A model for methylation of cytosine: a potential marker for cancer

Written at Rhodes College

Zoë Catherine Clark

Department of Chemistry

Rhodes College

Memphis, Tennessee

2012

Submitted in partial fulfillment of the requirements for the

Bachelor of Science degree with Honors in Chemistry

This Honors paper by Zoë Clark has been read

and approved for Honors in Chemistry.

Dr. Mauricio Cafiero

Project Advisor

_____

Dr. Larryn Peterson

Second Reader

_____

Dr. Carolyn Jaslow

Extra-Departmental Reader

_____

Dr. Darlene Loprete

Department Chair

_____

CONTENTS

ABSTRACT

The nucleic acid bases in DNA rely on hydrogen-bonding to maintain the structure that defines their function. When a hydrogen atom of a nucleic acid base is substituted by a methyl group, the resulting change in sterics disrupts intrahelical and interstrand bonding, thus altering DNA structure and function. Substitutions on cytosine, as in the case of 5-methylcytosine and hydroxymethylcytosine, are the only known modified nucleic acid bases that exist in eukaryotes, and both interfere with transcription of DNA. Methylation of benzene and cytosine was modeled with and without an additional aromatic molecule in a sandwiched conformation using HCTH/6-31+g*. This additional pi-stacked molecule models both intercalation and the natural DNA environment. The methylation reaction was studied via two mechanisms; a Friedel-Crafts alkylation and the mechanism used in the methyltransferase enzyme. Structures and energies were calculated for reactants, transition states, and products. Results showed that the presence of the stacked aromatic ring increased the enthalpy of reaction and lowered the activation energy.

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

Signed_____

Zoë Clark

Date _____

1. Introduction

1.1 Overview

Early detection is one of the best defenses against cancer. The work of Korshunova, *et al.* suggests a very simple and noninvasive test that samples human serum for the early signs of tumorigenesis (Korshunova, *et al.*, 2007, 23). The sample would be tested for patterns of methylation--a modification of DNA wherein a methyl group is substituted for a hydrogen atom of a nucleic acid base--to indicate the malfunctioning cellular machinery that is typical of cancer. In their work, Korshunova, *et al.* searched for methylation patterns in patients with and without cancer, hoping to find a definitive marker among each group. However, methylation occurs naturally at every locus in healthy nucleobases and little is known about these methylation patterns (Korshunova, *et al.*, 2007, 23). One piece of information that is clear is that DNA methylation patterns change in a patient with cancer; what was methylated becomes demethylated, methylation-free areas are modified, and there is greater variation in methylation patterns (Rodríguez López, *et al.*, 2010, 9103). As of yet, no definite patterns are known to mark the DNA of healthy patients or those with cancer (Korshunova, *et al.*, 2007, 27). This vein of early detection is being explored and Cramer, *et al.*, have found the biomarkers cancer antigen 125 (CA125) and human epididymis protein 4 (HE4) that are being used to monitor the progression of ovarian cancer (Cramer, *et al.*, 2012, 365). It is not yet useful in indicating whether or not the disease is present.

Stepping down to the molecular level, a malfunctioning cell that is typical of disease, such as cancer, can include improper or malfunctioning proteins that were

translated and transcribed from a misreading of the DNA backbone (Rodríguez López, *et al.*, 2010, 9100). Substitutions on cytosine, as in the case of 5-methylcytosine and hydroxymethylcytosine, are the only known modified nucleic acid bases that exist in eukaryotes, and both have been shown to interfere with the reading of the DNA backbone (Everts, 2011, 40; Colot and Rossignol, 1999, 406). Additional evidence of the importance of modified nucleic acid bases in the creation of cancerous cells lies in the discovery of N^6-methyladenine and N^4-methylcytosine in bacterial, plant, and protist DNA. N^6-Methyladenine regulates virulence and controls the DNA functions of replication, repair, expression, and transposition (Ratel, *et al.*, 2006, 309). Misreading of the DNA backbone that ultimately leads to the translation of malfunctioning proteins occurs because of altered sterics that change the structure of the DNA helix, and this change in sterics is caused by something as simple as the replacement of a hydrogen atom of a nucleobase with a methyl group, a process known as methylation (Rodríguez López, *et al.*, 2010, 9100).

The goal of this research has been to contribute to the body of knowledge surrounding the formation of methylcytosine in hopes of ultimately adding to the discovery of a pattern to indicate tumorigenesis or healthy DNA. Specifically, the enthalpy of reaction and activation energy of the reaction with and without a generalized pi-stacked ring were studied to better understand how readily this modification occurs. Finally, this mechanism was compared to that used by the dengue virus methyltransferase enzyme in order to understand how this kind of substitution occurs *in vivo*.

1.2 Methylation

Methylation is widespread among many forms of life and can vary greatly between these organisms (Colot and Rossignol, 1999, 402). Patterns of methylation are established during embryogenesis and are later remodeled in differentiating cells (Brenet, *et al.*, 2011, 1). It functions in regulation of DNA expression through activities such as inhibiting transcription initiation and acting as an imprinting signal (Colot and Rossignol, 1999, 408). Methylation controls gene expression by altering the structure of the DNA within the binding site of the transcription protein machinery (Hodges-Garcia and Hagerman, 1992, 7595). Methylation also functions as an imprinting signal by wrapping marked DNA tightly around the cores of histones and forming transcriptionally inactive heterochromatin (Lee and Lee, 2011, 173). This modification typically takes place in locations with CpG (cytosine-phospho-guanosine dinucleotide) islands where the cytosine will be methylated to 5-methylcytosine (Lee and Lee, 2011, 173 and Brenet, *et al.*, 2011, 1).

Abnormal methylation patterns and hypermethylation can lead to serious disease due to the silencing of important tumor suppressor genes or by silencing genes that are necessary for healthy development (Bird, 1989, 209; Esteller, 2007, 286; and Robertson, 2002, 5361). This abnormal methylation, termed an increase in methylation entropy by Xie, *et al.*, is characteristic of tumor cells, which have been found to have greater methylation entropy throughout the entire genome (Xie, *et al.*, 2011, 4099). Furthermore, Xie, *et al.* has shown that this greater variety in methylation suggests that tumor development and progression are frequently found alongside a disorder in the machinery

that is responsible for the establishment of the initial DNA methylation patterns, and may be an early indicator of serious disease (Xie, *et al.*, 2011, 4107).

Methyltransferase enzymes methylate nucleobases *in vivo*. There are two classes of this type of enzyme: one that methylates bases that have not been methylated before (*de novo* DNA methyltransferase) and one that maintains methylation of bases after replication (maintenance DNA methyltransferase) (Pradhan and Esteve, 2003, 6). The methyltransferase enzyme works by flipping out the base to be methylated, and then catalyzing a methyl group transfer from S-adenosyl-L-methionine to cytosine bases in DNA (Pradhan and Esteve, 2003, 6). S-Adenosyl-L-methionine will be converted to S-adenosyl-L-homocysteine in the process (Yap, *et al.*, 2010, 1). The methyl group's presence on the nucleobase does not change the nucleotide sequence, but it still marks the base and confers epigenetic information to the DNA sequence (Pradhan and Esteve, 2003, 6). This information is important in determining the health of a patient because it indicates whether or not genes, such as important tumor suppressors, are silenced.

1.3 Quantum chemistry

In biological systems, such as the one modeled in this work, noncovalent interactions, such as hydrogen bonds and induction/dispersion forces, are key to protein folding and nucleobase stacking and arrangement (Zhao and Truhlar, 2007, 289). An accurate prediction of these conformations is important in the assessment of overall DNA, RNA, and protein structure, and these noncovalent interactions need to be carefully and accurately modeled. An accurate picture of these structures is vital in assessing the function of the biological systems that are modeled because structure and

function are closely intertwined (Riley and Hobza, 2011, 3). Modeling noncovalent interactions is difficult, however, due to managing the compromise between affordability and accuracy (Zhao and Truhlar, 2007, 289).

Accuracy in computational chemistry is important because the bond energies that are modeled are small, and a relatively small margin of error can still result in large miscalculations of the energy and furthermore, the development of inaccurate models. For comparison, covalent interactions have bond energies on the order of 100 kcal/mol, ionic bonds have binding energies around 50 kcal/mol, hydrogen bonds exhibit the range of 5 to 20 kcal/mol interaction energies, and induction/dispersion forces have the smallest energies at 5 kcal/mol (Utkov, *et al.*, 2010, 98). Covalent bonds may be modeled by molecular mechanics, semi-empricial, and the Hartree-Fock *ab initio* methods with good accuracy (Utkov, *et al.*, 2010, 98). However, in this work induction/dispersion forces are responsible for the structures and configurations of the molecules. These systems require high accuracy for their results to be significant because induction/dispersion forces are so small. Density Functional Theory (DFT) is an *ab initio* method that can, in principle, model structures and energetics with relatively high accuracy, including all electron correlation effects like induction and dispersion (Utkov, *et al.*, 2010, 98). DFT is also relatively inexpensive.

The system modeled in this research is dominated by induction and dispersion forces being that it is a DNA and nucleobase system (Riley and Hobza, 2011, 4). HCTH (a DFT method due to Hamprecht, Cohen, Tozer, and Handy), is the method of choice given that it is successful in accurately predicting the structures and energies of systems governed by induction and dispersion forces (Hamprecht *et al.*, 1998, 6264). HCTH is

also of particular use due to its ability to accurately model transition states and organometallics (Rappe and Li, 2003, 11188).

2. Methods

According to Bruice, Friedel-Crafts alkylation substitutes an alkyl group for a hydrogen atom on an aromatic ring (Bruice, 2007, 661). In the first step of the reaction, an electrophile is formed from the reaction of an alkyl halide (CH_3F) with AlF_3. This produces a carbocation ($+CH_3$) that will be attacked by the nucleophilic pi bond of the aromatic ring, resulting in a bond between the electrophile and the ring. To restore its lost aromaticity, a proton will be abstracted by a base, resulting in methylbenzene and hydrogen fluoride (Bruice, 2007, 661-662).

A Friedel-Crafts alkylation reaction was used to model the attachment of a methyl group to benzene and cytosine. Initially, benzene was used to model the methylation of cytosine, and then cytosine was substituted once the optimized benzene structures were available. The calculations were repeated for both benzene and cytosine with an additional pi-stacked ring (Figure 3) to simulate the natural environment of DNA. The enthalpy of reaction and activation energy were noted for each calculation to gather information about the substitution. The enthalpy of reaction was used to determine the spontaneity of the reaction. The Gibbs free energy has contributions from the enthalpy and the entropy, and since the entropy of the systems studied does not change from system to system, then the only variable with useful information was the enthalpy of reaction. This value was found by subtracting the energy of the reactants from the energy of the products and the activation energy was found by subtracting the energy of the

reactants from the energy of the transition state. The activation energy was measured as an assessment of how easily the reaction takes place, as the activation energy is proportional to the rate constant.

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta H_{reaction} = E_{products} - E_{reactants}$$

$$E_A = E_{transition} - E_{reactants}$$

Structures and energies from the models were compared to the conformation of the nucleobase and pi-stacked ring in the crystal structure of the dengue virus methyltransferase enzyme (PDB ID: 2XBM). In this enzyme, which was chosen for use because its crystal structure is accessible, the nucleobase bends outward to be methylated by S-adenosyl-*L*-methionine (SAM), which will then convert to S-adenosyl-*L*-homocysteine (SAH). Therefore, in this model, the angle between the nucleobase and the pi-stacked ring was manipulated to find where the energy of the system was at a minimum. Finally, to model the action of the enzyme, the pi-stacked ring was replaced with guanine, and calculations of the enthalpy of reaction and activation energy were repeated. In cases where cytosine was the methylated molecule, corrections were performed to take into account the interactions between different parts of the molecule that were not present in the methylation of benzene.

The method used for all calculations is a Density Functional Theory (DFT) method called HCTH (Boese, *et al.*, 2000, and Boese and Handy, 2001). DFT is an electronic structure method that can be parameterized to model different types of chemical environments well. HCTH has been shown to have good accuracy for

organometallics and transition states. Rappe and Li (2003, 11188) have shown that HCTH can model transition states of reactions between organic molecules and metals bound to halogens in good agreements with experimental results. The reaction modeled in this work occurs between an organic molecule and a metal bound to a halogen, so HCTH should be a good method to get accurate results. A medium-sized estimate for the quality of the molecular orbitals, 6-31+g*, was used (Rappe and Li, 2003, 11188). This basis set, 6-31+g*, provides a mathematically satisfactory description of the shapes of the molecular orbitals, which is a compromise between accuracy and time considerations (a more accurate description would take much more time to calculate).

It should be noted that many of the calculations could not initially be run with 6-31+g* as the basis set. Although all of the reported values are determined using this basis set, pre-optimization cycles had to be completed with smaller basis sets to better approximate the values of the large basis set.

3. Results and Discussion

3.1 Benzene

Initial work focused on modeling the Friedel-Crafts alkylation of cytosine. However, using cytosine as a reactant proved challenging because it was difficult to get the incoming methyl group to interact with the appropriate carbon atom, C5, of the cytosine molecule instead of the more negatively charged nitrogen atoms. Therefore, initial modeling work was completed using benzene as a reactant instead of cytosine. Benzene shares pi electrons evenly over the molecule and makes it easier to attach a methyl group in the appropriate location (Figure 1). Friedel-Crafts alkylation was used to

model this reaction. The enthalpy of reaction for this first step using the aforementioned methods was -13.512 kcal/mol.

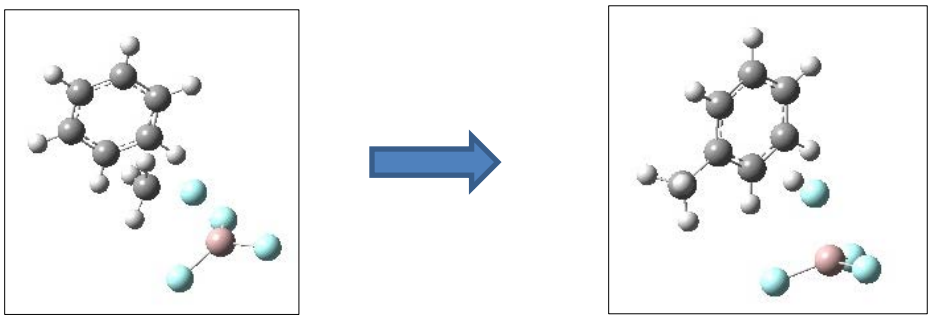

Figure 1. Diagram showing reactants (C_6H_6, CH_3F; left) and products (methylbenzene, HF; right) for the methylation of benzene using Friedel-Crafts alkylation and AlF_3 as a catalyst.

The calculation of the transition state is another important step in modeling the methylation of cytosine because it indicates the energy of activation of the reaction. Using benzene as the neutral starting molecule, the structure was found where a proton is removed and the methyl group begins to bond, but hydrogen fluoride has not yet formed since all four fluorine atoms are bonded to aluminum (Figure 2). The transition state structure was modeled to discover the amount of energy required to initiate the reaction; in this case 7.93 kcal/mol.

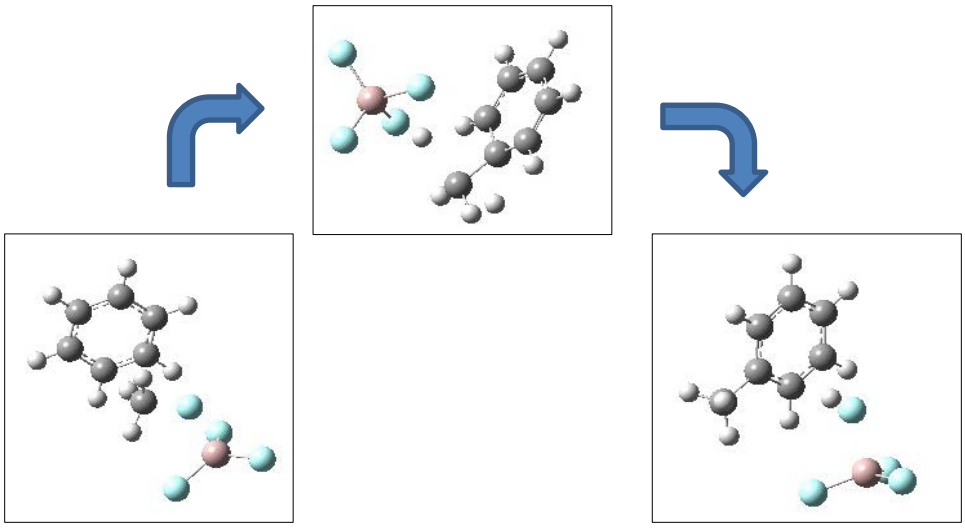

Figure 2. Diagram showing reactants (C_6H_6, CH_3F; left), transition state (top), and products (methylbenzene, HF; right) for the methylation of benzene using Friedel-Crafts alkylation and AlF_3 as a catalyst.

To simulate the natural pi-stacked environment of the nucleic acid bases, the first two calculations were repeated with one additional pi-stacked aromatic ring (Figure 3, Figure 4). Not only does this calculation model the natural environment of the DNA helix (Yap, *et al.*, 2010, 12836), but it also mimics the activity of carcinogens. Some carcinogens are intercalants, flat aromatic rings for which benzene is a prototypical form, that disrupt the shape of DNA (Leng, Chaires, and Waring; 2003, 6191). Therefore, this calculation is important to see if a carcinogen will speed up the reaction, increase methylation, and possibly mark the early signs of cancer. The enthalpy of the methylation reaction with an additional stacked benzene ring that mimics the activity of a generalized pi-stacked ring was -14.32 kcal/mol and therefore indicates a more spontaneous reaction. An additional pi-stacked aromatic ring was also added to the calculation of the activation energy of the transition state. The transition state of the pi-stacked system is similar to that of the first system, where a proton has been removed from benzene and the methyl group has attached in its place, but hydrogen fluoride has not formed since all four

fluorine atoms are attached to aluminum (Figure 4). From this calculation, the activation energy decreased to 4.18 kcal/mol, which is roughly half of the energy of the system without pi-stacking.

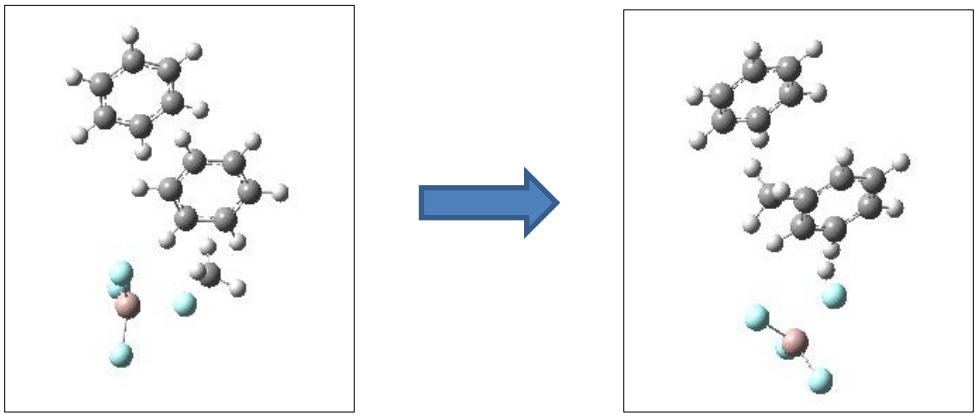

Figure 3. Diagram showing reactants (C_6H_6, CH_3F; left) and products (methylbenzene, HF; right) for the methylation of benzene using Friedel-Crafts alkylation and AlF_3 as a catalyst. Included is an additional pi-stacked aromatic ring.

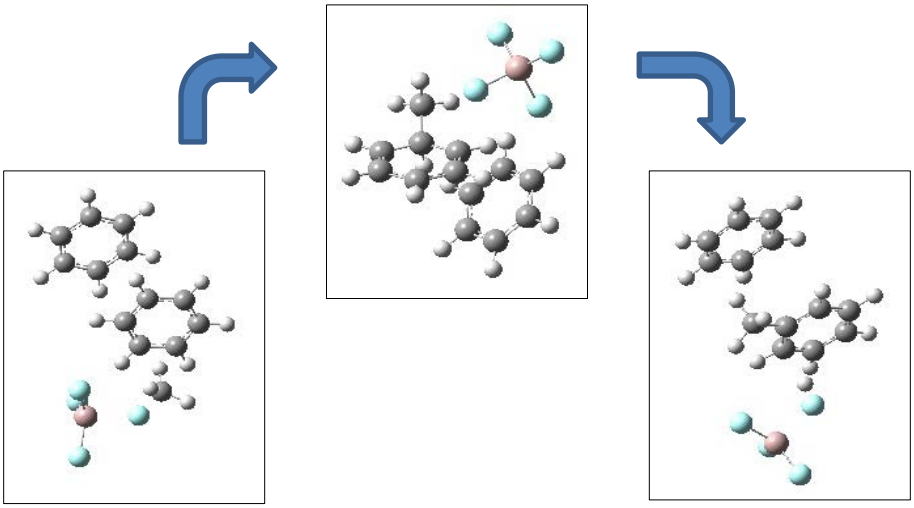

Figure 4. Diagram showing reactants (C_6H_6, CH_3F; left), transition state (top), and products (methylbenzene, HF; right) for the methylation of benzene using Friedel-Crafts alkylation and AlF_3 as a catalyst. Included is an additional pi-stacked aromatic ring.

The enthalpy of reaction and activation energy of the methylation of benzene with and without pi-stacking show that the reaction is both more spontaneous and occurs more easily in the presence of an additional pi-stacked ring (Table 1). This is encouraging because it indicates that in an imitation of the natural environment of DNA and carcinogen activity the reaction is more likely to occur. Cytosine will replace benzene in these optimized geometries to facilitate modeling of methylation of the nucleobase using Friedel-Crafts alkylation.

Table 1. Enthalpy of reaction and activation energy (both in kcal/mol) for the Friedel-Crafts alkylation of benzene using CH_3F with and without an additional pi-stacked aromatic ring. Calculations are performed using HCTH/6-31+g* and the values are single values.

	Enthalpy of Reaction	Activation Energy
Benzene	-13.51 kcal/mol	7.93 kcal/mol
Benzene and pi-stacking	-14.32 kcal/mol	4.18 kcal/mol

3.2 Cytosine

Once the methylation of benzene was completely optimized, cytosine was substituted into the benzene geometries. The calculation was run once to appropriately fit cytosine in to benzene's previous location, but the calculation was not fully optimized. In the transition state depicted below, the methyl group is attached to the five position on the cytosine molecule and four fluorine atoms are near enough to be considered bonded to aluminum, although the bonds drawn show that one fluorine is bonded to the methyl group and leaving to bond to aluminum (Figure 5). The proton is still bonded to cytosine. The enthalpy of reaction was found to be -6.34 kcal/mol and the activation energy was

83.82 kcal/mol for cytosine in the benzene structures (Table 3). Compared to the previous values, neither of these energies is indicative of a spontaneous and easy reaction, which was not expected because the methylation of cytosine takes place so frequently in eukaryotic DNA. However, this step was necessary to bridge the optimized benzene structures to optimized cytosine structures. Even if the values for enthalpy of reaction and activation energy are not as low as expected, there are at least values to report and that was not the case with the initial attempts to methylate cytosine.

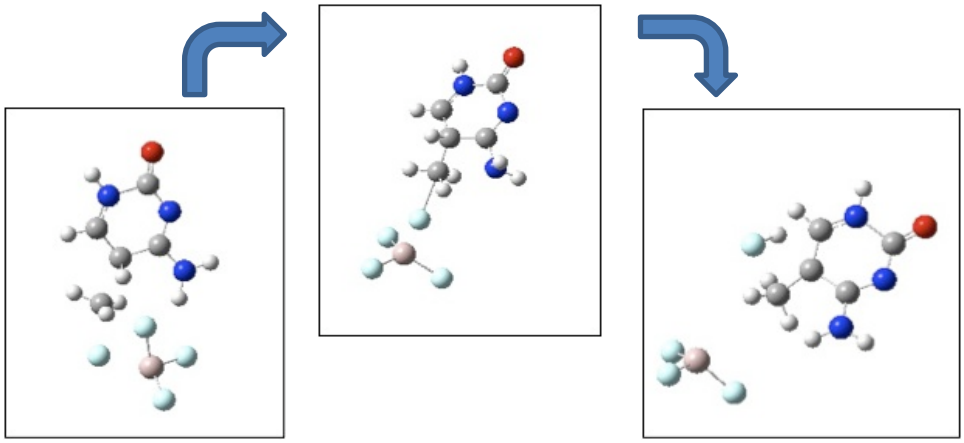

Figure 5. A depiction of the methylation of cytosine showing the reactants (cytosine, CH_3F; left), transition state (top), and products (methylcytosine, HF; right). Friedel-Crafts alkylation, with AlF_3 as a catalyst, was used to attach the methyl group.

The calculations of the enthalpy of reaction and activation energy that were completed for benzene were repeated with cytosine (Figure 6, Figure 7). These geometries were optimized from the original benzene structures in which benzene had been replaced by cytosine. In this way, the reaction is easier to model, and the initial difficulty of getting the methyl group to attach at the appropriate site was avoided. From the calculation depicted in Figure 6, the enthalpy of reaction of -14.63 kcal/mol was found, and the diagram of the transition state (Figure 7) in which the methyl group is

moving to bond with cytosine, aluminum has all four fluorine atoms attached, and cytosine retains its proton, the activation energy was discovered to be 7.76 kcal/mol. These values are very similar to the values that were found for the methylation of benzene. However, before these numbers were found to be so similar additional interaction energies had to be taken into account. Each structure was corrected for interaction energies that did not exist in the system when the benzene model was used. In the methylation of benzene, no atom of the aromatic ring pulls the methyl group toward itself with stronger electronegativity. That is just the opposite in the case of the methylation of cytosine. Here, extraneous interaction energies were subtracted from the overall enthalpy of reaction and activation energy. Specifically, the hydrogen bond between nitrogen on cytosine and hydrogen fluoride was subtracted from the system. Interactions energies for hydrogen fluoride and fluorine with aluminum trifluoride were also removed from the calculated energies of the products.

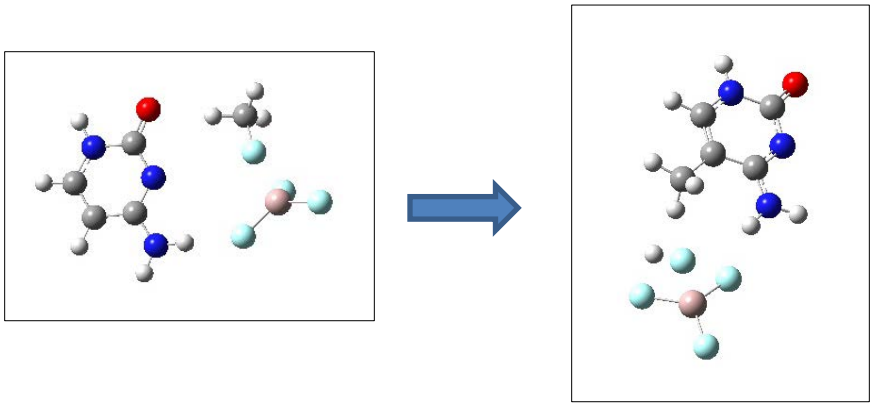

Figure 6. This image shows the methylation of cytosine from reactants (cytosine, CH_3F; left) to products (methylcytosine, HF; right) using Friedel-Crafts alkylation and AlF_3 as a catalyst. Notice that in the reactants the methyl group is on an entirely different side of the cytosine than where it will eventually bond in the products. This was the initial difficulty in methylating cytosine because the methyl group was near the wrong locus.

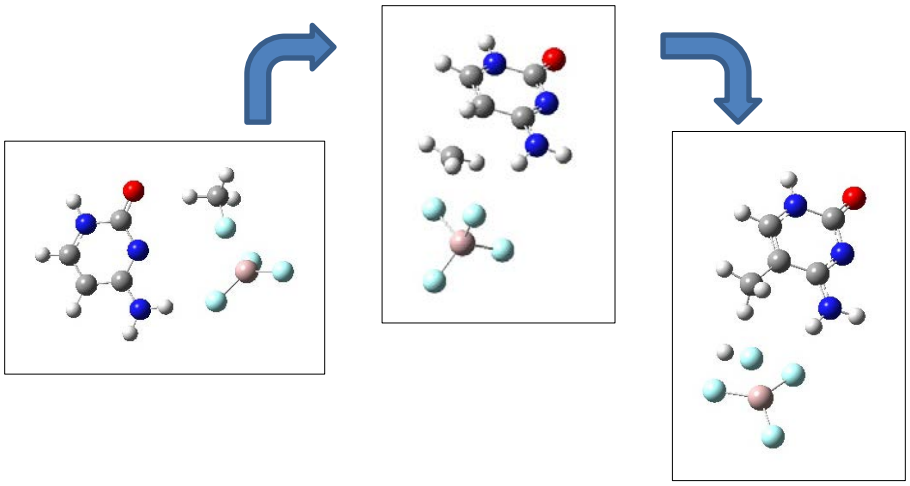

Figure 7. This diagram shows the methylation of cytosine, including reactants (cytosine, CH_3F; left), transition state (top), and products (methylcytosine, HF; right) using Friedel-Crafts alkylation and AlF_3 as a catalyst. Again, notice the movement of methylfluoride as it moves from the more electronegative side of cytosine toward the 5 locus where it will attach.

As in the case of the methylation of benzene, the calculations for the enthalpy of reaction and activation energy were repeated with cytosine with the addition of a pi-stacked aromatic ring that mimics the activity of generalized pi-stacking (Figure 8, Figure 9). Just as with the methylation of cytosine without pi-stacking, the hydrogen bond energies of hydrogen fluoride with nitrogen and fluoride with aluminum trifluoride were subtracted from the enthalpy of reaction and activation energy for cytosine with an extra aromatic ring. An enthalpy of reaction of -15.27 kcal/mol was found in the presence of an additional pi-stacked aromatic ring (Figure 8). The activation energy of the transition state in which cytosine has formed a bond with the methyl group, the proton remains attached to the nucleobase, and aluminum is bonded to all four fluorine atoms was 5.35 kcal/mol (Figure 9). The trends of decreased enthalpy of reaction and activation energy are consistent between cytosine and benzene. This indicates that the reaction becomes more spontaneous and occurs more easily in the presence of a pi-stacked ring in both the

benzene and cytosine systems, which is encouraging because that is thought to be an adequate simulation of the natural DNA environment (Table 2).

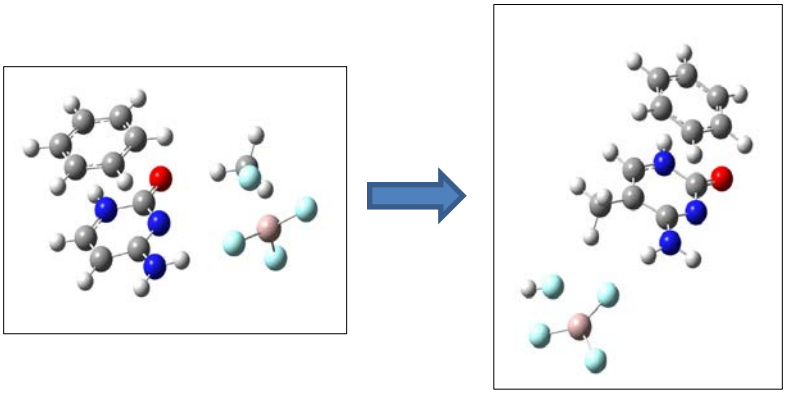

Figure 8. This image depicts the methylation of cytosine from reactants (cytosine CH_3F; left) to products (methylcytosine, HF; right) using Friedel-Crafts alkylation with AlF_3 as the catalyst. This image also shows the reaction with an additional pi-stacked aromatic ring that models the structure of DNA.

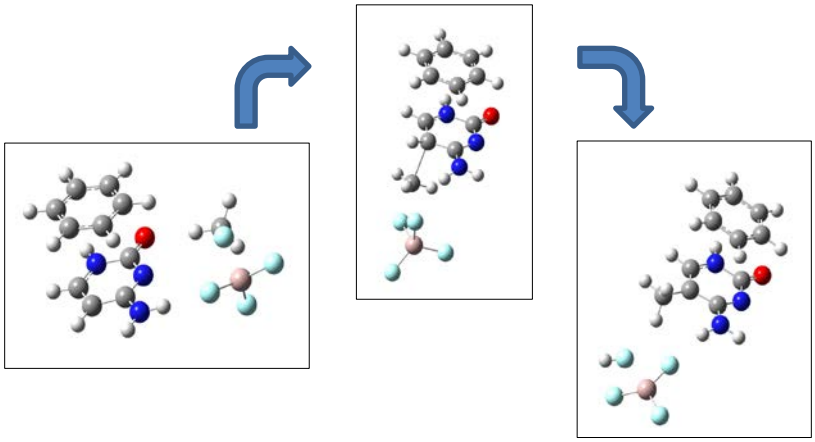

Figure 9. Here, the methylation of cytosine is shown with reactants (cytosine, CH_3F; left), transition state (top), and products (methylcytosine, HF; right) using Friedel-Crafts alkylation with AlF_3 as the catalyst. This image also shows the reaction with an additional pi-stacked aromatic ring that models the structure of DNA.

Table 2. Enthalpy of reaction and activation energy (both in kcal/mol) for Friedel-Crafts alkylation of cytosine using CH_3F with and without an additional pi-stacked aromatic ring. Calculations are performed using HCTH/6-31+g*.

	Enthalpy of Reaction	Activation Energy
Cytosine	-14.63 kcal/mol	7.76 kcal/mol
Cytosine and pi-stacking	-15.27 kcal/mol	5.35 kcal/mol

Table 3 illustrates the reliability of our models for the methylation of both benzene and cytosine. The values for enthalpy of reaction and activation energy are very similar, and, although not visible on this particular table, they decrease by similar amounts from the non pi-stacked to the pi-stacked models. The evidence in the methylation calculations for benzene and cytosine and the comparisons shown in Table 3 indicate that the reaction becomes more spontaneous and occurs more easily in the presence of an additional pi-stacked aromatic ring in a hypothetical environment. Next, the activity of the dengue virus methyltransferase enzyme, the *in vivo* situation, was investigated and compared to the results of the Friedel-Crafts models to see what similarities and differences exist.

Table 3. Comparison of the enthalpy of reaction and activation energy (both in kcal/mol) for Friedel-Crafts alkylation of benzene uaing CH_3F and of cytosine with CH_3F.

	Enthalpy of Reaction	Activation Energy
Benzene	-13.51 kcal/mol	7.93 kcal/mol
Benzene and pi-stacking	-14.32 kcal/mol	4.18 kcal/mol
Cytosine at benzene geometries	-6.34 kcal/mol	83.82 kcal/mol
Cytosine	-14.63 kcal/mol	7.76 kcal/mol
Cytosine and pi-stacking	-15.27 kcal/mol	5.35 kcal/mol

3.3 *In vivo* mechanism of methylation

The crystal structure for the dengue virus methyltransferase enzyme was solved by Yap, *et al.*, (2010). SAM (N-terminal S-adenosyl-*L*-methionine) methylates the nucleobase and is converted into SAH (S-adenosyl-*L*-homocysteine); however, in this structure the base is too far from SAM to be methylated (Yap, *et al.*, 2010, 1-5). Furthermore, these is no crystal structure showing the action of methylation, leaving Yap, *et al.* to assume that the nucleobase moves out of conformation in the backwards opening motion (Figure 10, right). Given this information from Yap, *et al*., a comparison of the angles of an additional pi-stacked aromatic ring with the nucleobase was investigateded to discover the optimal tilt angle of the pi-stacked ring with the nucleobase. To simulate the outward bending motion of the pi-stacked ring, the atom opposite the atom to be methylated was constrained at a distance of 3.5 Å from its counterpart atom in the nucleobase beneath the pi-stacked ring. The distance between the atom to be methylated

and its counterpart atom on the nucleobase was lengthened in increments of 0.1 Å until the lowest energy structure was found (Figure 11).

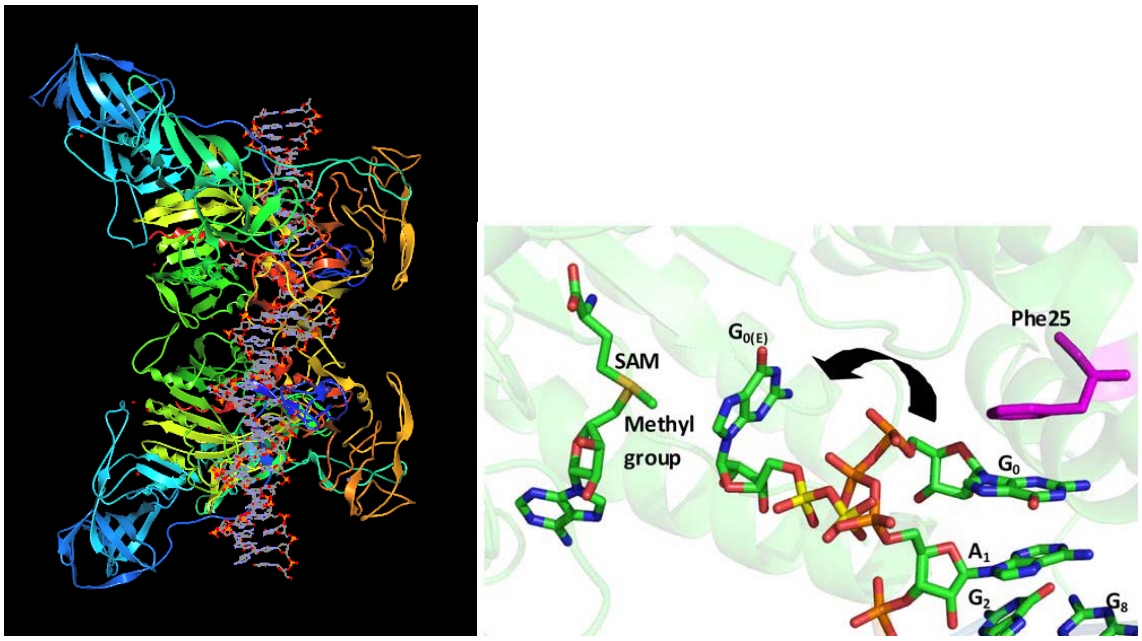

Figure 10. The methyltransferase enzyme complexed with a DNA helix (left), and the action that is thought to induce the methylation of cytosine, wherein the base flips out of conformation to reach SAM (N-terminal S-adenosyl-*L*-methionine), which will become SAH (S-adenosyl-*L*-homocysteine) once the methylation has taken place (right). Taken from Yap, *et al.*, 2010 (PDB ID: 2XBM).

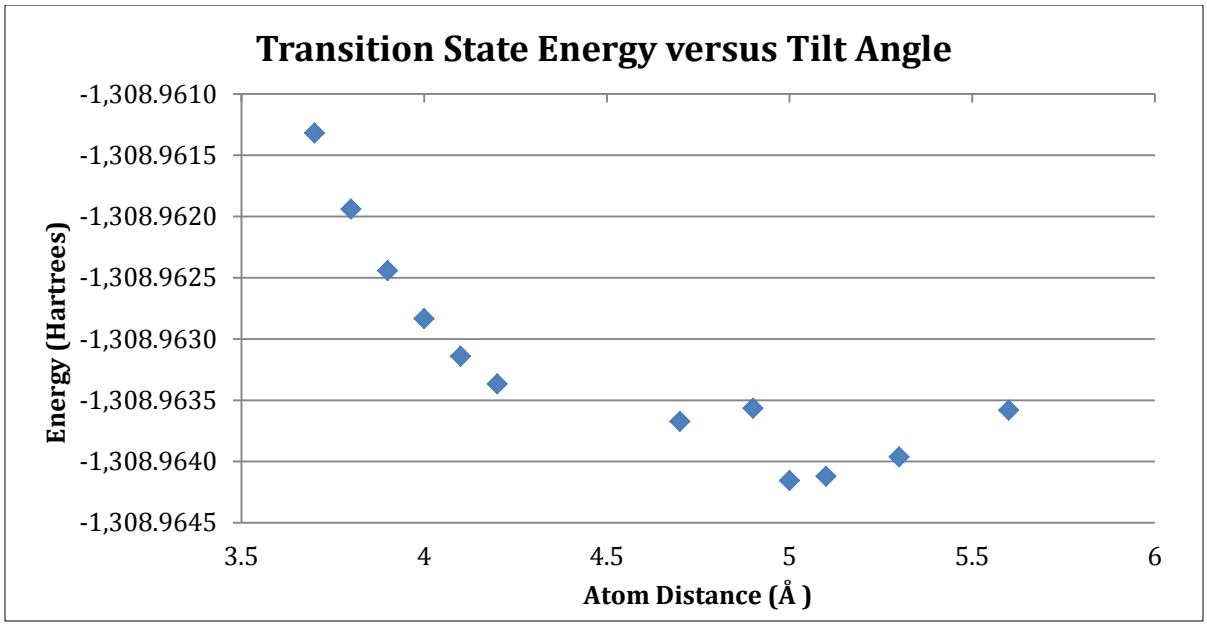

Figure 11. The optimal tilt angle of the pi-stacked aromatic ring above cytosine, the nucleobase being methylated. The distance between the atom being methylated and its complementary atom on the aromatic ring above is compared to the energy of the transition state. The distance between the atoms on the back of each molecule remained constant at 3.5 Å.

Of the distances explored, 5.0 Å between the locus of the methylation activity and the corresponding atom of the pi-stacked aromatic ring above proved to be the most comfortable conformation with the lowest energy (Figure 12). Since this value represents the minimum in Figure 11, it is possibly the conformation in which the nucleobase becomes methylated. Distances of 4.2 Å and 5.6 Å are also shown for comparison of the angle, although they both have higher energy (Figure 12). Without further information about the conformation of the methyltransferase enzyme or nucleobase during methylation, this comparison of distances to find the most suitable position of the bases as they receive methylation is all that has been found to show that a conformational change must take place so that DNA can be methylated most easily.

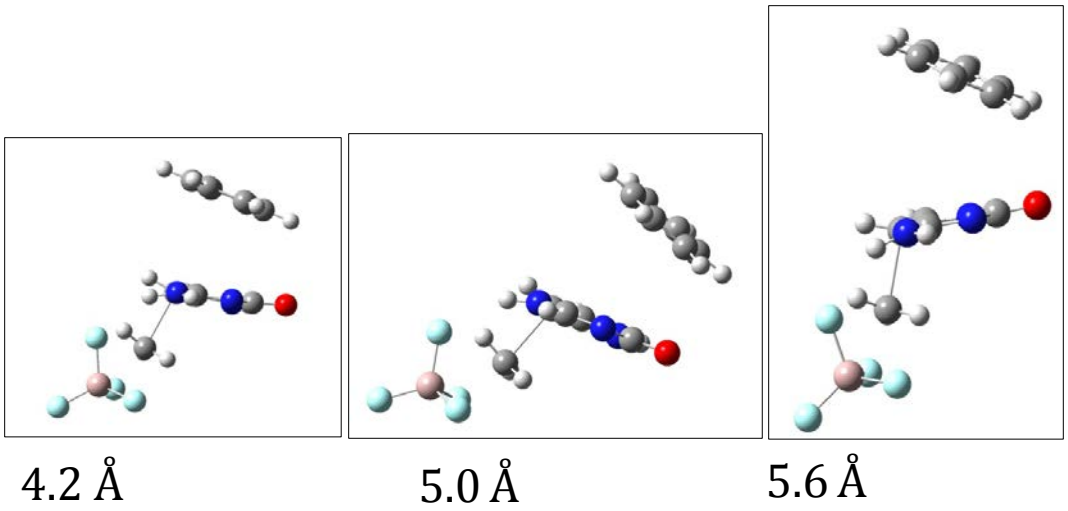

Figure 12. The change in distance between the methylated atom and its counterpart on the aromatic ring above. The images show a range of distances with different energies, 5.0 Å being the most comfortable conformation.

3.4 Full model

The final step of this project was to complete the enthalpy of reaction and activation energy calculations one more time, but using guanine as the pi-stacked ring instead of a generalized benzene ring. Methylation frequently takes place at CpG islands, so this model is an effort to get close to the in-vivo scenario (Yap, *et al.*, 2010). The cytosine-guanine product was the only calculation that ran to completion with the specified basis set (6-31+g*) (Figure 13). As outlined in the methods, smaller basis sets have been used to approximate the structure of the cytosine-guanine reaction. Once they were fully optimized, they were run again with increasingly larger basis sets until, finally, 6-31+g* was used. This is similar to the method that was used to approximate the methylated cytosine model. Initially, benzene was used to get close-to-comfortable structures for the molecules, and then cytosine was used after benzene indicated the correct conformation. The many difficulties with the calculations were caused the

changing electron density of the pi clouds around cytosine and guanine and prevented the optimization of the cytosine-guanine reactants and transition state. Therefore, there are no enthalpy of reaction or activation energy values to report.

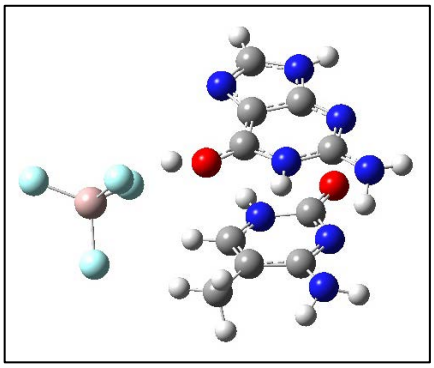

Figure 13. The optimized products of the Friedel-Crafts alkylation of cytosine with guanine as the pi-stacked ring above.

4. Conclusion

The methylation of cytosine to form 5-methylcytosine was modeled in this work. The calculations of the methylation of both benzene and cytosine had lower enthalpy of reaction and activation energy when the ring to be methylated was in complex with another pi-stacked aromatic ring. Since stepping toward an *in vivo* methylation model by including the pi-stacked aromatic rings decreased the activation energy and enthalpy of reaction, studying an actual *in vivo* situation with the dengue virus methyltransferase enzyme strengthened the conclusion that the modification takes places more easily when conditions more closely replicate the natural environment. Additionally, the energy necessary to transfer the methyl group from SAM to the nucleobase was at a minimum when the pi-stacked ring and the nucleobase were at an angle to one other, and that optimum angle was calculated and found to be one where the distance between the atom

to be methylated and its counterpart on the aromatic ring was 5 Å. Knowing these conditions, clarification of the mechanism of methylation should become easier. Elucidation of the mechanism will hopefully bring light to certain activities or conditions that may induce or increase methylation. In the absence of a cure for cancer, determination of the conditions that are potentially dangerous to the health of a patient are valuable. Even more valuable is the development of a screening process in which cancer may be discovered in the earliest stages, so that therapeutic methods may have a greater impact in the initial stages of the disease.

Literature Cited

1. Boese D., *et al*., *J Chem. Phys.* **112**, 1670 (2000).
2. Boese D. and N.C. Handy, *J. Chem. Phys.* **114,** 5497 (2001).
3. Brenet F., *et al*., "DNA methylation of the first exon is tightly linked to transcriptional silencing", *PLoS ONE*, **6**, e14524, 1-11 (2011).
4. Bird A., *Nature*, **321**, 209 (1986).
5. Bruice P.Y., Organic Chemistry, 5th ed., Pearson Education, Inc., 661-662 (2007).
6. Colot V. and J. L. Rossignol, "Eukaryotic DNA methylation as an evolutionary device", *Bioessays*, **21**, 402-410 (1999).
7. Esteller M., *Nat. Rev. Genet.*, **8**, 286 (2007).
8. Everts S., "New base on the block", *Science & Technology*, 40-41 (2011).
9. Hamprecht F.A., *et al*., *J. Chem. Phys.*, **109**, 6264 (1998).
10. Hodges-Garcia Y. and P.J. Hagerman, "Cytosine methylation can induce local distortions in the structure of duplex DNA", *Biochemistry*, **31**, 7595-7599 (1992).
11. Korshunova Y., *et al*., "Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA", *Genome Research*, **18**, 19-29 (2008).
12. Lee J.Y. and T.H. Lee, "Effects of DNA methylation of the structure of nucleosomes", *J. Am. Chem. Soc.*, **134**, 173-175 (2011).
13. Leng F., J.B. Chaires, and M. J. Waring, "Energetics of echinomycin binding to DNA", *Nucleic Acids Research*, **31**, (6191).
14. Pradhan S. and P.O. Esteve, "Mammalian DNA (cytosine-5) methyltransferases and their expression", *Clinical Immunology*, **109**, 6-16 (2003).
15. Rappe A.K. and S. Li, "Mechanism of chromyl chloride epoxidation", *J. Am. Chem. Soc.*, **37**, 11188-11189 (2003).
16. Ratel D., *et al*., "N6-methyladenine: the other methylated base of DNA", *Bioessays*, **3**, 309-315 (2006).
17. Riley K.E. and P. Hobza, "Noncovalent interactions in biochemistry", *WIREs Computational and Molecular Science*, **1**, 3-17 (2011).
18. Robertson K.D., *Oncogene*, **21**, 5361 (2002).
19. Rodríguez López C.M., *et al*., "Direct detection and quantification of methylation in nucleic acid sequences using high-resolution melting analysis", *Anal. Chem.*, **82**, 9100-9108 (2010).
20. Utkov H., *et al*., "Using density functional theory methods for modeling induction and dispersion interactions in ligand-protein complexes", *Ann. Rep. Comput. Chem.*, **16**, 97-112 (2010).
21. Xie H., *et al.*, "Genome-wide quantitative assessment of variation in DNA methylation patterns", *Nucleic Acids Research*, **39**, 4099-4108 (2011).
22. Yap L.J., *et al*., "Crystal Structure of the Dengue Virus Methyltransferase Bound to a 5'-Capped Octameric RNA", *PLoS ONE*, **5**, 1-9 (2010).
23. Zhao Y. and D. G. Truhlar, "Density functionals for noncovalent interaction energies of biological importance", *J. Chem. Theory Comput.*, **3**, 289-300 (2007).