





**ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE**

---

**SYMMETRY-ADAPTED PERTURBATION THEORY POTENTIALS  
FOR DNA BASES**



**M.Sc. THESIS**

**Armağan KARATOSUN**

**Computational Science and Engineering Department**

**Computational Science and Engineering Master Programme**

**JUNE 2018**



**ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE**

---

**SYMMETRY-ADAPTED PERTURBATION THEORY POTENTIALS  
FOR DNA BASES**



**M.Sc. THESIS**

**Armağan KARATOSUN  
(702141001)**

**Computational Science and Engineering Department**

**Computational Science and Engineering Master Programme**

**Thesis Advisor: Assoc. Prof. Adem Tekin**

**JUNE 2018**



**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ**

**DNA BAZLARI İÇİN SİMETRİ UYUMLU PERTURBASYON  
KURAMI POTANSİYELLERİ**

**YÜKSEK LİSANS TEZİ**

**Armağan KARATOSUN  
(702141001)**

**Hesaplamalı Bilim ve Mühendislik Anabilim Dalı**

**Hesaplamalı Bilim ve Mühendislik Yüksek Lisans Programı**

**Tez Danışmanı: Assoc. Prof. Adem Tekin**

**HAZİRAN 2018**





Armağan KARATOSUN, a M.Sc. student of ITU Informatics Institute Engineering and Technology 702141001 successfully defended the thesis entitled “SYMMETRY-ADAPTED PERTURBATION THEORY POTENTIALS FOR DNA BASES”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

**Thesis Advisor :**    **Assoc. Prof. Adem Tekin** .....  
Istanbul Technical University

**Jury Members :**    **Assoc. Prof. Adem Tekin** .....  
Istanbul Technical University

**Prof. Fethiye Aylin Sungur** .....  
Istanbul Technical University

**Assoc. Prof. Arzu Hatipoğlu** .....  
Yıldız Technical University

**Date of Submission :**    **4 May 2018**

**Date of Defense :**        **6 June 2018**





*To my supportive family,*



## **FOREWORD**

First, I would like to thank my advisor, Adem Tekin, for giving me the privilege to work with him. Secondly, I would like to thank TUBITAK for funding our project, which will eventually become the basis of this thesis. Finally, I would like to thank my family for their unconditional love and support.

June 2018

Armağın KARATOSUN  
(Engineer)





## TABLE OF CONTENTS

	<u>Page</u>
<b>FOREWORD</b> .....	<b>ix</b>
<b>TABLE OF CONTENTS</b> .....	<b>xi</b>
<b>ABBREVIATIONS</b> .....	<b>xiii</b>
<b>LIST OF TABLES</b> .....	<b>xv</b>
<b>LIST OF FIGURES</b> .....	<b>xvii</b>
<b>SUMMARY</b> .....	<b>xix</b>
<b>ÖZET</b> .....	<b>xxi</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>2. METHOD</b> .....	<b>3</b>
2.1 Ab-initio Methods .....	3
2.1.0.1 Moeller-Plesset perturbational theory .....	3
2.1.0.2 Coupled cluster .....	4
2.1.1 Density functional theory .....	5
2.1.2 Symmetry adapted perturbation theory and DFT-SAPT .....	6
2.1.3 Investigation of DNA bases via ab-initio methods .....	6
2.2 Fitting Details .....	8
2.3 Global and Local Geometry Optimization .....	10
2.3.1 Simulated annealing .....	10
<b>3. RESULTS AND DISCUSSION</b> .....	<b>13</b>
3.1 Selection of the Theoretical Method for PES.....	13
3.1.1 Adenine dimers.....	13
3.1.2 Thymine dimers.....	17
3.1.3 Adenine-Thymine dimers .....	20
3.2 Fitting Potential Energy Surfaces of DNA Bases .....	24
3.2.1 Adenine intermolecular function.....	24
3.2.2 Thymine and Adenine-Thymine intermolecular functions .....	25
3.3 Global Optimization of the DNA Bases via Simulated Annealing .....	28
3.3.1 Adenine clusters .....	28
3.3.1.1 Dimer .....	28
3.3.1.2 Trimer .....	32
3.3.1.3 Tetramer .....	38
3.3.2 Thymine clusters .....	41
3.3.2.1 Dimer .....	41
3.3.2.2 Trimer .....	44
3.3.2.3 Tetramer .....	47
3.3.3 Adenine-Thymine clusters.....	51
3.3.3.1 Dimer .....	51

3.3.3.2 Trimer .....	54
3.3.3.3 Tetramer .....	57
<b>4. CONCLUSION .....</b>	<b>61</b>
<b>REFERENCES.....</b>	<b>63</b>
<b>CURRICULUM VITAE.....</b>	<b>71</b>





## ABBREVIATIONS

<b>DNA</b>	: Deoxyribonucleic acid
<b>DFT</b>	: Density Functional Theory
<b>MD</b>	: Molecular Dynamics
<b>PES</b>	: Potential Energy Surface
<b>STM</b>	: Scanning Tunnelling Microscopy
<b>HF</b>	: Hatree-Fock
<b>CC</b>	: Coupled Cluster
<b>CCSD(T)</b>	: Single, Double and Perturbative Triple Couple Cluster
<b>MPPT</b>	: Møller-Plesset Perturbation Theory
<b>SCS-MP2</b>	: Spin Component Scaled MP2
<b>CP</b>	: Counterpoise
<b>BSSE</b>	: Basis Set Superposition Error
<b>SA</b>	: Simulated Annealing
<b>CMS</b>	: Center of Mass
<b>CBS</b>	: Complete Basis Set
<b>PEC</b>	: Potential Energy Curve
<b>PES</b>	: Potential Energy Surface



## LIST OF TABLES

	<u>Page</u>
<b>Table 3.1</b> : Interaction Energies of Chosen Adenine Dimers. H-bonded Dimer A (C.m.s. 6.13 Å ) and Stacked B (C.m.s. 3.25 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 .....	16
<b>Table 3.2</b> : Adenine A and B isomers for aug-cc-pvdz basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation. ....	17
<b>Table 3.3</b> : Thymine A and B isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation. ....	20
<b>Table 3.4</b> : Interaction Energies of Chosen H-bonded Tymine A (C.m.s. 6.59 Å ) and Stacked B (C.m.s. 3.52 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 .....	21
<b>Table 3.5</b> : Adenine-Thymine A and B isomers for aug-cc-pvdz basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation.....	23
<b>Table 3.6</b> : Interaction Energies of Chosen H-bonded Tymine A (C.m.s. 6.59 Å ) and Stacked B (C.m.s. 3.52 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 3.7</b> : Fitted parameters for Adenine 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 adenine monomer, respectively. ....	25
<b>Table 3.8</b> : Fitted parameters for Thymine 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 adenine monomer, respectively. ....	27
<b>Table 3.9</b> : Fitted parameters for Adenine-Thymine 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 adenine monomer, respectively. ....	27

<b>Table 3.10</b> : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine dimers shown in Figure 3.13.....	<b>31</b>
<b>Table 3.11</b> : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine trimers shown in Figure 3.14.....	<b>35</b>
<b>Table 3.12</b> : Total ( $\Delta E_{int}^{CP}$ ), two-body ( $\Delta^2 E_{int,AB}^{CP}$ , $\Delta^2 E_{int,AC}^{CP}$ and $\Delta^2 E_{int,BC}^{CP}$ ) and three-body ( $\Delta^3 E_{int,ABC}^{CP}$ ) interaction energies (calculated at B3LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 using aVDZ) of adenine trimers shown in Figure 3.14.....	<b>36</b>
<b>Table 3.13</b> : Total ( $\Delta E_{int}^{CP}$ ), two-body ( $\Delta^2 E_{int,AB}^{CP}$ , $\Delta^2 E_{int,AC}^{CP}$ and $\Delta^2 E_{int,BC}^{CP}$ ) and three-body ( $\Delta^3 E_{int,ABC}^{CP}$ ) interaction energies (calculated at B3LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 using aVDZ) of adenine trimers shown in Figure 3.14.....	<b>37</b>
<b>Table 3.14</b> : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine tetramers shown in Figure 3.15.....	<b>40</b>
<b>Table 3.15</b> : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of thymine dimers shown in Figure 3.16.....	<b>43</b>
<b>Table 3.16</b> : Interaction energy calculations (at AMBER) using aVDZ) of thymine trimers shown in Figure 3.17. ....	<b>46</b>
<b>Table 3.17</b> : Interaction energy calculations (at AMBER) using aVDZ) of thymine tetramers shown in Figure 3.18. and 3.19.....	<b>50</b>
<b>Table 3.18</b> : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine-thymine dimers shown in Figure 3.20.....	<b>53</b>
<b>Table 3.19</b> : Interaction energy calculations (at AMBER) using aVDZ) of adenine-thymine trimers shown in Figure 3.21. ....	<b>56</b>
<b>Table 3.20</b> : Interaction energy calculations (at AMBER) using aVDZ) of adenine-thymine tetramers shown in Figure 3.22.....	<b>59</b>

## LIST OF FIGURES

	<u>Page</u>
<b>Figure 1.1</b> : DNA Bases a)Adenine, b)Thymine.....	2
<b>Figure 2.1</b> : Potential energy curve of H-bonded adenine dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black.....	8
<b>Figure 2.2</b> : A six-dimensional model. R represents the distances between CMS, $\Theta$ and $\Phi$ are polar angels and $\theta$ , $\phi$ and $\psi$ are euler rotation angels. ....	9
<b>Figure 2.3</b> : Global and Local Minimum Points. ....	11
<b>Figure 3.1</b> : Adenine dimers: a) H-bonded A, b) Stacked B. ....	15
<b>Figure 3.2</b> : Potential energy curves of Adenine Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black.....	15
<b>Figure 3.3</b> : DFT-SAPT(LPBE0AC) energy contributions for the H-bonded and Stacked Adenine dimers. Long-dashed lines represent $E_{el}^{(1)}$ and $E_{exch}^{(1)}$ , short-dashed $E_{ind}^{(2)}$ and $E_{exch-ind}^{(2)}$ , dotted $E_{disp}^{(2)}$ and $E_{exch-disp}^{(2)}$ , dotted long-dashed $\delta(HF)$ , and the solid line $E_{int}$ . ....	16
<b>Figure 3.4</b> : Thymine dimers: a) H-bonded A, b) Stacked .....	18
<b>Figure 3.5</b> : Basis set dependence of DFT-SAPT(LPBE0AC) energy components for Adenine and Thymine dimers.....	19
<b>Figure 3.6</b> : Potential energy curves of Thyime Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black.....	19
<b>Figure 3.7</b> : Adenine-Thymine dimers: a) H-bonded A, b) Stacked .....	21
<b>Figure 3.8</b> : Potential energy curves of Adenine-Thyime Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black.....	22
<b>Figure 3.9</b> : Basis set dependence of DFT-SAPT(LPBE0AC) energy components for Adenine-Thymine Dimers. ....	22
<b>Figure 3.10</b> Adenine Monomer.....	25
<b>Figure 3.11</b> Comparison and deviations of the Adenine model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies.....	26
<b>Figure 3.12</b> Comparison and deviations of the Thymine model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies.....	26
<b>Figure 3.13</b> Adenine dimer structures found by the SA approach. ....	30

<b>Figure 3.14</b>	Adenine trimer structures found by the SA approach. ....	<b>34</b>
<b>Figure 3.15</b>	Adenine tetramer structures found by the SA approach. ....	<b>39</b>
<b>Figure 3.16</b>	Thymine dimer structures found by the SA approach. ....	<b>42</b>
<b>Figure 3.17</b>	Thymine trimer structures found by the SA approach. ....	<b>45</b>
<b>Figure 3.18</b>	Thymine tetramer structures found by the SA approach. ....	<b>48</b>
<b>Figure 3.19</b>	Thymine tetramer structures found by the SA approach. ....	<b>49</b>
<b>Figure 3.20</b>	Adenine-Thymine dimer structures found by the SA approach. ....	<b>52</b>
<b>Figure 3.21</b>	Adenine-Thymine trimers found by the SA approach. ....	<b>55</b>
<b>Figure 3.22</b>	Adenine-Thymine tetramer structures found by the SA approach. ....	<b>58</b>



# SYMMETRY-ADAPTED PERTURBATION THEORY POTENTIALS FOR DNA BASES

## SUMMARY

Deoxyribonucleic acid (DNA) is considered as the fundamental structure of life and it contains the all genetic information of known living beings and some viruses. Therefore, DNA is the one of the most important and heavily inspected biological molecule. The diversity it can produce, however, rests upon the combination of the four different molecules; adenine, cytosine, thymine, and guanine. These DNA bases are stabilized by non-bonded interactions, especially with hydrogen bonding and stacking. This unique structure gives the DNA the ability to form triplexes, quadruplexes and many other complex structures. These DNA bases can also be embedded on metal surfaces and the resulting metal-DNA complexes can be used for many special purposes such as biochip sensors, organic semiconductors, and organic photovoltaic tools. In this study, we developed a new force field specialized for adenine, thymine and adenine-thymine oligomers. For this purpose, the first step of this study will be to compare the performances of various ab initio methods for calculating the interaction energies. MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D and DFT-SAPT (PBE0AC and LPBE0AC) was used to calculate Potential Energy Curves (PEC) with Dunning's aug-cc-pVXZ (aVXZ) base sets ( $X = D, T$  and in some cases  $Q$ ) to find the most accurate and computationally cheap method. After this step, acquired results compared with the CCSD(T), which is considered as the most accurate but computationally expensive method. The best resultant method will be used to calculate the potential energy surfaces (PES) and used as an alternative to CCSD(T). As a next step, the resulting interaction forces fitted to an analytical function to develop force fields. Then, the model function was used to perform a global search to find the local and global minima of adenine, thymine and adenine-thymine dimers, trimers, and tetramers by using the one of the most successful and robust global optimization algorithm, Simulated Annealing. Our structure optimizations showed us that the new potential is able to reproduce the 2-D filament structures and some new non-planar adenine, thymine and adenine-thymine clusters. For the final step, we compared the well-known AMBER with our generated force field, for all considered systems. Although AMBER force field generally agrees with ab initio methods and proposed model, it can fail to produce the correct interaction energies for certain cases while proposed model potential performs better. Therefore, we suggest our proposed force field is suitable to perform molecular dynamics simulations of much larger DNA systems.





## DNA BAZLARI İÇİN SİMETRİ UYUMLU PERTURBASYON KURAMI POTANSİYELLERİ

### ÖZET

DNA bir nükleik asit olup, yaşayan tüm canlıların ve bazı virüslerin temel yapı taşı olması sebebiyle bilimsel literatürde en çok incelenen yapılardan biridir. Oluşturabildiği çeşitliliği, 4 temel bileşeni olan adenin, timin, sitozin ve guanin'in farklı kombinasyonlar halinde bir araya gelmesiyle oluşturan DNA, kovalent olmayan etkileşimler sayesinde yapılarını kararlı hale getirmektedir. Özellikle elektrostatik (O-H ve N-H arasında oluşan hidrojen bağları) ve  $\pi$ -yığılma ( $\pi$ -stacking) etkileşimleri, bu sistemlerin kararlılığını etkileyen önemli faktörlerden biridir. Bir pürin (A-T) bazı ve bir pirimidin (S-G) bazının eşleşmesiyle oluşan DNA'nın çift sarmallı yapısı, bu etkileşimler sayesinde kararlı hale gelir. Bu eşsiz yapı, DNA'ya tripleksler gibi kompleks yapıları oluşturma yeteneği kazandırır. Ayrıca, DNA'yı oluşturan bazlar, çeşitli metallerin üzerine yerleştirilerek kendilerine organik yarıiletkenler, bioçip sensörleri ve organik fotovoltajik araçlar gibi uygulamalar da bulabilir. Bu uygulama alanlarının yanı sıra, altın gibi metal yüzeylere yerleşen DNA bazları da literatürde oldukça irdelenmiş ve DNA bazlarının metal yüzeylerdeki konumları, hem taramalı tünelleme mikroskopları (STM) hem de moleküler dinamik (MD) simülasyonları kullanılarak incelenmiştir. STM ile küçük organik moleküllerin tespiti zor olduğundan, yoğunluk fonksiyon teorisi (DFT) gibi teorik kuantum kimya hesapları bu tarz yapıların aydınlatılmasında oldukça yararlı olmuştur. Özellikle tek ve çift eksitasyonlu perturbatif üçlü eksitasyon düzeltmelerini içeren coupled cluster (CCSD(T)) metodu, neredeyse tüm sistemler için yüksek doğrulukta sonuçlar verebilmektedir. Fakat CCSD(T)'nin, hesaplama gücü açısından çok pahalı olması, onu, bu çalışmanın son bölümlerinde de aydınlatmak isteyeceğimiz küme yapıları (cluster) gibi kompleks sistemler için uygulanamaz kılmaktadır. Bu nedenden ötürü CCSD(T), DNA küme yapıları için maalesef kullanılamamaktadır. Bu sorunu aşmak için analitik ifadeleri kullanarak bir potansiyel enerji yüzeyi (PEY) oluşturulabilir ve elde edilecek enerji değerleri, bir analitik fonksiyona fitlenebilir. Fakat bu durumda da, elde edeceğimiz analitik fonksiyonun doğruluğu, doğrudan potansiyel enerji yüzeyini hesapladığımız teorik hesap seviyesine bağlı olacaktır. Bu sebepten ötürü, potansiyel enerji yüzeyini oluştururken kullandığımız hesaplama seviyesinin seçimi, modelimizin en uygun (doğruya en yakın ve en hızlı) şekilde çalışması için büyük önem taşımaktadır. Bu amaçla çalışmamızın ilk adımı, PEY'leri oluşturmak için çeşitli ab-initio yöntemlerin, ilgili dimerlerdeki etkileşim enerjilerini hesaplayabilme performanslarını karşılaştırmak olacaktır. En yüksek doğrulukta ve en düşük hesaplama maliyeti ile etkileşim enerjilerini hesaplayabilen yöntemi bulabilmek için Potansiyel Enerji Eğrileri (PEE) oluşturulacak ve bu eğrilerin sonuçlarına göre PEY'lerin hesaplanmasında kullanılacak olan yöntem seçilecektir. Etkileşim enerjilerinin hesaplanmasında supermoleküler hesaplama ya da perturbasyon teorisi kullanılabilir. Supermoleküler hesaplama, herhangi bir teorik seviye kullanılarak yapılabilir ve dimer enerjisinden, monomerlerin enerjilerinin

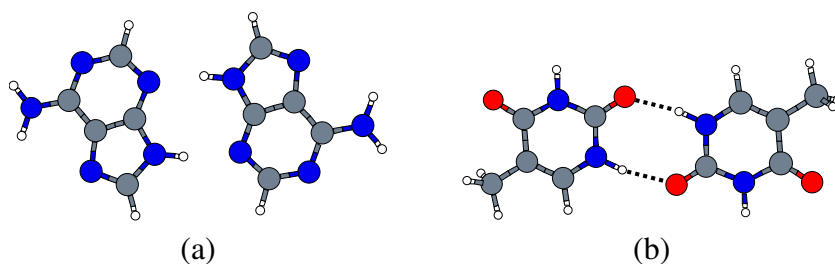
analitik olarak çıkarılmasıyla elde edilir. Bu hesaplamalarda karşılaşılabilecek bir sorun baz seti superpozisyon hatasıdır (basis set superposition error) (BSSE). Kuantum kimya hesaplamaları, baz seti süperpozisyon hatasına karşı hassastırlar. Etkileşen moleküllerin (veya aynı molekülün farklı bölümlerinin) atomları birbirine yaklaştıkça, temel işlevleri üst üste gelir. Her monomer, diğer yakın bileşenlerden baz setlerini "ödünç" alır ve baz setini etkin bir şekilde artırır. Bunun sonucunda, ilgili enerjiler yanlış hesaplanacak ve bu uyumsuzluk bir hata getirecektir. BSSE, karşılıklı hesaplama (counterpoise (CP)) yaklaşımı ile giderilebilir. CP yaklaşımında monomerlerin toplam enerjisi, toplam dimer baz seti kullanılarak hesaplanmaktadır. Perturbasyon teorisinde ise, etkileşim enerjileri fiziksel kökenleri farklı olan birden çok terimin toplamı şeklinde ifade edilir ve BSSE hatası da içermemektedir. Bahsedilen yöntemlere ek olarak, hidrojen bağlı yapılarda iyi çalıştığı bilinen MP2 ve MP2'nun iki farklı varyantı olan SCS-MP2 ve SCS-MI-MP2 da PEY'in hesaplanması için uygun yöntemler olabilir. Özellikle düşük hesaplama maliyeti ile gelen SCS-MI-MP2, PEY'lerin hesaplanması için harcanan süreyi bir hayli kısaltacak olması açısından oldukça avantajlıdır. Bu çalışmada, adenin, timin ve adenin-timin bazları ele alınacak ve ilgili bazların dimer potansiyel enerji yüzeyleri (PEY), CCSD(T) ile en uyumlu teorik metot ile hesaplanıp bir analitik fonksiyona fitlenecektir. PEY'in hesaplanmasından önce MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D ve DFT-SAPT (PBE0AC ve LPBE0AC) metotları kullanılarak potansiyel enerji eğrileri (PEE) oluşturulacak ve bu metotlar, hakem metot olan CCSD(T) ile karşılaştırılarak potansiyel enerji yüzeylerini (PEY) hesaplamak için en uygun ve en hızlı metot seçilecektir. Dimerler için elde edilmiş bu kuvvet alanlarının oluşturulmasından sonra, adenin, timin ve adenin-timin küme yapıları aydınlatılmaya çalışılacaktır. Grubumuz tarafından, guanin, stozin ve stozin-guanin dimerleri için yapılmış önceki çalışmalar, PEY'lerin hesaplanması için bahsi geçen yöntemleri karşılaştırmış ve DFT-SAPT'ın, CCSD(T)'ye en yakın sonuçları verdiğini tespit etmiştir. Ayrıca DFT-SAPT'ın, etkileşim enerjilerini elektrostatik, dispersiyon, indüksiyon ve bunlara ait itici terimlerin bir toplamı şeklinde vermesi, sistemi stabilize eden faktörlerin belirlenmesi açısından fayda sağlamaktadır. DFT-SAPT'ın bu performansı, adenin, timin ve adenin-timin dimerleri için de beklenmektedir. Dimerler için tasarlanmış bir kuvvet alanının, örneğin bir trimerin dispersiyon etkileşimlerini hesaplariken, etkileşimleri monomerlerin toplamları (AB+AC+BC) şeklinde ifade edilmesinden ötürü daha büyük sistemlerde beklenildiği kadar doğru çalışmayabilir. Fakat yinede hesaplanan kuvvet alanının performansının, literatürde sıklıkla kullanılan AMBER kuvvet alanına kıyasla daha doğru çalışması beklenmektedir. Adenin, timin ve adenin-timin dimerleri için PEY'lerin geliştirilmesinden sonraki adım, benzetilmiş tavlama (simulated annealing) yöntemi ile PEY'in küresel minimumlarını tahmin etmek olacaktır. Benzetilmiş tavlama, sezgisel bir küresel eniyileme (global optimization) yöntemi olup, grubumuz tarafından yapılan daha önceki çalışmalarda da küresel minimumların tahmin edilmesinde başarıyla kullanılmıştır. Burada karşılaşılabilecek bir sorun, tüm küresel eniyileme yöntemlerinde de bulunan küresel minimuma ulaşma sorunudur. Bu sorun, küresel eniyileme algoritmalarının, küresel minimuma ulaşmadan önce, yerel bir minimum noktasına takılarak, küresel minimum noktasını ıskalaması sonucunda meydana gelir. Benzetilmiş tavlama yöntemini, daha öncesinden farkında olduğumuz bu problemten kurtarmak için birden fazla kez çalıştırarak, algoritmanın küresel minimum noktasına gitmesini sağlayabiliriz. Hesapsal maliyetinin de görece düşük olmasından dolayı benzetilmiş tavlama yöntemi bu şekilde uygulanarak adenin, timin ve adenin-timin küme yapılarının aydınlatılmasında kullanılabilir. Çalışmamızın son

aşamasında adenin, timin ve adenin-timin dimerleri için geliştirilen kuvvet alanları kullanılarak, benzetilmiş tavlama yöntemi ile küresel minimum noktalarına ulaşmaya çalışılacak, bu kuvvet-alanlarının, dimer ve diğer düzlemsel yada düzlemsel olmayan oligomerlerinin (en fazla 4 tane nükleik asit bazı içeren) yapıları tahmin etme gücü belirlenecek ve benzetilmiş tavlama yöntemi ile bulunan yapılar, literatürde bulunan diğer sonuçlarla karşılaştırılacaktır. Buradaki ana hedef, hem literatürde bulunan yapılara ulaşmak, hem de (eğer varsa) bu yapılardan daha düşük enerjili yapıları tespit etmektir. Ayrıca litetartürde irdelenen ve taramalı tünelleme mikroskobu (Scanning tunneling microscope (STM)) görüntüleri ile tespit edilen metal yüzeylerle yerleşik DNA bazlarının, benzetilmiş tavlama yöntemi ve geliştirilen kuvvet alanı kullanılarak da tespiti, diğer bir odak noktamızdır. Sitozin, guanin ve sitozin-guanin için yapılmış olan önceki çalışmalar, bu tarz yapıların aydınlatılmasında izlediğimiz yolun başarılı olduğunu ortaya koymuştur. Burada, ilgili sistemlerin toplam enerjisi, sistemdeki tüm ikili etkileşimlerin toplamı şeklinde ifade edilecektir. Yinede hesaplanan kuvvet alanı performansları, literatürde sıklıkla kullanılan AMBER kuvvet alanı ile kıyaslanacak ve bu kuvvet alanına kıyasla daha doğru çalışması hedeflenecektir. Çalışmalarımızın sonucunda, adenin, timin ve adenin timin bazları için PEE oluşturulmuş ve bu eğrilerin yardımı ile PEY yüzeylerini oluşturmak için kullanılacak olan teorik metot DFT-SAPT olarak belirlenmiştir. Fakat, DFT-SAPT'ın performansına oldukça yakın bir performans gösteren SCS-MI-MP2 yöntemi de gelecek çalışmalar için kullanılması muhtemel bir yöntem olarak göze çarpmıştır. PEY'lerin, oluşturulmasında adenin, timin ve adenin-timin bazları için sırasıyla 7286, 4412 ve 6390 tane nokta kullanılmıştır. İlgili PEY'lerin oluşturulmasından sonra sıra benzetilmiş tavlama yöntemini, her bir baz için çalıştırmak olmuştur. Bunun için, PEY'lerin sonuçları ele alınarak benzetilmiş tavlama yönteminde kullanılacak olan parametreler belirlenmiş ve ilgili DNA bazlarının küme yapıları, benzetilmiş tavlama yöntemi ile aydınlatılmıştır. Benzetilmiş tavlama metodu, küresel minimum noktasına ulaşmasını sağlamak adına ardışık olarak çalıştırılmış ve hem literatürde hali hazırda rapor edilmiş adenin, timin ve adenin-timin bazlarını ve kümelerini yakalamayı başarmış hem de dimer, trimer ve tetramerler için yeni yapılar bulmayı başarmıştır. Bu sonuçların ışığında ve çalışmalarımızın sonucunda geliştirilmiş olan kuvvet alanının, hali hazırda kullanılan AMBER kuvvet alanına göre daha iyi sonuçlar verebildiği saptanmış ve DNA bazlarını içeren yapıların aydınlatılmasında kullanılacak yeni bir kuvvet alanı olarak sunulmuştur. Ayrıca ulaştığımız bu sonuçlar, ilgili bazların daha büyük küme yapılarının da incelenmesi açısından da umut vaatmektedir.



## 1. INTRODUCTION

Deoxyribonucleic acid (DNA) is considered as the fundamental structure of life and it contains the all genetic information of known living beings and some viruses. DNA is used for passing the information to the offsprings and DNA degeneration can cause serious harms, such as cancer, cardiovascular diseases, and diabetes. Therefore, DNA has become the one of the most important and heavily inspected biological molecule in the scientific community [1–10]. The diversity it can produce however, rests upon the combination of the four fundamental molecules; adenine, cytosine, thymine, and guanine. These DNA bases are stabilized by non-bonded interactions, especially via hydrogen bonding and stacking. This unique structure gives DNA the ability to form triplexes, quadruplexes and many other complex structures [1, 2, 11–19]. DNA bases can also be embedded on metal surfaces and the resulting metal-DNA complexes can be used in various applications such as biochip sensors, organic semiconductors and organic photovoltaic tools [20–22]. For example, adenine supermolecular structures was observed on Cu(111) [8–10, 23–25], Cu(110) [26, 27], Au(111) [6, 28, 29], graphite [30–34], molybdenite [31] and Ag terminated Si(111) [35] surfaces for their role on self-assembled structures. Cytosine [36–39], guanine [37, 40] and thymine [37, 41, 42] were also deposited on various metal surfaces. Previous STM images showed that all DNA bases can form one or two dimensional supermolecular planar networks on metal surfaces and DNA metal surfaces can act as a platform to sustain the DNA bases [36]. Computational methods can be implemented to widen our knowledge on this subject and have been used previously to identify the existing structural motifs in STM images. [43]. In such studies, one must compute the base-base and base-surface interactions via force field which are not specialized for DNA bases and this can be misleading. Maleki et al. [44] used a combination of well known Lennard-Jones potential with AMBER parameters and Coulomb term to overcome these complications. In the current study, we used our posteriori knowledges coming from guanine [45] and cytosine [46] oligomers and implemented a new force field specialized for adenine, thymine, and adenine-thymine oligomers with symmetry-adapted perturbation theory



**Figure 1.1** : DNA Bases a)Adenine, b)Thymine.

combined with density functional theory (DFT-SAPT) developed by Hasselmann and Jansen [47–50]. Then, the model function was used to search the structures of adenine, thymine and adenine-thymine dimers, trimers, and tetramers. Similar to the previous works in cytosine [46] and guanine [45], our current results are quite promising and it shows that our model potential might be an asset for further MD simulations. The outline of this thesis is as follows; Section II presents the details of our computational methods and quantum chemical calculations performed by DFT-SAPT theory and DFT. We also compared our selected computational methods with other high accurate quantum chemical methods and also AMBER force field. This section also contains pieces of information about our fitting strategy. In Section III, we compare the efficiencies of all methods and we used our force field to generate adenine, thymine and adenine-thymine clusters. Finally, Section IV summarises the study based on our findings.

## 2. METHOD

### 2.1 Ab-initio Methods

As a term, ab-initio means "from the first principles" or "from the beginning" and first used by Robert Parr and his coworkers in a study on the excited states of benzene [51]. Ab-initio calculations are based on quantum mechanical laws, thermodynamics and some physical constants such as speed of light. Ab-initio methods produce correct theoretical results but as a nature of this process, they require intense computing resources. Because of this demand, ab-initio methods are only applicable for smaller or slightly bigger systems. Hartree-Fock (HF) theory is the most basic ab-initio method available, however in HF, electron correlation effects are omitted. These effects are crucial for calculating the dispersion forces on intermolecular interactions and for this purpose, post-HF methods such as Møller-Plesset (MP) methods has been developed. According to its contained number of terms, MP can be classified from second to fourth order or more orders. Also, there are other methods available such as coupled cluster (CC) and configuration interaction (CI) methods which can produce more accurate results than Møller-Plesset. Single and double excitation coupled cluster theory including perturbative triple excitations (CCSD(T)) is known as one of the most reliable and accurate methods and is considered as a reference calculation level. But calculating the intermolecular energies with CCSD(T) is not feasible because of its high computational demands. In addition, CI methods can also be considered as very accurate and if the wave function can be determined as the sum of all possible stimulations, then it becomes a Full-CI, the most accurate method. As a result, we can order the accuracy level of the ab-initio methods as follows; HF < MP < CCSD(T) < Full-CI.

#### 2.1.0.1 Moeller-Plesset perturbrational theory

As a Post-HF method, Møller-Plesset Perturbational Theory improves the HF method by adding the electron correlations effects. According to its contained number of terms,

MP can be classified as MP2, MP3, MP4 and so on. MP2 can be considered as the most popular one among the other MP methods because of its computationally feasible nature. MP3 and MP4 are also used quite widely and accuracy of the MP4 is very close to CISD. Usage of the MP5 is very limited, mostly because of its high computational demand. MP2 energies are calculated by the sum of the anti-parallel and parallel spin correlation energies and can be seen in the following equation 2.1.

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

Since anti-parallel and parallel spin correlation energies do not evenly contribute the total energy, equation 2.1 can be scaled with two scaling parameters,  $c_1$  and  $c_2$  respectively. This scaled version of equation 2.1 can be seen in equation 2.2.

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

This scaled MP2 method is called as Spin Component Scaled MP2 (SCS-MP2) and generally gives the better and more accurate results than MP2 energies [52]. The optimized scaling parameters are 1.2 for antiparallel spin and 1/3 for parallel spin. Distasio et al. [53] further optimized the spin-component scaling parameters shown in equation 2.2 and the resultant model, SCS-MI-MP2, emerged as a new and more accurate method that corrects the MP2 errors which mostly seen in hydrogen-bonded structures. The resulting SCS-MI-MP2 parameters with respect to their original values are 1.29 for antiparalel and 0.40 for the parallel spin.

### 2.1.0.2 Coupled cluster

Coupled Cluster (CC) method is a numerical technique to describe many-body systems and it takes the HF molecular orbital method and uses the exponential cluster operator to construct multi-electron wavefunctions. CC method is often used for reference calculation method because of its high accuracy and it is the perturbative variant of the Many-Electron Theory (MET). CC methods are classified by their highest number of excitations allowed in the definition of cluster operator,  $T$ .

$$T = T_1 + T_2 + T_3 + \dots, \quad (2.3)$$



where  $T_1$ ,  $T_2$  and  $T_3$  are the operators for single, double and triple excitations respectively. CC methods begin with the abbreviation "CC" followed by "S", "D", "T" or "Q" for single, double, triple and quadruple excitations, respectively.

### 2.1.1 Density functional theory

In Density Functional Theory (DFT), the interaction energies of many-electron systems can be determined by using functionals which can be defined as "function of another function". The name DFT comes from the electron densities, which used to calculate the interaction energies via density functionals, instead of wave functions [54]. DFT requires an initial functional to proceed and this makes the process of determining an appropriate functional as a crucial step. DFT method is considered as a faster and computationally feasible method especially when compared to the ab-initio methods such as HF theory and its descendants that include electron correlations. The main setback of the DFT is that its inability to calculate the interaction energies such as dispersion term accurately enough. This setback has been overcome by the dispersion correction term,  $E_{disp}$  on DFT-D.

$$E_{DFT-D} = E_{KS-DFT} + E_{disp}. \quad (2.4)$$

where,  $E_{KS-DFT}$  is the Kohn-Sham DFT functional. Dispersion correction term,  $E_{disp}$  can be expanded as a sum of dispersion parameters for the  $ij$  pairs and can be seen in equation 2.5,

$$E_{disp.} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} f_{dmp}(R_{ij}) \quad (2.5)$$

where  $N_{at}$  is the number of the atoms,  $s_6$  is the scaling factor for specific density functional,  $R_{ij}$  is the distance between the atoms and  $f_{dmp}$  is the damping function to prevent singularities in the short distances.

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{\frac{R_{ij}}{R_r - 1}}} \quad (2.6)$$

Here  $R_r$  is the sum of the Van der Waals radius.

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

Symmetry Adapted Perturbation Theory (SAPT) was developed to describe non-covalent interactions and it provides a perturbative expression for interaction energies. SAPT decomposes the interaction energies into first-order electrostatic  $E_{el}^{(1)}$ , second-order induction  $E_{ind}^{(2)}$ , dispersion  $E_{disp}^{(2)}$  and repulsive terms resulting from electron exchange among monomers  $E_{exch}^{(1)}$ ,  $E_{exch-ind}^{(2)}$  and  $E_{exch-disp}^{(2)}$  respectively.  $\delta(HF)$  represents the effect of second-order terms and it is the difference between the supermolecular Hartree-Fock (HF) interaction energy and the electrostatic, induction, and their exchange counterpart energies obtained from HF monomer density matrices and HF response functions. SAPT generally was considered as a computationally expensive [55] method and to overcome this bottleneck, monomer properties can be calculated at the DFT level resulting to DFT-SAPT [47–50, 56] and SAPT(DFT) [57] methods. Interaction energies in DFT-SAPT calculated with the sum of physical terms;

$$E_{int} = E_{el}^{(1)} + E_{exch}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} + \delta(HF) \quad (2.7)$$

Third and higher order contributions are defined in  $\delta(HF)$  and above equation 2.7 can be reformulated as:

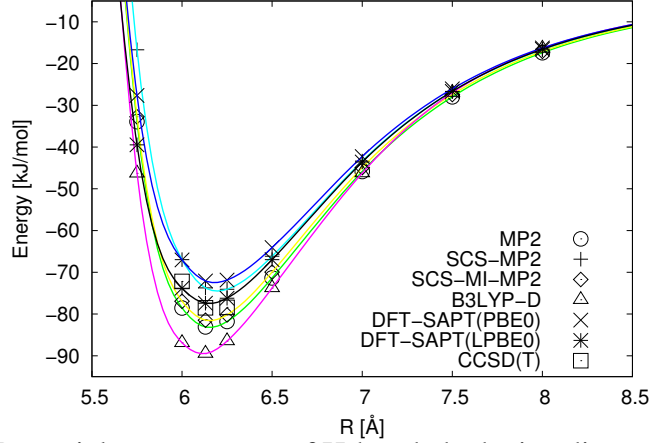
$$\delta(HF) = E_{int}(HF) + E_{el}^{(1)}(HF) + E_{exch}^{(1)}(HF) + E_{ind}^{(2)}(HF) + E_{exch-ind}^{(2)}(HF) \quad (2.8)$$

Here  $E_{int}(HF)$  is the Counterpoise-corrected supermolecular HF interaction energy. DF-DFT-SAPT is a computationally faster variant of DFT-SAPT and currently implemented on MOLPRO [58] program package. In DF-DFT-SAPT, PBE0AC [56] and LPBE0AC [59] density functionals was used to obtain the monomer attributes and the efficiency of DFT is increased with this approach when it compared to CCSD(T).

### 2.1.3 Investigation of DNA bases via ab-initio methods

Selection of the theoretical method to compute the Potential Energy Surfaces (PES) of adenine, thymine and adenine-thymine dimers has a critical importance to improve the predictive power of the generated force fields and our posterior knowledge coming from guanine [45] and cytosine [46] oligomers had also proved this fact. CCSD(T)

is known as the most reliable and accurate, but also computationally intense method. Calculating the intermolecular energies with CCSD(T) is not feasible, therefore, we must find the most accurate and computationally feasible method which can be an alternative to CCSD(T). The second order Møller-Plesset perturbation theory (MP2), spin-component scaled MP2 (SCS-MP2) [52], SCS-MI-MP2 [60] and DFT-D were frequently used for supermolecular approaches instead of computationally intense CCSD(T). Here, B3LYP-D [61, 62] was chosen as a DFT-D approach and it has been observed that its performance depends on the considered system similar to PBE-D and meta-generalized gradient approximation functionals M05-2X and M06-2X, e.g., even it produces interaction energies which are similar to CCSD(T) for pyrazin [63–65] and parallel-displaced benzene [66], it is shown that this approach fails for stacked benzene and H<sub>2</sub>S-benzene dimer interactions [66]. As an alternative to supermolecular approach, one can also use DFT-SAPT to obtain the interaction energies as a sum of physically distinct components such as electrostatics, dispersion, induction and their exchange counterparts. It has been shown that DFT-SAPT is in quite good agreement with CCSD(T) [45, 46, 65, 67, 68]. In this study, we compared the performances of DFT-SAPT and counterpoise (CP)-corrected supermolecular MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D and CCSD(T) methods for the stacked and hydrogen-bonded adenine, thymine and adenine-thymine dimers to find the best and most suitable alternative for CCSD(T). Adenine, thymine and adenine-thymine dimers which were used to calculate Potential Energy Curves (PEC) taken from well known studies [6, 7, 69, 70]. Interaction energy computations were performed using MOLPRO [58] and density fitting implementation of DFT-SAPT (DF-DFT-SAPT) [71] has been used. We showed that DFT-SAPT (LPBE0) is in quite good agreement with CCSD(T). In all these computations, Dunning's aug-cc-pVXZ (aVXZ) basis sets with X = D, T and Q were employed. The developed force field for adenine, thymine and adenine-thymine dimers were used for the prediction of the corresponding oligomers. Some of the resulting structures were further relaxed at PBE, SCS-MP2 and CP-SCSMP2 levels with aug-cc-pVDZ basis set. For the larger cluster sizes, geometry optimizations were carried out only at SCS-MP2/aug-cc-pVDZ level. TURBOMOLE V6.1 program package [72] was used for all these computations. In addition to the comparisons performed using ab-initio calculations, we also utilize the well known AMBER non-bonded empirical potential for the calculation of the interaction energies of the



**Figure 2.1** : Potential energy curve of H-bonded adenine dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black

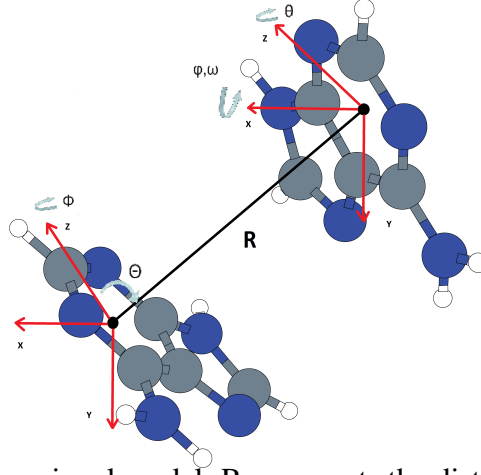
adenine, thymine and adenine-thymine clusters. Only Van der Waals and Coulombic terms were used in AMBER force field [73] as described in Morgado et al. [74].

## 2.2 Fitting Details

The second step of this study is to fit a potential model to interaction energies which acquired by the calculation of PES with the chosen theoretical method. Recently, our work group successfully developed a first principles potential for the acetylene [75], cytosine [46] and guanine [45] using DFT-SAPT(PBE0AC) and resulting interaction energies were fitted to a site-site model potential. We used the same approach for the adenine, thymine and adenine-thymine bases. First, we create a PES by using a six-dimensional grid shown in Figure 2.2 for each DNA bases and we eliminate the symmetry-redundant and close-contact orientations. Resulting grid was calculated with DFT-SAPT (LPBE0)/aVDZ and Levenberg-Marquardt nonlinear weighted least squares method were used to fit the model. The functional form which has been used previously on cytosine [46] and guanine [45] dimers was also used for the fitting of the adenine, thymine and adenine-thymine dimers and can be seen bellow;

$$V = \sum_{i \in A} \sum_{j \in B}^{sites} \left\{ \alpha_{ij} \exp(-\beta_{ij} r_{ij}) + \frac{C_{ij}}{(r_{ij}^6 + c_{ij}^6)} + f_0(\delta_0^{ij}, r_{ij}) \frac{q_i q_j}{r_{ij}} \right\}, \quad (2.9)$$

$r_{ij}$  indicates the distances between  $i$  and  $j$ , which corresponds the monomers A and B, respectively.  $\alpha$ ,  $\beta$ ,  $C$  are defined as the fitting parameters. The each element in adenine dimers for example, was assumed to be a different site and this gives a total of 6 pair



**Figure 2.2** : A six-dimensional model.  $R$  represents the distances between CMS,  $\Theta$  and  $\Phi$  are polar angles and  $\theta$ ,  $\phi$  and  $\psi$  are euler rotation angles.

interactions (C-C, C-H, C-N, H-H, H-N and N-N). Each of these interactions can be defined with 3 parameters as shown in the above equation and this gives a total of 18 fit parameters. In equation 2.9,  $q_i$  and  $q_j$  are the ESP fitted partial charges and  $f_0$  is the Tang-Toennies damping function [76] with  $\delta = 1$  and  $n = 0$ ,

$$f_n(\delta, r) = 1 - \exp(-\delta r) \sum_{m=0}^n \frac{(\delta r)^m}{m!}, \quad (2.10)$$

which eliminates the divergence of the Coulomb interaction term for  $r_{ij} \rightarrow 0$ . Manually adjusted  $c_{ij}$  parameters were also used for damping the dispersion interaction term, to avoid unphysically large values. Finally, Levenberg-Marquardt non-linear weighted least squares method was used to determine the fitting parameters by minimizing the equation 2.11,

$$\chi^2 = \sum_{i=1}^N \sigma_i (y_0(x_i) - y(x_i; \alpha_{ij}, \beta_{ij}, C_{ij}))^2, \quad (2.11)$$

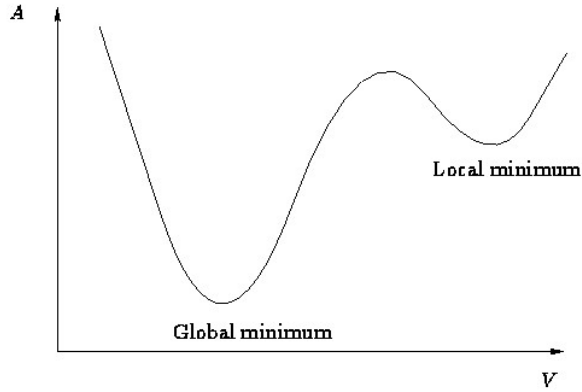
where the ab-initio and model energy defined as  $y_0$  and  $y$  at dimer geometry  $x_i$ . Individual weight term,  $\sigma_i$ , was also determined to each dimer geometries based on its interaction energy, ( $E_{int}$ ). And for the last,  $\sigma_i$  was set to  $1/y_0$  for  $E_{int} > 1$  mH (2.6 kJ/mol) and to  $\exp((1-y_0)/5)$  for  $E_{int} \leq 1$  mH.

## 2.3 Global and Local Geometry Optimization

In computational chemistry, the geometry optimization (also called energy minimization) is a process to find an arrangement in the space of atoms where the net inter-atomic force on each atom is close to zero. Figure 2.3 shows an example of an global and local minimum points and with given vector  $r$ , which describes the position of atoms, one can describe the concept of energy as a function of the positions,  $E(r)$ . By doing this, geometry optimization problem becomes a mathematical optimization and can be used to find the value of  $r$  for which  $E(r)$  is at the local minimum. That is, the derivative of the energy which respect to the positions of atoms. The main problem in locating the global and local minima of a cluster is the presence of a large number of local minima. These local points exponentially grows with the number of atoms in the searched space. Therefore, the main goal of the global and local geometry optimization is to escape from these local minima in a most efficient way. Computational demands of local optimization algorithms are expensive, therefore, heuristic algorithms such as Genetic Algorithm or Simulated Annealing (SA) are often used to locate global minimum regions because of their relatively low computational demands. To resolve the problem in locating the global minima, SA can manipulate the search domain to find the best solution. In this work, we used SA algorithm to find the global minima of adenine, thymine and adenine-thymine dimers and oligomers. Since SA might stuck within the local minimum and ignore the global minimum, we recursively execute SA search with a random set of initial points from our six-dimensional model to find the global minimum of PES.

### 2.3.1 Simulated annealing

Simulated Annealing is a metaheuristic algorithm to approximate the global optimum of a given function in large and discrete search spaces. Name Simulated Annealing comes from annealing process in metallurgy which involves heating and controlled cooling of a material to achieve desired crystal structure and material properties. SA algorithm mimics the annealing and cooling process to achive global minimum of given functions. The process starts with an initial temperature and this temperature decreases down until the minimum temperature was achieved, as in real annealing.



**Figure 2.3** : Global and Local Minimum Points.

In Global Optimization, this process determined by Metropolis criteria [77] seen in equation 2.12.

$$p = e^{-\frac{f_i - f'}{T_k}} . \quad (2.12)$$

In equation 2.12,  $f_i$ ,  $f'$  and  $T_k$  represents candidate point, current minimum and temperature respectively.  $p$  determines the acceptance process and if  $p$  is lower or higher than randomly generated value in the range of 0-1, than the new point will accepted. Also,  $T_k$  decreases each  $N_s * N_t$  interactions so higher number of  $N_s$  and  $N_t$  means higher possibilities to locate global minimum. New candidate points will be evaluated in  $N_s$  iterations with stepsize  $v$  and these evaluation repeated itself for  $N_t$  times.

$$v' = v \left( 1 + c \frac{n/N_s - 0.6}{0.4} \right) \quad (2.13)$$

Pseudo-code for SA algorithm can be seen bellow:

1. COMPUTE  $x' = x_i + rv_k$
2. IF  $f(x') < f_i$ , *ACCEPT*, ELSE, compute  $p'$
3. IF  $p' < p$ , *ACCEPT*, ELSE, *REJECT*,  $j = j + 1$
4. IF  $j < N_s$  *GO TO* 1, ELSE, *CONT.*

5. UPDATE  $v$  and  $k = k + 1$
6. IF  $k < N_t$  GO TO 1, ELSE, CONT.
7. reduce  $T$ ,
8. apply termination criteria,
9. IF *ACCEPT*, stop.
10. ELSE, GO TO 1





### 3. RESULTS AND DISCUSSION

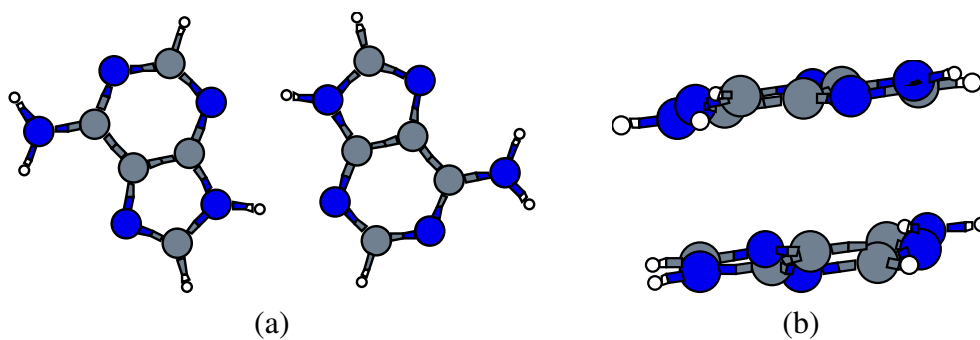
#### 3.1 Selection of the Theoretical Method for PES

Selection of the theoretical method for calculating PES is vital for our future steps. For this reason, we calculated the interaction energies of adenine, thymine and adenine-thymine dimers with using MP2, SCS-MP2, B3LYP-D, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC). CCSD(T) calculations were used for comparison of the performances of each method. The most accurate and fast method will be selected for our future investigations.

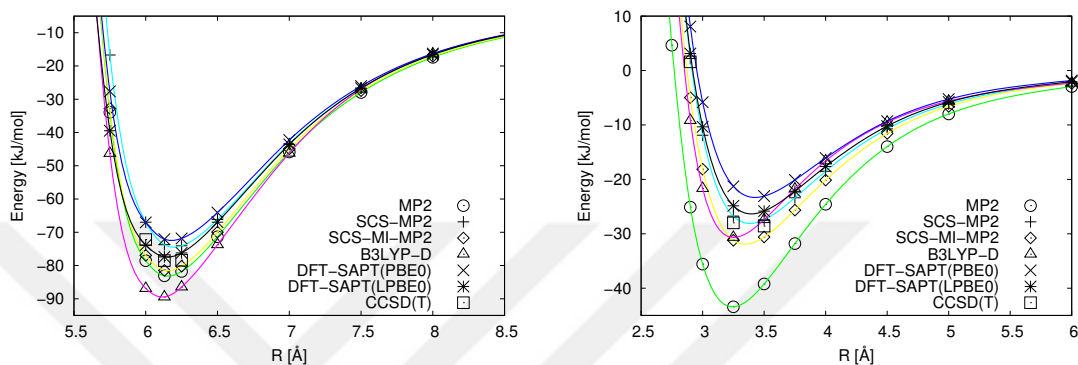
##### 3.1.1 Adenine dimers

Interaction energies were found with using CP-corrected supermolecular methods (MP2,SCS-MP2, SCS-MI-MP2, B3LYP-D and CCSD(T)) and DFT-SAPT, as a function of the intermolecular distance R. For H-bonded adenine dimer shown in Figure 3.1. B3LYP-D overestimates interaction energies compared to CCSD(T) while DFT-SAPT (PBE0AC) and SCS-MP2 gives better results. It still underestimates the energies compared to CCSD(T) and SCS-MI-MP2 performs better than SCS-MP2. DFT-SAPT (LPBE0AC) and SCS-MI-MP2 are in good agreement with CCSD(T). For the stacked adenine shown in Figure 3.1, however, the performance of the used methods differs. MP2 generally overestimates the interaction energies compared to CCSD(T) and such performance of MP2 has already been reported for cytosine [46], acetylene-furan [78], pyrazine and triazine dimers [79]. Even though SCS-MP2 corrects MP2, it still overestimates the interaction energies compared to CCSD(T). SCS-MI-MP2 performs better than SCS-MP2 and with DFT-SAPT (LPBE0AC), they are still in good agreement with CCSD(T).When these results were evaluated together, we found that the closest results to CCSD(T) were obtained by LPBE0AC and it was followed by SCS-MI-MP2. Table 3.2 shows us the interpolated energies and the corresponding R distances of the PECs shown in Figure 3.2. From Figure 3.2, we can

clearly see that the best alternative method for CCSD(T) is DFT-SAPT(LPBE0AC). DFT-SAPT(LPBE0AC) shows minor differences for H-bonded and stacked dimers, and followed by SCS-MI-MP2. Table 3.1 shows us the dependence between interaction energies and basis set for the H-bonded and stacked dimers. In Table 3.1, we also showed that the extrapolated interaction energies to the complete basis set (CBS) limit using aVTZ and aVQZ energies from the two-point formula of Bak et al. [80]. Because of its very demanding nature, we could not employ these calculations to the CCSD(T) and we employ a different method to overcome this, which is shown at Janowski et al. [81]. For obtaining the extrapolated interaction energies, we assumed that the difference between CCSD(T) and SCS-MP2 interaction energies was constant and the CBS limit for CCSD(T) was determined by the addition of this difference to the extrapolated SCS-MP2 energies. We assume that the using MP2 energies instead of SCS-MP2 energies creates negligible effect in CCSD(T) energies. For calculating the DFT-SAPT energies, we used only  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$  from the extrapolation formula and the remaining terms were taken from aVQZ level. This approach was in good agreement as shown in Figure 3.5. In this figure, we can clearly see that except  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$ , the other energy components converges at the aug-cc-pVTZ basis set. We extrapolated the electron correlation ( $E_{corr}$ ) and total interaction energy with MP2 (SCS-MP2 and SCS-MI-MP2) and B3LYP-D, respectively. As it is shown in Table 3.1, the energies obtained from the MP2 (SCS-MP2 and SCS-MI-MP2) differs from the aVTZ to aVQZ and from aVQZ to the CBS limit, respectively 4-7 kJ/mol for H-bonded dimers. We also showed that the use of larger basis set in B3LYP-D has no effect on the calculation of the interaction energies. For H-Bonded dimers, it is shown that the SCS-MI-MP2 gives the best performance among the others. But for stacked dimers, DFT-SAPT (LPE0AC) was more compatible with CCSD(T) than SCS-MI-MP2. All these results indicate that DFT-SAPT (LPE0AC) can be used in the calculation of potential energy surface (PES). There was also one feature of DFT-SAPT that we think important; its ability to decompose the total energy into their components. This gives us the easy way to determine which forces were responsible for the stabilizing factors in the dimers. Figure 3.3 shows this effects. As indicated in this figure,  $E_{ind}^{(2)}$  and  $E_{disp}^{(2)}$  were always been negative for H-bonded dimer but this changes for higher cms values for the stacked dimer. It is shown that the magnitude of the contributions were changed

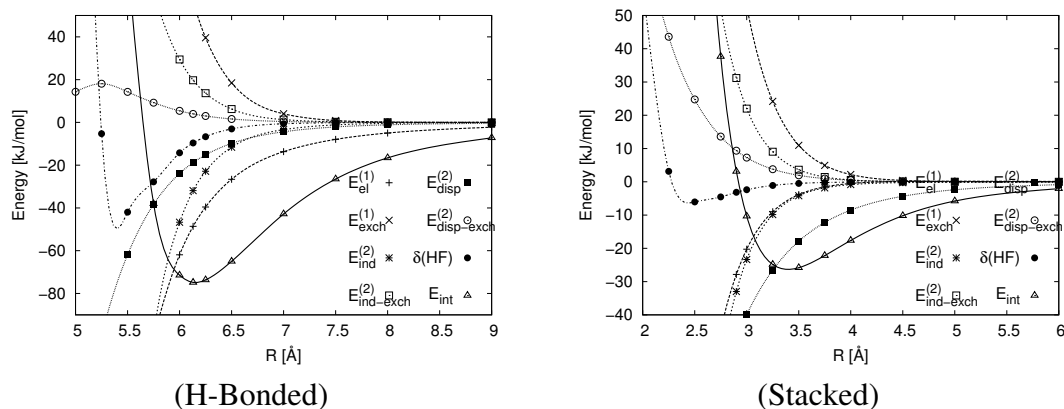


**Figure 3.1** : Adenine dimers: a) H-bonded A, b) Stacked B.



**Figure 3.2** : Potential energy curves of Adenine Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black

according to the following order,  $E_{exch-disp}^{(2)} \leq E_{exch-ind}^{(2)} \leq E_{exch}^{(1)} \leq E_{exch}^{(1)}$ . The share of  $E_{exch}^{(1)}$  in these terms is more than %70.



**Figure 3.3** : DFT-SAPT(LPBE0AC) energy contributions for the H-bonded and Stacked Adenine dimers. Long-dashed lines represent  $E_{el}^{(1)}$  and  $E_{exch}^{(1)}$ , short-dashed  $E_{ind}^{(2)}$  and  $E_{exch-ind}^{(2)}$ , dotted  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$ , dotted long-dashed  $\delta(HF)$ , and the solid line  $E_{int}$ .

**Table 3.1** : Interaction Energies of Chosen Adenine Dimers. H-bonded Dimer A (C.m.s. 6.13 Å) and Stacked B (C.m.s. 3.25 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

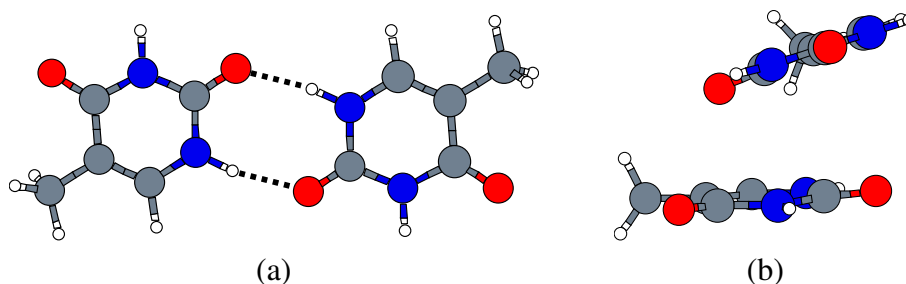
Isomer	X	$E_{int}$						
		MP2	SCSMP2	MI-MP2	B3LYPD	PBE0	LPBE0	CCSD(T)
A	D	-83.09	-73.74	-81.46	-89.47	-72.40	-77.40	-78.51
	T	-87.76	-78.27	-84.74	-89.62	-73.45	-78.99	
	Q	-89.47	-80.08	-85.71	-89.74	-74.42	-80.01	
	cbs*	-90.66	-81.35	-86.35	-89.83	-75.07	-80.68	-83.4
B	D	-43.42	-26.75	-31.19	-30.66	-23.64	-27.22	-27.99
	T	-46.79	-29.65	-33.86	-31.66	-24.41	-28.15	
	Q	-47.98	-30.76	-34.68	-31.69	-25.48	-29.25	
	cbs*	-48.80	-31.53	-35.24	-31.71	-26.21	-30.01	-32.04

**Table 3.2 :** Adenine A and B isomers for aug-cc-pvdz basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation.

Dimer	Method	C.m.s. Å	<i>kJ/mol</i>
A	MP2	6.13	-83.10
	SCS-MP2	6.21	-74.51
	MI-MP2	6.13	-81.47
	B3LYP-D	6.11	-89.70
	PBE0AC	6.20	-69.88
	LPBE0AC	6.13	-74.84
	CCSD(T)	6.19	-79.19
B	MP2	3.23	-43.52
	SCS-MP2	3.41	-28.05
	MI-MP2	3.37	-31.72
	B3LYP-D	3.23	-30.73
	PBE0AC	3.44	-23.38
	LPBE0AC	3.42	-26.34
	CCSD(T)	3.41	-28.85

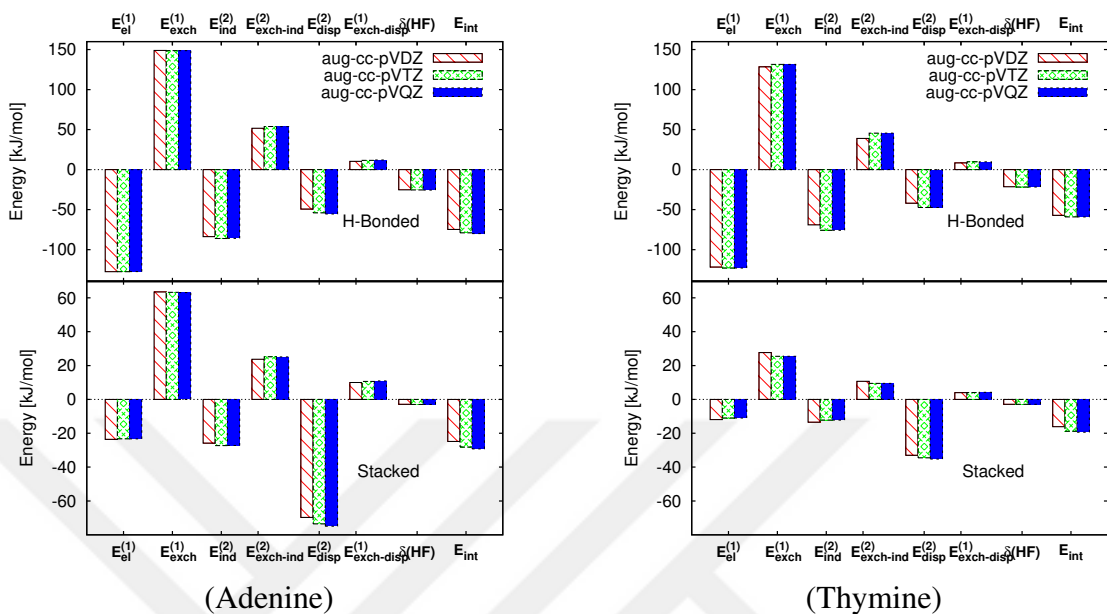
### 3.1.2 Thymine dimers

We used the same approach used in adenine dimers for calculating the PECs of the thymine dimers. For H-bonded thymine dimer A shown in Figure 3.4, B3LYP-D still overestimates interaction energies compared to CCSD(T), just like we saw in the adenine H-bonded dimer. SCS-MP2 and DFT-SAPT(PBE0AC) are very close to each other and gives similar results, but they both underestimates energies compared to CCSD(T). SCS-MI-MP2 gives better results compared to B3LYP-D, SCS-MP2 and DFT-SAPT(PBE0AC). Still, MP2 and DFT-SAPT(LPBE0AC) are the best methods when we compare the interaction energies with the CCSD(T), as we observed in adenine dimers. For stacked dimer B, however, SCS-MI-MP2 differs from CCSD(T) by just 0.72kJ/mol and gives better results than MP2. DFT-SAPT(LPBE0AC) is still the most accurate method and it differs from CCSD(T) by 0.39kJ/mol. When these results were evaluated together, we showed that our previous observations in adenine dimers still holds for the thymine dimers as well. The closest results to CCSD(T) were obtained by LPBE0AC and it followed by SCS-MI-MP2. Table 3.3 shows us the

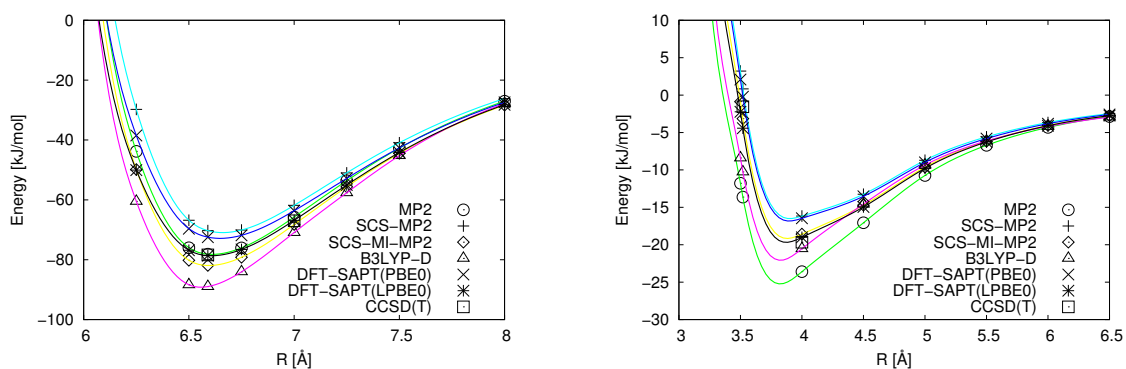


**Figure 3.4** : Thymine dimers: a) H-bonded A, b) Stacked

interpolated energies and the corresponding R distances of the PECs shown in Figure 3.6. For calculating the DFT-SAPT energies, we used our previous approach shown in adenine dimer and only used  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$  from the extrapolation formula. Remaining terms were taken from aVQZ level as we did in adenine. This approach was also in good agreement when it comes to thymine dimers and can be seen in Figure 3.6. The energy components other than the  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$  converges on the aug-cc-pVTZ basis set as we observed in adenine dimers. We extrapolated the electron correlation, ( $E_{corr}$ ), for thymine bases as well and the total interaction energy with MP2 (SCS-MP2 and SCS-MI-MP2) and B3LYP-D. As it is shown in Table 3.4, the energies obtained from MP2 (SCS-MP2 and SCS-MI-MP2) differs from the aVTZ to aVQZ and from aVQZ to the CBS limit, respectively 1-6 kJ/mol for H-bonded dimers. We also showed again that the use of larger basis sets in B3LYP-D has no effect on the calculation of the interaction energies in thymine dimers as well. Also, we have shown that the SCS-MI-MP2 gives the best performance among others for H-Bonded dimers. But for stacked dimers, DFT-SAPT (LPE0AC) was more compatible with CCSD(T) than SCS-MI-MP2. All these results indicate that DFT-SAPT (LPE0AC) can also be used in the calculation of potential energy surface (PES) of thymine.



**Figure 3.5** : Basis set dependence of DFT-SAPT(LPBE0AC) energy components for Adenine and Thymine dimers.



**Figure 3.6** : Potential energy curves of Thymine Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black

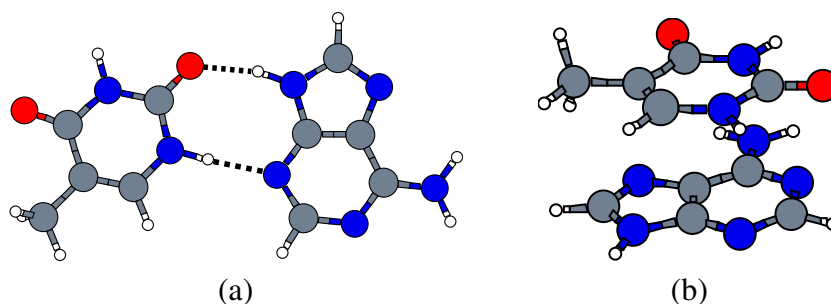
**Table 3.3** : Thymine A and B isomers for aug-cc-pVDZ basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation.

Dimer	Method	C.m.s. Å	<i>kJ/mol</i>
A	MP2	6.58	-78.12
	SCS-MP2	6.58	-70.10
	MI-MP2	6.60	-81.86
	B3LYP-D	6.57	-89.25
	PBE0AC	6.68	-70.78
	LPBE0AC	6.50	-74.87
	CCSD(T)	6.59	-78.38
B	MP2	3.96	-23.92
	SCS-MP2	3.93	-15.88
	MI-MP2	3.99	-18.63
	B3LYP-D	3.98	-20.71
	PBE0AC	4.30	-15.01
	LPBE0AC	4.00	-18.96
	CCSD(T)	3.99	-19.35

### 3.1.3 Adenine-Thymine dimers

The same approach used in adenine and thymine bases were also used on adenine-thymine and our results are similar to the previous applications. For H-bonded adenine-thymine dimer A shown in Figure 3.7, MP2, SCS-MI-MP2 and DFT-SAPT (LPBA0AC) gives the best results. B3LYP-D still overestimates the interaction energies while MP2 and DFT-SAPT(PBE0AC) underestimate in contrast to CCSD(T). In stacked B shown in Figure 3.7 however, MP2 performs worse than B3LYP-D as we also observed in adenine and thymine dimers. Table 3.5 shows us the interpolated energies and the corresponding R distances of the PECs shown in Figure 3.8. the DFT-SAPT energies calculated as the same way with adenine and thymine dimers and achieved results can be seen in Figure 3.9. Remaining terms were taken from aVQZ level as we did in adenine and thymine dimers. This approach was again proved itself and in good agreement when it comes to adenine-thymine dimers as well. Figure 3.9 shows this fact clearly and as it can be seen, the energy components other than the  $E_{disp}^{(2)}$  and  $E_{exch-disp}^{(2)}$  converges on the aug-cc-pVTZ basis set. Electron correlation energy



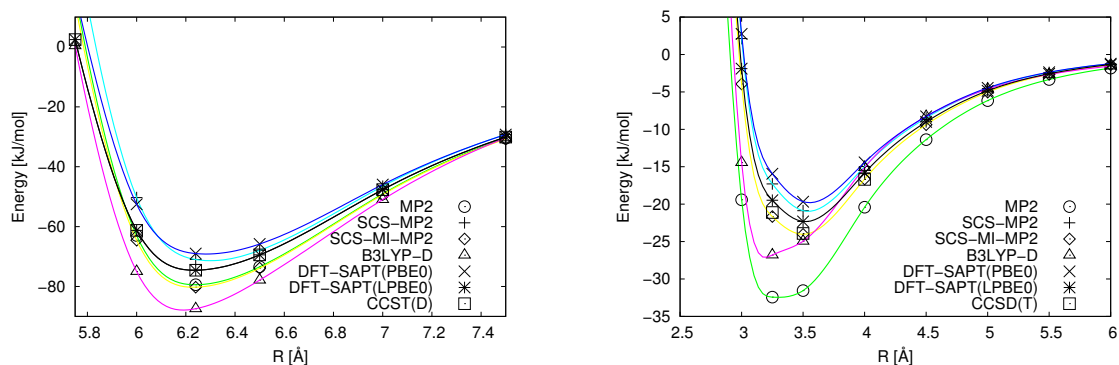


**Figure 3.7** : Adenine-Thymine dimers: a) H-bonded A, b) Stacked

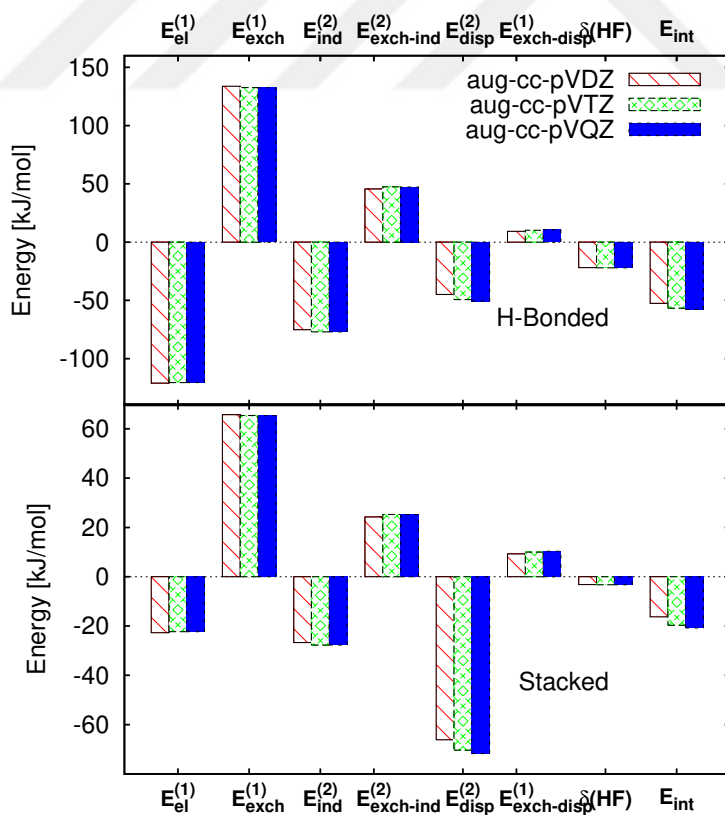
**Table 3.4** : Interaction Energies of Chosen H-bonded Tyimine A (C.m.s. 6.59 Å ) and Stacked B (C.m.s. 3.52 Å ) with basis sets aug-cc-pVXZ(X=D, T and Q). aug-cc-pVTZ and aug-cc-pVQZ were used for the extrapolation of cbs

Isomer	X	$E_{int}$						
		MP2	SCSMP2	MI-MP2	B3LYPD	PBE0	LPBE0	CCSD(T)
A	D	-78.14	-70.28	-81.93	-88.96	-70.38	-76.51	-78.37
	T	-82.93	-74.78	-85.39	-89.58	-74.55	-80.95	
	Q	-84.84	-76.78	-86.56	-89.90	-75.71	-82.22	
	cbs*	-86.01	-78.02	-87.21	-90.12	-76.31	-82.82	-83.4
B	D	-13.63	0.84	-3.15	-10.28	-0.22	-4.42	-4.08
	T	-18.55	-3.43	-6.97	-10.62	-4.77	-9.22	
	Q	-20.08	-4.86	-7.99	-10.67	-6.18	-10.68	
	cbs*	-21.13	-5.87	-8.70	-10.72	-7.18	-11.73	-9.63

( $E_{corr}$ ) extrapolated for adenine-thymine bases as well, and total interaction energy with MP2 (SCS-MP2 and SCS-MI-MP2) and B3LYP-D was calculated with the same fashion. As it is shown in Table 3.4, the energies obtained from the MP2 (SCS-MP2 and SCS-MI-MP2) differs from the aVTZ to aVQZ and from aVQZ to the CBS limit, respectively by 4-7 kJ/mol for H-bonded and by 2-5 kJ/mol for stacked dimers. Usage of larger basis sets in B3LYP-D still has no effect on the calculation of the interaction energies as we observed in adenine and thymine dimers.



**Figure 3.8** : Potential energy curves of Adenine-Thymine Dimers. H-bonded A and Stacked B dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black



**Figure 3.9** : Basis set dependence of DFT-SAPT(LPBE0AC) energy components for Adenine-Thymine Dimers.

**Table 3.5** : Adenine-Thymine A and B isomers for aug-cc-pvdz basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation.

Dimer	Method	C.m.s. Å	<i>kJ/mol</i>
A	MP2	6.22	-79.53
	SCS-MP2	6.25	-71.15
	MI-MP2	6.23	-80.43
	B3LYP-D	6.207	-87.66
	PBE0AC	6.29	-68.97
	LPBE0AC	6.22	-74.63
	CCSD(T)	6.34	-79.41
B	MP2	3.42	-32.66
	SCS-MP2	3.34	-19.39
	MI-MP2	3.48	-24.26
	B3LYP-D	3.48	-26.01
	PBE0AC	3.46	-19.88
	LPBE0AC	3.51	-22.23
	CCSD(T)	3.49	-23.93

**Table 3.6** : Interaction Energies of Chosen H-bonded Thymine A (C.m.s. 6.59 Å ) and Stacked B (C.m.s. 3.52 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

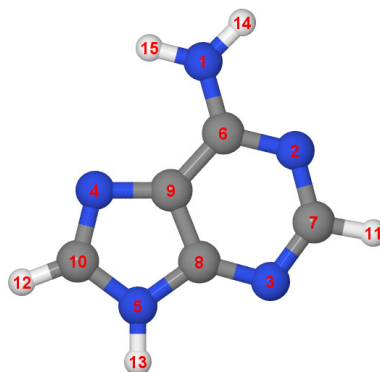
Isomer	X	<i>E<sub>int</sub></i>						
		MP2	SCSMP2	MI-MP2	B3LYPD	PBE0	LPBE0	CCSD(T)
A	D	-79.42	-70.99	-80.24	-87.44	-69.01	-74.56	-77.32
	T	-83.97	-75.33	-83.52	-87.80	-72.96	-78.79	
	Q	-85.70	-77.15	-84.55	-88.08	-73.99	-79.90	
	cbs*	-88.83	-78.35	-85.18	-88.24	-74.65	-80.37	-82.26
B	D	-31.55	-20.87	-24.03	-24.96	-19.67	-22.31	-23.92
	T	-33.72	-22.73	-25.72	-25.54	-21.64	-24.39	
	Q	-35.59	-24.55	-25.58	-26.29	-21.87	-24.67	
	cbs*	-36.38	-25.32	-25.52	-26.61	-22.47	-25.27	-25.63

## 3.2 Fitting Potential Energy Surfaces of DNA Bases

In summary, for all considered adenine, thymine, and adenine-thymine dimers, the most compatible method with CCSD(T) was found to be DFT-SAPT(LPBE0AC), which was followed by SCS-MI-MP2. Due to this superior performance of DFT-SAPT (LPBE0AC), potential energy surfaces of these dimers will be calculated by this method.

### 3.2.1 Adenine intermolecular function

In order to calculate the six dimensional potential energy surface (PES) of adenine dimer, we used CC2/aug-cc-pVTZ optimized fixed geometry of adenine monomer as shown in Figure 3.10. The first dimension,  $R$ , defines the center of mass(cms)distances between monomers.  $R$  was defined as following set of distances  $R=3.0, 4.0, 5.0, 6.0, 7.0, 9.0$  and  $12.0 \text{ \AA}$ . Two polar angles with  $[0, 45 \text{ and } 90^\circ]$  and  $[29.29, 88.08, 152.34, 171.93, 203.84, 249.03, 271.03, 313.16 \text{ and } 350.80^\circ]$  was defined to vary the position of the center of mass vector. As it can be seen, the second polar angle has an unregular grid. The reason of the employment such grid is to allow the second adenine to settle around the hydrogens and nitrogens of the first adenine. In order to define the orientation of the second adenine monomer, we defined three polar angles,  $[0, 45, 90 \text{ and } 270^\circ]$  and  $[0, 45, 90, 135, 180, 225, 270 \text{ and } 315^\circ]$ . This angles allows us to generate h-bonded and stack adenine dimers. After the elimination processes of symmetric and close-contact orientations, this grid generated 7286 dimer geometries. The interaction energies of these geometries in the computational grid were calculated with the DFT-SAPT(LPBE0AC)/aVDZ. We used the Levenberg-Marquardt nonlinear weighted least squares method for fitting. In the end of these calculations, we obtained parameters listed in Table 3.7. When all the grid points were taken into account, the standard deviation of fit was obtained as 16.01 kJ/mol. But when the interaction energy  $E_{int}$  is less than or equal to 2.6 kJ/mol, which is true for 5897 dimer structures, the standard deviation of this fit was obtained as 1.70 kJ / mol. In this case, the standard deviation for the other 1386 dimer structures was obtained as 36.51 kJ/mol. Figure 3.11 and 3.11 compares the model potential with DFT-SAPT (LPBE0AC)/aug-cc-pVDZ. As it can be seen from these figures, our fitting is quite successful, especially for the



**Figure 3.10** : Adenine Monomer.

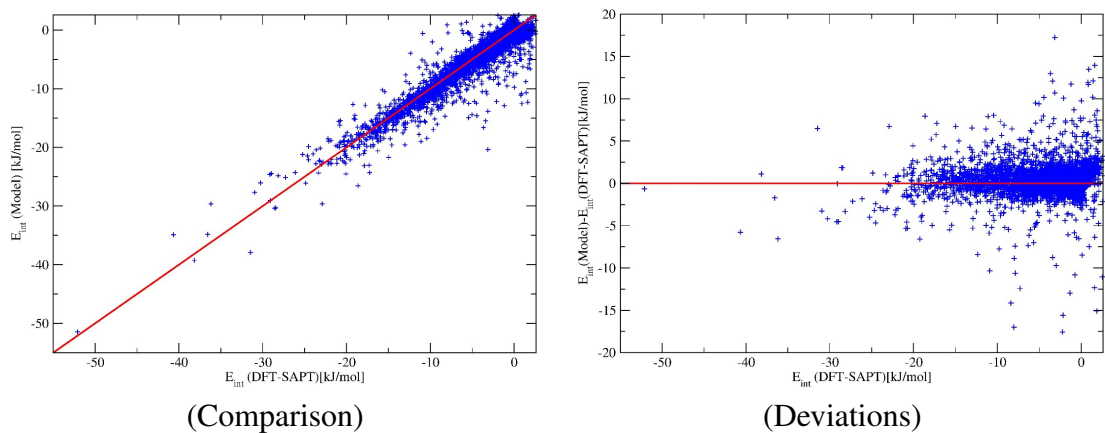
**Table 3.7** : Fitted parameters for Adenine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of first and second adenine monomer, respectively.

i-j	$\alpha[H]$	$\beta[b^{-1}]$	$C[Hb^6]$	$c[b]$
C-C	-2.177	0.901	430.967	1.60
C-H	-0.224	0.944	26.554	1.20
C-N	-0.002	0.082	97.777	1.75
N-H	0.004	0.097	4.571	1.25
N-N	-0.383	0.657	239.596	1.00
H-H	-0.007	0.140	6.395	1.25

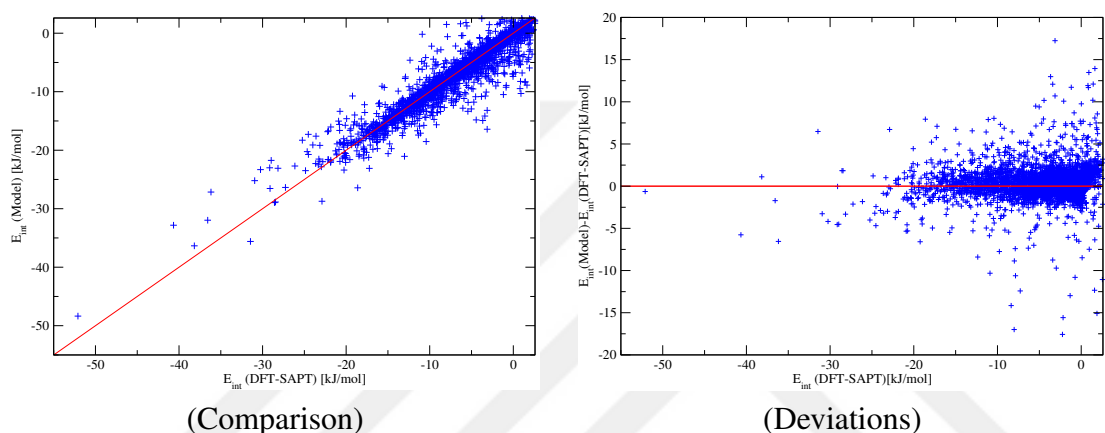
low-energy dimer structures. However, for structures with an interaction energy  $E_{int} > 30$  kJ/mol, the fitting performance has begun to fall. This showed us that our fitting is promising, especially for the low-energy dimer structures.

### 3.2.2 Thymine and Adenine-Thymine intermolecular functions

The same fitting approach used in adenine dimers was also used for thymine and adenine-thymine dimers. The same R distances  $R=3.0, 4.0, 5.0, 6.0, 7.0, 9.0$  and  $12.0 \text{ \AA}$  was used for both thymine and mobile thymine monomer of adenine-thymine dimers. Two polar angles defined as  $[0, 45 \text{ and } 90^\circ]$  and  $[3.95, 61.91, 127.92, 195.74, 257.46, 310.37 \text{ and } 335.19^\circ]$  for thymine and  $[0, 45 \text{ and } 90^\circ]$  and  $[29.29, 88.08, 152.34, 171.93, 203.84, 249.03, 271.03, 313.16 \text{ and } 350.79^\circ]$  for mobile thymine to vary the position of the center of mass vector. Three polar angles,  $[0, 45, 90 \text{ and } 270^\circ]$  and  $[0, 45, 90, 135, 180, 225, 270 \text{ and } 315^\circ]$  defines the orientation of the second thymine and mobile thymine monomers as we also discussed in adenine intermolecular function. After the elimination process of symmetric and close-contact orientations, thymine grid generated 4412 dimer geometries and adenine-thymine grid generated 6390 dimer



(Comparison) (Deviations)  
**Figure 3.11** : Comparison and deviations of the Adenine model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies.



(Comparison) (Deviations)  
**Figure 3.12** : Comparison and deviations of the Thymine model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies.

geometries. Table 3.8 and Table 3.9 show the obtained parameters for thymine and adenine-thymine dimers. Also, Figure 3.12 compares the thymine model potential with DFT-SAPT (LPBE0AC)/aug-cc-pVDZ.

**Table 3.8** : Fitted parameters for Thymine 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 adenine monomer, respectively.

i-j	$\alpha[H]$	$\beta[b^{-1}]$	$C[Hb^6]$	$c[b]$
C-C	-0.293	0.702	224.806	1.00
C-H	-1.862x10e-5	0.046	10.781	0.25
C-N	-2.855x10e-5	0.252	16.456	1.00
C-O	-0.068	0.661	85.961	1.00
N-H	0.031	0.579	21.321	0.75
H-O	-0.337	1.111	12.449	1.00
N-N	-3.471	0.996	440.497	1.00
N-O	-0.326	0.674	200.234	1.00
H-H	-0.036	0.799	6.087	1.00
O-O	-0.789	0.966	161.944	1.15

**Table 3.9** : Fitted parameters for Adenine-Thymine 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 adenine monomer, respectively.

i-j	$\alpha[H]$	$\beta[b^{-1}]$	$C[Hb^6]$	$c[b]$
C-C	-0.238	0.785	107.88	1.00
C-H	-0.290	1.03	30.18	1.00
C-N	-0.002	0.26	44.26	1.00
N-H	-0.012	0.54	6.68	1.00
N-N	-1.08	0.78	443.90	1.00
H-H	-0.009	0.67	5.90	1.00
O-H	-0.007	0.42	5.09	1.00
O-N	-2.59	0.968	394.00	1.00
O-C	-0.227	0.72	127.13	1.00

### 3.3 Global Optimization of the DNA Bases via Simulated Annealing

Global optimization of the DNA bases is the main objective of this thesis and previous steps are essential for achieving to this target. From now on, we use the fitting parameters for each DNA base to uncover the most stable adenine, thymine and adenine-thymine dimers and oligomers. We used SA [82] implementation to search the resulting 6-dimensional PES calculated with DFT-SAPT and the approval of the movement on the PES determined by Metropolis criterion [77] and the annealing temperature was taken as half of the previous steps.

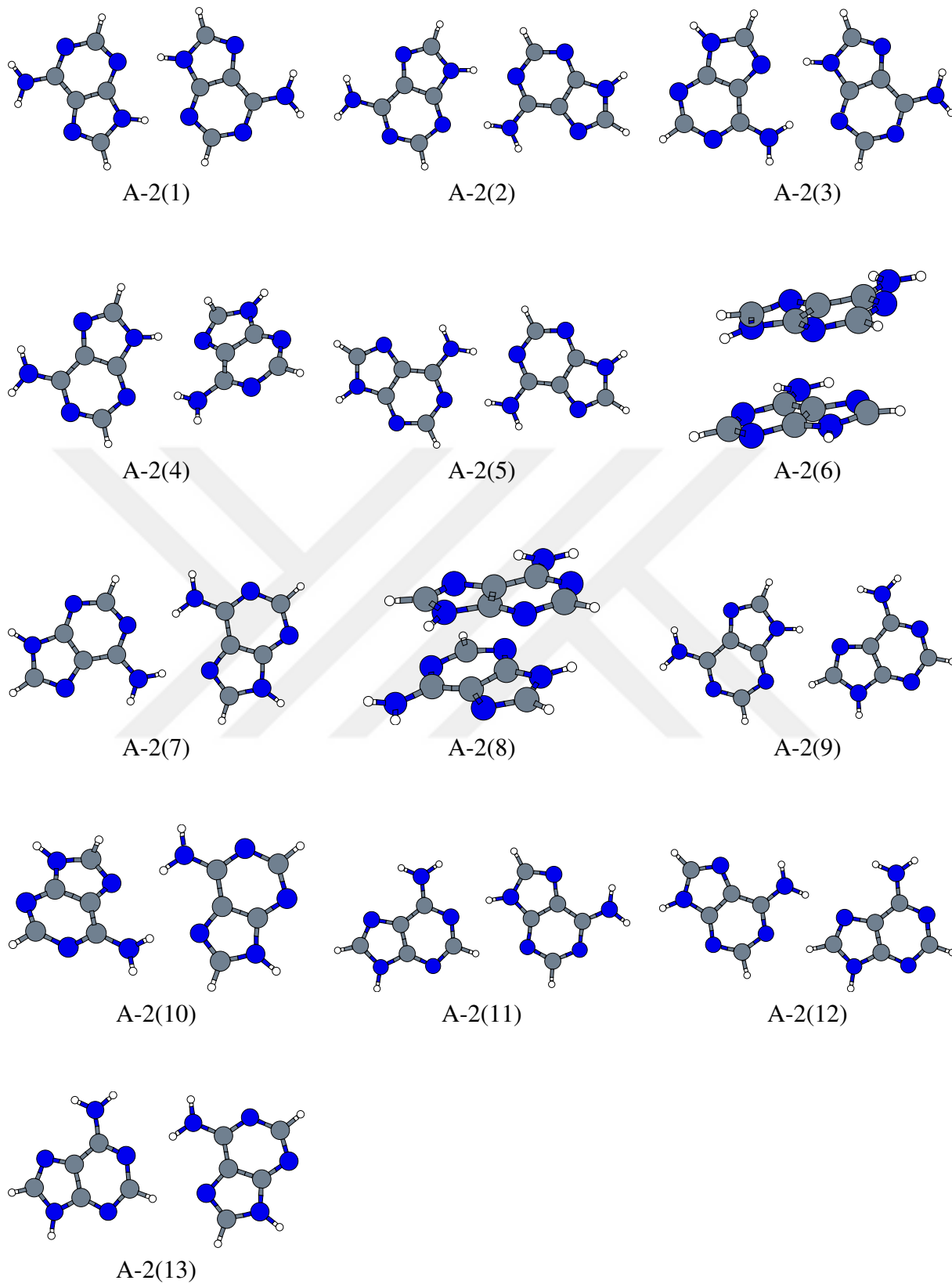
#### 3.3.1 Adenine clusters

##### 3.3.1.1 Dimer

We detected 12 different adenine dimer structures with our model shown in Figure 3.13. We achieve our primary objective and have found the previously reported structures in the literature [6, 7, 69, 83]. Our model also be able to detect reported as Self-assembled adenine-dimer chains on Cu (110) in Preuss et al. [84] and these orientations named as A-2(1), A-2(10) and A-2(11). Proposed force field even manages to generate new two stacked and six H-bonded structures, A-2(6), A-2(8) and A-2(4), A-2(10) respectively. This is a sign that our current potential works very well for adenine dimers. Resulting adenine dimer structures further relaxed with PBE, SCS-MP2 and CP-SCS-MP2 using aug-cc-pVDZ basis set to clarify the quality of the predicted orientations. Structures obtained from SA were slightly changed upon the ab-initio relaxations, however, A-2(4) and A-2(8) changes their orientations again, after the SCS-MP2/aVDZ optimization. A-2(4) transformed into stacked orientation after SCS-MP2 and CP-SCS-MP2, while A-2(8) transformed into a H-bonded orientation. There was also a notable difference in A-2(7) after SCS-MP2/aVDZ relaxations, but it preserved its planarity. In Table 3.10, we showed that the interaction energies obtained from B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ basis set. We showed that B3LYP-D overestimates the interaction energies and MP2 and SCS-MI-MP2 were quite good agreement for H-bonded structures. For stacked



geometries, B3LYP-D generally overestimates the interaction energies about 10 kJ/mol for such orientations and SCS-MP2 gives the best results in between MP2-based methods. LPBE0AC is in agreement with SCS-MI-MP2 while generating lower interaction energies than PBE0AC. Generally, the magnitude of the interaction energy was obtained in the following order:  $B3LYP-D < MP2 \approx SCS-MI-MP2 \approx DFT-SAPT(LPBE0AC) < SCS-MP2 \approx DFT-SAPT(PBE0AC)$ . Our model predicted A-2(1) as the lowest energy orientation, lower than 10 and 15 kJ/mol from the closest orientations A-2(2) and A-2(3) respectively, and this was unchanged for all ab-initio calculations. The lowest energy order remains the same for all ab-initio calculations and this can be seen in Table 3.10. This agreement can give us an idea about the prediction power of our model potential. In order to confirm this, we further calculated the interaction energies of dimers that optimized in CP-SCS-MP2/aVDZ. As it can be seen in Table 3.10, A-2(1) is still the lowest energy after the CP-SCS-MP2 relaxation. A-2(1) was followed by A-2(4), A-2(3) and A-2(2). Result of these calculations shows that the lowest energy orientation is indeed A-2(1) and, we can say that our model successfully located the structures in the literature [6, 7, 69, 83] and their corresponding energy ordering generally stays same when compared with CP-SCS-MP2 relaxed geometries. For adenine dimers, AMBER and our model are in good agreement for most cases but AMBER performs slightly worse for H-bonded geometries like A-2(9) and A-2(12) when we compare it to our force field, which is in good agreement with ab-initio calculations. When the suggested geometries relaxed with CP-SCS-MP2/aVDZ level, AMBER has been able to give more accurate results in contrast to its previous performance. These results give us the opportunity to extend our model for the larger DNA homo-oligomers.



**Figure 3.13** : Adenine dimer structures found by the SA approach.

**Table 3.10** : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine dimers shown in Figure 3.13.

Dimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	PBE0AC	LPBE0AC	AMBER	Model
A-2(1) <sup>a</sup>	-80.28	-74.76	-66.16	-72.92	-62.38	-67.07	-66.89	-60.99
A-2(2) <sup>a</sup>	-64.95	-58.17	-50.70	-55.85	-48.61	-52.43	-55.47	-50.57
A-2(3) <sup>a</sup>	-58.03	-52.69	-45.40	-50.38	-43.82	-47.38	-47.67	-45.00
A-2(4) <sup>a</sup>	-53.58	-49.11	-41.44	-46.68	-40.33	-43.64	-47.31	-43.03
A-2(5) <sup>a</sup>	-54.04	-58.17	-39.16	-43.20	-30.23	-41.59	-46.20	-42.51
A-2(6) <sup>a</sup>	-38.16	-49.77	-35.11	-39.27	-35.70	-33.27	-41.57	-40.30
A-2(7) <sup>a</sup>	-34.61	-42.45	-36.03	-39.94	-26.92	-38.51	-39.92	-38.04
A-2(8) <sup>a</sup>	-53.58	-45.50	-31.42	-35.48	-38.60	-29.97	-40.00	-36.89
A-2(9) <sup>a</sup>	-46.52	-42.36	-35.66	-41.37	-35.05	-38.39	-16.98	-35.02
A-2(10) <sup>a</sup>	-42.16	-38.03	-31.33	-35.40	-31.29	-33.81	-35.99	-34.03
A-2(11) <sup>a</sup>	-18.76	-37.07	-31.17	-35.95	-30.65	-32.96	-30.83	-29.85
A-2(12) <sup>a</sup>	-34.25	-29.28	-23.42	-28.00	-24.27	-26.97	-10.17	-27.99
A-2(13) <sup>a</sup>	-25.13	-20.10	-14.31	-18.72	-16.05	-18.69	-10.73	-23.08
A-2(1) <sup>b</sup>	-92.39	-86.45	-76.96	-84.84	-71.54	-76.89	-69.68	-60.25
A-2(2) <sup>b</sup>	-71.61	-65.74	-57.79	-63.69	-55.46	-59.61	-59.19	-46.45
A-2(3) <sup>b</sup>	-72.10	-66.42	-58.57	-64.35	-55.56	-59.72	-53.40	-46.48
A-2(4) <sup>b</sup>	-71.97	-66.60	-58.83	-64.59	-62.59	-61.93	-37.43	-47.18
A-2(5) <sup>b</sup>	-63.28	-55.37	-48.76	-52.98	-31.28	-33.69	-43.69	-42.93
A-2(6) <sup>b</sup>	-44.48	-56.91	-38.86	-44.32	-33.45	-37.04	-35.66	-40.89
A-2(7) <sup>b</sup>	-57.17	-51.66	-45.73	-49.37	-44.73	-47.92	-53.37	-38.19
A-2(8) <sup>b</sup>	-45.97	-58.89	-40.06	-46.18	-38.27	-39.28	-48.79	-37.87
A-2(9) <sup>b</sup>	-53.05	-49.45	-42.88	-49.12	-31.28	-44.77	-34.52	-32.63
A-2(10) <sup>b</sup>	-53.42	-48.78	-42.06	-46.43	-41.39	-44.15	-41.51	-36.42
A-2(11) <sup>b</sup>	-46.90	-42.50	-36.60	-41.49	-41.33	-38.57	-33.86	-28.44
A-2(12) <sup>b</sup>	-40.73	-36.81	-31.66	-35.57	-31.28	-33.69	-27.31	-26.44
A-2(13) <sup>b</sup>	-33.95	-30.34	-25.45	-29.02	-25.80	-27.98	-27.29	-24.81

<sup>a</sup> Using model geometries.

<sup>b</sup> Using CP-SCS-MP2/aVDZ geometries.

### 3.3.1.2 Trimer

For adenine, thymine and adenine-thymine trimers and beyond, the total interaction energies in the SA approach was calculated by summing all possible dimer interactions in the system as shown in the following formula:

$$\Delta E_{int}^{CP} = E_{ABC}^{ABC} - E_A^{ABC} - E_B^{ABC} - E_C^{ABC} \quad (3.1)$$

These total trimer interaction energy may be broken down into two- and three-body components as follows:

$$\Delta^2 E_{int,AB}^{CP} = E_{AB}^{ABC} - E_A^{ABC} - E_B^{ABC} \quad (3.2)$$

$$\Delta^2 E_{int,AC}^{CP} = E_{AC}^{ABC} - E_A^{ABC} - E_C^{ABC} \quad (3.3)$$

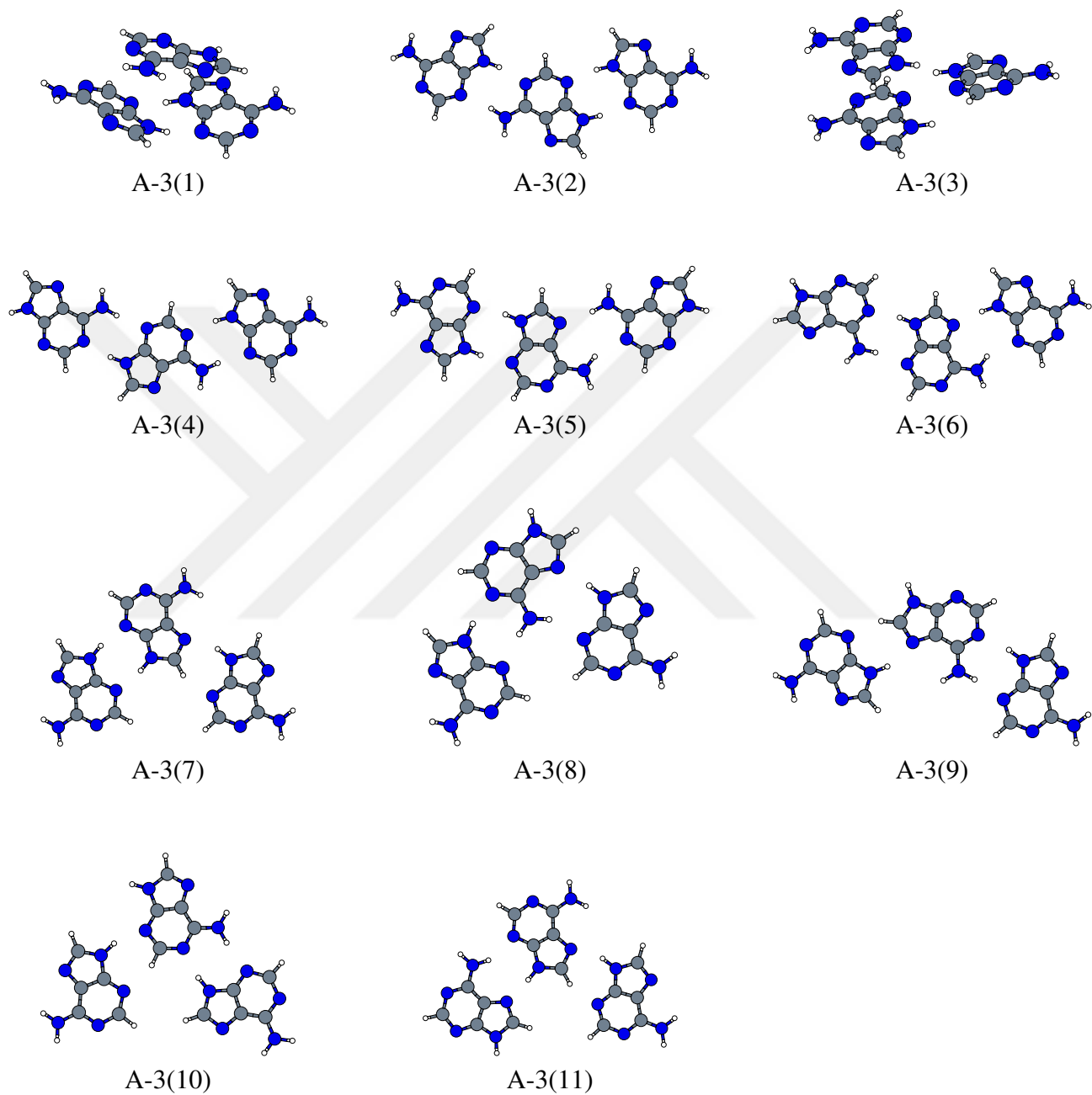
$$\Delta^2 E_{int,BC}^{CP} = E_{BC}^{ABC} - E_B^{ABC} - E_C^{ABC} \quad (3.4)$$

The total interaction energy is then written as a sum of these two-body interaction energies plus a three-body interaction energy,

$$\Delta E_{int}^{CP} = \Delta^2 E_{int,AB}^{CP} + \Delta^2 E_{int,AC}^{CP} + \Delta^2 E_{int,BC}^{CP} + \Delta^3 E_{int,ABC}^{CP} \quad (3.5)$$

where  $\Delta E_{int}^{CP}$  is the total trimer interaction energy,  $\Delta^2 E_{int,AB}^{CP}$ ,  $\Delta^2 E_{int,AC}^{CP}$  and  $\Delta^2 E_{int,BC}^{CP}$  are the two-body components and  $\Delta^3 E_{int,ABC}^{CP}$  is the three-body term. Our model predicted 11 adenine trimers shown in Figure 3.14 and predicted structures were relaxed with only SCS-MP2/aVDZ because of the high computational demand. Amongst them, the most stable adenine trimers, A-3(1), A-3(2) and A-3(7) are a combination of the lowest energy adenine dimers A-2(1), A-2(2), A-2(3), respectively. Other trimer structures also contain at least one of the dimer motif. All found structures are planar except A-3(1) and A-3(5) which contains stacking interactions. This structures were further relaxed at SCS-MP2 levels employing aVDZ basis set and no significant change has been observed except A-3(5) which transform itself to the already located,

A-3(1). Interaction energies of the SCS-MP2 corrected geometries were calculated at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 levels with aVDZ basis set and this was listed in Table 3.11. To sum it up, we can say that our model has predicted A-3(1) as the lowest energy in respect to A-3(2) by 0.38 kJ/mol. Ab-initio interaction energy computations also showed us that A-3(2) is slightly lower in energy than the predicted geometry, A-3(1). SCS-MP2 optimized structures in the interaction energy computations also changed our ordering and it agrees with the previously reported ab-initio calculations. As we stated above, we neglected the many-body effects for the calculation of the total interaction energy. But we computed the individual two- and three-body interactions of our trimers shown in Figure 3.14. Neglecting the  $\Delta^3 E_{int,ABC}^{CP}$  term has shown no significant effect and  $\Delta^3 E_{int,ABC}^{CP}$  is only negative if three adenine monomers are close to each other. AMBER, on the other hand prefers A-3(1) for the lowest energy like our model, in contrast to ab-initio calculations which prefer A-3(2). Also, AMBER still overestimates H-bonded geometries like A-3(7), A-3(9) and A-3(11) and corrects its predictions when SCS-MP2/aVDZ relaxed geometries taken into an account, just like we observed in adenine dimers.



**Figure 3.14** : Adenine trimer structures found by the SA approach.

**Table 3.11** : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine trimers shown in Figure 3.14.

Trimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
A-3(1) <sup>a</sup>	-128.46	-136.15	-106.94	-120.26	-124.64	-112.07
A-3(2) <sup>a</sup>	-143.25	-131.07	-114.99	-126.72	-120.48	-111.69
A-3(3) <sup>a</sup>	-108.50	-117.27	-90.60	-102.13	-112.44	-109.75
A-3(4) <sup>a</sup>	-129.11	-115.75	-100.75	-111.04	-110.12	-100.20
A-3(5) <sup>a</sup>	-131.57	-119.65	-104.61	-115.24	-107.60	-99.15
A-3(6) <sup>a</sup>	-126.20	-114.11	-99.19	-109.44	-104.35	-97.13
A-3(7) <sup>a</sup>	-125.98	-116.44	-101.1	-113.56	-84.78	-96.69
A-3(8) <sup>a</sup>	-120.04	-109.16	-93.85	-103.90	-102.0	-95.70
A-3(9) <sup>a</sup>	-112.25	-101.24	-86.98	-97.91	-74.47	-86.66
A-3(10) <sup>a</sup>	-108.92	-98.03	-83.78	-94.21	-83.34	-86.66
A-3(11) <sup>a</sup>	-106.49	-97.36	-82.93	-93.82	-65.00	-82.53
A-3(1) <sup>b</sup>	-150.43	-157.83	-121.39	-138.87	-124.25	-115.60
A-3(2) <sup>b</sup>	-166.36	-153.58	-136.16	-149.67	-127.14	-118.35
A-3(3) <sup>b</sup>	-150.43	-157.83	-121.39	-138.87	-124.24	-115.60
A-3(4) <sup>b</sup>	-153.26	-139.57	-123.50	-135.30	-117.39	-107.35
A-3(5) <sup>b</sup>	-154.53	-141.46	-124.87	-137.35	-114.14	-104.85
A-3(6) <sup>b</sup>	-154.06	-140.82	-124.34	-136.67	-113.55	-105.86
A-3(7) <sup>b</sup>	-143.56	-133.77	-117.60	-131.27	-103.33	-100.22
A-3(8) <sup>b</sup>	-142.44	-131.09	-114.76	-126.14	-110.60	-103.1
A-3(9) <sup>b</sup>	-129.52	-118.91	-104.31	-115.81	-93.91	-91.93
A-3(10) <sup>b</sup>	-124.78	-114.35	-99.64	-110.60	-93.94	-89.98
A-3(11) <sup>b</sup>	-127.30	-117.92	-102.92	-114.89	-88.82	-87.85

<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.

**Table 3.12** : Total ( $\Delta E_{int}^{CP}$ ), two-body ( $\Delta^2 E_{int,AB}^{CP}$ ,  $\Delta^2 E_{int,AC}^{CP}$  and  $\Delta^2 E_{int,BC}^{CP}$ ) and three-body ( $\Delta^3 E_{int,ABC}^{CP}$ ) interaction energies (calculated at B3LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 using aVDZ) of adenine trimers shown in Figure 3.14.

Method	Interaction	A-3(1) <sup>a</sup>	A-3(2) <sup>a</sup>	A-3(3) <sup>a</sup>	A-3(4) <sup>a</sup>	A-3(5) <sup>a</sup>
B3LYP-D	$\Delta E_{int}^{CP}$	-128.45	-143.25	-108.50	-129.11	-131.57
	$\Delta^2 E_{int,AB}^{CP}$	-20.15	-64.93	-42.25	-65.01	-49.02
	$\Delta^2 E_{int,AC}^{CP}$	-76.39	-80.27	-49.95	0.84	-80.29
	$\Delta^2 E_{int,BC}^{CP}$	-32.83	1.90	-18.76	-64.92	-0.91
	$\Delta^3 E_{int,ABC}^{CP}$	0.92	0.05	2.45	-0.018	1.34
MP2	$\Delta E_{int}^{CP}$	-136.15	-131.07	-117.27	-115.75	-119.95
	$\Delta^2 E_{int,AB}^{CP}$	-28.40	-58.17	-40.96	-58.27	-42.63
	$\Delta^2 E_{int,AC}^{CP}$	-71.64	-74.78	-48.07	0.82	-74.76
	$\Delta^2 E_{int,BC}^{CP}$	-37.57	2.02	-30.12	-58.14	-0.94
	$\Delta^3 E_{int,ABC}^{CP}$	1.47	-0.14	1.89	-0.16	-1.32
SCS-MP2	$\Delta E_{int}^{CP}$	-106.94	-114.99	-90.60	-100.75	-104.60
	$\Delta^2 E_{int,AB}^{CP}$	-17.84	-50.71	-35.07	-50.80	-36.23
	$\Delta^2 E_{int,AC}^{CP}$	-62.76	-66.20	-41.43	0.85	-66.12
	$\Delta^2 E_{int,BC}^{CP}$	-27.60	2.04	-15.73	-50.66	-0.91
	$\Delta^3 E_{int,ABC}^{CP}$	1.25	-0.12	1.64	-0.14	-1.35
SCS-MI-MP2	$\Delta E_{int}^{CP}$	-120.26	-126.72	-102.13	-111.04	-115.24
	$\Delta^2 E_{int,AB}^{CP}$	-20.51	-55.85	-38.76	-55.95	-40.12
	$\Delta^2 E_{int,AC}^{CP}$	-69.49	-72.94	-45.89	0.84	-72.91
	$\Delta^2 E_{int,BC}^{CP}$	-31.68	2.13	-0.90	-55.84	-73.02
	$\Delta^3 E_{int,ABC}^{CP}$	1.43	-0.06	1.86	-0.11	-1.31
		A-3(1) <sup>b</sup>	A-3(2) <sup>b</sup>	A-3(3) <sup>b</sup>	A-3(4) <sup>b</sup>	A-3(5) <sup>b</sup>
B3LYP-D	$\Delta E_{int}^{CP}$	-150.43	-166.36	-150.43	-153.26	-154.53
	$\Delta^2 E_{int,AB}^{CP}$	-25.97	-77.11	-36.46	-77.00	-59.74
	$\Delta^2 E_{int,AC}^{CP}$	-89.69	-91.49	-89.69	1.01	-91.91
	$\Delta^2 E_{int,BC}^{CP}$	-36.45	2.06	-25.97	-77.21	-1.03
	$\Delta^3 E_{int,ABC}^{CP}$	1.69	0.18	1.69	-0.05	-1.86
MP2	$\Delta E_{int}^{CP}$	-157.83	-153.58	-157.83	-139.57	-141.46
	$\Delta^2 E_{int,AB}^{CP}$	-34.23	-70.18	-42.45	-70.09	-52.84
	$\Delta^2 E_{int,AC}^{CP}$	-83.56	-85.53	-83.55	0.99	-85.70
	$\Delta^2 E_{int,BC}^{CP}$	-42.45	2.19	-34.24	-70.23	-1.07
	$\Delta^3 E_{int,ABC}^{CP}$	2.40	-0.068	2.40	-0.24	-1.85
SCS-MP2	$\Delta E_{int}^{CP}$	-121.39	-136.16	-121.39	-123.50	-124.87
	$\Delta^2 E_{int,AB}^{CP}$	-20.61	-62.21	-29.37	-62.10	-45.90
	$\Delta^2 E_{int,AC}^{CP}$	-73.46	-76.11	-73.45	1.03	-76.04
	$\Delta^2 E_{int,BC}^{CP}$	-29.37	2.21	-20.60	-62.21	-1.04
	$\Delta^3 E_{int,ABC}^{CP}$	2.04	-0.05	2.035	-0.21	-1.88
SCS-MI-MP2	$\Delta E_{int}^{CP}$	-138.87	-149.67	-138.87	-135.30	-137.35
	$\Delta^2 E_{int,AB}^{CP}$	-24.66	-68.07	-34.69	-68.00	-50.37
	$\Delta^2 E_{int,AC}^{CP}$	-81.82	-83.95	-81.81	1.02	-84.13
	$\Delta^2 E_{int,BC}^{CP}$	-34.69	2.31	-24.66	-68.14	-1.03
	$\Delta^3 E_{int,ABC}^{CP}$	2.30	0.04	2.30	-0.18	-1.83

<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.



**Table 3.13** : Total ( $\Delta E_{int}^{CP}$ ), two-body ( $\Delta^2 E_{int,AB}^{CP}$ ,  $\Delta^2 E_{int,AC}^{CP}$  and  $\Delta^2 E_{int,BC}^{CP}$ ) and three-body ( $\Delta^3 E_{int,ABC}^{CP}$ ) interaction energies (calculated at B3LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 using aVDZ) of adenine trimers shown in Figure 3.14.

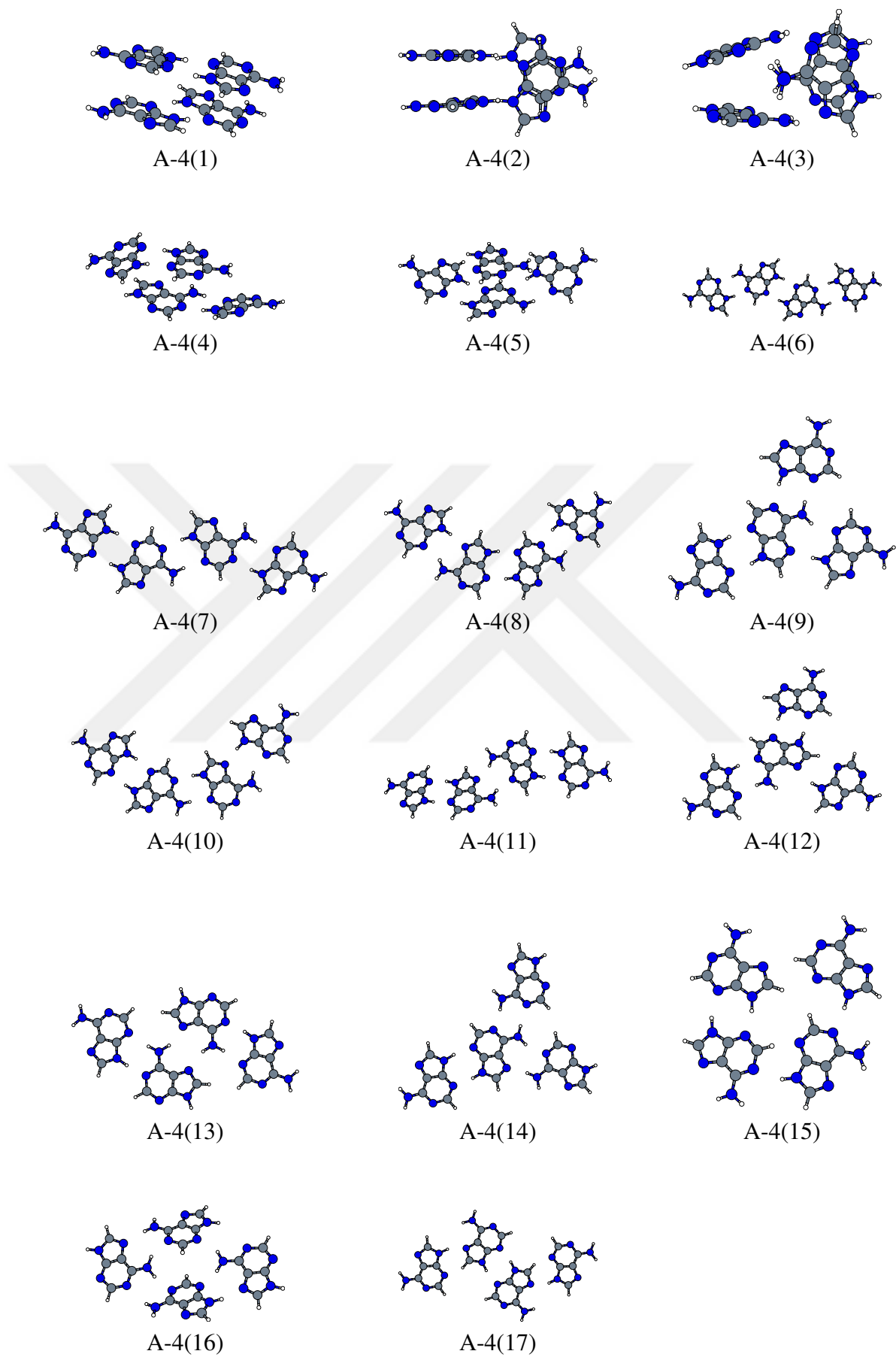
Method	Interaction	A-3(6) <sup>a</sup>	A-3(7) <sup>a</sup>	A-3(8) <sup>a</sup>	A-3(8) <sup>a</sup>	A-3(10) <sup>a</sup>	A-3(11) <sup>a</sup>
B3LYP-D	$\Delta E_{int}^{CP}$	-126.20	-125.98	-120.042	-112.25	-108.92	-106.49
	$\Delta^2 E_{int,AB}^{CP}$	-58.08	-46.76	-1.10	-41.40	-65.28	-46.33
	$\Delta^2 E_{int,AC}^{CP}$	-1.33	1.53	-57.70	-73.91	-5.64	-1.13
	$\Delta^2 E_{int,BC}^{CP}$	-65.00	-80.34	-65.03	-46.69	-39.12	-58.07
	$\Delta^3 E_{int,ABC}^{CP}$	-1.79	-0.41	3.79	0.45	1.13	-0.96
MP2	$\Delta E_{int}^{CP}$	-114.11	-116.44	-109.16	-101.24	-98.03	-97.36
	$\Delta^2 E_{int,AB}^{CP}$	-52.73	-42.70	1.53	-1.32	-58.59	-42.14
	$\Delta^2 E_{int,AC}^{CP}$	-1.44	1.25	-52.90	-57.92	-6.35	-1.78
	$\Delta^2 E_{int,BC}^{CP}$	-58.17	-74.86	-58.43	-42.57	-34.62	-52.66
	$\Delta^3 E_{int,ABC}^{CP}$	-1.77	-0.13	0.64	0.58	1.53	-0.78
SCS-MP2	$\Delta E_{int}^{CP}$	-99.19	-101.10	-93.85	-86.98	-83.78	-82.94
	$\Delta^2 E_{int,AB}^{CP}$	-45.38	-36.06	2.26	-1.29	-51.01	-35.45
	$\Delta^2 E_{int,AC}^{CP}$	-1.39	1.44	-45.61	-50.40	-5.84	-1.45
	$\Delta^2 E_{int,BC}^{CP}$	-50.63	-66.26	-51.05	-35.84	-28.37	-45.24
	$\Delta^3 E_{int,ABC}^{CP}$	-1.78	-0.22	0.54	0.54	1.45	-0.79
SCS-MI-MP2	$\Delta E_{int}^{CP}$	-109.44	-113.56	-103.90	-97.91	-94.21	-93.81
	$\Delta^2 E_{int,AB}^{CP}$	-50.41	-41.70	2.09	-1.28	-56.24	-41.14
	$\Delta^2 E_{int,AC}^{CP}$	-1.39	1.36	-50.58	-55.58	-5.91	-1.51
	$\Delta^2 E_{int,BC}^{CP}$	-55.84	-73.02	-56.13	-41.57	-33.55	-50.35
	$\Delta^3 E_{int,ABC}^{CP}$	-1.80	-0.19	0.72	0.53	1.49	-0.81
		A-3(7) <sup>b</sup>	A-3(8) <sup>b</sup>	A-3(9) <sup>b</sup>	A-3(20) <sup>b</sup>	A-3(11) <sup>b</sup>	A-3(12) <sup>b</sup>
B3LYP-D	$\Delta E_{int}^{CP}$	-154.06	-143.56	-142.44	-129.52	-124.79	-127.30
	$\Delta^2 E_{int,AB}^{CP}$	-72.30	-52.86	2.09	-1.25	-75.92	-53.05
	$\Delta^2 E_{int,AC}^{CP}$	-1.45	1.28	-70.24	-76.5	-6.22	-1.14
	$\Delta^2 E_{int,BC}^{CP}$	-77.72	-91.22	-75.69	-52.42	-43.75	-71.41
	$\Delta^3 E_{int,ABC}^{CP}$	-2.60	-0.75	1.39	0.74	1.1	-1.69
MP2	$\Delta E_{int}^{CP}$	-140.82	-133.77	-131.09	-118.91	-114.35	-117.92
	$\Delta^2 E_{int,AB}^{CP}$	-66.28	-49.29	1.28	-1.33	-69.16	-49.37
	$\Delta^2 E_{int,AC}^{CP}$	-1.57	1.09	-64.58	-69.57	-7.01	-1.63
	$\Delta^2 E_{int,BC}^{CP}$	-70.41	-85.14	-68.87	-48.90	-39.66	-65.46
	$\Delta^3 E_{int,ABC}^{CP}$	-2.56	-0.42	1.09	0.88	1.48	-1.46
SCS-MP2	$\Delta E_{int}^{CP}$	-124.34	-117.60	-114.76	-104.31	-99.64	-102.92
	$\Delta^2 E_{int,AB}^{CP}$	-58.12	-42.73	1.95	-1.29	-61.06	-42.72
	$\Delta^2 E_{int,AC}^{CP}$	-1.52	1.23	-56.71	-61.54	-6.42	-1.40
	$\Delta^2 E_{int,BC}^{CP}$	-62.13	-75.57	-60.99	-42.31	-33.55	-57.32
	$\Delta^3 E_{int,ABC}^{CP}$	-2.58	-0.53	1.00	0.84	1.38	-1.48
SCS-MI-MP2	$\Delta E_{int}^{CP}$	-136.67	-131.27	-126.14	-115.81	-110.59	-114.89
	$\Delta^2 E_{int,AB}^{CP}$	-64.23	-48.39	1.80	-1.28	-66.94	-48.47
	$\Delta^2 E_{int,AC}^{CP}$	-1.52	1.16	-62.42	-67.43	-6.52	-1.47
	$\Delta^2 E_{int,BC}^{CP}$	-68.32	-83.54	-66.72	-47.93	-38.58	-63.42
	$\Delta^3 E_{int,ABC}^{CP}$	-2.60	-0.51	1.20	0.84	1.46	-1.53

<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.

### 3.3.1.3 Tetramer

We predicted 17 different adenine tetrads, including stacked, filament(chain) and planar geometries shown in Figure 3.15. All of the resulting tetramers were further relaxed at SCS-MP2/aVDZ level and the interaction energies were listed in Table 3.14. After the relaxation, only A-4(16) changes its angular orientation while others preserved their geometries. The model potential preferred the H-bonded A-4(1) over the filament and planar structures. A-4(1) differs from A-4(2), A-4(3) and A-4(4) by -6.95, -23.27 and -25.57 kJ/mol respectively. B3LYP-D agrees with our model and favours A-4(1), however, for MP2, SCS-MP2 and SCS-MP2, A-4(1) takes the second place, after A-4(2). AMBER still has problems with H-bonded structures, especially for A-4(7), A-4(8) and A-4(13), which are lower in energy in contrast to our model and all ab-initio calculations. Employment of SCS-MP2 optimized structures altered the energy ordering both at the model and other ab-initio levels. A-4(16) has changed its angular orientation after the employment of SCS-MP2 relaxation and becomes a much more stable structure. Our model still prefers A-4(1) as the lowest energy but all ab-initio calculations show otherwise. A-4(2) was preferred as the lowest energy structure in all ab-initio calculations and it was followed by A-4(1). MP2 and our model favoured A-4(5) over A-4(6) which becomes reversed in the terms of energy in B3LYP-D, SCS-MP2 and SCS-MI-MP2. In contrast to all methods, AMBER prefers the stacked A-4(3), A-4(7), A-4(8) and A-4(16) as the most stable isomers while all ab-initio calculations and our model disagree.



**Figure 3.15** : Adenine tetramer structures found by the SA approach.

**Table 3.14** : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine tetramers shown in Figure 3.15.

Tetramer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
A-4(1) <sup>a</sup>	-225.52	-238.44	-188.26	-212.29	-245.02	-205.83
A-4(2) <sup>a</sup>	-224.30	-240.18	-190.88	-214.59	-243.25	-198.88
A-4(3) <sup>a</sup>	-200.02	-201.38	-154.58	-173.10	-230.12	-182.56
A-4(4) <sup>a</sup>	-211.21	-214.94	-172.96	-193.94	-226.94	-180.26
A-4(5) <sup>a</sup>	-192.53	-193.26	-154.64	-174.08	-198.10	-171.90
A-4(6) <sup>a</sup>	-205.95	-187.09	-163.55	-180.20	-172.12	-163.27
A-4(7) <sup>a</sup>	-207.26	-188.46	-164.87	-181.71	-230.59	-162.26
A-4(8) <sup>a</sup>	-203.66	-186.09	-164.87	-179.33	-216.63	-159.81
A-4(9) <sup>a</sup>	-201.27	-184.97	-160.95	-177.65	-166.47	-159.59
A-4(10) <sup>a</sup>	-203.92	-186.40	-162.87	-179.66	-168.21	-158.62
A-4(11) <sup>a</sup>	-206.74	-191.48	-167.43	-185.10	-171.78	-157.13
A-4(12) <sup>a</sup>	-187.37	-171.13	-148.10	-165.66	-134.61	-149.81
A-4(13) <sup>a</sup>	-179.63	-163.36	140.48	-154.46	-205.38	-141.58
A-4(14) <sup>a</sup>	-183.28	-164.29	-142.24	-156.64	-151.31	-139.27
A-4(15) <sup>a</sup>	-167.30	-151.10	-126.50	-145.23	-100.61	-135.13
A-4(16) <sup>a</sup>	-156.17	-140.65	-115.44	-129.78	-179.71	-134.70
A-4(17) <sup>a</sup>	-171.80	-158.26	-136.14	-154.31	-120.84	-133.96
A-4(1) <sup>b</sup>	-255.73	-272.77	-209.09	-239.21	-232.26	-210.79
A-4(2) <sup>b</sup>	-261.67	-278.32	-215.56	-246.63	-243.30	-207.14
A-4(3) <sup>b</sup>	-229.66	-233.10	-179.82	-201.37	-250.44	-202.06
A-4(4) <sup>b</sup>	-249.72	-254.45	-200.48	-227.56	-231.47	-187.37
A-4(5) <sup>b</sup>	-225.32	-228.80	-179.69	-204.59	-205.16	-176.24
A-4(6) <sup>b</sup>	-240.90	-221.41	-196.11	-215.16	-182.57	-137.61
A-4(7) <sup>b</sup>	-242.43	-222.92	-197.46	-216.83	-244.56	-172.47
A-4(8) <sup>b</sup>	-242.22	-223.23	-197.51	-217.20	-234.55	-172.47
A-4(9) <sup>b</sup>	-236.62	-219.16	-192.97	-212.78	-177.43	-170.15
A-4(10) <sup>b</sup>	-242.79	-223.77	-197.91	-217.75	-180.49	-170.52
A-4(11) <sup>b</sup>	-242.50	-224.83	-198.37	-219.19	-182.09	-167.15
A-4(12) <sup>b</sup>	-217.92	-201.27	-177.16	-196.40	-160.11	-167.75
A-4(13) <sup>a</sup>	-206.91	-189.54	-164.76	-180.80	-216.80	-156.56
A-4(14) <sup>a</sup>	-213.41	-193.54	-169.59	-185.97	-160.17	-150.54
A-4(15) <sup>a</sup>	-196.19	-181.39	-155.68	-175.40	-153.34	-143.79
A-4(16) <sup>a</sup>	-228.54	-140.65	-178.20	-129.78	-239.67	-174.89
A-4(17) <sup>a</sup>	-195.56	-158.26	-159.23	-154.31	-172.12	-138.18

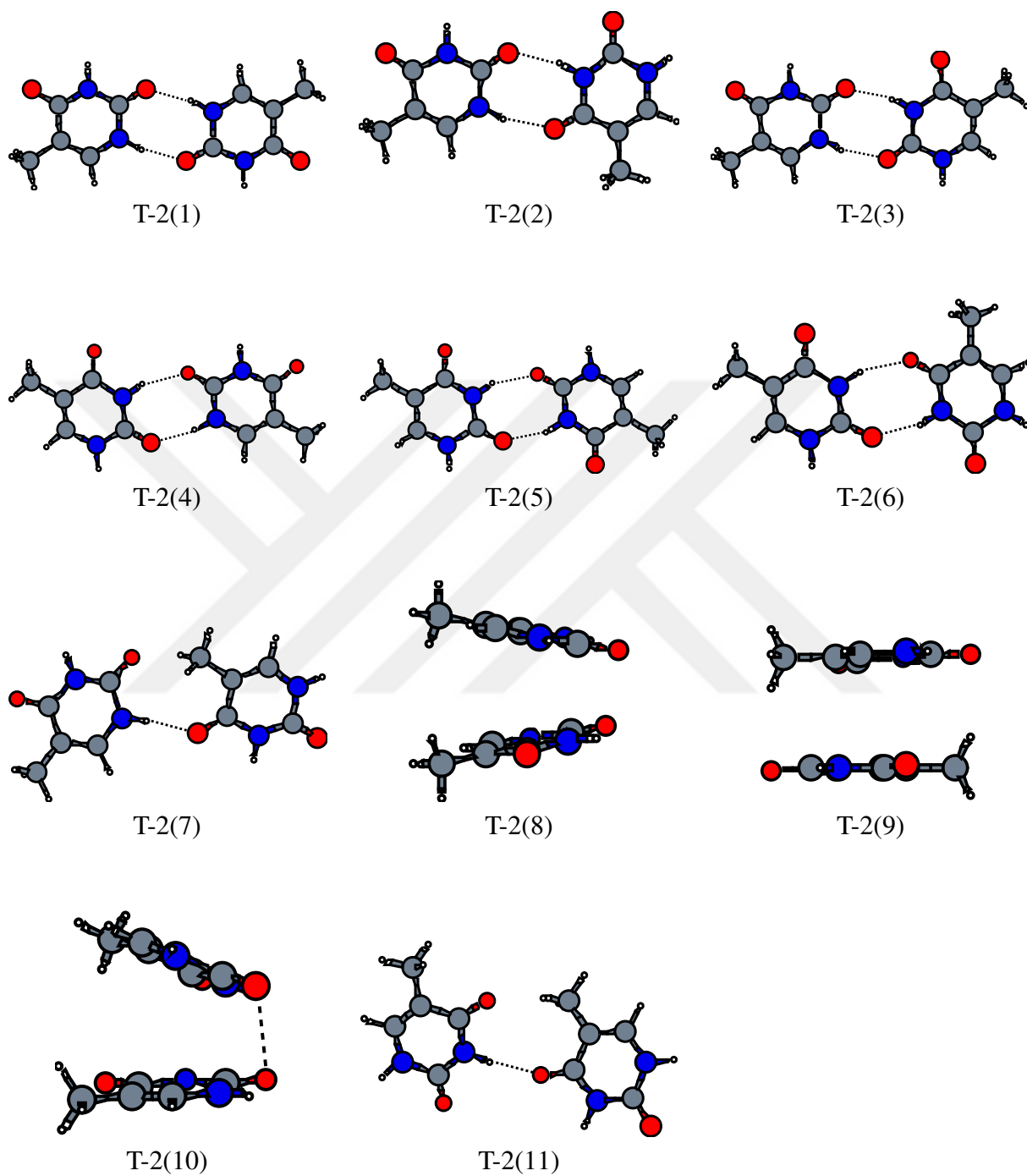
<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.

### 3.3.2 Thymine clusters

#### 3.3.2.1 Dimer

Our model detected 11 different thymine dimers shown in Figure 3.16. H-Bonded T-2(1), T-2(2), T-2(3) and stacked T-2(8), T-2(9) and T-2(10) was reported in the literature by Bravaya et al. [70] and Sun et al. [85] used T-2(1) and T-2(10) to calculate ionization potentials and electron affinities. In addition to that, T-2(1) and T-2(5) can be seen in the STM images of the supramolecular nanopatterns, assembled by adenine-thymine quartets reported by Mamdouh et al. [86]. Our force field even manages to generate new stacked and H-bonded dimers such as T-2(4) and T-2(6). Resulting thymine dimers were further relaxed with PBE, SCS-MP2 and CP-SCS-MP2 using aug-cc-pVDZ basis set and Table 3.15 shows the interaction energies obtained from B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ basis set. B3LYP-D still overestimates the interaction energies while MP2 and SCS-MI-MP2 were in quite good agreement for H-bonded structures. LPBE0AC is in agreement with SCS-MI-MP2 while generating lower interaction energies than PBE0AC. Our model predicted T-2(1) as the lowest energy orientation, lower than 10.56 and 10.89 kJ/mol from the closest orientations T-2(2) and T-2(3), respectively. The lowest energy order also remains same for all ab-initio calculations and this can be seen in Table 3.15. AMBER fails to give an accurate result when single H-bonded geometries such as T-2(7) and T-2(11) taken into an account but manages to correct this error when the geometries optimized with CP-SCS-MP2. Also, T-2(1) was still found as the lowest energy orientation when CP-SCS-MP2 optimized geometries taken into an account. As a result, we can say that our model successfully located the structures in the literature [70, 85, 86] and their corresponding energy ordering generally stays same with CP-SCS-MP2 relaxed geometries. This gives us the opportunity to extend our model for the larger thymine homo-oligomers.



**Figure 3.16** : Thymine dimer structures found by the SA approach.

**Table 3.15** : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of thymine dimers shown in Figure 3.16.

Dimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	PBE0AC	LPBE0AC	AMBER	Model
T-2(1) <sup>a</sup>	-72.25	-64.84	-58.68	-67.91	-59.93	-64.41	-63.99	-62.28
T-2(2) <sup>a</sup>	-61.21	-54.54	-48.50	-56.94	-50.12	-53.82	-54.73	-51.72
T-2(3) <sup>a</sup>	-58.82	-53.21	-47.59	-55.01	-48.03	-51.77	-54.33	-51.39
T-2(4) <sup>a</sup>	-57.12	-51.64	-46.00	-53.37	-46.60	-50.30	-53.85	-51.04
T-2(5) <sup>a</sup>	-47.79	-43.47	-38.23	-44.45	-38.48	-41.56	-47.91	-43.85
T-2(6) <sup>a</sup>	-48.84	-43.93	-38.42	-45.20	-39.35	-42.39	-47.62	-43.38
T-2(7) <sup>a</sup>	-42.39	-36.23	-29.89	-37.82	-33.75	-37.47	-11.94	-42.80
T-2(8) <sup>a</sup>	-36.93	-37.62	-25.76	-31.69	-27.35	-30.79	-34.11	-41.74
T-2(9) <sup>a</sup>	-34.68	-34.68	-22.61	-28.31	-25.41	-29.57	-37.17	-36.49
T-2(10) <sup>a</sup>	-34.21	-34.55	-23.81	-29.43	-25.03	-28.26	-30.42	-35.81
T-2(11) <sup>a</sup>	-29.33	-25.26	-19.33	-25.74	-22.48	-25.47	-0.99	-41.49
T-2(1) <sup>b</sup>	-91.51	-80.68	-72.84	-84.75	-73.07	-79.24	-65.89	-61.73
T-2(2) <sup>b</sup>	-75.54	-65.89	-58.36	-69.04	-59.21	-64.17	-55.66	-50.50
T-2(3) <sup>b</sup>	-71.89	-63.88	-57.01	-66.23	-56.48	-61.42	-55.37	-50.58
T-2(4) <sup>b</sup>	-71.88	-63.87	-57.01	-66.22	-56.48	-61.41	-55.38	-50.59
T-2(5) <sup>b</sup>	-57.17	-51.33	-45.26	-52.55	-44.55	-48.38	-48.76	-43.36
T-2(6) <sup>b</sup>	-59.37	-52.48	-45.93	-54.11	-45.96	-49.84	-48.38	-42.49
T-2(7) <sup>b</sup>	-51.43	-45.01	-39.08	-46.88	-42.28	-45.84	-36.66	-42.01
T-2(8) <sup>b</sup>	-47.63	-47.53	-32.99	-40.03	-34.55	-38.66	-41.36	-41.14
T-2(9) <sup>b</sup>	-44.30	-45.06	-31.50	-38.01	-32.56	-36.36	-40.33	-35.72
T-2(10) <sup>b</sup>	-44.60	-45.54	-31.03	-37.54	-31.17	-35.25	-34.89	-34.56
T-2(11) <sup>b</sup>	-37.22	-32.98	-27.04	-32.92	-28.98	-31.69	-31.86	-29.53

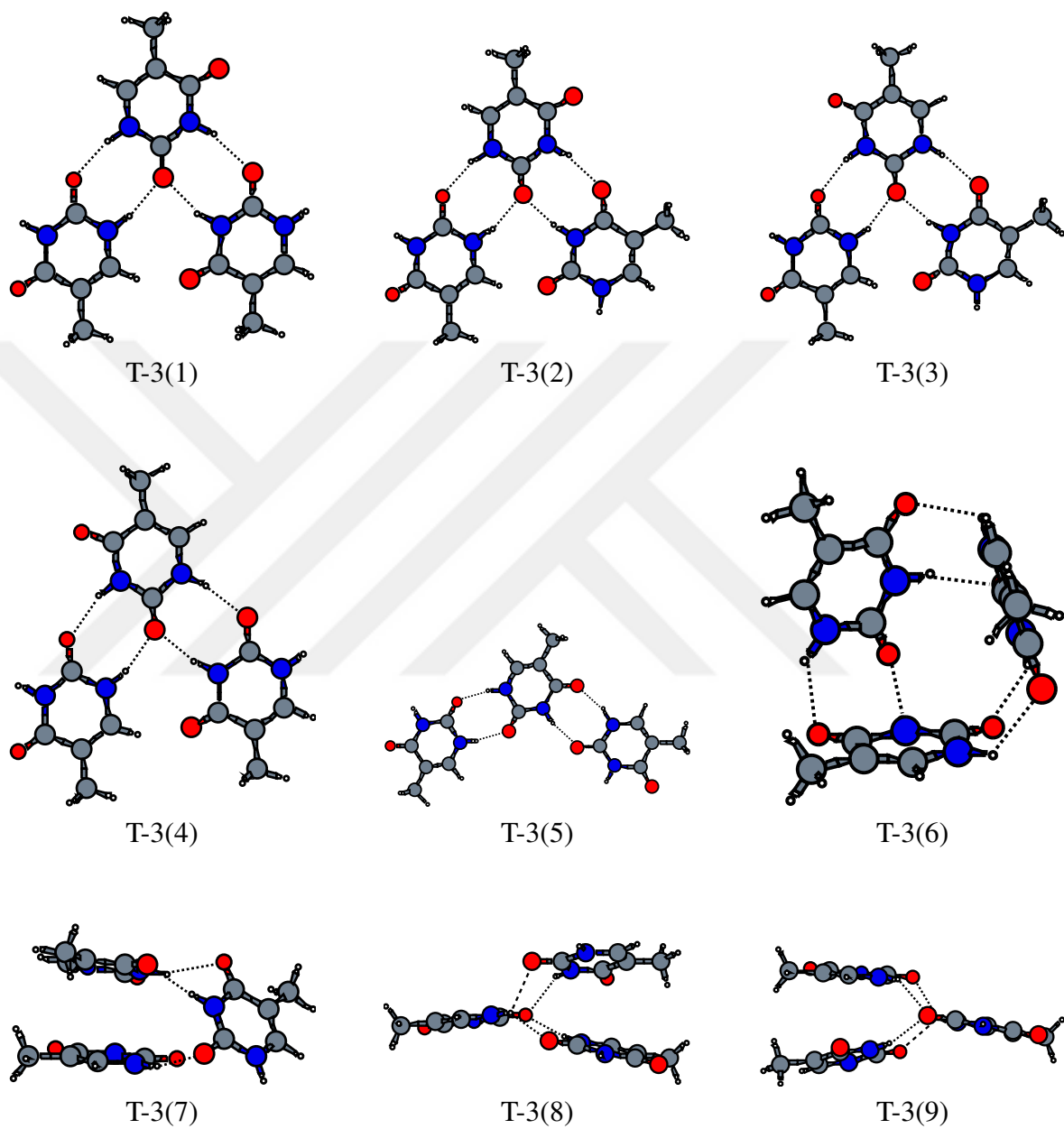
<sup>a</sup> Using model geometries.

<sup>b</sup> Using CP-SCS-MP2/aVDZ geometries.

### 3.3.2.2 Trimer

For trimer and beyond, we neglected the many-body effects and implemented the same technique that we used on adenine oligomers. With using this approach, we manage to detect 9 different planar, stacked and ribbon thymine trimers as shown in Figure 3.17. As it can be seen in Table 3.16, our model has predicted T-3(1) as the lowest energy in respect to T-3(2) by -0.36 kJ/mol and T-3(3) by -2.73 kJ/mol. In addition, our model has been able to locate structures which have both stacked and H-bonded interactions, T-3(6) and T-3(9), respectively. T-3(5) orientated as ribbon, which is a unique feature with respect to other isomers, and was reported previously by Chiriki et al. [87] as having the lowest energy. Compared to T-3(5), our model can be able to locate T-3(1), T-3(2), T-3(3) and T-3(4) which all have lower energies than T-3(5) by -7.73 kJ/mol, -7.37 kJ/mol, 5 kJ/mol and 4.41 kJ/mol, respectively. AMBER in the other hand, fails to come in an agreement with all ab-initio calculations and prefers T-3(5) as the lowest energy. Generally, AMBER overestimates the interaction energies of T-3(5), T-3(6), T-3(7) and T-3(8) while all ab-initio calculations and our model shows otherwise. As we observed in adenine, AMBER still has problems with H-bonding interactions when we considered thymine clusters. All structures reported in Figure 3.17 were further relaxed at SCS-MP2 level, employing aVDZ basis set and no significant change has been observed. Interaction energies of the SCS-MP2 corrected geometries were calculated at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 levels and this was listed in Table 3.11. In Table 3.11, we can see that all ab-initio calculations prefer T-3(5) over T-3(1) by small margins but our model still prefers T-3(1) over T-3(5) by -6.79 kJ/mol. Also the lowest energy isomer has changed for both our model and AMBER while all ab-initio calculations disagree with this prediction. For ab-initio calculations, now T-3(5) has the lowest energy but our model and AMBER prefers T-3(6), which has both stacked and H-bonded interactions.





**Figure 3.17** : Thymine trimer structures found by the SA approach.

**Table 3.16** : Interaction energy calculations (at AMBER) using aVDZ) of thymine trimers shown in Figure 3.17.

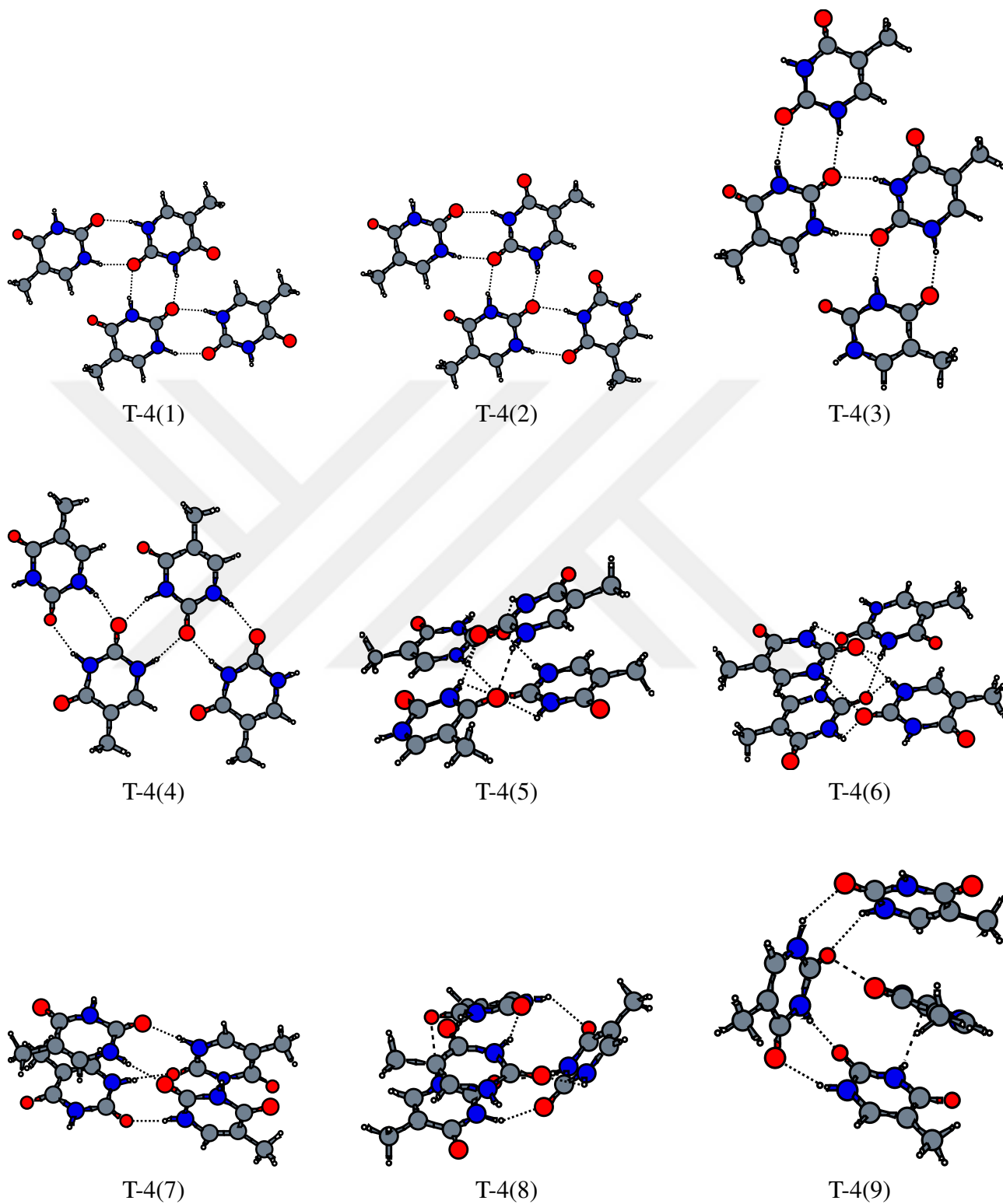
Trimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
T-3(1) <sup>a</sup>	-131.56	-116.77	-102.08	-121.88	-90.63	-124.83
T-3(2) <sup>a</sup>	-131.01	-116.26	-101.66	-121.27	-89.43	-124.47
T-3(3) <sup>a</sup>	-127.01	-112.63	-98.08	-117.31	-89.24	-122.10
T-3(4) <sup>a</sup>	-124.70	-110.91	-96.40	-115.36	-89.45	-121.51
T-3(5) <sup>a</sup>	-135.79	-121.40	-109.03	-127.12	-121.57	-117.10
T-3(6) <sup>a</sup>	-116.31	-104.88	-82.94	-102.74	-119.09	-115.84
T-3(7) <sup>a</sup>	-111.10	-103.96	-85.06	-101.27	-110.63	-107.88
T-3(8) <sup>a</sup>	-107.97	-101.61	-82.06	-97.97	-110.51	-107.49
T-3(9) <sup>a</sup>	-114.86	-108.44	-86.13	-102.70	-77.73	-104.84
T-3(1) <sup>b</sup>	-165.62	-147.08	-130.96	-153.18	-123.02	-120.58
T-3(2) <sup>b</sup>	-165.90	-142.20	-130.97	-153.34	-122.47	-119.90
T-3(3) <sup>b</sup>	-160.65	-142.16	-126.01	-147.91	-120.63	-116.87
T-3(4) <sup>b</sup>	-158.76	-141.27	-125.42	-146.60	-120.61	-116.92
T-3(5) <sup>b</sup>	-169.17	-147.88	-131.98	-155.39	-123.54	-113.79
T-3(6) <sup>b</sup>	-149.71	-133.61	-109.36	-133.45	-132.09	-125.09
T-3(7) <sup>b</sup>	-149.48	-130.38	-104.92	-124.71	-110.32	-98.11
T-3(8) <sup>b</sup>	-141.68	-131.05	-105.33	-125.90	-116.34	-106.39
T-3(9) <sup>b</sup>	-146.72	-134.84	-105.07	-128.17	-121.88	-107.42

<sup>a</sup> Using model geometries.

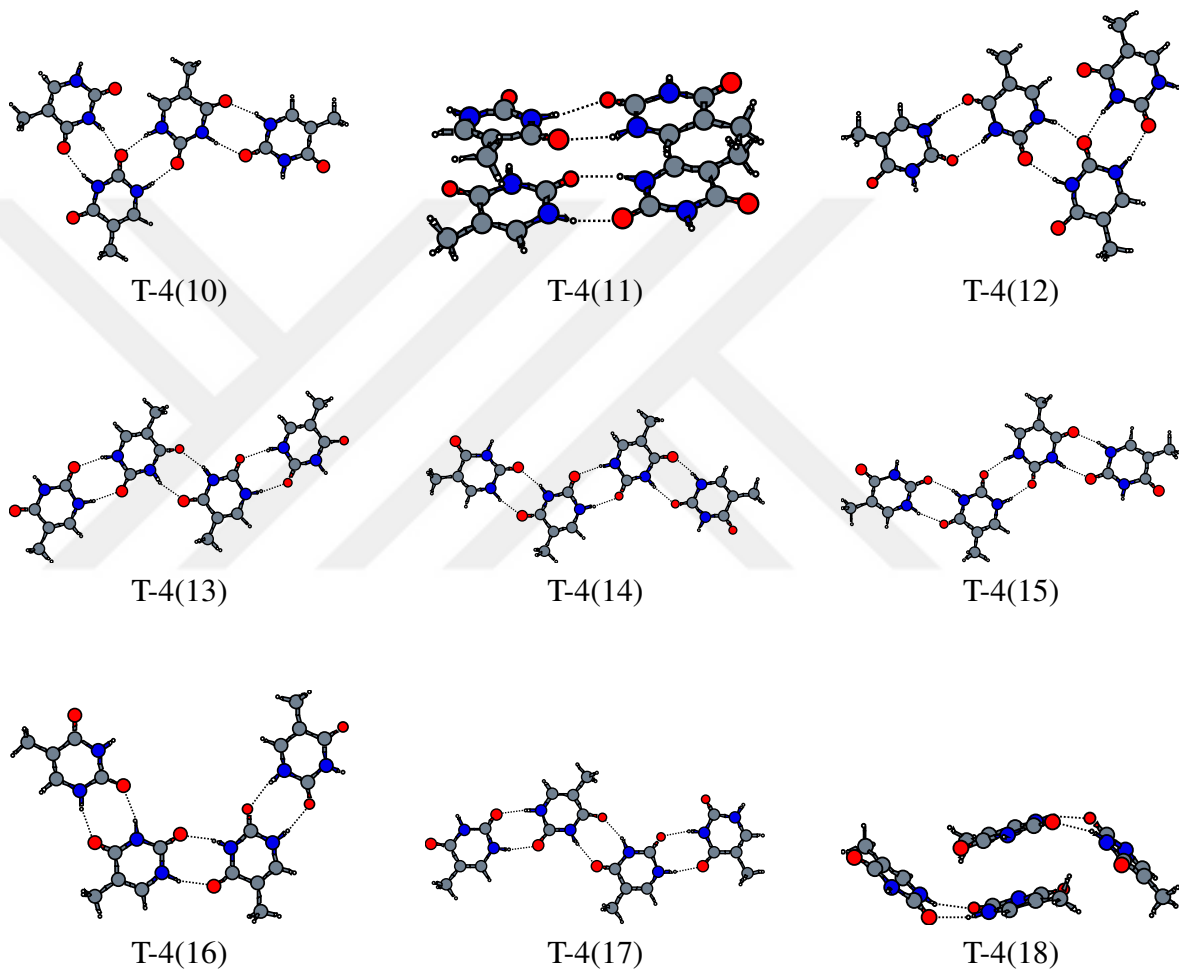
<sup>b</sup> Using SCS-MP2/aVDZ geometries.

### 3.3.2.3 Tetramer

When we apply the same approach for searching thymine tetramers, our method can be able to detect 18 unique structures including thymine tetrads, ribbons and stacked geometries as shown in Figure 3.18 and Figure 3.19 . Amongst them, T-4(1), T-4(2), T-4(3) and T-4(9) includes T-3(1), T-3(2), T-3(4) and with one additional thymine dimer. Amongst them, T-4(1) was reported in Chiriki et al. [87] as having the lowest energy. Also, our model has been able to locate additional orientations which have both stacked and H-bonded interactions, T-4(4), T-4(8), T-4(10) and T-4(17), ribbons, T-4(12) to T-4(16) and stacked, T-4(4) and T-4(5). All structures shown in Figure 3.18 were also further relaxed with the same level and basis set which had been used for adenine and almost all structures preserved their orientations except T-4(5) and T-4(8). SCS-MP2 relaxed geometries again were calculated at B3LYP-D, MP, SCS-MP2, SCS-MI-MP2 levels and results can be seen in Table 3.14. In Table 3.14, we can see that both our model and all ab-initio calculations favoured T-4(1) as the lowest energy but AMBER prefers T-4(7) over T-4(1), by -16.25 kJ/mol. For our model, however, T-4(1) has the lowest energy over T-4(7) by -5.35 kJ/mol and ab-initio calculations also confirm this prediction. In addition to that, T-4(6), T-4(7) and T-4(10) overestimated by AMBER while T-4(3) T-4(4) and T-4(5) underestimated in contrast to all ab-initio calculations and our model.



**Figure 3.18** : Thymine tetramer structures found by the SA approach.



**Figure 3.19** : Thymine tetramer structures found by the SA approach.

**Table 3.17** : Interaction energy calculations (at AMBER) using aVDZ) of thymine tetramers shown in Figure 3.18. and 3.19

Tetramer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
T-4(1) <sup>a</sup>	-211.60	-186.95	-163.02	-196.12	-153.47	-205.05
T-4(2) <sup>a</sup>	-194.16	-171.43	-148.04	-178.86	-132.28	-192.48
T-4(3) <sup>a</sup>	-191.73	-169.69	-146.37	-176.81	-131.97	-191.82
T-4(4) <sup>a</sup>	-178.75	-164.04	-131.02	-159.63	-148.95	-188.68
T-4(5) <sup>a</sup>	-117.14	-162.98	-130.23	-158.18	-148.71	-187.72
T-4(6) <sup>a</sup>	-203.30	-198.54	-162.91	-190.00	-203.02	-186.33
T-4(7) <sup>a</sup>	-189.60	-175.89	-141.51	-170.55	-199.97	-185.84
T-4(8) <sup>a</sup>	-189.32	-177.12	-140.73	-169.36	-184.47	-181.56
T-4(9) <sup>a</sup>	-194.84	-173.16	-152.34	-180.87	-147.89	-179.77
T-4(10) <sup>a</sup>	-192.24	-185.61	-149.05	-175.77	-185.86	-178.74
T-4(11) <sup>a</sup>	-186.93	-166.37	-145.68	-173.25	-143.22	-173.97
T-4(12) <sup>a</sup>	-199.47	-178.42	-159.92	-186.57	-179.93	-173.17
T-4(13) <sup>a</sup>	-199.13	-177.80	-159.21	-186.16	-178.82	-172.15
T-4(14) <sup>a</sup>	-196.08	-175.16	-156.60	-183.38	-178.04	-171.23
T-4(15) <sup>a</sup>	-197.54	-176.38	-147.81	-184.59	-177.53	-170.74
T-4(16) <sup>a</sup>	-185.15	-165.45	-147.17	-172.54	-168.03	-159.88
T-4(17) <sup>a</sup>	-171.69	-160.15	-129.12	-153.15	-168.37	-159.55
T-4(18) <sup>a</sup>	-150.60	-133.01	-113.13	-137.31	-117.18	-151.18
T-4(1) <sup>b</sup>	-274.24	-243.16	-217.07	-254.38	-203.09	-197.69
T-4(2) <sup>b</sup>	-249.22	-221.15	-196.01	-230.11	-187.49	-184.07
T-4(3) <sup>b</sup>	-247.12	-220.14	-195.34	-228.60	-187.35	-184.11
T-4(4) <sup>b</sup>	-245.25	-224.26	-181.93	-217.88	-192.67	-178.26
T-4(5) <sup>b</sup>	-246.22	-225.37	-182.37	-217.43	-196.56	-177.03
T-4(6) <sup>b</sup>	-248.13	-235.97	-190.36	-225.43	-194.59	-174.48
T-4(7) <sup>b</sup>	-257.47	-236.75	-191.32	-231.92	-219.34	-192.34
T-4(8) <sup>b</sup>	-238.77	-222.72	-172.08	-209.35	-194.63	-181.81
T-4(9) <sup>b</sup>	-245.18	-215.95	-191.66	-225.73	-180.21	-173.19
T-4(10) <sup>b</sup>	-240.30	-223.94	-176.92	-211.83	-184.19	-166.75
T-4(11) <sup>b</sup>	-235.25	-207.65	-183.85	-216.27	-175.23	-167.30
T-4(12) <sup>b</sup>	-248.81	-217.66	-193.91	-228.41	-182.76	-167.97
T-4(13) <sup>b</sup>	-248.12	-216.35	-192.42	-227.47	-181.71	-166.95
T-4(14) <sup>b</sup>	-248.11	-216.35	-192.42	-227.47	-181.72	-166.96
T-4(15) <sup>b</sup>	-246.42	-214.90	-191.02	-225.79	-180.02	-165.17
T-4(16) <sup>b</sup>	-230.36	-200.93	-177.60	-210.37	-170.25	-154.30
T-4(17) <sup>b</sup>	-219.88	-203.77	-164.81	-196.56	-175.32	-155.75
T-4(18) <sup>b</sup>	-188.76	-168.57	-147.89	-173.95	-152.19	-144.38

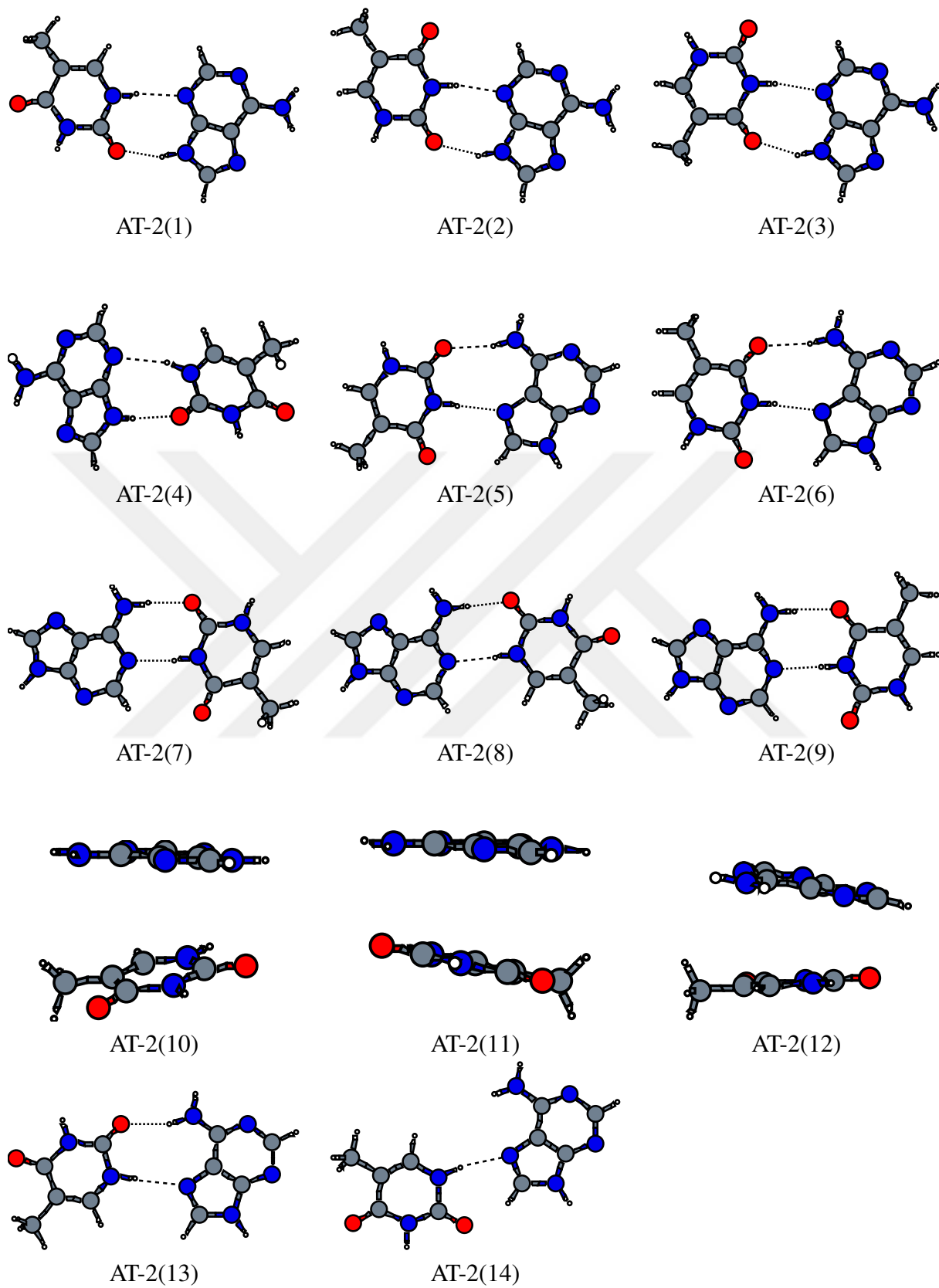
<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.

### 3.3.3 Adenine-Thymine clusters

#### 3.3.3.1 Dimer

For adenine-thymine dimers and oligomers, we implemented the same approach used in the adenine and thymine case. With this approach, our model can be able to detect 14 different adenine-thymine dimers shown in Figure 3.20, including AT-2(1) and AT-2(8), which have been reported previously [70, 88]. In addition to AT-2(1) and AT-2(8), we have been able to locate Watson-Crick base pairing AT-2(9), which was heavily inspected in the literature [89–92] because of its importance in DNA replication. Our force field also able to generate other H-bonded and stacked geometries as well, such as H-bonded AT-2(2), AT-2(5), AT-2(6), AT-2(7), AT-2(8), AT-2(13) which also inspected by Monajjemi et al. [89] and stacked AT-2(10), AT-2(11) and AT(12). Based on Monajjemi et al. [89], AT-2(1) has the lowest interaction energy, -72.21 kJ/mol, followed by AT-2(3) and AT-2(2). But Kratochil et al.'s interaction energy computations [90] show that AT-2(1) has the lowest energy, -65.19 kJ/mol, followed by AT-2(2) with -6.20 kJ/mol difference. However, our model was favored AT-2(2) over AT-2(1) by -2.63 kJ/mol, agreeing with Monajjemi et al. [89]. Resulting adenine dimer structures further relaxed with PBE, SCS-MP2 and CP-SCS-MP2 using aug-cc-pVDZ basis set to clarify the quality of the predicted orientations and structures obtained from SA were not changed upon the ab-initio relaxations. After relaxations, AT-2(2) was still favored over AT-2(1) for our model by -2.82 kJ/mol and was followed by AT-2(3) and AT-2(4). Ab-initio calculations however, favoured AT(2)-1 over AT(2)-2 with small differences. AMBER also agrees with ab-initio calculations except for AT-(14), which has been underestimated in according to all ab-initio calculations and our model.



**Figure 3.20** : Adenine-Thymine dimer structures found by the SA approach.



**Table 3.18** : Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of adenine-thymine dimers shown in Figure 3.20.

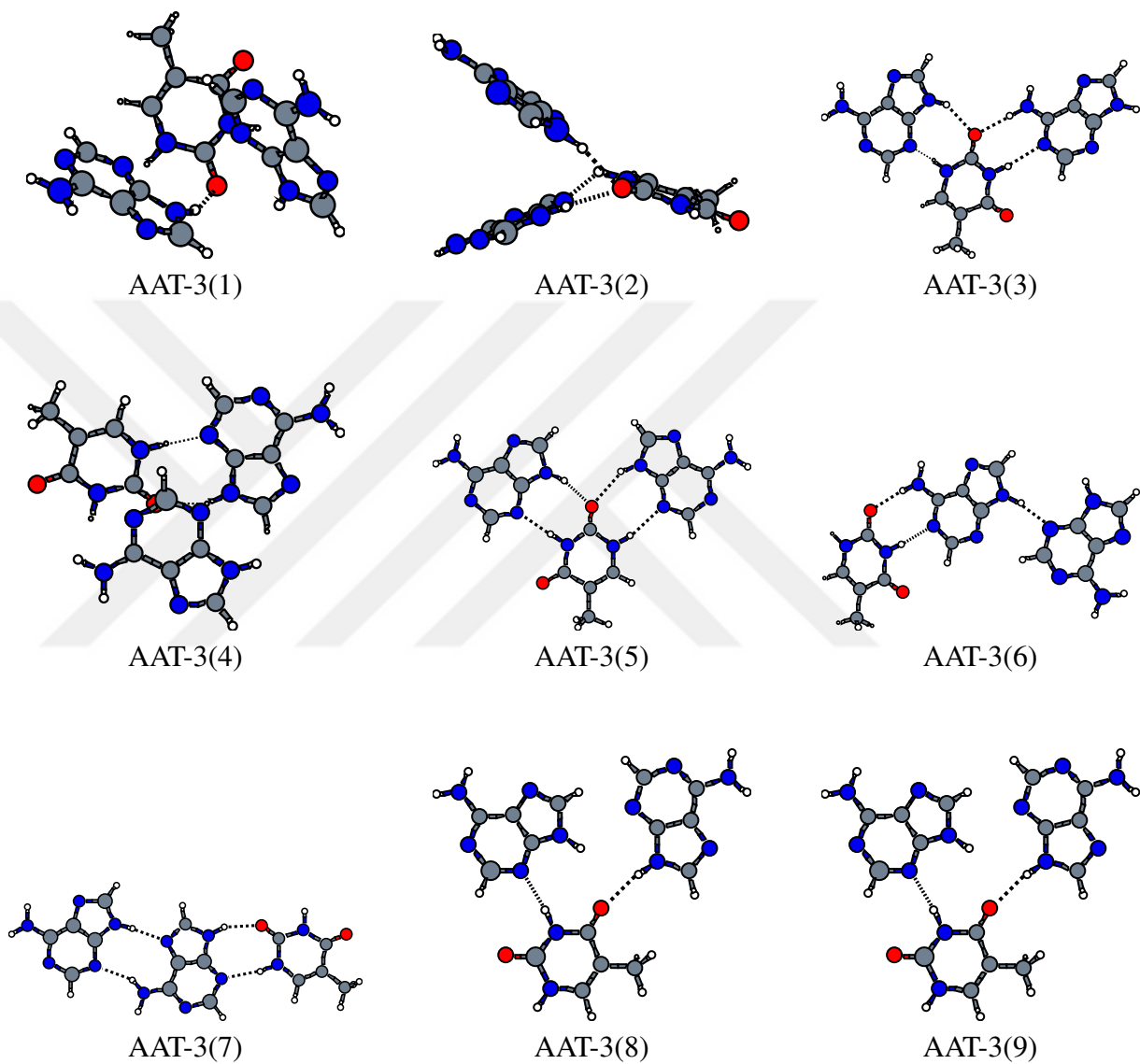
Dimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	PBE0AC	LPBE0AC	AMBER	Model
AT-2(1) <sup>a</sup>	-72.94	-67.90	-61.58	-68.91	-60.03	-63.99	-63.23	-50.63
AT-2(2) <sup>a</sup>	-64.66	-60.40	-54.16	-60.48	-52.71	-56.25	-58.64	-52.00
AT-2(3) <sup>a</sup>	-66.25	-61.37	-54.91	-61.87	-54.42	-57.83	-58.56	-49.97
AT-2(4) <sup>a</sup>	-62.82	-58.85	-52.13	-58.57	-51.42	-55.07	-57.38	-46.67
AT-2(5) <sup>a</sup>	-62.62	-57.23	-50.94	-56.91	-51.04	-54.20	-51.05	-46.01
AT-2(6) <sup>a</sup>	-63.49	-57.81	-51.42	-57.59	-51.88	-57.98	-50.79	-45.51
AT-2(7) <sup>a</sup>	-59.27	-53.01	-47.08	-52.58	-46.82	-49.95	-52.30	-46.23
AT-2(8) <sup>a</sup>	-62.66	-56.38	-50.66	-55.81	-49.97	-53.34	-52.52	-42.10
AT-2(9) <sup>a</sup>	-60.42	-53.75	-50.66	-53.57	-47.99	-51.01	-52.19	-45.04
AT-2(10) <sup>a</sup>	-38.76	-42.55	-31.05	-35.95	-30.82	-33.99	-42.17	-36.86
AT-2(11) <sup>a</sup>	-38.72	-43.65	-31.50	-36.46	-31.14	-34.17	-43.29	-37.30
AT-2(12) <sup>a</sup>	-35.34	-39.59	-28.92	-33.66	-28.47	-31.47	-39.54	-35.41
AT-2(13) <sup>a</sup>	-55.66	-50.82	-45.41	-50.11	-45.11	-48.09	-44.10	-36.23
AT-2(14) <sup>a</sup>	-42.09	-37.49	-32.36	-38.01	-33.08	-36.04	-18.03	-31.34
AT-2(1) <sup>b</sup>	-89.87	-82.09	-73.62	-73.63	-82.09	-77.22	-68.69	-50.58
AT-2(2) <sup>b</sup>	-77.76	-71.08	-63.19	-63.19	-71.08	-66.35	-62.68	-52.82
AT-2(3) <sup>b</sup>	-80.36	-72.09	-64.21	-64.21	-72.59	-68.32	-62.92	-50.38
AT-2(4) <sup>b</sup>	-62.82	-58.85	-52.13	-52.13	-57.85	-55.07	-57.38	-46.67
AT-2(5) <sup>b</sup>	-74.15	-67.01	-59.04	-59.04	-67.01	-63.62	-53.35	-47.13
AT-2(6) <sup>b</sup>	-74.15	-66.69	-58.94	-58.94	-66.68	-63.78	-53.44	-46.83
AT-2(7) <sup>b</sup>	-71.75	-62.72	-55.23	-55.23	-62.72	-59.66	-55.53	-47.43
AT-2(8) <sup>b</sup>	-77.83	-68.78	-61.27	-61.27	-68.78	-65.20	-58.30	-42.57
AT-2(9) <sup>b</sup>	-73.46	-63.70	-55.93	-55.93	-63.70	-60.80	-55.72	-48.07
AT-2(10) <sup>b</sup>	-50.29	-55.35	-36.79	-36.79	-55.35	-39.28	-38.42	-33.29
AT-2(11) <sup>b</sup>	-50.55	-54.82	-36.87	-36.87	-54.82	-40.30	-42.79	-37.16
AT-2(12) <sup>a</sup>	-48.60	-53.38	-35.99	-35.99	-53.38	-38.94	-41.75	-36.43
AT-2(13) <sup>a</sup>	-72.43	-64.54	-56.87	-56.87	-64.45	-61.41	-50.65	-37.18
AT-2(14) <sup>a</sup>	-52.20	-46.57	-40.42	-40.42	-46.57	-44.53	-33.41	-29.16

<sup>a</sup> Using model geometries.

<sup>b</sup> Using CP-SCS-MP2/aVDZ geometries.

### 3.3.3.2 Trimer

For trimer and beyond, we adopted our assumption of neglecting the many-body effects and implemented the same technique that we used on adenine and thymine oligomers. But for adenine-thymine trimers, we must search two different trimer pairings. Adenine-thymine trimers consist of either two adenine and one thymine, AAT or two thymine and one adenine, ATT. These geometries can be seen in Figure 3.21. Amongst them, AT-3(1), AT-3(2) and AT-3(8) are not planar and formed by AT-2(3), AT-2(2) and AT-2(1) with one additional adenine dimer, respectively. AT-2(11) has both H-bonded and stacked interactions while AT-3(18) has formed by adenine-adenine stacked with one additional thymine dimer on top of them. Corresponding to AT-3(18), AT-3(9) has formed by adenine-thymine stacked with one additional adenine dimer besides them. Other found structures shown in Figure 3.21 are H-bonded ribbons. All resulting structures in Figure 3.21 were further relaxed at SCS-MP2 level, employing aVDZ basis set and no significant change has been observed. Interaction energies of the SCS-MP2 corrected geometries were calculated at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 levels and these were listed in Table 3.19. From Table 3.19, we can see that AT-3-(2) has the lowest energy, followed by AT-3(1), AT-3(13) and AT-3(3) by -1.06 kJ/mol, -2.25 kJ/mol and -3.87 kJ/mol, respectively. All ab-initio calculations favored AT-3(1) over AT-3(2) with small differences while SCS-MP2 relaxed model energies still favour AT-3(2) over AT-3(1) by -1.69 kJ/mol. Also for SCS-MP2 relaxed ab-initio calculations, AT-3(10) has the lowest energy while our model and AMBER still favour other orientations. AMBER favours AT-3(1) as the lowest energy, followed by AT-3(2) with -1.02 kJ/mol difference and our model favours AT-3(2) over AT-3(1).



**Figure 3.21** : Adenine-Thymine trimers found by the SA approach.

**Table 3.19** : Interaction energy calculations (at AMBER) using aVDZ) of adenine-thymine trimers shown in Figure 3.21.

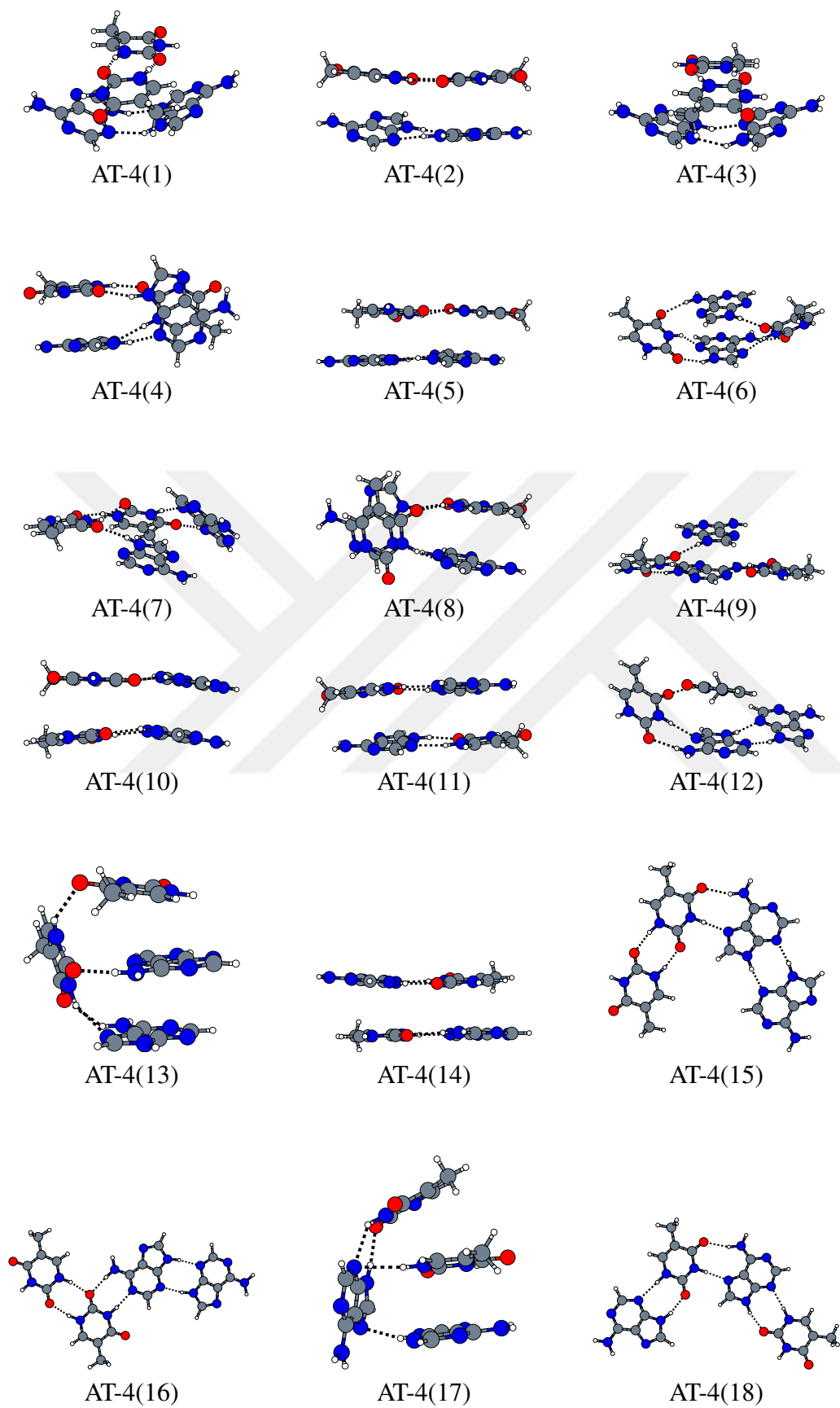
Tetramer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
AT-3(1) <sup>a</sup>	-127.10	-133.01	-106.25	-121.07	-132.26	-111.24
AT-3(2) <sup>a</sup>	-126.15	-132.22	-105.55	-120.18	-131.63	-112.30
AT-3(3) <sup>a</sup>	-117.88	-121.50	-96.23	-109.63	-123.62	-108.43
AT-3(4) <sup>a</sup>	-125.90	-126.35	-100.40	-114.41	-124.21	-93.84
AT-3(5) <sup>a</sup>	-142.85	-130.46	-115.60	-128.35	-122.48	-92.65
AT-3(6) <sup>a</sup>	-142.53	-132.22	-119.02	-133.04	-124.58	-103.66
AT-3(7) <sup>a</sup>	-144.65	-133.50	-118.50	-131.47	-118.60	-90.12
AT-3(8) <sup>a</sup>	-124.25	-125.17	-101.80	-115.77	-122.29	-94.28
AT-3(9) <sup>a</sup>	-119.31	-121.21	-98.01	-111.96	-116.18	-92.27
AT-3(10) <sup>a</sup>	-134.52	-123.82	-110.99	-123.65	-112.32	-99.77
AT-3(11) <sup>a</sup>	-129.19	-116.75	-102.85	-114.20	-107.20	-84.04
AT-3(12) <sup>a</sup>	-131.11	-120.82	-107.94	-120.46	-109.96	-93.46
AT-3(13) <sup>a</sup>	-127.01	-116.76	-102.90	-115.09	-108.00	-110.05
AT-3(14) <sup>a</sup>	-124.97	-111.93	-98.45	-109.14	-109.58	-85.46
AT-3(15) <sup>a</sup>	-125.46	-114.05	-100.22	-111.28	-105.15	-81.83
AT-3(16) <sup>a</sup>	-120.70	-107.51	-94.48	-105.47	-106.30	-92.04
AT-3(17) <sup>a</sup>	-109.99	-99.48	-87.04	-98.00	-71.57	-76.63
AT-3(18) <sup>a</sup>	-76.61	-92.49	-65.70	-74.80	-85.09	-73.93
AT-3(1) <sup>b</sup>	-155.87	-161.88	-124.96	-146.18	-140.67	-117.95
AT-3(2) <sup>b</sup>	-154.56	-160.44	-123.99	-144.90	-139.65	-119.64
AT-3(3) <sup>b</sup>	-138.82	-144.71	-110.71	-129.39	-130.14	-109.06
AT-3(4) <sup>b</sup>	-151.73	-151.33	-116.98	-135.65	-134.67	-96.08
AT-3(5) <sup>b</sup>	-164.89	-149.53	-132.28	-148.13	-127.44	-94.78
AT-3(6) <sup>b</sup>	-170.69	-155.13	-137.80	-157.06	-132.54	-102.37
AT-3(7) <sup>b</sup>	-168.70	-154.20	-136.20	-152.94	-123.25	-91.61
AT-3(8) <sup>b</sup>	-150.33	-148.01	-115.97	-135.72	-128.28	-91.73
AT-3(9) <sup>b</sup>	-146.34	-143.95	-113.68	-132.31	-118.26	-92.96
AT-3(10) <sup>b</sup>	-163.56	-147.84	-130.80	-148.65	-120.34	-100.21
AT-3(11) <sup>b</sup>	-153.94	-138.52	-121.98	-136.70	-112.57	-85.65
AT-3(12) <sup>b</sup>	-160.30	-145.33	-128.43	-145.93	-116.88	-93.42
AT-3(13) <sup>b</sup>	-156.33	-142.01	-125.18	-141.30	-117.93	-115.56
AT-3(14) <sup>b</sup>	-149.91	-133.56	-117.77	-131.52	-115.88	-87.94
AT-3(15) <sup>b</sup>	-147.43	-133.66	-117.44	-131.43	-111.01	-83.76
AT-3(16) <sup>b</sup>	-143.15	-127.49	-112.40	-126.08	-112.65	-94.35
AT-3(17) <sup>b</sup>	-130.02	-117.14	-103.00	-116.06	-93.21	-78.18
AT-3(18) <sup>b</sup>	-88.54	-108.15	-69.27	-82.68	-70.08	-64.88

<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.

### 3.3.3.3 Tetramer

Our model can predict more complex structures such as AT-tetrads so we further tested our model to obtain these geometries. As a result, our model generated 18 different AT-tetramers including ribbons, stacked geometries, and AT-tetrads. AT-tetrads, especially, inspected heavily in the literature by various researchers [85, 93, 94]. For example, AT-4(14) was previously inspected by Gil et al. [94] and Bravaya et al. [93] calculate the cooperative effect of H-Bonding and  $\pi$ -stacking on the ionization energy of adenine, which contains two AT-2(9) stacked with adenine dimers. AT-4(1) and AT-4(2) both contain T-2(1) stacked with adenine but while AT-4(2) align itself horizontally on adenine dimer, AT-4(1) prefers to align vertically. AT-4(3) and AT-4(5) are very similar to AT-4(1) and AT-4(2) but with different alignments. AT-4(6) contains ATT ribbon with one additional adenine dimers align them vertically. AT-4(4) and AT-4(8) contains two stacked adenine and thymine dimers combined while other found AT tetramers including AT-4(15), AT-4(16) and AT-4(18) form ribbons. All resulting geometries in Figure 3.21 were further relaxed with only SCS-MP2 level, employing aVDZ basis set and no significant change in the geometries was observed. Interaction energies of the SCS-MP2 corrected geometries were calculated at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 levels and this was listed in Table 3.20. From Table 3.20, we can see that AT-4(6) is favoured as the lowest energy by both our model and AMBER, while ab-initio calculations differ. For example, SCS-MP2 favours AT-4(4) as the lowest energy followed by AT-4(2) and AT-4(8) while B3LYP-D, MP2, and SCS-MI-MP2 favours AT-4(8) as the lowest energy followed by AT-4(6) with difference between them is -12.08 kJ/mol, -11.83 kJ/mol and 9.15 kJ/mol, respectively. In conclusion, both AMBER and model favoured AT-4(6) as the lowest energy followed by AT-4(1), with difference -9.36 kJ/mol and -0.76 kJ/mol, respectively.



**Figure 3.22** : Adenine-Thymine tetramer structures found by the SA approach.

**Table 3.20** : Interaction energy calculations (at AMBER) using aVDZ) of adenine-thymine tetramers shown in Figure 3.22.

Tetramer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model
AT-4(1) <sup>a</sup>	-202.72	-197.55	-157.54	-181.94	-208.14	-175.91
AT-4(2) <sup>a</sup>	-211.58	-210.84	-171.51	-195.35	-208.42	-172.36
AT-4(3) <sup>a</sup>	-199.86	-194.62	-155.57	-179.71	-201.96	-170.55
AT-4(4) <sup>a</sup>	-214.37	-217.09	-177.94	-201.68	-211.70	-179.94
AT-4(5) <sup>a</sup>	-205.32	-206.10	-164.78	-188.86	-203.05	-164.77
AT-4(6) <sup>a</sup>	-197.44	-196.47	-159.27	-182.43	-201.21	-185.86
AT-4(7) <sup>a</sup>	-205.82	-204.23	-167.35	-192.54	-210.20	-174.17
AT-4(8) <sup>a</sup>	-208.80	-210.82	-171.21	-195.78	-203.87	-173.37
AT-4(9) <sup>a</sup>	-186.04	-184.89	-148.96	-171.59	-193.18	-176.58
AT-4(10) <sup>a</sup>	-205.62	-210.44	-170.25	-193.25	-204.73	-177.78
AT-4(11) <sup>a</sup>	-214.01	-216.79	-178.90	-201.34	-206.24	-176.64
AT-4(12) <sup>a</sup>	-202.54	-182.13	-139.02	-162.20	-194.18	-155.82
AT-4(13) <sup>a</sup>	-174.77	-182.13	-139.02	-162.20	-184.70	-162.42
AT-4(14) <sup>a</sup>	-199.08	-203.09	-163.81	-186.08	-193.88	-168.98
AT-4(15) <sup>a</sup>	-214.26	-196.01	-174.55	-196.36	-181.58	-152.35
AT-4(16) <sup>a</sup>	-206.18	-187.38	-166.20	-187.53	-180.74	-153.14
AT-4(17) <sup>a</sup>	-171.05	-176.33	-136.52	-153.15	-168.37	-159.55
AT-4(18) <sup>a</sup>	-209.21	-193.46	-174.30	-193.79	-176.43	-151.45
AT-4(1) <sup>b</sup>	-247.06	-235.64	-190.34	-221.23	-218.77	-179.93
AT-4(2) <sup>b</sup>	-261.72	-258.35	-203.62	-236.87	-201.80	-165.10
AT-4(3) <sup>b</sup>	-246.24	-233.91	-188.82	-220.07	-210.96	-172.08
AT-4(4) <sup>b</sup>	-275.14	-273.13	-115.75	-250.84	-207.94	-166.97
AT-4(5) <sup>b</sup>	-240.39	-241.69	-188.13	-218.67	-199.85	-154.73
AT-4(6) <sup>b</sup>	-245.66	-244.45	-192.79	-225.86	-219.53	-189.29
AT-4(7) <sup>b</sup>	-243.75	-240.39	-191.80	-225.87	-207.86	-160.21
AT-4(8) <sup>b</sup>	-257.74	-256.28	-200.29	-235.01	-200.97	-165.40
AT-4(9) <sup>b</sup>	-232.97	-231.64	-183.51	-213.15	-199.44	-168.68
AT-4(10) <sup>b</sup>	-250.95	-249.81	-194.18	-227.76	-206.61	-175.45
AT-4(11) <sup>b</sup>	-256.15	-252.99	-200.28	-232.03	-210.18	-169.14
AT-4(12) <sup>b</sup>	-242.83	-232.76	-188.61	-219.08	-202.42	-159.91
AT-4(13) <sup>b</sup>	-212.24	-219.64	-161.76	-192.19	-184.22	-166.76
AT-4(14) <sup>b</sup>	-236.09	-236.75	-183.25	-213.94	-190.06	-155.49
AT-4(15) <sup>b</sup>	-259.85	-234.70	-208.35	-236.86	-188.16	-152.81
AT-4(16) <sup>b</sup>	-252.46	-226.84	-201.45	-228.82	-187.66	-154.46
AT-4(17) <sup>b</sup>	-221.44	-218.90	-164.61	-194.69	-194.01	-156.14
AT-4(18) <sup>b</sup>	-256.06	-231.98	-206.16	-233.95	-189.48	-150.75

<sup>a</sup> Using model geometries.

<sup>b</sup> Using SCS-MP2/aVDZ geometries.





## 4. CONCLUSION

In this thesis, we aim to generate an intermolecular potential function force field for adenine, thymine and adenine-thymine dimers and oligomers which is in a good agreement with the well known AMBER force field. Our first step to achieve this objective to find the most suitable ab-initio calculation method to generate the PESs. For this purpose, we compare the performances of B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) on the dimer geometries of adenine, thymine and adenine-thymine to find the most suitable and accurate method. Results of ab-initio calculations in adenine, thymine, and adenine-thymine showed us that we can generate a suitable force field via symmetry-adapted perturbation theory (DFT-SAPT). In addition to its accuracy, DFT-SAPT is useful in determining the stabilizing factors of the given system by giving the interaction energy as a sum of electrostatic, dispersion, induction and repulsive terms. In order to proceed and generate PESs, we used DFT-SAPT calculations on 7286, 4412 and 6390 grid points for adenine, thymine, and adenine-thymine, respectively. Results of this calculations were used in the fitting process and our proposed fitting was very successful and the performance of the model potential was tested multiple times with a potential energy searches of dimers, trimers and tetramers via Simulated Annealing. Our proposed model was able to reproduce the previously reported and well-known adenine, thymine and adenine-thymine dimers as well as new orientations. Cluster structures generated from our model also been relaxed at CP-SCS-MP2/aVDZz (for dimer) and SCS-MP2/aVDZ (trimer, tetramer) levels. We showed that there is a slight difference between our model and the corresponding ab-initio calculations. Although the proposed model and AMBER force fields are in agreement for most cases, AMBER can fail to calculate the correct interaction energies for certain cases such as A-2(9), A-2(12), T-2(12) and T-3(9). For all considered DNA bases, our proposed model can able to give more consistent results

than AMBER, therefore, our model can be applied in all simulations of much larger adenine, thymine and adenine-thymine clusters.



## REFERENCES

- [1] Mergny, J.L., Lacroix, L., Han, X., Leroy, J.L. and Hélène, C. (1995). Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an i-DNA Motif, *J. Am. Chem. Soc.*, 117(35), 8887–8898.
- [2] Kypr, J., Kejnovská, I., Renčiuk, D. and Vorličová, M. (2009). Circular dichroism and conformational polymorphism of DNA, *Nuclear Acids Research*, 37(6), 1713–1725.
- [3] Vesenka, J., Marsh, T., Henderson, E. and Vellandi, C. (1998). THE DIAMETER OF DUPLEX AND QUADRUPLEX DNA MEASURED BY SCANNING PROBE MICROSCOPY, *Scanning Microscopy*, 12(2), 329–342.
- [4] Kelly, R.E.A., Lee, Y.J. and Kantorovich, L.N. (2005). Homopairing Possibilities of the DNA Bases Cytosine and Guanine: An ab Initio DFT Study, *The Journal of Physical Chemistry B*, 109(46), 22045–22052, <http://pubs.acs.org/doi/abs/10.1021/jp055207z>, pMID: 16853862, <http://pubs.acs.org/doi/pdf/10.1021/jp055207z>.
- [5] Kelly, R.E.A., Lukas, M., Kantorovich, L.N., Otero, R., Xu, W., Murali, M., Laegsgaard, E., Stensgaard, I. and Besenbacher, F. (2008). Understanding the disorder of the DNA base cytosine on the Au(111) surface, *Journal of Chemical Physics*, 129, 184707.
- [6] Maleki, A., Alavi, S. and Najafi, B. (2011). Molecular Dynamics Simulation Study of Adsorption and Patterning of DNA Bases on the Au (111) Surface, *The Journal of Physical Chemistry C*, 115(45), 22484–22494.
- [7] Camargo, A., Baumgärtel, H. and Donner, C. (2003). Coadsorption of the DNA bases thymine and adenine at the Au (111) electrode, *Physical Chemistry Chemical Physics*, 5(8), 1657–1664.
- [8] Tanaka, H., Nakagawa, T. and Kawai, T. (1996). Two-dimensional self-assembly of DNA base molecules on Cu (111) surfaces, *Surface science*, 364(2), L575–L579.
- [9] Kawai, T., Tanaka, H. and Nakagawa, T. (1997). Low dimensional self-organization of DNA-base molecules on Cu (111) surfaces, *Surface science*, 386(1-3), 124–136.
- [10] Furukawa, M., Tanaka, H. and Kawai, T. (2001). The role of dimer formation in the self-assemblies of DNA base molecules on Cu (111) surfaces: A

scanning tunneling microscope study, *The Journal of Chemical Physics*, 115(7), 3419–3423.

- [11] **Lacroix, L., Mergny, J.L., Leroy, J.L. and Hélène, C.** (1996). Inability of RNA To Form the i-Motif: Implications for Triplex Formation, *Biochemistry*, 35, 8715–8722.
- [12] **Borbone, N., Amato, J., Oliviero, G., D’Atri, V., Gabelica, V., Pauw, E.D., Piccialli, G. and Mayol, L.** (2011). d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads, *Nucleic Acids Research*, 39(17), 7848–7857.
- [13] **Pagba, C.V., Lane, S.M. and Wachsmann-Hogiu, S.** (2011). Conformational changes in quadruplex oligonucleotide structures probed by Raman spectroscopy, *BIOMEDICAL OPTICS EXPRESS*, 2(2), 207.
- [14] **Guéron, M. and Leroy, J.L.** (2000). The i-motif in nucleic acids, *Current Opinion in Structural Biology*, 10(3), 326–331, <http://www.sciencedirect.com/science/article/pii/S0959440X00000919>.
- [15] **Egli, M.** (2004). Nucleic acid crystallography: current progress, *Current Opinion in Chemical Biology*, 8, 580–591.
- [16] **Patel, P., Bhavesh, N.S. and Hosur, R.** (2000). NMR Observation of a Novel C-Tetrad in the Structure of the SV40 Repeat Sequence GGGCGG, *Biochemical and Biophysical Research Communications*, 270(3), 967–971, <http://www.sciencedirect.com/science/article/pii/S0006291X00924793>.
- [17] **Patel, P., Bhavesh, N.S. and Hosur, R.** (2000). Cation-Dependent Conformational Switches in d-TGGCGGC Containing Two Triplet Repeats of Fragile X Syndrome: NMR Observations, *Biochemical and Biophysical Research Communications*, 278, 833–838.
- [18] **Li, T., Ackermann, D., Hall, A.M. and Famulok, M.** (2012). Input-Dependent Induction of Oligonucleotide Structural Motifs for Performing Molecular Logic, *J. Am. Chem. Soc.*, 134, 3508–3516.
- [19] **Pan, B., Shi, K. and Sundaralingam, M.** (2006). Base-tetrad swapping results in dimerization of RNA quadruplexes: Implications for formation of the i-motif RNA octaplex, *PNAS*, 103(9), 3130–3134.
- [20] **Boon, E.M., Ceres, D.M., Drummond, T.G., Hill, M.G. and Barton, J.K.** (2000). *Nature biotechnology*, 18(10), 1096–1100.
- [21] **Wang, J.** (2000). *Nucleic Acids Research*, 28(16), 3011–3016.
- [22] **McKendry, R., Zhang, J., Arntz, Y., Strunz, T., Hegner, M., Lang, H.P., Baller, M.K., Certa, U., Meyer, E., Güntherodt, H.J. and Gerber, C.** (2002). *Proceedings of the National Academy of Sciences*, 99(15), 9783–9788.

- [23] **Furukawa, M., Tanaka, H. and Kawai, T.** (1997). Scanning tunneling microscopy observation of two-dimensional self-assembly formation of adenine molecules on Cu (111) surfaces, *Surface science*, 392(1-3), L33–L39.
- [24] **Furukawa, M., Tanaka, H., Sugiura, K.i., Sakata, Y. and Kawai, T.** (2000). Fabrication of molecular alignment at the specific sites on Cu (111) surfaces using self-assembly phenomena, *Surface science*, 445(1), L58–L63.
- [25] **Furukawa, M., Tanaka, H. and Kawai, T.** (2000). Formation mechanism of low-dimensional superstructure of adenine molecules and its control by chemical modification: a low-temperature scanning tunneling microscopy study, *Surface science*, 445(1), 1–10.
- [26] **Chen, Q. and Richardson, N.** (2003). Enantiomeric interactions between nucleic acid bases and amino acids on solid surfaces, *Nature materials*, 2(5), 324–328.
- [27] **Chen, Q., Frankel, D.J. and Richardson, N.V.** (2002). Self-assembly of adenine on Cu (110) surfaces, *Langmuir*, 18(8), 3219–3225.
- [28] **Tao, N., DeRose, J. and Lindsay, S.** (1993). Self-assembly of molecular superstructures studied by in situ scanning tunneling microscopy: DNA bases on gold (111), *The Journal of Physical Chemistry*, 97(4), 910–919.
- [29] **Lukas, M., Kelly, R.E., Kantorovich, L.N., Otero, R., Xu, W., Laegsgaard, E., Stensgaard, I. and Besenbacher, F.** (2009). Adenine monolayers on the Au (111) surface: Structure identification by scanning tunneling microscopy experiment and ab initio calculations, *The Journal of chemical physics*, 130(2), 024705.
- [30] **Tao, N. and Shi, Z.** (1994). Monolayer guanine and adenine on graphite in NaCl solution: a comparative STM and AFM study, *The Journal of Physical Chemistry*, 98(5), 1464–1471.
- [31] **Sowerby, S.J., Heckl, W.M. and Petersen, G.B.** (1996). Chiral symmetry breaking during the self-assembly of monolayers from achiral purine molecules, *Journal of Molecular Evolution*, 43(5), 419–424.
- [32] **Freund, J., Edelwirth, M., Kröbel, P. and Heckl, W.** (1997). Structure determination of two-dimensional adenine crystals on graphite, *Physical Review B*, 55(8), 5394.
- [33] **Uchihashi, T., Okada, T., Sugawara, Y., Yokoyama, K. and Morita, S.** (1999). Self-assembled monolayer of adenine base on graphite studied by noncontact atomic force microscopy, *Physical Review B*, 60(11), 8309.
- [34] **Mamdouh, W., Dong, M., Kelly, R.E., Kantorovich, L.N. and Besenbacher, F.** (2007). Coexistence of homochiral and heterochiral adenine domains at the liquid/solid interface, *The Journal of Physical Chemistry B*, 111(42), 12048–12052.

- [35] **Perdigão, L.M., Staniec, P.A., Champness, N.R., Kelly, R., Kantorovich, L. and Beton, P.H.** (2006). Experimental and theoretical identification of adenine monolayers on Ag-terminated Si (111), *Physical Review B*, 73(19), 195423.
- [36] **Otero, R., Lukas, M., Kelly, R.E.A., Xu, W., Lægsgaard, E., Stensgaard, I., Kantorovich, L.N. and Besenbacher, F.** (2008). *Science*, 319(5861), 312–315.
- [37] **Furukawa, M., Tanaka, H. and Kawai, T.** (2001). *The Journal of Chemical Physics*, 115, 3419.
- [38] **Frankel, D.J., Chen, Q. and Richardson, N.V.** (2006). *The Journal of chemical physics*, 124, 204704.
- [39] **Kelly, R.E.A., 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.
- [40] **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.
- [41] **Xu, W., Kelly, R.E.A., Otero, R., Schöck, M., Lægsgaard, E., Stensgaard, I., Kantorovich, L.N. and Besenbacher, F.** (2007). *Small*, 3(12), 2011–2014.
- [42] **Krull, C., Valencia, S., Pascual, J.I. and Theis, W.** (2009). *Applied Physics A*, 95(1), 297–301.
- [43] **Kelly, R.E.A. and Kantorovich, L.N.** (2006). *Journal of Materials Chemistry*, 16(20), 1894–1905.
- [44] **Maleki, A., Alavi, S. and Najafi, B.** (2011). *The Journal of Physical Chemistry C*, 115(45), 22484–22494.
- [45] **Manukyan, A. and Tekin, A.** (2017). The intermolecular dimer potential for guanine, *The Journal of Chemical Physics*, 147(15), 154311.
- [46] **Manukyan, A. and Tekin, A.** (2015). First principles potential for the cytosine dimer, *Physical Chemistry Chemical Physics*, 17(22), 14685–14701.
- [47] **Jansen, G. and Håbelmann, A.** (2001). *J. Phys. Chem. A*, 105, 11156.
- [48] **Håbelmann, A. and Jansen, G.** (2002). *Chem. Phys. Lett.*, 357, 464.
- [49] **Håbelmann, A. and Jansen, G.** (2002). *Chem. Phys. Lett.*, 362, 319.
- [50] **Håbelmann, A. and Jansen, G.** (2003). *Chem. Phys. Lett.*, 367, 778.
- [51] **Parr, R.G., Craig, D.P. and Ross, I.G.** (1950). Molecular orbital calculations of the lower excited electronic levels of benzene, configuration interaction included, *The Journal of Chemical Physics*, 18(12), 1561–1563.

- [52] **Grimme, S.** (2003). *J. Chem. Phys.*, *118*, 9095.
- [53] **Distasio Jr, R.A. and Head-Gordon, M.** (2007). Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies, *Molecular Physics*, *105*(8), 1073–1083.
- [54] **Hohenberg, P. and Kohn, W.** (1964). *Phys Rev B* *136*: 864. doi: 10.1103/PhysRev.136.B864.
- [55] **Jeziorski, B., Moszynski, R. and Szalewicz, K.** (1994). Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes, *Chemical Reviews*, *94*(7), 1887–1930.
- [56] **Williams, H.L. and Chabalowski, C.F.** (2001). Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions, *The Journal of Physical Chemistry A*, *105*(3), 646–659.
- [57] **Misquitta, A.J., Podaszwa, R., Jeziorski, B. and Szalewicz, K.** (2005). Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations, *The Journal of chemical physics*, *123*(21), 214103.
- [58] **Werner, H.J., Knowles, P.J., Lindh, R., Manby, F.R., Schütz, M., Celani, P., Korona, T., Amos, G.R.R.D., Bernhardsson, A., Berning, A., Cooper, D.L., Deegan, M.J.O., Dobbyn, A.J., Hampel, F.E.C., Hetzer, G., Lloyd, A.W., McNicholas, S.J., Mura, W.M.M.E., Nicklass, A., P, P.P., Pitzer, R., Schumann, U., Stoll, H., AJ, A.J.S., Tarroni, R. and Thorsteinsson, T.**, MOLPRO, version 2009.1, a package of ab initio program, <http://www.molpro.net>.
- [59] **Hesselmann, A., Jansen, G. and Schütz, M.** (2005). Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies, *The Journal of chemical physics*, *122*(1), 014103.
- [60] **Distasio, R.A. and Head-Gordon, M.** (2007). Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies, *Molecular Physics*, *105*, 1073.
- [61] **Grimme, S.** (2004). *J. Comput. Chem.*, *25*, 1463.
- [62] **Grimme, S.** (2006). *J. Comput. Chem.*, *27*, 1787.
- [63] **Busker, M., Svartsov, Y.N., Häber, T. and Kleinermanns, K.** (2009). *Chem. Phys. Lett.*, *467*, 255.
- [64] **Mishra, B.K., Arey, J.S. and Sathyamurthy, N.** (2010). *J. Phys. Chem. A*, *114*, 9606.
- [65] **Sütay, B., Tekin, A. and Yurtsever, M.** (2012). Intermolecular interactions in nitrogen-containing aromatic systems, *Theoretical Chemistry Accounts*, *131*, 1–13, <http://dx.doi.org/10.1007/s00214-012-1120-3>.

- [66] **Sherrill, C.D., Takatani, T. and Hohenstein, E.G.** (2009). *J. Phys. Chem. A*, *113*, 10146.
- [67] **Haßelmann, A. and Jansen, G.** (2003). *Phys. Chem. Chem. Phys.*, *5*, 5010.
- [68] **Tekin, A. and Jansen, G.** (2007). How accurate is the density functional theory combined with symmetry-adapted perturbation theory approach for CH-[small pi] and [small pi]-[small pi] interactions? A comparison to supermolecular calculations for the acetylene-benzene dimer, *Phys. Chem. Chem. Phys.*, *9*, 1680–1687, <http://dx.doi.org/10.1039/B618997K>.
- [69] **Morgado, C.A., Jurečka, P., Svozil, D., Hobza, P. and Šponer, J.** (2010). Reference MP2/CBS and CCSD (T) quantum-chemical calculations on stacked adenine dimers. Comparison with DFT-D, MP2.5, SCS (MI)-MP2, M06-2X, CBS (SCS-D) and force field descriptions, *Physical Chemistry Chemical Physics*, *12*(14), 3522–3534.
- [70] **Bravaya, K.B., Kostko, O., Ahmed, M. and Krylov, A.I.** (2010). The effect of  $\pi$ -stacking, H-bonding, and electrostatic interactions on the ionization energies of nucleic acid bases: adenine–adenine, thymine–thymine and adenine–thymine dimers, *Physical Chemistry Chemical Physics*, *12*(10), 2292–2307.
- [71] **Haßelmann, A., Jansen, G. and Schütz, M.** (2005). *J. Chem. Phys.*, *122*, 014103.
- [72] **Ahlrichs, R., Bär, M., Häser, M., Horn, H. and Kölmel, C.** (1989). *Chem. Phys. Lett.*, *162*, 165.
- [73] **Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spelimeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A.** (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc.*, *117*, 5179.
- [74] **Morgado, C.A., Jurečka, P., Svozil, D., Hobza, P. and Šponer, J.** (2010). Reference MP2/CBS and CCSD(T) quantum-chemical calculations on stacked adenine dimers. Comparison with DFT-D, MP2.5, SCS(MI)-MP2, M06-2X, CBS(SCS-D) and force field descriptions, *Phys. Chem. Chem. Phys.*, *12*, 3522.
- [75] **Leforestier, C., Tekin, A., Jansen, G. and Herman, M.** (2011). First principles potential for the acetylene dimer and refinement by fitting to experiments, *The Journal of chemical physics*, *135*(23), 12B603.
- [76] **Tang, K. and Toennies, J.P.** (1984). An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients, *The Journal of chemical physics*, *80*(8), 3726–3741.
- [77] **Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E.** (1953). Equation of state calculations by fast computing machines, *The journal of chemical physics*, *21*(6), 1087–1092.



- [78] **Sánchez-García, E., Mardyukov, A., Tekin, A., Crespo-Otero, R., Montero, L.A., Sander, W. and Jansen, G.** (2008). Ab initio and matrix isolation study of the acetylene–furan dimer, *Chemical Physics*, *343*(2), 168–185.
- [79] **Sütay, B., Tekin, A. and Yurtsever, M.** (2012). Intermolecular interactions in nitrogen-containing aromatic systems, *Theoretical Chemistry Accounts*, *131*(2), 1120.
- [80] **Bak, K.L., Jørgensen, P., Olsen, J., Helgaker, T. and Klopper, W.** (2000). Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations, *The Journal of Chemical Physics*, *112*(21), 9229–9242.
- [81] **Janowski, T. and Pulay, P.** (2007). High accuracy benchmark calculations on the benzene dimer potential energy surface, *Chemical Physics Letters*, *447*(1), 27–32.
- [82] **Corona, A., Marchesi, M., Martini, C. and Ridella, S.** (1987). Minimizing multimodal functions of continuous variables with the simulated annealing algorithm, *ACM Transactions on Mathematical Software*, *13*(3), 262–280.
- [83] **Rajabi, K., Theel, K., Gillis, E.A., Beran, G. and Fridgen, T.D.** (2009). The structure of the protonated adenine dimer by infrared multiple photon dissociation spectroscopy and electronic structure calculations, *The Journal of Physical Chemistry A*, *113*(28), 8099–8107.
- [84] **Preuss, M. and Bechstedt, F.** (2008). Self-assembly of adenine-dimer chains on Cu (110): Driving forces from first-principles calculations, *Surface Science*, *602*(9), 1643–1649.
- [85] **Sun, H., Zhang, S. and Sun, Z.** (2015). Applicability of optimal functional tuning in density functional calculations of ionization potentials and electron affinities of adenine–thymine nucleobase pairs and clusters, *Physical Chemistry Chemical Physics*, *17*(6), 4337–4345.
- [86] **Mamdouh, W., Dong, M., Xu, S., Rauls, E. and Besenbacher, F.** (2006). Supramolecular nanopatterns self-assembled by adenine–thymine quartets at the liquid/solid interface, *Journal of the American Chemical Society*, *128*(40), 13305–13311.
- [87] **Chiriki, S., Dagar, A. and Bulusu, S.S.** (2015). Structural evolution of nucleobase clusters using force field models and density functional theory, *Chemical Physics Letters*, *634*, 166–173.
- [88] **Plützer, C., Hünig, I., Kleinermanns, K., Nir, E. and de Vries, M.S.** (2003). Pairing of isolated nucleobases: double resonance laser spectroscopy of adenine–thymine, *ChemPhysChem*, *4*(8), 838–842.
- [89] **Monajjemi, M., Chahkandi, B., Zare, K. and Amiri, A.** (2005). Study of the hydrogen bond in different orientations of adenine–thymine base pairs: an ab initio study, *Biochemistry (Moscow)*, *70*(3), 366–376.

- [90] **Kratochvíl, M., Šponer, J. and Hobza, P.** (2000). Global Minimum of the Adenine Thymine Base Pair Corresponds Neither to Watson- Crick Nor to Hoogsteen Structures. Molecular Dynamic/Quenching/AMBER and ab Initio beyond Hartree- Fock Studies, *Journal of the American Chemical Society*, 122(14), 3495–3499.
- [91] **Shishkin, O., Šponer, J. and Hobza, P.** (1999). Intramolecular flexibility of DNA bases in adenine–thymine and guanine–cytosine Watson–Crick base pairs, *Journal of molecular structure*, 477(1-3), 15–21.
- [92] **Mazurkiewicz, K., Harańczyk, M., Gutowski, M. and Rak, J.** (2007). Can an excess electron localize on a purine moiety in the adenine–thymine Watson–Crick base pair? A computational study, *International Journal of Quantum Chemistry*, 107(12), 2224–2232.
- [93] **Bravaya, K.B., Epifanovsky, E. and Krylov, A.I.** (2012). Four bases score a run: ab initio calculations quantify a cooperative effect of H-bonding and  $\pi$ -stacking on the ionization energy of adenine in the AATT tetramer, *The journal of physical chemistry letters*, 3(18), 2726–2732.
- [94] **Gil, A., Branchadell, V., Bertran, J. and Oliva, A.** (2009). An analysis of the different behavior of DNA and RNA through the study of the mutual relationship between stacking and hydrogen bonding, *The Journal of Physical Chemistry B*, 113(14), 4907–4914.

## CURRICULUM VITAE



**Name Surname:** Armağan Karatosun

**Place and Date of Birth:** Istanbul, 16.08.1992

**E-Mail:** karatosuna@be.itu.edu.tr

### Education:

- **M.Sc.:Computational Science and Engineering** 2018, Istanbul Technical University, Informatics Institute, Computational Science and Engineering Department
- **B.Sc.:Mechanical Engineering** 2014, Dogus University, Engineering Faculty, Mechanical Engineering Department

### Publications:

- **Armagan Karatosun, Adem Tekin.** The Thymine Dimer Potential From a Combined Density Functional Theory and Symmetry-Adapted Perturbation Theory. 52nd Symposium on Theoretical Chemistry, Chemistry in Solution (STC 2016), Bochum, Germany. September 2016