

# Rhodes College Digital Archives - DLynx

## Ab Initio and DFT Calculations of Increasingly Complex Models of Ligand-Nucleic Acid Binding

Item Type	Thesis
Authors	Shroyer, Michelle
Publisher	Memphis, Tenn. : Rhodes College
Rights	Rhodes College owns the rights to the archival digital objects in this collection. Objects are made available for educational use only and may not be used for any non-educational or commercial purpose. Approved educational uses include private research and scholarship, teaching, and student projects. For additional information please contact <a href="mailto:archives@rhodes.edu">archives@rhodes.edu</a> . Fees may apply.
Download date	2026-04-15 13:04:05
Link to Item	<a href="http://hdl.handle.net/10267/9709">http://hdl.handle.net/10267/9709</a>

Ab Initio and DFT Calculations of Increasingly Complex Models of Ligand-Nucleic Acid  
Binding

Written at Rhodes College

Michelle Catherine Shroyer

Department of Chemistry

Rhodes College

Memphis, Tennessee

2011

Submitted in partial fulfillment of the requirements for the  
Bachelor of Science degree with Honors in Chemistry

This Honors paper by Michelle Shroyer has been read  
and approved for Honors in Chemistry.

Dr. Mauricio Cafiero

Project Advisor

---

Dr. David Y. Jeter

Second Reader

---

Dr. Deseree Meyer

Extra-Departmental Reader

---

Dr. Darlene Loprete

Department Chair

---

## CONTENTS

Signature page	ii
Contents	iii
List of Figures and Tables	iii
Abstract	v
Introduction	1
Methods	11
Results and Discussion	12
Conclusion	21
Works Cited	23
Figure 1: The process of an intercalant binding in between the rungs of the DNA ladder and the subsequent deformation of the DNA molecule.	3
Table 1: Counterpoise corrected interaction energies between a free complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.	13
Table 2: Counterpoise corrected interaction energies between a free complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.	13
Table 3: Counterpoise corrected interaction energies for the forced sandwich complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.	14
Figure 2: Comparison of free and forced structures for Adenine-Indole-Thymine (AIT).	14
Figure 3: Comparison of free and forced structures for Cytosine-Indole-Guanine (CIG).	15
Figure 4: Comparison of the free and forced structures for Adenine-Indole-Adenine (AIA).	16

Table 4: MP2 results for the single-stranded ONIOM counterpoise corrected interaction energies (kcal/mol)	16
Table 5: SVWN Results for the single-stranded ONIOM counterpoise corrected interaction energies (kcal/mol)	17
Figure 5: Single-stranded ONIOM interactions; two bases with indole	18
Figure 6: Single-stranded ONIOM structures; three bases with indole	18
Table 6: Counterpoise corrected interaction energies using MP2 for double-stranded ONIOM structures	19
Table 7: Counterpoise corrected interaction energies using SVWN for double-stranded ONIOM structures	20
Figure 7: Double-stranded ONIOM structures	20
Table 8: Comparison of the single and double structure interaction energies (kcal/mol) using MP2/6-31+g*.	21

## ABSTRACT

## Ab Initio and DFT Calculations of Increasingly Complex Models of Ligand- Nucleic Acid Binding

by

Michelle Catherine Shroyer

This work focuses on the binding of a model intercalant (indole) between the pairings of the four DNA bases to ascertain the most likely location of the binding of a carcinogen or chemotherapeutic drug. This knowledge will aid in the design and targeting of future chemotherapy drugs. Ten complexes (DNA base-intercalant-DNA base) were optimized into free structures and sandwich structures using MP2/6-31g. Interaction energies were computed for these complexes using the MP2 and DFT methods with the 6-31+g\* and 6-311+g\* basis sets. Comparison of the stability of the free structures to the stability of the sandwich structures allowed estimation of the distorting force acting on that segment of DNA. A short strand of DNA with an intercalant bound in between two bases was then studied using the ONIOM method. MP2 and DFT methods were used to model the interactions between the bases and the intercalant, while AM1 was used to model the DNA backbone. These calculations are crucial because the manner in which DNA distorts depends heavily on the backbone and the presence of the neighboring DNA bases. Finally, models of the intercalant bound between pairs of two bases in double-stranded DNA were modeled using ONIOM. The double-stranded results show interactions around -40 kcal/mol for all pairings of DNA bases, and this result is consistent with the single-stranded DNA and free nucleic acid base models.

I give permission for public access to my Honors paper and for any copying of digitization to be done at the discretion of the College Archivist and/or the College Librarian.

Signed \_\_\_\_\_

Michelle Shroyer

Date \_\_\_\_\_

## 1. Introduction

Cancer is a disease that affects millions of lives each year. It is difficult to find an individual who has not witnessed the devastating effects of cancer, either through personal experience or through the experience of a friend or family member. Cancer is already a common disease, yet statistics portray that in 2010 there was an estimated total of 1,529,560 new cancer cases with 569,490 cancer related deaths in the United States alone (Jemal *et al.*, 2010). Although there are still new cancer cases diagnosed each year, overall there has been a decrease in the number of cancer-related deaths over the last 16 years (Jemal *et al.*, 2010). This decrease is due to early detection, preventative measures, and the improvement of drug therapies (Jemal *et al.*, 2010). Even with this decrease, cancer is still prevalent in the United States, and it is evident that drug designers can improve the lives of those who are diagnosed with cancer. This goal of continually working for better cancer treatments is a driving force for increased research in an effort to cure and prevent this life threatening disease.

There are many characteristics that define cancer, as there are more than one hundred different forms of this disease. Cancer begins when cells no longer respond to stimulatory and inhibitory signals regulating their growth. Cells can then develop the ability to spread from their origin into other tissues, as they grow more aggressive over time. Once invasive cancers form they can not only invade tissues but also blood vessels in order to support their growth through angiogenesis. Cancer tumors eventually prove to be deadly when they begin disrupting the functions of vital organs in the body. (Weinberg, 1996)

Surgery, radiation treatment, and chemotherapeutic drugs are all used to treat cancer. Large tumors are often removed using surgery, while smaller more spread out

cancers are treated through radiation. Chemotherapeutic drugs are used to target hard to reach or lingering cancerous tissue. Chemotherapeutic drugs often work by binding directly to DNA through two different methods of action: groove binding and intercalation. Groove binding occurs when crescent shaped molecules bind to the minor groove of DNA. This binding does not result in major conformational changes and mirrors the standard lock and key model of ligand binding. Intercalation results in more disruptive DNA alterations than groove binding. Intercalation is the process where a planar molecule inserts itself between the DNA base pairs. This insertion leads to a decrease in the helical twist of the DNA. The DNA is then elongated. As the structure of the DNA is disrupted, the DNA's function is altered (Palchadhuri, 497-300). This intercalation method of DNA binding has been the focus of this work.

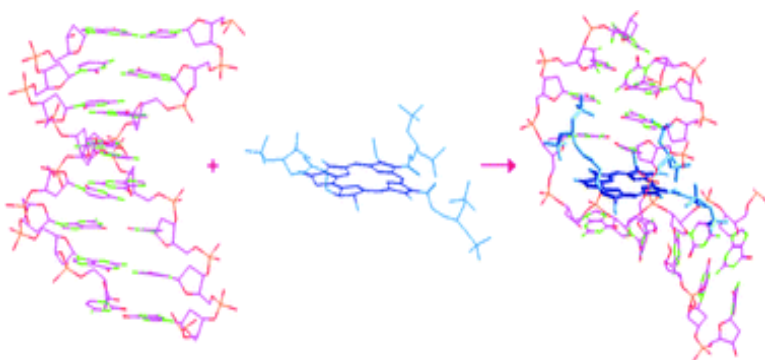
### **1.1 Intercalation**

In recent years more information has become available on the sequence of the human genome along with better understanding of the molecular basis of cancer. This combination allows for more insight into novel genetic targets that can be utilized for chemotherapy drug design. This approach depends on the creation of drugs capable of discriminating between binding sites in DNA. In order to understand these binding sites, more work is needed on how small molecules recognize and bind reversibly to DNA through intercalation. Through computational research these binding sites can be identified and molecules can be designed that are specialized for these sites prior to synthesizing new drugs for site specific chemotherapy. (Haq, 1-15)

DNA structure is known to be vital for its function. When the intercalant molecule binds to the DNA, it distorts the vital structure of the DNA rendering it unrecognizable

by the enzymes that wind and unwind DNA during replication (topoisomerases). When replication is disrupted, apoptosis can occur which kills the cell using energy from ATP. Intercalation also leads to lengthening of the DNA helix through unwinding, extending, and stiffening the helix (Long and Barton, 271-273). The lengthening disrupts the hydrogen bonding of the nucleic acid base pairs. According to Haq, intercalation normally results in a lengthening of about 1 base pair, approximately 3.4Å. These structural changes in the DNA can be used to predict the binding mode of the intercalant. (Haq, 1-15)

**Figure 1: The process of an intercalant binding in between the rungs of the DNA ladder and the subsequent deformation of the DNA molecule.**



Intercalants commonly used in current chemotherapy research are utilized because of their basic structure. Intercalant molecules are often polycyclic aromatic rings with either heteroatoms or polar groups attached. This hetero-aromatic ring structure allows for pi-pi stacking between the intercalant and DNA. Noncovalent interactions like pi-pi stacking between aromatic molecules contribute greatly to the stability of DNA (El Gogary and Koehler, 57-64). Intercalants have been traditionally used as antitumor, antibiotic, antifungal, antineoplastic, and antimalarial agents; however, not all intercalators are genotoxic. In order to ensure that the agent is genotoxic the presence of basic, electrophilic, or cationic groups are vital (Palchaudhuri, 497-300). Numerous

current chemotherapeutic drugs-for example Ellipticine-are known to act by the intercalant method of action. Other common intercalants include ethidium, daunorubicin, propidium, and adriamycin. (Ren, 8439-8447)

In order to determine the relationship between the intercalant and its target DNA, it is necessary to calculate the binding free energy ( $\Delta G_{\text{obs}}$ ). This energy contains contributions from a minimum of five major sources, and these components can be added together to find the binding free energy.

eq. 1

$\Delta G_{\text{conf}}$  is the energy contribution due to conformational transitions in the intercalant and the DNA and is an unfavorable contribution.  $\Delta G_{\text{r+t}}$  is defined as the free energy cost due to the restriction of translational and rotational freedom of the intercalant. This value is normally a large positive value as it is restricting force, not a force that leads to tighter binding of the intercalant.  $\Delta G_{\text{pe}}$  denotes the polyelectrolyte contribution.  $\Delta G_{\text{hyd}}$  is the hydrophobic interaction; intercalants are commonly lipophilic and hydrophobic, so when the intercalant is able to wedge itself between the nucleic acids, it shields the intercalant from the water, and this reaction is favorable.  $\Delta G_{\text{mol}}$  is the contribution from all other major molecular forces including H-bonding, and van der Waals interactions that hold the intercalant to the DNA. This interaction energy is normally a large negative number demonstrating stabilization of the complex. (Ren, 8439-8447)

Previous work on intercalants has revealed the importance of  $\Delta G_{\text{mol}}$  for each specific intercalant in the stabilization of intercalants in the helix. It is believed that  $\Delta G_{\text{mol}}$  is the most important contributor to the  $\Delta G_{\text{obs}}$ , as all the other contributions roughly

cancel each other out (Ren, 8439-8447). The negative contribution from  $\Delta G_{\text{mol}}$  is responsible for the negative total  $\Delta G_{\text{obs}}$ ; in this research we focus on calculating  $\Delta G_{\text{mol}}$ .

## 1.2 Intermolecular Forces and Intercalation

The role of dispersion, hydrogen bonding, and other noncovalent forces in the formation of molecular complexes is not as clearly understood as covalent interactions. These forces play a vital role in biomolecular structures that include nucleic acids, ligands, and proteins. H-bonding and dispersion/induction are believed to be equally important in the binding of these molecular structures. H-bonding is an intermolecular force of attraction between a small electronegativity charged atom and a hydrogen atom. The two sides of the DNA double helix are held together by these hydrogen bonds. Dispersion is the weakest intermolecular force. Dispersion forces occur between nonpolar molecules due to the momentary dipoles created from uneven electron distribution in adjacent molecules. Both of these forces play a vital role in the binding of intercalants and the stabilization of DNA (Riley, 1-17).

Ellipticine is a four ring aromatic compound that has shown promise as an effective intercalator. Ellipticine's stabilization with nucleobases is shown to be largely due to London dispersion forces. In order to accurately describe these forces with quantum chemistry, high level *ab initio* methods are required such as MP2 and CCSD(T). These methods will be described in Section 2. A previous study, done by Lin *et al.*, examines the role of ellipticine as a DNA binding intercalant. They calculated the interaction energy for the intercalator and DNA using the following equation:

$$E_{\text{int}} = E_{\text{whole complex}} - E_{\text{ellipticine}} - E_{\text{4-nucleobase complex}} \quad \text{eq. 2}$$

In order to calculate the three component interaction energy ( $\Delta^{(3)}$ ) for only two nucleic acids and the intercalant they utilized the following equation:

$$\Delta^{(3)} = E_{ABI} - \Delta E_{A-I} - \Delta E_{B-I} - \Delta E_{A-B} \quad \text{eq. 3}$$

Using MP2, B3LYP, and DCACP, they calculated the interaction energy for one side of the helix with ellipticine as the intercalator. For the two nucleic acid pairs (GC and AT), they found interaction energies ( $\Delta^{(3)}$ ) of around -20kcal/mol (Lin *et al.*, 14348-14350). Ren *et al.* use calorimetry to measure binding energies of several intercalants and from this information extract the interaction energies. Their experimental values for several intercalants are all around -15 kcal/mol (Ren, 8439-8447).

When analyzing stacking complexes, the complex frequently deviates from the sandwich structure and forms a t-shaped structure due to the tendency of the complex to form H-bonds. Work by El- Gogary and Koehler revealed that forming stacked structures was quite difficult because the complexes could form more stable structures through H-bonding. Their data depicts the interaction between two complexes of nucleic acids with an intercalant in between in the stacked position to range from -16.16 to -19.93 kcal/mol. (El Gogary and Koehler, 57-64)

Although there have been previous studies on the interaction of intercalant molecules with aromatic rings or single DNA bases, there has been little high accuracy computational research done on the nature of the binding of intercalant molecules between pairs of DNA bases. It is known that many intercalants work by simultaneously binding between two different sites of the DNA, yet this phenomenon has not been modeled using computation. Previous work by this author focused on the interaction of complexes of six-carbon-rings with variable numbers of double bonds using density

functional theory (DFT) and *ab initio* methods. This work, which has culminated in a recently published article, studied the same interactions present in intercalation. (Van Sickle, 78)

## 2. Computational methods

Dispersion energies are vital for ring stacking interactions. Cerny *et al.* have shown that calculations of the structure of DNA done without dispersion have an elongated pitch between the DNA ladder with a distance of 7Å instead of the normal 3.6Å (Cerny, 16055). Clearly these dispersion forces are necessary for the structure and functionality of DNA. The first dispersion method used for these calculations was the second order Moller Plessett theory (MP2). This method takes into account contributions to the wave function and molecular energy due to doubly excited determinants:

$$\psi_{MP2}(r_1, r_2, \dots, r_n) = \psi_{HF} + \sum_{i < j} \sum_{a < b} C_{ij}^{ab} \Phi_{ij}^{ab} \quad \text{eq. 4}$$

MP2 is one of the most frequently used wave function theory methods (WFT). This method can provide accurate estimations for hydrogen bonded systems, yet it is not the best method for stacking calculations with large systems because MP2 overestimates the binding energy. The reason for this over-binding is due to the uncoupled Hartree Fock (HF) dispersion energy that overestimates the interaction energies by between 10-20%. To account for this overestimation in the calculations, MP2 can be used with a smaller basis set. When MP2's overcompensation of the interaction energies is coupled with a smaller basis set, it results in a fairly accurate calculation. MP2 is the least expensive high accuracy method in terms of processing time and computing resources. (Riley, 1-17)

There are other more accurate and also more expensive post-HF methods available. One group of these methods is coupled cluster, CCSD(T), which is the industry

standard for high accuracy calculations. CCSD(T) methods are not practical for large protein/ligand complexes because of the excessive amount of resources and time required for the calculations. In this work, MP2 is the standard method used and can be considered to be accurate. Although, it is not as precise as CCSD(T), in this work differences in energies are considered, not absolute energies. (Riley, 1-17)

A major focus of this research is to identify single-determinant methods that are fairly accurate and yet are fast and require a reasonable amount of resources. Single determinant methods are methods that do not make explicit reference to excited states such as HF. These methods take about as much time to run as HF, but not nearly as much time as MP2. These methods are called Density Functional Theory (DFT) methods. The DFT methods (specifically the Kohn-Sham implementation of DFT) which are used use an exchange/correlation potential,  $v_{xc}$ , to add electron correlation energy within a single determinant framework;

$$E[\rho, \nabla\rho] = E_T[\rho] + E_J[\rho] + E_{ext}[\rho] + E_{xc}[\rho, \nabla\rho] \quad \text{eq. 5}$$

$$\rho = 2\sum\phi_{KS}^2 \quad \text{eq. 6}$$

In this equation  $\rho$  is the electron density obtained from summing over the squares of the Kohn-Sham orbitals.  $\phi_{KS}$  and the subscripts in the equation refer to T, the kinetic energy. J is the Coulomb energy, ext is the external potential, and XC is the exchange/correlation energy. Kohn-Sham DFT is a formally exact theory, yet the form of the XC potential is not known. Because of this lack of knowledge, different approximations are used for correlation. These correlation functionals of the electron density work well in some cases, but not in all. (Riley, 1-17)

DFT methods are able to provide the exact solution to the Schrodinger equation including the long range correlation. However, DFT provides poor results for noncovalent interactions that focus on dispersion such as stacking interactions. DFT methods are more accurate with local interactions such as covalent bonds and hydrogen bonds, but they are not as accurate with nonlocal interactions such as dispersion. DFT does have a number of advantages over wave function based methods (WFT), and because of these advantages DFT methods have recently been focused on in an attempt to overcome the problems concerning dispersion. (Riley, 1-17)

B3LYP is an improved DFT method that utilizes reparameterization of current density functionals. It is a widely used functional in much organic and bio-organic computational research. B3LYP is derived from the Becke's exchange functional with the correlation functional by Lee, Yang, and Parr. This method was designed without considering noncovalent complexes; therefore, van der Waals attractions are not well-represented by this method. This method has been proven to be inaccurate when calculating the interaction energies between stacked molecules. Previous work utilizing B3LYP portrays its' inability to accurately describe noncovalent interactions. Work by Hofto, van Sickle, and Cafiero modeled intercalation through measuring sandwich-type interactions between polyaromatic molecules and benzene. They calculated the interaction energy for each of fourteen complexes using MP2, HCTH407, SVWN, B3LYP and HF. As predicted MP2 and SVWN produced accurate results that were in agreement with the favorable distances between the molecules. B3LYP, HCTH407, and HF did not model the systems well (Hofto, 112). The recent work by Zhao and Truhlar

also confirms this result. They found that B3LYP fails to describe dispersion interactions in their experimentation (Zhao and Truhlar, 289-300).

A more accurate DFT method for modeling dispersion dominated stacking interactions is SVWN. SVWN is a local exchange functional that uses unrestricted wavefunctions to find the local spin. SVWN is also called LSDA or local spin density approximation. Research by Zhao and Truhlar reveals that "LSDA, BHand H, M05-2X, PWB6K, and MP2 are the best performers for the prediction of interaction energies of the dispersion-dominated (or dispersion-like-dominated) complexes" (Zhao and Truhlar, 289-300). Their previous research demonstrates that LSDA gives accurate predictions for stacked dimers, yet errors result when attempting to calculate hydrogen bonding, dispersion, charge-transfer complexes, and dipole interactions. Previous work in this research group has also shown that for molecular complexes wherein one molecule in the complex contains an aromatic moiety, SVWN mimics the values and trends of the MP2 method (Van Sickle, 78).

## **2.1 ONIOM Calculations**

For large systems consisting of a hundreds of atoms, methods such as ONIOM are often applied. The ONIOM method allows the system that is being studied to be divided into different levels to treat each level with a different accuracy of theory. The upper level is treated with a high accuracy level of theory, while the lower level is treated with a lower accuracy method of theory. The upper level is normally treated with an ab initio or DFT method. The lower level requires a less demanding method like HF or molecular mechanics. This use of multiple methods for one system allows one to save computing

resources and time through using a faster, less demanding method on the area of the complex that is not as important when calculating the binding energy. (Utkov)

### **3. Methods**

#### **3.1 Unattached Nucleic Acid Base Models**

Ten complexes of all the possible pairings of DNA bases (Cytosine, Adenine, Guanine, and Thymine) with an indole molecule in between were created. Indole was used as a model intercalant because it has a double ring and contains a heteroatom, and this represents a basic archetype of a chemotherapeutic drug molecule. Each complex was optimized using B3LYP/6-31G to find the lowest energy conformation using analytical Hessians. Each complex was also optimized into a sandwich configuration using MP2/6-31+g\*. Counterpoise corrected interaction energies for each complex in both configurations were calculated using MP2/6-31+g\* and MP2/6-311+g\*. All calculations were repeated using fast DFT methods: SVWN and B3LYP and the 6-311+g\* basis set.

#### **3.2 Single-Stranded DNA Model**

Small sections of RNA with three nucleic acid bases were isolated from a crystal structure (Goto-Ito). Protons were added to this system, and the indole molecule was placed in between two of the nucleic acid bases. The structure of the entire complex was then optimized using ONIOM with SVWN for the indole molecule and nucleic acid bases, while AM1 was used for the sugar and phosphate portions of the backbone. Interaction energies were calculated for this complex using MP2 and SVWN with the 6-31+g\* basis set, as the difference between the energy of the entire complex and the intercalant plus DNA.

### 3.3 Double-Stranded DNA Model

Small sections of DNA with four nucleic acid bases (two complementary pairs per side) were isolated from a crystal structure (Goto-Ito). Protons were added to this system, and the indole molecule was placed in between the four nucleic acid bases. The entire complex was then optimized using ONIOM with SVWN for the indole molecule and nucleic acid bases, while AM1 was used for the sugar and phosphate portions of the backbone. Interaction energies were calculated for this complex using MP2 and SVWN with the 6-31+g\* basis set, as the difference between the energy of the entire complex and the intercalant plus DNA. The DNA binding energy was calculated as well and subtracted from the total binding energy to get the intercalant binding energy.

All calculations were made using either the Gaussian '03 or the PQS Software (Gaussian and PQS).

## 4. Results and Discussion

### 4.1 Free Nucleic Acid Base Models

Table 1 shows the counterpoise corrected interaction energies between each free optimization of two nucleic acid bases with the model intercalant, indole, in between (AIA, AIG, AIT, CIC, CIA, CIG, CIT, GIG, GIT, and TIT). These interaction energies ranged from -6.09 (AIA) to -20.58 (GIG). These results show that the weakest bonding was typically between complexes involving adenine or thymine. Tighter binding was found when guanine was present. B3LYP is fairly accurate for most of these complexes because these are by design dominated by hydrogen bonds.

**Table 1: Counterpoise corrected interaction energies between a free complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.**

6-31+g*	aia	Aig	ait	cic	cia	cig	cit	gig	git	tit
MP2	-6.09	-14.79	-17.89	-15.72	-18.79	-15.49	-12.11	-20.58	-15.70	-13.31
SVWN	-2.59	-19.58	-25.86	-20.56	-24.62	-22.27	-16.84	-28.57	-22.65	-19.56
B3LYP	1.95	-8.58	-12.48	-11.93	-13.24	-11.88	-8.98	-14.75	-10.98	-8.53

Table 2 shows the counterpoise corrected interaction energies for the free optimization of the same ten complexes, but with a larger basis set, 6-311+g\*. These values ranged from -6.40 (AIA) to -21.22 (GIG) and demonstrated the similar trends as found when using the 6-31+g\* basis set. This agreement between the smaller and the larger basis set allows the use of only the smaller basis for later calculations.

**Table 2: Counterpoise corrected interaction energies between a free complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.**

6-311+g*	aia	aig	ait	cic	cia	cig	cit	gig	git	tit
MP2	-6.40	-15.20	-18.40	-16.03	-19.24	-15.78	-12.14	-21.22	-16.25	-13.65
SVWN	-2.44	-19.85	-26.39	-20.86	-24.97	-22.58	-16.96	-29.12	-23.11	-19.92
B3LYP	2.02	-8.82	-12.97	-12.22	-13.24	-12.13	-9.16	-15.15	-11.34	-8.80

## 4.2 Forced Sandwich Structures

The interaction energies for the forced sandwich structures are shown in Table 3. These values range from -8.62 (AIT) to -20.20 (CIG). Overall, these values were lower than those for the free energy structures. When analyzed individually, the values for the forced structures were larger than the values for the free structures with the same basis set, 6-311+g\*, for AIA, CIG, CIT, and TIT.

**Table 3: Counterpoise corrected interaction energies for the forced sandwich complex of bases X and Y with an intercalant (I) in between. Format is XIY. Energies are in kcal/mol.**

6-311+g*	aia	aig	ait	cic	cia	cig	cit	gig	git	tit
MP2	-10.22	-12.00	-8.62	-10.73	-16.60	-20.20	-19.36	-12.32	-12.03	-18.38
SVWN	-5.89	-13.74	-5.78	-8.38	-18.28	-22.03	-21.46	-13.58	-8.69	-22.50
B3LYP	1.07	19.69	0.32	-0.54	9.77	3.09	1.72	18.59	-1.33	5.70

Figures 1, 2, and 3 show the structures of three of the complexes in the free and forced formations that depict the overall trends of the ten complexes analyzed. Figure 1 shows Adenine-Indole-Thymine (AIT). This complex was more stable in its free structure where it formed a T-shape in order to form hydrogen bonds. The interaction energy of the t-shaped AIT complex is more stable than the AIT sandwich complex by around -10kcal/mol. This implies that when an intercalant is placed between adenine and thymine the bases would want to adopt a hydrogen bonded structure in order to gain 10 kcal/mol of stability. AIG, CIC, CIA, GIG, and GIT all favor the hydrogen bonded structure implying that all of these structures would have a large distorting force acting on them.

**Figure 2: Comparison of free and forced structures for Adenine-Indole-Thymine (AIT).**

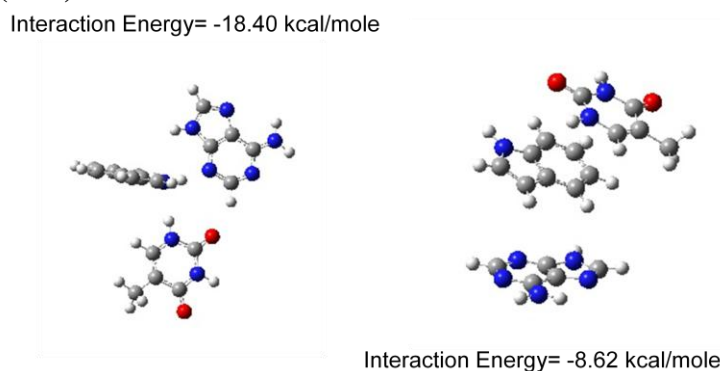
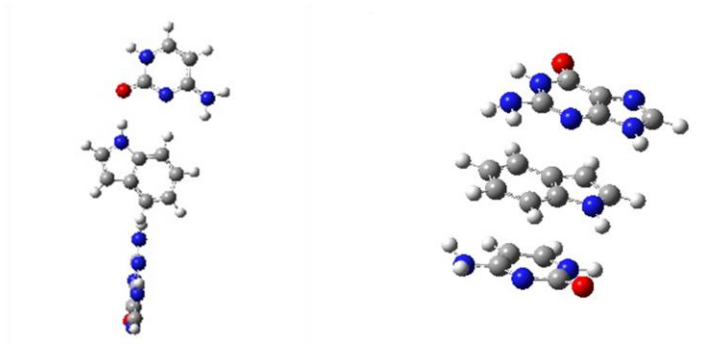


Figure 2 compares the two complex formations for Cytosine-Indole-Guanine (CIG). The CIG complex is more stable in the forced sandwich structure, yet in its free

structure it was not in a sandwich. The molecules in this complex opened up to allow for hydrogen bonds between the indole and nucleic acids. The sandwich structure is around -5 kcal/mol more stable than the free structure. This stabilization in the sandwich structure shows that an intercalant binding here would not lead to as great of distortion in the DNA as the above mentioned complexes. CIT and TIT also follow this pattern.

**Figure 3: Comparison of free and forced structures for Cytosine-Indole-Guanine (CIG).**

Interaction Energy= -15.78 kcal/mole

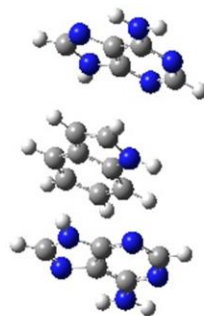
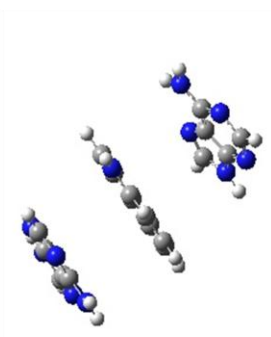


Interaction Energy= -20.20kcal/mole

The results for Adenine-Indole-Adenine, (AIA) was unique compared to the other structures and interaction energies calculated. AIA was the only complex to stay in a sandwich structure in its free structure. It also had a slightly tighter binding energy for its forced structure. Because AIA stays in a sandwich structure, it can be predicted that in between two adenines would not be a good binding site for a chemotherapeutic agent because it would not lead to significant disruption of DNA's vital structure.

**Figure 4: Comparison of the free and forced structures for Adenine-Indole-Adenine (AIA).**

Interaction Energy= -6.40 kcal/mole



Interaction Energy= -10.22 kcal/mole

### 4.3 Single-stranded ONIOM interactions

Interaction energies were calculated for the single-stranded complexes using the ONIOM method. The MP2 results show binding energies for the complexes of around -20 kcal/mol. TITG had the weakest binding energy of -15.78 kcal/mol. GITT had a very tight binding energy of -28.87 kcal/mol. These interaction energies are shown in Table 4. In general these values are a little larger than the values for the unattached complexes. These larger values indicate that the backbone accounts for a small amount of attraction. The complexes containing guanine tended to have larger binding energies than the complexes without guanine.

**Table 4: MP2 results for the single-stranded ONIOM counterpoise corrected interaction energies (kcal/mol)**

MP2:AM1	ACIC	GIGG	AIAA	GITT	CIAA	TITG	AITA	GGIA	GICC	TICC
Binding Energy	-22.40	-24.22	-22.60	-28.87	-22.81	-15.78	-20.04	-27.21	-22.70	-16.99

Table 5 depicts the interaction energies calculated for the single-stranded ONIOM structures using the DFT method, SVWN. These results predicted an overall weaker

binding compared to the MP2 results, yet the same general trends were seen in the data.

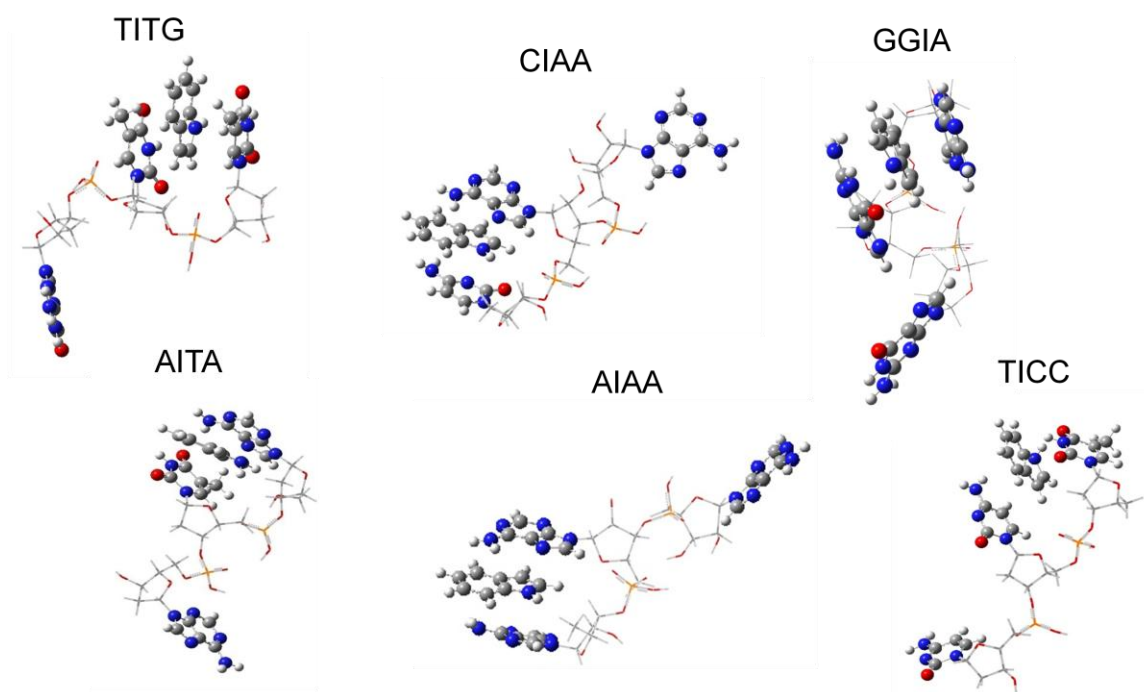
The SVWN results varied from the MP2 by 0.09 to 9.88 kcal/mol.

**Table 5: SVWN Results for the single-stranded ONIOM counterpoise corrected interaction energies (kcal/mol)**

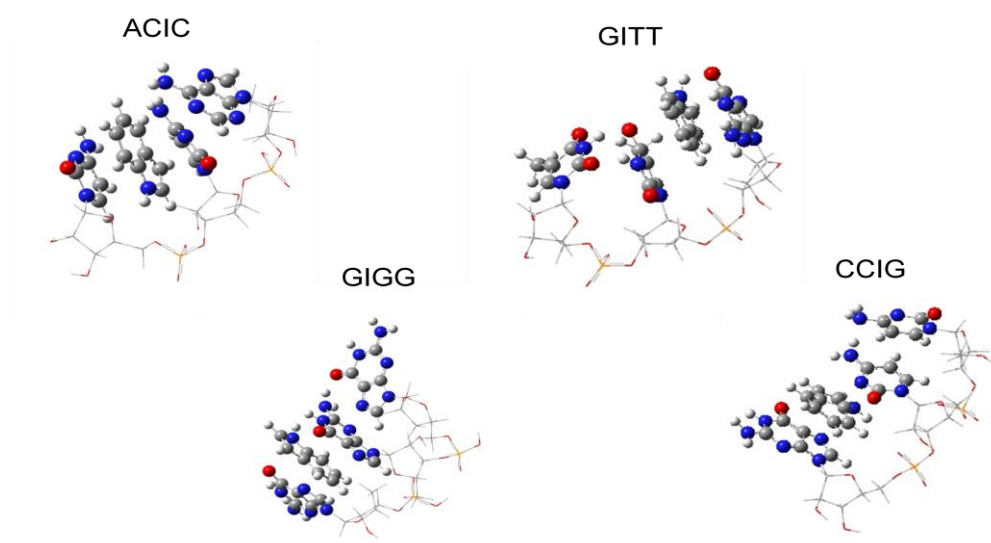
SVWN:AM1	ACICIE	GIGGIE	AIAA	GITTIE	CIAA	TITGIE	AITAIE	GGIAIE	GICCIE	TICCIE
Binding Energy	-12.51	-17.94	-16.79	-28.78	-18.82	-13.38	-16.45	-26.13	-15.49	-21.34

Single-stranded ONIOM complexes were optimized into sandwich structures with the intercalant in between and the third nucleic acid base separate from the three molecule complex, XIY. This ideal arrangement was accomplished with 6 of the 10 complexes, TITG, CIAA, GGIA, AITA, AIAA, and TICC. In these complexes both nucleic acid bases are able to use their entire electron density to react with the indole. These structures are depicted in Figure 4. For ACIC, GITT, GICC, and CCIG, the complexes were only stable in their optimization when the third nucleic acid base was present. For these complexes, the presence of the third nucleic acid base is involved in binding to one of the nucleic acid bases which is binding to indole. This binding of the third base thus reduces the electron density in that base available to interact with indole. This is shown in Figure 5.

**Figure 5: Single-stranded ONIOM interactions; two bases with indole**



**Figure 6: Single-stranded ONIOM structures; three bases with indole**



#### 4.4 Double-stranded ONIOM interactions

Interaction energies were calculated for the double-stranded ONIOM structures using MP2 for the bases and the indole and AM1 for the sugar and phosphate backbone. The intercalant binding energies were around -40 kcal/mol for the five complexes, ATIT, ATIGC, GCIGC, GCITA, and GGICC. These MP2 results are depicted in Table 6. The interaction energies for all five complexes are remarkably uniform around -40 kcal/mol. The presence of the full DNA has a leveling effect on the interaction of the intercalant. It is possible that the intercalant used is small in comparison to the pitch between the DNA bases, and this is what leads to the uniformity of the interaction energies. A larger intercalant may alter these uniform results.

**Table 6: Counterpoise corrected interaction energies using MP2 for double-stranded ONIOM structures**

MP2:AM1	ATITAie	ATIGCie	GCIGCie	GCITAie	GGICCie
DNA Binding Energy	-490.695	-66.173	-70.1979	-69.1068	-77.6739
Intercalant Binding Energy	-40.126	-39.0398	-39.776	-41.1533	-39.5022

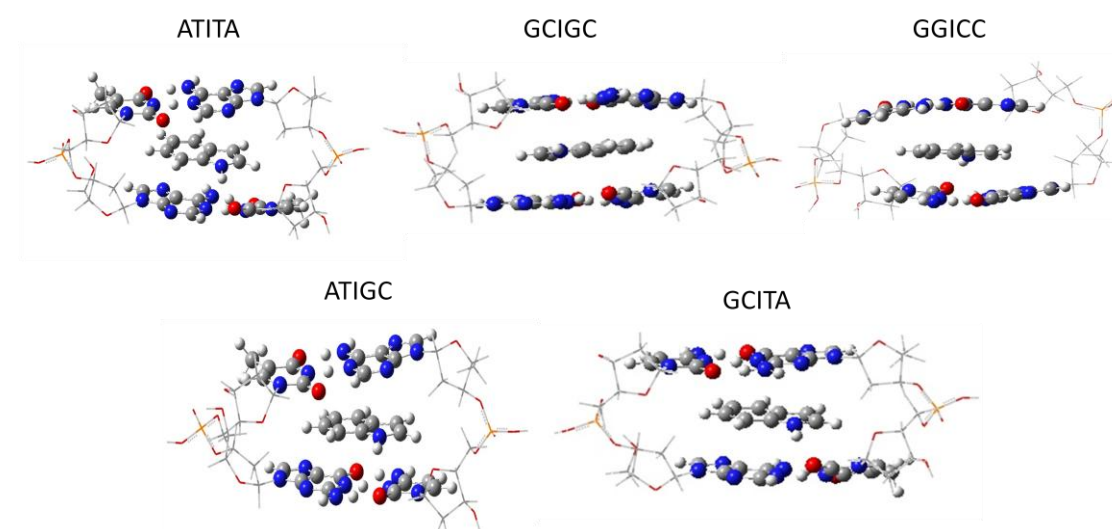
Interaction energies were calculated for the double-stranded ONIOM structures using SVWN for the bases and the indole and AM1 for the sugar and phosphate backbone. These values are much smaller than those collected for MP2, as they are around -23 kcal/mol. These SVWN results are depicted in Table 7. For the unattached bases and for the single-stranded DNA SVWN had been very similar to MP2. This method's failure in the double-stranded DNA is likely due to the fact that SVWN is a local DFT method and this very large complex has significant contributions from non local electron density.

**Table 7: Counterpoise corrected interaction energies using SVWN for double-stranded ONIOM structures**

SVWN:AM1	ATITAie	ATIGCie	GCIGCie	GCITAie	GGICCie
DNA Binding Energy	-518.217	-97.2801	-103.962	-99.0623	-110.87
Intercalant Binding Energy	-23.6135	-21.4741	-22.124	-23.9173	-24.0118

The structures for the double-stranded ONIOMs, AITA, GCIGC, GGICC, ATIGC, and GCITA are shown in Figure 6.

**Figure 7: Double-stranded ONIOM structures**



#### 4.5 ONIOM Comparison

The values for the single, double, and free structures are compared in Table 8. A ratio was calculated for the double-stranded ONIOMs to 2 times the value of the single-stranded. The ratio for ATITA was 1.00 depicting that the single-stranded calculation accurately predicted the double-stranded interaction energy. The other energies varied around 0.85, which shows a fairly accurate modeling of the double-stranded interaction with the single-stranded. This will allow for use of simpler models such as single-stranded or unattached complexes in place of double-stranded models.

**Table 8: Comparison of the single and double structure interaction energies (kcal/mol) using MP2/6-31+g\*.**

MP2:AM1	ATITAie	ATIGCie	GCIGCie	GCITAie	GGICCie
Double Strand ONIOM MP2	-40.13	-39.04	-39.78	-41.15	-39.50
2x Single Strand ONIOM MP2	-40.08	-44.21	-46.62	-51.67	-45.40
Ratio- DS:2xSS	1.00	0.88	0.85	0.80	0.87

## 5. Conclusions

The free structures, overall, had greater binding energies than the forced structures with the exception of AIA, CIG, CIT, and TIT. These results depict the nucleic acids naturally tendency to spread out and rearrange themselves in order to form hydrogen bonds. The AIA complex was a sandwich in both its free and forced structure, yet the forced had a stronger interaction energy. The AIA complex stays in a sandwich structure no matter what optimization is performed, thus it can be predicted that this would not be an optimal binding site for a chemotherapeutic agent. If a chemotherapeutic agent was to bind in this location, it is predicted that the structure of the DNA would not be greatly altered.

The SVWN results are qualitatively similar to the MP2 results, yet SVWN tends to overestimate the interaction energies. The B3LYP results tend to underestimate the interaction energies and are in some cases qualitatively incorrect. There was not a dramatic difference between the interaction energies using the two different basis sets, 6-31+g\* and 6-311+g\*. The same trends were observed for the MP2 results. The MP2 values between the two basis sets normally differed by less than 1 kcal/mol. The accuracy of the smaller basis set allowed 6-31+g\* to be used later for ONIOM calculations.

The interaction energy results for the single-stranded ONIOM calculations were around -20 kcal/mol for all complexes, as the backbone increased the overall binding energies and made the energies more consistent across the series. For the double-stranded ONIOM calculations, the interaction energies were roughly the sum of the two pairs. A ratio of the interaction energies for the double-stranded complexes to twice the single-stranded complexes was calculated. This ratio portrays the manner in which the single-stranded interactions were able to accurately predict the values for the double-stranded ONIOM calculations.

These results correlate with other experimental and theoretical results for  $\Delta G_{\text{mol}}$ . Previous calculations on other intercalants bound between two bases show values of between -15 to -30 kcal/mol (Lin *et al.*, 14346-14354). This research mirrors the work of El Gogary *et al.* who also found values of around -20 kcal/mol for complexes consisting of two nucleic acids and an intercalant. Finally, Ren *et al.* report experimental values for the  $\Delta G_{\text{mol}}$  of around -15 kcal/mol. These values are derived from calorimetry measurements and may in fact be an underestimation of the actual value. In the future calculations will be done with other intercalants larger than indole to see if these computational results continue to mirror those obtained through other methods of research.

## 6. Works Cited

1. Jemal A, Siegel R, Xu J, Ward E. *CA Cancer J Clin* **60**, 277-300 (2010).
2. Weinberg RA. *Scientific American*. 62-70 (1996).
3. Palchaudhuri and Hergenrother, *Current Opinion in Biotechnology* **18**, 497 (2007).
4. Haq, I. *Archives of Biochemistry and Biophysics*. **403**, 1 (2002).
5. Long and Barton. *Accounts of Chemical Research*, **90**, 271-273 (1990).
6. El-Gogary and Koehler, *THEOCHEM* **895**, 57 (2009).
7. Ren, J., Jenkins, T., and Chaires, J. *Biochemistry* **39**, 8439-8447 (2000).
8. Riley, K., and Hobza, P. *WIRES Computational Molecular Science*. **1**, 1-17 (2011).
9. Lin, I., Lilienfeld, O., Coutinho-neto, M., Tavernelli, I., and Rothlisberger, U. *Journal of Physical Chemistry B*. **111**, 14346-14354 (2007).
10. Van Sickle, Shroyer, M., and Cafiero, M, *THEOCHEM* **941**, 78 (2010).
11. Cerny, J., Kabelac, P., and Hobza, P. *Journal of the American Chemistry Society*. **130**, 16055 (2008).
12. Hofto, L.R., Van Sickle, K., Cafiero, M., *Int. J. Quantum Chem.*, **107**, 112 (2007).
13. Zhao, Y., Truhlar D., *Journal Chemical Theory and Computation*. **3**, 289-300 (2007).
14. Utkov, H., Price, A., and Cafiero, M. *Computational and Theoretical Chemistry*. In Press.
15. Goto-Ito et al. *NAT.STRUCT.MOL.BIOL.* (2009).
16. PQS version 3.1, Parallel Quantum Solutions, 2013 Green Acres Road, Fayetteville, Arkansas 72703.
17. Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.