynuclear Aromatic Carcinogens. In *Chemical Carcinogens*, Ed. Searle, C. E., Vol. 1, 2nd Ed., pp. 41-163, American Chem. Soc., Washington, D.C.
4. Grasso, P. (1984) Carcinogens in food. In *Chemical Carcinogens*, Ed. Searle, C. E., Vol. 1, 2nd ed., pp. 41-163, American Chem. Soc., Washington, D.C.
5. Grimmer, G. (1983) *Environmental Carcinogens: Polycyclic Aromatic Hydrocarbons*, CRC Press, Boca Raton, FL.
6. Peltonen, K. and Dipple, A. (1997) Polycyclic Aromatic Hydrocarbons: chemistry of DNA adduct formation. *J. Occup. Environ. Med.* **37**, 52-58.
7. Case, R. A., Hosker, M., McDonald, D. B., and Pearson, J. T. (1954) Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Parts I and II. *Br. J. Indust. Med.* **11**, 75, 213.
8. Rehn, L. (1895) Blasengeschwulste bei fuchsin arbeitern. *Arch. Klin. Chir.* **50**, 588-600. Reprinted in translation in Shinkin, M. B. (1980) *Some Classics of Experimental Oncology: 50 Selections, 1775-1965*, NIH Publication No. 80-2150, pp.45-51.
9. Kriek, E. (1974) Carcinogenesis by aromatic amines. *Biochemica et Biophysica Acta* **355**, 177-203.
10. Nesnow, S., Argus, M., Bergman, H., Chu, K., Frith, C., Helmes, T., McGaughy, R., Ray, V., Slaga, T.J., Tennant, R. and Weisburger, E. (1988) Chemical carcinogens. A review and analysis of the literature of selected chemicals and the establishment of the Gene-Tox Carcinogen Data Base. The

- U.S. Environmental Protection Agency Gene-Tox Program. *Mutation Res.* **185**, 1-195.
11. Weisburger, J. H. (1988) Past, present and future role of carcinogen and mutagenic N-substituted aryl compounds in human cancer causation. In *Carcinogenic and Mutagenic Response to Aromatic Amines, and Nitroarenes*, Eds. King, C. M., Romano, L. J., and Schueltze, D., pp. 3-19, Elsevier, New York.
 12. Clayson, D. B. (1962) *Chemical Carcinogenesis*, Little, Brown and Co., Boston.
 13. Hueper, W. C. (1969) *Occupational and Environmental Cancers of the Urinary System*, Yale University Press, New Haven.
 14. Scott, T. S. (1962) *Carcinogenic and Chronic Toxic Hazards of Aromatic Amines*, Elsevier, Amsterdam.
 15. Weisburger, E. K. and Weisburger, J. H. (1958) *Adv. Cancer Res.* **5**, 331.
 16. Miller, E. C. and Miller, J. A. (1955) *J. Natl. Cancer Inst.* **15**, 1571-1590.
 17. Arcosm, J. C. and Argus, M. F. (1974) *Chemical Induction of Cancer*, Vol. 11B, Academic, New York.
 18. Weisburger, E. K. and Thorgeirsson, S. S., Eds. (1981) Carcinogenic and Mutagenic N-Substituted Aryl Compounds. NCI Monograph 58 (U.S. Dept. of Health and Human Services, National Cancer Inst., Bethesda, MD, NIH Publ. No. 81-2379).
 19. Beland, F. A. and Kadlubar, F. F., Eds. (1985) Formation and persistence of arylamine DNA adducts *in vivo*. *Environ. Health Perspect.* **62**, 1.
 20. Lotlikar, P. O. (1981) In *Carcinogens in Industry and the Environment*, ed. J. M. Santag, pp. 345-438, New York.
 21. Weisburger, J. H. and Fiala, E. S. (1981) Mechanisms of species, strain, and dose effects in arylamine carcinogenesis. NCI Monograph 58, 41-48.
 22. Irving, C. C. (1973) In *Methods of Cancer Research*, ed. Busch, H., Vol. 7, pp. 189-224, Academic Press, New York.

23. Miller, E. C. (1978) Some current perspectives on chemical carcinogenesis in humans and experimental animals: presidential address. *Cancer Research* **38**, 1479-1496.
24. Garner, R. C., Martin, C. N., and Clayson, D. B. (1984) Carcinogenic aromatic amines and related compounds. In *Chemical Carcinogens*, Ed. Searle, C. S., ACS Monograph 182, Vol. 1, pp. 175-276, American Chemical Society, Washington D.C.
25. Beland, F. A., and Kadlubar, F. F. (1985) Formation and persistence of arylamine DNA adducts *in vivo*. *EHP, Environmental Health Perspectives* **62**, 19-30.
26. Duane, M. P., Westhof, E., Koffel-Schwartz, N., and Fuchs, R. P. P. (1985) Covalent binding of a carcinogen as a probe for the dynamics of deoxyribonucleic acid. *Biochemistry* **24**, 2275.
27. Gorrod, J. W. and Damani, L. A., Eds. (1985) *Biological Oxidation of Nitrogen in Organic Molecules*. Ellis Horwood, Chichester, England.
28. Adamson, R. H. and Sieber, S. M. (1983) In *Organ and Species Specificity in Chemical Carcinogenesis*. Eds. Langenbach, R., Nesnow, S., Rice, J. M., pp. 129-156, Plenum, New York.
29. Weisburger, E. K. (1983) In *Organ and Species Specificity in Chemical Carcinogenesis*. Eds. Langenbach, B., Nesnow, S., Rice, J. M., pp. 23-48, Plenum, New York.
30. Bishop, J. M. (1990) Retroviruses and Oncogenes. *Bioscience Reports* **10**, 473-491.
31. Braithwaite, E., Wu, X., Wang, Z. (1999) Repair of DNA lesions: mechanisms and relative repair efficiencies. *Mutation Research* **424(1,2)**, 207-219.
32. Zhang, Y., Yuan, F., Wu, X., Taylor, J., and Wang, Z. (2001) Response of human DNA polymerase ϵ to DNA lesions. *Nucleic Acids Research* **29(4)**, 928-935.
33. Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J., Geacintov, N. E., and Wang, Z. (2000) Error-free and error-prone lesion bypass by

- human DNA polymerase κ *in vitro*. *Nucleic Acids Research* **28(21)**, 4138-4146.
34. Luo, C., Krishnasamy, R., Basu, A. K., and Zou, Y. (2000) Recognition and incision of site-specifically modified C8 guanine adducts formed by 2-aminofluorene, N-acetyl-2-aminofluorene and 1-nitropyrene by UvrABC nuclease. *Nucleic Acids Research* **28(19)**, 3719-3724.
 35. Yamasaki, H., Pulkrabek, P., Grunberger, D., and Weinstein, I. B. (1977) Differential excision from DNA of the C-8 and N² Guanosine Adducts of N-acetyl-2-aminofluorene by single strand-specific endonucleases. *Cancer Research* **37**, 3756-3760.
 36. McGregor, W. G., Wei, D., Chen, R., Maher, V. M., and McCormick, J. J. (1997) Relationship between adduct formation, rates of excision repair and the cytotoxic and mutagenic effects of structurally-related polycyclic aromatic carcinogens. *Mutation Research* **376**, 143-152.
 37. Kriek, E. (1992) Fifty years of research on N-acetyl-2-aminofluorene, one of the most versatile compounds in experimental cancer research. *J. Cancer Res. Clin. Oncol.* **118**, 481-489.
 38. Wilson, R.H., DeEds, F. and Cox, A.J. (1941) The toxicity and carcinogenic activity of 2-acetaminofluorene. *Cancer Research* **1**, 595-608.
 39. Beland, F.A. and Kadlubar (1990) Metabolic Activation and DNA Adducts of Aromatic Amines and Nitroaromatic Hydrocarbons. In *Handbook of Experimental Pharmacology: Chemical Carcinogenesis and Mutagenesis*. Eds. Cooper, C.S. and Grover, P.L., Vol. 94/1, pp. 267-325, Springer Verlag, Heidelberg.
 40. Heflich, R.H. and Neft, R.E. (1994) Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutation Research* **318**, 73-174.
 41. Basu, A. K., and Essigmann, J. M. (1988) Site-specifically modified oligodeoxynucleotides as probes for the structural and biological effects of DNA-damaging agents. *Chem. Res. Toxicol.* **1**, 1-18.

42. Cramer, J. W., Miller, J. A., and Miller, E. C. (1960) *J. Biol. Chem.* **235**, 885-888.
43. Herlinglake R., Kiese, M., Renner, G., Wenz, W. (1960) N-Oxydation von 2-Naphthylamin *in vivo* und Wirkungen von Oxydationsprodukten des 2-Naphthylamins. *Naunyn-Schmiedebergs Arch. Pharmacol. Exp. Pathol.* **239**, 370-382.
44. Clayson, D. B., Garner, R. C. (1976) Carcinogenic aromatic amines and related compounds. In *Chemical Carcinogens*. Ed. Searle, C. E., pp. 366-461, American Chemical Society, Washington, D.C.
45. Kadlubar, F. F. and Hammons, G. J. (1987) The role of cytochrome P-450 in the metabolism of chemical carcinogens. In *Mammalian cytochromes P-450*. Ed. Guengerich, F. P., Vol II, pp. 81-130, CRC Press, Boca Raton, FL.
46. Irving, C. C. and Veazey, R. A. (1969) Persistent binding of 2-acetylaminofluorene to rat liver DNA *in vivo* and consideration of the mechanism of binding of N-hydroxy-2-acetylaminofluorene to rat liver nucleic acids. *Cancer Research* **29**, 1799-1804.
47. Kriek, E. (1969) On the mechanism of action of carcinogenic aromatic amines. I. Binding of 2-acetylaminofluorene and N-hydroxy-2-acetylaminofluorene to rat liver nucleic acids *in vivo*. *Chem.-Biol. Interactions* **1**, 3-17.
48. Meerman, J. H. N and Mulder, G. J. (1981) Prevention of the hepatotoxic action of N-hydroxy-2-acetylaminofluorene in the rat by inhibition of N-O-sulfation by pentachlorophenol. *Life Sciences* **28**, 2361-2365.
49. Meerman, J. H. N, Beland, F. A., and Mulder, G. J. (1981) Role of sulfation in the formation of DNA adducts from N-hydroxy-2-acetylaminofluorene in rat liver *in vivo*. Inhibition of N-acetylated aminofluorene adduct formation by pentachlorophenol. *Carcinogenesis* **2**, 413-416.
50. Visser A., Westra J. G. (1981) Partial persistency of 2-aminofluorene and N-acetyl-2-aminofluorene in rat liver DNA. *Carcinogenesis* **2**, 737-740.

51. Beland, F. A., Dooley, K. L., and Jackson, C. D. (1982) Persistence of DNA adducts in rat liver and kidney after multiple doses of the carcinogen N-hydroxy-2-acetylaminofluorene. *Cancer Research* **42**, 1348-1354.
52. Beland, F. A., Beranek, D. T., Dooley, K. L., Heflich, R. H., and Kadlubar, F. F. (1983) Arylamine-DNA adducts *in vitro* and *in vivo*: their role in the bacterial mutagenesis and urinary bladder carcinogenesis. *Environ. Health Perspect.* **49**, 125-134.
53. Poirier, M. C., True, B., Laishes, B. A. (1982) Formation and removal of (guan-8-yl)DNA-2-acetylaminofluorene adducts in liver and kidney of male rats given dietary 2-acetylaminofluorene. *Cancer Research* **42**, 1317-1321.
54. Poirier, M. C., Hunt, J. M., True, B., Laishes, B. A., Young, J. F., Beland, F. A. (1984) DNA adduct formation, removal and persistence in rat liver during one month feeding 2-acetylaminofluorene. *Carcinogenesis* **5**, 1591-1596.
55. Poirier, M. C., Fullerton, N. F., Beland, F. A. (1988) DNA adduct formation and removal chronic dietary administration of 2-acetylaminofluorene. In *Carcinogenic and Mutagenic Responses to Aromatic Amines and Nitroarenes*. Eds. King, C. M., Romano, L. J., and Schuetzle, D., pp. 321-328, Elsevier, New York.
56. Allaben, W. T., Weeks, C. E., Weis C. C., Burger, G. T., and King, C. M. (1982) Rat mammary gland carcinogenesis after local injection of N-hydroxy-N-acyl-2-aminofluorenes: relationship to metabolic activation. *Carcinogenesis* **3**, 233-240.
57. Gupta, R. C. and Dighe, N. R. (1984) Formation and removal of DNA adducts in rat liver treated with N-hydroxy derivatives of 2-acetylaminofluorene, 4-acetylaminobiphenyl, and 2-acetylaminophenanthrene. *Carcinogenesis* **5**, 343-349.
58. Lai C-C, Miller, J. A., Miller, E. C., Liem, A. (1985) Sulfoöxy-2-aminofluorene is the major ultimate electrophilic and carcinogenic metabolite of N-hydroxy-2-acetylaminofluorene in the livers of infant male C57BL/6J x C3H/HeJ F₁ (B6C3F₁) mice. *Carcinogenesis* **6**, 1037-1045.

59. Lai, C-C, Miller, E. C., Miller, J. A., Liem, A. (1987) Initiation of hepatocarcinogenesis in infant male B6C3F₁ mice by N-hydroxy-2-aminofluorene or N-hydroxy-2-acetylaminofluorene depends primarily on metabolism to N-sulfoöxy-2-aminofluorene and formation of DNA-(deoxyguanosin-8-yl)-2-aminofluorene adducts. *Carcinogenesis* **8**, 471-478.
60. King, C. M. (1974) Mechanism of reaction, tissue distribution, and inhibition of aryl-hydroxamic acid acyltransferase. *Cancer Research* **34**, 1503-1515.
61. King, C. M., Traub, N. R., Lortz, Z. M., Thissen, M. R. (1979) Metabolic activation of aryl-hydroxamic acids by N-O-acyltransferase of rat mammary gland. *Cancer Research* **39**, 3369-3372.
62. King, C. M., Glowinski, I. B. (1983) Acetylation, deacetylation and acyltransfer. *Environ Health Perspect.* **49**, 43-50.
63. Allaben, W. T., Weis, C. C., Fullerton, N. F., and Beland, F. A. (1983) Formation and persistence of DNA adducts from the carcinogen N-hydroxy-2-acetylaminofluorene in rat mammary gland *in vivo*. *Carcinogenesis* **4**, 1067-1070.
64. Shirai, T., Fysh, J. M., Lee, M-S, Vaught J. B., and King, C. M. (1981) Relationship of metabolic activation of N-hydroxy-N-acylarylamines to biological response in the liver and mammary gland of the female CD rat. *Cancer Research* **41**, 4346-4353.
65. Lotlikar, P. D. and Luha, L. (1971) Acetylation of the carcinogen N-hydroxy-2-acetylaminofluorene by acetyl coenzyme A to form a reactive ester. *Molecular Pharmacology* **7**, 381-388.
66. Howard, P.C., Casciano, D.A., Beland, F.A. and Shaddock, J.G. (1981) The binding of N-hydroxy-2-acetylaminofluorene to DNA and repair of the adducts in primary rat hepatocyte cultures. *Carcinogenesis* **2**, 97-102.
67. King, C. M. and Allaben, W. T. (1978) The role of arylhydroxamic acid N-O-acyltransferase in the carcinogenicity of aromatic amines. In *Conjugation Reactions in Drug Biotransformation*. Ed. Aitio, A., pp. 431-441, Elsevier Press.

68. Kriek, E. (1972) Persistent binding of a new reaction product of the carcinogen N-hydroxy-N-2-acetylaminofluorene with guanine in rat liver DNA *in vivo*. *Cancer Research* **32**, 2042-2048.
69. Westra, J., Kriek, E., and Hittenhausen, H. (1976) Identification of the persistently bound form of the carcinogen N-acetyl-2-aminofluorene to rat liver DNA *in vivo*. *Chem.-Biol. Interact.* **15**, 149-164.
70. Grunberger, D., Nelson, J. H., Cantor, C. R., Weinstein, I. B. (1970) Coding and conformational properties of oligonucleotides modified with the carcinogen N-2-acetylaminofluorene. *Proc. Natl. Acad. Sci. USA* **66**, 488-494.
71. Fuchs, R. and Duane, M. (1972) Physical studies on deoxyribonucleic acid after covalent binding of a carcinogen. *Biochemistry* **11**, 2659-2666.
72. Hingerty, B. E., and Broyde, S. (1982) Conformation of the deoxydinucleoside monophosphate d(CpdG) modified at carbon 8 of guanine with 2-(acetylaminofluorene). *Biochemistry* **21**, 3243-3252.
73. Neidle, S., Kiaroda, L., Broyde, S., Hingerty, B. E., Levine, L. A., Miller, D. W., Evans, F. E. (1984) Studies on the conformation and dynamics of the C8-substituted guanine adduct of the carcinogen acetylaminofluorene; model for a possible Z-DNA modified structure. *Nucleic Acids Research* **12**, 8219-8233.
74. Hingerty, B. E., and Broyde, S. (1982) Base displacement in AAF and AF modified dCpdG: *syn* and *anti* guanine. *Int. J. Quantum Chem., Quantum Biol. Symp.* **9**, 125-136.
75. Broyde, S. and Hingerty, B. (1985) Base displacement in AAF-modified Z-DNA. *Carcinogenesis* **6**, 151-154.
76. Sage, E. and Leng, M. (1980) Conformation of poly(dG-dC).poly(dG-dC) modified by the carcinogens N-acetoxy-N-acetyl-2-aminofluorene and N-hydroxy-N-2-aminofluorene. *Proc. Natl. Acad. Sci. USA* **77**, 4597-4601.
77. Santella, R. M., Grunberger, D., Broyde, S., and Hingerty, B. E. (1981) Z-DNA conformation of N-2-acetylaminofluorene modified by poly(dG-dC).poly(dG-dC) determined by reactivity with anti cytidine antibodies and minimized potential energy calculations. *Nucleic Acids Research* **9**, 5459-5467.

78. Santella, R. M., Grunberger, D., Weinstein, I. B., and Rich, A. (1981) Induction of the Z conformation in poly(dG-dC).poly(dG-dC) by binding of N-2-acetylaminofluorene to guanine residues. *Proc. Natl. Acad. Sci. USA* **78**, 1451-1455.
79. Fuchs, R. P. P. and Duane, M. P. (1974) Dynamic structure of DNA modified with the carcinogen N-acetoxy-N-2-acetylaminofluorene. *Biochemistry* **13**, 4435-4440.
80. Lavine, A. F., Fink, L. M., Weinstein, I. B., and Grunberger, D. (1974) Effect of N-2-acetylaminofluorene modification on the conformation of nucleic acids. *Cancer Research* **34**, 319-327.
81. Fuchs, R. P. P. (1975) *In vitro* recognition of carcinogen-induced local denaturation sites in native DNA by S₁ endonuclease from *Aspergillus oryzae*. *Nature* **257**, 151-152.
82. Beland, F. A. (1978) Computer-generated graphic models of the N²-substituted deoxyguanosine adducts of 2-acetylaminofluorene and benzo[*a*]pyrene and the O⁶-substituted deoxyguanosine adduct of 1-naphthylamine in the DNA double helix. *Chem. Biol. Interact.* **22**, 329-339.
83. Fritsch, V., Westhof, B. (1991) Minimization and molecular dynamics studies of guanosine and Z-DNA modified by N-2-acetylaminofluorene. *J. Comp. Chem.* **12**, 147-166.
84. Besson, M. and Mihalek, C. L. (2001) Total energy of deoxyguanosine bonded to N-2-acetylaminofluorene by the semi-empirical modified-neglect of differential diatomic overlap method. *Mutation Research* **473**, 211-217.
85. Topper, R. Q., Chung, K., Boelke, C., Louie, D., Kang, J. S., Hannan, R., Kiang, T., and Chan, L. H. (accepted, 2002) Computational structural determination and energy landscape analysis of the hepatic carcinogen 2-(acetylamino)fluorene. To be published in the ECCC Proceedings issue of *Theoretical Chemistry Accounts*.
86. Shapiro, R., Hingerty, B. E., and Broyde, S. (1989) Minor-groove binding models for acetylaminofluorene modified DNA. *J. Biomol. Struct. Dyn.* **7**, 493-513.

87. Broyde, S., and Hingerty, B. E. (1983) Conformation of 2-aminofluorene-modified DNA. *Biopolymers* **22**, 2423-2441.
88. Shapiro, R., Sidawi, D., Miao, Y., Hingerty, B. E., Schmidt, K. E., Moskowitz, J., and Broyde, S. (1994) Conformation of Amine-Modified DNA: 2-Aminofluorene- and 2-(Acetylamino)fluorine-Modified Deoxydinucleoside Monophosphates with All Possible Nearest Neighbors. A Comparison of Search and Optimization Methods. *Chemical Research in Toxicology* **7**, 239-253.
89. O'Handley, S. F., Sanford, D. G., Xu, R., Lester, C. C., Hingerty, B. E., Broyde, S., Krugh, T. R. (1993) Structural Characterization of an N-acetyl-2-aminofluorene (AAF) Modified DNA Oligomer by NMR, Energy Minimization, and Molecular Dynamics. *Biochemistry* **32**, 2481-2497.
90. Grad, R., Shapiro, R., Hingerty, B. E., Broyde, S. (1997) A molecular mechanics and dynamics study of the minor adduct between DNA and the carcinogen 2-(acetylamino)fluorene (dG-N-2-AAF). *Chem. Res. Toxicol.* **10**, 1123-1132.
91. Shibutani, S., and Grollman, A. P. (1993) On the mechanism of frameshift (deletion) mutagenesis *in vitro*. *J. Biol. Chem.* **268**, 11703-11710.
92. (1997) *SPARTAN Version 5.0 User's Guide*. Wavefunction, Inc., Irvine, CA.
93. Hehre, W. J., Huang, W. W., Klunzinger, P. E., Deppmeier, B. J., and Driessen, A. J. (1997) *A SPARTAN Tutorial*. Wavefunction, Inc., Irvine, CA.
94. Hehre, W. J., Huang, W. W. and Nelson, J. E. (1997) *A Guide to Graphical Models and Graphical Modeling in SPARTAN*. Wavefunction, Inc., Irvine, CA.
95. Hehre, W. J., Yu, J. and Klunzinger, P. E. (1997) *A Guide to Molecular Mechanics and Molecular Orbital Calculations in SPARTAN*. Wavefunction, Inc., Irvine, CA.
96. Hehre, W. J. and Lou, L. (1997) *A Guide to Density Functional Calculations in SPARTAN*, Wavefunction, Inc., Irvine, CA.
97. Hehre, W. J. (1995) *Practical Strategies for Electronic Structure Calculations*, Wavefunction, Inc., Irvine, CA.

98. Reynolds, C. H. (1997) Semiempirical MO methods: the middle ground in molecular modeling. *J. Mol. Struct. (Theochem)* **401**, 267.
99. Levine, I. N. (2000) *Quantum Chemistry*, 5th ed., p. 665, Prentice Hall, Upper Saddle River, NJ.
100. Rappé, A. K. and Casewit, C. J. (1997) *Molecular Mechanics across Chemistry*, p. 8, University Science Books, Sausalito, CA.
101. Lennard-Jones, J. E. (1924) *Proc. R. Soc. London, Ser. A* **106**, 463.
102. Rappé, A. K. and Casewit, C. J. (1997) *Molecular Mechanics across Chemistry*, pp. 15-20, University Science Books, Sausalito, CA.
103. Burkert, U. and Allinger, N. L. (1982) *Molecular Mechanics*. ACS Monograph No. 177, American Chemical Society.
104. Allinger, N. L. and Yan, L. (1993) Molecular mechanics (MM3). Calculations of furan, vinyl ethers, and related compounds. *J. Am. Chem. Soc.* **115**, 11918.
105. Allinger, N. L., Chen, K., Lii, J-H. (1996) An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **17**, 642.
106. Allinger, N. L., Chen, K., Katzenellenbogen, J. A., Wilson, S. R., Anstead, G. M. (1996) Hyperconjugative effects on carbon-carbon bond lengths in molecular mechanics (MM4). *J. Comput. Chem.* **17**, 747.
107. Nevins, N., Chen, K., Allinger, N. L. (1996) Molecular mechanics (MM4) calculations on alkenes. *J. Comput. Chem.* **17**, 669.
108. Nevins, N., Lii, J-H., Allinger, N. L. (1996) Molecular mechanics (MM4) calculations on conjugated hydrocarbons. *J. Comput. Chem.* **17**, 695.
109. Nevins, N., Allinger, N. L. (1996) Molecular mechanics (MM4) vibrational frequency calculations for alkenes and conjugated hydrocarbons *J. Comput. Chem.* **17**, 730.
110. Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A. (1986) An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **7**, 230.
111. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., Weiner, P. (1984) A new force field for molecular

- mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765.
112. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179.
113. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187.
114. MacKerell, A. D., Jr., Wiorkiewicz-Kuczera, J., Karplus, M. (1995) An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **117**, 11946.
115. MacKerell, A. D., Jr., Bashford, D., Bellott, R. L., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., III, Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **102**, 3586.
116. Jorgensen, W. L. and Tirado-Rives, J. (1988) The OPLS (optimized potentials for liquid simulations) potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657.
117. Pranata, J., Wierschke, S. G., Jorgensen, W. L. (1991) OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. *J. Am. Chem. Soc.* **113**, 2810.
118. Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225.
119. Hwang, M. J., Stockfish, T. P., and Hagler, A. T. (1994) Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field,

- CFF93, for the Alkyl Functional Group and Alkane Molecules. *J. Am. Chem. Soc.* **116**, 2515.
120. Maple, J. R., Hwang, M. -J., Jalkanen, K. J., Stockfisch, T. P., Hagler, A. T. (1998) Derivation of class II force fields. V. Quantum force field for amides, peptides, and related compounds. *J. Comput. Chem.* **19**, 430.
121. Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., III, Skiff, W. M. (1992) UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024.
122. Clark, M., Cramer, R. D., III, Van Opdenbosch, N. (1989) Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **10**, 982.
123. Halgren, T. A. (1996) Merck Molecular Force Field. Part I. Basis Form, Scope, Parametrization, and Performance of MMFF94. *J. Comp. Chem.* **17(5,6)**, 490-519.
124. Halgren, T. A. (1996) Merck Molecular Force Field. Part II. *J. Comp. Chem.* **17(5,6)**, 520-552.
125. Halgren, T. A. (1996) Merck Molecular Force Field. Part III. *J. Comp. Chem.* **17(5,6)**, 553-586.
126. Halgren, T. A. and Nachbar, R. B. (1996) Merck Molecular Force Field. Part IV. *J. Comp. Chem.* **17(5,6)**, 587-615.
127. Halgren, T. A. (1996) Merck Molecular Force Field. Part V. *J. Comp. Chem.* **17(5,6)**, 616-641.
128. ftp.wiley.com/public/journals/jcc/suppmat/17/490.
129. Levine, I. N. (2000) *Quantum Chemistry*, 5th ed., pp. 629-630, 652-664, Prentice Hall, Upper Saddle River, NJ.
130. Klopman, G. A. and Evans, R. C. (1977) Chapter 2 in *Semiempirical Methods of Electronic Structure Calculation, Part A*, Ed. Segal, G. A., Plenum.
131. Dewar, M. J. S. and Thiel, W. (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **99**, 4899.

132. Dewar, M. J. S. and Thiel, W. (1977) Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J. Am. Chem. Soc.* **99**, 4907.
133. Dewar, M. J. S. and Rzepa, H. S. (1978) Ground states of molecules. 40. MNDO results for molecules containing fluorine. *J. Am. Chem. Soc.* **100**, 58.
134. Dewar, M. J. S. and Rzepa, H. S. (1978) Ground states of molecules. 45. MNDO results for molecules containing beryllium. *J. Am. Chem. Soc.* **100**, 777.
135. Dewar, M. J. S. and Rzepa, H. S. (1978) Calculations of electron affinities using the MNDO semiempirical SCF-MO method. *J. Am. Chem. Soc.* **100**, 784.
136. Dewar, M. J. S. and McKee, M. L. (1977) Ground states of molecules. 41. MNDO results for molecules containing boron. *J. Am. Chem. Soc.* **99**, 5231.
137. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart, J. J. P. (1985) Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902.
138. Stewart, J. J. P. (1989) Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **10**, 209.
139. Stewart, J. J. P. (1989) Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem.* **10**, 221.
140. Stewart, J. J. P. (1990) Reply to "Comments on a comparison of AM1 with the recently developed PM3 method". *J. Comput. Chem.* **11**, 543.
141. Stewart, J. J. P. (1991) Optimization of parameters for semiempirical methods. III. Extension of PM3 to beryllium, magnesium, zinc, gallium, germanium, arsenic, selenium, cadmium, indium, tin, antimony, tellurium, mercury, thallium, lead, and bismuth. *J. Comput. Chem.* **12**, 320.
142. Levine, I. N. (2000) *Quantum Chemistry*, 5th ed., pp.573-592, Prentice Hall, Upper Saddle River, NJ.
143. Hohenberg, P. and Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev.* **136**, B864.

144. Kohn, W. and Sham, L. J. (1965) Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133.
145. Perdew, J. P. (1986) Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33**, 8822.
146. Becke, A. D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3089.
147. Becke, A. D. (1988) A multicenter numerical integration scheme for polyatomic atoms. *J. Chem. Phys.* **88**, 2547.
148. Versleis, L. Ziegler, T. (1988) The determination of molecular structures by density functional theory. The evaluation of analytical energy gradients by numerical integration. *J. Chem. Phys.* **88**, 322.
149. Delley, B. (1990) An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.* **92**, 508.
150. Van Meersche, M., Germain, G., Declercq, J. P., and Touillaux, R. (1980) 2-acetylaminofluorene, C₁₅H₁₃NO. *Cryst. Struct. Commun.* **9**, 515-518.
151. Haisa, M., Kashino, S., Ueno, T., Shinozaki, N., and Matsuzaki, Y. (1980) The structures of N-aromatic amides: p-acetanisidide, N-2-naphthylacetamide, and N-2-fluorenylacetamide. *Acta Crystallogr. B* **36**, 2306-2311.
152. Voet, D., Voet, J. G., and Pratt, C. W. (1999) *Fundamentals of Biochemistry*. p. 732, John Wiley and Sons, Inc., New York.
153. Fritsch, V., Westhof, B. (1991) Minimization and molecular dynamics studies of guanosine and Z-DNA modified by N-2-acetylaminofluorene. *J. Comp. Chem.* **12**, 147-166.