





**GLOBAL GEOMETRY OPTIMIZATION OF DNA BASES  
VIA AN INTERMOLECULAR POTENTIAL ENERGY FUNCTION**

**M.Sc. THESIS**

**Artür MANUKYAN**

**Department of Computational Science and Engineering**

**Computational Science and Engineering Programme**

**JUNE 2013**



**GLOBAL GEOMETRY OPTIMIZATION OF DNA BASES  
VIA AN INTERMOLECULAR POTENTIAL ENERGY FUNCTION**

**M.Sc. THESIS**

**Artür MANUKYAN  
(702101002)**

**Department of Computational Science and Engineering**

**Computational Science and Engineering Programme**

**Thesis Advisor: Asst. Prof. Dr. Adem TEKİN**

**JUNE 2013**



**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ**

**MOLEKÜLLER ARASI POTANSİYEL ENERJİ FONKSİYONU İLE  
DNA BAZ YAPILARININ KÜRESEL ENİYİLENMESİ**

**YÜKSEK LİSANS TEZİ**

**Artür MANUKYAN  
(702101002)**

**Hesaplamalı Bilim ve Mühendislik Anabilim Dalı**

**Hesaplamalı Bilim ve Mühendislik Programı**

**Tez Danışmanı: Asst. Prof. Dr. Adem TEKİN**

**HAZİRAN 2013**





**Artür MANUKYAN**, a M.Sc. student of ITU Informatics Institute 702101002 successfully defended the thesis entitled “**GLOBAL GEOMETRY OPTIMIZATION OF DNA BASES VIA AN INTERMOLECULAR POTENTIAL ENERGY FUNCTION**”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :**     **Asst. Prof. Dr. Adem TEKİN** .....  
Istanbul Technical University

**Jury Members :**     **Assoc. Prof. Dr. Aylin Fethiye KONUKLAR** .....  
Istanbul Technical University

**Prof. Dr. Mine YURTSEVER** .....  
Istanbul Technical University

.....

**Date of Submission :**   **3 May 2013**

**Date of Defense :**     **5 June 2009**



## **FOREWORD**

This thesis is based on the 1001 TUBITAK research project "Intermolecular Interactions in DNA Bases". I would like to thank TUBITAK for the two years of funding. I also would like to thank UHEM (National Computing Center(In Turkish)) for the technical support and for providing computational resources required both the project and my thesis. But more importantly, I would like to thank my Master advisor for his valuable insight and directions. Because of his patience and unending guidance, I have a deeper understanding in the purpose of science than ever before.

June 2013

Artür MANUKYAN



## TABLE OF CONTENTS

|   | <u>Page</u> |
|---|-------------|
| <b>FOREWORD</b> .....   | <b>vii</b>  |
| <b>TABLE OF CONTENTS</b> .....  | <b>ix</b>   |
| <b>ABBREVIATIONS</b> .....  | <b>xi</b>   |
| <b>LIST OF TABLES</b> .....   | <b>xiii</b> |
| <b>LIST OF FIGURES</b> .....  | <b>xv</b>   |
| <b>SUMMARY</b> .....  | <b>xix</b>  |
| <b>ÖZET</b> .....   | <b>xxi</b>  |
| <b>1. INTRODUCTION</b> .....  | <b>1</b>    |
| <b>2. METHODS</b> .....   | <b>7</b>    |
| 2.1 Ab-initio Methods .....   | 7           |
| 2.1.1 Post-HF methods .....   | 8           |
| 2.1.1.1 Moeller-Plesset perturbation theory.....                                    | 8           |
| 2.1.1.2 Coupled cluster .....   | 9           |
| 2.1.2 Density functional theory .....   | 10          |
| 2.1.3 Symmetry adapted perturbation theory and DFT-SAPT.....                        | 10          |
| 2.1.4 Calculating the interactions in DNA bases via ab-initio methods.....          | 11          |
| 2.2 Fitting .....   | 12          |
| 2.2.1 The Levenberg-Marquardt method.....   | 14          |
| 2.3 Optimizations .....   | 16          |
| 2.3.1 Simulated annealing .....   | 16          |
| 2.3.2 Powell's method .....   | 18          |
| <b>3. RESULTS AND DISCUSSION</b> .....  | <b>19</b>   |
| 3.1 Evaluation of Theoretical Methods .....   | 19          |
| 3.1.1 Cytosine dimer.....   | 19          |
| 3.1.2 Guanine dimer .....   | 24          |
| 3.1.3 Cytosine-Guanine dimer.....   | 26          |
| 3.2 Fitting Surfaces of Homo and Hetero Dna Base Dimers .....                       | 29          |
| 3.2.1 Cytosine intermolecular energy function .....                                 | 30          |
| 3.2.2 Guanine intermolecular energy function .....                                  | 35          |
| 3.2.3 Cytosine-Guanine intermolecular energy function.....                          | 39          |
| 3.3 Global Optimization of DNA Dimer and Oligomers via Simulated<br>Annealing ..... | 43          |
| 3.3.1 Cytosine dimer and oligomers.....   | 44          |
| 3.3.2 Guanine dimer and oligomers .....   | 48          |
| 3.3.3 Cytosine-Guanine dimer and oligomers.....                                     | 54          |
| <b>4. CONCLUSION</b> .....  | <b>59</b>   |
| <b>REFERENCES</b> .....   | <b>61</b>   |

**CURRICULUM VITAE..... 65**

## ABBREVIATIONS

|                |   |
|----------------|---|
| <b>DNA</b>     | : Deoxyribonucleic acid                                 |
| <b>MD</b>      | : Molecular Dynamic                                     |
| <b>PES</b>     | : Potential Energy Surface                              |
| <b>STM</b>     | : Scanning Tunneling Microscopy                         |
| <b>HF</b>      | : Hartree-Fock  |
| <b>MP</b>      | : Moeller-Plesset                                       |
| <b>CC</b>      | : Coupled Cluster                                       |
| <b>CI</b>      | : Configuration Interaction                             |
| <b>CCSD</b>    | : Single and Double Excited Coupled Cluster             |
| <b>CCSD(T)</b> | : Single, Double and Perturbative Triple Couple Cluster |
| <b>CISD</b>    | : Single and Double Excited Configuration Interaction   |
| <b>MPPT</b>    | : Moeller-Plesset Perturbrational Theory                |
| <b>SCS-MP2</b> | : Spin Component Scaled MP2                             |
| <b>MET</b>     | : Many Electron Theory                                  |
| <b>SAPT</b>    | : Symmetry Adapted Perturbation Theory                  |
| <b>DF</b>      | : Density-Fitting                                       |
| <b>RI</b>      | : Resolution of Identity                                |
| <b>CP</b>      | : Counterpoise  |
| <b>BSSE</b>    | : Basis Set Superposition Error                         |
| <b>SA</b>      | : Simulated Annealing                                   |
| <b>PEC</b>     | : Potential Energy Curve                                |
| <b>CMS</b>     | : Center of Mass  |
| <b>CBS</b>     | : Complete Basis Set                                    |





## LIST OF TABLES

|   | <u>Page</u> |
|---|-------------|
| <b>Table 3.1</b> : CPU-time [seconds] for the computation of dimers, stacked (S) and anti-stacked (AS), in Figure 3.2. ....   | 20          |
| <b>Table 3.2</b> : Cytosine A and B isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and CMS distances are obtained via spline interpolation. ....                                   | 22          |
| <b>Table 3.3</b> : Energies of Dimer A (CMS 5.31 Å ) and B (CMS 3.50 Å ) with basis sets aug-cc-pVXZ(X=D, T and Q). Aug-cc-pVTZ and aug-cc-pVQZ were used for the extrapolation of CBS. ....  | 23          |
| <b>Table 3.4</b> : Interaction energy calculations for PBE[a], SCS-MP2[b] and CP-SCS-MP2[c] optimized guanine dimers A, B and C. ....   | 26          |
| <b>Table 3.5</b> : Guanine A, B, C, D and E isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and CMS distances are obtained via spline interpolation. ....                           | 28          |
| <b>Table 3.6</b> : Ab-initio calculations for PBE, SCS-MP2 and CP-SCS-MP2 optimized cytosine-guanine dimer A, Watson-Crick base pair. ....  | 29          |
| <b>Table 3.7</b> : Interaction energies of three cytosine-guanine dimers calculated at MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) using aug-cc-pvdz basis set Minimum energies and CMS distances were obtained via a spline interpolation. .... | 31          |
| <b>Table 3.8</b> : The range of six-dimensions. ....  | 31          |
| <b>Table 3.9</b> : Unique sites in cytosine dimer and the required parameter number for each unique site. ....  | 32          |
| <b>Table 3.10</b> : Atomic charges of cytosine molecule. ....   | 32          |
| <b>Table 3.11</b> : Fitted parameters for cytosine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of first and second cytosine monomer, respectively. ....               | 33          |
| <b>Table 3.12</b> : Unique sites in guanine dimer and the required parameter number for each unique site. ....  | 36          |
| <b>Table 3.13</b> : Atomic charges of guanine molecule. ....  | 37          |
| <b>Table 3.14</b> : Fitted parameters for guanine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of first and second guanine monomer, respectively. ....                 | 37          |
| <b>Table 3.15</b> : Unique sites in cytosine-guanine dimer and the required parameter number for each unique site. ....   | 40          |

|  |    |
|--|----|
| <b>Table 3.16:</b> Fitted parameters for cytosine-guanine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of cytosine and guanine monomer, respectively..... | 41 |
| <b>Table 3.17:</b> Comparison of the most important distances obtained from SA and other quantum mechanical methods for cytosine dimer. ....   | 45 |
| <b>Table 3.18:</b> Interaction energy calculations at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels for cytosine dimers A,B, and C shown in Figure 3.27.....   | 45 |
| <b>Table 3.19:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for cytosine trimer. ....   | 46 |
| <b>Table 3.20:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for cytosine tetramer. ....   | 47 |
| <b>Table 3.21:</b> Interaction energy calculations at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels guanine dimers A,B, C, D and E. ....   | 49 |
| <b>Table 3.22:</b> Comparison of the most important distances obtained from SA and other quantum mechanical methods for guanine dimer. Note(*): Dimer C and D transforms into Dimer A and B as a result of PBE, respectively.....                              | 50 |
| <b>Table 3.23:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for guanine trimer. ....  | 52 |
| <b>Table 3.24:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for guanine tetramer.....   | 53 |
| <b>Table 3.25:</b> Model and interaction energies obtained at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels for cytosine-guanine dimers A,B, C, D, E and F. ....   | 54 |
| <b>Table 3.26:</b> Comparison of the most important distances obtained from SA and other quantum mechanical methods for guanine dimer. Note(*): Dimer E transforms into Dimer F. ....  | 55 |
| <b>Table 3.27:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for GCC and GGC trimers.....  | 57 |
| <b>Table 3.28:</b> Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for GCGC tetramers. ....  | 58 |

## LIST OF FIGURES

|   | <u>Page</u> |
|---|-------------|
| <b>Figure 1.1</b> : Representation of DNA bases a) Cytosine-Guanine dimer b) Thymine-Adenine dimer. ....  | 1           |
| <b>Figure 1.2</b> : $\pi$ -stacked interactions among DNA bases. ....   | 2           |
| <b>Figure 1.3</b> : Various DNA sequences a) rich cytosine b) rich guanine c) cytosine-guanine. ....  | 3           |
| <b>Figure 1.4</b> : Various DNA interactions a) cytosine b) guanine c) cytosine-guanine. ....   | 4           |
| <b>Figure 1.5</b> : DNA interactions on gold surfaces a) planar guanine b) planar cytosine. ....  | 4           |
| <b>Figure 2.1</b> : A six-dimensional cytosine model. Here, $R$ represent the distances between center of masses. $\Theta$ and $\Phi$ are polar angles as the first monomer is centered, and $\theta$ , $\phi$ and $\psi$ are euler rotation angles as the second monomer is centered. .... | 13          |
| <b>Figure 2.2</b> : Global and Local Minimum Points. ....   | 17          |
| <b>Figure 3.1</b> : Employed configurations of cytosine dimers a) dimer A b) dimer B. ....  | 19          |
| <b>Figure 3.2</b> : DNA dimer a) stacked b) anti-stacked. ....  | 20          |
| <b>Figure 3.3</b> : Potential Energy Curve of Cytosine dimer A. ....  | 21          |
| <b>Figure 3.4</b> : Potential Energy Curve of Cytosine dimer B. ....  | 21          |
| <b>Figure 3.5</b> : DFT-SAPT(LPBE0AC) and DFT-SAPT(PBE0AC) components for cytosine dimers in aug-cc-pVXZ ( $X=D, T$ and $Q$ ) levels. A and B stands for dimer A and dimer B. ....  | 24          |
| <b>Figure 3.6</b> : Employed configurations of guanine dimers a) dimer A b) dimer B c) dimer C d) dimer D e) Dimer E. ....  | 25          |
| <b>Figure 3.7</b> : Potential Energy Curves of guanine dimers a) dimer A b) dimer B c) dimer C d) dimer D e) Dimer E. ....  | 27          |
| <b>Figure 3.8</b> : Employed configurations of cytosine-guanine dimers a) dimer A b) dimer B c) dimer C. ....   | 28          |
| <b>Figure 3.9</b> : PECs of cytosine-guanine dimers a) dimer A b) dimer B c) dimer C  | 30          |
| <b>Figure 3.10</b> : Atom numbering in cytosine molecule. ....  | 32          |
| <b>Figure 3.11</b> : Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for Cytosine Dimer. ....  | 33          |
| <b>Figure 3.12</b> : Fit errors in kJ/mol for cytosine dimer. ....  | 34          |
| <b>Figure 3.13</b> : Cytosine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C. ....   | 34          |
| <b>Figure 3.14</b> : Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for cytosine Dimer A, B and C. ....   | 35          |
| <b>Figure 3.15</b> : Partial potential energy surface of cytosine dimer. ....   | 36          |
| <b>Figure 3.16</b> : Atom numbering of guanine molecule. ....   | 36          |

|  |    |
|--|----|
| <b>Figure 3.17:</b> Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for guanine dimer. ....   | 38 |
| <b>Figure 3.18:</b> Fit errors in kJ/mol for guanine dimer. ....   | 38 |
| <b>Figure 3.19:</b> Guanine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C d) Dimer D e) Dimer E. ....  | 39 |
| <b>Figure 3.20:</b> Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for guanine Dimer A, B, C, D and E. ....  | 39 |
| <b>Figure 3.21:</b> Partial potential energy surface of guanine dimer. ....  | 40 |
| <b>Figure 3.22:</b> Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for cytosine-guanine dimer. ....  | 42 |
| <b>Figure 3.23:</b> Fit errors in kJ/mol for cytosine-guanine dimer. ....  | 42 |
| <b>Figure 3.24:</b> Cytosine-guanine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C. ....   | 42 |
| <b>Figure 3.25:</b> Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for Cytosine-Guanine Dimer A, B and C. ....   | 43 |
| <b>Figure 3.26:</b> Partial potential energy surface of cytosine-guanine dimer. ....   | 43 |
| <b>Figure 3.27:</b> Three cytosine dimers located by SA a) Dimer A b) Dimer B c) Dimer C. ....   | 44 |
| <b>Figure 3.28:</b> Three H-bonded and two stacked cytosine trimer conformations: a) cyttrimer 10 b) cyttrimer 8 c) cyttrimer 2 d) cyttrimer 7 e) cyttrimer 9. ....  | 46 |
| <b>Figure 3.29:</b> Stacked H-bonded and three planar tetramer conformations located by SA (top) and their relaxations at PBE/TZVP (bottom) a) cyttetra 1 b) cyttetra 2 c) cyttetra 3 d) cyttetra 4. ....    | 47 |
| <b>Figure 3.30:</b> C-tetrad structure obtained from a) Powell and b) further relaxation of Powell geometry with PBE/TZVP. ....  | 48 |
| <b>Figure 3.31:</b> Two cytosine filament pentamer structures a) Cypenta 1 b) Cypenta 2. ....  | 48 |
| <b>Figure 3.32:</b> Five guanine dimers located by SA approach a) Dimer A b) Dimer B c) Dimer E d) Dimer D e) Dimer C. ....  | 49 |
| <b>Figure 3.33:</b> Three H-bonded and two stacked guanine trimer conformations: a) guatrimer 9 b) guatrimer 5 c) guatrimer 1 d) guatrimer 6 e) guatrimer 2. ....  | 51 |
| <b>Figure 3.34:</b> Two stacked H-bonded and two planar tetramer conformations of model (top) and their PBE/TZVP (bottom) optimized geometries a) guatetra 1 b) guatetra 2 c) guatetra 3 d) guatetra 4. .... | 52 |
| <b>Figure 3.35:</b> G-quartet optimization a) Powell b) PBE/TZVP. ....   | 53 |
| <b>Figure 3.36:</b> Two guanine stacked pentamer structures a) Guapenta 1 b) Guapenta 2. ....  | 53 |
| <b>Figure 3.37:</b> Six cytosine-guanine dimers located by SA approach a) Dimer A b) Dimer B c) Dimer C d) Dimer D e) Dimer E f) Dimer F. ....   | 54 |
| <b>Figure 3.38:</b> GGC and GCC planar trimer conformations: a) cytguatrimer 5 b) cytguatrimer 4 c) cytguatrimer 3 d) cytguatrimer 10 e) cytguatrimer 9 f) cytguatrimer 8. ....                              | 56 |

|  |    |
|--|----|
| <b>Figure 3.39:</b> A stacked H-bonded and three planar tetramer conformations of model (top) and its PBE/TZVP (bottom) optimized geometries a) cygutetra 1 b) cygutetra 2 c) cygutetra 3 d) cygutetra 4. .... | 57 |
| <b>Figure 3.40:</b> A stacked and a filament hexamer structures a) GCGCGC 1 b) GCGCGC 2.....   | 58 |



# GLOBAL GEOMETRY OPTIMIZATION OF DNA BASES VIA AN INTERMOLECULAR POTENTIAL ENERGY FUNCTION

## SUMMARY

Non-covalent interactions stabilize biochemically significant complexes such as Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA). Especially, electrostatic (hydrogen bonds between O-H and N-H) interaction are the most important stabilizing factors in these systems. In the helical structure of DNA, complementary base pairing occurs between cytosine – guanine and adenine – thymine bases. In this study, potential energy surfaces (PES) of cytosine, guanine and cytosine – guanine dimers will be computed and the corresponding interaction energies will be fitted to an analytical expression to develop the corresponding force fields. For this purpose, first, to determine the theoretical method which will be used to compute the PES, potential energy curves (PEC) of selected dimers will be calculated employing different methods such as single and double excitation coupled cluster theory including perturbative triple excitations (CCSD(T)), second-order Møller-Plesset (MP2), spin-component scaled MP2 (SCS-MP2), dispersion augmented density functional theory (DFT-D) and density functional theory combined with symmetry adapted perturbation theory (DFT-SAPT). It has been found that interaction energies obtained from dft-sapt (lpbeoac) are in very good agreement with reference CCSD(T). Therefore, this level of theory was included to compute the pes of homo and hetero dimers of dna bases. This will be followed by fitting the interaction energies to an analytic expression containing electrostatics, dispersion and repulsion terms. Using these expressions, it is possible to perform a global search, employing one of the most successful global optimizer, Simulated Annealing, to find the interesting local and global minima of both dimers and DNA base oligomers including cytosine and guanine.





## MOLEKÜLLER ARASI POTANSİYEL ENERJİ FONKSİYONU İLE DNA BAZ YAPILARININ KÜRESEL ENİYİLENMESİ

### ÖZET

Kovalent olmayan etkileşimler Deoksiribonükleik asit (DNA) ve Ribonükleik asit (RNA) gibi biyolojik önemi yüksek olan komplekslerin yapılarını kararlı hale getirmektedir. Özellikle elektrostatik (O-H ve N-H arasında oluşan hidrojen bağları) etkileşimleri bu sistemlerin kararlılığını etkileyen en önemli aktörlerdir. DNA'nın heliksel yapısında tamamlayıcı baz eşleşmesi sitozin – guanin ve adenin – timin bazları arasında gerçekleşir. Bu çalışmada DNA bazlarından sitozin, guanin ve sitozin-guanin dimerlerinin potansiyel enerji yüzeyleri (PEY) hesaplanarak elde edilen etkileşim enerjileri analitik bir fonksiyona fitlenerek kuvvet alanları geliştirilecektir. Öncelikle PEY hesaplarının hangi teorik seviyede yapılacağına belirlenmesi için tek ve çift eksitasyonları ve perturbatif üçlü eksitasyon düzeltmelerini içeren coupled cluster (CCSD(T)), ikinci dereceden Møller-Plesset (MP2), ölçeklendirilmiş dönme bileşenli MP2 (SCS-MP2), dispersiyon eklenmiş yoğunluk fonksiyonel teori (DFT-D) ve yoğunluk fonksiyonel teori ile iliştilmiş simetri adaptasyonlu perturbasyon teorisi (DFT-SAPT) kullanılarak seçilen çeşitli dimerler için potansiyel enerji eğrileri (PEE) hesaplanacaktır. Yakın zamanda, DFT-SAPT(LPBE0AC) seviyesinde hesaplanan etkileşim enerjilerinin CCSD(T) seviyesine yakın sonuçlar verdiği gözlemlenmiştir. Bu hususta, özellikle DFT-SAPT(LPBE0AC), homo ve hetero DNA baz dimerlerinin etkileşim enerjilerinin hesaplanmasında diğer teorik seviyelerle birlikte kullanılacaktır ve tüm teorik hesaplama seviyeleri karşılaştırılacaktır. Sonrasında, elde edilen etkileşim enerjileri itme, dispersiyon ve elektrostatik kuvvetlerin bir toplamı şeklinde analitik bir forma fitlenecektir. Bu formlar yardımıyla, hem dimer hem de DNA baz oligomerlerinin PEY'leri en başarılı küresel optimizasyon metodlarından biri olan benzetimli tavlama metodu ile incelenerek önemli küresel ve yerel minimum yapıları tayin edilecektir.

Helikal DNA molekülünde nükleik asit bazlarının Watson-Crick eşleşmesi yapıları yanında, bu bazların telomerlerde hidrojen bağlı quartetler de (4 tane aynı DNA bazının biraraya gelmesiyle oluşan yapı) oluşturdukları bilinmektedir. Ayrıca, günümüzde tek-molekül (single-molecule) teknikleri oldukça gelişmiş olup, atomik kuvvet mikroskoplarının yardımıyla bireysel DNA zincirleri dahi metal yüzeylerin üzerine yerleştirilebilmektedir ve bunlar biyoçip sensörleri, organik yalı iletkenler veya organik fotovoltajik araçları olarak kullanılabilirler. Metal yüzeyine, özellikle altın, yerleşen DNA bazlarının yüzeyde nasıl konumlandıkları hem taramalı tünelleme mikroskopisi (STM) hemde moleküler dinamik (MD) simülasyonları ile incelenebilmektedir.

Günümüzde kuantum kimya hesaplama metodlarına “çözülmüş” bir problem sınıfı olarak kabul edebiliriz. Örneğin tek ve çift eksitasyonları ve perturbatif üçlü eksitasyon düzeltmelerini içeren coupled cluster (CCSD(T)) metodu tüm sistemler için oldukça doğru değerler veren ve hakem hesaplama seviyesi olarak da kabul

edilen bir metottur. Halbuki CCSD(T) metodunu ancak çok küçük sistemler için büyük baz setleri kullanarak çalıştırabiliriz. CCSD(T)'den daha düşük sistem kaynaklarına gereksinim duyan diğer metotları yine ancak çok küçük ölçekli küme yapılarını incelemek için kullanabiliriz. Moleküler Dinamik (MD) simülasyonları gibi içinde binlerce molekül yada atom içeren sistemler için en düşük teori dahi kullanılamıyacaktır. Bu durumda moleküller arası etkileşimleri küme yada MD boyutunda incelemek için tek yolumuz analitik ifadeleri kullanmaktan geçmektedir. Literatürde nükleik asit bazlarının özellikle dimerleri yoğun bir şekilde incelenmiştir. Halbuki, quartetler, i-motifler, şeritler ve diğer farklı şekildeki örgüler hesaplamalı olarak yeterince incelenememiştir. Bunun temel nedenlerinden birisi bu oligomerlerin oldukça büyük olmalarından dolayıdır. Bu durumda, MD metodu vazgeçilmez bir araç olarak karşımıza çıkmaktadır. Fakat, Lennard-Jones (LJ) potansiyel gibi enerji formülleri bu da hesaplamalardaki doğruluğu oldukça düşürmektedir. Bu problemi ortadan kaldırmanın bir yolu, homo ve hetero DNA bazları arasındaki etkileşimleri ifade eden kuvvet alanları geliştirmektir.

Bu çalışmada, DNA bazlarından sitozin ve guanin homo- ve hetero-oligomerlerindeki etkileşimleri ifade edebilecek kuvvet alanları geliştirilmiştir. Bu amaç için öncelikle sitozin, guanin ve sitozin-guanin dimer yüzeyleri hesaplanarak sonrasında analitik bir fonksiyona fitlenmiştir. CCSD(T) gibi referans bir hesaplama metodu PEY hesaplamalarında çok büyük hesaplama kaynaklarına ihtiyaç duyduğu için kullanılmamıştır. Projedeki ilk hedefimiz, CCSD(T)'ye yakın etkileşim enerjileri veren teorik metotları bulmaktır. Bu amaçla MP2, SCS-MP2, B3LYP-D ve DFT-SAPT (PBE0AC ve LPBE0AC) metotlarında aug-cc-pVXZ (X=D, T yada Q) baz setiyle bu dimerlerin en önemli hidrojen bağlı ve istiflenmiş geometrileri için etkileşim enerjilerini hesapladık. Her 3 dimer için, MP2'nun hidrojen bağlı sistemler için CCSD(T)'ye çok yakın değerler verdiğini fakat istiflenmiş yapılar için hem etkileşim enerjilerini daha düşük hem de monomerler arası minimum mesafeleri daha düşük bulduğunu gözlemledik. SCS-MP2, MP2'ü düzeltmekte fakat genellikle etkileşim enerjilerini daha büyük vermeye başlamaktadır. B3LYP-D ise, MP2 gibi davranmakla birlikte hidrojen bağlı yapılar için de daha düşük enerjiler vermektedir. Bunun yanında DFT-SAPT(PBE0AC), SCS-MP2 gibi hareket ederek CCSD(T) yakın sonuçlar vermektedir. Fakat, CCSD(T) ile uyumlu en iyi bulgular DFT-SAPT(LPBE0AC) metodu ile bulunmuştur. LPBE0AC xc-fonksiyonelin bu başarısından dolayı MP2 ve B3LYP-D hesaplamalarına göre daha fazla cpu-zamanı istemesine karşın bu dimer sistemlerinin potansiyel enerji yüzeyleri (6000 – 7000 hesaplama noktası içermektedir) bu metot kullanılarak hesaplandı. Elde edilen etkileşim enerjileri, mevki-mevki bir Buckingham tipli, itme, dispersiyon ve elektrostatik terimlerinden oluşan bir potansiyel formuna Levenberg-Marquardt lineer olmayan en küçük karalar metodu kullanılarak fitlendi. Herbir dimer fitlemesi için elde edilen standart sapmalar 1 mH etkileşim enerjisinden daha küçük oryantasyonlar için 0.44, 0.72 ve 0.51 mH olarak sırasıyla sitozin, guanin ve sitozin-guanin için bulundu. Her üç dimer içinde fitlemeyle elde edilmiş modelin ürettiği etkileşim enerjileri LPBE0AC ile oldukça uyumlu bulundu. Kuvvet alanları sonrasında, sitozin, guanin ve sitozin-guaninden oluşmuş dimerlerin BT metoduyla küresel eniyilenmesiyle birçok dimer izomeri bulunmuştur. Bunlardan en düşük enerjili olanı, B3LYP-D, MP2, SCS-MP2 ve DFT-SAPT ile aynıdır. Yüksek enerjili izomerlerde ise kuantum mekanik metotlardan elde edilen sıralamalarda farklılıklar oluşmaktadır. Dimer geometrilerindeki, dimerler arasındaki

en önemli uzunluklara baktığımızda ise model CP-SCS- MP2 ile oldukça uyumlu mesafeler üretmektedir.

Kuvvet alanları homo ve hetero sitozin ve guanin trimerlerine uygulandığında, birçok yeni izomer elde edilmiştir. Bulunan model trimerlerin geometrileri, PBE/TZVP seviyesinde eniyilendiğinde, istiflenmiş yapılar haricinde, geometrilerin çok değişmediği gözlenmiştir. PBE/TZVP, istiflenmiş dimer geometrilerinde, bu yapıları düzlemsel hidrojen bağlı olacak şekilde değiştirme eğilimindeydi. Fakat, bu durum, SCS-MP2 ve CP-SCS-MP2 metotlarında görülmemektedir. Bundan dolayı, PBE/TZVP'nin istiflenmiş trimerleri düzlemsel hale dönüştürmesi bu metodun bir zayıflığı olarak kabul edilebilir.

Sitozin ve guanin homo ve hetero tetramerleri deneysel olarakta gözlemlendikleri için (quartetler) büyük önem arz etmektedirler. Bu yapılar dört DNA sarmalının uygun bir şekilde konumlanmasıyla oluşmaktadır ve daha çok guanin ve sitozin-guanin için bulunmuşlardır. Model, çeşitli sitozin tetramerlerini ve deneysel olarakta varlığı bilinen c-tetrad yapısının enerjisini diğer izomere göre daha yüksek tahmin etmektedir. Daha önemli olan, guanin ve sitozin-guanin tetramerlerinde de model çok çeşitli ve deneysel olarakta varlıkları bilinen yapıları bulabilmekte olup, bunların enerjilerinin de kuantum mekanik değerlere yakın olarak vermektedir.

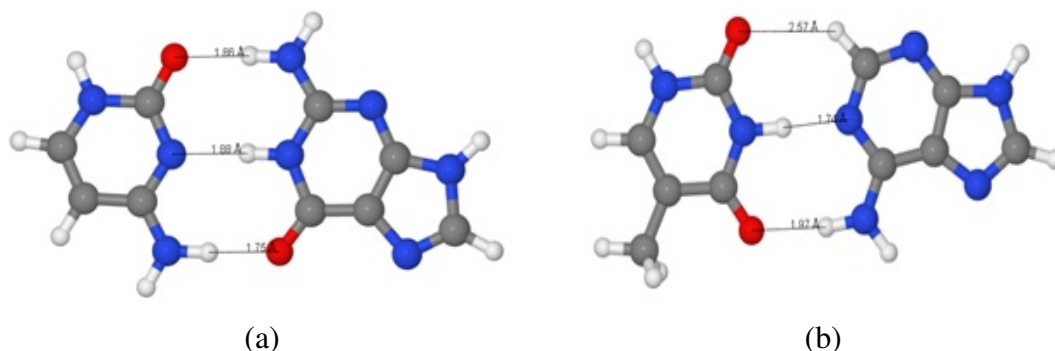
Elde edilen bu başarılı sonuçlar, yapıları hakkında daha az bilgi sahibi olduğumuz daha büyük oligomerlerin yapılarını aydınlatmakta da bu kuvvet alanlarını kullanabileceğimizi göstermektedir.



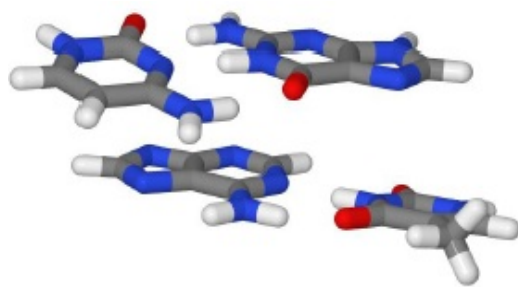
## 1. INTRODUCTION

Deoxyribonucleic acid (DNA) contains required genetic information of all known living beings and some viruses. Thus, DNA is considered to be the most important biological molecule. Non-covalent interactions ensure the stabilization of DNA and RNA. Especially, electrostatic (Hydrogen bonding between O-H and N-H) and  $\pi$ -stacking interactions are the most important stabilization factors. There are four different bases in DNA. These are Adenine (A), Cytosine (C), Guanine (G), Thymine (T) bases. Here, A-G and T-C are purine and pyrimidine bases, respectively. A purine base is paired with a pyrimidine base to establish a DNA double helix. As shown in Figure 1.1, cytosine-guanine bases and adenine-thymine bases are paired via hydrogen bonds. Moreover,  $\pi$ -stacking interactions also ensure the dispersion based stabilization of hydrogen bonded base pairs on DNA systems, as shown in Figure 1.2.

In quantum chemistry, there are many methods used to compute the interaction energies. Amongst them, single and double excitations and perturbative triple excitation corrected couple cluster (CCSD(T)) is a successful and reliable method. However, we can use CCSD(T) for only small chemical systems with high-level basis sets. Hence, it is inappropriate and time-consuming to employ CCSD(T) for the investigation of DNA systems and clusters. Furthermore, some other supermolecular methods, which require less computational resources than CCSD(T), could only be applied to small cluster systems. And more importantly, using some of these methods



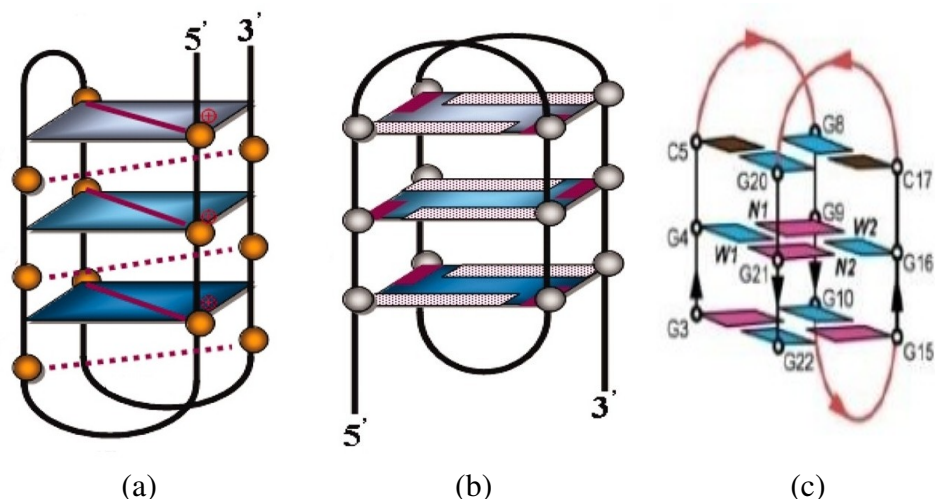
**Figure 1.1:** Representation of DNA bases a) Cytosine-Guanine dimer b) Thymine-Adenine dimer.



**Figure 1.2:**  $\pi$ -stacked interactions among DNA bases.

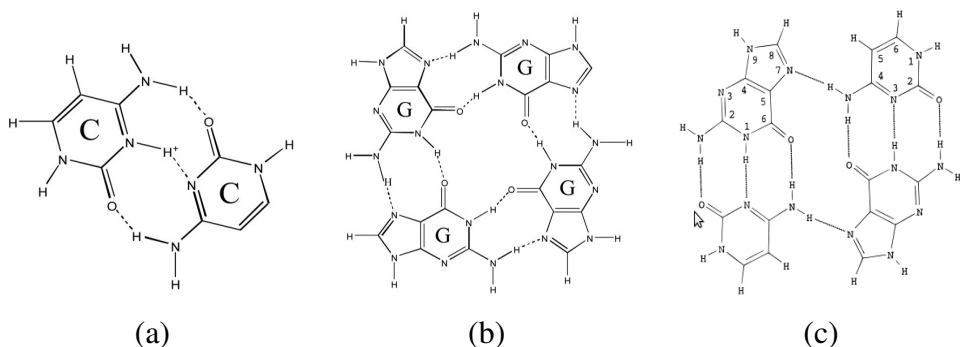
for Molecular Dynamic (MD) simulations, where thousands of molecules and atoms reside, is very unlikely. In this manner, analytical expressions could be very effective for MD simulations and cluster computations. A way of constructing an analytical expression is to compute the entire Potential Energy Surface (PES) by using dimer structures; and then fit the corresponding PES to an analytical energy function. Here, the accuracy of the theoretical methods are crucial since the method will be used to generate the data required for the fitting of PES. The method is required to be both fast and accurate to obtain better predictions.

Investigation of DNA bases in theoretical levels covers a wide area in the literature. For instance, 17 stacked and 4 hydrogen bonded cytosine dimers have been investigated via MP2 and CCSD(T) levels with aug-cc-pVDZ and aug-cc-pVTZ basis sets [1]. These dimers were optimized via MP2 with 6-31G\*\* basis sets. In this study, CCSD(T) energies have been produced by the extrapolation of MP2 energies to complete basis sets rather than directly computed [1]. Furthermore, variational-perturbational stacking components of 63 cytosine dimers in B-DNA crystals have been investigated on MP2/aug-cc-pVDZ levels [2]. Via variational-perturbational method, as in DFT-SAPT, energy components are observed individually in order to understand which force is more superior on stabilization of dimer structures. As a result, dispersion was found to be the most stabilizing factor, and electrostatic contribution has been found positive for the most dimer conformations [2]. A similar study have also been conducted for guanine dimers [3]. Here, results showed that dispersion is also the most effective stabilizing factor for guanine dimers. As also for cytosine-guanine dimers, stacked conformations can be found in the literature along with H-bonded dimers [4]. Here, the most stabilized H-bonded conformation is the Watson-Crick base pair which consists of three hydrogen bondings [5].

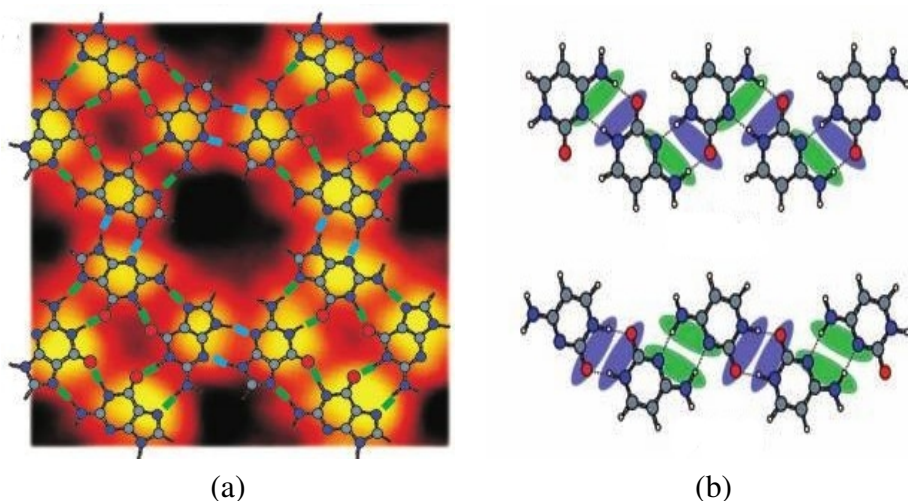


**Figure 1.3:** Various DNA sequences a) rich cytosine b) rich guanine c) cytosine-guanine.

Along with Watson-Crick paired helical DNA structures, it is known that these DNA bases form triplexes, quadruplexes and many other complex structures [6]. Cytosine, guanine and cytosine-guanine interactions form various quadruplex conformations, as shown in Figure 1.3, where these structures can be observed by scanning tunneling microscopy (STM) [7, 8]; and, as shown in Figure 1.4, these quadruplexes are formed via unique hydrogen bonds. Especially, guanine quartets has been intensively investigated both experimentally and theoretically at BLYP-D/TZ2P level. Some guanine quartets includes cytosine tetrads as a form of "sandwich" via the interaction of four parallel d-(TGGGCGGT) DNA helices. Experimental and theoretical evidences of cytosine tetrads are also available [9, 10]. It has been emphasized that this planar cytosine tetrad is mostly stabilized by hydrogen bonds. As for GCGC interactions shown in Figure 1.4, Watson-Crick interactions appear to be the most important factor [11]. Two Watson-Crick pairs form fully hydrogen bonded GCGC quadruplexes [12]. Here, GCGC formations were found even more stabilized than guanine quadruplexes. Nowadays, single-molecule methods are quite advanced, and this allows to embed individual DNA structures on metallic surfaces. Resulting metallic-DNA hybrids are used to build biochip sensors, organic semi-conductors, and organic photovoltaic tools [13–15]. DNA bases on Au(111) surfaces have been observed with STM, and further investigated with MD simulations. For example, in Figure 1.5, STM images of embedded guanine bases on Au(111) surface are shown. Here, guanine bases form guanine quadruplexes to establish a planar guanine layer on Au(111) [16]. As also



**Figure 1.4:** Various DNA interactions a) cytosine b) guanine c) cytosine-guanine.



**Figure 1.5:** DNA interactions on gold surfaces a) planar guanine b) planar cytosine.

indicated in Figure 1.5, cytosine molecules form planar filaments on Au(111) [17]. In these studies, both DNA bases effectively adsorbed onto Au(111) using recent single-molecule techniques [17, 18]. Apparently, hydrogen bonded conformations ensures the planar interactions. Similar researches on metallic-DNA surfaces [19, 20] exist for pure cytosine [21, 22], guanine [16], adenine [23–26], and thymine [27, 28]. Moreover, MD simulations of the absorption of DNA bases on Au(111) surfaces have been defined via Lennard-Jones potential (related AMBER force field parameters are used) and Coulomb term, and DNA-Au(111) interactions have been computed with Born-Mayer potential [29].

Although DNA base dimers are extensively examined in the literature; filaments, motifs, quartets and other more complicated DNA structures have not been studied theoretically. The main reason behind this is that these oligonucleotides are quite large structures, and the calculation of interaction energies results in a high computational complexity. Hence, MD simulations could be a solution to this time-consuming



problem. Generally, Lennard-Jones potential is employed for this purpose and this reduces the accuracy of the calculations. An alternative is to use force fields especially derived for the interactions between DNA bases. In this study, cytosine and guanine homo and hetero dimers will be computed and the corresponding force fields will be developed by fitting the interaction energies to analytical energy functions. For this aim, the first step is to find the best theoretical method which is agree with ccsdt to calculate the interaction energies. Then, these energies will be fitted to analytical functions.. Thereafter, we will use the resulting force fields to investigate structure of dna oligomers. Indeed, a dimer based force field might not be quite effective to be applied to DNA oligomers. However, this novel analytical expression could be more efficient compared to Lennard-Jones and AMBER force fields.



## 2. METHODS

### 2.1 Ab-initio Methods

The name of "ab-initio" has a latin origin which means "from the beginning". Ab-initio methods, or "first principles", which are electronic structure methods, produce correct theoretical results, and they require intensive high parallel computing resources. As a term, ab-initio calculations are based on quantum mechanical laws (charges, mass and atomic core of electrons), statistical thermodynamics, and some physical constants (speed of light and Planck constant). However, a big disadvantage of ab-initio methods that they are only applicable for small systems. Nevertheless, these systems using approximations such as Density Fitting and Resolution of Identity (RI).

For an ab-initio calculation, the theoretical method and the basis set must be set. The most basic ab-initio method is the Hartree-Fock (HF) theory. However, electron correlation effects cannot be determined with HF. This effect is crucial for the determination of dispersion forces on intermolecular interactions. For this purpose, post-HF methods have been developed to overcome this drawback. The first post-HF method is Moeller-Plesset (MP) perturbation theory. Based on the included number of terms, MP can be classified from second order MP (MP2) to forth order MP (MP4) or more orders. Apart from these, there exist methods, such as coupled cluster (CC), and configuration interaction (CI), that produce more accurate results than MP.

Especially, CC with single, double and perturbative triple excitations (CCSD(T)) is considered as a reference method. However, it is impossible to employ CCSD(T) for big systems and high basis sets. CC with single and double excitations (CCSD) is a lot faster than CCSD(T). Similar to CC methods, CI methods are named according to excitations. The fastest CI is the one including single and double excitations (CISD). If the wave function can be determined as the sum of all possible excitations, it becomes a full-CI: the most accurate method. Hence, the order of accuracy is: HF < MP2 < CISD  $\approx$  MP4  $\approx$  CCSD < CCSD(T) < CCSDT < Full-CI.

### 2.1.1 Post-HF methods

HF method expresses an average electron repulsion instead of a full definition. In HF theory, possibility of an existing electron around an atom is determined by the distance from the core, but instead, should be determined by considering the distances from the other electrons also. This is of course physically wrong.

To avoid such an error, post-HF methods, first, calculate the HF energy; and then, they try to fix the electron correlation. Some of these approaches are Moeller-Plesset perturbation theory (MPn, n=2,3,4, n is the order of correction), Generalized Valence-bond method, Multi Configurational Self-consistent Field, Configurational Interaction, and Coupled Cluster. Especially, electron correlation is important on systems where dispersion interactions are superior.

#### 2.1.1.1 Moeller-Plesset perturbation theory

Moeller-Plesset Perturbation Theory (MPPT) is calculated via adding electron correlation to HF energy. Moreover, increasing the number of terms results in the calculation of second order MP (MP2), third order MP (MP3), fourth order MP (MP4) and so on. MP2 has the minimum number of correlations added to HF, and is the most popular method among MP methods. Moreover, MP3 and MP4 are also quite widely used methods. Here, the accuracy of MP4 is close to CISD, and the use of MP5 is rare due to immense need of computational resources.

MP2 calculations are not variational apart from HF. Hence, MP2 energies could be produced lower than the actual energies. Regarding the nature of chemical system, the employment of higher perturbation level might increase or decrease the energy more than necessary, or make the energy converge to the real energy; for example, MP2 can produce a lower energy, MP3 can produce a higher energy, and MP4 might, again, generate a lower energy than the real energy. On MPPT, perturbed Hamiltonian is obtained by adding the perturbation term  $V$  to the HF Hamiltonian. Here,  $V$  is the correlation energy, and is actually multiplied by the undimensional parameter,  $\lambda$ . If there is no perturbation,  $\lambda$  becomes 0; and if there is a full perturbation,  $\lambda$  becomes 1. MP2 energy, as in Equation 2.1, is calculated by the sum of anti-parallel and parallel spin correlation energies:

$$E_{corr}(MP2) = E_{corr}(\downarrow\downarrow) + E_{corr}(\uparrow\downarrow) \quad (2.1)$$

However, HF theory, as a spin correlation, contains high amount of electron correlation with parallel spins. Hence, energy scaling is required for anti-parallel electron correlation since they don't evenly contribute to the total energy. In Equation 2.2, a scaled version of Equation 2.1 is presented [30]:

$$E_{corr}(MP2) = c_1 E_{corr}(\downarrow\downarrow) + c_2 E_{corr}(\uparrow\downarrow) \quad (2.2)$$

Here,  $c_1$  and  $c_2$  are scaling parameters. This corrected MP2 method is called spin component scaled MP2 (SCSMP2). According to Equation 2.1, MP2 generally produce lower energies than the expected value. With SCS-MP2, correlation energy is calculated lower, and the results are generally more accurate. Mostly, MP2 energies generate higher interaction energies than SCS-MP2.

### 2.1.1.2 Coupled cluster

Coupled Cluster (CC) theory, which is a variant of Many Electron Theory (MET), employs a non-linear to construct the multi-electron wavefunction. Here, this operator is  $e^{\hat{T}}$ , and  $\hat{T}$  is known as the excitation operator. By expanding  $e^{\hat{T}}$  into Taylor series, this exponential definition is transformed into a summation of linear expressions to solve the time-independent Schrodinger equation:

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Phi_0\rangle \quad (2.3)$$

, and:

$$|\Psi_{CC}\rangle = \left(1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \dots\right)|\Phi_0\rangle \quad (2.4)$$

Moreover, the solution is improved by defining higher order of excitations where the sum of these excitations is  $\hat{T}$ :

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots\hat{T}_n \quad (2.5)$$

Here, n is the highest order of excitation; for example, inclusion of the second order excitation to CC results in the following:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 \quad (2.6)$$

The most popular and considered to be the most accurate CC variant is the single, double, and perturbative triple excitation CC (CCSD(T)).

### 2.1.2 Density functional theory

In Density Functional Theory (DFT), total energies are calculated by density functionals instead of wave functions [31]. Here, a functional is defined as a function of a function, and the density functional is the function of electron density. A density function is used to calculate the energy of electron density. Hence, it is believed that electron density might contain enough information about the system. However, since the method requires a functional, determining an appropriate functional is a crucial step. Although the method is considerably faster than wave function based methods, it agrees well with these methods. But unfortunately, DFT, which is an effective theoretical method, fails to calculate interaction energies: unsuccessfully calculates dispersion term where electron correlation is important. However, this drawback have been overcome via the empirical dispersion correction, DFT-D. In DFT-D, total interaction energy is given by:

$$E_{DFT-D} = E_{KS-DFT} + E_{disp}. \quad (2.7)$$

Here,  $E_{KS-DFT}$  is a self-consistent Khon-Sham energy, and  $E_{disp}$  is the empirical dispersion correction:

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} f_{dmp}(R_{ij}) \quad (2.8)$$

Here,  $N_{at}$  is the number of atoms,  $C_6^{ij}$  is the dispersion parameter for the  $ij$  pair,  $s_6$  is the general scaling factor based on density functional, and  $R_{ij}$  is the distance between atoms. In addition, it is important to include the  $f_{dmp}$  damping function to avoid singularity on short  $R_{ij}$  distances:

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{\frac{R_{ij}}{R_r-1}}} \quad (2.9)$$

Here,  $R_r$  is the sum of van der Waals radius.

### 2.1.3 Symmetry adapted perturbation theory and DFT-SAPT

Interaction energies can be calculated by two different approaches: supermolecular calculations in any theoretical level, and symmetry adapted perturbation theory (SAPT). A major feature of SAPT is its ability to decompose the interaction energy

into first-order electrostatic,  $E_{el}^{(1)}$ , second-order induction,  $E_{ind}^{(2)}$ , and dispersion  $E_{disp}^{(2)}$  terms. Also, terms;  $E_{exch}^{(1)}$ ,  $E_{exch-ind}^{(2)}$ , and  $E_{exch-disp}^{(2)}$ ; accompanies to these terms as a result of electron exchange among monomers. The effect of higher terms than the second is represented by  $\delta(HF)$ , and this term can be defined as the subtraction of HF interaction energy and the summation of electrostatic, induction, and their exchange energies in HF level. Since SAPT [32] could be computationally expensive, a two hybrids of DFT and SAPT have recently been developed: DFT-SAPT [33–36] and SAPT(DFT) [37]. In DFT-SAPT, interaction energy,  $E_{int}$ , is calculated via a series of physical terms:

$$E_{int} = E_{el}^{(1)} + E_{exch}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} + \delta(HF) \quad (2.10)$$

Here,  $E_{el}^{(1)}$  and  $E_{exch}^{(1)}$  are the first-order Coulomb ve exchange energies,  $E_{ind}^{(2)}$  and  $E_{exch-ind}^{(2)}$  are the second-order induction and its exchange counterpart, and  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$  are the second-order dispersion and its exchange counterparts. Moreover, the third and higher order contributions are called as  $\delta(HF)$  and which is:

$$\delta(HF) = E_{int}(HF) + E_{el}^{(1)}(HF) + E_{exch}^{(1)}(HF) + E_{ind}^{(2)}(HF) + E_{exch-ind}^{(2)}(HF) \quad (2.11)$$

Here,  $E_{int}(HF)$  is the CP-corrected supermolecular Hartree-Fock (HF) interaction energy. Other terms are calculated with HF density matrix and coupled-perturbated HF (CPHF) response-density matrix.

Density-fitting variant of DFT-SAPT, DF-DFT-SAPT (a computationally faster approach), recently implemented into MOLPRO [38] quantum chemistry package. In DF-DFT-SAPT, monomer properties are obtained by PBE0AC [33] and LPBE0AC [39] density functionals. Here, PBE0AC is the combination of %25 exact change included by PBE exchange-correlation (xc) potential and asymptotically corrected LB94 xc potential. The SAPT becomes computationally more efficient when the monomer properties accounted by DFT. Moreover, it could be accelerated more via DF-DFT-SAPT. DFT-SAPT, especially, produces accurate results for CH- $\pi$  and  $\pi - \pi$  interactions in agreement to CCSD(T).

#### 2.1.4 Calculating the interactions in DNA bases via ab-initio methods

Compared to monomer energies, interaction energies are lower in four magnitudes. Therefore, wide calculation basis sets are required, and the most appropriate basis

sets are the Dunning’s augmented basis sets (aug-cc-pVXZ; X=D, T or Q). As for theoretical methods, electron correlations included MP2, SCS-MP2 [40] and CCSD(T) methods can be used. Even the lowest level of augmented basis set, aug-cc-pVDZ, is computationally expensive for small monomers; for example, cytosine. Our goal is to, first, find the most accurate and computationally feasible method alternative to CCSD(T). Prior to calculate the interactions energies, hydrogen bonded and stacked dimers will be optimized by DFT (with PBE functional), SCS-MP2 and CP-SCS-MP2 methods via TURBOMOLE [41]. Afterwards, we will apply MP2, SCS-MP2, B3LYP-D [42, 43], DFT-SAPT(PBE0), DFT-SAPT(LPBE0) and CCSD(T) methods with aug-cc-pVXZ basis sets where X=D, T or Q to calculate the interaction energy in DNA bases. Density Fitting (DF) approach has been used for supermolecular HF, MP2 and SCS-MP2 calculations. Additionally, in DF-HF and DF-DFT-SAPT, cc-pV(X+1)Z JK-fitting basis set has been employed [44]. Moreover, aug-cc-pVXZ MP2-fitting basis set have been employed in the calculations of DF-MP2 and DF-SCS-MP2 [45].

Expect for DFT-SAPT, counterpoise (CP) correction has been used for the supermolecular calculation methods (to avoid Basis set superposition error, BSSE). In The CP scheme, interaction energy is calculated by the following formula:

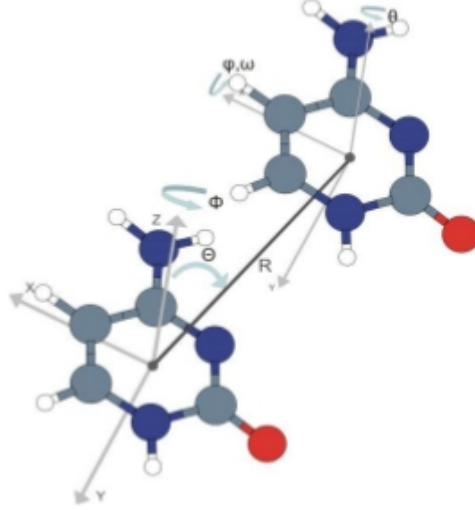
$$V_{AB}^{CP} = E_{AB}(AB) - E_A(AB) - E_B(AB) \quad (2.12)$$

Here,  $V_{AB}^{CP}$  is the CP corrected interaction energy,  $E_{AB}(AB)$  is the total energy of AB dimer using the total basis set of monomers A and B,  $E_A(AB)$  and  $E_B(AB)$  are the total energies.

## 2.2 Fitting

The second phase of this work is to fit the interaction energies to a model potential energy function. These energies will be acquired by the calculation of PES with a chosen theoretical method. Levenberg-Marquardt method will be used to fit the model [46]. For this purpose, six dimensional, e.g., cytosine dimer shown in Figure 2.1 is defined, and the same approach is followed for the other dimers, guanine and cytosine-guanine. Since the DNA bases are not symmetric molecules, the resulting PES can contain thousands of dimer conformations. PES of DNA base dimers are generated with a python script which is able to remove the symmetric orientation. Also, dimers with their closest atom-atom distance smaller than 1.5 are removed.





**Figure 2.1:** A six-dimensional cytosine model. Here,  $R$  represent the distances between center of masses.  $\Theta$  and  $\Phi$  are polar angles as the first monomer is centered, and  $\theta$ ,  $\phi$  and  $\psi$  are euler rotation angles as the second monomer is centered.

Success of the analytical potential fitting depends on the characteristics of the model function. Regarding the fitting of PES, LJ, Morse, Buckingham and Varshni models were frequently employed [47–50]. However, none of these were efficient for acetylene dimers. Hence, a novel model form has been established by adding new terms to the function, and it has resulted in a success [50]. In this study, we will apply the same potential function for DNA dimers. This site by site interaction model is the summation of buckingham type repulsion, dispersion and the electrostatic terms:

$$E_{int} = \sum_{i=1}^N \sum_{j=1}^N u_{ij}(r_{ij}). \quad (2.13)$$

Here,  $u_{ij}$  is the potential energy between  $i$ th site of the first monomer and  $j$ th site of the second monomer of a dimer structure.  $u_{ij}$  is a function of the distance between two sites,  $r_{ij}$ , and can be formulated as:

$$u_{ij}(r_{ij}) = \alpha_{ij} e^{(-\beta_{ij} r_{ij})} + \frac{C_{ij}}{r_{ij}^6 + c_{ij}^6} + f(\theta_{ij}, r_{ij}) \frac{q_i q_j}{r_{ij}}. \quad (2.14)$$

Here,  $\alpha$ ,  $\beta$ ,  $C$  and  $\delta$  are the fitting parameters.  $c_{ij}$  has been added to the dispersion term in order to avoid errors resulted by the effect of close distance between atoms on dispersion and electrostatic terms. Moreover, the special case of tang-toennies damping function ( $\delta=n=1$ ):

$$f(\theta_{ij}, r_{ij}) = 1 - e^{-\theta r} \sum_{m=0}^n \frac{(\theta r)^m}{m!}, \quad (2.15)$$

has been multiplied with the electrostatic term. This function converges to 1 for increasing value of  $r_{ij}$ , and to 0, vice versa. In the fitting, the main goal is to minimize the weighted  $\chi^2$  term:

$$\chi^2 = \sum_{i=1}^N (\sigma_i (y_i - y(x_i; \alpha_{ij}, \beta_{ij}, C_{ij})))^2 \quad (2.16)$$

where  $y_i$  and  $y$  are theoretical and model interaction energies. Weighting term,  $\sigma_i$ , is treated separately for each dimer orientation. Thus, negative interaction energies are tried to be estimated more accurately while the accuracy of positive energies are ignored.

### 2.2.1 The Levenberg-Marquardt method

Levenberg-Marquardt fitting algorithm is considered to be a non-linear least square fitting method. The fitting method is a hybrid of the Steepest Descent and Newton's methods [51]. The dumping parameter,  $\lambda$ , controls which is more dominant. While Steepest Descent is more effective when the solution is away from the minimum, Newton's method accelerates the solution when the solution vector converges to the minimum. The goal is the minimization of sum of square error between the dependant variable and its approximation:

$$\chi^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.17)$$

Here,  $\hat{y}_i$  is the approximation of  $y_i$ . Dependant variable can also be formulated as:

$$\hat{y}_i = \hat{y}_i(x_1, x_2, \dots, x_p; \beta_1, \beta_1, \dots, \beta_k) \quad (2.18)$$

where  $x$  and  $\beta$  are independent variables and fitting parameters, respectively. Minimizing  $\chi^2$  results in fortifying the approximation of  $y_i$ . For this purpose, it is important to first express the Taylor series expansion of  $y_i$ .

$$y(\mathbf{x}, \boldsymbol{\beta} + \mathbf{b}) = y(\mathbf{x}, \boldsymbol{\beta}) + P^T \mathbf{b}. \quad (2.19)$$

Here,  $b$  is the converging value of the estimate of parameter  $\beta$ . At each iteration, the solution vector,  $\delta$ , is added to  $b$  until the  $\chi^2$  is minimized. Standard least square approximation solves the linear set of equations:

$$A\delta = \mathbf{g} \quad (2.20)$$

$$\mathbf{g} = P^T(\mathbf{y} - \hat{\mathbf{y}}), \quad A = P^T P \quad \text{and} \quad P = \left( \frac{\partial \hat{y}_j}{\partial b_j} \right). \quad (2.21)$$

Levenberg-Marquardt adds a damping variable,  $\lambda$ , to the least square problem as shown in the following equation [52]:

$$(A + \lambda I)\delta = \mathbf{g} \quad \text{or} \quad (P^T P + \lambda I)\delta = P^T(\mathbf{y} - \hat{\mathbf{y}}). \quad (2.22)$$

For the corresponding iteration, solution of the new equation leads us to find the stepsize and the search direction,  $\delta$ . With current  $\delta$ , Equation 2.22 is being formulated and solved. Then, solution is added to the current  $\beta_i$ . What can be observed from Equation 2.20 and 2.22 is that, for a  $\lambda$  value that is closer to zero, Equation 2.22 resembles Equation 2.20. Here,  $\lambda$  serves a role of changing the behaviour of convergence and improving the convergence rate to be much faster. In each step, the value of  $\lambda$  is altered in order to minimize  $\chi^2$  effectively. For a given initial set of parameters  $b_0$ , the algorithm runs as follows:

1. Compute  $\chi^2(b_i)$
2. Pick a value for  $\lambda$
3. Solve Equation 2.22 and find  $\delta$
4. Update:  $b_{k+1} = b_k + \delta_k$
5. If  $\chi^2(b_{k+1}) > \chi^2(b_k)$ , increase  $\lambda$ , and go to step 3
6. If  $\chi^2(b_{k+1}) < \chi^2(b_k)$ , decrease  $\lambda$ , and go to step 3

If current  $\chi^2$  is less than the previous one, our problem converges to an standart least square fitting, Newton's method, as shown in Equation 2.20. On the other hand, with the increasing  $\chi^2$ , value of  $\lambda$  also increases to accelerate the convergence rate where  $\chi^2$  gets away from the minimum by using the Steepest Descent Method.

## 2.3 Optimizations

Heuristic search algorithms are based on locating an approximate solution for a numerical problem. Especially, for global numerical optimizations, heuristic algorithms are often used to locate global minimum regions where local optimization algorithms are inappropriate. Here, it is known that such heuristic algorithm cannot precisely locate the actual global minimum. However, with advanced techniques, such as Simulated Annealing (SA), it is possible to manipulate search directives to fortify the minimization process.

In this study, we aim to optimize DNA dimer and oligomer geometries globally. Here, it is important to uncover the most stabilized DNA structures to enhance our knowledge on DNA sequences and metallic-DNA relationships. Moreover, global optimizations are quite efficient due to the employment of previously fitted intermolecular potential functions. Thus, numerous SA runs can be simulated to find the global minimum of PES. At each iteration, we generate a random set of initial points from our six-dimensional dimer model, and then, execute a SA search. Since it is known that SA might ignore the global minimum region and stuck on a neighboring local minimum, it is essential to initiate successive SA searches.

Furthermore, although global search algorithms allow us to elucidate some important stabilized DNA geometries, it might be crucial to test our potential function whether the fitted PES could locate some local DNA conformations or not. For this purpose, we have employed the local optimization method of Powell. The method optimizes objective function where either a derivatable function exists or not. In this study, we have employed a potential function with the distance between atoms are input variables, but the optimized parameters are from the six-dimension dimer model. However, the gradient of our function can be obtained with finite differences.

### 2.3.1 Simulated annealing

Employing random optimizing parameters is the only way of locating different minimum regions. Local optimization methods rejects parameters which increase the value of objective function. However, by accepting some of these rejected values, current minimum region might be avoided, and parameters can "jump" to a new minimum region where a lower objective minimum value resides (see

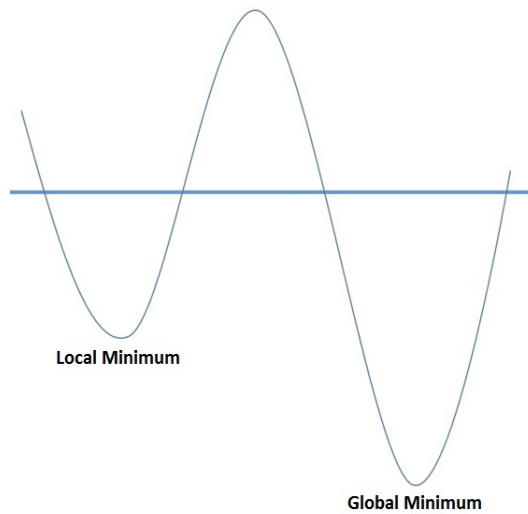
Figure 2.2). Here, this jumping operation is conducted by a controlled random generation mechanism, "temperature". The process starts by assigning an initial temperature value, and it slowly decreases the temperature till it cools down: reaching to zero. Throughout the global search, new points are accepted when the new point either has a lower value or a reasonable value according to Metropolis criteria [53]:

$$p = e^{-\frac{f_i - f'}{T_k}} . \quad (2.23)$$

Here,  $f_i$  and  $f'$  are candidate and old minimum points, respectively; and,  $T_k$  is the current temperature. More importantly,  $p$  value determines if the point will be accepted or not. If  $p$  is lower or higher than a randomly generated value between  $[0, 1]$ , the point is accepted. Moreover, the  $T_k$  decreases after  $N_s * N_t$  function evaluations. Thus, new points start to get hardly accepted due to increasing  $p$ . Here, new candidate points are evaluated in  $N_s$  iterations, and the stepsize,  $v$ , is altered at the end; for example:

$$v' = v \left( 1 + c \frac{n/N_s - 0.6}{0.4} \right) \quad (2.24)$$

Later on, these evaluation is repeated for  $N_t$  times. As it is understood, higher number of  $N_s$  and  $N_t$  results in higher possibilities of locating the actual global minimum since more random points are considered with more iterations. These random numbers are multiplied with  $v$  and added to the current set of optimizing parameters to generate new parameters.



**Figure 2.2:** Global and Local Minimum Points.

Generally, SA consists of point evaluation, altering stepsize and decrease of temperature:

1. Compute  $x' = x_i + rv_k$
2. if  $f(x') < f_i$  accept; else compute  $p'$ . If  $p' < p$  accept, else reject and  $j = j + 1$
3. if  $j < N_s$  go to Step 1, else continue
4. update  $v$  and  $k = k + 1$
5. if  $k < N_T$  go to Step 1, else continue
6. reduce  $T$ , and apply termination criteria. If accepted, stop; else, go to Step 1

### 2.3.2 Powell's method

Gradient based local optimizers are easy to be employed due to their effectiveness of finding the appropriate direction to locate the nearby minimum points. However, it is often hard to define a derivatable objective function or take a derivative. Powell's method is one of these non-gradient local optimizers. The idea is to generate  $N$  number of direction vectors with a size of  $N$ . Here,  $N$  is also the number of optimizing parameters. By defining  $N$  linearly independant directions, one can search for a minimum along the first direction and find the minimum; and from there, search along the second direction, and continue untill no significant improvement can be done. However, it is essential to find efficient linearly independant directions. One solution could be an identity matrix,  $I$ , but it is doubtful that the identity matrix would work for all applications. Hence, in Powell's method, selection of successful direction vectors is challenging.

In Powell's method, searching a minimum along a direction is called line minimization [46]. In two steps, line minimization can be carried out by, first, bracketing the minimum region, and then, locating the minimum point via Brent's one-dimensional minimization method [54]. Here, Brent's method is a combination of bisection and secant method. Brent's method can also be fortified by defining the first-derivatives.

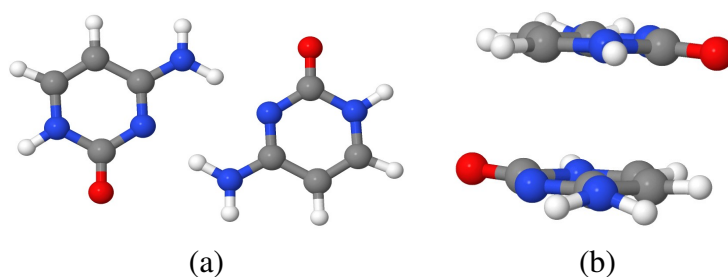
### 3. RESULTS AND DISCUSSION

#### 3.1 Evaluation of Theoretical Methods

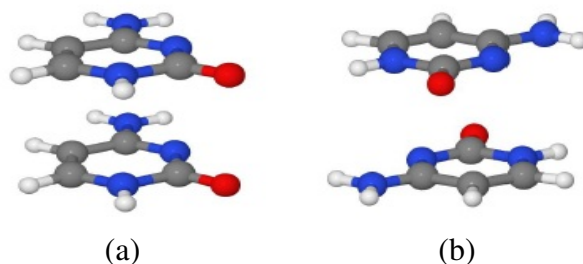
In this part of our work, an accurate theoretical method will be chosen to establish the PES of cytosine, guanine and cytosine-guanine dimers. For this purpose, energies of some prototype dimer conformations will be calculated in MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) and CCSD(T) levels.

##### 3.1.1 Cytosine dimer

Two of the considerably important, a hydrogen bonded and a stacked, cytosine dimers have been shown in Figure 3.1 [1]. To obtain the potential energy curves (PEC) of these geometries, we have considered several distances between the center of masses of two monomers within a range of 2.50 and 14 Å, and 2.50 and 12 Å for H-bonded and stacked conformations, respectively. Unique 20 H-bonded and 13 stacked conformations have been selected from the mentioned range, and energies of these dimers have been computed with theoretical methods. Since CCSD(T) calculations with aug-cc-pVDZ basis set is quite time-consuming, only three points near PEC minimum zone have been calculated with CCSD(T). Prior to the PEC, we have calculated the interaction energies of two stacked dimers shown in Figure 3.2, and the cpu time required for all methods are listed in Table 3.1. Clearly, it takes a large amount of time to calculate a CCSD(T) compared to other ab-initio methods. CCSD(T) is followed by LPBE0AC, PBE0AC, B3LYP-D, SCS-MP2 and MP2.



**Figure 3.1:** Employed configurations of cytosine dimers a) dimer A b) dimer B.



**Figure 3.2:** DNA dimer a) stacked b) anti-stacked.

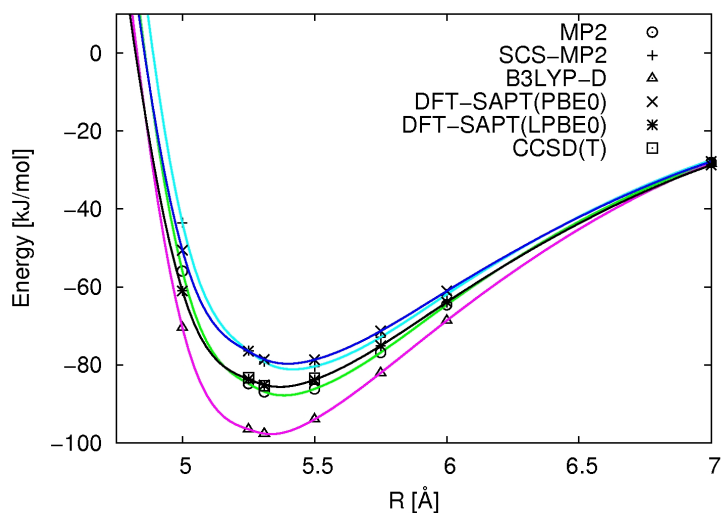
**Table 3.1:** CPU-time [seconds] for the computation of dimers, stacked (S) and anti-stacked (AS), in Figure 3.2.

| Dimer | MP2  | SCS-MP2 | B3LYP-D | DFT-SAPT<br>(PBE0AC) | DFT-SAPT<br>(LPBE0AC) | CCSD(T) |
|-------|------|---------|---------|----------------------|-----------------------|---------|
| S     | 7628 | 7628    | 10365   | 30139                | 40934                 | 2371265 |
| AS    | 7424 | 7424    | 10103   | 29946                | 40843                 | 1480066 |

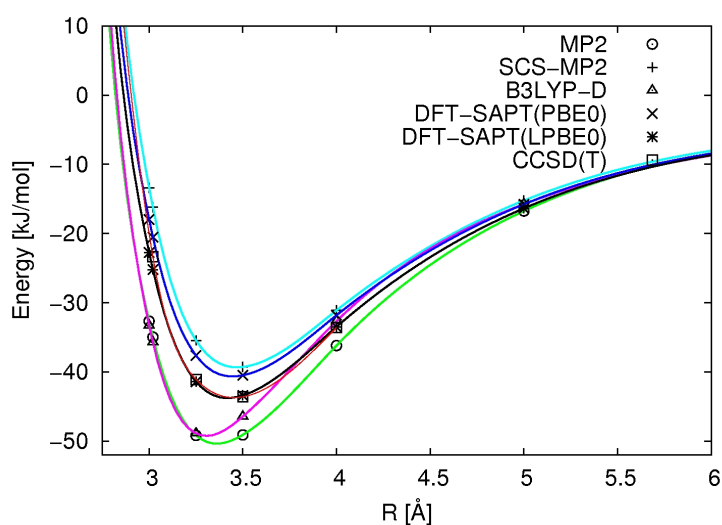
To analyze the performance of all considered computation levels, two cytosine dimers shown in Figure 3.1 were selected and the corresponding PEC's are shown in Figure 3.3 and Figure 3.4. Here, H-bonded dimer is twice as stabilized as than the stacked dimer for all theoretical calculations. For example, dimer A is 41.64 kJ/mol more stabilized than dimer B in CCSD(T) level. The main reason is the superiority of N-H and O-H interactions over CH- $\pi$  and  $\pi - \pi$  interactions. Similar findings were also observed for pyrazine and triazine dimers [55]. For pyrazine and triazine, however, stacked dimers can be found more stabilized than H-bonded ones. Nevertheless, dimer A is also more stabilized than dimer B in MP2 and SCS-MP2 levels. Although MP2 produces lower interaction energies as indicated in Figure 3.4, it agrees well with CCSD(T). The same behaviour of MP2 calculations have also been seen on research of acetylene-benzene [56], pyrazine and triazine dimers [55]. Mostly, SCS-MP2 corrects MP2 results, but this time, SCS-MP2 results in higher interaction energies than CCSD(T). Also for dimer A and B, B3LYP-D acts similar to MP2. Apparently, experimental dispersion correction of B3LYP-D overestimates the dispersion contribution. Especially for dimer B, MP2 and B3LYP-D give the same results on short center of mass (CMS) distances. However, it is the opposite for long distances: while MP2 calculates lower interaction energies, B3LYP-D begins to give high energies, as observed in H<sub>2</sub>S-benzene, metan-benzene and metan dimers [57]. In



DFT-SAPT, PBE0AC produces lower interaction energies than SCS-MP2 for dimer A, and vice-versa for dimer B. However, local version of xc potential corrects PBE0AC results, and it is seen that LPBE0AC gives the closest energy values to CCSD(T) calculations [55].



**Figure 3.3:** Potential Energy Curve of Cytosine dimer A.



**Figure 3.4:** Potential Energy Curve of Cytosine dimer B.

Table 3.2 shows the smoothed minimum CMS distance and corresponding interaction energies of the points shown in Figure 3.3 and Figure 3.4. Again, MP2 produce similar results to CCSD(T) for dimer A; however, for dimer B, the CMS distance and the interaction energy decrease by 0.06 Å and 6.44 kJ/mol, respectively. SCS-MP2 gives 4.71 and 4.32 kJ/mol lower interaction energies than CCSD(T) for dimer A and B, but CMS distance is 0.06 Å longer for dimer A. Nevertheless, for dimer B, SCS-MP2 has

the same CMS distance with CCSD(T). For both dimer A and B, B3LYP-D has shorter distances: 0.08 and 1.5 Å shorter, respectively [55,57]. Moreover, B3LYP-D gives the lowest interaction energies compared to CCSD(T). Smoothed energies of B3LYP-D are 11.99 and 4.94 kJ/mol lower than CCSD(T). CMS distance of PBE0AC is 0.04 Å lower than CCSD(T) for dimer A, but distances are the same for dimer B. Especially, PBE0AC produces the highest energies among dimer A calculations. However, for dimer B, PBE0AC agrees well with CCSD(T) regarding the energies. On the other hand, among all calculations, distances and energies of LPBE0AC are the closests to CCSD(T) for both dimers. Differences in distances and energies for dimer A and B are 0.01 and 0.02 Å, and 0.42 and 0.07 kJ/mol, respectively.

**Table 3.2:** Cytosine A and B isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and CMS distances are obtained via spline interpolation.

| Dimer | Method  | CMS [Å] | Eint [kJ/mol] |
|-------|---------|---------|---------------|
| A     | MP2     | 5.40    | -87.43        |
|       | SCS-MP2 | 5.44    | -80.98        |
|       | B3LYP-D | 5.30    | -97.68        |
|       | PBE0AC  | 5.42    | -79.45        |
|       | LPBE0AC | 5.37    | -85.27        |
|       | CCSD(T) | 5.38    | -85.69        |
| B     | MP2     | 3.38    | -50.50        |
|       | SCS-MP2 | 3.44    | -39.74        |
|       | B3LYP-D | 3.29    | -49.00        |
|       | PBE0AC  | 3.43    | -41.04        |
|       | LPBE0AC | 3.42    | -44.13        |
|       | CCSD(T) | 3.44    | -44.06        |

In Table 3.3, basis set dependency in the interaction energies of dimer A and B is given. Additionally, Table 3.3 includes complete basis set (CBS) limit energies which are obtained by the extrapolation of aug-cc-pVTZ and aug-cc-pVQZ results [58]. Extrapolation is done by the use of contiguous basis sets; for example, aug-cc-pVDZ and aug-cc-pVTZ, or aug-cc-pVTZ and aug-cc-pVQZ. However, in Table 3.3, extrapolated results of CCSD(T) are omitted since aug-cc-pVTZ calculations are required even for the most simple CBS, and its computation is extremely time-consuming for CCSD(T). As shown in Table 3.3, MP2 energy of dimer A decrease by 1.95 kJ/mol from aug-cc-pVTZ to aug-cc-pVQZ, and 1.36 kJ/mol from

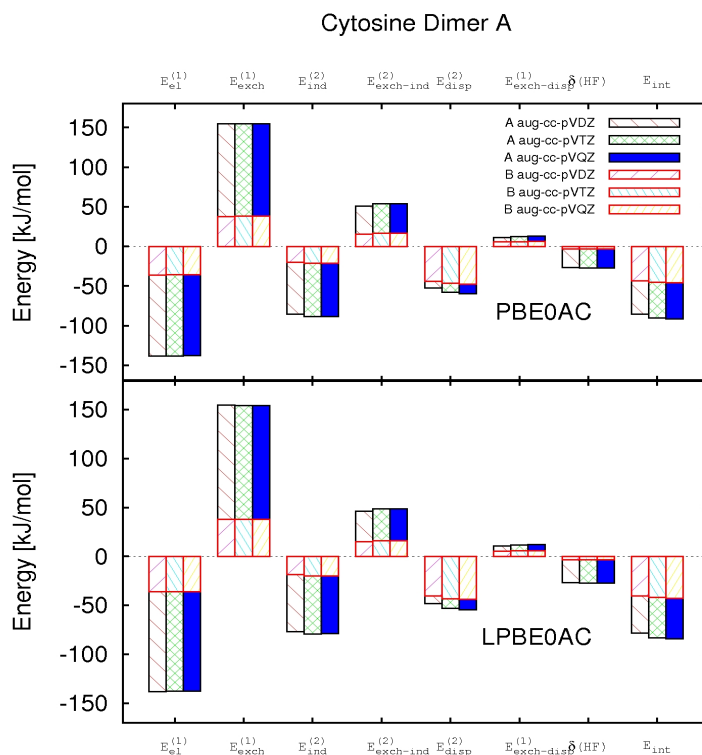
aug-cc-pVQZ to CBS. And for dimer B, a similar picture was obtained. For dimer A, SCSMP2 energy decreases by 2.05 and 1.44 kJ/mol and for dimer B, similar results to MP2 case were obtained. All basis sets of B3LYP-D calculations are close to each other. And, both PBE0AC and LPBE0AC calculations represent similar results to MP2 and SCS-MP2. Even though CCSD(T) calculations in higher basis sets are ignored, it is believed that results would have been similar to MP2 and SCS-MP2 [55]. Here, it is important to mention that it is quite challenging to calculate CCSD(T)/aug-cc-pVTZ even for little monomers, such as benzene, pyrazine and triazine.

**Table 3.3:** Energies of Dimer A (CMS 5.31 Å ) and B (CMS 3.50 Å ) with basis sets aug-cc-pVXZ(X=D, T and Q). Aug-cc-pVTZ and aug-cc-pVQZ were used for the extrapolation of CBS.

| Dim. | X   | $E_{int}$ |            |         |                      |                       |         |
|------|-----|-----------|------------|---------|----------------------|-----------------------|---------|
|      |     | MP2       | SCS<br>MP2 | B3LYP-D | DFT-SAPT<br>(PBE0AC) | DFT-SAPT<br>(LPBE0AC) | CCSD(T) |
| A    | D   | -86.92    | -79.23     | -97.65  | -78.60               | -85.15                | -85.86  |
|      | T   | -92.38    | -84.47     | -97.97  | -83.21               | -90.11                |         |
|      | Q   | -94.33    | -86.52     | -98.12  | -84.33               | -91.29                |         |
|      | CBS | -95.69    | -87.96     | -98.23  | -85.09               | -91.29                |         |
| B    | D   | -49.11    | -39.30     | -46.46  | -40.50               | -43.41                | -43.61  |
|      | T   | -51.44    | -41.32     | -46.77  | -42.16               | -45.24                |         |
|      | Q   | -52.39    | -42.25     | -46.76  | -42.89               | -45.99                |         |
|      | CBS | -53.05    | -42.90     | -46.75  | -43.36               | -46.50                |         |

With DFT-SAPT, interaction energies can be decomposed into several physical components and their exchange terms. Here, we have demcomposed the energies of dimer A and B in DFT-SAPT (PBE0AC and LPBE0AC) aug-cc-pVXZ (X=D, T or Q) levels as shown in Figure 3.5.  $E_{el}^{(1)}$ ,  $E_{ind}^{(2)}$ ,  $E_{disp}^{(2)}$  are all attractive and all exchange terms are repulsive: the order of exchange energies are  $E_{exch-disp}^{(2)} < E_{exch-ind}^{(2)} < E_{exch}^{(1)}$ . Furthermore,  $E_{el}^{(1)}$  is dominant term for dimer A, and  $E_{disp}^{(2)}$  is the dominant component for dimer B. Here, we can say that  $E_{el}^{(1)}$  and  $E_{disp}^{(2)}$  terms ensure the stabilization of H-bonded and stacked conformations, respectively. In DFT-SAPT calculations, using LPBE0AC as a exchange correlation functional has resulted in a decrease in the energy compared to PBE0AC, and because of that reason, LPBE0AC is actually agrees well with CCSD(T). Moreover, dispersion contribution is lower for LPBE0AC; for example, for cytosine dimer A (5.31 Å) and dimer B (3.02 Å), LPBE0AC xc potential produced dispersion energies which are lower than PBE0AC by 9.19 and 5.00 kJ/mol,

respectively for dimer A and B. And apparently, all components other than  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$ , are converged with aug-cc-pVDZ.

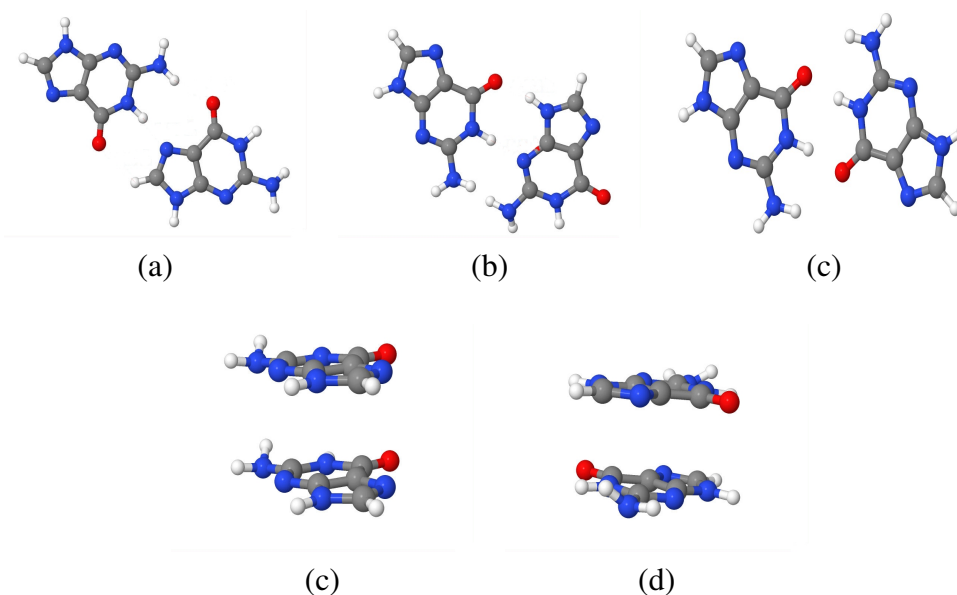


**Figure 3.5:** DFT-SAPT(LPBE0AC) and DFT-SAPT(PBE0AC) components for cytosine dimers in aug-cc-pVXZ (X=D, T and Q) levels. A and B stands for dimer A and dimer B.

### 3.1.2 Guanine dimer

Three H-bonded and two stacked guanine dimer conformations shown in Figure 3.6 were considered to test the performance of different methodologies. Here, dimer A, B and C forms three, two and four hydrogen bonds; and dimer D and E are full-stacked and anti-stacked conformations, respectively.

Here, H-bonded conformations have been structurally optimized with DFT (with PBE functional), SCS-MP2 and CP-SCS-MP2 methods using aug-cc-pVDZ basis sets. As a result, totally nine optimized H-bonded conformations have been generated, and interaction energies of these stabilized geometries have been computed with different ab-initio methods using aug-cc-pVDZ basis set. Table 3.4 shows the results. Interaction energies of dimers A,B and C are in the ascending order of SCS-MP2, CP-SCS-MP2 and PBE. Here, MP2 level of optimizations are omitted, but it is known



**Figure 3.6:** Employed configurations of guanine dimers a) dimer A b) dimer B c) dimer C d) dimer D e) Dimer E.

that MP2 results in higher energies than SCS-MP2 and CP-SCS-MP2 for pyrazine dimers [55]. As it can be seen from Table 3.4, PBE energies produce the lowest energies for all obtained optimized geometries. Moreover, regarding CMS distances of PBE, SCS-MP2 and CP-SCS-MP2 geometries, PBE produces the geometries with the shortest CMS distance. For all three dimers, the differences in interaction energies of SCS-MP2 and CP-SCS-MP2 geometries are approximately 1 kJ/mol. However, this difference is nearly 2 kJ/mol when comparing CP-SCS-MP2 and PBE for dimer A and B. Moreover, the difference increases to 4.5 kJ/mol for dimer C. In conclusion, it has been decided that CP-SCS-MP2 method should be used to optimize hydrogen bonded dimer structures due to the CP correction term; but since PBE generates geometries with lower energies than CP-SCS-MP2, this could be an intriguing situation.

It has been mentioned before that CCSD(T) is too time-consuming to be employed for the establishment of the force fields. Especially, for guanine, this complexity increases even more. Hence, CCSD(T) computation are omitted for the investigation of guanine dimer PEC; however, it is possible to compare the remaining ab-initio methods to observe if DFT-SAPT(LPBE0AC) behaves similar to cytosine dimer as shown in Figure 3.3 and Figure 3.4. PEC of five guanine dimers are given in Figure 3.7, and the minimum of the curves are given in Table 3.5. H-bonded conformations, dimer A, B and C, are two times or more stabilized than stacked dimers, dimer D and E; and dimer

**Table 3.4:** Interaction energy calculations for PBE[a], SCS-MP2[b] and CP-SCS-MP2[c] optimized guanine dimers A, B and C.

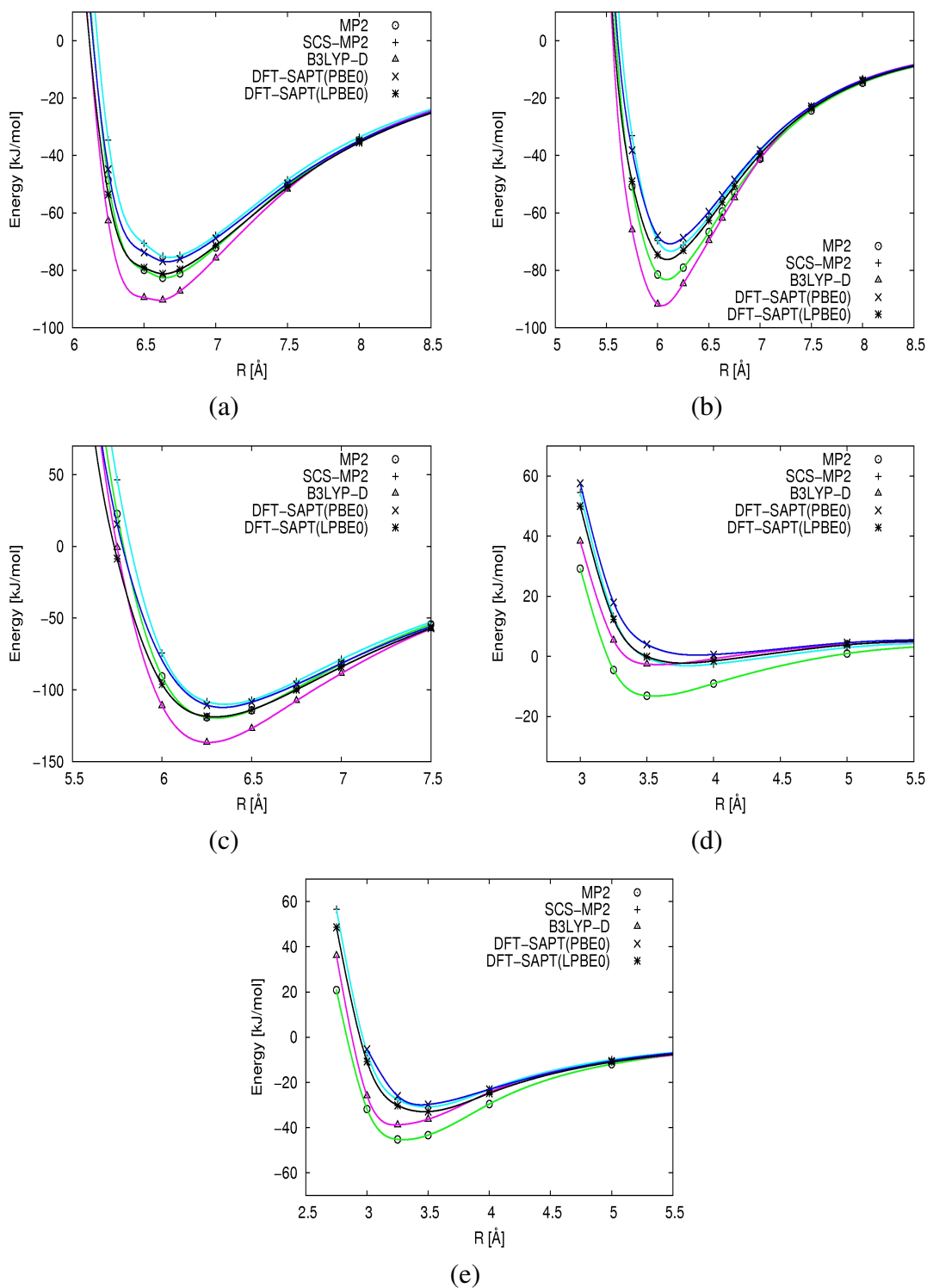
| Dimer | MP2     | SCS-MP2 | B3LYP-D | DFT-SAPT<br>(PBE0AC) | DFT-SAPT<br>(LPBE0AC) | CMS  |
|-------|---------|---------|---------|----------------------|-----------------------|------|
| A[a]  | -84.91  | -76.96  | -93.27  | -78.80               | -83.51                | 6.62 |
| A[b]  | -81.87  | -74.28  | -89.76  | -76.15               | -80.56                | 6.64 |
| A[c]  | -82.61  | -75.14  | -90.35  | -76.90               | -81.25                | 6.63 |
| B[a]  | -84.71  | -73.57  | -94.24  | -70.92               | -77.21                | 6.03 |
| B[b]  | -81.94  | -71.50  | -90.86  | -69.25               | -75.04                | 6.08 |
| B[c]  | -82.78  | -72.45  | -91.48  | -70.09               | -75.84                | 6.08 |
| C[a]  | -124.87 | -114.30 | -142.21 | -115.91              | -123.78               | 6.28 |
| C[b]  | -118.15 | -108.24 | -134.82 | -110.16              | -117.34               | 6.32 |
| C[c]  | -119.82 | -109.95 | -136.39 | -111.77              | -118.96               | 6.31 |

C has the lowest interaction energy due to four O-H bondings. Apparently, MP2 calculations of guanine dimers acts similar to cytosine dimers: Although MP2 results are similar to LPBE0AC for H-bonded dimers, it generates extremely low interaction energies for stacked dimers. And again, SCS-MP2 corrects MP2 results, and increase the interaction energy to be more than LPBE0AC. As for B3LYP-D, low interaction energies are obtained. SCS-MP2, PBE0AC and LPBE0AC give similar energies for stacked dimers. However, for all five dimers, LPBE0AC calculations are consistent, similar to cytosine dimers. Therefore, LPBE0AC will be employed to calculate of PES of guanine dimer

### 3.1.3 Cytosine-Guanine dimer

To investigate the PEC of Cytosine-Guanine dimers, we have employed a Watson-Crick pair [5], dimer A and two stacked dimers, B and C, as shown in Figure 3.8. Both stacked conformations have unique anti-stacked structures [4]. Similar to cytosine and guanine dimers, first, we have optimized the H-bonded dimer A with PBE, SCS-MP2 and CP-SCS-MP2, and calculated the interaction energy in MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC). Later on, PECs of three dimer conformations have been calculated to choose the best ab initio method to fit the function of cytosine-guanine dimers. Again, for cytosine-guanine dimers, CCSD(T) calculations are ignored.

Watson-crick base pair, dimer A, has been optimized with three geometry optimization method in Table 3.6. Here, PBE geometry have the shortest CMS distance, and the

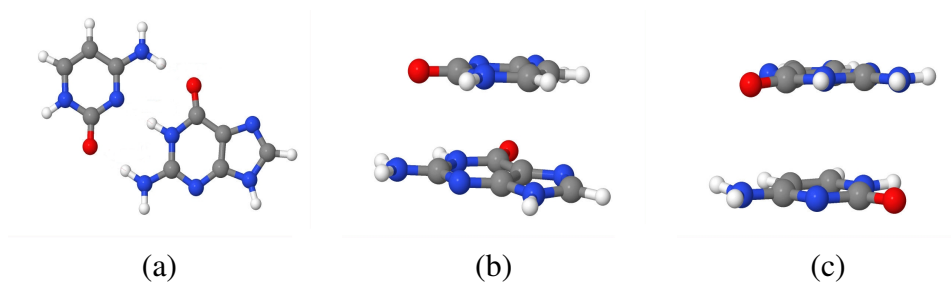


**Figure 3.7:** Potential Energy Curves of guanine dimers a) dimer A b) dimer B c) dimer C d) dimer D e) Dimer E.

interaction energies of PBE optimized dimer A is the most stabilized geometry for all ab-initio calculations. Although CP-SCS-MP2 has CP correction and likely to produce more stabilized geometries, it seems, both in guanine and cytosine-guanine dimers (Figure 3.4 and Figure 3.6), PBE gives clearly lower interaction energies. However,

**Table 3.5:** Guanine A, B, C, D and E isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and CMS distances are obtained via spline interpolation.

| Dimer | Method  | CMS [ $\text{\AA}$ ] | Eint [ $\text{kJ/mol}$ ] |
|-------|---------|----------------------|--------------------------|
| A     | MP2     | 6.62                 | -82.64                   |
|       | SCS-MP2 | 6.69                 | -74.42                   |
|       | B3LYP-D | 6.60                 | -90.67                   |
|       | PBE0AC  | 6.63                 | -76.89                   |
|       | LPBE0AC | 6.62                 | -81.30                   |
| B     | MP2     | 6.11                 | -81.93                   |
|       | SCS-MP2 | 6.17                 | -72.55                   |
|       | B3LYP-D | 5.99                 | -91.91                   |
|       | PBE0AC  | 6.16                 | -70.02                   |
|       | LPBE0AC | 6.12                 | -75.25                   |
| C     | MP2     | 6.25                 | -119.15                  |
|       | SCS-MP2 | 6.38                 | -109.63                  |
|       | B3LYP-D | 6.23                 | -136.86                  |
|       | PBE0AC  | 6.37                 | -111.21                  |
|       | LPBE0AC | 6.25                 | -118.52                  |
| D     | MP2     | 3.48                 | -13.16                   |
|       | SCS-MP2 | 3.90                 | -2.76                    |
|       | B3LYP-D | 3.49                 | -2.62                    |
|       | PBE0AC  | 3.94                 | 0.52                     |
|       | LPBE0AC | 3.89                 | -1.78                    |
| E     | MP2     | 3.25                 | -45.25                   |
|       | SCS-MP2 | 3.45                 | -31.14                   |
|       | B3LYP-D | 3.24                 | -38.84                   |
|       | PBE0AC  | 3.45                 | -30.06                   |
|       | LPBE0AC | 3.45                 | -33.31                   |



**Figure 3.8:** Employed configurations of cytosine-guanine dimers a) dimer A b) dimer B c) dimer C.

dimer A have been optimized with CP-SCS-MP2 (it is believed that CP corrected SCS-MP2 is more accurate).



**Table 3.6:** Ab-initio calculations for PBE, SCS-MP2 and CP-SCS-MP2 optimized cytosine-guanine dimer A, Watson-Crick base pair.

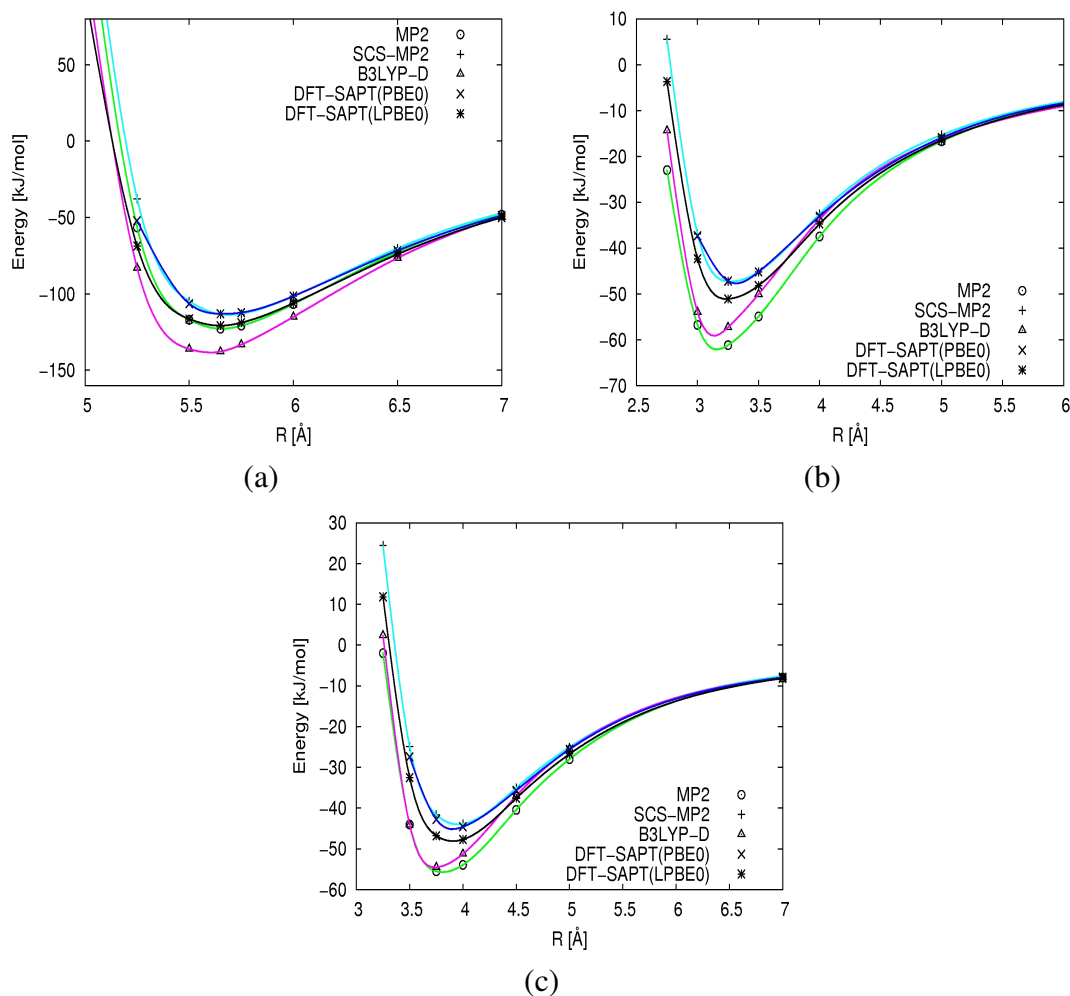
| Geo.opt.   | MP2     | SCS-MP2 | B3LYP-D | DFT-SAPT<br>(PBE0AC) | DFT-SAPT<br>(LPBE0AC) | CMS  |
|------------|---------|---------|---------|----------------------|-----------------------|------|
| PEC        | -127.26 | -116.72 | -144.24 | -116.33              | -125.11               | 5.59 |
| SCS-MP2    | -121.85 | -112.25 | -137.12 | -112.09              | -119.96               | 5.64 |
| CP-SCS-MP2 | -122.81 | -113.36 | -137.81 | -113.00              | -120.79               | 5.65 |

In Figure 3.9 and Table 3.7, ab-initio energies and the minimum of the PECs are given. It is obvious that cytosine-guanine dimers are the most stabilized dimer structures compared to guanine and cytosine dimers (see Figure 3.3, Figure 3.4 and Figure 3.7). Especially, cytosine-guanine dimer A is lower in energy than guanine dimer C by 3.71 kJ/mol in MP2 level, 4.05 kJ/mol in SCS-MP2 level, 1.72 kJ/mol in B3LYP-D level, 1.87 kJ/mol in DFT-SAPT(PBE0AC) level and 2.35 kJ/mol in DFT-SAPT(LPBE0AC) level. MP2 energies shown in Figure 3.9 suggests that MP2 produces low interaction energies for stacked dimers but give similar results to LPBE0AC for H-bonded dimer A. Also, according to Table 3.7 MP2 has the lowest CMS distances. Moreover, SCS-MP2 increases the MP2 energies, and B3LYP-D generates the lowest interaction energies. Especially, SCS-MP2 and DFT-SAPT(PBE0AC) results shows similarity for dimers A and B. As it can be seen from table 3.7, comparing PBE0AC and LPBE0AC shows that LPBE0AC energies are slightly lower than PBE0AC. Overall, all findings for cytosine-guanine dimers resembles to that of cytosine and guanine dimers.

For all three systems, cytosine, guanine and cytosine-guanine dimer, it is obvious that MP2 might be a good alternative to CCSD(T) for the H-bonded orientations. However, it is the opposite for stacked conformations: MP2 is far lower than CCSD(T). SCS-MP2 increases the MP2 energies, and B3LYP-D often behaves similar to MP2. Among both DFT-SAPT calculations, PBE0AC mostly resembles to SCS-MP2. In conclusion, DFT-SAPT(LPBE0AC) is the most effective method to be used instead of CCSD(T) since LPBE0AC has the closest interaction energies and center of mass distances to CCSD(T). Thus, PES of cytosine-guanine dimer will also becalculated with DFT-SAPT(LPBE0AC).

### 3.2 Fitting Surfaces of Homo and Hetero Dna Base Dimers

In this section, The PES of cytosine, guanine and cytosine-guanine dimers will be fitted to an intermolecular potential function using the interaction energies calculated



**Figure 3.9:** PECs of cytosine-guanine dimers a) dimer A b) dimer B c) dimer C

by DFT-SAPT(LPBE0AC)/aug-cc-pVDZ. Model and LPBE0AC energies will be compared to observe the goodness of the fit.

### 3.2.1 Cytosine intermolecular energy function

6-dimensional PES of cytosine dimer has been generated by a python script. The values of these 6 dimensions are given in Table 3.8. This selection of computation grid produced a total of 28224 points. After the elimination of symmetric points and the points with closest atom distance less than  $1.5 \text{ \AA}$ , 6140 unique cytosine dimer conformations have been produced to be calculated with DFT-SAPT(LPBE0AC).

These 6140 interaction energies were fitted to the analytic function by considering 10 unique interactions shown in Table 3.9 for 169 total interactions. This model has resulted in 30 fit parameters and 10 adjustable constants,  $c$ . So, for all these individual unique interactions, same fit parameters were used. Also, atomic

**Table 3.7:** Interaction energies of three cytosine-guanine dimers calculated at MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) using aug-cc-pvdz basis set. Minimum energies and CMS distances were obtained via a spline interpolation.

| Dimer | Method  | C.m.s. Å | <i>kJ/mol</i> |
|-------|---------|----------|---------------|
| A     | MP2     | 5.64     | -122.86       |
|       | SCS-MP2 | 5.70     | -113.68       |
|       | B3LYP-D | 5.61     | -138.58       |
|       | PBE0AC  | 5.69     | -113.08       |
|       | LPBE0AC | 5.63     | -120.87       |
| B     | MP2     | 3.19     | -61.83        |
|       | SCS-MP2 | 3.25     | -47.31        |
|       | B3LYP-D | 3.17     | -58.31        |
|       | PBE0AC  | 3.30     | -47.30        |
|       | LPBE0AC | 3.24     | -51.14        |
| C     | MP2     | 3.84     | -55.64        |
|       | SCS-MP2 | 3.94     | -44.35        |
|       | B3LYP-D | 3.73     | -54.58        |
|       | PBE0AC  | 3.92     | -45.21        |
|       | LPBE0AC | 3.92     | -48.44        |

**Table 3.8:** The range of six-dimensions.

| Dim.     | Range                              |
|----------|------------------------------------|
| R        | 3.0, 4.0, 5.0, 6.0, 7.0, 9.0, 12.0 |
| $\Theta$ | 0, 45, 90                          |
| $\Phi$   | 28, 45, 92, 154, 208, 260, 318     |
| $\theta$ | 0, 45, 90, 270                     |
| $\phi$   | 0, 45, 90, 270                     |
| $\psi$   | 0, 45, 90, 135, 180, 225, 270, 315 |

charges required in the electrostatic term were calculated using the ESP method with PBE1PBE(PBE0)/aug-cc-pVDZ, via GAUSSIAN programme. Resulting atomic charges are shown in Table 3.10, and the atom numbering is shown in Figure 3.10.

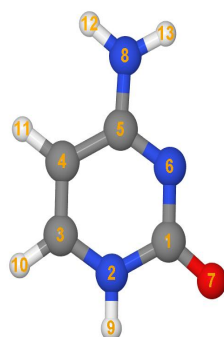
For cytosine dimers, weighting term,  $\sigma_i$ , is set to  $\sigma_i = \frac{1}{y_i^2}$  for interaction energies greater than 1 mH, and  $\sigma_i = e^{\frac{1-y_i}{5}}$  for energies lower than 1 mH. In this sense, the most crucial important points of PES can be estimated more accurately. The comparison of model and LPBE0AC energies are presented in Figure 3.11 and Figure 3.12, and the optimized fit is given in Table 3.11. Our fitting model is quite successful except a few points in the close-contact region of the PES. For orientations with low energies, our model is able to produce DFT-SAPT(LPBE0AC) with an approximate error of 5

**Table 3.9:** Unique sites in cytosine dimer and the required parameter number for each unique site.

| Interaction Type | Interaction Number | Parameter Number |
|------------------|--------------------|------------------|
| C-C              | 16                 | 3                |
| C-H              | 20                 | 3                |
| C-O              | 8                  | 3                |
| C-N              | 24                 | 3                |
| O-O              | 1                  | 3                |
| O-N              | 6                  | 3                |
| O-H              | 10                 | 3                |
| N-N              | 9                  | 3                |
| H-N              | 30                 | 3                |
| H-H              | 25                 | 3                |

**Table 3.10:** Atomic charges of cytosine molecule.

| Atom number | Atom type | Charge    |
|-------------|-----------|-----------|
| 1           | C         | 0.957370  |
| 2           | N         | -0.628865 |
| 3           | C         | 0.256891  |
| 4           | C         | -0.746416 |
| 5           | C         | 1.025022  |
| 6           | N         | -0.808602 |
| 7           | O         | -0.627164 |
| 8           | N         | -1.054862 |
| 9           | H         | 0.357512  |
| 10          | H         | 0.137243  |
| 11          | H         | 0.242960  |
| 12          | H         | 0.440158  |
| 13          | H         | 0.448753  |

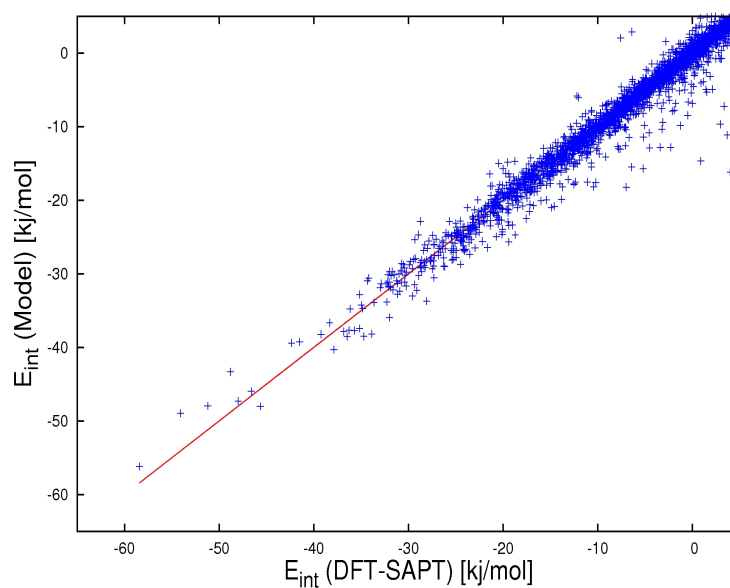


**Figure 3.10:** Atom numbering in cytosine molecule.

kJ/mol. Moreover, standard deviation for 4511 interaction energies which are lower than 1 mh is only 1.144 kJ/mol. In the case of 1000 interaction energies which are greater than 1 mh, the standard deviation was found to be 58.100 kJ/mol. The overall standard deviation was calculated to be 11.49 kJ/mol.

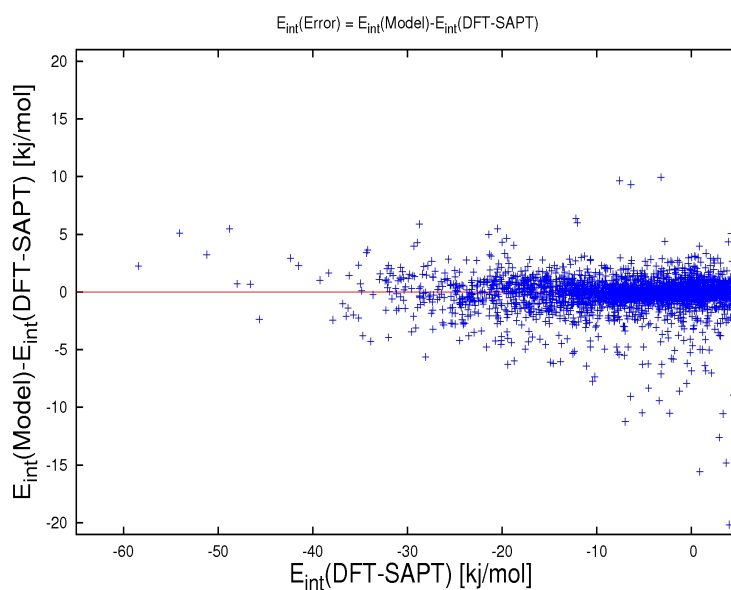
**Table 3.11:** Fitted parameters for cytosine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of first and second cytosine monomer, respectively.

| i-j | $\alpha[H]$ | $\beta[b^{-1}]$ | $C[Hb^6]$ | $c[b]$ |
|-----|-------------|-----------------|-----------|--------|
| C-C | -0.9838     | 0.8425          | 291.68    | 1.00   |
| C-H | -0.4861     | 0.4786          | 22.89     | 1.70   |
| C-O | -1.6241     | 0.9819          | 143.65    | 1.80   |
| C-N | -0.1255     | 0.4717          | 57.26     | 1.30   |
| O-O | -0.3008     | 0.1832          | 2.01      | 1.25   |
| O-N | -0.7271     | 0.8186          | 121.79    | 0.75   |
| O-H | -3.0980     | 0.7106          | 211.52    | 1.50   |
| N-N | -0.2079     | 0.3262          | 6.08      | 2.50   |
| H-N | -0.1735     | 0.4321          | 45.94     | 0.50   |
| H-H | -0.3031     | 0.5422          | 1.25      | 1.50   |



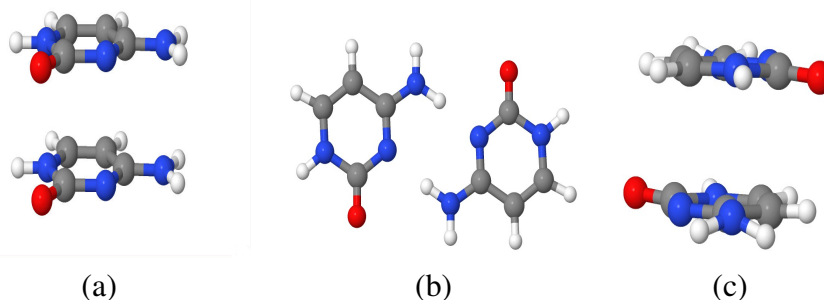
**Figure 3.11:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for Cytosine Dimer.

Our model successfully estimates the most important stacked and H-bonded conformations. Comparison of model and LPBE0AC calculations of dimer A,B and C shown in Figure 3.13 has been depicted in Figure 3.14. Here, dimer B and C seem to



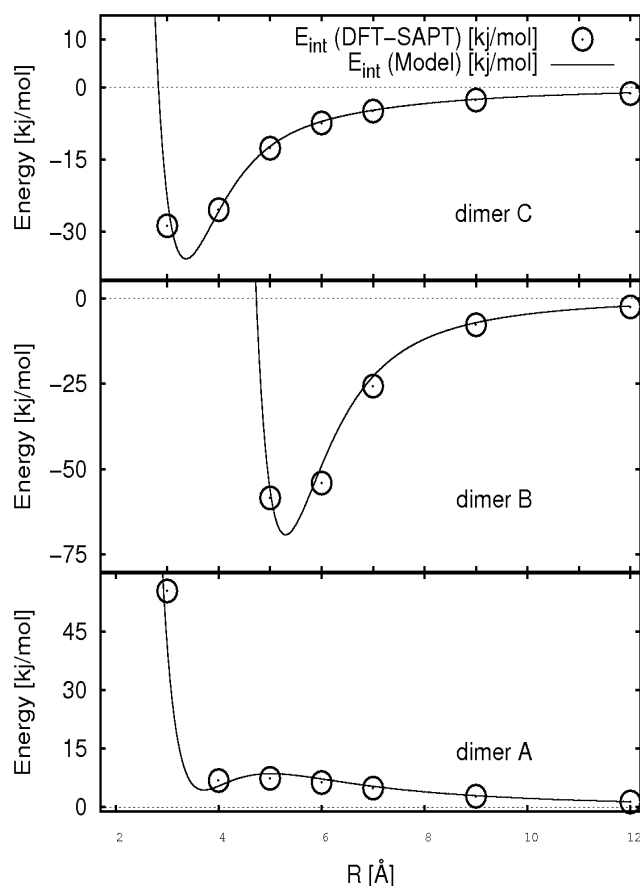
**Figure 3.12:** Fit errors in kJ/mol for cytosine dimer.

establish dimerization. For all three dimers, our model constructs an appropriate PEC, and it efficiently estimates the LPBE0AC energies.



**Figure 3.13:** Cytosine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C.

It is of course impossible to visualize a six dimensional PES. Instead, three dimensions are fixed to elucidate low energy regions (hydrogen bonded):  $\Theta = \theta = 90$  and  $\phi = 90, 270$ . Furthermore, the CMS distance is optimized (1th dimension) for all unique set of degrees (other 2-5th dimensions). Finally, we have generated the PES shown in Figure 3.15 in terms of changing three dimensions:  $\Phi$ ,  $\psi$ , and  $\phi$ . Where  $\phi = 90$ , two minimum regions can be reached. These two regions both represents the hydrogen bonded dimer B. There are also two additional minima for  $\phi = 270$ , and these two minima are different hydrogen bonded conformations.

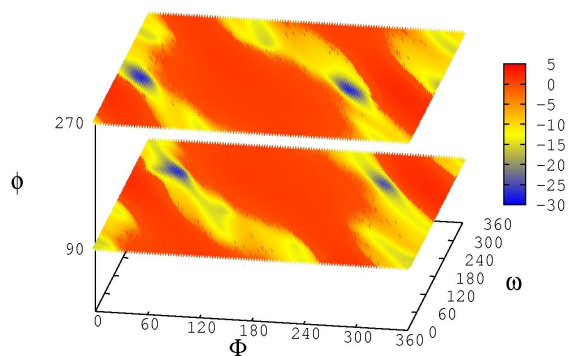


**Figure 3.14:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for cytosine Dimer A, B and C.

### 3.2.2 Guanine intermolecular energy function

The PES of the guanine dimer has been established with the same python script which was used for cytosine dimer, and the points of the PES has been obtained via a 6-dimensional grid as shown in Table 3.8 (similar to cytosine case). After the elimination of symmetric points and the points with closest atom distance less than 1.5 Å, 5916 unique guanine dimer conformations have been produced to be calculated with DFT-SAPT(LPBE0AC).

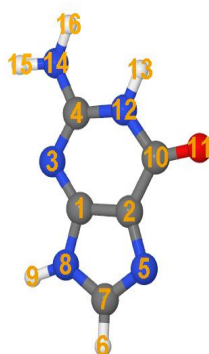
These 5916 interaction energies were fitted to an analytic function by considering 10 unique interactions shown in Table 3.12 for 256 total interactions. Similar to cytosine fit, 30 fit parameters are required for the fit of guanine dimer PES. Also again, atomic charges in the electrostatic term were calculated using the ESP method. Resulting atomic charges are shown in Table 3.13, and the atom numbering guanine molecule is shown in Figure 3.16.



**Figure 3.15:** Partial potential energy surface of cytosine dimer.

**Table 3.12:** Unique sites in guanine dimer and the required parameter number for each unique site.

| Interaction Type | Interaction Number | Parameter Number |
|------------------|--------------------|------------------|
| C-C              | 25                 | 3                |
| C-H              | 50                 | 3                |
| C-O              | 10                 | 3                |
| C-N              | 50                 | 3                |
| O-O              | 1                  | 3                |
| O-N              | 10                 | 3                |
| O-H              | 10                 | 3                |
| N-N              | 25                 | 3                |
| H-N              | 50                 | 3                |
| H-H              | 25                 | 3                |



**Figure 3.16:** Atom numbering of guanine molecule.

For guanine dimers, weighting term,  $\sigma_i$ , is set to  $\sigma_i = \frac{1}{y_i}$  for interaction energies bigger than 5 mH, and  $\sigma_i = e^{\frac{1-y_i}{4}}$  for energies lower than 5 mH. The comparison of model and LPBE0AC energies are presented in Figure 3.17 and Figure 3.18, and optimized fit is



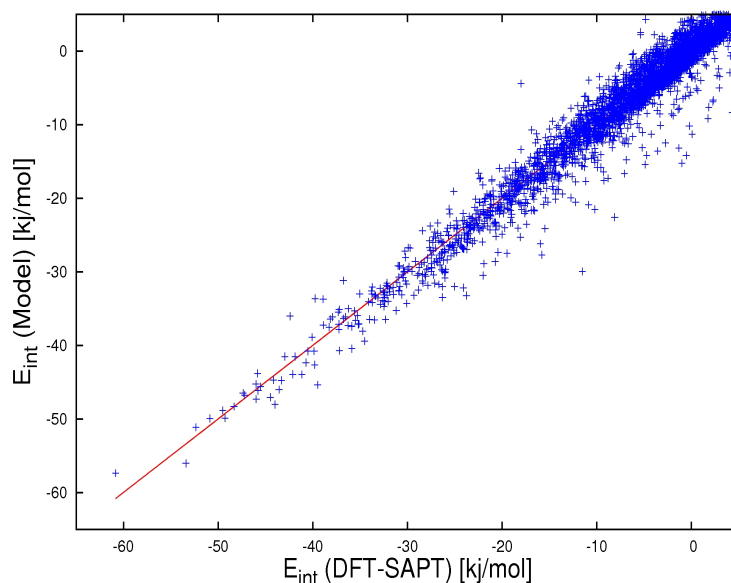
**Table 3.13:** Atomic charges of guanine molecule.

| Atom number | Atom type | Charge    |
|-------------|-----------|-----------|
| 1           | C         | 0.422291  |
| 2           | C         | 0.202353  |
| 3           | N         | -0.599975 |
| 4           | C         | 0.548248  |
| 5           | N         | -0.547018 |
| 6           | H         | 0.131875  |
| 7           | C         | 0.209237  |
| 8           | N         | -0.517636 |
| 9           | H         | 0.387561  |
| 10          | C         | 0.488148  |
| 11          | O         | -0.522023 |
| 12          | N         | -0.569437 |
| 13          | H         | 0.377703  |
| 14          | N         | -0.705275 |
| 15          | H         | 0.356716  |
| 16          | H         | 0.3400    |

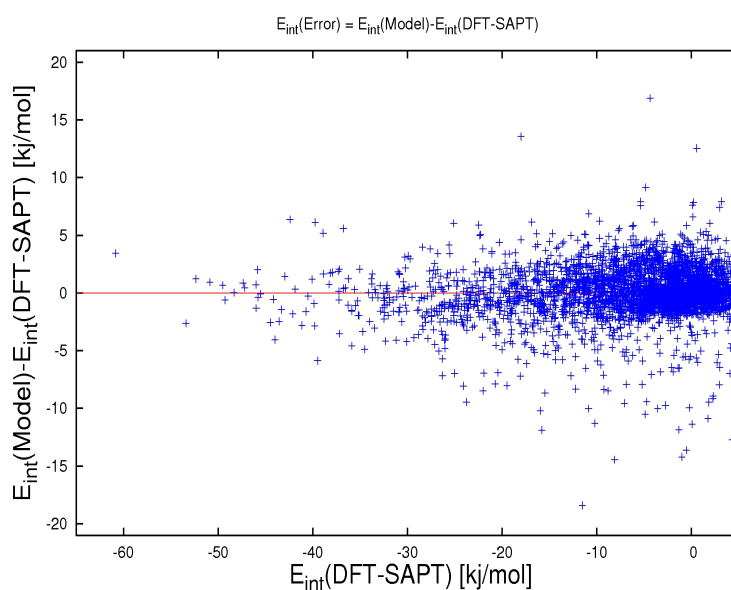
given in Table 3.14. Our fitting model is quite successful except a few points in the close-contact region of the PES. For orientations with low energies, our model is able to produce DFT-SAPT(LPBE0AC) with an approximate error of 5 kJ/mol. Moreover, standard deviation for 4903 interaction energies which are lower than 5 mh is only 1.89 kJ/mol. In the case of 1013 interaction energies which are greater than 5 mh, the standard deviation was found to be 47.15 kJ/mol. The overall standard deviation was calculated to be 19.58 kJ/mol.

**Table 3.14:** Fitted parameters for guanine potential energy function. Here, parameters have been given in "bohr" and shown as " $b$ ". " $H$ " is used for "Hartree". " $i$ " and " $j$ " represent the site of first and second guanine monomer, respectively.

| i-j | $\alpha[H]$ | $\beta[b^{-1}]$ | $C[Hb^6]$ | $c[b]$ |
|-----|-------------|-----------------|-----------|--------|
| C-C | 319.5093    | 2.28            | -26.0366  | 1.00   |
| C-H | 12.8207     | 2.16            | -3.0344   | 1.70   |
| O-C | 29.8200     | 1.77            | -30.9674  | 1.80   |
| N-C | 8.5277      | 1.44            | -49.5212  | 1.30   |
| N-H | 10.5716     | 1.92            | -16.0053  | 1.25   |
| O-N | 47.7828     | 1.81            | -40.7508  | 0.75   |
| N-N | 32.9922     | 1.62            | -70.3108  | 1.50   |
| O-H | 9.6686      | 1.81            | -26.9892  | 1.50   |
| O-O | 60.4017     | 1.87            | -70.5276  | 1.50   |
| H-H | 0.0008      | 0.28            | 3.1691    | 1.50   |



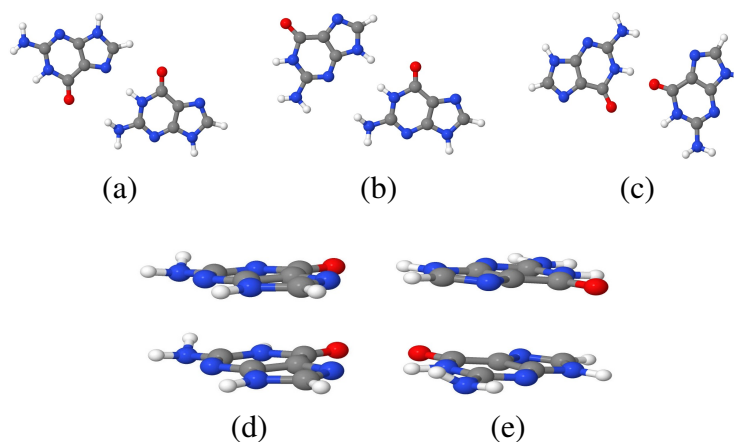
**Figure 3.17:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for guanine dimer.



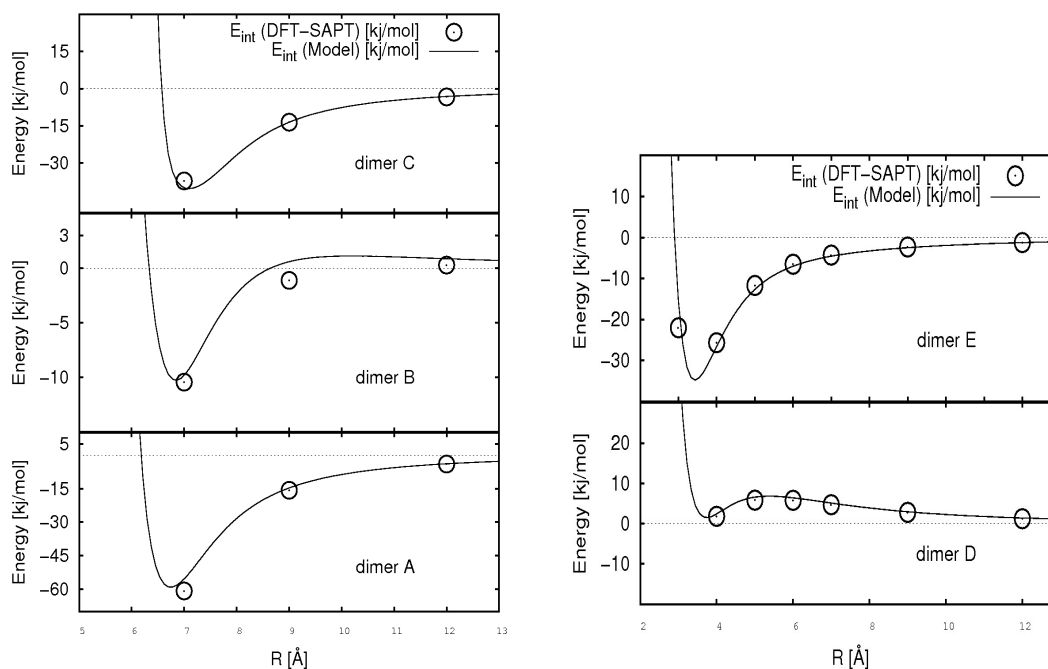
**Figure 3.18:** Fit errors in kJ/mol for guanine dimer.

Our model successfully estimates the most important stacked and H-bonded conformations. In Figure 3.20, model and LPBE0AC energies were compared for the five guanine dimers shown in Figure 3.19. Here, interaction energy of dimer D is repulsive and the others are attractive. For all five dimers, our model constructs appropriate PECs, and it efficiently estimates the LPBE0AC energies, similar to cytosine dimers.

By fixing the same dimensions with the same values as done for cytosine dimer, a partial PES of guanine dimer has been constructed as shown in Figure 3.21. Here,



**Figure 3.19:** Guanaine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C d) Dimer D e) Dimer E.

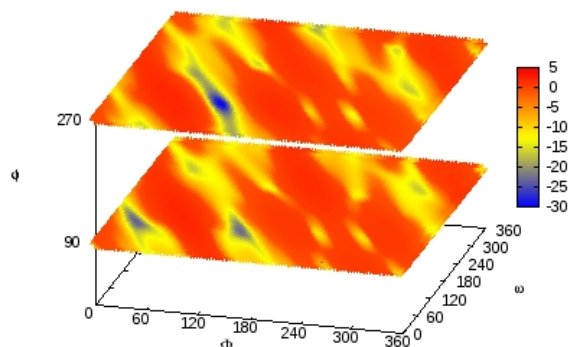


**Figure 3.20:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for guanine Dimer A, B, C, D and E.

only three dimensions:  $\Phi$  and  $\psi$ , and  $\phi$ , are allowed to change. When  $\phi=90$ , two minima regions were located, and, for  $\phi=270$ , there is an additional minimum region. These three minimum points are expected to be found by the global optimizations on the following sections.

### 3.2.3 Cytosine-Guanine intermolecular energy function

For our final dimer, we have used the same python script to establish the 6-dimensional PES of cytosine and guanine. 6-dimensional grid values are listed in Table 3.8. After



**Figure 3.21:** Partial potential energy surface of guanine dimer.

the elimination of symmetric points and the points with closest atom distance less than 1.5 Å, 7131 unique cytosine-guanine dimer conformations have been produced to be calculated with DFT-SAPT(LPBE0AC). The number of unique conformations are slightly more than the number of cytosine and guanine dimers due to obvious asymmetry of cytosine-guanine dimers.

These 7131 interaction energies were fitted to the analytic function by considering 10 unique interactions shown in Table 3.15 for 208 total interactions. Again, similar to cytosine and guanine fitting, the number of fit parameters are 30, and unique interactions are similar to cytosine and guanine cases. Also again, atomic charges required in the electrostatic term are taken from Figure 3.10 and Figure 3.16, and Table 3.10 and Table 3.13.

**Table 3.15:** Unique sites in cytosine-guanine dimer and the required parameter number for each unique site.

| Interaction Type | Interaction Number | Parameter Number |
|------------------|--------------------|------------------|
| C-C              | 20                 | 3                |
| C-H              | 45                 | 3                |
| C-O              | 9                  | 3                |
| C-N              | 35                 | 3                |
| O-O              | 1                  | 3                |
| O-N              | 80                 | 3                |
| O-H              | 10                 | 3                |
| N-N              | 15                 | 3                |
| H-N              | 40                 | 3                |
| H-H              | 25                 | 3                |

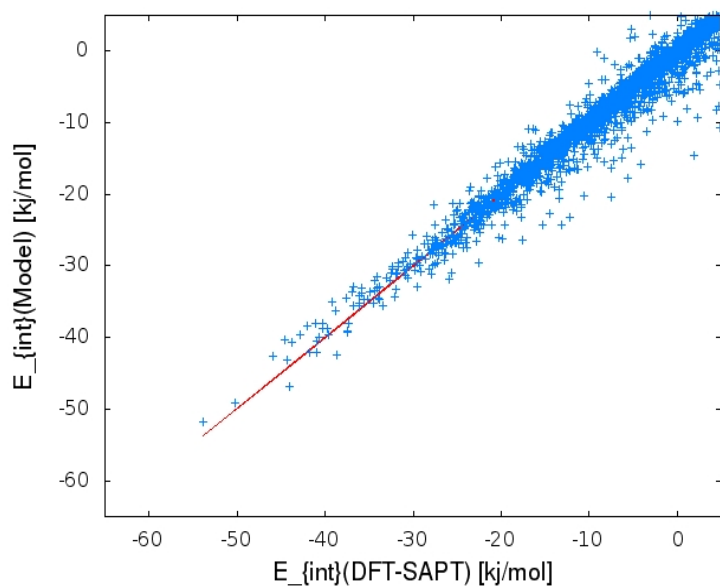
For cytosine-guanine dimers, weighting term,  $\sigma_i$ , is set to  $\sigma_i = \frac{1}{y_i}$  for interaction energies bigger than 1 mH, and  $\sigma_i = e^{\frac{1-y_i}{5}}$  for energies lower than 1 mH. The comparison of model and LPBE0AC energies are presented in Figure 3.22 and Figure 3.23, and the optimized fit is given in Table 3.16. Our fitting model is quite successful except a few points in the close-contact region of the PES. For orientations with low energies, our model is able to produce DFT-SAPT(LPBE0AC) with an approximate error of 5 kJ/mol. Moreover, standard deviation for 5056 interaction energies which are lower than 1 mH is only 1.33 kJ/mol. In the case of 2075 interaction energies which are greater than 1 mH, the standard deviation was found to be 35.59 kJ/mol. The overall standard deviation was calculated to be 19.20 kJ/mol.

**Table 3.16:** Fitted parameters for cytosine-guanine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of cytosine and guanine monomer, respectively.

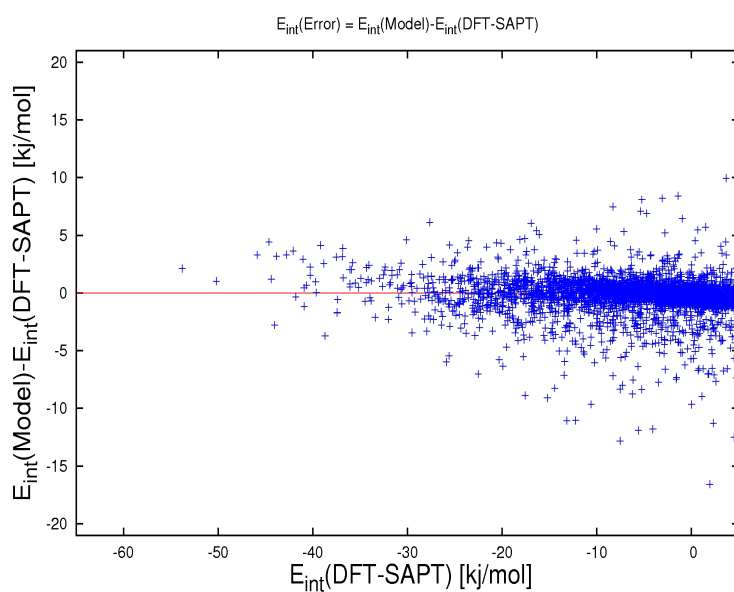
| i-j | $\alpha[H]$ | $\beta[b^{-1}]$ | $C[Hb^6]$ | $c[b]$ |
|-----|-------------|-----------------|-----------|--------|
| C-C | -.27428     | .75053          | 146.249   | 1      |
| C-H | -.00274     | .37100          | 13.257    | 1.7    |
| O-C | -.00274     | .23794          | 49.801    | 1.8    |
| N-C | -2.0473     | .42745          | 31.088    | 1.3    |
| N-H | -.11700     | .86773          | 20.650    | 1.25   |
| O-N | -.81960     | .79027          | 297.260   | 0.75   |
| N-N | -.54810     | .73655          | 316.805   | 1.5    |
| O-H | -.36016     | 1.02064         | 30.683    | 1.5    |
| O-O | -1.58809    | -.10347         | 2.825     | 1.5    |
| H-H | .34106      | 1.2575          | -5.803    | 1.5    |

Our model successfully estimates the most important stacked and H-bonded conformations. Comparison of model and LPBE0AC energies of dimer A,B and C shown in Figure 3.24 is illustrated in Figure 3.25. Here, all dimers have attractive interaction energies. For all three dimers, our model constructs appropriate PECs, and it efficiently estimates the LPBE0AC energies, similar to cytosine and guanine dimers.

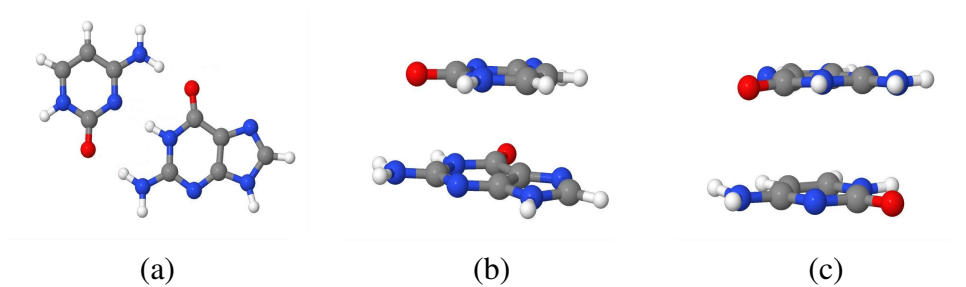
Similar to cytosine and guanine dimers, a partial PES of cytosine-guanine dimer is illustrated in Figure 3.26. Here, only the following dimensions are allowed to change:  $\Phi$  and  $\psi$ , and  $\phi$ . When  $\phi=90$ , a minimum region is located, and, for  $\phi=270$ , there is an additional minimum region. These two minimum points are expected to be found by the global optimizations on the following sections.



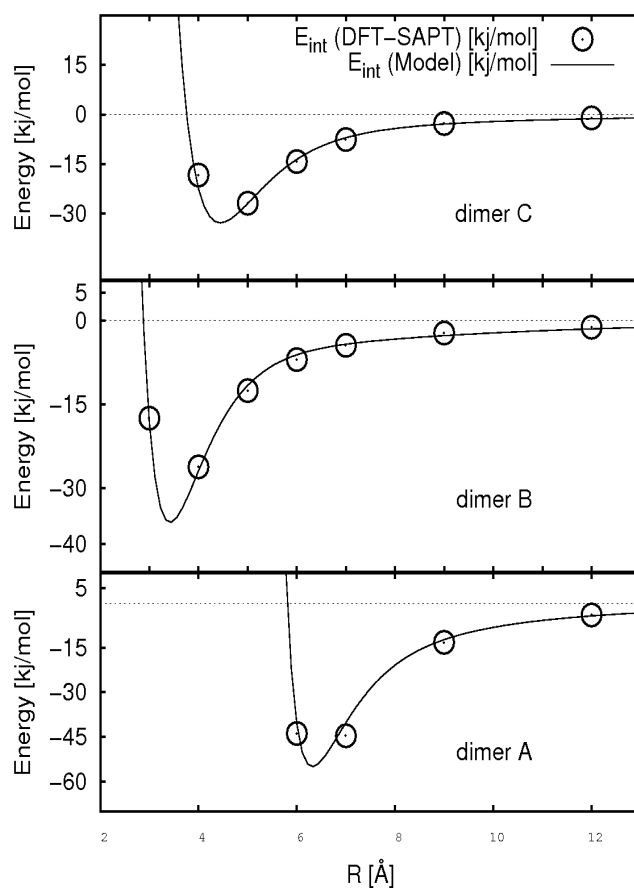
**Figure 3.22:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for cytosine-guanine dimer.



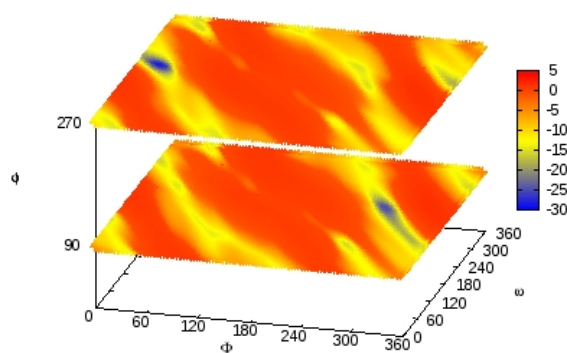
**Figure 3.23:** Fit errors in kJ/mol for cytosine-guanine dimer.



**Figure 3.24:** Cytosine-guanine dimer conformations used to compare model and ab-initio energies a) dimer A b) dimer B c) dimer C.



**Figure 3.25:** Comparison of model and DFT-SAPT(LPBE0AC) energies in kJ/mol for Cytosine-Guanine Dimer A, B and C.



**Figure 3.26:** Partial potential energy surface of cytosine-guanine dimer.

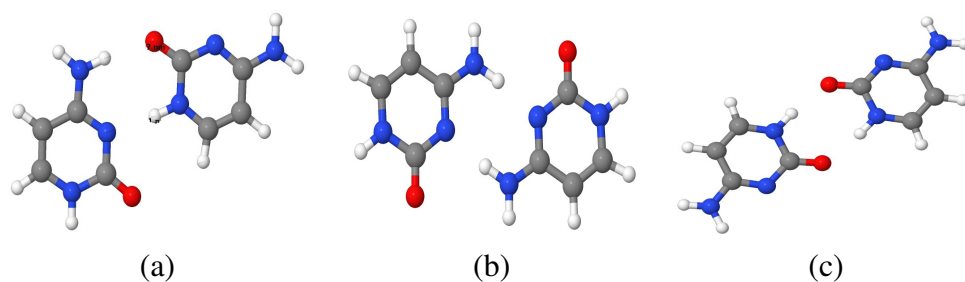
### 3.3 Global Optimization of DNA Dimer and Oligomers via Simulated Annealing

Previous phases of this work is the backbone to be able to perform global structure optimizations. Especially, by the definition of potential functions of three individual dimer structures, it is possible to uncover the most stabilized DNA dimers and oligomers. Throughout this section, while homo guanine and cytosine oligomers

are treated by a single corresponding function, cytosine-guanine oligomers uses the conjunction of three different potential functions. By comparing the results of global and local searches with ab-initio calculations, we will determine the efficiency of our potential energy functions.

### 3.3.1 Cytosine dimer and oligomers

The minimum points of PES shown in Figure 3.15 have been successfully located via our SA optimizations. In Figure 3.27, we present these dimer structures. Here, three dimers have been further optimized by the ab-initio methods of PBE/AVDZ, SCS-MP2/AVDZ and CP-SCS-MP2/AVDZ. In Table 3.17, H-bonding and CMS distances obtained from SA geometries are compared with the results of ab-initio structures. Apparently, the shortest bond distances are generated by PBE/AVDZ, while SCS-MP2 and CP-SCS-MP2 produce similar geometries. Among all SA results, dimer B seems to be the most successfully located dimer: N-H bonding difference of CP-SCS-MP2 and SA for dimer B is 0.04 Å, while the same difference is 0.09 Å for dimer A. Moreover, bonding distance of SA and ab-initio results for dimer C is longer compared to dimer A and B. Overall, SA achieves to find stable dimer structures which agree with ab-initio methods.



**Figure 3.27:** Three cytosine dimers located by SA a) Dimer A b) Dimer B c) Dimer C.

In Figure 3.18, interaction energies obtained for the SA structures shown in Figure 3.19 are compared with B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) calculations of three SA dimers. Dimer A is the most stabilized structures according to both the model and ab-initio calculations. The difference between model and LPBE0AC energy for dimer A is 9 kJ/mol. Although theoretical calculations of dimer B and C are similar, dimer C is the second most stabilized structure. However, the model produces the energy of dimer B 6.54 kJ/mol



**Table 3.17:** Comparison of the most important distances obtained from SA and other quantum mechanical methods for cytosine dimer.

| Dimer | Bond | SA   | PBE/TZVP | SCS-MP2 | CP-SCS-MP2 |
|-------|------|------|----------|---------|------------|
| A     | O-H  | 1.97 | 1.72     | 1.80    | 1.80       |
|       | N-H  | 1.91 | 1.77     | 1.81    | 1.82       |
|       | O-H  | 2.95 | 2.94     | 2.91    | 2.92       |
|       | CMS  | 5.42 | 5.35     | 5.38    | 5.39       |
| B     | N-H  | 1.96 | 1.84     | 1.91    | 1.92       |
|       | N-H  | 1.96 | 1.84     | 1.91    | 1.92       |
|       | CMS  | 5.38 | 5.38     | 5.39    | 5.39       |
| C     | O-H  | 1.92 | 1.66     | 1.72    | 1.72       |
|       | O-H  | 1.92 | 1.66     | 1.72    | 1.72       |
|       | CMS  | 6.25 | 5.98     | 6.02    | 6.03       |

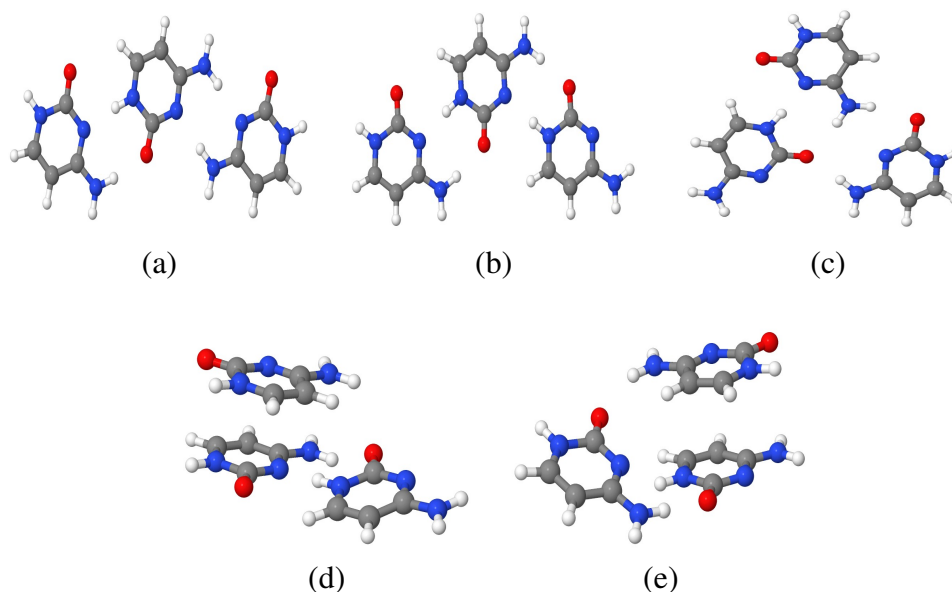
**Table 3.18:** Interaction energy calculations at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels for cytosine dimers A,B, and C shown in Figure 3.27.

| Dimer | B3LYP-D | MP2    | SCS-MP2 | PBE0AC | LPBE0AC | Model  |
|-------|---------|--------|---------|--------|---------|--------|
| A     | -90.50  | -82.22 | -75.18  | -75.98 | -81.14  | -71.81 |
| B     | -84.62  | -76.00 | -69.37  | -69.01 | -74.56  | -70.96 |
| C     | -85.61  | -78.84 | -72.64  | -73.28 | -77.45  | -64.42 |

lower than dimer C. Nevertheless, this error can be acceptable when the standard deviation of cytosine dimer model shown in Figure 3.11 is considered

Especially in Table 3.17, SA results agree well with ab-initio methods regarding dimer geometries. Thus, it is also possible to employ our model for trimer and tetramer structures also. For the cytosine trimer, in total 11 unique conformations have been located by SA. Amongst them, nine of them are planar hydrogen bonded and the remaining two are stacked orientations. All these geometries were further structurally relaxed at PBE/TZVP level. In Figure 3.28, three planar and two stacked of them have been given. Furthermore, MP2, SCS-MP2 and B3LYP-D energies of 11 trimers have been compared with model energies in Table 3.19. MP2 generates the lowest interaction energies, while SCS-MP2 and B3LYP-D are 10-20 kJ/mol and 20-30 kJ/mol higher in energy, respectively. Model agrees well with B3LYP-D rather than SCS-MP2. For all methods, including the model and PBE/TZVP, cytrimer 8 is the most stabilized trimer, and cytrimer 9 and 7 are the structures with the highest

energy. However, energy order of theoretical methods and model is slightly different for some of the orientations such as cyttrimer 4 and cyttrimer 6.

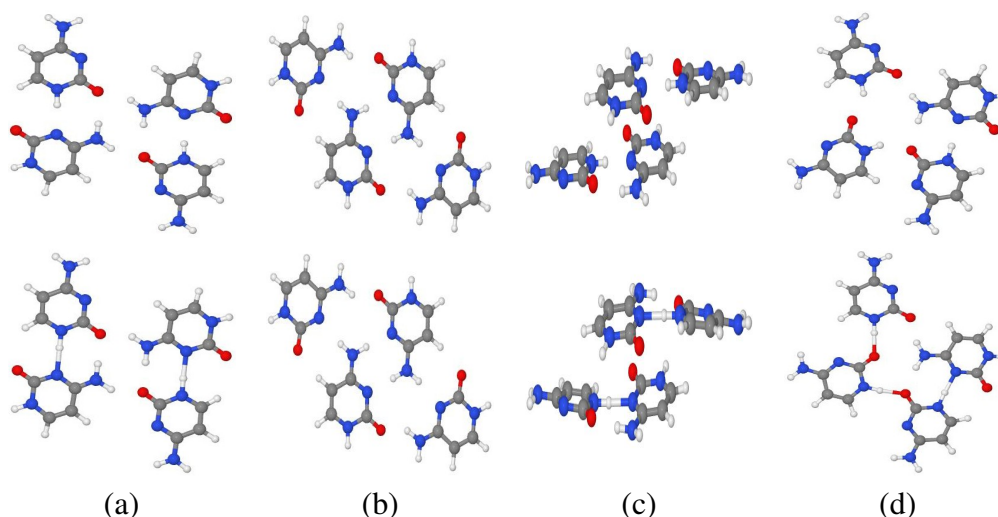


**Figure 3.28:** Three H-bonded and two stacked cytosine trimer conformations: a) cyttrimer 10 b) cyttrimer 8 c) cyttrimer 2 d) cyttrimer 7 e) cyttrimer 9.

**Table 3.19:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for cytosine trimer.

| Cyttrimer | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|-----------|---------|---------|---------|---------|--------------------|
| 1         | -159.32 | -145.93 | -133.50 | -133.39 | -1183.71715        |
| 2         | -167.99 | -154.93 | -142.71 | -131.46 | -1183.720453       |
| 3         | -155.73 | -142.39 | -129.56 | -131.35 | -1183.716267       |
| 4         | -159.42 | -145.16 | -131.18 | -136.78 | -1183.717237       |
| 5         | -170.76 | -156.28 | -142.96 | -132.27 | -1183.725825       |
| 6         | -164.92 | -149.92 | -137.08 | -132.15 | -1183.722971       |
| 7         | -143.76 | -137.87 | -117.13 | -125.85 | -1183.703239       |
| 8         | -177.20 | -161.01 | -146.77 | -140.49 | -1183.725496       |
| 9         | -141.00 | -137.30 | -117.39 | -125.96 | -1183.704879       |
| 10        | -170.06 | -153.49 | -139.78 | -139.47 | -1183.722346       |
| 11        | -149.35 | -137.62 | -125.54 | -121.45 | -1183.715636       |

Four located tetramer structures, one stacked and three planar conformations, and their PBE/TZVP geometries have been presented in Figure 3.29. Table 3.20 compares the model, MP2, SCS-MP2 and B3LYP-D interaction energies for the cytosine tetramers. It is clear from Figure 3.29 model and PBE geometries are quite similar to each other. Here, Cyttetra 1 and 3 are the most stabilized structures. Cyttetra 1 was found to be the lowest energy isomer with model, SCSMP2 and B3LYP-D. In contrast, MP2 favors



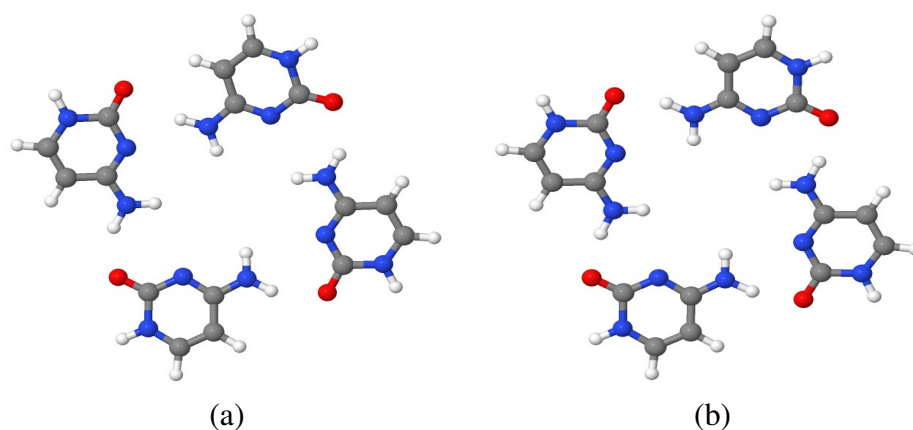
**Figure 3.29:** Stacked H-bonded and three planar tetramer conformations located by SA (top) and their relaxations at PBE/TZVP (bottom) a) cyttetra 1 b) cyttetra 2 c) cyttetra 3 d) cyttetra 4.

cyttetra3 as the most stable orientation. Similar to planar dimer and trimers, stability of tetramers is due to the hydrogen bonds. Regarding the order of energies, model agrees with MP2, while the order is the opposite for SCS-MP2 and B3LYP-D.

**Table 3.20:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for cytosine tetramer.

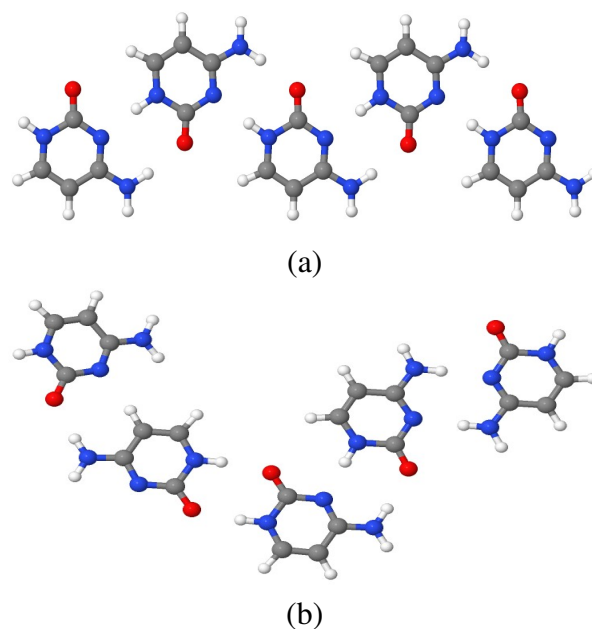
| Cyttetra | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|----------|---------|---------|---------|---------|--------------------|
| 1        | -234.74 | -214.75 | -255.32 | -207.31 | -1578.600777374    |
| 2        | -220.04 | -198.57 | -239.92 | -203.36 | -1578.595733991    |
| 3        | -237.26 | -205.69 | -250.1  | -211.91 | -1578.586760838    |
| 4        | -220.87 | -202.69 | -237.51 | -190.1  | -1578.597448346    |

The parallel interaction of four d-(TGGGCGGT) DNA sequences and the resulting H-bonded C-tetrad structure is a considerably stabilized point in the tetramer PES. Since SA executions are carried out by generated random numbers, and SA might avoid some minimum points to locate other ones; it is often possible to miss some important local minima which are close to global minimum. Unfortunately, C-tetrad has not been located with SA. Then, local minimizations employing Powell's algorithm has been tried to locate the C-tetrad. Later on, we have additionally optimized the powell result with PBE/TZVP, and compared these two structures in Figure 3.30. Similar to other SA tetramer structures, powell optimization of C-tetrad also seems in agreement with PBE/TZVP results.



**Figure 3.30:** C-tetrad structure obtained from a) Powell and b) further relaxation of Powell geometry with PBE/TZVP.

By increasing the number of cytosine monomers in a oligomer system, pentamer and even hexamer structures can be globally optimized with our potential function. In our SA simulations, we were able to locate two unique pentamer structures which are shown in Figure 3.31. These two planar filament pentamers consist of our previously investigated dimer interactions: cytpenta 1 includes the most stabilized dimer interaction shown in Figure 3.27, and cytpenta 2 includes four unique dimer interactions.

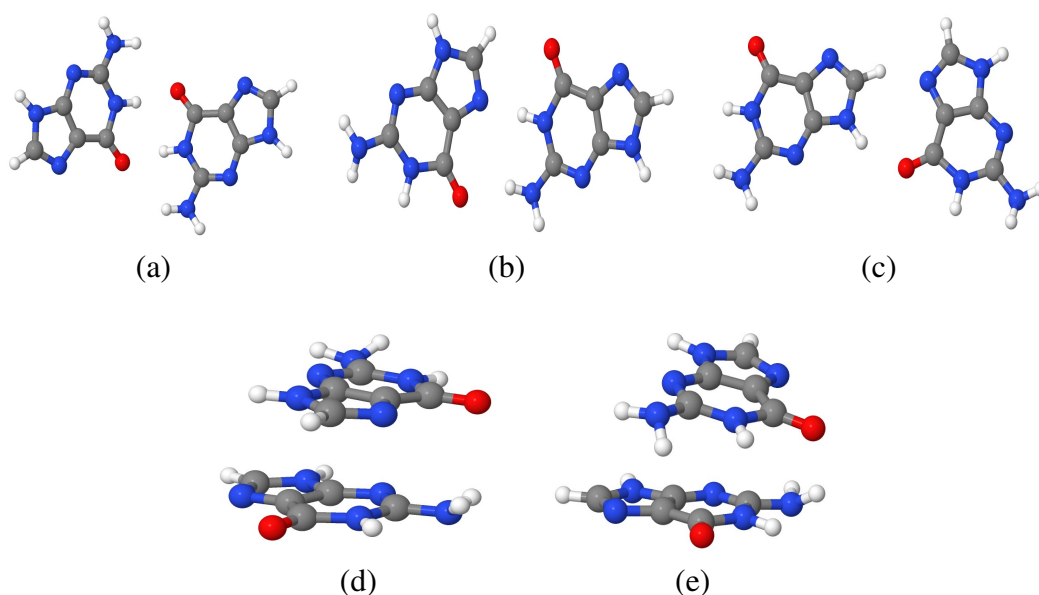


**Figure 3.31:** Two cytosine filament pentamer structures a) Cyt-penta 1 b) Cyt-penta 2.

### 3.3.2 Guanine dimer and oligomers

PES of cytosine-guanine dimer has been investigated by the SA approach and five local minima including both H-bonded and stacked orientations shown in Figure 3.32

have been located. Their corresponding interaction energies were compared with the ab-initio outcomes in Table 3.21. Among the considered methods, B3LYP-D generates the lowest interaction energies. Here, in contrast to cytosine case, order in model interaction energies were found to be the same with ab-initio methods. Within the structures shown in Figure 3.32, dimer A, which consists of four H-bonds, is the most stabilized dimer structure. Dimer A is followed by dimer C,B,D and E in the order of interaction energies. From the located cytosine-guanine isomers, especially H-bonded dimers, have been reported before [59–63]. Energy ordering reported in these studies were in agreement with our findings. However, these reported ab-initio energies were lower in energy than the model.



**Figure 3.32:** Five guanine dimers located by SA approach a) Dimer A b) Dimer B c) Dimer E d) Dimer D e) Dimer C.

**Table 3.21:** Interaction energy calculations at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels guanine dimers A,B, C, D and E.

| Dimer | B3LYP-D | MP2    | SCS-MP2 | PBE0AC | LPBE0AC | Model  |
|-------|---------|--------|---------|--------|---------|--------|
| A     | -99.24  | -88.26 | -81.31  | -84.12 | -88.59  | -82.68 |
| B     | -78.58  | -71.87 | -65.13  | -67.0  | -70.78  | -67.56 |
| C     | -80.55  | -80.96 | -65.95  | -67.83 | -72.07  | -73.62 |
| D     | -65.71  | -71.23 | -55.67  | -54.74 | -58.65  | -63.02 |
| E     | -37.69  | -34.06 | -27.96  | -30.16 | -33.57  | -43.36 |

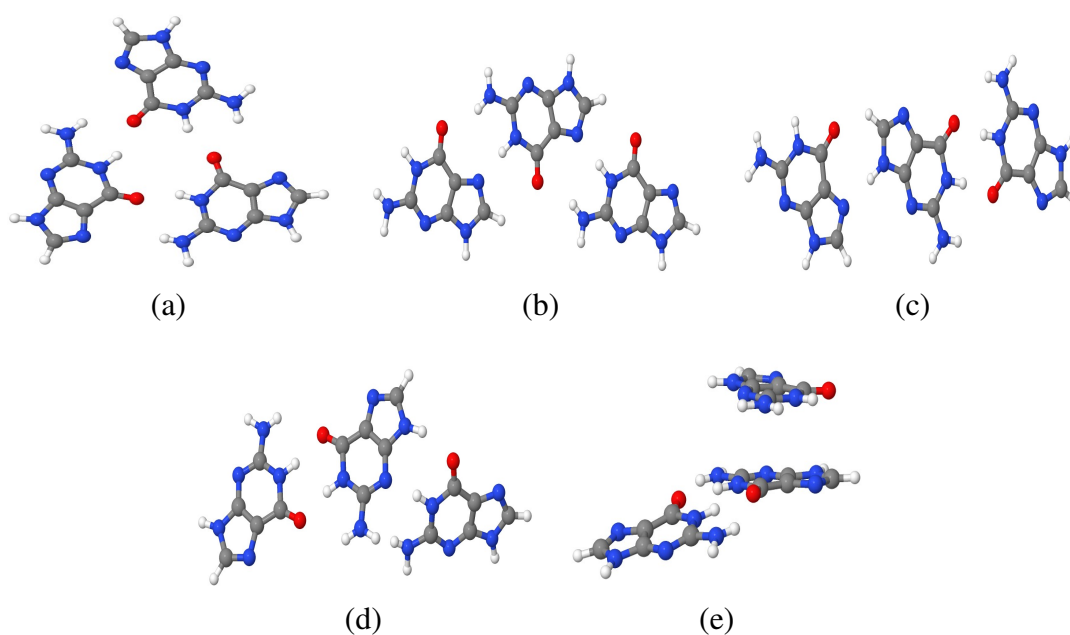
Table 3.22 compares the model geometries with their relaxed forms. Here, the most important distances in the dimers have been highlighted. Generally, PBE produces structures with shorter H-bond distances. Interestingly, PBE transforms the stacked C and D dimers to the h-bonded planar orientations in contrast to SCSMP2 and CP-SCSMP2 which are in agreement with the model. Commonly, H-bonds of Dimer B and E behave similar: some of the H-bonds get shorter, while some gets longer. For example, second O-H bond of dimer B found to be the same in all ab-initio methods. O-H bond of dimer E decreases by 0.14 Å and N-H bond increases by 0.5 Å compared to the model. In conclusion, it has been observed that the model is in better agreement with CP-SCSMP2 and the model achieves to find both local H-bonded and stacked dimers.

**Table 3.22:** Comparison of the most important distances obtained from SA and other quantum mechanical methods for guanine dimer. Note(\*): Dimer C and D transforms into Dimer A and B as a result of PBE, respectively.

| Dimer | Bond | SA   | PBE/TZVP | SCSMP2 | CP-SCS-MP2 |
|-------|------|------|----------|--------|------------|
| A     | O-H  | 1.96 | 1.66     | 1.75   | 1.75       |
|       | O-H  | 1.99 | 1.66     | 1.75   | 1.75       |
|       | CMS  | 6.16 | 6.18     | 6.21   | 6.18       |
| B     | O-H  | 2.78 | 2.51     | 2.51   | 2.50       |
|       | N-H  | 1.91 | 1.80     | 1.84   | 1.85       |
|       | O-H  | 2.20 | 2.10     | 2.19   | 2.20       |
|       | CMS  | 6.29 | 6.33     | 6.36   | 6.35       |
| C     | O-H  | 2.85 | 1.66*    | 2.64   | 2.60       |
|       | O-H  | 2.94 | 1.66*    | 2.64   | 2.60       |
|       | CMS  | 3.40 | 6.18*    | 3.08   | 3.20       |
| D     | O-H  | 2.76 | 3.58*    | 2.61   | 2.65       |
|       | O-H  | 3.44 | 2.51*    | 3.29   | 3.33       |
|       | CMS  | 3.34 | 6.34*    | 3.07   | 3.14       |
| E     | O-H  | 2.05 | 1.87     | 1.90   | 1.91       |
|       | N-H  | 2.01 | 2.61     | 2.57   | 2.61       |
|       | CMS  | 6.92 | 6.99     | 6.93   | 6.94       |

Following the global structure optimizations of guanine dimers, SA approach was used to search the PES of guanine trimer. 11 different trimer isomers, only 5 of them (four planar and one stacked) shown in Figure 3.33, have been located with this method. In Table 3.23, total energies obtained from PBE/TZVP level and the interaction energies calculated with different ab-initio methods using the model guanine trimers have been listed. The resulting PBE geometries are in agreement with the model ones. According

to all interaction and PBE total energies, guatrimer 9 is the most stabilized isomer. Similar to cytosine trimer interaction energies shown in Table 3.19, model energies of guanine trimers were obtained higher in energy than the ab-initio ones. Amongst all the methods, MP2 calculations produce the lowest interaction energies. Guatrimer 9 is followed by guatrimer 8 regarding the order of energy. However, there is a disagreement between model and ab-initio methods for the third lowest energy isomer: cyttrimer 2 and 7 were preferred with model and ab-initio methods, respectively. On the other hand, the order of interaction energy also differs among ab-initio results themselves.

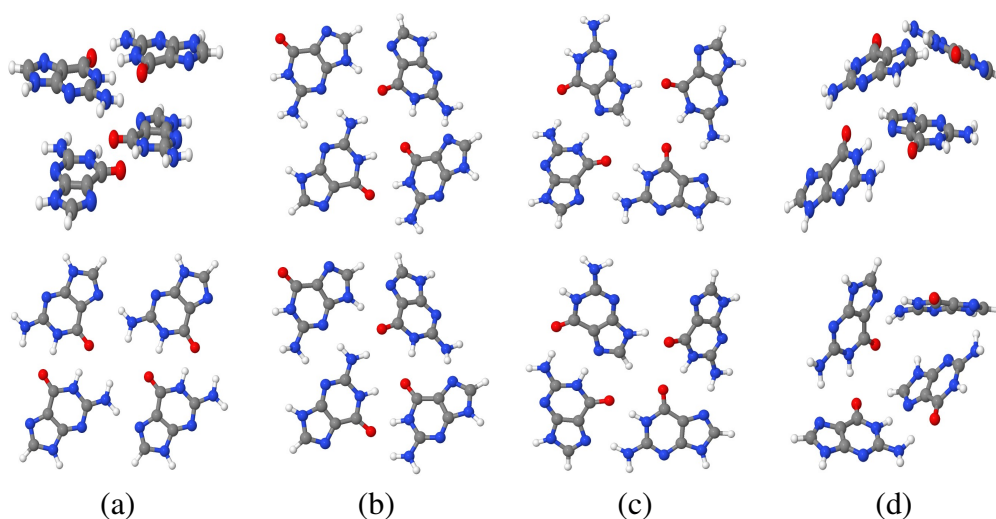


**Figure 3.33:** Three H-bonded and two stacked guanine trimer conformations: a) guatrimer 9 b) guatrimer 5 c) guatrimer 1 d) guatrimer 6 e) guatrimer 2.

Four guanine tetramers have been found with our SA simulations, and their relaxed geometries with PBE/TZVP have been given in Figure 3.34 and compared with SA ones. Similar to guanine stacked dimers, stacked guatetra 1 structure transformed into an full planar conformation after the PBE/TZVP geometry relaxation. It has been previously shown that SCSMP2 and CP-SCSMP2 do not convert the stacked guanine dimers to the planar counterparts. Even though, SCSMP2 and CP-SCSMP2 geometry optimizations have not been performed for the guanine tetramers, guanine dimer computations might be an indication that these methods might still prefer the stacked orientations for the guanine tetramers. When the interaction energies of

**Table 3.23:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for guanine trimer.

| Guatramer | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|-----------|---------|---------|---------|---------|--------------------|
| 1         | -153.74 | -140.49 | -128.05 | -124.29 | -1626.19695        |
| 2         | -154.98 | -152.94 | -126.53 | -143.40 | -1626.19386        |
| 3         | -155.08 | -149.99 | -128.36 | -131.66 | -1626.20589        |
| 4         | -157.67 | -144.98 | -130.60 | -136.46 | -1626.20158        |
| 5         | -159.05 | -147.70 | -128.91 | -131.73 | -1626.19434        |
| 6         | -164.23 | -150.22 | -137.20 | -130.80 | -1626.19863        |
| 7         | -176.27 | -158.34 | -142.67 | -139.01 | -1626.20831        |
| 8         | -179.73 | -165.93 | -151.24 | -148.69 | -1626.20550        |
| 9         | -195.99 | -182.11 | -171.32 | -154.15 | -1626.21994        |



**Figure 3.34:** Two stacked H-bonded and two planar tetramer conformations of model (top) and their PBE/TZVP (bottom) optimized geometries a) guatetra 1 b) guatetra 2 c) guatetra 3 d) guatetra 4.

guanine tetramers are considered, it has been seen that the order in the ab-initio and model varies: while guatetra 3 is the most stabilized tetramer for SCS-MP2, remaining methods suggest that guatetra 1 has the lowest interaction energy. For the guanine tetramers, our model agrees well with SCS-MP2 with an approximately 15 kJ/mol error.

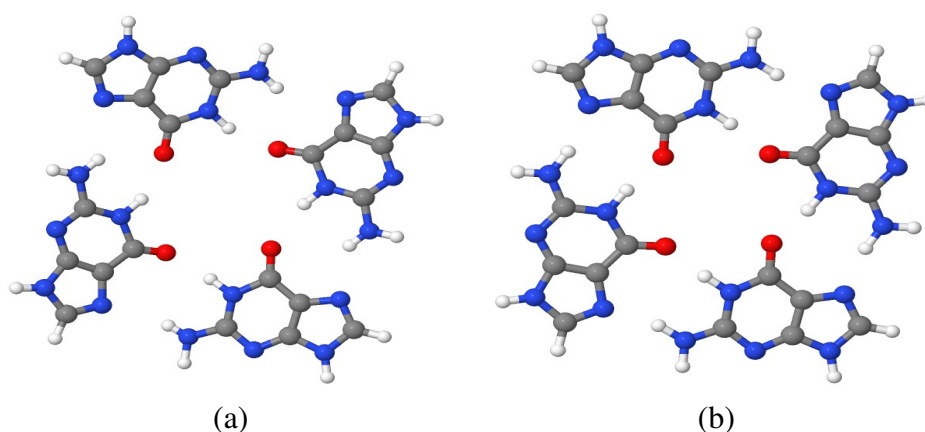
Similar to C-tetrad, guanine tetramer, G-quartet, is also a fully planar H-bonded tetramer that exist either in conjunction of cytosine, thymine and adenine tetrads. Moreover, both experimental and theoretical investigations of this conformation is available [64]. In Figure 3.35, the PBE/TZVP and powell optimization results have



**Table 3.24:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for guanine tetramer.

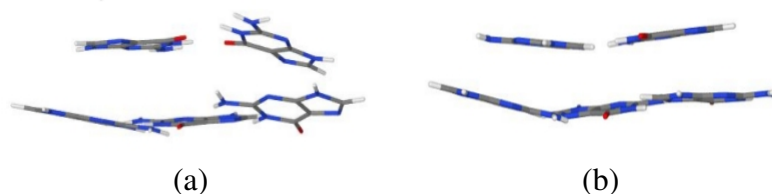
| Guatetra | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|----------|---------|---------|---------|---------|--------------------|
| 1        | -281.71 | -275.44 | -232.63 | -238.24 | -2168.296035       |
| 2        | -271.87 | -244.93 | -221.25 | -214.29 | -2168.290295       |
| 3        | -275.80 | -255.18 | -235.75 | -217.08 | -2168.296623       |
| 4        | -268.07 | -262.39 | -220.15 | -235.80 | -2168.282553       |

been compared. In PBE/TZVP orientation, there are two H-bonded interactions between the monomers in contrast to model where O-H bonding can be clearly seen but not the N-H.



**Figure 3.35:** G-quartet optimization a) Powell b) PBE/TZVP.

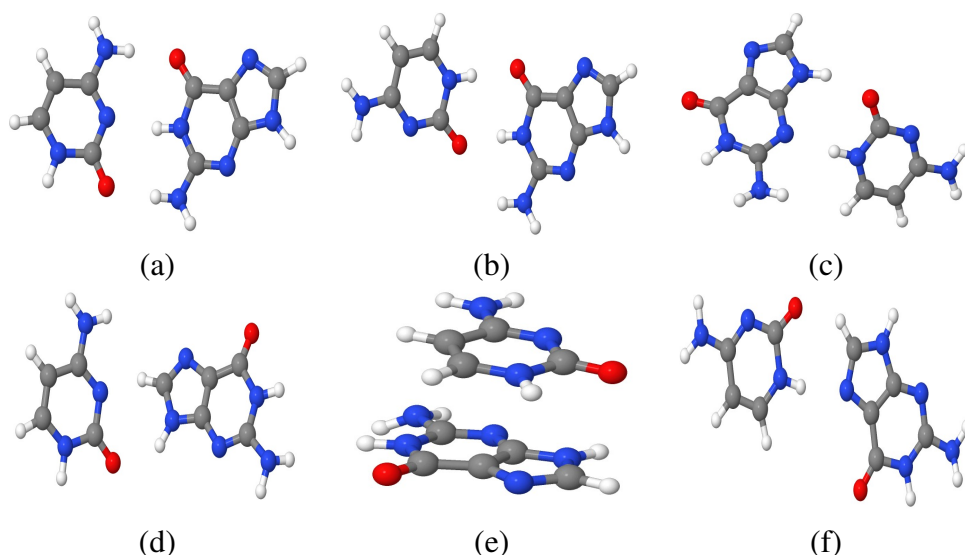
Finally, we have located two stacked guanine pentamers where two H-bonded guanine monomers are stacked with a H-bonded trimer. Interaction energies of guapenta 1 (contains the interaction seen in dimerA) and guapenta 2 (contains the interaction seen in dimer B) are -328.64 and -310.73 kJ/mol, respectively. Here, guapenta 1 is more stabilized due to superiority of dimer A over dimer B as shown in Figure 3.32.



**Figure 3.36:** Two guanine stacked pentamer structures a) Guapenta 1 b) Guapenta 2.

### 3.3.3 Cytosine-Guanine dimer and oligomers

In Figure 3.37, SA results of six unique cytosine-guanine dimers are given. Here, dimer A is a Watson-Crick base pair, and dimer E is the only stacked conformation. A Watson-Crick base pair consists of two O-H and one N-H bonds, and the remaining dimers have two H-bonds each. Table 3.25 shows a comparison of interaction energies obtained for six dimers. Similar to the results indicated in Table 3.18 and 3.21, B3LYP-D generates the lowest interaction energies; for example, dimer A is 10 kJ/mol lower in energy in B3LYP-D compared to LPBE0AC. In general, SCSMP2 and MP2 energies are close to that of PBE0AC and LPBE0AC, respectively. Similar to guanine dimers, model agrees with ab-initio calculations in terms of energy ordering. However, the model produces the highest energies, and difference between the model and LPBE0AC decreases for some high energy dimers such as dimer E or F.



**Figure 3.37:** Six cytosine-guanine dimers located by SA approach a) Dimer A b) Dimer B c) Dimer C d) Dimer D e) Dimer E f) Dimer F.

**Table 3.25:** Model and interaction energies obtained at B3LYP-D, MP2, SCS-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) levels for cytosine-guanine dimers A, B, C, D, E and F.

| Dimer | B3LYP-D | MP2    | SCS-MP2 | PBE0AC | LPBE0AC | Model  |
|-------|---------|--------|---------|--------|---------|--------|
| A     | -93.98  | -87.69 | -102.46 | -88.93 | -93.82  | -77.82 |
| B     | -81.67  | -76.16 | -89.12  | -77.81 | -81.61  | -66.70 |
| C     | -57.37  | -51.81 | -59.58  | -50.61 | -53.89  | -43.66 |
| D     | -50.09  | -38.90 | -44.31  | -38.57 | -41.70  | -38.42 |
| E     | -52.48  | -48.19 | -56.02  | -48.74 | -51.28  | -38.87 |
| F     | -40.12  | -36.22 | -42.19  | -37.49 | -40.10  | -37.56 |

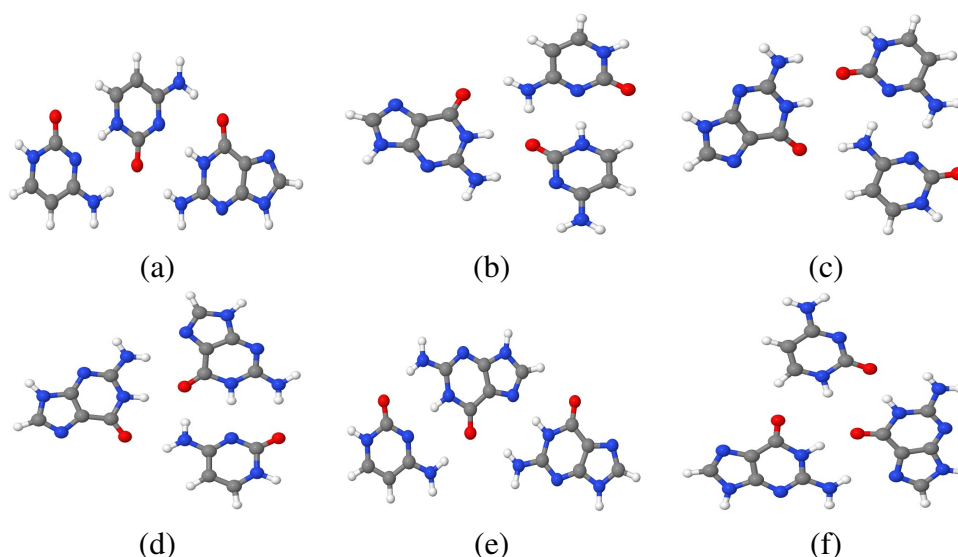
Geometry optimization results of dimers shown in Table 3.26 suggest that PBE produces the shortest bond distances, and SCS-MP2 and CP-SCS-MP2 seem to agree with each other. As seen in previous sections, we had similar results regarding homo cytosine and guanine dimers. Thus, it is obvious that PBE behaves in an exact same manner for all these DNA bases. Moreover, similar to guanine C and D dimers, stacked cytosine-guanine E orientation was transformed into H-bonded F at the PBE level, but both of SCSMP2 and MP2 preserved the stacked form. According to geometry optimizations, cytosine-guanine dimer D and F acts like guanine dimer B and E: in both dimer structures, O-H bond distance increases in ab-initio levels compared to the model. Overall, it has been seen that SA successfully produces local minima of the cytosine-guanine dimer surface.

**Table 3.26:** Comparison of the most important distances obtained from SA and other quantum mechanical methods for guanine dimer. Note(\*): Dimer E transforms into Dimer F.

| Dimer | Bond | SA   | PBE/TZVP | SCSMP2 | CP-SCS-MP2 |
|-------|------|------|----------|--------|------------|
| A     | O-H  | 2.07 | 1.67     | 1.77   | 1.78       |
|       | N-H  | 2.11 | 1.84     | 1.91   | 1.92       |
|       | O-H  | 2.11 | 1.85     | 1.92   | 1.93       |
|       | CMS  | 5.25 | 5.09     | 5.15   | 5.15       |
| B     | O-H  | 1.98 | 1.63     | 1.70   | 1.70       |
|       | O-H  | 2.10 | 1.66     | 1.76   | 1.76       |
|       | CMS  | 5.82 | 5.63     | 5.67   | 5.67       |
| C     | O-H  | 2.01 | 1.68     | 1.76   | 1.75       |
|       | N-H  | 2.31 | 1.88     | 1.93   | 1.94       |
|       | CMS  | 4.96 | 4.80     | 4.78   | 4.78       |
| D     | O-H  | 2.46 | 2.87     | 2.83   | 2.89       |
|       | N-H  | 2.20 | 1.83     | 1.89   | 1.88       |
|       | CMS  | 5.43 | 5.57     | 5.30   | 5.32       |
| E     | O-H  | 3.34 | 3.93*    | 3.21   | 3.30       |
|       | O-H  | 3.38 | 2.21*    | 3.15   | 3.16       |
|       | CMS  | 3.06 | 5.91*    | 2.77   | 2.82       |
| F     | O-H  | 2.22 | 2.23     | 2.28   | 2.28       |
|       | N-H  | 2.22 | 1.83     | 1.88   | 1.88       |
|       | CMS  | 6.18 | 5.90     | 5.93   | 5.92       |

DNA trimers consist of either two guanines and one cytosine, GGC, or 2 cytosine and one guanine, GCC. Some of these orientations are given in Figure 3.38. For cytosine-guanine oligomers, here, we have employed three unique potential energy functions together: cytosine, guanine and cytosine-guanine. Thus, we were able

to investigate this complex systems in depth, and elucidate crucial oligonucleotides. We have located five GGC and five GCC minima via our SA simulations, and optimized their geometries with PBE/TZVP. We observed that resulting PBE/TZVP geometries are in agreement with model. Interaction energies of GGC and GCC trimers shown in Table 3.27 present that GGC conformations are more stabilized than GCCs. CyGutrimer 10 and 5 has the lowest interaction energies among GGC and GCC structures, respectively. However, a minor difference between CyGutrimer 9 and 4 can be observed on their interaction energy order. Model suggests that CyGutrimer 4 is more stabilized than trimer 9, but B3LYP-D, MP2 and SCS-MP2 suggest that energy of CyGutrimer 4 is approximately 5 kJ/mol lower than CyGutrimer 9. In addition, for high interaction energies, the order of interaction energies varies in all calculations: both model and ab-initio. Nevertheless, our model is able to locate the most stabilized trimers.



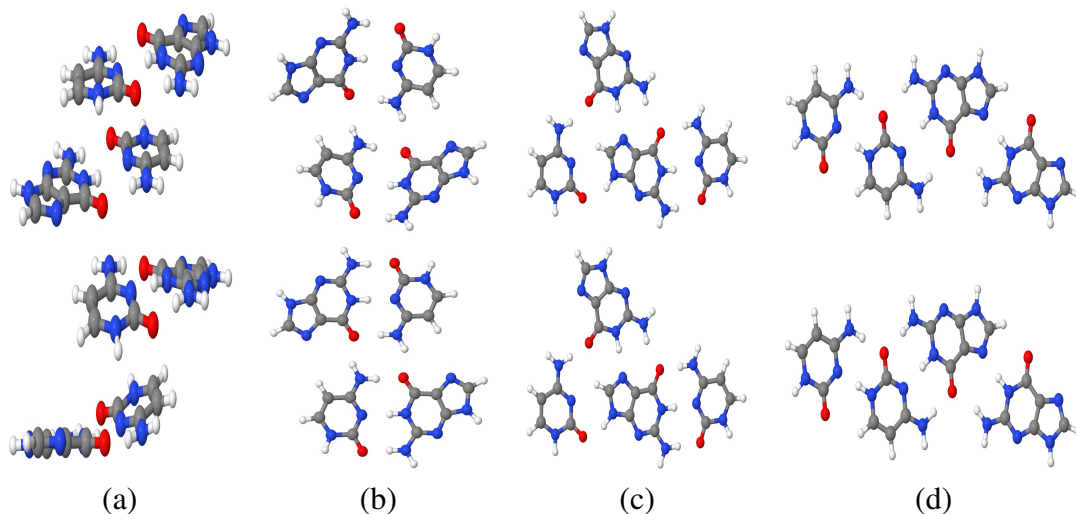
**Figure 3.38:** GGC and GCC planar trimer conformations: a) cytguatrimer 5 b) cytguatrimer 4 c) cytguatrimer 3 d) cytguatrimer 10 e) cytguatrimer 9 f) cytguatrimer 8.

In this study, tetramer conformations with two cytosine and two guanine, GCGC, have been searched with SA approach and then the resulting structures were further relaxed at PBE/TZVP level and shown in Figure 3.39. Interaction energies of SA geometries calculated at various methods are listed in Table 3.28. In these tetramers, CyGutetra 1 and 2 contain a combination of two watson-crick base pairs [65], and are the most stabilized geometries. CyGutetra 2 is a fully planar H-bonded conformation, GCGC quartet. Apparently, our model can be considered successful due to its ability to

**Table 3.27:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for GCC and GGC trimers.

| trimer | type | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|--------|------|---------|---------|---------|---------|--------------------|
| 1      | GCC  | -140.35 | -137.25 | -115.73 | -122.92 | -1331.20374        |
| 2      | GCC  | -153.15 | -142.97 | -132.50 | -108.50 | -1331.21481        |
| 3      | GCC  | -168.94 | -154.40 | -142.42 | -135.85 | -1331.21627        |
| 4      | GCC  | -177.27 | -161.78 | -149.41 | -138.54 | -1331.21776        |
| 5      | GCC  | -189.00 | -172.85 | -159.84 | -142.76 | -1331.22785        |
| 6      | GGC  | -164.15 | -159.10 | -136.34 | -140.36 | -1478.70175        |
| 7      | GGC  | -167.81 | -153.75 | -141.90 | -127.17 | -1478.71068        |
| 8      | GGC  | -178.68 | -164.97 | -153.88 | -138.09 | -1478.71630        |
| 9      | GGC  | -181.46 | -166.40 | -153.99 | -137.61 | -1478.71719        |
| 10     | GGC  | -196.78 | -180.04 | -168.13 | -147.54 | -1478.72482        |

uncover watson-crick pairs as a result of global optimizations. Energy ordering was obtained different in ab-initio methods: while MP2 and model favor cygutetra 1 as the lowest energy conformation, SCSMP2 and B3LYP-D find cygutetra 2 as the lowest energy minimum structure.

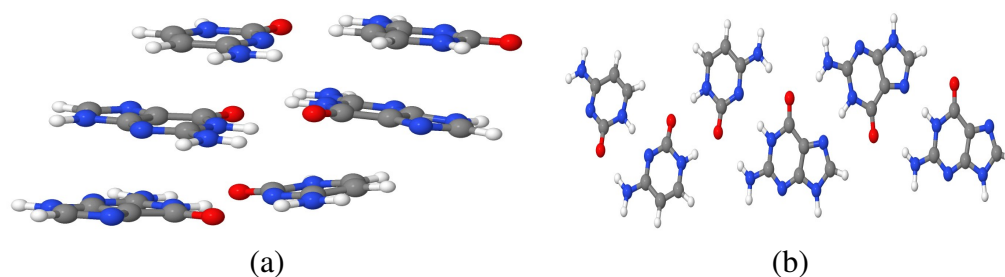


**Figure 3.39:** A stacked H-bonded and three planar tetramer conformations of model (top) and its PBE/TZVP (bottom) optimized geometries a) cygutetra 1 b) cygutetra 2 c) cygutetra 3 d) cygutetra 4.

Finally, cytosine-guanine oligomers which consist of three guanine and three cytosine, GCGCGC, are also investigated by our SA algorithm. As a result, a GCGCGC conformation with three stacked dimers, GCGCGC 1, and a fully planar filament, GCGCGC 2, have been located and presented in Figure 3.40. Here, GCGCGC 1

**Table 3.28:** Total energies obtained at PBE/TZVP level and interaction energies calculated with model, B3LYP-D, MP2, SCS-MP2 levels [kJ/mol] for GCGC tetramers.

| CyGutetra | B3LYP-D | MP2     | SCS-MP2 | Model   | PBE/TZVP [hartree] |
|-----------|---------|---------|---------|---------|--------------------|
| 1         | -280.97 | -275.32 | -238.44 | -226.22 | -1873.29979        |
| 2         | -286.87 | -266.98 | -246.56 | -213.16 | -1873.31751        |
| 3         | -268.99 | -249.48 | -231.11 | -202.57 | -1873.30460        |
| 4         | -269.26 | -246.16 | -226.51 | -206.53 | -1873.30355        |



**Figure 3.40:** A stacked and a filament hexamer structures a) GCGCGC 1 b) GCGCGC 2.

is a combination of guanine dimer A and B, and cytosine dimer A. These dimer conformations are found to be the most stabilized dimers on their corresponding surfaces. Also, GCGCGC 2 includes planarly merged guanine and cytosine trimers. The interaction energies of GCGCGC 1 and 2 are -360.16 and -330.40 kJ/mol, respectively.

## 4. CONCLUSION

In this study, Force fields of cytosine, guanine and cytosine-guanine interactions have been developed. For this purpose, first, three unique dimer surface have been fitted to an analytical function. Here, dimer surfaces could not be calculated with CCSD(T) due to the requirement of massive computational resources. Hence, at the first stage, benchmark interaction energy computations were completed using MP2, SCSMP2, B3LYP-D and DFT-SAPT to find the method which is in best agreement with CCSD(T). Although MP2 gives similar results to CCSD(T) for H-bonded conformations, its performance is not good for the stacked orientations. SCS-MP2 seems to correct MP2 results, but overestimates energies compared to CCSD(T). Amongst the considered methods, it has been found that DFT-SAPT(LPBE0AC) agrees as best with CCSD(T). Following the selection of the method, dimer surfaces containing 6000-7000 points were calculated with DFT-SAPT(LPBE0AC) using aug-cc-pVDZ basis set.

Resulting interaction energies have been used to fit the parameters of intermolecular potential function which includes repulsion, dispersion and electrostatic terms. We have employed the non-linear least square method of Levenberg-Marquardt to fit the surfaces. Fitting of dimer surfaces are resulted in 0.44, 0.72 and 0.52 mH standart deviations for orientations that has an energy lower than 1 mH for cytosine, guanine and cytosine-guanine, respectively. It has also been observed that model is in very good agreement with DFT-SAPT energies.

After the generation of force fields, PES of these energy expressions were globally searched with SA approach. These SA optimizations were completed for dimer, trimer, tetramer and even bigger clusters of three surfaces. The resulting SA structures were then further relaxed at pbe, scsmp2 and cpsscsm2 levels using TZVP basis set in order to see the quality of the SA structures. Furthermore, SA model interaction energies were compared to that of mostly MP2, SCSMP2 and B3LYP-D. For each cluster, SA was able to find many different local minima. However, energy ordering

in these minima were found to be slightly different in each method. Most of the cases, model interaction energies were in agreement with SCSMP2. The globally optimized structures mostly contain H-bonded interactions and there were only a few stacked orientations which were transformed to the planar ones after PBE/TZVP geometry optimizations whereas scsmp2 and CP-SCSMP2 kept the stacked orientations as they are. Overall, SA structures were in quite in agreement with ab-initio optimized ones.

Along with global optimizations of oligonucleotides, we have locally optimized some important experimentally determined DNA structures: C-tetrad, G-quartet and GCGC-quartet. All these conformations have been optimized with our powell algorithm, and it has been observed that this local minima exist in the corresponding PES with a small structural differences.

The successful results of our intermolecular interaction models show that such potential functions could be used to elucidate the structure of more bigger oligonucleotides. Furthermore, they can also be exploited to investigate the interaction between metal and DNA bases.



## REFERENCES

- [1] **Jurecka, P., Šponer, J. and Hobza, P.** (2004). *The Journal of Physical Chemistry B*, **108**(17), 5466–5471.
- [2] **Czyżnikowska, Ż. and Zaleśny, R.** (2009). *Biophysical Chemistry*, **139**(2), 137–143.
- [3] **Cysewski, P., Czyżnikowska, Ż., Zaleśny, R. and Czeleń, P.** (2008). *Physical Chemistry Chemical Physics*, **10**(19), 2665–2672.
- [4] **Acosta-Silva, C., Branchadell, V., Bertran, J. and Oliva, A.** (2010). *The Journal of Physical Chemistry B*, **114**(31), 10217–10227.
- [5] **Schnier, P.D., Klassen, J.S., Strittmatter, E.F. and Williams, E.R.** (1998). *Journal of the American Chemical Society*, **120**(37), 9605–9613.
- [6] **Guerra, C.F., van der Wijst, T., Poater, J., Swart, M. and Bickelhaupt, F.M.** (2010). *Theoretical Chemistry Accounts*, **125**(3-6), 245–252.
- [7] **Lim, K.W., Alberti, P., Guédin, A., Lacroix, L., Riou, J.F., Royle, N.J., Mergny, J.L. and Phan, A.T.** (2009). *Nucleic acids research*, **37**(18), 6239–6248.
- [8] **Cheung, W., Pontoriero, F., Taratula, O., Chen, A.M. and He, H.** (2010). *Advanced drug delivery reviews*, **62**(6), 633–649.
- [9] **Gu, J. and Leszczynski, J.** (2002). *Chemical physics letters*, **351**(5), 403–409.
- [10] **Patel, P., Bhavesh, N.S. and Hosur, R.** (2000). *Biochemical and biophysical research communications*, **270**(3), 967–971.
- [11] **Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R.A. and Patel, D.J.** (1998). *Journal of molecular biology*, **282**(3), 619–636.
- [12] **Gu, J. and Leszczynski, J.** (2000). *The Journal of Physical Chemistry A*, **104**(31), 7353–7358.
- [13] **Boon, E.M., Ceres, D.M., Drummond, T.G., Hill, M.G. and Barton, J.K.** (2000). *Nature biotechnology*, **18**(10), 1096–1100.
- [14] **Wang, J.** (2000). *Nucleic Acids Research*, **28**(16), 3011–3016.
- [15] **McKendry, R., Zhang, J., Arntz, Y., Strunz, T., Hegner, M., Lang, H.P., Baller, M.K., Certa, U., Meyer, E., Güntherodt, H.J. et al.** (2002). *Proceedings of the National Academy of Sciences*, **99**(15), 9783–9788.

- [16] Otero, R., Schöck, M., Molina, L.M., Lægsgaard, E., Stensgaard, I., Hammer, B. and Besenbacher, F. (2005). *Angewandte Chemie International Edition*, **44**(15), 2270–2275.
- [17] Otero, R., Lukas, M., Kelly, R.E., Xu, W., Lægsgaard, E., Stensgaard, I., Kantorovich, L.N. and Besenbacher, F. (2008). *Science*, **319**(5861), 312–315.
- [18] Erdmann, M., David, R., Fornof, A.R. and Gaub, H.E. (2010). *Nature chemistry*, **2**(9), 745–749.
- [19] Furukawa, M., Tanaka, H. and Kawai, T. (2001). *The Journal of Chemical Physics*, **115**, 3419.
- [20] Kelly, R. and Kantorovich, L. (2006). *Journal of Materials Chemistry*, **16**(20), 1894–1905.
- [21] Frankel, D., Chen, Q. and Richardson, N. (2006). *The Journal of chemical physics*, **124**, 204704.
- [22] Kelly, R.E., Lukas, M., Kantorovich, L.N., Otero, R., Xu, W., Mura, M., Lægsgaard, E., Stensgaard, I. and Besenbacher, F. (2008). *The Journal of chemical physics*, **129**, 184707.
- [23] Kelly, R.E., Xu, W., Lukas, M., Otero, R., Mura, M., Lee, Y.J., Lægsgaard, E., Stensgaard, I., Kantorovich, L.N. and Besenbacher, F. (2008). *Small*, **4**(9), 1494–1500.
- [24] Lukas, M., Kelly, R.E., Kantorovich, L.N., Otero, R., Xu, W., Laegsgaard, E., Stensgaard, I. and Besenbacher, F. (2009). *The Journal of chemical physics*, **130**, 024705.
- [25] Perdigão, L.M., Staniec, P.A., Champness, N.R., Kelly, R., Kantorovich, L. and Beton, P.H. (2006). *Physical Review B*, **73**(19), 195423.
- [26] Preuss, M. and Bechstedt, F. (2008). *Surface Science*, **602**(9), 1643–1649.
- [27] Xu, W., EA Kelly, R., Otero, R., Schöck, M., Lægsgaard, E., Stensgaard, I., Kantorovich, L.N. and Besenbacher, F. (2007). *Small*, **3**(12), 2011–2014.
- [28] Krull, C., Valencia, S., Pascual, J. and Theis, W. (2009). *Applied Physics A*, **95**(1), 297–301.
- [29] Maleki, A., Alavi, S. and Najafi, B. (2011). *The Journal of Physical Chemistry C*, **115**(45), 22484–22494.
- [30] Grimme, S. (2003). *The Journal of chemical physics*, **118**, 9095.
- [31] Hohenberg, P. and Kohn, W. (1964). *Physical Review*, **136**(3B), B864.
- [32] Jeziorski, B., Moszynski, R. and Szalewicz, K. (1994). *Chemical Reviews*, **94**(7), 1887–1930.

- [33] **Williams, H.L. and Chabalowski, C.F.** (2001). *The Journal of Physical Chemistry A*, **105**(3), 646–659.
- [34] **Heßelmann, A. and Jansen, G.** (2002). *Chemical physics letters*, **357**(5), 464–470.
- [35] **Heßelmann, A. and Jansen, G.** (2002). *Chemical physics letters*, **362**(3), 319–325.
- [36] **Heßelmann, A. and Jansen, G.** (2003). *Chemical physics letters*, **367**(5), 778–784.
- [37] **Misquitta, A.J., Podeszwa, R., Jeziorski, B. and Szalewicz, K.** (2005). *The Journal of chemical physics*, **123**, 214103.
- [38] **Werner, H., Knowles, P., Lindh, R., Manby, F., Schütz, M., Celani, M., Korona, T., Mitrushenkov, A., Rauhut, G., Adler, T. et al.** MOLPRO, version 2009.1; A Package of Ab Initio Programs; University College Cardiff Consultants Limited: Wales, UK, 2009.
- [39] **Hesselmann, A., Jansen, G. and Schütz, M.** (2005). *The Journal of chemical physics*, **122**, 014103.
- [40] **Grimme, S.** (2003). *The Journal of chemical physics*, **118**, 9095.
- [41] **Ahlrichs, R., Bär, M., Häser, M., Horn, H. and Kölmel, C.** (1989). *Chemical Physics Letters*, **162**(3), 165–169.
- [42] **Grimme, S.** (2004). *Journal of computational chemistry*, **25**(12), 1463–1473.
- [43] **Grimme, S.** (2006). *Journal of computational chemistry*, **27**(15), 1787–1799.
- [44] **Weigend, F.** (2002). *Physical Chemistry Chemical Physics*, **4**(18), 4285–4291.
- [45] **Weigend, F., Köhn, A. and Hättig, C.** (2002). *The Journal of chemical physics*, **116**, 3175.
- [46] **Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.** (1992). *Numerical Recipes in FORTRAN 77: Volume 1, Volume 1 of Fortran Numerical Recipes: The Art of Scientific Computing*, volume 1, Cambridge university press.
- [47] **Sathyamurthy, N.** (1985). *Computer Physics Reports*, **3**, 1–69.
- [48] **Varshni, Y.P.** (1957). *Reviews of Modern Physics*, **29**(4), 664.
- [49] **Steele, D., Lippincott, E.R. and Vanderslice, J.T.** (1961). Comparative study of empirical internuclear potential functions, **Technical Report**, Maryland. Univ., College Park; Maryland. Univ., College Park. Inst. of Molecular Physics.
- [50] **Leforestier, C., Tekin, A., Jansen, G. and Herman, M.** (2011). *The Journal of chemical physics*, **135**(23), 234306–234306.
- [51] **Arora, J.** (2004). *Introduction to optimum design*, Academic Press.

- [52] **Marquardt, D.W.** (1963). *Journal of the Society for Industrial & Applied Mathematics*, **11**(2), 431–441.
- [53] **Corana, A., Marchesi, M., Martini, C. and Ridella, S.** (1987). *ACM Transactions on Mathematical Software (TOMS)*, **13**(3), 262–280.
- [54] **Brent, R.P.** (1973). *Algorithms for minimization without derivatives*, Courier Dover Publications.
- [55] **Sütay, B., Tekin, A. and Yurtsever, M.** (2012). *Theoretical Chemistry Accounts*, **131**(2), 1–13.
- [56] **Sánchez-García, E., Mardyukov, A., Tekin, A., Crespo-Otero, R., Montero, L.A., Sander, W. and Jansen, G.** (2008). *Chemical Physics*, **343**(2), 168–185.
- [57] **Sherrill, C.D., Takatani, T. and Hohenstein, E.G.** (2009). *The Journal of Physical Chemistry A*, **113**(38), 10146–10159.
- [58] **Bak, K.L., Jørgensen, P., Olsen, J., Helgaker, T. and Klopper, W.** (2000). *The Journal of Chemical Physics*, **112**, 9229.
- [59] **Kelly, R., Lee, Y. and Kantorovich, L.** (2005). *The Journal of Physical Chemistry B*, **109**(46), 22045–22052.
- [60] **Šponer, J., Leszczynski, J. and Hobza, P.** (1996). *The Journal of Physical Chemistry*, **100**(13), 5590–5596.
- [61] **Kabelác, M. and Hobza, P.** (2001). *The Journal of Physical Chemistry B*, **105**(24), 5804–5817.
- [62] **Šponer, J., Jurecka, P. and Hobza, P.** (2004). *Journal of the American Chemical Society*, **126**(32), 10142–10151.
- [63] **Šponer, J., Leszczynski, J. and Hobza, P.** (1996). *The Journal of Physical Chemistry*, **100**(5), 1965–1974.
- [64] **Setnicka, V., Novy, J., Bohm, S., Sreenivasachary, N., Urbanova, M. and Volka, K.** (2008). *Langmuir*, **24**(14), 7520–7527.
- [65] **Špacková, N., Berger, I. and Šponer, J.** (2001). *Journal of the American Chemical Society*, **123**(14), 3295–3307.

## CURRICULUM VITAE



**Name Surname:** Artür Manukyan

**Place and Date of Birth:** ISTANBUL 1987

**Address:** Ergenekon str. Celebigil apt. No: 59 Level: 16 Pangalti Şişli Istanbul

**E-Mail:** artur-man@hotmail.com

**B.Sc.:** Statistics, Yıldız Technical University

### PUBLICATIONS/PRESENTATIONS ON THE THESIS

- Manukyan A., Tekin A., 2012: Potential Energy Surface of Cytosine Dimer *Theoretical and Computational Chemistry for the Modeling of Biochemical Systems: From Theory to Applications*, July, 2012, Girona, Spain.
- Manukyan A., Tekin A., 2012: Intermolecular Interactions in Cytosine Dimer *26th Annual National Conference of Chemistry*, October, 2012, Mugla, Turkey.