

**Quantum Chemical Investigations of Structural and  
Photophysical Properties of Emissive RNA Nucleobase  
Analogues**

by

**Melis Gedik**

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

**Doctor of Philosophy**

**Department of Chemistry**

**University of Alberta**

©Melis Gedik, 2017

# Abstract

The photostability of nucleic acid constituents is crucial in maintaining the integrity of our genetic code. This is achieved by the essential mechanism of ultrafast radiationless decay of the singlet excited state to the electronic ground state of nucleobases. While this mechanism provides resilience against mutations to the genetic code, it results in the non-emissive nature of the canonical building blocks of nucleic acids. Consequently, the conformational dynamics of DNA and RNA are difficult to study by fluorescence spectroscopy and imaging. Investigations on the structure and function of these biomolecules can be achieved with fluorescent probes that mimic the natural components of nucleic acids without perturbing the overall structure. Hence, the design of such "isomorphic" nucleoside analogues is highly desired in the field of bioimaging.

The photophysical properties of nucleoside analogues are very sensitive to structural modifications. In particular, the nature of the low-lying excited electronic states, presence of long range excitations, conformational flexibility, tautomerization, and access to non-radiative deactivation pathways such as conical intersections (CIs), have significant impact on fluorescence. The vast majority of fluorescent analogues are also sensitive to their local environment. Base stacking, base sequence dependence and base pairing, i.e., hydrogen bonding interactions, all play pivotal roles in the incorporation of analogues into oligomers and helical structures. This thesis outlines computational investigations of these properties for a variety of isomorphic nucleoside analogues.

Initial work in this thesis (Chapter 2) involves the evaluation of the photophysical properties of thieno-modified analogues with time-dependent density functional

theory (TD-DFT). These preliminary investigations shed light on the energetics of CI pathways. Meanwhile, the presence of stable tautomers confirmed with computational studies (Chapter 3) helps explain the unique spectral features observed in experimental results. Further investigations on base pairing interactions (Chapter 5) of these analogues illustrate the structural similarities to their natural counterparts.

As the analogues studied in this work have potential utility in bioimaging, it is important that they be biocompatible for *in vivo* applications. Light penetration into biological media is very poor at short wavelengths due to increased absorption and scattering. As a result, multi-photon absorption techniques can be utilized to overcome the disadvantages of one-photon absorption. Multi-photon excitation reduces out-of-focus photobleaching, ensures localization of the incident light, and absorption occurs at higher wavelengths which allows for deeper penetration into biological tissue. The utility of analogues for two-photon fluorescence spectroscopy is analyzed by computation of the two-photon absorption (TPA) cross sections in Chapter 4. Studies on the conformational flexibility of select analogues uncover differences in TPA properties that could account for experimental discrepancies.

The ultimate goal of the work done in this thesis is the utilization of the knowledge obtained from computational studies in the design of new analogues with the desired physical and spectral properties. While the aim is to provide computational insight for advances in the development of tuneable isomorphous nucleoside analogues, it is equally important to elucidate the reasons behind undesired properties of analogues. This is a vibrant field of research which requires collaboration between synthetic chemists, spectroscopists, biologists, as well as theoretical chemists.

# Preface

A version of chapter 2 is published as M. Gedik and A. Brown, *J. Photochem. Photobiol. A*, 2013, **259**, 25-32. I ran all the computations, analyzed the data and was the primary author of the manuscript. A. Brown contributed scientifically and editorially to the paper.

Chapter 3 is adapted from work done in collaboration with undergraduate student Cat-Tuong Huynh. I performed the data analysis as well as the manuscript composition. Cat-Tuong Huynh assisted with the electronic structure computations, data collection and wrote the abstract for submission to *Chem Phys. Chem.* (paper being revised after reviewers' comments). Manuscript edits were done by A. Brown, the supervisory author.

A version of Chapter 4 has been prepared for submission to *Phys. Chem. Chem. Phys.*. I ran all the computations, analyzed the data and was the primary author of the manuscript. A. Brown contributed scientifically and editorially to the manuscript.

I performed all computations and data analysis for Chapter 5, and was the sole author. A. Brown contributed scientifically and editorially to this chapter.

Not included in this thesis is a book review published as M. Alaraby Salem, Melis Gedik and Alex Brown, "Two-photon Absorption: Theory and Applications For Biological Imaging" *Handbook of Computational Chemistry*, Ed. Jerzy Leszczynski. Springer Netherlands, 2016, 1-19. I wrote the section on nucleic acid bases and contributed editorially to the entire manuscript. All other sections were written by M. Alaraby Salem.

I dedicate this thesis to my mother, who has given me the strength to forge my own path and to all those who are willing to see in me an unquenchable thirst for knowledge.

# Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor, Professor Alex Brown, for his encouragement and support throughout my Ph.D. Thank you for inspiring me to become a computational chemist. Your enthusiasm for this field and passion for teaching (which I first experienced in CHEM 282), is the reason I am here today.

To Professor Mariusz Klobukowski: thank you for making computational chemistry a truly inspiring adventure and for our random, impromptu discussions.

To Professor Glen Loppnow: the knowledge I acquired in your course has truly been invaluable. I have learned so much about teaching (and learning) from you.

I would like to extend my thanks to all members of the Brown Group whom I had the pleasure of working with: Ryan, Stephanie, Ekadashi, Shuai, M. Alaraby, Mohammad. A special thank you to Yusuf, for his expert guidance on writing my thesis. I would also like to thank Cassandra, Amelia, and Meagan for their guidance and support throughout my academic career.

I am indebted to the undergraduate students who I mentored (Evan, Lisa, Christian, Cat-Tuong, and Tong) and taught over my career. Thank you for teaching me that I have as much to learn from young bright minds as I have knowledge to pass on.

I wish to thank my amazing mother, Sena Kalay. Thank you for supporting me throughout my journey, I would have never been where I am standing now without your encouragement. A special mention to my dearest friend, Dilara thank you for emotional support and comraderie, you are my family and I will forever be grateful for you. I also wish to thank all my extended family (especially Lovona) and friends

who have supported me along the way. There are too many of you to list, but you all know you are.

To my spouse and personal hero, Jordan: I am so thankful for you. Thank you for being my eternal cheerleader, for our long daily sit-reps and especially for helping me find the humour in even the most difficult circumstances.

# Table of Contents

<b>1 Chapter 1</b>	<b>1</b>
1.1 Theory of Fluorescence . . . . .	2
1.1.1 Absorption, Fluorescence and Competing Processes . . . . .	2
1.1.2 Parameters Affecting Fluorescence . . . . .	5
1.2 Fluorescence Spectroscopy . . . . .	6
1.2.1 Multi-photon Absorption . . . . .	7
1.3 Nucleic Acids - DNA and RNA . . . . .	8
1.3.1 Isomorphous and Emissive Nucleoside Analogues . . . . .	12
1.4 Theoretical Investigations . . . . .	14
1.5 Scope of Thesis . . . . .	16
<b>2 Chapter 2: Computational Study of the Excited State Properties of Modified RNA Nucleobases</b>	<b>18</b>
2.1 Introduction . . . . .	19
2.2 Computational Methods . . . . .	21
2.3 Results and Discussion . . . . .	22
2.3.1 Vertical Excitation Energies . . . . .	23
2.3.2 Emission Energies . . . . .	33
2.4 Conclusions . . . . .	34
<b>3 Chapter 3: Computational Insights Into Stable Tautomers of Emissive Thiophene-Derived RNA Nucleobases</b>	<b>35</b>
3.1 Introduction . . . . .	36
3.2 Computational Methods . . . . .	37
3.3 Results and Discussion . . . . .	40
3.3.1 <sup>th</sup> Guanine . . . . .	43
3.3.2 <sup>th</sup> Cytosine . . . . .	45
3.3.3 <sup>th</sup> Adenine . . . . .	46
3.3.4 <sup>th</sup> Uracil . . . . .	47
3.4 Conclusions . . . . .	47

<b>4</b>	<b>Chapter 4: Two-Photon Emissive Nucleobase Analogues: A Computational Investigation of Structural Effects on Photophysical Properties</b>	<b>48</b>
4.1	Introduction . . . . .	49
4.2	Computational Methods . . . . .	50
4.2.1	Computation of Two-Photon Absorption Cross Sections . . . . .	51
4.3	Results and Discussion . . . . .	53
4.3.1	Photophysical Properties . . . . .	53
4.4	Conclusions . . . . .	60
<b>5</b>	<b>Chapter 5: Base-pairing Interactions of Thiophene-Derived RNA Nucleobases</b>	<b>63</b>
5.1	Introduction . . . . .	64
5.1.1	Basis Set Superposition Error . . . . .	65
5.1.2	Interaction Energies . . . . .	66
5.2	Computational Methods . . . . .	68
5.3	Results and Discussion . . . . .	70
5.3.1	Structural Characteristics of the Watson-Crick Base pairs . . . . .	70
5.3.2	Binding Energies . . . . .	71
5.3.3	AIM Results . . . . .	73
5.3.4	Photophysical Properties of the Modified Nucleobase Pairs . . . . .	76
5.4	Conclusions . . . . .	81
<b>6</b>	<b>Conclusions</b>	<b>82</b>
6.1	Summary of Thesis Research . . . . .	83
6.2	Future Perspectives . . . . .	86
	<b>Bibliography</b>	<b>88</b>
<b>A</b>	<b>Appendix to Chapter 2</b>	<b>102</b>
A1	Vertical Excitation Energies . . . . .	103
A1.1	$\Lambda$ Diagnostic Parameter . . . . .	104
A2	Coordinates . . . . .	105
A2.1	Ground State Geometries of the Modified Nucleobases . . . . .	105
A2.2	Excited State Geometries of the Modified Nucleobases . . . . .	112
<b>B</b>	<b>Appendix to Chapter 3</b>	<b>118</b>
B1	Thermochemistry . . . . .	119
B1.1	CCSD(T) Energies . . . . .	119
B2	IR Spectra . . . . .	120
B3	Vertical Excitation Energies of Low-Lying Singlet States . . . . .	127
B4	Ground State Geometries of the Modified Nucleobases and their Tautomers . . . . .	136

B4.1	Gas-Phase . . . . .	136
B4.2	Water (PCM) . . . . .	140
B5	Ground State Geometries of the Modified Nucleosides and their Tautomers . . . . .	144
B5.1	Gas Phase . . . . .	144
B5.2	Water . . . . .	151
<b>C</b>	<b>Appendix to Chapter 4</b>	<b>158</b>
C1	Nucleoside Structures . . . . .	159
C2	Photophysical Properties . . . . .	161
C2.1	Charge Transfer Character of Low-Lying States . . . . .	161
C2.2	Two-Photon Absorption Properties for $S_1$ . . . . .	171
C3	Two-photon Absorption Properties for all Low-Lying States . . . . .	173
C4	Coordinates . . . . .	179
C4.1	Nucleobases . . . . .	179
C4.2	Nucleosides . . . . .	183
<b>D</b>	<b>Appendix to Chapter 5</b>	<b>189</b>
D1	Binding Energies . . . . .	190
D2	Intermolecular Distances . . . . .	192
D3	Excitation Properties of the Modified Nucleobase Pairs . . . . .	194
D4	Coordinates . . . . .	196
D4.1	Base Pairs . . . . .	196
D4.2	Bases . . . . .	220

# List of Tables

1.1	Radiative and non-radiative relaxation pathways and their relative timescales (Lifetimes) . . . . .	4
1.2	Photophysical properties of the natural nucleobases and nucleosides . . . . .	11
2.1	Vertical excitation energies (in eV) of the lowest lying singlet state, $S_1$ of the modified nucleobases at $S_0$ equilibrium geometry. Computed with TD-DFT using the specified functional and the 6-311++G(2df,2p) basis set. Oscillator strengths are given in parentheses. . . . .	24
2.2	Vertical excitation energies (in eV) and corresponding oscillator strengths (given in parentheses) of the modified nucleobases computed using TD-DFT with the specified functional and the 6-311++g(2df,2p) basis set, compared to the natural nucleobases in water. . . . .	26
2.3	Emission energies (in eV) of the modified nucleobases at the $S_1$ minimum. All computations utilized TD-DFT with the specified functional and the 6-311++G(2df,2p) basis set. . . . .	33
3.1	Relative electronic, $\Delta E$ , and free energies, $\Delta G$ , (in kcal/mol) for modified nucleobase (NB) and nucleoside (NS) tautomers in the gas phase and water(PCM). <sup>a</sup> Values are given relative to the canonical tautomer. . . . .	41
3.2	Vertical excitation energies (in eV) and corresponding oscillator strengths to $S_1$ of modified nucleobase tautomers <sup>a</sup> computed at the PBE0/6-311++G(2df,2p) level of theory in water. <sup>b</sup> Previous theoretical results also included. <sup>c</sup> . . . . .	43
4.1	Excitation energies (in eV), oscillator strengths ( $f$ ) and TPA cross sections ( $\sigma^{TPA}$ in GM) of the modified nucleobases and nucleosides computed with B3LYP/6-31++G(d,p) in water (PCM) . . . . .	55
4.2	Excitation energies (in eV), oscillator strengths ( $f$ ), $\Lambda$ diagnostic <sup>118</sup> and TPA cross sections ( $\sigma^{TPA}$ in GM) of selected singlet states of <i>th</i> 5-6azaU, <i>th</i> 5-2U, 5-furan2U and 7-amino nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water(PCM). Also provided is the dominant character of the transition. . . . .	57

4.3	Maximum TPA cross sections obtained from rotation of $\theta$ and the corresponding relative energies ( $\Delta E$ ), excitation energies ( $\omega_0$ ) and oscillator strengths ( $f$ ). . . . .	59
5.1	Intermolecular distances (in Å) of the natural and modified nucleobases calculated with $\omega$ B97xD/6-31++G(d,p) in the gas phase. . . . .	71
5.2	BSSE and ZPE corrected binding energies in (kcal/mol) of the modified base pairs computed with B3LYP, M06-2X, (+D3 corrections) and $\omega$ B97XD and the 6-31++G(d,p) basis set in the gas phase. . . . .	72
5.3	Binding energies in (kcal/mol) of the modified base pairs calculated with $\omega$ B97XD/6-31++G(d,p) in gas phase . . . . .	73
5.4	Hydrogen bond critical points (HBCPs), electron density ( $\rho_{BCP}$ ), Laplacian ( $\nabla^2\rho_{BCP}$ ) and the total electron energy density ( $H_{BCP}$ ) determined with QTAIM . . . . .	75
5.5	Vertical excitation energies of low-lying states of the modified base pairs calculated with B3LYP/6-31++G(d,p) in gas phase and water . . . . .	78
5.6	Vertical excitation energies and oscillator strengths of select states of the bases and base pairs calculated with B3LYP/6-31++G(d,p) in water . . . . .	79
A1	Vertical excitation energies of the modified nucleobases computed with TD-CAM-B3LYP/6-311++G(2df,2p) in dioxane. . . . .	103
A2	Vertical excitation energies of the modified nucleobases computed with TD-PBE0/6-311++G(2df,2p) in dioxane. . . . .	103
A3	Vertical excitation energies of the modified nucleobases computed with TD-B3LYP/6-311++G(2df,2p) in dioxane. . . . .	104
A4	Vertical excitation energies and $\Lambda$ diagnostic of the modified nucleobases computed with GAMESS . . . . .	104
B1	Relative thermodynamic parameters (kcal/mol) for <sup>th</sup> Guanine nucleobase tautomers in in the gas phasegas phase and water. Relative energies were computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	119
B2	Relative ground state energies (kcal/mol) for the abundant tautomers of the modified nucleobases in the gas phase and solution computed at CCSD(T)/cc-pVDZ level of theory in comparison with the B3LYP/6-311++G(2df,2p) level of theory. . . . .	119
B3	Vertical excitation energies of the singlet state $S_1$ of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	127
B4	Vertical excitation energies of the singlet state $S_2$ of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	127

B5	Vertical excitation energies of the singlet state $S_3$ of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	128
B6	Vertical excitation energies of the singlet state $S_1$ of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	128
B7	Vertical excitation energies of the singlet state $S_2$ of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	129
B8	Vertical excitation energies of the singlet state $S_3$ of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory. . . . .	129
B9	Vertical excitation energies of the singlet state $S_1$ of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	130
B10	Vertical excitation energies of the singlet state $S_2$ of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	130
B11	Vertical excitation energies of the singlet state $S_3$ of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	131
B12	Vertical excitation energies of the singlet state $S_1$ of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	131
B13	Vertical excitation energies of the singlet state $S_2$ of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	132
B14	Vertical excitation energies of the singlet state $S_3$ of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory. . . . .	132
B15	Vertical excitation energies of the singlet state $S_1$ of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	133
B16	Vertical excitation energies of the singlet state $S_2$ of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	133
B17	Vertical excitation energies of the singlet state $S_3$ of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	134
B18	Vertical excitation energies of the singlet state $S_1$ of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	134

B19	Vertical excitation energies of the singlet state $S_2$ of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	135
B20	Vertical excitation energies of the singlet state $S_3$ of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory. . . . .	135
C1	Excitation energies, oscillator strengths ( $f$ ), $\Lambda$ diagnostic and TPA cross sections ( $\sigma^{TPA}$ in GM) of low-lying singlet states of the thio-phenene nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water (PCM). Also provided is the dominant character of the transition. . . . .	161
C2	Excitation energies, oscillator strengths ( $f$ ), $\Lambda$ diagnostic and TPA cross sections ( $\sigma^{TPA}$ in GM) of low-lying singlet states of the isothiazole nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water (PCM). Also provided is the dominant character of the transition. . . . .	162
C3	Two-photon absorption properties of the first excited state, $S_1$ for the modified nucleobases computed with B3LYP/6-31++g(d,p) in water (PCM) . . . . .	171
C4	Two-photon absorption properties of the first excited state, $S_1$ for the modified nucleosides computed with B3LYP/6-31++g(d,p) in water (PCM) . . . . .	171
C5	Two-photon absorption properties of the first excited state, $S_1$ for the modified nucleobases computed with CAM-B3LYP/6-31++g(d,p) in water (PCM) . . . . .	172
C6	Two-photon absorption properties of the first excited state, $S_1$ for the modified nucleosides computed with CAM-B3LYP/6-31++g(d,p) in water (PCM) . . . . .	172
C7	Two-photon absorption properties of the low-lying states of $^{th}5-6azaU$ , $^{th}5-2U$ , 5-furan2U and 7-amino computed with B3LYP/6-31++g(d,p) in water (PCM) . . . . .	173
C8	Two-photon absorption properties of the low-lying states of $^{th}A$ , $^{th}C$ , $^{th}G$ and $^{th}U$ computed with B3LYP/6-31++g(d,p) in water (PCM) . . . . .	174
C9	Two-photon absorption properties of the low-lying states of $^{tz}A$ , $^{tz}C$ , $^{tz}G$ and $^{tz}U$ computed with B3LYP/6-31++g(d,p) in water (PCM) . . . . .	175
C10	TPA properties of the rotation of $^{th}5-6azaU$ about $\theta$ . . . . .	176
C11	TPA properties of the rotation of $^{th}5-2U$ about $\theta$ . . . . .	177
C12	TPA properties of the rotation of 5-furan2U about $\theta$ . . . . .	178
D1	Binding energies in (kcal/mol) of the modified base pairs computed with B3LYP/6-31++G(d,p) in the gas phase. . . . .	190
D2	Binding energies in (kcal/mol) of the modified base pairs computed with dispersion B3LYP-D3/6-31++G(d,p) in the gas phase . . . . .	190

D3	Binding energies in (kcal/mol) of the modified base pairs computed with M06-2X/6-31++G(d,p) in the gas phase . . . . .	191
D4	Binding energies in (kcal/mol) of the modified base pairs computed with dispersion M062X-D3/6-31++G(d,p) in the gas phase . . . . .	191
D5	Binding energies in (kcal/mol) of the modified base pairs computed with $\omega$ B97XD/6-31++G(d,p) in the gas phase . . . . .	191
D6	Intermolecular distances of the natural and modified nucleobases computed with B3LYP/6-31++G(d,p) . . . . .	192
D7	Intermolecular distances of the natural and modified nucleobases computed with B3LYP-D3/6-31++G(d,p) . . . . .	192
D8	Intermolecular distances of the natural and modified nucleobases computed with M062X/6-31++G(d,p) . . . . .	193
D9	Intermolecular distances of the natural and modified base pairs computed with M062X-D3/6-31++G(d,p) . . . . .	193
D10	Vertical Excitation energies of low-lying states of the modified base pairs computed with B3LYP/6-31++G(d,p) in gas phase . . . . .	194
D11	Vertical excitation energies of low-lying states of the modified base pairs computed with B3LYP/6-31++G(d,p) in water . . . . .	195

# List of Figures

1.1	Jablonski diagram depicting radiative and non-radiative relaxation processes. . . . .	3
1.2	Schematic of degenerate and non-degenerate two-photon absorption. . . . .	8
1.3	Structures of the canonical building blocks of nucleic acids. . . . .	9
1.4	Ball and stick model of a RNA duplex, side and top view . . . . .	10
1.5	Structure of 2-aminopurine(2-AP). R = ribose or deoxyribose group. . . . .	12
2.1	Ground State Structures of the Modified Nucleobases . . . . .	22
2.2	Valence Molecular Orbitals of Modified Nucleobases . . . . .	27
2.3	Absorption Spectra of Modified Nucleobases in Water . . . . .	29
2.4	Absorption Spectra of Modified Nucleobases in Dioxane . . . . .	30
3.1	Structures of the Thiophene Nucleobases and their Major Tautomers . . . . .	39
3.2	Computed UV-Vis Absorption Spectra of <sup>th</sup> Guanine and <sup>th</sup> Cytosine . . . . .	46
4.1	Structures of the Two-Photon Emissive Nucleobase Analogues . . . . .	50
4.2	Rotation around the twist angle, $\theta$ , of <sup>th</sup> 5-6azaU . . . . .	58
4.3	TPA cross sections to S <sub>1</sub> as a function of rotation around the twist angle, $\theta$ , for <sup>th</sup> 5-6azaU, <sup>th</sup> 5-2U and 5-furan2U . . . . .	61
4.4	TPA cross sections, TPA strengths ( $\delta$ ), and oscillator strengths of S <sub>1</sub> as a function of rotation of the twist angle, $\theta$ . . . . .	62
5.1	Representation of the basis set superposition error for the <sup>th</sup> G- <sup>th</sup> C base pair . . . . .	66
5.2	Watson-Crick base pairing of the native and modified RNA nucleobases . . . . .	69
5.3	Computed absorption spectra of the modified base pairs in gas phase and water . . . . .	80
B1	Computed IR spectra of the <sup>th</sup> A tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	120
B2	Computed IR spectra in the OH region of the <sup>th</sup> A tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	121
B3	Computed IR spectra of the <sup>th</sup> C tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	122

B4	Computed IR spectra in the OH region of the <sup>th</sup> C tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	123
B5	Computed IR spectra of the <sup>th</sup> G tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	124
B6	Computed IR spectra of the <sup>th</sup> G enol tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	125
B7	Computed IR spectra in the NH region of the <sup>th</sup> G enol tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM). . . . .	126
C1	Structures of the modified nucleoside and ribonucleosides. . . . .	159
C2	Structures of the modified thiophene nucleosides. . . . .	160
C3	Structures of the modified isothiazole nucleosides. . . . .	160
C4	Valence molecular orbitals of <sup>th</sup> 5-2U computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	163
C5	Valence molecular orbitals of <sup>th</sup> 5-6azaU computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	164
C6	Valence molecular orbitals of 5-furan2U computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	165
C7	Valence molecular orbitals of 7-amino computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	166
C8	Valence molecular orbitals of <sup>tz</sup> A computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	167
C9	Valence molecular orbitals of <sup>tz</sup> C computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	168
C10	Valence molecular orbitals of <sup>tz</sup> G computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	169
C11	Valence molecular orbitals of <sup>tz</sup> U computed at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. . . . .	170

# List of Abbreviations and Acronyms

2-AP	2-aminopurine
5-furan2U	5-(furan-2-yl)-2'-deoxyuridine
6-31++G(d,p)	Double-zeta (split valence) Pople basis set with d and p polarization functions on heavy and light atoms, respectively and diffuse s and p functions
6-311++G(2df,2p)	Triple-zeta (split valence) Pople basis set with improved polarization, and diffuse s and p functions
6-MAP	4-amino-6-methyl- 8-(2'-deoxy- $\beta$ -D-ribofuranosyl)-7(8H)-pteridone
6-MI	6-methylisoxanthopterin
7-amino	7-amino-1-ribose-quinazoline-2,4(1H,3H)-dione
A	Adenine
AIM	Atoms in molecules
B3LYP	Becke, three-parameter, Lee-Yang-Parr hybrid functional
BCP	Bond critical point
BSIE	Basis set incompleteness errors
BSSE	Basis set superposition error
C	Cytosine
CAM-B3LYP	Coulomb-attenuating method with the B3LYP hybrid functional
CASSCF	Complete active space self consistent field
CASPT2	Complete active space (second order) perturbation theory
CI	Conical intersection

CI	Configuration interaction
CC	Coupled cluster
CC2	Second order approximate coupled cluster singles and doubles
CCS	Coupled cluster, singles
CCD	Coupled cluster, doubles
CCSD	Coupled cluster singles and doubles
CCSD(T)	Coupled cluster singles and doubles and perturbative triples
cc-pVDZ	Correlation-consistent valence double-zeta basis set
CD	Circular dichroism
CP	Counterpoise
CT	Charge transfer
D3	Grimme's empirical dispersion model
DNA	Deoxyribonucleic acid
DFT	Density functional theory
DFT-B	Density functional tight binding
EOM	Equation of motion
ES	Excited state
eV	Electron volt
FC	Franck-Condon
FRET	Förster Resonance energy transfer
fs	Femtoseconds
FWHM	Full width at half maximum
G	Guanine
GM	Göppert-Mayer
GS	Ground state
HBCP	Hydrogen bond critical point
HF	Hartree-Fock
HFX	Hartree-Fock exchange
HOMO	Highest occupied molecular orbital
HWHM	Half width at half maximum

IC	Internal conversion
ICT	Internal charge transfer
ISC	Intersystem crossing
IEF-PCM	Integral equation formalism variant of PCM
LUMO	Lowest unoccupied molecular orbital
MAD	Mean absolute difference
MBPT	Many-body perturbation theory
MD	Molecular dynamics
M06-2X	Minnesota 2006 density functional with 54% Hartree-Fock exchange
MP	Møller-Plesset
NIR	Near-infrared
NBO	Natural bond orbitals
ns	Nanoseconds
OPA	One-photon absorption
OPM	One-photon microscopy
OS	Oscillator strength
PBE0	Hybrid Perdew, Burke, and Ernzerhof exchange-correlation functional
PCM	Polarizable continuum model
PES	Potential energy surface
ps	Picoseconds
QM/MM	Quantum mechanics/Molecular mechanics
RI	Resolution of identity
RI-CC2	Resolution of identity - second order approximate coupled cluster
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SOC	Spin-orbit coupling
SOS	Sum over states
SSFS	Steady-state fluorescence spectroscopy
T	Thymine
th	thieno[3,4-d]

TPFM	Two-photon fluorescence microscopy
TRFS	Time-resolved fluorescence spectroscopy
tz	isothiazolo[4,3-d]
<i>th</i> 5-2U	5-(thiophen-2-yl)-2'-deoxyuridine
<i>th</i> 5-6azaU	5-(thiophen-2-yl)-6-aza-uridine
TD	Time-dependent
TD-DFT	Time-dependent density functional theory
TPA	Two-photon absorption
U	Uracil
VR	Vibrational relaxation
$\omega$ B97xD	Becke's omega separated 1997 hybrid functional with short-range HFX and dispersion.
XC	Exchange correlation
ZPE	Zero-point energy
ZPVE	Zero-point vibrational energy

# List of Symbols

$\sigma^{\text{TPA}}$	Two-photon absorption cross section
$\delta^{\text{OPA}}$	One-photon absorption transition moment
$\delta^{\text{TPA}}$	Two-photon absorption transition moment
$\langle \delta^{\text{TPA}} \rangle$	Rotationally averaged two-photon absorption strength
$\omega$	(photon) Energy
$\mu$	Dipole moment
$a_0$	Bohr radius
$\alpha$	Fine structure constant
$c$	Speed of light
$\Gamma$	Lifetime broadening
$S_{ab}$	Element of the two-photon transition matrix
$S_n$	$n^{\text{th}}$ singlet electronic level (state)
$P$	Induced polarization
$\epsilon$	Molar absorptivity (Molar extinction coefficient)
$\hbar$	Reduced Planck constant
$\lambda$	Wavelength
$\Lambda$	Lambda diagnostic parameter (degree of spatial overlap)
$f$	Oscillator strength
$\Phi$	Quantum yield
$\tau_0$	Excited state lifetime
$T_1$	First excited triplet state
$\nu$	Frequency

# Chapter 1

## Introduction

*This dark brightness that falls from the stars.*

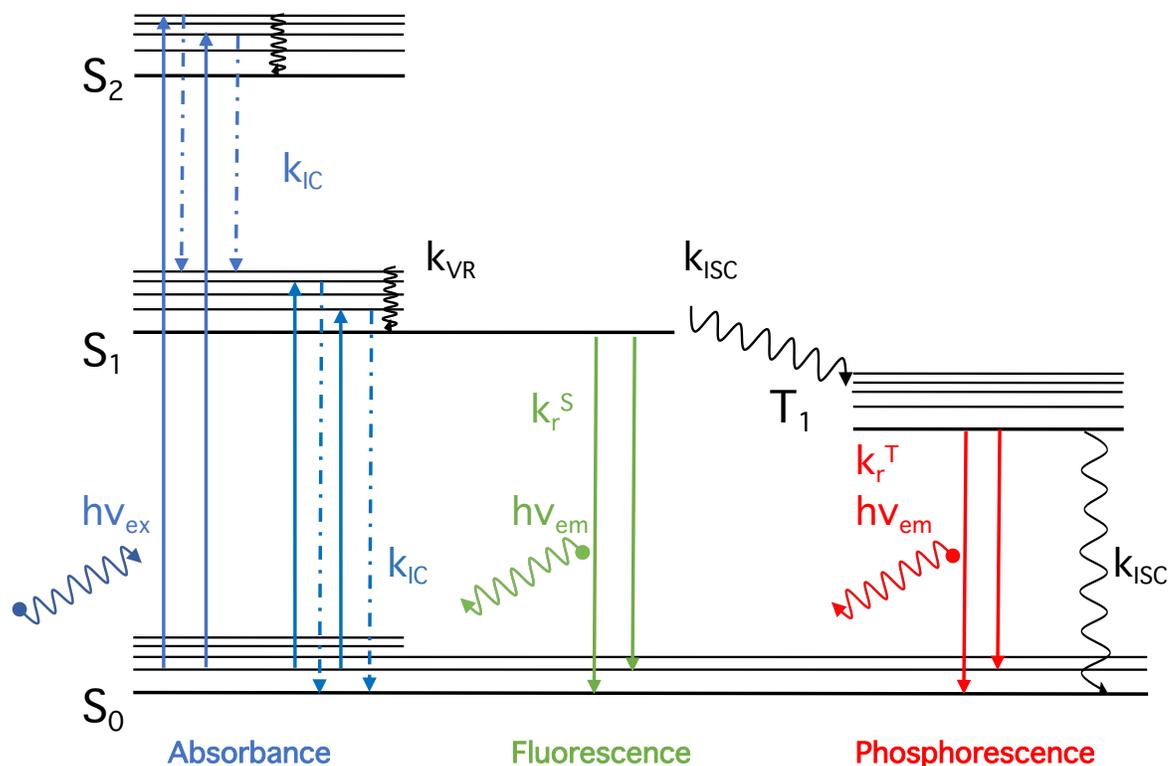
– Pierre Corneille

## 1.1 Theory of Fluorescence

Fluorescence is a type of photoluminescence that is generated by the absorption and subsequent emission of a photon (or photons).<sup>1</sup> Since this thesis encompasses the study of the photophysical properties of small molecule fluorescent probes, a brief overview of fluorescence and fluorescence detection techniques is provided.

### 1.1.1 Absorption, Fluorescence and Competing Processes

Optical excitation of the chromophore (light absorbing region of a molecule) occurs with incident light with sufficient energy. The efficiency of this instantaneous absorption process is related to the molar absorptivity (or extinction coefficient),  $\epsilon$  of the molecule at the excitation wavelength. After excitation, the chromophore generally resides in one of the many vibrational levels of an excited singlet state ( $S_1$  in Figure 1.1). Most chromophores reside in their excited state (ES) for 0.5-20 nanoseconds,<sup>2</sup> although some chromophores such as the building blocks of DNA and RNA have much shorter lifetimes, e.g. tens of femtoseconds. This excited state lifetime,  $\tau_0$ , reflects the sum of radiative and non-radiative processes that the chromophore undergoes as it relaxes back to the ground state (GS). Upon excitation to higher singlet states, molecules relax through internal conversion (IC) to higher vibrational levels of  $S_1$  and the rate of this process ( $k_{IC}$ ) occurs on the picosecond-femtosecond timescale. Furthermore, molecules residing in the higher vibrational levels of  $S_1$  will rapidly fall to the lowest vibrational energy level by losing energy to other molecules through collisions. This non-radiative process is termed vibrational relaxation, and occurs much faster than internal conversion (also a non-radiative process).



**Figure 1.1:** *Jablonski diagram depicting radiative and non-radiative relaxation processes.*

Alternatively, the chromophore may be excited into higher singlet states by the absorption of a second photon which presents a possible pathway for photobleaching. Photobleaching is an irreversible reaction which results in a loss of fluorescence due to covalent modification to the structure of the chromophore. Quenching is a term broadly applied to any process which results in the loss of fluorescence intensity and typically occurs as a consequence of interactions/collisions of the chromophore with other molecules. The highly reactive nature of excited states is responsible for the occurrence of these processes. Some molecules will undergo radiative depopulation from  $S_1$  by spontaneous emission of a photon (fluorescence). As a consequence of the loss of energy through non-radiative processes, fluorescence emissions are shifted to lower energies, and thus longer wavelengths, compared to the absorption process. This phenomena is referred to as the Stoke's shift and is the basis for the high sensi-

tivity of fluorescence spectroscopy. For the context of this chapter, molecules which exhibit fluorescence will be referred to as fluorophores. Another de-excitation pathway is possible by intersystem crossing (ISC), which occurs due to the spin-flip of an excited electron of a molecule in  $S_1$ . For many chromophores, this transition to the triplet state is inefficient and spin-forbidden. The probability of ISC increases if the vibrational levels of the singlet and triplet states overlap. The presence of heavy atoms can substantially increase the transition probability of this process due to an increase in the magnitude of spin-orbit coupling (SOC). The spontaneous emission of a photon from  $T_1 \rightarrow S_0$  results in phosphorescence. As in the case of singlet states, triplet states can also be excited to higher electronic energy levels, and hence access photobleaching and quenching pathways. The relative timescales for the processes outlined above are tabulated in Table 1.1.

**Table 1.1:** *Radiative and non-radiative relaxation pathways and their relative timescales (Lifetimes)*

Process	Transition	Rate	Timescale ( $s^{-1}$ )
Internal Conversion (IC)	$S_n \rightarrow S_1$	$k_{IC}$	$10^{10} - 10^{14}$
Internal Conversion (IC)	$S_1 \rightarrow S_0$	$k_{IC}$	$10^6 - 10^7$
Vibrational Relaxation (VR)	$S_1(\nu=n) \rightarrow S_1(\nu=0)$	$k_{VR}$	$10^{10} - 10^{12}$
Singlet Absorption	$S_0 \rightarrow S_n$	$k_{ex}$	$10^{15}$
	$S_n \rightarrow S_m$		
Fluorescence	$S_1 \rightarrow S_0$	$k_r^S$	$10^7 - 10^9$
Intersystem Crossing (ISC)	$S_1 \rightarrow T_1$	$k_{ISC}$	$10^5 - 10^8$
	$S_n \longleftrightarrow T_n$		
Phosphorescence	$T_1 \rightarrow S_0$	$k_r^T$	$10^{-2} - 10^3$

\*Absorption, fluorescence and phosphorescence are radiative, whereas VR, IC, and ISC are non-radiative processes.

The measure of a molecule's fluorescence intensity is usually given by the quantum yield, which represents the fraction of fluorophores that relax by fluorescence. In the equation below,  $k_{nr}$  represents all possible non-radiative deactivation pathways,

including IC, ISC, and other quenching mechanisms.

$$\Phi = \frac{k_r}{k_r + \sum k_{nr}} = k_r \tau. \quad (1.1)$$

Ideally, the quantum yield of a fluorophore should be as close to unity for optimal fluorescence. Fluorophores such as fluorescein<sup>3</sup> and rhodamine<sup>4</sup> exhibit large absolute quantum yields of 0.95 and 0.94, respectively, and are among the most commonly used fluorescent probes. These highly fluorescent compounds are typically utilized as standards for (relative) quantum yield measurements of other fluorophores. For example, the quantum yields of tyrosine and tryptophan<sup>5</sup> measured against the quinine sulfate<sup>6</sup> standard are 0.14 and 0.13, respectively. Quantum yields of common fluorophores range from 0.1-1.0<sup>1,7</sup> and are very sensitive to experimental conditions such as pH, temperature and solvent.

### 1.1.2 Parameters Affecting Fluorescence

The fluorescence efficiency of a chromophore is highly dependent on its molecular structure. In general, conjugation of aromatic compounds shifts the excitation and emission energies to longer wavelengths (bathochromic shift) and consequently increase the emission intensity (hyperchromic shift). Electron donating groups such as -OH and -NH<sub>2</sub> also cause hyperchromicity. Another example of structural effects can be observed with heterocycles which contain oxygen, sulfur or nitrogen. These molecules tend to have relatively high quantum yields as a result of the orientation of the  $\pi$  electron system to the nonbonding orbitals of O, S or N.

In addition to their structure, the effect of environmental parameters such as pH, solvent polarity, viscosity and temperature on the chromophores is crucial to their relative fluorescence. Due to their high degrees of freedom and complex vibrational profiles, coupled with the interactions between solvent molecules, fluorophores exhibit a large number of energetically different transition probabilities of their relaxation to the ground state. Furthermore, owing to the different properties of ground and excited states, the dipole moment changes upon excitation and this also is dependent on solvent polarity. The effect of solvent polarity on the photophysical properties of

chromophores is known as solvatochromism. Solvatochromism of chromophores is also observed as a result of charge transfer (CT) transitions. Internal charge transfer (ICT) states can occur as a result of charge separation within a fluorophore after excitation; this causes an instantaneous change in the dipole moment of the molecule. This change is not in equilibrium with respect to the solvent molecules and is enhanced with increased solvent polarity. Consequently, a shift in the emission energy (bathochromic) due increasing solvent polarity is indicative of the charge transfer character of the excited state.

## 1.2 Fluorescence Spectroscopy

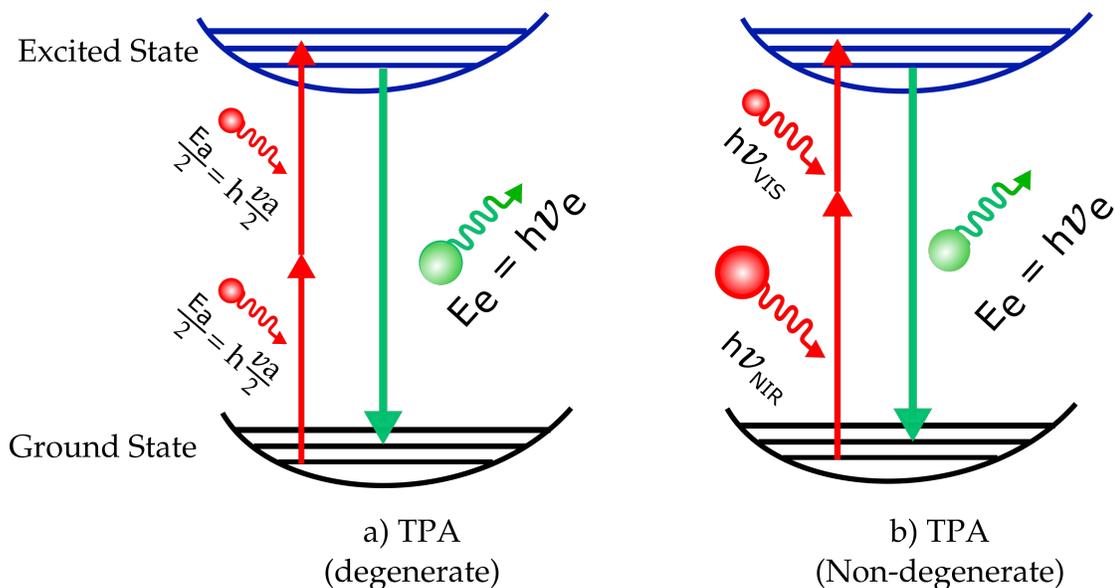
Fluorescence spectroscopy has long been a valuable tool for the study of biomolecules such as proteins and nucleic acids, and their interactions with their microenvironment. The sensitivity, variety in types of experimentation, and continuous instrument development are among many of the reasons why this methodology is widely utilized by researchers. Many advancements have been made in fluorescence spectroscopy - increased sensitivity, multi-photon excitation, automation and single-molecule detection (SMD).<sup>8</sup> Hence, some of these methods and their potential applications in the field of biotechnology is discussed herein.

Steady-state fluorescence spectroscopy (SSFS) is a method which produces an emission spectrum upon excitation of a fluorophore (typically at the  $\lambda_{max}$ ) with a light source providing a constant photon flow. At low concentrations, linear proportionality between the emission intensity and concentration of the chromophore is observed.<sup>2</sup> Although the emission intensity is inherent to the particular chromophore, it is sensitive to the microenvironment as discussed previously. Consequently, the effect of solvent polarity on the Stoke's shift can be measured with this method.<sup>9,10</sup> SSFS has been utilized in various studies investigating the effects of local polarity in proteins and DNA.<sup>11-13</sup> While SSFS gives an average emission profile of all excited fluorophores, time-resolved fluorescence spectroscopy (TRFS) is concentration independent and can provide insight into the decay pathways, or more generally,

ES dynamics, of a chromophore.<sup>14</sup> Specifically, TRFS measurements provide excited state lifetimes and can distinguish between collisional/dynamic (lifetime affected) and static quenching (lifetime unaffected). Furthermore, deconvolution of spectra can provide information about different decay pathways.<sup>15</sup> Another fluorescence technique involves the process of quenching, which is utilized in Förster Resonance Energy Transfer (FRET),<sup>16</sup> a method useful for determining distances within or between biomolecules. FRET has also been used in studies exploring folding and dynamics of RNA.<sup>17</sup> Fluorescence based imaging techniques both *in vivo*<sup>18,19</sup> and *ex vivo* have also made significant advancements.<sup>20</sup> These studies utilize near-infrared (NIR) probes,<sup>21</sup> as well as modified amino acids<sup>22</sup> and nucleosides.<sup>23</sup>

### 1.2.1 Multi-photon Absorption

Two-photon absorption involves the simultaneous absorption of two photons and was first characterized theoretically by Maria Göppert-Mayer in 1931.<sup>24</sup> After 30 years and with the invention of the laser, the first experiments were conducted for TPA.<sup>25</sup> TPA can occur with photons of the same frequency (degenerate) or with different frequencies (non-degenerate) as depicted in Figure 1.2. It is also important to note that one-photon and two-photon absorption have different selection rules, and hence TPA allowed transitions can be OPA forbidden and vice-versa. For example in centrosymmetric molecules, only  $g \leftrightarrow u$  OPA transitions are allowed, whereas this transition is a forbidden transition in TPA, while the  $g \leftrightarrow g$  and  $u \leftrightarrow u$  are allowed. Since TPA involves the interaction of two photons, the strength of the transition depends on the square of the incident light intensity. In contrast, one-photon absorption has a linear dependence on light intensity. This property of TPA has yielded a wide range of applications, particularly with advances in laser technology.



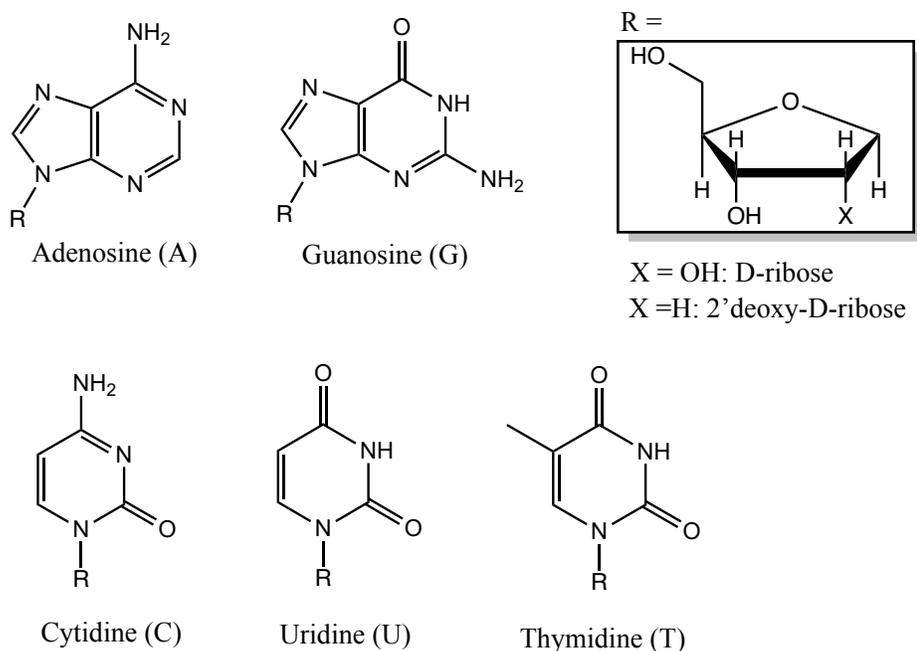
**Figure 1.2:** *Schematic of degenerate and non-degenerate two-photon absorption.*

Due to the nature of biological tissues, which strongly scatter light, high resolution imaging of deep tissue is very challenging. However, the quadratic dependence on the light intensity in TPA provides better focus and consequently, deep penetration into biological tissues.<sup>26</sup> In particular, two-photon induced fluorescence microscopy (TPFM)<sup>27</sup> has emerged as an essential tool for biological imaging. TPFM has significant advantages over OPM (one-photon microscopy), including higher spatial resolution, deeper tissue penetration, and even longer observation times in biophysical assays. To further facilitate the utilization of TPFM, there is a strong need for the development of two-photon probes. For a comprehensive review of TPA theory and experimental techniques, we refer the reader to a recent review.<sup>28</sup> The theoretical considerations for TPA as well as the computational approaches are extensively discussed in Chapter 4 (see Section 4.1).

### 1.3 Nucleic Acids - DNA and RNA

The native nucleic acids are biopolymers with the sugar moiety (D-ribose or D-deoxyribose) and phosphate groups acting as a backbone to the helical structure. Deoxyribonucleic acid (DNA)<sup>29-33</sup> is composed of two antiparallel stands of linked

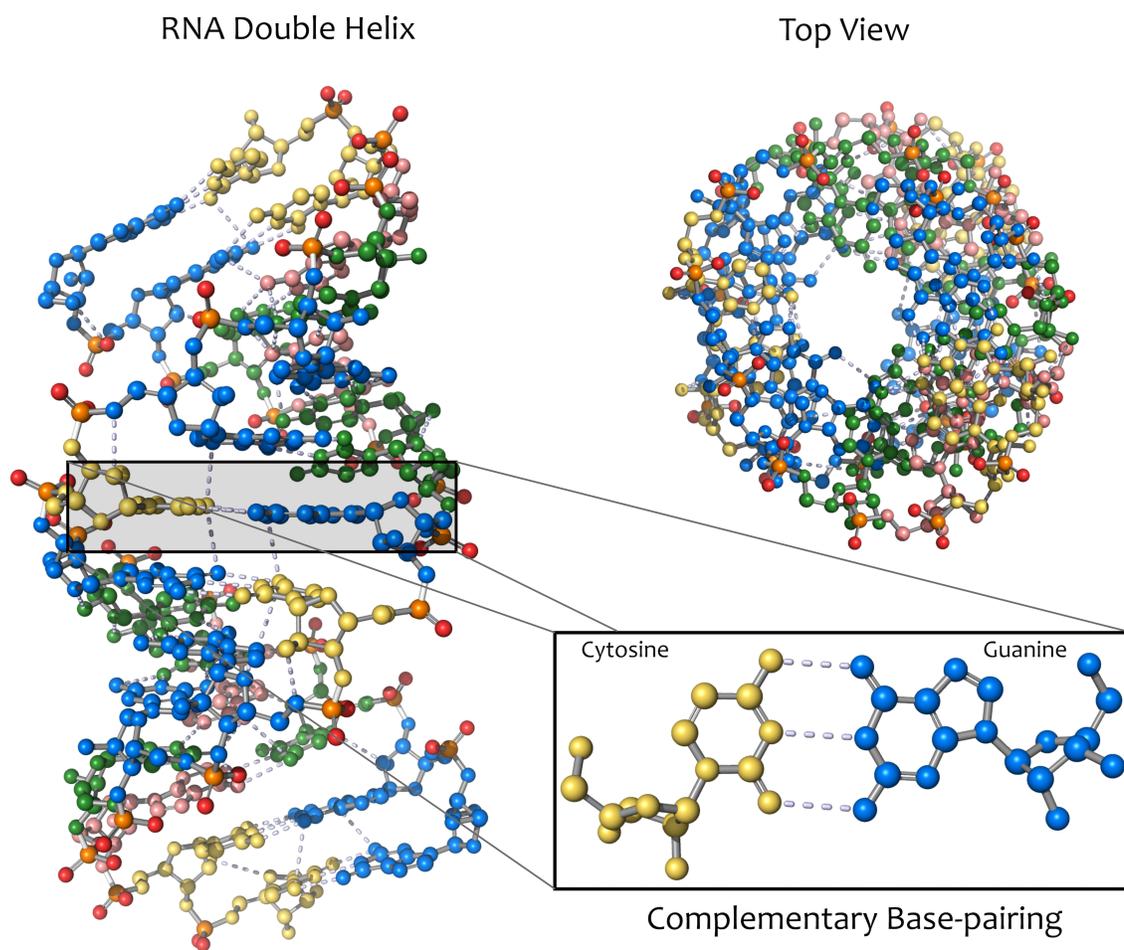
nucleotide units. Nucleotides are constructed from three subunits: a deoxyribose sugar group, a nitrogenous base (used interchangeably as nucleobase), and a phosphate group. The four main bases in DNA are adenine (A), cytosine (C), guanine (G), and thymine (T). A and T are denoted as pyrimidines, while G and C are known as purines and consist of a pyrimidine ring fused to an imidazole ring. The same nucleotides make up ribonucleic acid (RNA); however, the thymine base is replaced with the pyrimidine, uracil (U), and the sugar moiety is a ribose instead of deoxyribose. In addition to representing the nucleobases, the terms A, C, G, T and U will also refer to the nucleoside structures (nucleobase + sugar group). The majority of the work contained within this thesis relates to nucleobase or nucleoside structures (see Figure 1.3), and hence a clear distinction between the two forms is provided where appropriate throughout the thesis.



**Figure 1.3:** Structures of the canonical building blocks of nucleic acids.

Typically, RNA exists as a single helix structure, but has duplex (double helical) conformations as well. Figure 1.4 depicts a duplex comprised of 14 bases<sup>34</sup> connected by hydrogen bonds to their respective complementary base, referred to as a base pair. Complementary G-C bases form 3 H-bonds, while the A-T (or U) bases are

connected by 2 H-bonds. These are the favourable pairings and are referred to as Watson-Crick base pairs. Other types of pairings such as Hoogsteen, or wobble, base-pairs are possible, but are beyond the scope of this thesis. These alternative structures do not offer as much stability; however, allow for other conformations, such as a triple helix. The base-pairing interactions along with the electrostatic and hydrophobic interactions due to base stacking contribute to duplex stability. The degree of stabilization is sequence dependent as certain combinations of base pairs have more favourable interactions.



**Figure 1.4:** Ball and stick model of a RNA helix, taken from X-ray crystal structure of 14 basepair RNA duplex (PDB ID: 433D). Top and side views are shown along with complementary base-pairing of G-C. Nucleoside residues: A=pink, C=yellow, G=blue, and U=green.

Biological activity performed by nucleic acids involves significant structural changes due to interactions with regulatory proteins and enzymes. As a result, investigations of nucleic acid structure, dynamics and recognition can be achieved with fluorophores that can "probe" the local environment.<sup>35,36</sup> The majority of biomolecules and their constituents are not inherently fluorescent. RNA and DNA nucleobases (the chromophores of nucleic acids) are no exception, and exhibit short excited state lifetimes and low quantum yields ( $\Phi$ ). Consequently, artificial chromophores need to be introduced into nucleic acids for detection in fluorescence experiments. The non-emissive nature of nucleobases is due to their efficient non-radiative relaxation pathways. Specifically, the bases have access to conical intersections (CIs), which serve as the primary deactivation pathways.<sup>37-40</sup> CIs are crossing points on the potential energy surface where the energies of (two or more) states become degenerate. Furthermore, the native nucleobases absorb light in the ultraviolet (UV) range. The photophysical properties of the nucleosides as determined from circular dichroism(CD) and fluorescence experiments<sup>41-48</sup> are summarized in Table 1.2.

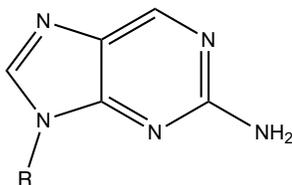
**Table 1.2:** *Photophysical properties of the natural nucleobases and nucleosides*

Code	Nucleobase	Nucleoside	$\lambda_{max}$ (nm)	$\lambda_{em}$ (nm)	$\tau$ (ps)	$\Phi_{NS}$ (x 10 <sup>-4</sup> )	$\Phi_{NB}$ (x 10 <sup>-4</sup> )
A	Adenine	Adenosine	260	310	0.53	0.5	2.6
C	Cytosine	Cytidine	271	324	0.76	0.7	0.8
G	Guanine	Guanosine	253	346	0.69	-	3.0
T	Thymine	Thymidine	267	327	0.70	1.0	1.0
U	Uracil	Uridine	259	309	0.25	-	0.5

Absorption energies and emission energies of nucleosides.<sup>44-47</sup>  
Quantum yields of nucleobases ( $\Phi_{NB}$ )<sup>42,43</sup> and of nucleosides ( $\Phi_{NS}$ ).<sup>45,47</sup> Lifetimes of nucleosides ( $\tau$ ) obtained by fluorescence up-conversion.<sup>48</sup>

The inherent lack of fluorescent properties of the natural nucleosides was overcome with the discovery of 2-aminopurine (2-AP).<sup>49</sup> 2-AP has been extensively utilized in

studies of DNA structure and dynamics<sup>50,51</sup> due its high sensitivity to its microenvironment. When incorporated into oligonucleotides and duplexes, 2-AP can exhibit several orders of magnitude reduction in fluorescence intensity. This characteristic is due to quenching by neighbouring bases (and stacking).<sup>52</sup> Unfortunately, 2-AP emits significantly below the visible range (370nm),<sup>53,54</sup> and thus has limited use. Since then, a plethora of synthetic emissive nucleoside analogues have been designed to be utilized as probes.<sup>2,55-59</sup> Interestingly, the absence of fluorescence from DNA and RNA subunits actually offers an advantage in some biophysical assays since fluorescent signals from emissive analogues can be observed without any competing background signals. Various types of fluorescent nucleobase analogues have been designed in recent years; these include pteridines, expanded or extended nucleobases, and isomorphous nucleobases. A comprehensive review by Sinkeldam et al.<sup>2</sup> outlines the properties of a large number of fluorescent analogues. This thesis focuses on the study of isomorphous nucleoside analogues and 2-AP is an example of such an analogue that has been widely investigated (see Figure 1.5).



**Figure 1.5:** *Structure of 2-aminopurine(2-AP). R = ribose or deoxyribose group.*

### 1.3.1 Isomorphous and Emissive Nucleoside Analogues

Many commercially available fluorophores have the desired emissive properties for the studies outlined in the previous sections. Unfortunately, these probes cause structural perturbations when incorporated into oligonucleotides or nucleic acid structures.<sup>60</sup> Fluorescent nucleoside analogues can be utilized as probes while offering structural similarity to the natural nucleosides. This property is particularly advantageous if the probe can demonstrate minimal perturbations to secondary structure; this is defined as isomorphousity.<sup>61</sup> However, designing effective probes for incorporation into RNA and DNA is particularly challenging as nucleic acid stability is extremely sensitive

to modification of its building blocks. This sensitivity presents a limited landscape for modifications. For example, structural changes may disrupt hydrogen bonding to complementary base pairs, or the presence of functional groups in certain positions (near minor or major grooves) of the nucleobase could disrupt helix formation and ligand binding. In contrast, the majority of modified amino acids can readily replace native residues without producing deleterious effects on protein folding and function.<sup>62,63</sup> In addition to retaining structural requirements, analogues must fulfill photophysical requirements for fluorescence. The desired properties of fluorescent analogues can be summarized as follows:

- Selective excitation and emission energies, preferably bathochromic shifted energies relative to natural nucleobases.
- Adequate brightness, as a result of large quantum yields for fluorescence and/or molar absorptivity.
- Minimal perturbation to secondary structure.
- Sensitivity to microenvironment.

Naturally occurring fluorescent biomolecules typically have an extended (aromatic)  $\pi$  system, and hence, this structural commonality is often used as a basis for the design of probes. Introducing (or extending) a  $\pi$ -system to the native nucleobase structure results in a bathochromic shift to the absorption maximum, allowing for a strong  $\pi$ - $\pi^*$  excitation, and ultimately yielding the desired fluorescent properties. This particular modification will be the focus for the isomorphous analogues that are studied within the context of this thesis. As discussed in Section 1.1.2, the parameters affecting fluorescence are numerous, and, thus investigation of the photophysical properties as a function of the environment is crucial for probe design. The aim of this thesis is to shed light into these properties with the use of computational methods.

## 1.4 Theoretical Investigations

The applications of computational methods have become a crucial part of the study of chemical phenomena largely due to the improvements in accuracy, reliability, efficiency, versatility and, to some extent, their ease of use in recent years.<sup>64</sup> Furthermore, computational approaches may provide insight into properties of a system that would otherwise be inaccessible or difficult to determine through experimental studies. The fundamental principle underlying theoretical/computational methodology for molecules and molecular complexes (for static problems) is to solve the electronic Schrödinger equation (within the Born-Oppenheimer approximation) to obtain properties of the system of interest; the present thesis does not discuss the variety of classical, semi-classical, or quantum mechanical dynamics approaches available. These static methods are generally referred to as electronic structure calculations.<sup>65</sup> To obtain accurate energies and other properties, sophisticated electronic structure methods usually need to be employed. Unfortunately, using more advanced methods, such as those with electron correlation and large basis sets, is computationally expensive. Consequently, a balance must be achieved between the computational "cost" and accuracy. Hartree-Fock (HF) Theory is considered the starting point of *ab initio* methods; however it only describes average electron-electron interactions, i.e., most electron correlation is ignored. *Ab initio* methods are based on first principles of quantum mechanics and do not reference experimental data. To increase the accuracy of computations, the correlation between electrons needs to be included. Methods which account for this correlation require a multi-determinant, i.e., an expression which describes the anti-symmetric wavefunction and satisfies the Pauli-exclusion principle. Electron correlation can be achieved with configuration interaction (CI), Many-body perturbation theory (MBPT) and coupled cluster (CC) methods. Configuration interaction uses a linear combination of determinants (called configurational state functions or CSFs) and goes through an iterative procedure to minimize the energy of the system (based on the variational principle). Essentially, CI allows the electrons to occupy virtual orbitals and gives a better description of their correlation.

The number of excited state determinants added determines the level of CI, and the contributions they make to electronic correlation energy are significant for obtaining accurate properties of the system. MBPT-based methods such as nth-order Møller-Plesset (MPn perturbation theory accounts for electron correlation by introducing a perturbation on the Hamiltonian operator. The perturbed wavefunction and resulting energy are expressed as a power series with respect to the parameter ( $\lambda$ ) which controls the extent of the perturbation. The level of theory is determined by the highest power of  $\lambda$  used in the expression, denoted by the n in MPn. Coupled cluster theory is another post-HF method which uses an infinite expansion of contributions from a given excitation level to the wavefunction. For both CI and CC, truncation must be done in terms of the inclusion of excited determinants, otherwise the computational cost becomes unfeasible for large systems. The level of the CC method is established by the excitation level of the cluster operator. The lowest level of approximation is referred to as CCS, CCD, followed by CCSD, etc. Another approach is to build on CCSD by adding the triples (third excitation level) contribution evaluated with perturbation theory. This hybrid method is denoted as CCSD(T).

An alternative to *ab initio* methods is density functional theory (DFT),<sup>66</sup> in which the electron correlation is modelled by a function of the electron density. In the Kohn-Sham approach,<sup>67</sup> the theory was further developed with the introduction of orbitals (as in *ab initio* methods). The calculated energy is a function of the electron density,  $\rho(\mathbf{r})$ , which itself is a function of the Kohn-Sham orbitals,  $\psi_i(\mathbf{r})$ . As a result of the dependence of the function on a function, the electron density is a "functional" and represented by

$$\rho(r) = \sum_i^N |\psi_i(r)|^2 \quad (1.2)$$

where N denotes the number of electrons of the system. What is intriguing about this methodology is that the exact functional that should be used to obtain the energy (from the density) is not known. In particular, the exact exchange-correlation component is unknown and approximate correlation functionals need to be used. Consequently, a wide variety of functionals have been developed and tested for DFT.

The choice of the functional to be used in computational studies is crucial as each functional has its own advantages and disadvantages. Furthermore, there is no systematic way to improve on the accuracy by the choice of the functional. The main advantage of using DFT is the significant improvement over HF with comparable computational efficiency. In this work, several functionals have been utilized and compared in their ability to reproduce experimental results. The functionals used in this work include B3LYP,<sup>68,69</sup> CAM-B3LYP,<sup>70</sup> PBE0,<sup>71-74</sup>  $\omega$ B97XD<sup>75,76</sup> and M06-2X.<sup>77</sup> All of these functionals are hybrid functionals, due to their inclusion of the exact exchange obtained from HF theory. The amount of HF exchange varies from functional to functional. Many of these functionals include empirical parameters in the description of exchange correlation (XC). For example, the M06-2X functional, which has 32 empirical parameters, shows improved performance with systems where dispersion is significant.

## 1.5 Scope of Thesis

This thesis will expand on the efforts that have been made to gain insight into structural and photophysical properties of modified nucleoside analogues by using computational approaches. Details of the theoretical methods used to obtain properties of the systems under investigation are provided in the introduction of each chapter. The basic concepts behind computational approaches have been summarized in the previous section; a comprehensive account of the underlying principles of these methods is beyond the scope of this thesis.

As discussed previously, the design of emissive and isomorphous analogues poses unique challenges due to sensitivity of modifications on the native structure of nucleic acids. Computational studies can provide more efficient means of determining the desired properties of these analogues. The majority of the work in this thesis focuses on the isolated nucleobase structure (chromophoric component); however, the inclusion of the sugar moiety on the properties has also been considered and analyzed. In Chapter 2, the excited state properties of nucleobases designed with a thiophene

modification are determined with TD-DFT computations. In this work, three hybrid functionals, B3LYP, CAM-B3LYP and PBE0, are tested in their ability to reproduce experimental data and we discover that intensities are not well reproduced by the commonly used B3LYP functional. Moreover, we conclude that modified analogues show selective excitation due their bathochromic shifts in absorption energies compared to their natural counterparts. The tautomerization of a class of emissive RNA analogues is investigated in Chapter 3. The findings suggest that minor stable tautomers of two of the modified nucleobases display appreciable stability in solution. In Chapter 4, TPA cross sections for 12 modified uridine analogues are determined by TD-DFT and yield excellent agreement with experimentally measured values, where available. Chapter 5 explores the base-pairing capabilities of the thiophene analogues studied in Chapter 2, relative to their natural nucleobase counterparts. We find that binding energies and hydrogen acceptor-donor distances are well reproduced with dispersion corrected functionals, such as  $\omega$ B97XD, and that there are only modest energetic and structural changes caused by introducing the emissive thiophene analogue in place of the corresponding natural nucleobase.

## Chapter 2

# Computational Study of the Excited State Properties of Modified RNA Nucleobases <sup>1</sup>

*We are all connected; To each other, biologically. To the earth, chemically. To the rest of the universe atomically.*

– Neil DeGrasse Tyson

---

<sup>1</sup>This chapter is adapted (with minor corrections) from the published paper: M. Gedik and A. Brown, *J. Photochem. Photobiol. A*, 2013, **259**, 25-32.

## 2.1 Introduction

Ribonucleic acid has an essential role in biological processes and has emerged as a target for the control of cell function and study of ligand interactions.<sup>78,79</sup> However, the nucleobase constituents of RNA, which are the primary chromophores for probing these processes, are non-emissive, have ultrafast excited state lifetimes and low quantum yields.<sup>39,80–82</sup>

The excited-state properties of the nucleobase constituents of RNA and DNA have been extensively studied by both theoretical and experimental approaches<sup>40,83–101</sup> (see Ref.<sup>39,80–82</sup> for recent theoretical and experimental reviews). The inherent non-emissive nature of the naturally occurring nucleobases, however has driven the investigation into emissive analogues to the nucleobases. Emissive analogues of nucleobases have significant implications for the field of biotechnology when utilized as molecular probes.<sup>2,57,102</sup> Design of synthetic probes requires several criteria in order to be successfully implemented in biophysical assays. In general, the modified analogues need to achieve isomorphism with respect to their natural counterparts and thus display minimal structural and functional perturbations upon modification.<sup>2,61</sup> While structural and functional perturbation is an inevitable consequence of such modifications, the goal is to minimize these perturbations and retain the structural requirements necessary for base pairing and helix formation in the nucleic acid. It is therefore no surprise that the majority of the synthetic analogues designed show strong resemblance to the natural nucleobases. One example of such an analogue is 2-aminopurine and it has been extensively utilized in studies of DNA conformational dynamics due its high quantum yield and visible emission.<sup>50,54,103–106</sup> The structure of 2-aminopurine is remarkably similar to that of adenine and only shows differences in the placement of nitrogens in the aromatic ring. While initial success has been achieved with this particular analogue, it emits significantly below the visible range (350 nm) and is susceptible to quenching by neighbouring bases when incorporated into the nucleic acid sequence.<sup>107,108</sup> Consequently, another criterion is important for successful probe design: the resistance to quenching in the local environment.

The non-emissive nature of the nucleobases is a consequence of non-radiative relaxation through singlet electronic states<sup>39,80-82</sup> and in lieu of this, it is highly valuable to have a theoretical studies to interpret and analyze the excited state properties of these types of systems. The electronic structure of the individual nucleobases is complicated due to the presence of multiple electronic transitions in the low energy region. Experimentally, the existence of the  $n\pi^*$  state arising from the lone pairs of heteroatoms is difficult to determine due to their weak intensities and forbidden character. These transitions are overlapped by stronger  $\pi\pi^*$  transitions in the optical spectrum and are suggested to play a role in the radiationless decay through conical intersections.<sup>39,82</sup> Consequently the analogues should exhibit an excited state minimum on the potential energy surface that renders conical intersections less energetically accessible. From a theoretical standpoint, this characteristic would achieve the desired properties that promote fluorescence. As a result, an investigation of the potential energy surfaces of these molecules is essential to provide the complete picture on the emissive properties of the analogues. While this is beyond the scope of the present study, it represents the next step in order to determine the fluorescent properties of the analogues. The accurate description on the excited states of this system will lay the groundwork for further studies.

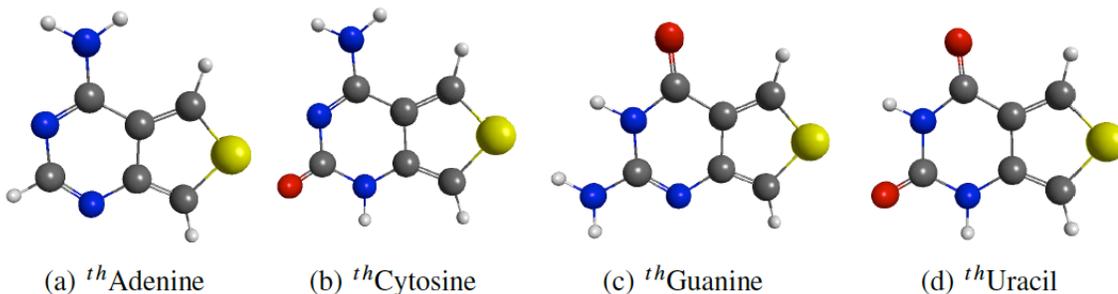
This computational study focuses on the absorptive and emissive properties of nucleobase analogues derived from thieno[3,4-d]-pyrimidine as the heterocyclic nucleus designed and characterized by Shin et al.<sup>109</sup> We limit this work to the isolated nucleobases and focus on an accurate photophysical description of their excited state properties. Since the computational cost scales with system size, analysis of ribonucleosides presents a challenge to study with the desired accuracy. The effect of the sugar on the emissive properties of nucleic acid analogues is suggested not to play a significant role<sup>82</sup> and thus the results obtained from this study are expected to reflect the desired properties of the system.

## 2.2 Computational Methods

The thiophene substituted nucleobase analogues ( $^{th}A$ ,  $^{th}C$ ,  $^{th}G$ , and  $^{th}U$ ) of RNA were characterized by density functional quantum chemical computations. These modified nucleobases in their relevant forms are depicted in Figure 2.1. Ground state geometries, vertical excitation energies and emission energies were determined using the CAM-B3LYP,<sup>70</sup> B3LYP,<sup>68,69</sup> and PBE0<sup>71-73</sup> hybrid and exchange correlation functionals with the 6-311++G(2df,2p) basis set<sup>110,111</sup> for all atoms. All computations were performed with the Gaussian 09 software package<sup>112</sup> employing the polarizable continuum model (PCM)<sup>113,114</sup> to account for solvation in water and dioxane, the two solvents in which experimental measurements were made.<sup>109</sup> Ground state geometries were computed using "Tight" convergence criteria, i.e., Maximum Force =  $1.5 \times 10^{-5}$  a.u., RMS Force =  $1.0 \times 10^{-5}$  a.u., Max displacement =  $6.0 \times 10^{-5}$ , and RMS displacement =  $4.0 \times 10^{-5}$ . The grid used for numerical integration in DFT is set to "Ultrafine", i.e., a pruned grid of 99 radial shells and 590 angular points per shell. The excited state geometries were determined using TD-DFT with equilibrium solvation implemented in Gaussian09.<sup>115-117</sup> Frequency analysis was performed for both ground and excited state equilibrium geometries at the same level of theory in order to verify that the optimized structures were true minima on the potential energy surface.

Ground state geometries were determined by sequential optimizations initially in the gas phase with modest basis size, then increasing basis set size and including solvation effects with the PCM method. Vertical excitation energies were computed at this fixed geometry in order to obtain the absorption energies. Next a state-specific calculation was done to optimize the first excited state in preparation for the excited state geometry optimization. Vibrational frequencies at the excited state geometry were also computed, followed by another state-specific calculation in order to obtain the emission energies.

In their experimental work,<sup>109</sup> Shin et al. suggest that these chromophores possess charge-transfer character. Computations to assess charge transfer character via the



**Figure 2.1:** DFT equilibrium structures of the thiophene substituted RNA nucleobases computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory (in water). Atomic representations: S=yellow, N=blue, O=red, C=gray and H=white. Cartesian coordinates are available in Appendix A. (a) <sup>th</sup>Adenine, (b) <sup>th</sup>Cytosine, (c) <sup>th</sup>Guanine, (d) <sup>th</sup>Uracil.

$\Lambda$  diagnostic,<sup>118</sup> were carried out with GAMESS-US<sup>119,120</sup> at the same level of theory and using a similar PCM solvation method. The  $\Lambda$  diagnostic measures the degree of overlap of the occupied and virtual orbitals involved in a given excitation. The numerical value of  $\Lambda$  ranges from 0 to 1, with values below approximately 0.5 representing long-range excitations (charge transfer or Rydberg-type) and values near unity for short-range valence-type excitations. Charge transfer (CT) type excitation energies are often underestimated by TD-DFT calculations utilizing conventional exchange-correlation functionals. In order to correct for this shortcoming, range-separated hybrid functionals such as CAM-B3LYP that are designed for CT excitations can be used as has been done in the present work.

## 2.3 Results and Discussion

Selected ground-state equilibrium geometries of the modified nucleobases determined by DFT are depicted in Figure 2.1. In the present work, the geometries have been optimized without the ribose sugar. Our studies have shown that the presence of the sugar ring has only a modest impact on the vertical excitation energies - this was also observed previously.<sup>121</sup> The coordinates of all the computed structures at the PBE0/6-311++G(2df,2p), B3LYP/6-311++G(2df,2p) and CAM-B3LYP/6-311++G(2df,2p)

levels of theory can be found in Appendix A. The  ${}^{th}C$  and  ${}^{th}U$  base analogues display planar ( $C_s$ ) symmetry while the  ${}^{th}A$  and  ${}^{th}G$  analogues break planar symmetry slightly due to the nitrogen in the amino group adopting a pyramidalized geometry. In all cases, the lowest-energy structure was utilized for subsequent calculations. Not surprisingly, the computationally determined structures are in excellent agreement with the measured X-ray data.<sup>109</sup>

### 2.3.1 Vertical Excitation Energies

Vertical excitation energies were computed with the TD-DFT method using three different functionals: CAM-B3LYP, PBE0, and B3LYP. A full analysis of the low-lying singlet excited states was performed and the nature of the transitions characterized for the modified analogues. Since no experimental gas phase spectra are available for comparison, all TD-DFT calculations were performed in water or dioxane - the two solvents used experimentally.<sup>109</sup> The computed vertical excitation energies and corresponding oscillator strengths for the nucleobases in both water and dioxane are given in Table 2.1. Also, given are the energies of the maximum absorptions from the experimental measurements<sup>109</sup> as well as recent TD-B3LYP/6-31++G(d,p) results for the vertical excitations in water.<sup>121</sup> The trend in excitation energies is:  ${}^{th}U > {}^{th}C > {}^{th}G > {}^{th}A$ , which is in keeping with the experimental absorption energies.

The first excited state is the highly allowed  $\pi\pi^*$  state dominated by the HOMO $\rightarrow$ LUMO transition for all the modified nucleobases. Reasonable agreement of  $S_1$  excitation energies to experimental absorption maxima ( $\lambda_{max}$ ) are obtained with the CAM-B3LYP and PBE0 functionals. The mean square errors (MSE) in water are 0.19 eV and 0.02 eV, respectively. The B3LYP functional shows the least deviation with MSE = 0.01eV. In dioxane, the MSEs are 0.26 eV, 0.05 eV, and 0.01 eV for the CAM-B3LYP, PBE0, and B3LYP functionals, respectively. The oscillator strengths from the PBE0 and B3LYP functionals are slightly smaller in comparison to those computed with the CAM-B3LYP functional. The vertical excitation energies reported from all the functionals are blue shifted in comparison to the absorption maxima as determined by experiment. The dioxane absorption energies are slightly red-shifted for  ${}^{th}A$  and

**Table 2.1:** Vertical excitation energies (in eV) of the lowest lying singlet state,  $S_1$  of the modified nucleobases at  $S_0$  equilibrium geometry. Computed with TD-DFT using the specified functional and the 6-311++G(2df,2p) basis set. Oscillator strengths are given in parentheses.

Base	Solvent	CAM-	PBE0	B3LYP	Previous	Expt <sup>b</sup>
		B3LYP			Work <sup>a</sup>	
		$\Delta E$ , eV( <i>f</i> )	$\Delta E$ , eV ( <i>f</i> )	$\Delta E$ , eV ( <i>f</i> )	$\Delta E$ , eV ( <i>f</i> )	eV
<sup>th</sup> <b>A</b>	Water	4.09 (0.173)	3.88 (0.146)	3.75 (0.137)	3.59 (0.200)	3.64
	Dioxane	4.06 (0.182)	3.85 (0.154)	3.72 (0.144)		3.60
<sup>th</sup> <b>C</b>	Water	4.31 (0.134)	4.00 (0.102)	3.84 (0.093)	3.85 (0.103)	3.88
	Dioxane	4.27 (0.153)	3.95 (0.117)	3.79 (0.107)		3.81
<sup>th</sup> <b>G</b>	Water	4.25 (0.121)	3.92 (0.090)	3.76 (0.081)	3.66 (0.134)	3.87
	Dioxane	4.31 (0.130)	3.99 (0.098)	3.83 (0.088)		3.73
<sup>th</sup> <b>U</b>	Water	4.56 (0.108)	4.24 (0.080)	4.07 (0.072)	4.07 (0.085)	4.08
	Dioxane	4.62 (0.126)	4.29 (0.093)	4.12 (0.083)		4.08

<sup>a</sup> Determined at TD-B3LYP/6-31++G(d,p) level of theory.<sup>121</sup>

<sup>b</sup> Absorption maxima energies.<sup>109</sup>

$^{th}C$  which is in agreement with experimental findings, however  $^{th}G$  and  $^{th}U$  do not demonstrate this trend. It should be noted that the energy differences between the water and dioxane calculations are very small ( $\sim 0.1$  eV) which is consistent with the experimental measurements.

The next two lowest lying excited states ( $S_2$  and  $S_3$ ) are either of  $\pi\pi^*$  or  $n\pi^*$  type - the relative ordering depends upon the system studied, see Table 2.2. The valence molecular orbitals which dominate the corresponding excitations to  $S_1$ ,  $S_2$ , and  $S_3$  are depicted in Figure 2.2 (see discussion below on individual nucleobases for further details).

In comparison to their natural counterparts, the thiophene analogues' first absorption peaks are red-shifted (bathochromic) and as a result can be selectively excited. The nature of the low-lying excited states of the natural nucleobases differs in comparison to the modified nucleobases. For example, the modified thiophene analogues always display a  $\pi\pi^*$  transition for the first excited state, however theoretical studies computed in water show uracil to have a  $n\pi^*$  transition for the first excited state.<sup>84,87</sup> A detailed comparison of excitation energies of the modified nucleobases with their natural counterparts is depicted in Table 2.2. In order to facilitate this comparison, we have determined the excitation energies of the natural nucleobases at an identical level of theory to that used here for the thiophene analogues, i.e., TD-PBE0/6-311++G(2df,2p) in PCM water.

Analysis of the low-lying excitations computed with the CAM-B3LYP functional in water reveals the  $n\pi^*$  state to have the most charge transfer character. For  $^{th}A$  and  $^{th}U$ , this corresponds to the second excited state, while for  $^{th}C$  and  $^{th}G$ , the third excited state. The  $\Lambda$  values are 0.371, 0.374, 0.393 and 0.378 for  $^{th}A$ ,  $^{th}U$ ,  $^{th}C$  and  $^{th}G$ , respectively. The  $\Lambda$  values along with the corresponding excitation energies computed with the GAMESS software can be found in Appendix A.

A comparison of theoretical and experimental absorption spectra in water is depicted in Figure 2.3 and the corresponding absorption spectra in dioxane in Figure 2.4. Overall, all three functionals perform reasonably well in reproducing the experimental spectra both in terms of peak positions and their relative intensities. In order

**Table 2.2:** Vertical excitation energies (in eV) and corresponding oscillator strengths (given in parentheses) of the modified nucleobases computed using TD-DFT with the specified functional and the 6-311++g(2df,2p) basis set, compared to the natural nucleobases in water.

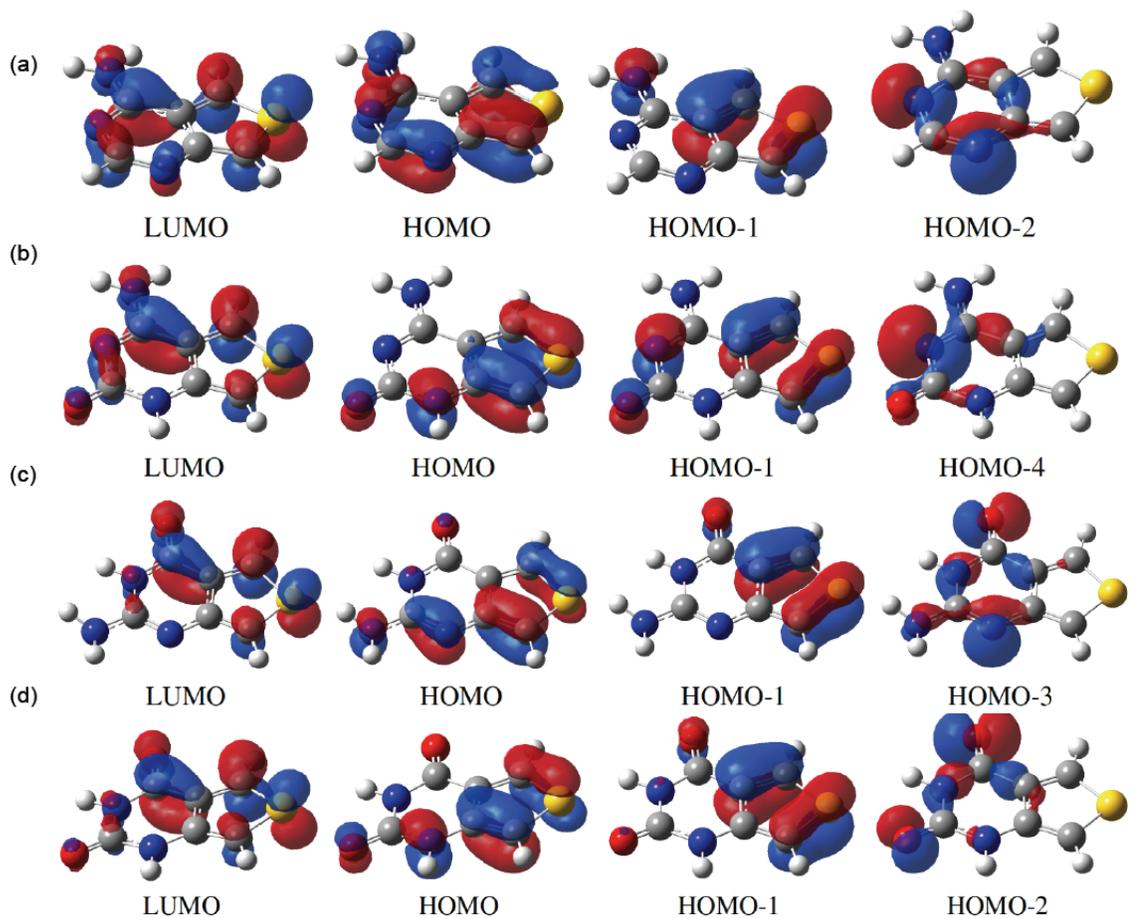
$S_n$	Base	Modified Nucleobases			Natural Bases	
		CAM-B3LYP	PBE0	B3LYP	This Work <sup>d</sup>	Previous Work <sup>e,f</sup>
S <sub>1</sub>	A	4.09 (0.173) <sup>a</sup>	3.88 (0.146) <sup>a</sup>	3.75 (0.137) <sup>a</sup>	5.11 (0.299) <sup>a</sup>	4.86 (0.443) <sup>a,e</sup>
	C	4.31 (0.134) <sup>a</sup>	4.00 (0.102) <sup>a</sup>	3.84 (0.093) <sup>a</sup>	4.98 (0.092) <sup>a</sup>	4.86 (0.178) <sup>a,e</sup>
	G	4.25 (0.121) <sup>a</sup>	3.86 (0.089) <sup>a</sup>	3.76 (0.081) <sup>a</sup>	4.93 (0.151) <sup>a</sup>	4.81 (0.240) <sup>a,e</sup>
	U	4.56 (0.108) <sup>a</sup>	4.24 (0.080) <sup>a</sup>	4.07 (0.072) <sup>a</sup>	5.08 (0.000) <sup>b</sup>	5.09 (0.000) <sup>b,f</sup>
S <sub>2</sub>	A	4.74 (0.002) <sup>b</sup>	4.38 (0.002) <sup>a</sup>	4.21 (0.001) <sup>a</sup>	5.25 (0.001) <sup>b</sup>	5.15 (0.001) <sup>b,e</sup>
	C	5.09 (0.305) <sup>a</sup>	4.90 (0.266) <sup>a</sup>	4.74 (0.253) <sup>a</sup>	5.44 (0.002) <sup>b</sup>	5.32 (0.007) <sup>b,e</sup>
	G	5.17 (0.029) <sup>a</sup>	4.96 (0.000) <sup>b</sup>	4.83 (0.004) <sup>b</sup>	5.29 (0.379) <sup>a</sup>	5.09 (0.554) <sup>a,e</sup>
	U	5.14 (0.000) <sup>b</sup>	4.85 (0.000) <sup>b</sup>	4.72 (0.000) <sup>b</sup>	5.23 (0.195) <sup>a</sup>	5.17 (0.190) <sup>a,f</sup>
S <sub>3</sub>	A	5.02 (0.106) <sup>a</sup>	4.89 (0.076) <sup>b</sup>	4.73 (0.075) <sup>b</sup>	5.35 (0.048) <sup>a</sup>	5.23 (0.084) <sup>a,e</sup>
	C	5.33 (0.002) <sup>b</sup>	5.01 (0.001) <sup>b</sup>	4.82 (0.000) <sup>b</sup>	5.67 (0.143) <sup>a</sup>	5.48 (0.234) <sup>a,e</sup>
	G	5.23 (0.000) <sup>b</sup>	5.04 (0.011) <sup>a</sup>	4.89 (0.014) <sup>a</sup>	5.36 (0.007) <sup>c</sup>	5.62 (0.007) <sup>c,e</sup>
	U	5.30 (0.177) <sup>a</sup>	5.18 (0.160) <sup>a</sup>	5.02 (0.157) <sup>a</sup>	6.17 (0.000) <sup>b</sup>	N/A

<sup>a</sup>  $\pi\pi^*$ , <sup>b</sup>  $n\pi^*$ , <sup>c</sup> Significant  $\pi\sigma^*$  contribution

<sup>d</sup> Computed with TD-PBE0/6-311++G(2df,2p) in Water (PCM)

<sup>e</sup> Reference<sup>83</sup> TD-B3LYP/6-311+G(d,p) in Water (PCM)

<sup>f</sup> Reference<sup>84</sup> TD-PBE0/6-311+G(2d,2p) in Water (PCM)



**Figure 2.2:** Valence molecular orbitals of (a)  $^{th}A$ , (b)  $^{th}C$ , (c)  $^{th}G$  and (d)  $^{th}U$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-311++G(2df,2p) level of theory with solvation in water. Visualization of orbitals was performed with GaussView V.5.<sup>122</sup>

to highlight the important points to be drawn from the determination of the vertical excitation energies, each of the nucleobases will be discussed individually.

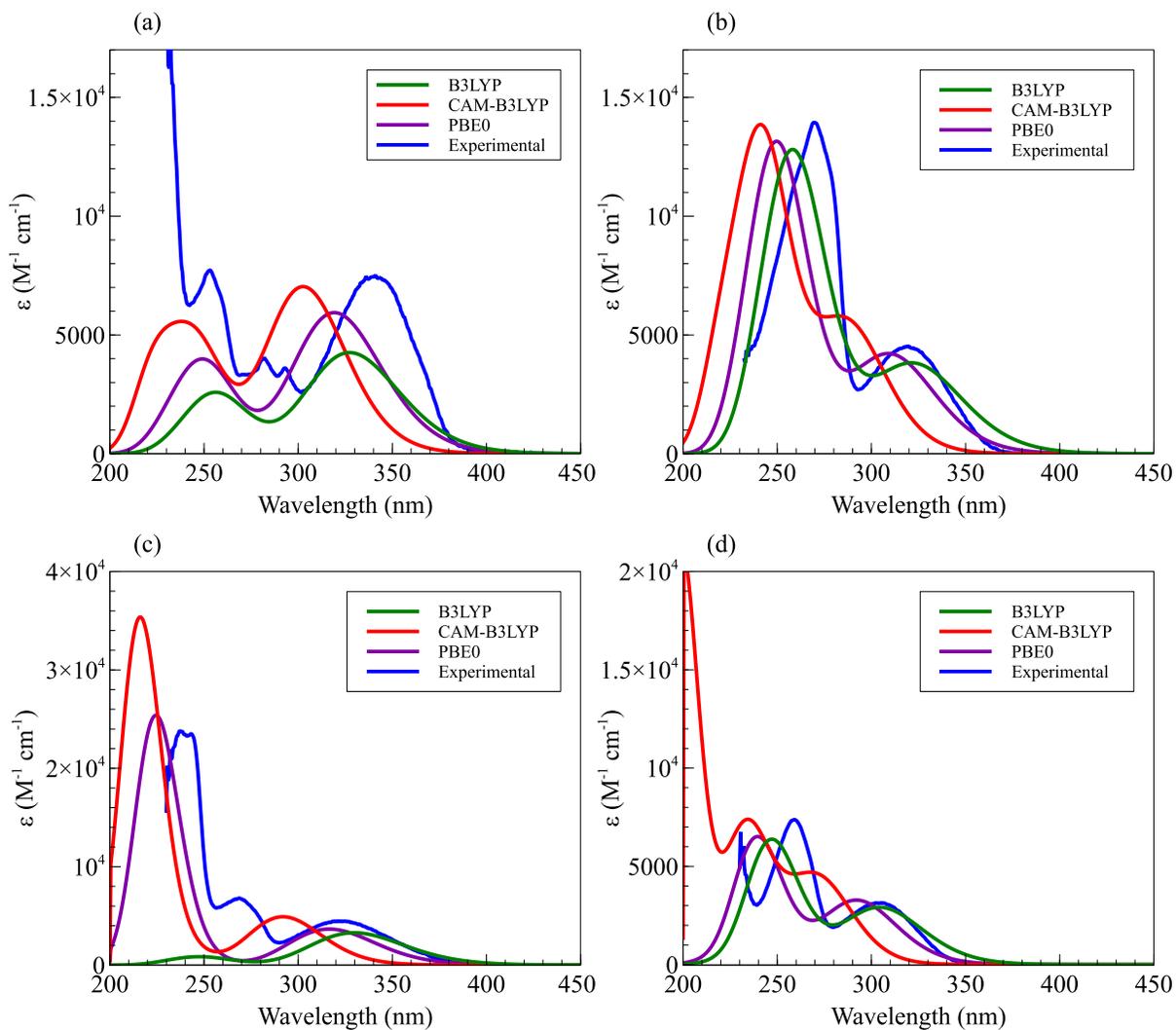
### *<sup>th</sup>Adenine*

The adenine analogue has two low-lying bright states ( $S_1$  and  $S_3$ ) with the first  $\pi\pi^*$  transition being the most intense – relative oscillator strengths are approximately 2:1, see Table 2.2. The excited states beyond  $S_1$  are a  $\pi\pi^*$  state dominated by the HOMO-1 $\rightarrow$ LUMO transition and a  $n\pi^*$  state involving the HOMO-2 $\rightarrow$ LUMO transition. The relative ordering of these states is functional dependent with PBE0 and B3LYP having the  $\pi\pi^*$  state as  $S_2$ , while it is  $S_3$  using CAM-B3LYP. Analysis of the dominant molecular orbitals in the latter  $n\pi^*$  transition shows significant evidence of charge transfer, see Figure 2.2(a), and the computed  $\Lambda$  value (0.371) is also consistent with this finding. The experimental spectrum, see Figure 2.3(a), appears to show some vibrational structure and this is not accounted for in the present work. As compared to its natural counterpart, <sup>th</sup>A exhibits a significantly larger energy gap between  $S_1$  and  $S_2$  (see Table 2.2).

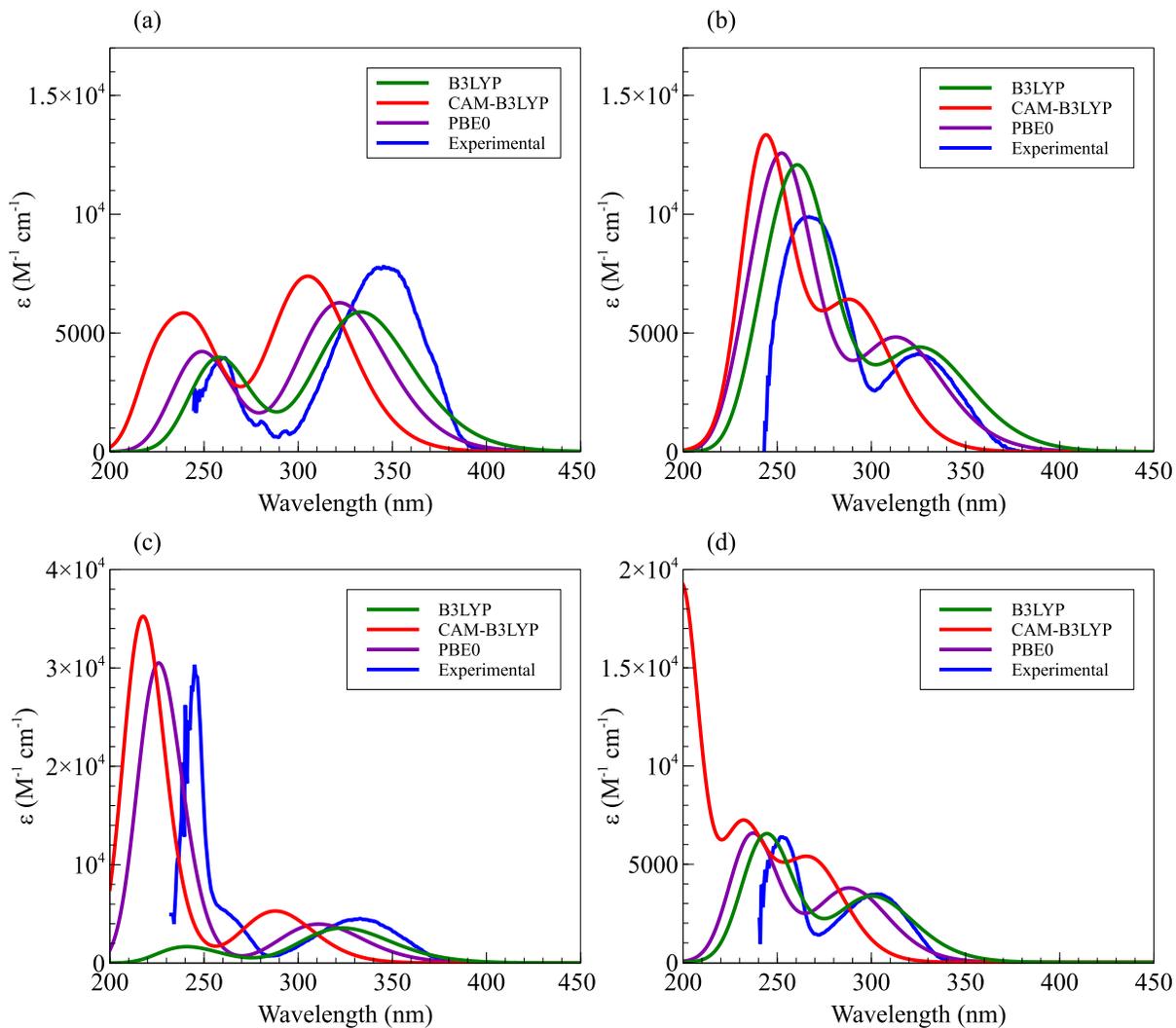
In the natural adenine, the excited state lifetime is short ( $\approx 1$  ps) due to direct rapid access to a conical intersection (CI) between  $S_1$  and  $S_0$ , where the C2-H group moves out of plane.<sup>39</sup> There is no structural reason that <sup>th</sup>A cannot access a similar CI, or the majority of the CIs that are energetically accessible, in adenine. Therefore, it is most likely energetics that make these CIs inaccessible (perhaps due to the large shift in the absorption maximum) in <sup>th</sup>A, and lead to the long, observed lifetime (3.9 ns in water).<sup>109</sup>

### *<sup>th</sup>Cytosine*

<sup>th</sup>C has three  $\pi\pi^*$  transitions (to  $S_1$ ,  $S_2$ , and  $S_4$ ) all below 5.5 eV with the two strongest excitations separated by approximately 1 eV. The most intense excitation for <sup>th</sup>C corresponds to the second  $\pi\pi^*$  state due to the HOMO-1 $\rightarrow$ LUMO transition with energies ranging from 4.7 to 5.0 eV depending on the functional used. This observation is consistent with the experimental measurement, see Figure 2.3(b). There



**Figure 2.3:** Theoretical and experimental absorption spectra of the modified RNA nucleobases in water. (a) <sup>th</sup>Adenine, (b) <sup>th</sup>Cytosine, (c) <sup>th</sup>Guanine, (d) <sup>th</sup>Uracil. Experimental data form Ref.<sup>109</sup> Theoretical absorption spectra generated with GaussView V.5 using a fixed Gaussian HWHM equal to 0.33 eV.<sup>122</sup>



**Figure 2.4:** *Theoretical and experimental absorption spectra of the modified RNA nucleobases in dioxane. (a) <sup>th</sup>Adenine, (b) <sup>th</sup>Cytosine, (c) <sup>th</sup>Guanine, (d) <sup>th</sup>Uracil. Experimental data form Ref.<sup>109</sup> Theoretical absorption spectra generated with GaussView V.5 using a fixed Gaussian HWHM equal to 0.33 eV.<sup>122</sup>*

are two additional  $n\pi^*$  type transitions below 6 eV (to  $S_3$  and  $S_5$ ) both of which have relatively small oscillator strengths ( $<0.002$ ). The lowest lying  $n\pi^*$  state is dominated by the HOMO-4 $\rightarrow$ LUMO transition, see Figure 2.2(h). For the PBE0 and B3LYP functionals, as the solvent polarity decreases, i.e., the solvent is changed from water to dioxane, a switch in energetic ordering between the  $S_2$  and  $S_3$  states is observed. This results in a  $n\pi^*$  state separating the two bright  $\pi\pi^*$  states, see Tables S1-S3. CAM-B3LYP results do not exhibit the switch in order of  $S_2$  and  $S_3$  states, but for CAM-B3LYP these states are nearly degenerate in dioxane (5.09 eV and 5.10 eV, respectively).

The excited state lifetime of natural cytosine is short ( $\approx 1$  ps), while that for  $^{th}C$  is much, much longer (15.2 ns in water).<sup>109</sup> In cytosine, the deactivation mechanism is suggested to be the most complex of the natural nucleobases with deactivation occurring via either a  $\pi\pi^*$  to  $n\pi^*$  to ground state via a CI mechanism or a  $\pi\pi^*$  to  $n\pi^*$  to  $\pi\pi^*$  to ground state via a CI one.<sup>39</sup> In  $^{th}C$ , the first mechanism should be structurally allowed as it involves puckering at N1 and N3. However, the second of these cannot proceed as it involves deformation at C6, which is unlikely (impossible) due to the thiophene substitution. Hence, the primary reason for the long lifetime of  $^{th}C$  lies in the energetics of the CI relative to the absorption maximum. This speculation remains to be confirmed by further computational determination of CIs.

### *$^{th}$ Guanine*

$^{th}G$  has a bright  $\pi\pi^*$  type state followed by  $n\pi^*$  and  $\pi\pi^*$  states which are nearly degenerate in energy (independent of the functional, see Table 2.2). The second  $\pi\pi^*$  state which is dominated by a HOMO-1 $\rightarrow$ LUMO orbital transition has a significantly lower intensity in comparison the first  $\pi\pi^*$  transition. For the  $n\pi^*$  state, both HOMO-2 $\rightarrow$ LUMO and HOMO-3 $\rightarrow$ LUMO transitions are contributors. The CAM-B3LYP and PBE0 functionals predict a very intense transition ( $S_5$ ) at 5.7 and 5.5 eV, respectively, and this strong peak is observed experimentally, see Figure 2.3(c). However the B3LYP functional fails to demonstrate this highly allowed transition, where the B3LYP spectrum exhibits only two (relatively) weak transitions between

200-400 nm, see Figure 2.3(c). The experimental spectrum appears to display a third transition at 270 nm that is not seen in any of the computed spectra.

Natural guanine displays the same deactivation mechanism as adenine via a CI involving puckering at C1. Again, this leads to a short excited state lifetime (<1ps). The long lifetime determined for  ${}^{th}G$  (14.8 ns) suggests that this CI is no longer energetically accessible. Note that the shift in energy required would need not be large as this CI is only  $\approx 0.4$  eV below the absorption maximum in natural guanine.

### ${}^{th}Uracil$

For  ${}^{th}U$  the observed transitions are as follows;  $\pi\pi^*$ ,  $n\pi^*$  followed by a more intense third  $\pi\pi^*$  state irrespective of the functional employed. The orbital transitions of the  $n\pi^*$  and third  $\pi\pi^*$  state are HOMO-2 $\rightarrow$ LUMO and HOMO-1 $\rightarrow$ LUMO, respectively (see Figure 2.3(d)). Changing the solvent polarity does not have any significant effect on the ordering of states as seen previously with  ${}^{th}C$ , see Tables S1-S3. There is approximately a 0.2 eV difference in excitation energies between the water and dioxane calculations when the same functional is employed. This is also consistent with the experimental findings for the absorption maxima of  ${}^{th}U$  in water compared to dioxane.

In natural uracil, the deactivation proceeds from  $S_2$  (the bright  $\pi\pi^*$  state) to the dark  $n\pi^*$  state ( $S_1$ ).<sup>39</sup> In  ${}^{th}U$ , the ordering of these two states is reversed. Thus, this pathway is no longer energetically accessible and a long excited state lifetime (11.5 ns in water) is observed.<sup>109</sup>

### Summary of Vertical Excitation Energy Data

In summary, the results of the TD-DFT calculations of these modified bases show good quantitative agreement with the excitation energies and relative intensities for the absorption peaks seen experimentally. Moreover, the changes to the spectra observed when going from water to dioxane solvent are faithfully reproduced. The energetics and nature of the excited states are suggestive of the reasons for the much longer excited state lifetimes (order of ns) observed relative to the natural nucleobases

(order of ps), see also discussion of emission energies below.

### 2.3.2 Emission Energies

The computed emission energies of the modified analogues are compared to the experimental emission maxima in Table 2.3. All involve emission from the lowest lying  $\pi\pi^*$  state. There is no straightforward computational method for determining quantum yields, and hence only energies are compared. The MSE in energies are 0.1 eV, 0.03 eV, and 0.01 eV in water and 0.12 eV, 0.02 eV, and 0.01 eV in dioxane for the CAM-B3LYP, PBE0, and B3LYP functionals, respectively. With the exception of  $^{th}\text{A}$ , there is a hypsochromic shift in energies going from water to dioxane and this is in keeping with the experimental findings.<sup>109</sup> Except for PBE0 in dioxane, where an excited state geometry could not be converged, there appears to be little role of charge transfer in the emissive state.

**Table 2.3:** *Emission energies (in eV) of the modified nucleobases at the  $S_1$  minimum. All computations utilized TD-DFT with the specified functional and the 6-311++G(2df,2p) basis set.*

Base	Solvent	CAM-B3LYP	PBE0	B3LYP	Experimental <sup>a</sup>
$^{th}\text{A}$	Water	3.36	3.28	3.16	2.95
	Dioxane	3.32	3.23	3.13	3.02
$^{th}\text{C}$	Water	3.15	2.93	2.77	2.89
	Dioxane	3.30	3.06	2.89	2.94
$^{th}\text{G}$	Water	2.98	2.73	2.57	2.74
	Dioxane	3.30	N/A <sup>b</sup>	2.91	2.92
$^{th}\text{U}$	Water	3.25	3.00	2.83	3.03
	Dioxane	3.62	3.37	3.19	3.28

<sup>a</sup> Emission maxima energies.<sup>109</sup>

<sup>b</sup> Not available due to problems with charge transfer for dioxane results.

## 2.4 Conclusions

The photophysical properties of the thieno[3,4-d]-pyrimidine nucleobase analogues have been investigated by TD-DFT calculations. Absorption and fluorescence energies of the modified analogues were determined using the CAM-B3LYP, PBE0, and B3LYP functionals. The theoretical results reveal that the B3LYP functional predicts energies closest to those measured by experiment and the computed energies from the CAM-B3LYP and PBE0 functionals were blue-shifted in comparison to B3LYP and experimental results. The characteristics of the absorption spectra are successfully reproduced by the TD-DFT method in terms of both peak position and intensity. The TD-B3LYP results however do not satisfactorily predict the intensities for the *th*G nucleobase. Thus, while TD-B3LYP describes energies of the singlet electronic transitions faithfully, the TD-PBE0 method gives the best overall representation compared to experimental findings. We further conclude that charge transfer state for this class of modified nucleobases is represented by the lowest-lying  $n\pi^*$  state.

From the energetics determined in the present work, the reasons for the long excited state lifetimes in the thieno-analogues compared to the natural nucleobases can be surmised. The CIs involved in the primary deactivation paths of the natural nucleobases are structurally allowed in the thieno-analogues. However, they most probably are no longer energetically accessible. The detailed study of the CIs in the thieno-analogues will form the basis of future work, and confirm this speculation, or perhaps lead to an alternate mechanism.

## Chapter 3

# Computational Insights Into Stable Tautomers of Emissive Thiophene-Derived RNA Nucleobases<sup>1</sup>

*Part of the inhumanity of the computer is that, once it is competently programmed and working smoothly, it is completely honest.*

– Isaac Asimov

---

<sup>1</sup>This chapter is adapted from the submitted paper M. Gedik, T. Huynh and A. Brown *Chem. Phys. Chem.*, submitted.

## 3.1 Introduction

The natural bases of ribonucleic acids, i.e., adenine (A), guanine (G), cytosine (C), and uracil (U), can have various tautomeric forms due to protonation and deprotonation processes. These processes are more prevalent in solution, owing to the role of solvent-exchangeable protons in the micro-environment. The presence of tautomers increases the structural diversity of nucleobases which can potentially offer insight into nucleic acid dynamics as well as mutagenesis. The keto and imino forms dominate the tautomeric equilibrium for all the natural nucleobases. Adenine can adopt the imino form, guanine and cytosine are able to form both enol and imino tautomers and combinations therein, whereas uracil only participates in keto-enol tautomerism. The design of nucleoside analogues with multiple tautomeric forms are desirable, as they can be utilized as probes in biochemical applications.<sup>123</sup> For a comprehensive discussion on applications, we refer the reader to the article by Singh et al.<sup>124</sup>

The existence of major and minor tautomers of natural adenine, cytosine and guanine have been demonstrated experimentally by supersonic jet cooled spectroscopy,<sup>125</sup> IR spectroscopy in helium nanodroplets,<sup>126,127</sup> and NMR spectroscopy.<sup>128</sup> Theoretical approaches are challenging as the computed relative energies between tautomers are extremely sensitive to the method employed, see the review and references therein.<sup>129</sup> For example, computations at the CCSD(T) level of theory are essential in capturing the relative ordering of tautomers of cytosine. Both theoretical investigations and experimental studies on the tautomers of nucleic acids reveal that the relative stabilities are highly dependent on the environment.<sup>89,129</sup> Factors that influence the tautomeric equilibrium of nucleobases include solvation, base pair formation, stacking interactions and interactions with metal ions. However, as a first step, it is important to understand the relative energies, and corresponding spectral signatures, in the gas phase and water.

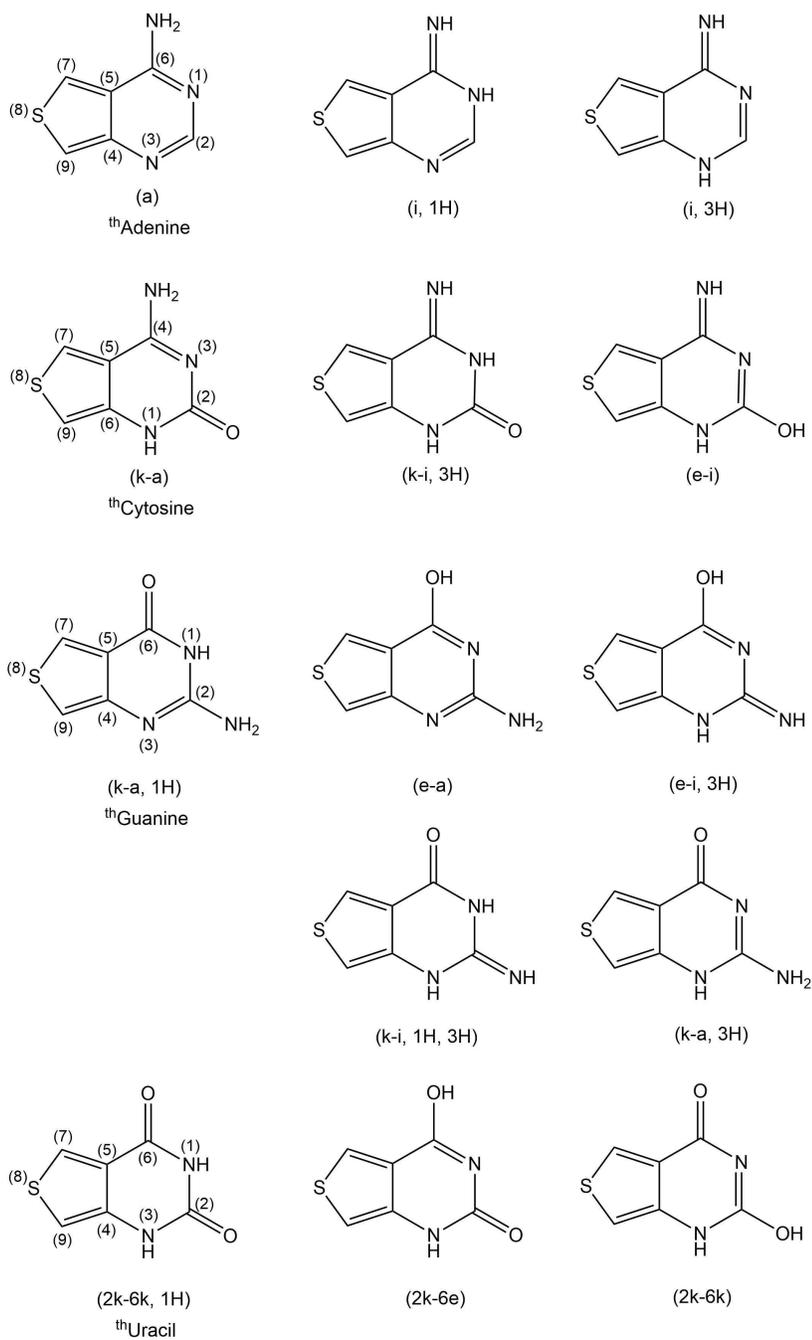
Upon photo-excitation, the natural nucleobases undergo ultrafast non-radiative relaxation, and hence cannot be used directly as spectroscopic probes. In 2011,<sup>109</sup> Tor and co-workers introduced a family of emissive RNA nucleobases derived from

a common thieno[3,4-d] pyrimidine core; these modified nucleobases will be referred to as  $^{th}A$ ,  $^{th}C$ ,  $^{th}G$ , and  $^{th}U$ . The photophysical properties of the canonical forms of the modified nucleobases have been previously studied computationally using time-dependent density functional theory (TD-DFT).<sup>121,130</sup> However, investigation of the tautomers of the thieno-analogues has been lacking until recently. In 2016, Sholokh et al. performed a joint theoretical and experimental investigation of the tautomers of thienoguanosine.<sup>131</sup> Two tautomers of thienoguanosine were identified ( $^{th}G$ -H1 and  $^{th}G$ -H3, which we designate as (k-a,1H) and (k-a,3H), respectively, see Figure 3.1 through contributions to both the absorption and emission spectra. Subsequent DFT computations at the PBE0/6-311+G(2d,2d)/PBE0/6-31G(d) level of theory in water revealed the  $^{th}G$ -H3 tautomer was only 0.11 eV (2.54 kcal/mol) higher in energy than the canonical  $^{th}G$ -H1 form. As has been seen previously for the natural nucleobases, their relative energy was sensitive to the inclusion of explicit water molecules; the energy difference between the tautomers was only 0.05 eV (1.15 kcal/mol) with the inclusion of 6 explicit water molecules within an implicit water environment. Corresponding TD-DFT computations successfully reproduced the shifts in the absorption and emission spectra observed experimentally. Due to the interest in the  $^{th}G$  tautomers, in the present work, the tautomers of  $^{th}A$ ,  $^{th}C$ , and  $^{th}U$  (as well as  $^{th}G$ ) are examined computationally for their relative stabilities, and predictions are made for their infrared (IR) and UV-Vis spectra to aid in experimental identification.

## 3.2 Computational Methods

Ground state geometries were optimized in the gas phase with the B3LYP<sup>132–135</sup> functional utilizing the 6-311++G(2df,2p)<sup>110,111</sup> basis set. Vibrational modes were computed for each of the tautomers at the same level of theory. The analysis of frequencies revealed the absence of imaginary values which verified that all optimized structures were minima. The predicted vibrational spectra will also be useful for potential identification of these tautomers. The grid used for numerical integration in DFT is set to "Ultrafine," i.e., a pruned grid of 99 radial shells and 590 angular

points per shell. The gas phase geometries were then re-optimized using the integral equation formalism variant of the polarizable continuum model (IEF-PCM, here simply referred to as PCM)<sup>113-117</sup> to account for (implicit) solvation in water. To assess the accuracy and reliability of the energies determined with DFT, we have also computed the relative energies with a single point energy computation with CCSD(T)<sup>136,137</sup>/cc-pVDZ<sup>138</sup> at the both the DFT optimized gas phase and solution-phase geometries. The vertical excitation energies were determined using TD-DFT with the B3LYP,<sup>68,69</sup> CAM-B3LYP,<sup>70</sup> and PBE0<sup>71-74</sup> functionals in both gas phase and water with the 6-311++G(2df,2p) basis set for all atoms. All quantum chemical calculations were performed with the Gaussian software package.<sup>139</sup> Geometry optimizations used default convergence criteria.



**Figure 3.1:** Structures of the thiophene nucleobases and their major tautomers. The canonical forms of each nucleobase are (a), (k-a), (k-a, 1H), and (2k-6k, 1H) for  $^{th}A$ ,  $^{th}C$ ,  $^{th}G$ , and  $^{th}U$ , respectively. The coordinates for all optimized structures in the gas phase and water are given in Appendix B.

### 3.3 Results and Discussion

The thiophene substituted nucleobase analogues ( $^{th}A$ ,  $^{th}C$ ,  $^{th}G$ ,  $^{th}U$ ) and their tautomers were characterized by density functional theory computations. The structures of the major tautomers are depicted in Figure 3.1. We considered both the amino-imino as well as the keto-enol structures that could be formed with consideration of the restrictions due to the position of the ribose moiety in the nucleoside; with the ribose connecting at C<sub>9</sub> in  $^{th}A$  and  $^{th}G$  and at N<sub>3</sub> in  $^{th}C$  and  $^{th}U$ .

After the optimized geometries have been determined, the relative stabilities of the tautomers can be assessed. The relative electronic ( $\Delta E$ ) and Gibbs free energies ( $\Delta G$  at 298.15 K) as compared to the lowest energy canonical form were determined at the B3LYP/6-311+G(2d,2p) level of theory for all the nucleobase (designated with the subscript NB) and nucleoside (NS) tautomers and are summarized in Table 3.1. Values in both the gas phase and in water (PCM) are provided as are Sholokh et al.’s previous DFT results for  $^{th}G$  (with the ribose replaced by methyl) as computed at the PBE0/6-311+(2d,2p)//PBE0/6-31G(d) level of theory in PCM (water).<sup>131</sup> Sholokh et al. also considered the role of solvation (gas phase versus dioxane versus water) including (up to) six explicit water molecules on the relative energies of the  $^{th}G$  tautomers.

To compare with the DFT results, CCSD(T)/cc-pVDZ single point computations were carried out at the gas phase and the solution-phase DFT optimized geometries for tautomers with relative energies (at the DFT level) within 10 kcal/mol of the canonical tautomer; the gas phase relative energies are given in Table 3.1, while those at the solution-phase geometries are given in Appendix B. We have omitted  $^{th}U$  from further CCSD(T)/cc-pVDZ computations as there are not low energy tautomers for this nucleobase. In general, the relative energies of the tautomers exhibit similar trends in the gas phase and water (as well as with and without the ribose); however, there are some notable exceptions (cf.  $^{th}G$  and  $^{th}C$ ) and these are discussed in detail in Sections 3.3.1 and 3.3.2. A comparison of the relative energies of the nucleobase with their nucleoside structures, in which the ribose moiety has been included has

**Table 3.1:** Relative electronic,  $\Delta E$ , and free energies,  $\Delta G$ , (in kcal/mol) for modified nucleobase (NB) and nucleoside (NS) tautomers in the gas phase and water(PCM).<sup>a</sup> Values are given relative to the canonical tautomer.

Tautomer	Gas phase					Water			
	$\Delta E_{NB}$	$\Delta E_{NB}^b$	$\Delta E_{NB}^{CC}$	$\Delta G_{NB}$	$\Delta G_{NS}$	$\Delta E_{NB}$	$\Delta E_{NB}^b$	$\Delta G_{NB}$	$\Delta G_{NS}$
(k-a, 3H)	7.99	9.22	8.14	7.81	12.82	1.79	2.54	1.38	3.85
(k-i, 1H,3H)	5.17	6.69	5.37	5.02	4.22	5.71	6.69	5.63	3.44
<sup>th</sup> G (e-a)	5.98	5.53	8.27	6.05	6.70	9.58	8.99	9.66	7.91
(e-i, 3H)	15.81	16.60	—	15.73	18.98	17.87	18.22	17.85	19.10
<sup>th</sup> C (k-i, 3H)	-0.97	—	-4.01	-0.76	-0.19	3.24	—	3.35	3.31
(e-i)	27.37	—	—	27.33	18.07	24.37	—	24.39	20.81
<sup>th</sup> A (i, 1H)	6.73	—	4.96	6.85	9.50	6.47	—	6.79	7.23
(i, 3H)	14.73	—	—	14.60	16.52	10.03	—	10.22	9.56
<sup>th</sup> U (2e-6k)	18.50	—	—	18.21	21.01	16.29	—	16.31	17.05
(2k-6e)	22.12	—	—	21.98	21.20	17.20	—	17.22	17.81

<sup>a</sup> Geometries in present work are computed at the B3LYP/6-311++G(2df,2p) level of theory. Energies are determined at the same level of theory, with the exception of <sup>CC</sup> which represents CCSD(T)/cc-pVDZ energies. <sup>b</sup> PBE0/6-311+G(2d,2p)//PBE0/6-31G(d) in water (PCM) from Shokolkh et al.<sup>131</sup> with ribose replaced with a methyl group.

also been undertaken. Based on the relative free energies of the nucleobases in water, the relative population of the tautomers can be determined from the Boltzmann distribution. At 298 K, the relative populations of the tautomers are expected to be negligible compared to the canonical forms save for the <sup>th</sup>G (k-a,3H) ( $\sim 10\%$ ) and <sup>th</sup>C (k-a,3H) ( $\sim 1\%$ ). However, as shown for <sup>th</sup>G, consideration of explicit solvation can reduce the differences in the relative energies and thus, increase the predicted relative populations. Although the energies do change, they are unlikely to shift such that other high energy tautomers (as predicted with PCM) will become significantly populated, and hence observable in experiments on isolated nucleobases (or nucleosides). Such a definitive conclusion may not necessarily be drawn in more complicated environments.

Since the thieno-modified nucleosides have been designed to be photo-excited such that they can be used as fluorescent probes, differences in UV-Vis absorption spectra

can serve as possible signatures for the different tautomers. For example, the experimental  ${}^{th}G$  absorption spectrum was deconvoluted into the two major contributing tautomers (k-a,1H) and (k-a,3H).<sup>131</sup> To assist with possible future identification, the excitation energies (and corresponding oscillator strengths) for the lowest energy  ${}^{th}G$ ,  ${}^{th}C$ , and  ${}^{th}A$  tautomers including the canonical forms explained previously,<sup>130</sup> are given in Table 3.2; vertical excitation energies are not included for the high-energy  ${}^{th}U$  tautomers.

All tautomers exhibit shifted vertical excitation energies to  $S_1$  compared to their most stable, canonical form. Implicit solvation in water yields blue-shifted energies up to 0.4 eV compared to gas phase results. However, the values in water are the most relevant for experimental measurements. Previous work<sup>130</sup> showed that differences in excitation energies between water and less polar dioxane were much smaller. For  ${}^{th}G$ (k-a, 1H) and  ${}^{th}G$ (k-a, 3H), Sholokh et al. demonstrated that inclusion of explicit water could shift the excitation to lower energy (longer wavelength), but the magnitude of the shift was strongly functional dependent; PBE0 predicted (perhaps incorrectly) larger shifts than M05-2X. With the exception of  ${}^{th}G$ (k-i, 1H 3H) tautomer, the differences in the excitation energies between the tautomers and the canonical form in water are on the order of 0.2-0.4 eV (20-40 nm) for all the functionals employed. There is a significant blue-shift (0.3-0.5 eV) in excitation energies to  $S_1$  for the CAM-B3LYP functional in comparison to B3LYP and PBE0 in the gas phase as well as water.

In the following sections, we discuss the tautomers for each of the thiophene nucleobases in turn. The results are compared and contrasted to those for the natural nucleobases, keeping in mind the thieno substitution strongly changes the accessible tautomers compared to natural counterparts. All energies in the following sections refer to the relative Gibbs free energies for the isolated nucleobase in water, unless otherwise noted.

**Table 3.2:** Vertical excitation energies (in eV) and corresponding oscillator strengths to  $S_1$  of modified nucleobase tautomers<sup>a</sup> computed at the PBE0/6-311++G(2df,2p) level of theory in water.<sup>b</sup> Previous theoretical results also included.<sup>c</sup>

Tautomer	gas phase		Water		Ref <sup>131</sup>	
	$\Delta E$	f	$\Delta E$	f		
<sup>th</sup> G (k-a,1H)	4.05	0.069	3.88	0.088	3.75 (0.09)	
	(k-a, 3H)	4.03	0.000	4.33	0.124	4.15 (0.12)
	(k-i,1H,3H)	3.99	0.066	3.89	0.080	
	(e-a)	3.60	0.065	3.61	0.086	
<sup>th</sup> C (k-a)	3.91	0.085	3.96	0.100		
	(k-i,3H)	4.32	0.102	4.27	0.124	
<sup>th</sup> A (a)	3.85	0.110	3.84	0.145		
	(i,1H)	4.29	0.152	4.19	0.206	
	(i,3H)	4.15	0.181	4.08	0.238	

<sup>a</sup> Values not computed for uracil tautomers due to their high relative energies compared to the canonical form.

<sup>b</sup> Detailed TD-DFT results for excitation to  $S_1$ ,  $S_2$ , and  $S_3$  for the PBE0, B3LYP, and CAM-B3LYP functionals are provided in Tables B3-B20 of Appendix B.

<sup>c</sup> Computed at the TD-PBE0/6-311+G(2d,2p) level of theory in PCM( $H_2O$ ) and only the two major forms considered.

### 3.3.1 <sup>th</sup>Guanine

The natural guanine nucleobase exhibits several stable tautomers due to the proton transfer from the heterocyclic ring nitrogen to the oxo or imino groups. These processes result in the formation of -OH (keto-enol tautomerism) and -NH<sub>2</sub> (imino-amino tautomerism). Theoretical approaches demonstrate that the relative ordering of the tautomers of guanine varies significantly depending on the microenvironment.<sup>140</sup> Particularly in the gas phase several tautomers (and rotamers) have been shown to have relative stabilities within 1 kcal/mol compared to the dominant keto-amino form.<sup>141</sup>

For <sup>th</sup>G, we have assessed the keto-enol tautomerism involving the N1 and O6

atoms, the N1-N3 amino-imino tautomerism and combinations thereof. There are four possible tautomers as depicted in Figure 3.1. In the gas phase, (k-i, 1H, 3H) and (e-a) are the most stable tautomers according to the DFT results for both nucleobase and nucleoside, see Table 3.2. However, their relative free energies depend on whether the nucleobase or nucleoside is considered with differences between nucleobase and nucleoside of  $\sim -0.8$  kcal/mol and  $+0.7$  kcal/mol, respectively. In the solution phase, the relative free energy ordering of the tautomers changes and depends on whether one considers the nucleobase or nucleoside. In the gas phase, the (k-a, 3H) tautomer is approximately 2-3 kcal/mol less stable compared to (k-i,3H,1H) and (e-a), but represents the most stable tautomer in water. The relative population of (k-a,3H) in water is calculated to be approximately 10%, which is in agreement with the previous computational results;<sup>131</sup> the experimental molar fraction was determined to be 0.44. Interestingly, when considering the nucleosides, the (k-i,3H,1H) and (k-i,3H) tautomers are nearly isoenergetic, albeit with relative energies  $\sim 3.5$  kcal/mol greater than the canonical form, i.e., 1% relative populations. Finally, <sup>th</sup>G-(e-i, 3H) is the least favoured tautomer in both gas phase and water.

Our findings for the relative energies of the <sup>th</sup>G nucleobase tautomers are consistent with the recently published work by Sholokh et al.,<sup>131</sup> where the computed electronic energy difference between the two lowest energy tautomers is 2.5 kcal/mol (0.11 eV) at the PBE0/6-31+G(2d,2p)//PBE0/6-31G(d) level of theory in water. Sholokh et al. demonstrated that the inclusion of explicit water (2 or 6 explicit H<sub>2</sub>O) decreased the relative energy to 1.5 kcal/mol (0.067 eV) and 1.1 kcal/mol (0.049 eV), respectively at the same level of theory. Using the M05-2X functional reduced these energies to 2.4, 1.2, and 0.5 kcal/mol (0.103, 0.054, and 0.023 eV), respectively. Therefore, it is feasible that inclusion of explicit water could reduce the relative energies of the (k-a,3H) and (k-i,1H,3H) tautomers, such that both could be present in solution. However, the presence of (k-i,3H,1H) was not observed by Sokolkh et al.<sup>131</sup>

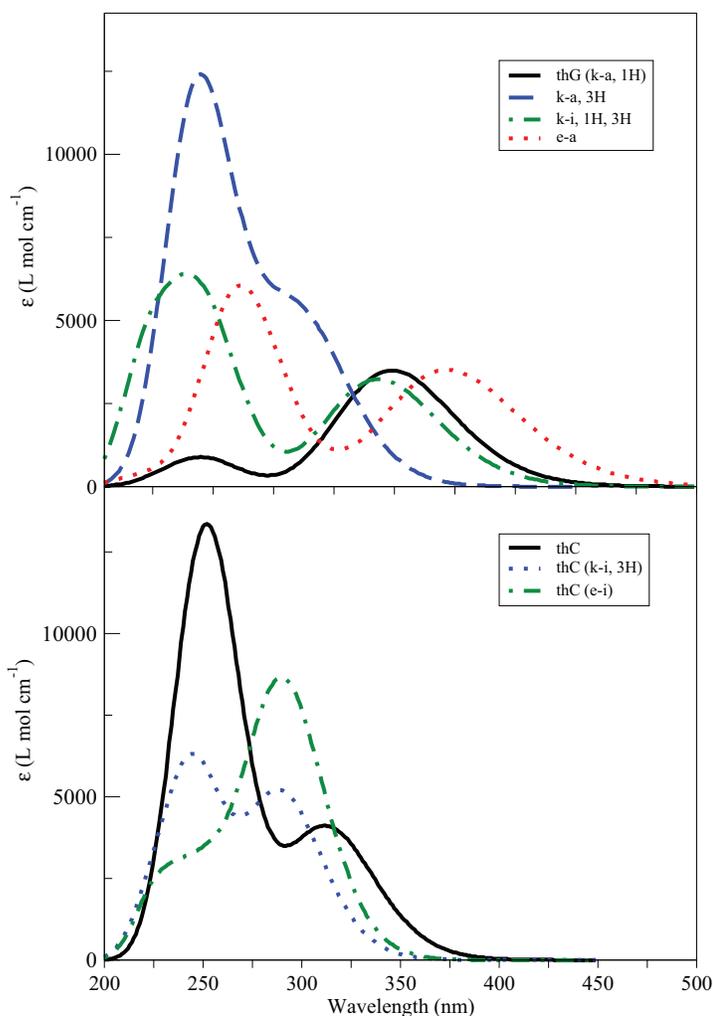
The vertical excitation energies in the gas phase are within 0.1 eV compared to solvation in water, with the exception of the most stable <sup>th</sup>G tautomer- (k-a,

3H), see Table 3.2.  ${}^{th}\text{G}(\text{k-a}, 3\text{H})$  is red-shifted by 0.26 eV in water. However, the solution phase results are the most experimentally relevant. The vertical excitation energies and intensities (oscillator strength) of  ${}^{th}\text{G}(\text{k-i}, 1\text{H}, 3\text{H})$  are nearly identical to that of the major  ${}^{th}\text{G}(\text{k-a}, 1\text{H})$  tautomer in the solution phase, see Table 3.2 and Figure 3.2. Hence, although the relative energies (based on the nucleoside model, see Table 3.1) suggest that  ${}^{th}\text{G}(\text{k-i}, 3\text{H}, 1\text{H})$  could be present in comparable amounts to the previously detected  $(\text{k-a}, 3\text{H})$  tautomer, it could not be distinguished by UV-absorption spectra alone. On the other hand, these three low-lying  ${}^{th}\text{G}$  tautomers could be readily distinguished by their IR spectra, see Figures B5, B6, B7 in the Appendix.

### 3.3.2 ${}^{th}\text{Cytosine}$

In the gas phase, the keto-imino form of  ${}^{th}\text{C}$  dominates over the canonical keto-amino conformation, whereas this is reversed in the aqueous environment. A similar trend is observed in natural cytosine, where the amino-hydroxy tautomer is preferred over the usual amino-oxo form both theoretically,<sup>142</sup> and experimentally.<sup>143</sup> It is important to note however, that the formation of the amino-hydroxy form of  ${}^{th}\text{C}$  is not considered in this study as the proton transfer from the N1 position would not be possible due to the ribose attachment at that position (see Figure 3.1).

Based on the relative free energies of either nucleobase or nucleoside, the  ${}^{th}\text{C}(\text{k-i}, 3\text{H})$  tautomer could be present at  $\sim 1\%$  population. As noted for the  ${}^{th}\text{G}$  tautomers,<sup>131</sup> this relative population (energy) increased (decreased) with the inclusion of explicit waters in the simulations and, even with the inclusion of explicit waters, the predicted population was less than that measured experimentally (although as noted in Section 3.3.1, the possible role of the  ${}^{th}\text{G}(\text{k-a}, 3\text{H}, 1\text{H})$  tautomer was not considered). From the predicted UV-Vis absorption, see Figure 3.2 and Table 3.2, the minor  ${}^{th}\text{C}(\text{k-i}, 3\text{H})$  tautomers could be readily distinguished from the canonical form. These spectroscopic differences are even more apparent in the predicted IR spectra. Hence, it should be possible to observe the presence of this tautomer.



**Figure 3.2:** Computed UV-Vis absorption spectra of (top) canonical  $^{th}$ Guanine and its three lowest energy tautomers, (bottom) canonical  $^{th}$ Cytosine and its lowest energy tautomers. Computed at the PBE0/6-311++G(2df,2p) level of theory in water (PCM). Detailed numerical results for the contributing excited electronic states are provided in Appendix B.

### 3.3.3 $^{th}$ Adenine

The amino-imino tautomerization of  $^{th}$ A produces two tautomers with relative energies that are in the range of 7-15 kcal/mol higher than that of the canonical form in the gas phase. Solvation in water does not significantly alter the relative energy for  $^{th}$ A(i, 1H); however, the energy of the less stable  $^{th}$ A(i, 3H) conformer is reduced

by approximately 4 kcal/mol in water. The relative populations of both tautomers are much less than 1% according to the Boltzmann distribution. If due to specific micro-environmental factors, these populations were increased, the tautomers exhibit significantly shifted UV-Vis absorption energies, see Table 3.2, as well as distinct signatures in their IR spectra, see Figures B1 and B2.

### 3.3.4 <sup>th</sup>Uracil

The tautomers of <sup>th</sup>Uracil do not have appreciable relative populations in the gas phase or solution. The relative free energies of the two most stable tautomers are 16-18 kcal/mol in water (19-22 kcal/mol in the gas phase) higher than canonical <sup>th</sup>U regardless of consideration of nucleobase or nucleoside. These large energy differences for the tautomers are also found with natural uracil by both experimental and theoretical studies. As mentioned previously, due to the large relative energies, further theoretical spectroscopic study (UV-Vis or IR) was not undertaken.

## 3.4 Conclusions

The tautomers of thiophene derived nucleobases have been investigated with DFT methods in the gas phase and in water (with the PCM solvation model). We considered 14 tautomers from this particular family of modified RNA nucleobases. Based on relative free energies in solution, the <sup>th</sup>G(k-a,3H) and <sup>th</sup>C(k-i,3H) tautomers are the only ones that give appreciable populations at 9% and 0.3% respectively; considering the nucleosides, the <sup>th</sup>G(k-a,3H,1H) tautomer may also have a modest population. While the <sup>th</sup>G(k-a,3H) tautomer, could, and has been,<sup>131</sup> identified by UV-Vis absorption, the <sup>th</sup>G(k-a,3H,1H) tautomer is not spectroscopically distinct (in UV-Vis absorption) from the canonical form. This tautomer could be observed though IR spectra. The <sup>th</sup>C(k-i,3H) tautomer could be observed via UV-Vis or IR spectra. However, it may have to be present in more significant population than that predicted from the present work. Overall, the present results suggest that further experimental investigation of the <sup>th</sup>G and <sup>th</sup>C species is warranted.

## Chapter 4

# Two-Photon Emissive Nucleobase Analogues: A Computational Investigation of Structural Effects on Photophysical Properties

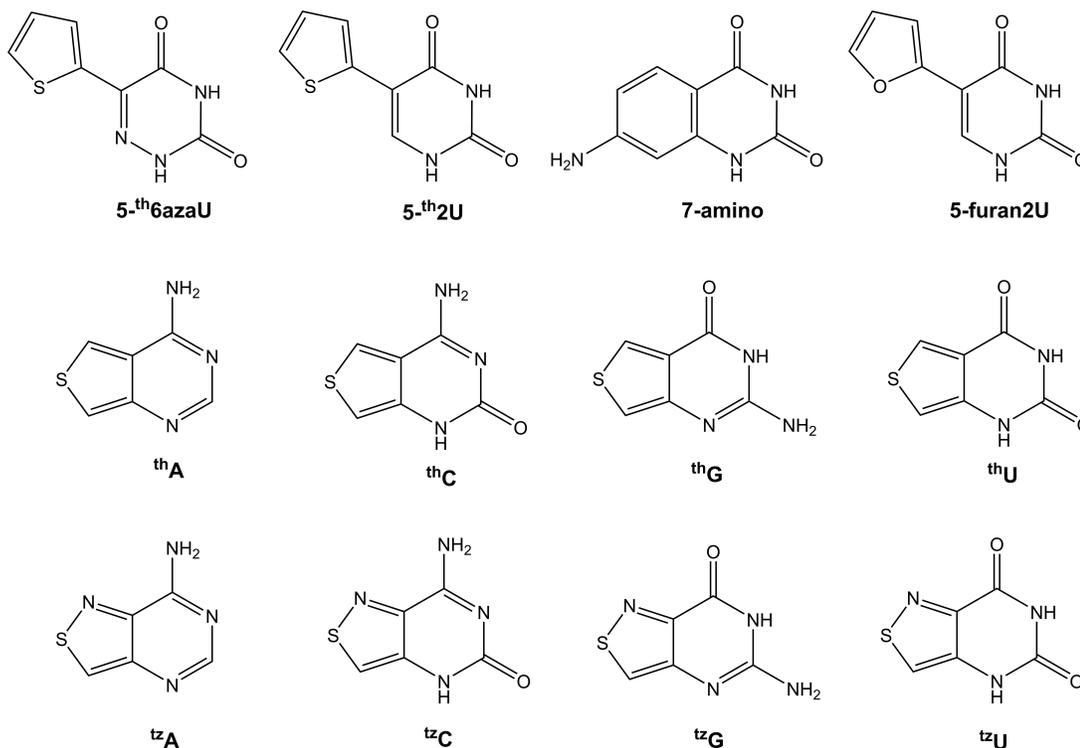
*We are reaching the stage where the problems we must solve are going to be insoluble without computers. I do not fear computers. I fear the lack of them.*

– Isaac Asimov

## 4.1 Introduction

The development of isomorphous fluorescent nucleoside analogues for single molecule detection in nucleic acids has gained significant interest due to their advantages over larger conjugated chromophores. In particular, these isomorphous "probes" exhibit sensitivity to their local environment, while minimizing perturbations to the overall native structure of nucleic acids.<sup>2</sup> Specifically, it is highly desirable to retain the Watson-Crick base pairing ability of nucleobase analogues while introducing small structural changes that enhance fluorescence. To develop suitable probes for single molecule detection studies, several factors need to be considered in their design. Most significantly, the probes need to be highly emissive, and hence should have large molar absorptivity and quantum yields. Analogues also need to be resistant to photobleaching, which is a common occurrence for chromophores that absorb in the UV range. To overcome limitations presented by poor penetration into biological tissues at the excitation wavelengths of fluorescent nucleoside analogues, multi-photon excitation in the near IR range for various analogues has been investigated by experimental approaches.<sup>144-146</sup> However, the fluorescence of the molecules studied is not adequate for single molecule detection. More recently, analogues derived from the thieno[3,4-d]pyrimidine heterocyclic nucleus,<sup>109</sup> along with several other analogues have been considered as two-photon probes.<sup>147</sup> Another class of molecules derived from the isothiazolo[4,3-d]pyrimidine moiety have also been synthesized;<sup>148</sup> however, to date only their one-photon absorption properties have been examined.<sup>149,150</sup> These isothiazole derivatives are designed to be "isofunctional" which is achieved by incorporating a basic site (addition of a nitrogen into the thiophene ring). This modification allows for hydrogen bonding at the basic site, which is observed in the natural purine nucleobases. In a recent computational study, Chawla and co-workers demonstrated that the one-photon absorption and emission spectra of the isothiazole bases are bathochromically shifted by 1.0 – 1.2 eV compared to their natural nucleobase counterparts.<sup>149</sup> This bathochromic shift was also observed for the thiophene derivatives in our previous work, see Chapter 2. In the present Chapter, computed one-photon

(OPA) and two-photon (TPA) absorption properties of 12 nucleoside analogues designed by Tor and co-workers are presented. These photophysical properties are obtained from time-dependent density functional theory (TD-DFT), and a comparison to experimental findings as well as previous theoretical investigations is provided.



**Figure 4.1:** Structures of the two-photon emissive nucleobase analogues (sugar moiety not depicted). The nucleoside structures displaying the sugar moiety and coordinates for all optimized structures are given in Appendix C

## 4.2 Computational Methods

The modified nucleobase analogues depicted in Figure 4.1, along with their corresponding nucleosides including the ribose or deoxyribose, were characterized by density functional theory computations. Where possible, X-ray crystallographic data<sup>109,148</sup> was used as a starting point for geometry optimizations. The ground state geometries of the nucleobases (excluding the ribose or deoxyribose groups) were first optimized

in the gas phase with the PBE0<sup>71-74</sup> functional and the 6-31++G(d,p)<sup>151-157</sup> basis set for all atoms. These structures were re-optimized using the same level of theory with the Polarizable Continuum Model (PCM)<sup>114,158-162</sup> to simulate solvation in water. Vibrational frequencies were computed and analyzed for all gas phase and solvated structures. The absence of imaginary frequencies confirmed that all optimized structures were true minima. An identical procedure was performed for all analogues with the relevant sugar moiety (ribose or deoxyribose) attached. The photophysical properties, including both the OPA and TPA cross sections were determined using time-dependent density functional theory (TD-DFT) utilizing linear and quadratic response theory. For the TD-DFT computations, the B3LYP<sup>68,69</sup> and CAM-B3LYP<sup>70</sup> functionals were used, employing the 6-31++G(d,p) basis set. All computations were performed with the GAMESS<sup>119,120</sup> (August 2016 version) software package.

#### 4.2.1 Computation of Two-Photon Absorption Cross Sections

The two-photon absorption (TPA) cross section can be related to experimental measurements according to<sup>163</sup>

$$\sigma^{TPA} = \frac{N\pi^2 a_0^5 \alpha \omega^2}{c \Gamma} \langle \delta^{TPA} \rangle g(2\omega, \omega_0, \Gamma). \quad (4.1)$$

Equation 4.1 reports the TPA cross section,  $\sigma^{TPA}$ , in Göppert-Mayer units (GM), which are equivalent to  $10^{-50}$  cm<sup>4</sup> s molecule<sup>-1</sup> photon<sup>-1</sup>. In this expression,  $\alpha$  is the fine structure constant, and  $a_0$  is the Bohr radius in atomic units. The speed of light,  $c$ , is in cm s<sup>-1</sup>. The photon energy is  $\omega$  whereas  $\omega_0$  represents the excitation energy, i.e. the energy difference between the initial and final states (both in atomic units).

To provide a comparison with experiment, the details concerning the parameters  $\langle \delta^{TPA} \rangle$ ,  $g(2\omega, \omega_0, \Gamma)$ ,  $N$  and  $\Gamma$  must be considered for the conversion to macroscopic units from computationally determined data, as clearly presented recently by Beerepoot et al.<sup>164</sup> (i) The rotational averaged TPA strength,  $\langle \delta^{TPA} \rangle$ , which is included in

the GAMESS response theory formalism is<sup>165–167</sup>

$$\langle \delta^{TPA} \rangle = \frac{1}{30} \sum_{ab} (F S_{aa} \bar{S}_{bb} + G S_{ab} \bar{S}_{ab} + H S_{aa} \bar{S}_{bb}) \quad (4.2)$$

where F, G and H are coefficients which depend on the polarization of the incident beam and the elements  $S_{ab}$  represent TPA transition moments between the initial and final state, defined as<sup>168,169</sup>

$$S_{ab}^{i \rightarrow f}(\omega_1, \omega_2) = \frac{1}{\hbar} \sum_{n \neq i} \left[ \frac{\langle i | \mu_a | n \rangle \langle n | \bar{\mu}_b | f \rangle}{\omega_{ni} - \omega_1} + \frac{\langle i | \mu_b | n \rangle \langle n | \bar{\mu}_a | f \rangle}{\omega_{ni} - \omega_2} \right] \quad (4.3)$$

$\langle i | \mu_a | n \rangle$  represents the transition dipole moment between the initial state  $i$  and intermediate state  $n$  in the Cartesian axis (a or b = x, y, or z), and  $\langle n | \bar{\mu}_a | f \rangle$  is the transition dipole moment between state  $n$  and final state  $f$ . For linearly polarized light (F = G = H = 2), and degenerate two-photon absorption, i.e.  $\omega_1 = \omega_2$ ,

$$\langle \delta^{TPA} \rangle = \frac{1}{15} \sum_{ab} (2 S_{ab} \bar{S}_{ab} + S_{aa} \bar{S}_{bb}). \quad (4.4)$$

(ii) A Gaussian lineshape function was chosen to account for broadening effects in solution.<sup>170</sup> The lineshape function equation is given by

$$G(2\omega) = \frac{\sqrt{\ln 2}}{\Gamma \sqrt{\pi}} \exp \left[ -\ln 2 \left( \frac{2\omega - \omega_0}{\Gamma} \right)^2 \right]. \quad (4.5)$$

Setting  $\omega = \frac{\omega_0}{2}$  and  $\Gamma$  as the HWHM, the maximum of the Gaussian function at the excitation energy,  $\omega_0$ , becomes

$$G(\omega_0) = \frac{\sqrt{\ln 2}}{\Gamma \sqrt{\pi}} \quad (4.6)$$

Hence, when this maximum is used in the equation for the lineshape function, it introduces a factor of  $\sqrt{\pi \ln 2}$  into the TPA cross section expression given by equation 4.1.

(iii) The factor of N is chosen as 4 to compare to the single-beam experiment employed by Lane et al.<sup>147</sup> (iv)  $\Gamma$  is the broadening factor. The value of  $\Gamma$  is set to 0.1 eV (HWHM), a standard damping width for chromophores in solution. Using the factors explained above the two-photon absorption cross section (in GM) is given by

$$\sigma_{TPA} = \frac{4\sqrt{\ln 2} \pi^{\frac{5}{2}} a_0^5 \alpha \omega^2}{c \Gamma} \langle \delta^{TPA} \rangle. \quad (4.7)$$

The corresponding values of  $\langle \delta^{TPA} \rangle$  in atomic units are reported in Appendix C to aide in any future comparison.

## 4.3 Results and Discussion

The optimized ground state geometries of the modified thiophene nucleosides all display an *anti* orientation at their glycosidic linkages to the ribose moiety which is consistent with the crystal structures as determined from experiment.<sup>109</sup> The isothiazole nucleosides <sup>tz</sup>**A** and <sup>tz</sup>**G** also adopt an anti orientation, while the <sup>tz</sup>**C** and <sup>tz</sup>**U** are found to be in the *syn* orientation. These findings are also in agreement with X-ray crystal structures.<sup>148</sup> Although crystal structures of the remaining four nucleosides (<sup>th</sup>5-6azaU, <sup>th</sup>5-2U, 5-furan2U and amino-7) are not available, these analogues are expected to adopt an anti orientation according to Sinkeldam et al.,<sup>171</sup> however, the computed equilibrium geometries of <sup>th</sup>5-6azaU and <sup>th</sup>5-2U display the *syn* orientation. It should also be noted that <sup>th</sup>5-6azaU, <sup>th</sup>5-2U and 5-furan2U can adopt different orientations with respect to the thiophene (or furan) ring to the nitrogenous six-membered ring. The computed ground state geometry of 5-furan2U depicts the oxygen atom of the furan ring pointing away from the carbonyl of the nitrogenous six-membered ring. This is also observed in <sup>th</sup>5-2U with the sulfur atom of the thiophene ring pointing away from the six-membered nitrogenous ring. Conversely, the sulfur atom of the thiophene in <sup>th</sup>5-6azaU faces away from the six-membered ring. While this work is limited to the conformations outlined above, other possible conformations with respect to both glycosidic linkage and ring orientation are currently being investigated.

### 4.3.1 Photophysical Properties

The excitation energies ( $\omega_0$ ), oscillator strengths ( $f$ ), and two-photon cross sections ( $\sigma^{TPA}$ ) for the first singlet excited state, S<sub>1</sub>, (in solution) for both the nucleobases and corresponding nucleosides of the 12 RNA analogues illustrated in Figure 4.1 are summarized in Table 4.1. The previously reported values of TPA energies and cross sections from Samanta and Pati<sup>172</sup> are also provided. The results reported by Samanta and Pati were computed at the B3LYP/6-31++G(d,p) level of theory with implicit solvation using PCM (with parameters for water) utilizing the DALTON 2013

software package.<sup>173</sup> It should be noted that the geometry optimizations and some of the TD-DFT results in the present work are computed at the same level of theory as Samanta and Pati; however, we have used the GAMESS quantum chemical program for our investigations. Since the previously reported results do not provide the exact details for the conversion to macroscopic units for the TPA cross sections, nor provide coordinates for the optimized nucleoside structures, we cannot directly compare the two computational approaches. However, we do note that the present results are of the correct magnitude when compared to the experimental measurements. The photophysical properties of the nucleobases are compared to the nucleosides to ascertain the effect of the inclusion of the ribose (or deoxyribose) group. For comparison, the experimentally determined absorption maxima and two-photon cross sections<sup>147</sup> are also given in Table 4.1.

The computed TPA cross sections are within the range of 0.2 – 4.9 GM, while the experimental  $\sigma^{TPA}$  range is from 0.2 – 7.6 GM. The largest deviation from the experimental results is seen with the *th*5-2U analogue, which has a significantly lower computed  $\sigma^{TPA}$  by a factor of 3.5. With the exception of *th*5-2U and *th*5-6azaU (which has a slightly lower TPA cross section), the computed  $\sigma^{TPA}$  from TD-DFT were higher compared to the experimental values (1 – 5 GM deviation). The inclusion of the ribose or deoxyribose group (sugar moiety) does not significantly alter (in terms of absolute change in GM) the cross sections of these analogues; relative, or percentage, change can be quite large due to the generally small values. The largest change in  $\sigma^{TPA}$ , when comparing the nucleobase to the nucleoside structure is seen in 5-furan2U with a difference of 0.9 GM (25% increase). Unless otherwise stated, the remainder of the discussion will consider the nucleoside results when referring to the analogues.

### Charge Transfer and Higher Lying States

The use of a long-range corrected functional, CAM-B3LYP, has been suggested for use in the determination of TPA;<sup>164,174,175</sup> however, CAM-B3LYP did not provide significantly different TPA cross sections for these set of molecules, see Table 4.2 which reports OPA and TPA properties that can be compared to those in Table

**Table 4.1:** Excitation energies (in eV), oscillator strengths ( $f$ ) and TPA cross sections ( $\sigma^{TPA}$  in GM) of the modified nucleobases and nucleosides computed with B3LYP/6-31++G(d,p) in water (PCM)

	This work				Ref <sup>172,a</sup>		Expt <sup>147</sup>	
	Nucleobase		Ribonucleoside		$\omega_0$	$\sigma^{TPA}$	$E_{max}^{b,c}$	$\sigma^{TPA}$
	$\omega_0$ ( $f$ )	$\sigma^{TPA}$	$\omega_0$ ( $f$ )	$\sigma^{TPA}$	$\omega_0$	$\sigma^{TPA}$		
<i>th</i> 5-6azaU	3.48 (0.291)	3.7	3.36 (0.380)	3.5	3.30	31.3	3.73	3.8
<i>th</i> 5-2U	3.82 (0.272)	2.4	3.76 (0.308)	2.2	3.64	18.2	3.95	7.6
5-furan2U	3.76 (0.279)	4.0	3.67 (0.344)	4.9	3.75	11.7	3.92	2.1
amino-7	4.29 (0.264)	2.3	4.27 (0.293)	3.1	4.10	72.6	3.92	1.8
<i>th</i> A	3.78 (0.162)	0.2	3.65 (0.202)	0.3	3.49	4.0	-	-
<i>th</i> C	3.88 (0.105)	0.5	3.89 (0.110)	1.3	3.78	11.3	-	-
<i>th</i> G	3.78 (0.094)	1.0	3.67 (0.133)	1.7	3.57	16.9	-	-
<i>th</i> U	4.10 (0.082)	0.7	4.12 (0.088)	1.4	4.01	12.8	4.08	0.2
<i>tz</i> A	3.80 (0.174)	0.6	3.75 (0.248)	0.4	-	-	-	-
<i>tz</i> C	3.87 (0.175)	0.7	3.84 (0.141)	1.0	-	-	-	-
<i>tz</i> G	3.69 (0.123)	1.3	3.70 (0.176)	1.9	-	-	-	-
<i>tz</i> U	4.08 (0.149)	0.8	4.07 (0.116)	1.6	-	-	-	-

<sup>a</sup> Values for the nucleoside as computed using B3LYP/6-31++G(d,p) in Water(PCM); see the main text for further discussion of discrepancies with the present work.

<sup>b</sup> Experimental TPA values are obtained from 690nm excitation, see Ref.<sup>147</sup>

<sup>c</sup> The maximum absorption peak determined from experiment.

4.1. Furthermore, the excitation energies are overestimated in comparison to experimental values, by as much as 0.5 eV. To assess whether CAM-B3LYP can provide more accurate OPA and TPA properties, we have also taken into consideration the charge transfer character of the low-lying electronic states by analyzing the lambda diagnostic<sup>118</sup> ( $\Lambda$ ) at the CAM-B3LYP/6-31++G(d,p) level of theory in GAMESS. The  $\Lambda$  parameter is a measure of the spatial overlap between the virtual and occupied orbitals for a given excitation.  $\Lambda$  values range from 0 to 1 with larger values representing maximal overlap between orbitals which signifies short-range or local

excitation. Long-range excitations such as charge transfer (CT) and Rydberg states typically have lower values of  $\Lambda$ . It should be noted that charge transfer states can have  $\Lambda$  values as high as 0.7 and a thorough analysis of the molecular orbitals involved must be performed. In this work, orbitals involved in the transitions were rendered and considered along with  $\Lambda$  for the assignment CT and Rydberg states (see Figures C4-C11 in Appendix C). The  $\Lambda$  parameters for low-lying ( $S_1$ - $S_5$ ) states along with their relative OPA and TPA properties of four nucleobase analogues are tabulated in Table 4.2. Analysis of this data reveals the presence of Rydberg excitations for 5-furan2U and 7-amino which occur at energies  $\geq 5$  eV and have characteristically low cross sections. The majority of transitions which exhibit charge transfer also have significantly low TPA cross sections; most likely due to the correspondingly low one-photon oscillator strengths (transition dipole moments). The exception to this trend is seen for the  $S_5$  states in *th*5-6azaU, *th*5-2U and 5-furan2U which have cross sections of 109, 20, 93 GM, respectively. While these values represent a significant improvement relative to the corresponding values to  $S_1$ , the energies are not experimentally useful (also, the veracity of the TD-DFT results may be questionable for these higher energy excitations). The analogues, *th*A and *th*U exhibit CT character for  $S_2$ , while *th*C and *th*G for the  $S_3$  excited state. These results are consistent with previously determined  $\Lambda$  values for the nucleobase structures using a larger basis set.<sup>130</sup> High lying Rydberg states are also observed for all four of these analogues and the remaining transitions are local excitations. The isothiazole analogues *tz*C, *tz*G also display charge transfer for  $S_2$  and  $S_4$ . For *tz*U, only the  $S_2$  state is observed as charge transfer type. No Rydberg type excitations are seen for the thiazole analogues. All observed CT transitions have  $\Lambda$  values  $< 0.5$ , and Rydberg transitions with values less than 0.3. Excitation parameters for the thiophene and isothiazole analogues can be found in Appendix C, Tables C1 and C2.

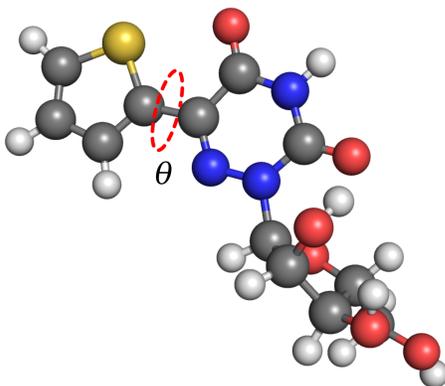
**Table 4.2:** Excitation energies (in eV), oscillator strengths ( $f$ ),  $\Lambda$  diagnostic<sup>118</sup> and TPA cross sections ( $\sigma^{TPA}$  in GM) of selected singlet states of *th*5-6azaU, *th*5-2U, 5-furan2U and 7-amino nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water(PCM). Also provided is the dominant character of the transition.

Nucleoside	$S_n$	Energy	$f$	$\Lambda$ Diagnostic	Transition	Type	$\sigma^{TPA}$
<b><i>th</i>5-6azaU</b>	$S_1$	3.73	0.485	0.70	H $\rightarrow$ L	L	3.9
	$S_2$	4.56	0.003	0.51	H-5 $\rightarrow$ L	CT	0.8
	$S_3$	4.67	0.039	0.41	H-1 $\rightarrow$ L	CT	3.9
	$S_4$	5.24	0.159	0.62	H $\rightarrow$ L+1	L	0.8
	$S_5$	5.45	0.060	0.47	H-6 $\rightarrow$ L	CT	109
<b><i>th</i>5-2U</b>	$S_1$	4.15	0.485	0.75	H $\rightarrow$ L	L	1.0
	$S_2$	5.06	0.003	0.66	H $\rightarrow$ L+1	L	3.5
	$S_3$	5.21	0.039	0.42	H-1 $\rightarrow$ L + H-6 $\rightarrow$ L	CT	0.7
	$S_4$	5.32	0.045	0.41	H-1 $\rightarrow$ L + H-5 $\rightarrow$ L	CT	0.9
	$S_5$	5.69	0.028	0.35	H-2 $\rightarrow$ L	CT	20.1
<b>5-furan2U</b>	$S_1$	4.06	0.473	0.73	H $\rightarrow$ L	L	3.7
	$S_2$	5.06	0.368	0.54	H $\rightarrow$ L +2	L	1.3
	$S_3$	5.21	0.001	0.49	H-4 $\rightarrow$ L	CT	0.0
	$S_4$	5.32	0.009	0.32	H $\rightarrow$ L+1	R	1.3
	$S_5$	5.91	0.072	0.48	H-1 $\rightarrow$ L	CT	92.7
<b>7-amino</b>	$S_1$	4.54	0.324	0.67	H $\rightarrow$ L	L	3.1
	$S_2$	4.83	0.125	0.59	H $\rightarrow$ L+2	L	3.0
	$S_3$	5.33	0.005	0.39	H-3 $\rightarrow$ L	R	0.0
	$S_4$	5.45	0.363	0.58	H-1 $\rightarrow$ L	L	0.8
	$S_5$	5.56	0.001	0.26	H $\rightarrow$ L+1	R	0.7

Types of transitions: R=Rydberg, CT= Charge Transfer, L=Local Orbitals; H= HOMO, L=LUMO

## Rotation Studies

To assess how the flexibility of the thiophene or furan ring can affect TPA cross sections, an investigation of rotation about the bond connecting the two rings was undertaken. Conformers of *th*5-6azaU, *th*5-2U and 5-furan2U nucleosides were generated by varying the angle between the five membered (thiophene or furan) ring connecting to the six-membered (diazine or triazine) ring by  $2^\circ$  in the range of  $0^\circ$  to  $360^\circ$ . All other atomic bond lengths and bond angles were fixed at their equilibrium values. This rotation for the *th*5-6azaU nucleoside is depicted in Figure 4.2. The excitation parameters for all conformers (180 for each nucleoside) were computed with the same level of theory as previously, i.e., B3LYP/6-31++G(d,p) with PCM solvation in water. The dependence of the TPA cross section on this twist angle,  $\theta$ , is presented in Figure 4.3 for these three nucleosides. Also provided are the corresponding electronic energies (in kJ/mol) relative to the equilibrium structure. It should be noted that the relative energy differences represent an upper bound for the barrier to rotation since geometries of conformers were not allowed to be "relaxed" (the remaining parameters were not optimized).



**Figure 4.2:** Rotation around the twist angle,  $\theta$ , of *th*5-6azaU

In general, the TPA cross sections for  $\theta$  values from  $0^\circ$  to  $180^\circ$  are symmetric to those from  $180^\circ$  to  $360^\circ$ . Hence the discussion will be limited to the  $0^\circ - 180^\circ$  range for  $\theta$ . The maximum TPA cross sections are obtained at  $\theta$  values of  $126^\circ$  for both *th*5-6azaU and *th*5-2U, and at  $128^\circ$  for 5-furan2U, see Table 4.3. Planar conformers ( $0^\circ$

or 180°) of *th*5-6azaU and 5-furan2U result in  $\sigma^{TPA}$  values in agreement to equilibrium values, since equilibrium geometries are planar themselves. The oscillator strengths of all three nucleosides show a decrease as the angles deviate from planarity, reaching a minimum at  $\theta=90^\circ$ , see Figure 4.4 which illustrates the oscillator strength as a function of  $\theta$ . The reverse trend is observed for the oscillator strengths with  $\theta$  values from 92°–180° as the structure becomes planar again. Relative ground state energies, vertical excitation energies, oscillator strengths, and TPA cross sections for all conformers at their corresponding  $\theta$  values, are tabulated in Tables C10-C12 of Appendix C.

**Table 4.3:** *Maximum TPA cross sections obtained from rotation of  $\theta$  and the corresponding relative energies ( $\Delta E$ ), excitation energies ( $\omega_0$ ) and oscillator strengths ( $f$ ).*

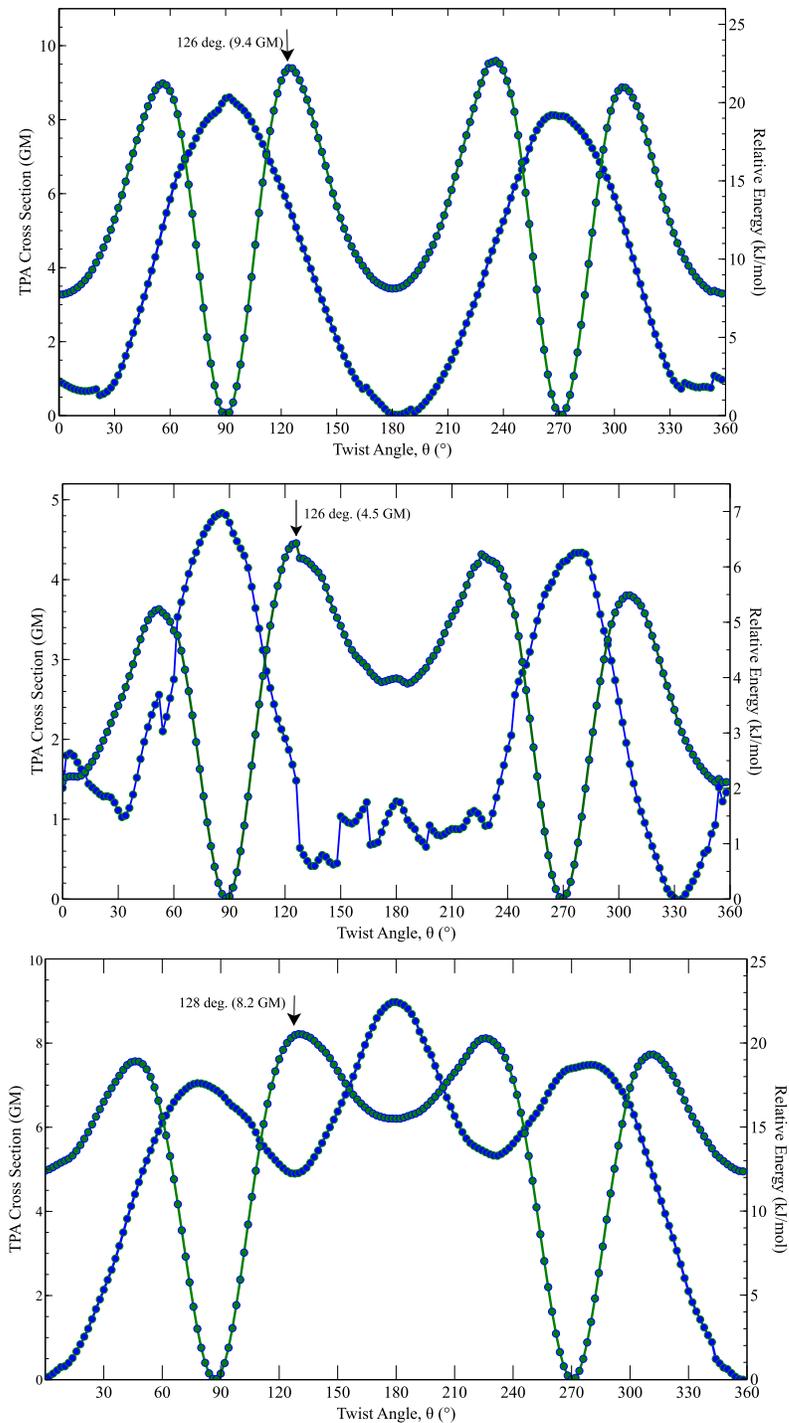
Nucleoside	$\theta$ (°)	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)
<i>th</i> 5-6azaU	126	12.8	3.46	0.188	9.4
<i>th</i> 5-2U	126	2.1	4.00	0.217	4.5
5-furan2U	128	12.2	3.82	0.196	8.2

The conformers of *th*5-6azaU and 5-furan2U display a relatively narrow range of excitation energies (within 0.1 eV and 0.3 eV respectively), while the conformers *th*5-2U span a much larger range with energy differences up to 0.5 eV. Nevertheless, these differences in excitation energies do not significantly alter the calculated TPA cross section values. Similar values of  $\sigma^{TPA}$  are obtained for the different conformers even if a fixed excitation energy is used in the calculation. It is evident that the TPA trend is driven by the TPA strength (denoted as Delta), see Figure 4.4. The  $\theta$  values ranging between 128°–148° for *th*5-2U result in significantly higher ranges of TPA cross sections (3.5 – 4.3 GM) in comparison to the optimized structure with corresponding  $\sigma^{TPA}$  of 2.5 GM. Conformers which have the smallest relative energies in reference to the equilibrium structure (within 1 kJ/mol), all display large TPA cross sections and excitation energies from 3.8–4.0 eV which are in excellent agreement with

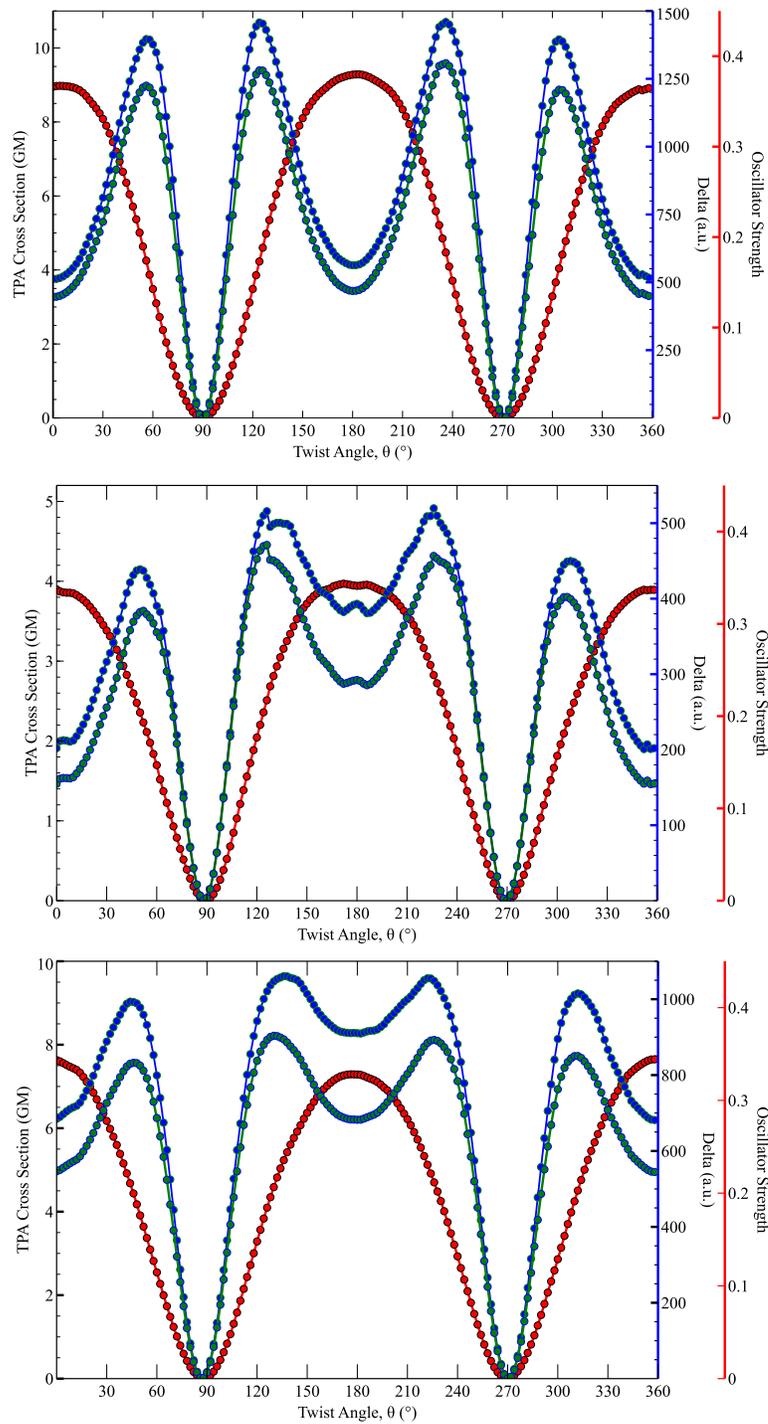
the experimental absorption maximum at 3.92 eV. In fact, all conformers of *th*5-2U have relative energies  $\leq 7$  kJ/mol suggesting that rotation about  $\theta$  readily occurs in solution. This finding suggests that our previously reported theoretical TPA cross section of *th*5-2U may be a lower limit, and shed insight into the discrepancy between theoretical and experimental values where the computational result at equilibrium underestimates (by a factor of 3.5), the experimentally measured value. In contrast, the relative ground state energies of the conformers of *th*5-6azaU and 5-furan2U span a much larger range, with values up to 22 kJ/mol, suggesting that rotation does not occur as readily for these nucleosides.

## 4.4 Conclusions

The two-photon absorption properties of 12 nucleoside analogues were investigated with TD-DFT employing the B3LYP and CAM-B3LYP functionals. The present results provide good agreement with 5 nucleosides that have been measured. We also provide the first predictions for the TPA cross sections for the isothiazole family of analogues. Regretfully, the predicted TPAs are low ( $< 2\text{GM}$ ) relative to the other analogues, see Table 4.1. The use of a long-range corrected functional, CAM-B3LYP does not offer any improvement on the computed TPA cross sections to  $S_1$ , and also gives overestimated excitation energies in comparison to experiment. In addition, there was no significant effect of the inclusion of the ribose (sugar) moiety on the TPA cross sections, this suggests future computational studies of these systems, e.g., in dimers or oligomers, may focus on the nucleobase. Furthermore, an investigation of the TPA cross section dependence on the rotation of the bond connecting the two rings ( $\theta$ ) for *th*5-6azaU, *th*5-2U and 5-furan2U nucleosides was performed. Analysis of the results obtained in this investigation reveal a strong dependence of  $\sigma^{TPA}$  on the orientation of the rings. Moreover, the relatively small differences in the energies of the conformers of *th*5-2U suggest a low rotational barrier for this nucleoside in solution and provides a possible explanation for the deviation from experimental TPA cross sections.



**Figure 4.3:** TPA cross sections (green) to  $S_1$  as a function of rotation around the twist angle,  $\theta$ , and relative ground state energies (blue) for (a)  ${}^{th}5\text{-}6\text{aza}U$ , (b)  ${}^{th}5\text{-}2U$  and (c)  $5\text{-furan}2U$ . The numerical values for the relative ground state energies, vertical excitation energies, oscillator strengths, and TPA cross sections are given in Appendix C.



**Figure 4.4:** TPA cross sections (green), TPA strength, delta (blue), and oscillator strengths (red) of  $S_1$  as a function of rotation of the twist angle,  $\theta$ , for <sup>th</sup>5-6azaU, <sup>th</sup>5-6azaU and 5-furan2U.

## Chapter 5

# Base-pairing Interactions of Thiophene-Derived RNA Nucleobases

*A subtle thought that is in error may yet give rise to fruitful inquiry that can establish truths of great value.*

– Isaac Asimov

## 5.1 Introduction

Ribonucleic acid has an essential role in biological processes and has emerged as a target for the control of cell function and study of ligand interactions.<sup>78,79</sup> However, the nucleobase constituents of RNA, which are the primary chromophores, are non-emissive, have ultrafast excited state lifetimes and low quantum yields.<sup>39,80–82</sup>

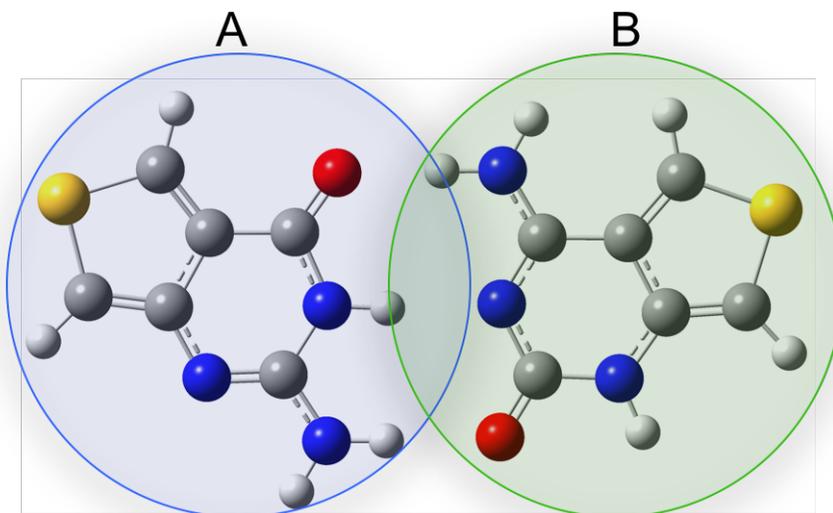
The excited-state properties of the nucleobase constituents of RNA and DNA have been extensively studied by both theoretical and experimental approaches (see Refs.<sup>39,80–82</sup> for recent theoretical and experimental reviews). The inherent non-emissive nature of the naturally occurring nucleobases has driven the investigation into emissive analogues to the nucleobases. Emissive analogues of nucleobases have significant implications for the field of biotechnology when utilized as molecular probes.<sup>2,57,102</sup> Design of synthetic probes requires several criteria to be successfully implemented in biophysical assays. In general, the modified analogues need to achieve isomorphism with respect to their natural counterparts and thus display minimal structural and functional perturbations upon modification.<sup>2,61</sup> While structural and functional perturbation is an inevitable consequence of such modifications, the goal is to minimize these perturbations and retain the structural requirements necessary for base pairing and helix formation in the nucleic acid. It is therefore no surprise that the majority of the synthetic analogues designed show strong resemblance to the natural nucleobases.

This chapter will focus on base pairing interactions of RNA nucleobase analogues derived from thieno[3,4-d]-pyrimidine as the heterocyclic nucleus designed and characterized by Shin *et al.*<sup>109</sup> We limit this work to the isolated nucleobases and exclude the ribose sugar. Since the computational cost scales with system size, analysis of nucleosides presents a challenge to study with the desired accuracy. Our previous work on the excited state properties of these analogues has demonstrated that the ribose group does not introduce a significant difference to the computed excitation energies,<sup>130</sup> see Chapter 2. Hence, the findings obtained from this study neglecting the ribose are expected to convey accurate results with respect to excitation energies,

and the presence of the ribose has little effect in the base-pairing. Watson-Crick base pairing interactions of both modified-modified and modified-natural nucleobases are determined in the present study and their relevant structures are illustrated in Figure 5.2. During the course of this research, a similar study appeared in the literature<sup>176</sup> and we provide comparison to this study where appropriate. The binding energies determined in Samanta and Pati's work<sup>176</sup> employ the same DFT functionals and basis set; however in this chapter, the dispersion corrections as well as an in-depth analysis on the choice of functional is provided. We also report AIM results to shed light on the strength of the hydrogen bonding between the nucleobases (see Section 5.3.3). Furthermore, the photophysical properties of all base-pairing schemes are provided and compared to their constituents (nucleobases) in the present work. In contrast, Samanta and co-workers have only provided computed absorption data for the modified-modified base pairs in a previous publication.<sup>121</sup> Since we are examining the interaction of monomers within a complex, we need to consider the relevant errors that may arise from quantum chemical computations on these types of systems. One type of error that needs to be accounted and corrected for is the basis set superposition error (BSSE). The following section gives a brief overview of BSSE.

### 5.1.1 Basis Set Superposition Error

Basis set superposition error<sup>177–179</sup> arises from the inconsistent treatment of monomers within a complex and causes artificial shortening of intermolecular distances and incidental strengthening of these interactions.



**Figure 5.1:** Representation of the basis set superposition error for the  ${}^{\text{th}}G$ - ${}^{\text{th}}C$  base pair. Monomer A refers to  ${}^{\text{th}}G$ , and monomer B is  ${}^{\text{th}}C$ .

Figure 5.1 depicts the interaction of basis functions of two monomers in a complex. As monomer A approaches monomer B, the complex/dimer is artificially stabilized. This is achieved due to monomer A utilizing extra basis functions from monomer B to describe its electron distribution. Hence, the monomers can access additional functions from each other at short intermolecular distances. It is important to note that BSSE errors are more significant at short intermolecular distances and with smaller basis sets. Moreover, the correction for BSSE does not account for the basis set incompleteness error (BSIE). BSIE is the result of basis set truncation and can only be eliminated in the limit of a complete basis set. In fact, at this limit, both BSSE and BSIE would be equal to zero.

### 5.1.2 Interaction Energies

The interaction of two monomers is calculated by subtracting the energies of the individual monomers from the energy of the dimer,

$$\Delta E_{\text{int}}(AB) = E_{AB}^{AB}(AB) - E_A^A(A) - E_B^B(B). \quad (5.1)$$

Since we need to include a correction for the BSSE, we need account for the differences in energies of the monomers in the dimer geometry. The energies of the monomers in the dimer geometry are represented by the  $E_{AB}^A(A)$  and  $E_{AB}^B(B)$  terms.  $E_{int}^{CP}$  is referred to as the counterpoise<sup>180</sup> corrected interaction energy,

$$\Delta E_{int}^{CP}(AB) = E_{AB}^{AB} - E_A^A(A) - E_B^B(B) - E_{AB}^{AB}(A) + E_{AB}^A(A) - E_{AB}^{AB}(B) + E_{AB}^B(B). \quad (5.2)$$

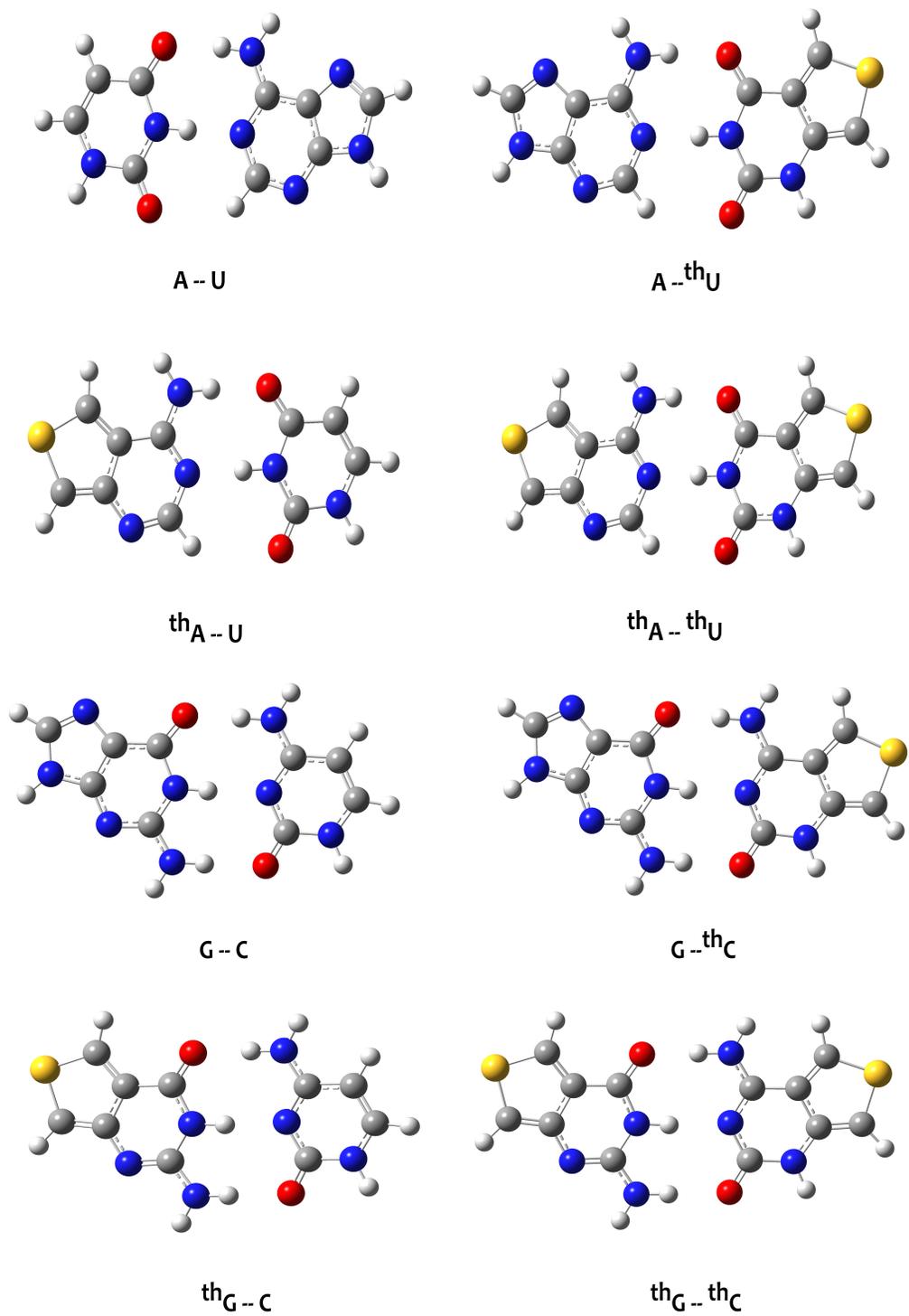
It should be noted that in equation 5.2, the superscript denotes the basis set used, the subscript indicates the geometry and the terms in brackets provide the system considered. The above expression includes the energy required to deform each of the monomers from their equilibrium geometries to their geometries within the dimer. This is referred to as the deformation energy, and equation 5.2 can be simplified to yield

$$\Delta E_{bind}^{CP}(AB) = E_{int}^{CP}(AB) + E_{def}^A(A) + E_{def}^B(B). \quad (5.3)$$

To account for errors as a result of basis set superposition, we can use counterpoise (CP) correction,<sup>181</sup> which Gaussian has implemented as a procedure. This method places all basis functions of one monomer on the atomic centre of that monomer while neglecting electrons and nuclear charges. The functions on this monomer are referred to as ghost functions or ghost atoms. This procedure is then repeated for the other monomer. For the counterpoise procedure, the number of fragments needs to be specified, and thus in theory one can use counterpoise for trimers or even larger complexes. From the counterpoise computation, we are able to extract the corrected energies for the dimer geometry. Thus, geometry optimizations for the individual monomers must also be performed to determine the deformation energy. In addition, a correction for the zero-point vibrational energies must be applied to obtain accurate interaction energies. The ZPE correction is obtained from the vibrational frequency computation (this is conventionally done after geometry optimizations to confirm that the structure is a true minima on the potential energy surface). This rather standard approach to obtain accurate binding energies of complexes<sup>182-185</sup> is utilized in this chapter.

## 5.2 Computational Methods

The thiophene substituted nucleobase analogues ( $^{th}A$ ,  $^{th}C$ ,  $^{th}G$ , and  $^{th}U$ ) of RNA were characterized by density functional quantum chemical computations to assess the base pairing interactions. The natural and modified nucleobases in their relevant base pairing schemes are depicted in Figure 5.2. X-ray crystallographic data<sup>109</sup> was used as a starting geometry, with the removal of the ribose moiety, which was replaced with a hydrogen atom. Ground state geometries were determined with the 6-31++G(d,p) basis set<sup>151-157</sup> for all atoms, employing the B3LYP,<sup>68,69</sup> M06-2X<sup>77</sup> and  $\omega$ B97xD<sup>75,76</sup> functionals in the gas phase. Additionally, density-independent dispersion (Grimme's D3) correction was added to the B3LYP and M06-2X functionals to assess the effect of long-range dispersion<sup>186</sup> (denoted as B3LYP-D3 and M06-2X-D3, respectively). The geometry optimizations were performed with the Gaussian 09 software package<sup>139</sup> using "Tight" convergence criteria, i.e., Maximum Force =  $1.5 \times 10^{-5}$  a.u., RMS Force =  $1.0 \times 10^{-5}$  a.u., Max displacement =  $6.0 \times 10^{-5}$  a.u., and RMS displacement =  $4.0 \times 10^{-5}$  a.u. The grid used for numerical integration in DFT is set to "Ultrafine". The vibrational modes were analyzed at the same level of theory to verify that the optimized structures were true minima on the potential energy surface. To account for basis set superposition error (BSSE), counterpoise correction<sup>177,181</sup> was performed to obtain the interaction energies for all base pair complexes. In addition to BSSE, further corrections were undertaken to account for the deformation energy (as discussed in the introduction), as well as the zero-point vibrational energy (ZPVE). Bader's quantum theory of atoms in molecules (QTAIM) is also utilized to investigate the hydrogen bonding profiles of the base pairs.<sup>187</sup> The topology of the electron density,  $\rho$ , is analyzed using the program package AIMAll<sup>188</sup> with default options. QTAIM analyses were carried out using wave-functions obtained at the  $\omega$ B97xD/6-31++G(d,p) level of theory. We also report the one-photon absorption properties of the modified-modified and modified-natural base pairs at the B3LYP/6-31++g(d,p) level of theory in both gas phase and in water utilizing the PCM<sup>113,114</sup> method.



**Figure 5.2:** Watson-Crick base pairing of the native and modified RNA nucleobases. Atomic representations: S=yellow, N=blue, O=red, C=gray and H=white. Cartesian coordinates are available in Appendix D.

## 5.3 Results and Discussion

### 5.3.1 Structural Characteristics of the Watson-Crick Base pairs

In general, the optimized structures of the natural and modified base pairs adopt a planar geometry. We have computed all the relevant base pairs depicted in Figure 5.2 in both  $C_s$  (planar) and  $C_1$  symmetry with B3LYP/6-31++G(d,p) and have used the lowest energy conformation for subsequent optimizations. The only base pairs which slightly break planarity are A-<sup>th</sup>U and <sup>th</sup>A-<sup>th</sup>U. It should be noted that the energy differences between  $C_1$  and  $C_s$  symmetry structures are negligible ( $\leq 0.005$  kcal/mol) for all the base pairs.

The intermolecular (H-bond acceptor-donor) distances computed at the  $\omega$ B97XD/6-31++G(d,p) level of theory are compared to natural base pairing in RNA in Table 5.1. The H-bond acceptor-donor distances are denoted as  $d_{NHO}$  and  $d_{NHN}$  for N-H  $\cdots$  O and N  $\cdots$  H-N, respectively, in A-U pairs. For the G-C pairs which have three H-bonds,  $d_{NHO}$ ,  $d_{NHN}$  and  $d_{OHN}$  correspond to the N-H  $\cdots$  O, N-H  $\cdots$  N and O  $\cdots$  H-N distances, respectively. In general, there is good agreement of intermolecular distances of the natural base pairs with respect to experiment. For dimers containing modified bases (either one or both), there is insignificant structural change (at least based on H-bond distances) compared to their natural counterparts with the use of all functionals, particularly with the inclusion of dispersion corrections. The full results, determined with the B3LYP, B3LYP-D3, M06-2X, and M06-2X-D3 functionals can be found in Appendix D. H-bond donor-acceptor distances of the modified-natural base pairs tend to be slightly larger in comparison the natural pairs. It is also apparent that the differences in the distances are less when we compare the modified-modified with the natural-natural pairs. The largest difference in distance of the modified-modified pairs in comparison to the natural-natural base pairs is 0.02 Å, while modified-natural pairs show differences up to 0.04 Å. Thus, there is greater deviation in base-pairing in the mixed natural-modified pairs. The computed intermolecular distances in comparison to the experimental findings for the natural-natural base pairs<sup>189</sup> also yield

good agreement, with the largest deviation of 0.08 Å, which is observed for N-H ... O of G-C pairs.

**Table 5.1:** Intermolecular distances (in Å) of the natural and modified nucleobases calculated with  $\omega$ B97xD/6-31++G(d,p) in the gas phase.

Base pair	$d_{NHO}$	$d_{NHN}$	$d_{OHN}$	$\Delta d$		
<i>Experimental</i>	$d_{NHO} = 2.93$		$d_{NHN} = 2.85$			
<b>A-U</b>	2.941	2.830	NA	(0.011)	(-0.020)	
<b>A-<sup>th</sup>U</b>	2.965	2.832	NA	0.024	0.002	
<b><sup>th</sup>A-U</b>	2.905	2.830	NA	-0.036	0.000	
<b><sup>th</sup>A-<sup>th</sup>U</b>	2.929	2.832	NA	-0.012	0.002	
<i>Experimental</i>	$d_{NHO} = 2.86$		$d_{NHN} = 2.95$		$d_{OHN} = 2.91$	
<b>G-C</b>	2.785	2.928	2.920	(-0.075)	(-0.022)	(0.010)
<b>G-<sup>th</sup>C</b>	2.759	2.935	2.903	-0.026	0.007	-0.017
<b><sup>th</sup>G-C</b>	2.827	2.934	2.940	0.042	0.006	0.020
<b><sup>th</sup>G-<sup>th</sup>C</b>	2.799	2.942	2.926	0.014	0.014	0.006

$\Delta d$  values represent the differences in computed distances between the natural-natural base pairs and the modified pairs, i.e.,  $d_{natural} - d_{modified}$ . Values in brackets are differences in computed and experimental distances from Refs.<sup>189,190</sup> for the natural-natural pairs.

### 5.3.2 Binding Energies

The ZPE and BSSE corrected binding energies of the Watson-Crick (WC) base pairs computed with all functionals are summarized in Table 5.2. As expected, B3LYP underbinds relative to the other functionals; however, the relative energies of the dimers are well reproduced. Further discussions focus on the results determined using the  $\omega$ B97XD functional to demonstrate the effects of both BSSE and ZPE corrections (see Table 5.3); however, all computed energies with their corresponding corrections can be found in Tables D1–D5 in Appendix D. Analysis of the correction terms to the binding energies reveal that the zero point energy correction is significant, accounting for approximately 7-10% of the total binding energy. The computed BSSE energies contribute slightly less than the ZPE, at approximately 5% of  $\Delta E_{bind}$ . Thus, it is

apparent that both of these corrections are essential to obtain accurate energies. The energy difference between modified pairs and natural pairs is referred to as the modification energy, or  $\Delta E_{mod}$  in the context of this chapter. Positive values of  $\Delta E_{mod}$  reflect a decrease in stability relative to the natural pairs, and negative values an increase in stability. The  $^{th}G-C$  and  $^{th}G-^{th}C$  pairs show more destabilization with values of 2.58 kcal/mol and 2.01 kcal/mol, respectively. The interaction strength of the  $^{th}A-^{th}U$  is remarkably close to the A-U pair, with a difference in energy ( $\Delta E_{mod}$ ) of only -0.11 kcal/mol. These findings with respect to the relative energies are also consistent with previous theoretical work done by Samanta and Pati,<sup>176</sup> save for the G- $^{th}C$  base pair. It should be noted that the structures of the base pairs in their work differ by the choice of a methyl group for capping the nucleobase at the ribose attachment rather than the H-capping method used here. The modification energies obtained by both studies are also provided in Table 5.3. Our findings demonstrate that the modified nucleobases are able to form WC base pairs with their natural counterparts with modest (0–15%) stabilization/destabilization.

**Table 5.2:** *BSSE and ZPE corrected binding energies in (kcal/mol) of the modified base pairs computed with B3LYP, M06-2X, (+D3 corrections) and  $\omega$ B97XD and the 6-31++G(d,p) basis set in the gas phase.*

Base Pair	B3LYP	B3LYP +D3	M06-2X	M06-2X +D3	$\omega$ B97XD	$\Delta E_{mod}$
A-U	-10.81	-15.02	-13.25	-13.72	-14.28	
A- $^{th}U$	-10.43	-14.69	-12.95	-13.70	-13.89	+ (0.0–0.4)
$^{th}A-U$	-11.87	-16.08	-14.16	-14.88	-15.33	- (1.0–1.2)
$^{th}A-^{th}U$	-11.23	-15.63	-13.61	-14.64	-14.75	- (0.4–0.9)
G-C	-23.46	-28.35	-25.93	-26.47	-27.58	
G- $^{th}C$	-24.26	-29.25	-26.81	-27.40	-28.55	- (0.8–1.0)
$^{th}G-C$	-21.26	-26.12	-23.61	-23.19	-25.21	+ (2.2–3.3)
$^{th}G-^{th}C$	-21.90	-26.87	-24.29	-24.90	-26.07	+ (1.5–1.6)

Positive values (+) indicate destabilization and negative values (-) stabilization with respect to the natural base pairs.

**Table 5.3:** Binding energies in (kcal/mol) of the modified base pairs calculated with  $\omega$ B97XD/6-31++G(d,p) in gas phase

Base pair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$ +BSSE	ZPE	$\Delta E_{bind}$ +BSSE+ZPE	$\Delta E_{mod}$	$\Delta E_{mod}^a$ (Ref <sup>176</sup> )
<b>A-U</b>	-16.36	0.71	-15.65	1.37	-14.28		
<b>A-<sup>th</sup>U</b>	-15.89	0.72	-15.17	1.27	-13.89	+0.47	+0.35
<b><sup>th</sup>A-U</b>	-17.09	0.74	-16.35	1.02	-15.33	-0.73	-0.89
<b><sup>th</sup>A-<sup>th</sup>U</b>	-16.47	0.75	-15.72	0.72	-14.75	-0.11	0.00
<b>G-C</b>	-30.47	0.98	-29.48	1.90	-27.58		
<b>G-<sup>th</sup>C</b>	-31.21	0.99	-30.22	1.66	-28.55	-0.74	+0.27
<b><sup>th</sup>G-C</b>	-27.89	0.97	-26.92	1.71	-25.21	+2.58	+3.52
<b><sup>th</sup>G-<sup>th</sup>C</b>	-28.46	0.97	-27.49	1.41	-26.07	+2.01	+3.10

<sup>a</sup> Computed at the  $\omega$ B97XD/6-31++G(d,p) level of theory.

### 5.3.3 AIM Results

At each bond critical point (BCP) in the hydrogen bonding region, the following characteristics were considered: the electron density of the BCP ( $\rho_{BCP}$ ), its Laplacian ( $\nabla^2\rho_{BCP}$ ), and the total electron energy density ( $H_{BCP}$ ), which is a sum of the potential electron energy density ( $V_{BCP}$ ) and the kinetic electron energy density ( $G_{BCP}$ ). These parameters are related to one another according to the following equations:

$$(1/4)\nabla^2\rho_{BCP} = 2G_{BCP} + V_{BCP} \quad (5.4)$$

$$H_{BCP} = G_{BCP} + V_{BCP}. \quad (5.5)$$

The Laplacian,  $\nabla^2\rho$ , allows for identifying spatial changes of electron concentrations (or depletions) rather than the electron density itself. The sign of the Laplacian determines the nature of the bond (shared or closed-shell interaction). The criteria for studying H-bonding interactions were proposed by Koch and Popelier.<sup>191</sup> These include the presence of bond critical points (BCPs), a bond path between hydrogen

donor and acceptor, as well as the positive value of the Laplacian of the electron density at this BCP. All of these criteria were met with the HBCPs analyzed, and the relevant parameters are tabulated in Table 5.4. In addition to these general criteria, Popelier reported that H-bonds should have a relatively high value of the electron density, 0.002 – 0.034 a.u., and the Laplacian of the electron density should be within 0.024 – 0.139 a.u.<sup>192</sup> The results obtained in this study fall within this range, with the exception of the N ··· H-N BCP, which displays slightly higher values for the  $\rho_{BCP}$ . This finding, along with the shorter distance in relation to the other H-bond in A-U pairs, indicates the presence of a stronger intermolecular H-bond.

**Table 5.4:** *Hydrogen bond critical points (HBCPs), electron density ( $\rho_{BCP}$ ), Laplacian ( $\nabla^2\rho_{BCP}$ ) and the total electron energy density ( $H_{BCP}$ ) as determined with QTAIM of base pairs calculated at the  $\omega$ B97XD/6-31++G(d,p) level of theory.*

Base pair	HBCP	$\rho_{BCP}$ (a.u.)	$\nabla^2\rho_{BCP}$ (a.u.)	$H_{BCP}$ (a.u.)
<b>A-U</b>	N-H ... O	0.02641	0.07580	-0.00034
	N ... H-N	0.04393	0.10013	-0.00023
<b>A-<sup>th</sup>U</b>	N-H ... O	0.02491	0.07120	-0.00040
	N ... H-N	0.04367	0.09993	-0.00228
<sup>th</sup> <b>A-U</b>	N-H ... O	0.02894	0.08368	-0.00022
	N ... H-N	0.04415	0.09994	-0.00245
<sup>th</sup> <b>A-<sup>th</sup>U</b>	N-H ... O	0.02732	0.07857	-0.00030
	N ... H-N	0.04381	0.09968	-0.00235
<b>G-C</b>	O ... H-N	0.02803	0.08070	0.00018
	N-H ... N	0.03370	0.08295	-0.00066
	N-H ... O	0.03901	0.11688	-0.00027
<b>G-<sup>th</sup>C</b>	O ... H-N	0.02890	0.08450	-0.00014
	N-H ... N	0.03315	0.08139	-0.00065
	N-H ... O	0.04188	0.12525	-0.00001
<sup>th</sup> <b>G-C</b>	O ... H-N	0.02664	0.07656	-0.00033
	N-H ... N	0.03312	0.08168	-0.00060
	N-H ... O	0.03526	0.10424	0.00010
<sup>th</sup> <b>G-<sup>th</sup>C</b>	O ... H-N	0.02734	0.07970	-0.00022
	N-H ... N	0.03250	0.07999	-0.00060
	N-H ... O	0.03783	0.11243	0.00013

### 5.3.4 Photophysical Properties of the Modified Nucleobase Pairs

It is of interest to compare the one-photon absorption of the base-pairs to the corresponding absorption properties of the isolated monomers. The photophysical properties of the modified base pairs computed in gas phase and water are summarized in Table 5.5. Figure 5.3 depicts the computed absorption spectra of the modified pairs in both gas phase and water at the B3LYP/6-31++G(d,p) gas phase optimized geometries. Results for higher lying states ( $S_4$  and higher) with negligible oscillator strengths ( $<0.01$ ) are excluded. For all computed photophysical properties, including excitation energies, oscillator strengths, and dominant transitions, see Tables D11 and D12 in Appendix D. The modified base pairs display excitation energies in the range of 300-350nm (3.5-4.0 eV) in the gas phase and slightly (10 nm) blue-shifted energies in water. It should be noted that weak transitions (oscillator strengths  $<0.01$ ) are observed for  $S_1$  in the gas phase for all the modified pairs except  ${}^{th}A-U$ . With the inclusion of solvation in water, the  $S_1$  transition displays much higher intensities (computed oscillator strengths). The excitation energies of the modified base pairs are comparable to those obtained for their monomeric units (isolated nucleobases) as determined at the B3LYP/6-311++g(2df,2p) level of theory in our previous work<sup>130</sup> as well as other theoretical<sup>83</sup> and experimental studies<sup>193</sup> of the natural nucleobases (See Chapter 2). A comparison of the excitation energies and oscillator strengths (for significant transitions) of the monomers and dimers are reported in Table 5.6. We note a particularly strong transition of  $G-{}^{th}C$  at 260 nm (4.77eV), which is very close to the previously computed  $S_2$  energy of  ${}^{th}C$ . We also note that the experimental (and theoretical<sup>83,130</sup>) absorption maximum of G in water is in this range, appearing at 252 nm. Similarly, the computed spectra of  $A-{}^{th}U$  and  ${}^{th}A-U$  pairs display peaks at 252 and 262 nm, which is characteristic of the computed  $S_2$  transition for  ${}^{th}A$  (262 nm), as well as the experimental peaks for A and U, at 259 and 262 nm respectively. Transitions in the 255-270 nm range are not observed for the modified-modified  ${}^{th}A-{}^{th}U$  pair. In contrast, the  ${}^{th}G-{}^{th}C$  pair displays two strong transitions in this range, at 263 and 260 nm, consistent with the computed  ${}^{th}C$   $S_2$  transition at 262 nm. It should be

noted that the intensities of higher energy (or lower wavelength) transitions of isolated  ${}^{th}\text{G}$  are underestimated with B3LYP according to our previous investigations (Chapter 2); hence, the relatively weak transitions observed in this region are expected to produce higher intensities with the use of other functionals such as PBE0. The  $\text{G}-{}^{th}\text{C}$ ,  ${}^{th}\text{G}-\text{C}$  and  ${}^{th}\text{G}-{}^{th}\text{C}$  pairs display low energy (at least 0.2 eV apart) excitations not related to the transitions of their constituents. In summary, the base pairs exhibit absorption profiles that correspond with characteristics of their monomeric units. In the case of the modified G-C pairs, we also note distinct red-shifted excitations for  $\text{S}_1$ .

**Table 5.5:** Vertical excitation energies of low-lying states of the modified base pairs calculated with B3LYP/6-31++G(d,p) in gas phase and water

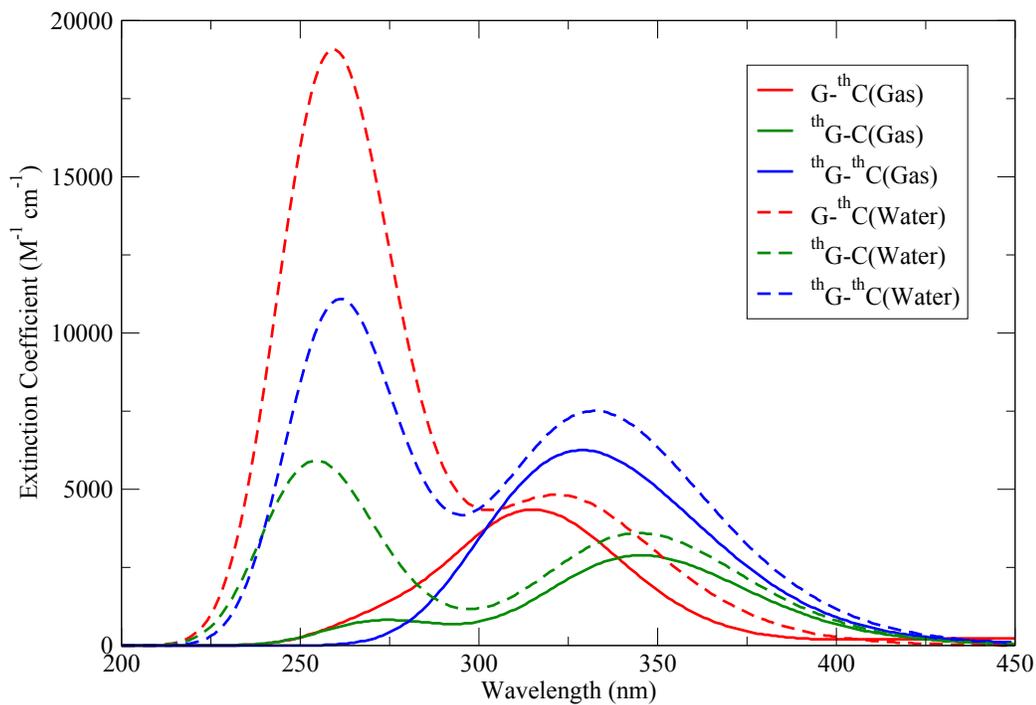
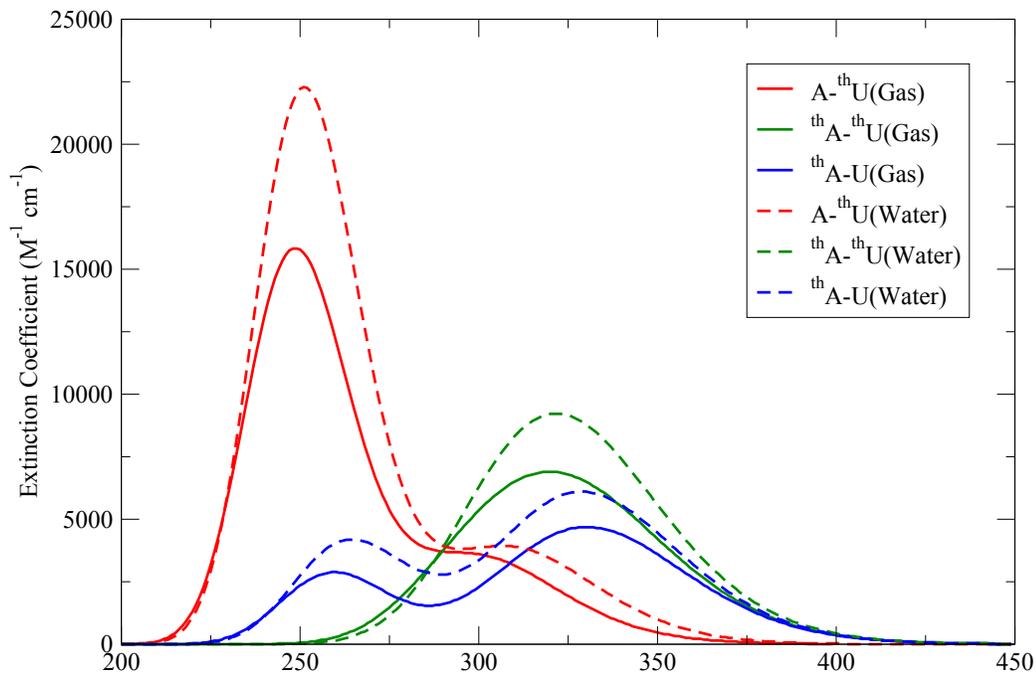
Base pair	Gas phase			Water		
	Transition	$\Delta E$ (eV)	$f$	Transition	$\Delta E$ (eV)	$f$
<b>A-<sup>th</sup>U</b>	H→L	3.78	0.002	H-1→L	4.00	0.092
	H-1→L	4.13	0.083	H→L	4.04	0.002
	H →L+1	4.95	0.290	H →L+1	4.98	0.448
<sup>th</sup> <b>A-U</b>	H→L	3.73	0.105	H→L	3.76	0.149
	H→L+1	3.98	0.015	H→L+1	4.30	0.006
	H-2→L	4.78	0.065	H-2→L	4.73	0.084
<sup>th</sup> <b>A-<sup>th</sup>U</b>	H→L	3.57	0.008	H→L	3.75	0.125
	H→L+1	3.77	0.116	H→L+1	3.88	0.037
	H-1→L	4.10	0.076	H-1→L	3.98	0.068
<b>G-<sup>th</sup>C</b>	H→L	2.80	0.005	H→L	3.57	0.015
	H-1→L	3.92	0.105	H-1→L	3.85	0.107
	H-3→L	4.50	0.023	H→L+1	4.77	0.302
<sup>th</sup> <b>G-C</b>	H→L	3.20	0.001	H→L	3.59	0.087
	H→L+1	3.59	0.071	H-1→L	4.68	0.040
	H-1→L	4.55	0.015	H-2→L	4.77	0.036
<sup>th</sup> <b>G-<sup>th</sup>C</b>	H→L	2.64	0.001	H→L	3.43	0.023
	H→L+1	3.56	0.075	H→L+1	3.62	0.077
	H-2→L	3.89	0.107	H-1→L	3.84	0.108
				H-2→L	4.72	0.105
				H-3→L	4.77	0.168

**Table 5.6:** Vertical excitation energies and oscillator strengths of select states of the bases and base pairs calculated with B3LYP/6-31++G(d,p) in water

A-U Pairs			G-C Pairs		
System	$\Delta E$ , eV (nm)	$f$	System	$\Delta E$ , eV (nm)	$f$
<b>A<sup>†</sup></b>	4.86 (255)	0.443	<b>G<sup>†</sup></b>	4.81 (258)	0.240
				5.09 (244)	0.554
<sup>th</sup> <b>A</b>	3.75 (330)	0.137	<sup>th</sup> <b>G</b>	3.76 (330)	0.081
	4.73 (262)	0.075		4.89 (254)	0.014
<b>U<sup>‡</sup></b>	5.17 (240)	0.190	<b>C<sup>†</sup></b>	4.86 (255)	0.178
<sup>th</sup> <b>U</b>	4.07 (305)	0.072	<sup>th</sup> <b>C</b>	3.84 (323)	0.093
	5.02 (247)	0.157		4.74 (262)	0.253
<b>A-<sup>th</sup>U</b>	4.00 (310)	0.092	<b>G-<sup>th</sup>C</b>	3.57 (347)	0.015
	4.98 (252)	0.448		3.85 (322)	0.107
	4.99 (249)	0.050		4.77 (260)	0.302
	5.02 (247)	0.057		4.96 (250)	0.089
<sup>th</sup> <b>A-U</b>	3.76 (329)	0.149	<sup>th</sup> <b>G-C</b>	3.59 (346)	0.087
	4.73 (262)	0.084		4.68 (265)	0.040
				4.87 (255)	0.036
<sup>th</sup> <b>A-<sup>th</sup>U</b>	3.75 (331)	0.125	<sup>th</sup> <b>G-<sup>th</sup>C</b>	3.43 (361)	0.023
	3.88 (320)	0.037		3.62 (342)	0.077
	3.98 (311)	0.068		3.84 (323)	0.108
	4.13 (300)	0.023		4.72 (263)	0.105
				4.77 (260)	0.168

<sup>†</sup> Computed at TD-B3LYP/6-311+G(d,p) in water (PCM)<sup>83</sup>.

<sup>‡</sup> Computed at TD-PBE0/6-311+G(2d,2p) in water (PCM)<sup>84</sup>.



**Figure 5.3:** *Computed absorption spectra of the modified A-U (top) and G-C (bottom) base pairs in gas phase and water.*

## 5.4 Conclusions

The base pairing stability and photophysical properties of the modified nucleobases based on the thieno[3,4-d]pyrimidine core, synthesized by Tor and co-workers<sup>109</sup> have been investigated in this chapter. We find that the inclusion of BSSE and ZPE corrections are essential to obtain accurate binding energies. Moreover, functionals with dispersion corrections (Grimme's D3 and  $\omega$ B97xD) yield results with good agreement to the natural base pairs. The computed structural and energetic properties demonstrate that these nucleobases have minimal impact on the geometry and relative stability of the canonical Watson-Crick pairs. Furthermore, AIM analyses reveal the relative H-bond strength between the WC pairs. We find that the H-bonding profiles of the modified base pairs are very similar to their natural counterparts. Hence, we propose that WC base pairing interactions of the modified bases would contribute to duplex stability. The computed absorption energies and intensities of the base pairs correlate well with the previous theoretically and experimentally determined energies of the individual constituents;<sup>83,84,130</sup> however distinct spectroscopic signatures are present for the modified G-C pairs, and possibly for A-<sup>th</sup>U.

# Chapter 6

## Conclusions

*An expert is a person who has found out by his own painful experience all the mistakes that one can make in a very narrow field.*

– Niels Bohr

## 6.1 Summary of Thesis Research

The main aim of this thesis has been to provide new insights into the structural and photophysical properties of modified nucleoside analogues. This has been achieved through the use of DFT and *ab initio* methods. This chapter will summarize the significant findings and provide a brief outlook on possible future investigations.

The first investigations performed were on the thieno[3,4-d]-pyrimidine nucleobase analogues. The photophysical properties of these analogues were determined using TD-DFT with various functionals in order to test their performance for reproducing experimental results. The commonly used B3LYP functional predicted absorption and emission energies with the best agreement to experiment, while the PBE0 and CAM-B3LYP energies were blue-shifted (hypsochromic) with respect to B3LYP as well as experiment. Although B3LYP reproduced energies successfully, it was deficient in predicting the absorption intensities for the <sup>th</sup>G analogue. The long-range corrected CAM-B3LYP functional yielded intensities best matching experiment. Ultimately, it was concluded that the PBE0 functional gave the best overall representation of the absorption and emission profiles for this class of analogues. Analysis of long range excitations using the  $\Lambda$  diagnostic with the range-separated CAM-B3LYP functional revealed that the lowest lying  $n\pi^*$  state of these analogues had charge transfer character. However, these CT states displayed significantly lower oscillator strengths than other excitations. The effect of solvent polarity was also assessed in this study by analyzing the computed properties in dioxane in comparison to water. Decreasing the solvent polarity resulted in higher computed emission energies, demonstrating the analogues' sensitivity to the microenvironment. This finding, along with the selective excitation energies of the analogues, shows that this class of analogues exhibit the desired properties for utility as probes.

The presence of unique spectral features from experimental measurements compared with subsequent absence of these features in computed spectra prompted an investigation of the tautomerization of the modified thiophene analogues. Chapter 3 examined the relative stabilities of the keto-enol and amino-imino tautomers with

DFT in the gas phase and bulk water. Based on the computed relative free energies of the 14 tautomers that were studied, only tautomers of  $^{th}G$  and  $^{th}C$  were predicted with appreciable stability. During the course of this investigation, another study<sup>131</sup> on the  $^{th}G$  tautomers confirmed our own findings of the stable  $^{th}G$  (k-a, 3H) tautomer. The use of explicit solvation in their work produced significantly lower relative energies and subsequent higher relative populations. This finding suggests that explicit solvation effects should be tested in future studies. The accuracy of DFT results was assessed by high accuracy CCSD(T) single-point energy computations at the DFT obtained geometries. Similar trends were observed in relative energies of the tautomers which suggests reliability of the DFT results. As in the previous study (Chapter 2), we utilized B3LYP, CAM-B3LYP and PBE0 functionals in the determination of photophysical properties of the tautomers. The  $^{th}G$ (k-a,3H) tautomer exhibited blue-shifted energies. It was also discovered that the tautomers of  $^{th}A$  display distinct absorption profiles; however, their relative populations were less than 1% in the gas phase and bulk water. The main outcome from the findings of this work is that both the  $^{th}G$ (k-a,1H) and  $^{th}G$ (k-a,3H) tautomers contribute to the spectral features of  $^{th}G$  in water.

Motivated by recent experimental studies on the two-photon excitation properties of nucleoside analogues, the TPA cross sections of 12 nucleoside analogues, including the thieno analogues were computed with TD-DFT in Chapter 4. Previous theoretical investigations<sup>172</sup> on eight of these analogues reported TPA cross sections several orders of magnitude higher in comparison to experiment; moreover, the exact computational approach used in determining the results was unclear. The TPA cross sections determined in the present work however were in much better agreement with respect to the experimental measurements. In addition, the first predictions for the TPA cross sections of the isothiazole analogues were determined, albeit resulting in low computed TPA values. Both B3LYP and CAM-B3LYP functionals were employed in this study; however no significant differences were observed between them. It was also concluded that the CAM-B3LYP functional did not provide any significant improvement to the computed TPA cross sections, as suggested by previous

studies.<sup>164</sup> Furthermore, there was no significant effect on the TPA cross section with the inclusion of the ribose group. Due to the conformational flexibility of three of the structures under study, the dependence on the rotation of the bond connecting the two rings on the TPA cross sections was analyzed. These rotation studies revealed a strong dependence of the TPA properties on the orientation of the rings. The energetics of the conformers generated by this rotation suggested a low rotational barrier for the *th*5-2U analogue; however a more detailed study of the potential energy surface is warranted. Specifically, a relaxed PES scan would provide more accurate energies for the barrier to rotation. We are also currently exploring the energetics of structures that yielded high TPA cross sections for the three analogues in question. Since the geometry optimizations are sensitive to the starting structures, it is important to ensure that all possible low energy conformers are located and their resulting photophysical properties analyzed. To our knowledge, the present work (Chapter 4) and the aforementioned previous study<sup>172</sup> are the only computational investigations on the two-photon absorption properties of nucleoside analogues. Consequently, this largely unexplored avenue of research requires further theoretical efforts.

In the last chapter, Watson-Crick base pairing interactions of the thieno analogues were investigated with the use of DFT functionals with added dispersion corrections in the gas phase-phase. To obtain accurate binding energies of the base pairs, we accounted for BSSE and ZPE corrections. These BSSE and ZPE energies accounted for 7-10% and 5% to the total binding energies respectively and thus demonstrate the significance of including such corrections. The structural and energy comparisons of modified pairs determined with the dispersion corrected functionals B3LYP-D3, M06-2X, and, in particular,  $\omega$ B97XD gave good agreement with respect to the natural base pairs. It was found that the *th*A-U, *th*A-*th*U and G-*th*C pairs are slightly stabilized compared to their natural counterparts while the other modified pairs are slightly destabilized. Our results were also compared with a similar study<sup>176</sup> that was published during the course of this research, and the two sets of results were consistent. With further investigations on the hydrogen bonding interactions with AIM analyses, significant similarities between the natural and modified pairs were

observed. These findings suggest that the incorporation of modified pairs through WC pairing would contribute to duplex stability. We concluded this study with the determination of the photophysical properties in the gas phase and water. These results revealed that the majority of the base pairs exhibit absorption profiles that correspond to characteristics of their monomeric units. We note the presence of distinct spectroscopic signatures for all the modified G-C pairs. Lastly, the two-photon absorption properties of the modified base pairs are under current investigation, as the TPA may be more sensitive to the base pairing than the computed one-photon absorption.

## 6.2 Future Perspectives

There is considerable scope for further investigations using the computational methodologies outlined in this thesis. Previous computations on the thieno analogues did not fully explore the excited-state potential energy surfaces beyond the Franck–Condon region. Future studies can address these issues with TD-DFT or with the use of high-accuracy multi-configurational methods such as CASSCF and CASPT2. Environmental effects could be accounted for with a combined quantum mechanical/molecular mechanical (QM/MM) computational protocol. Specifically this would require a realistic MM description of the ribose and water environment with an accurate QM treatment for the modified nucleobase (chromophore). A detailed study of the crossing points with this approach can elucidate whether conical intersections are energetically accessible.

The majority of the computations in this thesis considered the isolated nucleobase (or nucleoside) structures in bulk water without taking into account conformations of the chromophores in larger structures. Beyond the studies already mentioned, there would be great benefit in investigating conformations in oligomers or duplexes. The conformational changes and photophysical properties can be studied starting with base stacks, or dinucleotides, and moving up the longer oligonucleotides. For a complete picture, base-step parameters, backbone torsional angles and sugar pucker

conformations would need to be determined and compared to canonical structures. These parameters can be determined by conventional DFT for base stacks and dinucleotides, taking care to account for long-range interactions with the use of a dispersion corrected functionals. For larger structures, density functional tight binding (DFT-B) and TD-DFTB would potentially prove to be a suitable approach due its improved computational efficiency. However, careful benchmarking for structural and photophysical properties of the monomers and dimers must be undertaken.

We have yet to examine the role of tautomerization of the majority of the uridine analogues investigated in Chapter 4. More accurate theoretical data may be obtained by analyzing the tautomers of uridine analogues. The presence of stable tautomers may explain the discrepancies between experimentally and theoretically determined TPA cross sections, particularly in the case of <sup>th</sup>2-U, which is expected to have a significantly higher TPA cross section. Furthermore, more rigorous methods for obtaining TPA cross sections need to be investigated and compared with TD-DFT findings. In ongoing research, coupled cluster computations (CC2) have been employed for determining TPA cross sections, and could provide more accurate results as well as a comparison to TD-DFT. EOM-CCSD could also be utilized with the resolution-of-identity (RI) or Cholesky decomposition (CD) approximations to reduce computational costs and potentially yield comparable results to CC2. Future investigations could also include more detailed studies on the effect of the sugar moiety on the TPA cross section. Thus far, we have only analyzed the role of the sugar moiety in the fixed orientation from the optimized geometry; however, other stable conformations are possible and could alter photophysical properties. These conformations could be sampled using molecular dynamics approaches.

# Bibliography

- [1] Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer New York, 2006.
- [2] Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579–2619.
- [3] Brannon, J. H.; Magde, D. *J. Phys. Chem.* **1978**, *82*, 705–709.
- [4] Fischer, M.; Georges, J. *Chem. Phys. Lett.* **1996**, *260*, 115–118.
- [5] Chen, R. F. *Anal. Lett.* **1967**, *1*, 35–42.
- [6] Eastman, J. W. *Photochem. Photobiol.* **1967**, *6*, 55–72.
- [7] Valeur, B. *Molecular Fluorescence: Principles and Applications*; Wiley-VCH Verlag GmbH, 2001.
- [8] *Fluorescent Analogues of Biomolecular Building Blocks: Design and Applications*; John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016; pp 276–296.
- [9] Sinkeldam, R. W.; Tor, Y. *Org. Biomol. Chem.* **2007**, *5*, 2523–2528.
- [10] Reichardt, C. *Chem. Rev.* **1994**, *94*, 2319–2358.
- [11] Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *ChemBioChem* **2008**, *9*, 706–709.
- [12] Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. *Science* **2002**, *296*, 1700–1703.
- [13] Sundd, M.; Robertson, A. D. *Nat. Struct. Biol.* **2002**, *9*, 500–501.

- [14] Weiss, S. *Science* **1999**, *283*, 1676.
- [15] Birch, D. J. S.; Imhof, R. E. In *Topics in Fluorescence Spectroscopy: Techniques*; Lakowicz, J. R., Ed.; Springer US: Boston, MA, 1999; pp 1–95.
- [16] Förster, T. *Ann. d. Phys.* **1948**, *437*, 55–75.
- [17] Li, P. T. X.; Vieregg, J.; I. Tinoco, J. *Ann. Rev. Biochem.* **2008**, *77*, 77–100.
- [18] Frangioni, J. V. *Curr. Op. Chem. Biol.* **2003**, *7*, 626–634.
- [19] Hilderbrand, S. A.; Weissleder, R. *Curr. Opin. Chem. Biol.* **2010**, *14*, 71–79.
- [20] Goiffon, R. J.; Akers, W. J.; Berezin, M. Y.; Lee, H.; Achilefu, S. *J. Biomed. Opt.* **2009**, *14*, 020501–020501–3.
- [21] Sasaki, E.; Kojima, H.; Nishimatsu, H.; Urano, Y.; Kikuchi, K.; Hirata, Y.; Nagano, T. *J. Am. Chem. Soc.* **2005**, *127*, 3684–3685.
- [22] Kimura, R. H.; Cheng, Z.; Gambhir, S. S.; Cochran, J. R. *Cancer Res* **2009**, *69*, 2435.
- [23] Kricka, L. J.; Fortina, P. *Clin. Chem.* **2009**, *55*, 670.
- [24] Göppert-Mayer, M. *Ann. Phys.* **1931**, *401*, 273–294.
- [25] Kaiser, W.; Garrett, C. G. B. *Phys. Rev. Lett.* **1961**, *7*, 229–231.
- [26] Helmchen, F.; Denk, W. *Nat. Methods* **2005**, *2*, 932–940.
- [27] Denk, W.; Strickler, J.; Webb, W. *Science* **1990**, *248*, 73–76.
- [28] Pawlicki, M.; Collins, H.; Denning, R.; Anderson, H. *Angew. Chem., Int. Ed.* **2009**, *48*, 3244–3266.
- [29] Watson, J.; Crick, F. *Nature* **1953**, *171*, 737–738.
- [30] Wilkins, M.; Stokes, A.; Wilson, H. *Nature* **1953**, *171*, 738–740.

- [31] Franklin, R.; Gosling, R. *Nature* **1953**, *171*, 740–741.
- [32] Watson, J.; Crick, F. *Nature* **1953**, 964–967.
- [33] Franklin, R.; Gosling, R. *Nature* **1953**, *172*, 156–157.
- [34] Trikha, J.; Filman, D. J.; Hogle, J. M. *Nucleic Acids Res.* **1999**, *27*, 1728–1739.
- [35] Kimoto, M.; Cox, I., Robert Sidney; Hirao, I. *Expert Rev. Mol. Diagn.* **2011**, *11*, 321–331.
- [36] Tor, Y., Nakatani, K., Eds. *Modified Nucleic Acids*; Nucleic Acids and Molecular Biology; Springer, 2016; Vol. 31.
- [37] Improta, R.; Santoro, F.; Blancafort, L. *Chem. Rev.* **2016**, *116*, 3540–3593.
- [38] Kistler, K. A.; Matsika, S. In *Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications*; York, D. M., Lee, T.-S., Eds.; Springer Netherlands: Dordrecht, 2009; pp 285–339.
- [39] Barbatti, M.; Aquino, A. J. A.; Szymczak, J. J.; Nachtigallová, D.; Hobza, P.; Lischka, H. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 21453–21458.
- [40] Merchan, M.; Gonzalez-Luque, R.; Climent, T.; Serrano-Andres, L.; Rodriguez, E.; Reguero, M.; Pelaez, D. *J. Phys. Chem. B* **2006**, *110*, 26471–26476.
- [41] Sprecher, C. A.; Johnson, W. C. *Biopolymers* **1977**, *16*, 2243–2264.
- [42] Daniels, M.; Hauswirth, W. *Science* **1971**, *171*, 675.
- [43] Hauswirth, W.; Daniels, M. *Photochem. Photobiol.* **1971**, *13*, 157–163.
- [44] Fujiwara, T.; Kamoshida, Y.; Morita, R.; Yamashita, M. *J. Photochem, Photobiol. B* **1997**, *41*, 114–121.
- [45] Knighton, W. B. Excited State lifetimes of pyrimidines and purines in room temperature solution from fluorescence quantum yields and anisotropies. M.Sc. thesis, Montana State University, 1980.

- [46] Daniels, M. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2488–2491.
- [47] Callis, P. R. *Ann. Rev. Phys. Chem.* **1983**, *34*, 329–357.
- [48] Peon, J.; Zewail, A. H. *Chem. Phys. Lett.* **2001**, *348*, 255–262.
- [49] Bergstrom, D. E.; Inoue, H.; Reddy, P. A. *J. Org. Chem.* **1982**, *47*, 2174–2178.
- [50] Rachofsky, E. L.; Osman, R.; Ross, J. B. *Biochemistry* **2001**, *40*, 946–56.
- [51] Nordlund, T. M.; Andersson, S.; Nilsson, L.; Rigler, R.; Graeslund, A.; McLaughlin, L. W. *Biochemistry* **1989**, *28*, 9095–9103.
- [52] Bokacheva, L. P.; Semenov, S. G. *Opt. Spectrosc.* *75*, 708–711.
- [53] Smagowicz, J.; Wierzchowski, K. L. *J. Luminesc.* **1974**, *8*, 210–232.
- [54] Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem.* **1969**, *244*, 1228–1237.
- [55] Rist, M. J.; Marino, J. P. *Curr. Org. Chem.* **2002**, *6*, 775–793.
- [56] Asseline, U. *Curr. Org. Chem.* **2006**, *10*, 491–518.
- [57] Wilhelmsson, L. M. *Q. Rev. Biophys.* **2010**, *43*, 159–183.
- [58] Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res.* **2015**, *43*, e149–e149.
- [59] Tor, Y. *Tetrahedron* **2007**, *63*, 3415–3614.
- [60] Matsika, S. *Top. Curr. Chem.* **2015**, *355*, 209–243.
- [61] Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784–10785.
- [62] Krueger, A. T.; Imperiali, B. *ChemBioChem* **2013**, *14*, 788–799.
- [63] Connor, R. E.; Tirrell, D. A. *Polym. Rev.* **2007**, *47*, 9–28.
- [64] Young, D. C. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*; John Wiley & Sons, Inc., New York, 2001.

- [65] Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons, Ltd., West Sussex, England, 2007.
- [66] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.
- [67] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [68] Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242 – 250.
- [69] Becke, A. *J. Chem. Phys.* **1996**, *104*, 1040–1046.
- [70] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51 – 57.
- [71] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [72] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.
- [73] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6169.
- [74] Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.
- [75] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- [76] Chai, J.-D.; Head-Gordon, M. *J. Chem Phys.* **2008**, *128*, 084106.
- [77] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- [78] Tor, Y. *Angew. Chem., Int. Ed.* **1999**, *38*, 1579–1582.
- [79] Tor, Y. *Pure Appl. Chem.* **2009**, *81*, 263–272.
- [80] Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernández, C. E.; Kohler, B. *Annu. Rev. Phys. Chem.* **2009**, *60*, 217–39.
- [81] Crespo-Hernández, C. E.; Cohen, B.; Hare, P. M.; Kohler, B. *Chem. Rev.* **2004**, *104*, 1977–2020.
- [82] Kistler, K. A.; Matsika, S. In *Multi-scale Quantum Models for Biocatalysis*; York, D. M., Lee, T.-S., Eds.; Challenges and Advances in Computational Chemistry and Physics; Springer Netherlands, 2009; Vol. 7; pp 285–339.

- [83] Shukla, M. K.; Leszczynski, J. *Mol. Phys.* **2010**, *108*, 3131–3146.
- [84] Improta, R.; Barone, V. *J. Am. Chem. Soc.* **2004**, *126*, 14320–14321.
- [85] Mennucci, B.; Toniolo, A.; Tomasi, J. *J. Phys. Chem. A* **2001**, *105*, 4749–4757.
- [86] Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 11338–11346.
- [87] Shukla, M. K.; Leszczynski, J. *J. Comput. Chem.* **2004**, *25*, 768–778.
- [88] Barbatti, M.; Aquino, A. J. A.; Lischka, H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4959–4967.
- [89] Shukla, M. K.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2007**, *25*, 93–118.
- [90] Lorentzon, J.; Fuelscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 9265–73.
- [91] Zechmann, G.; Barbatti, M. *J. Phys. Chem. A* **2008**, *112*, 8273–8279.
- [92] Merchan, M.; Serrano-Andres, L. *J. Am. Chem. Soc.* **2003**, *125*, 8108–8109.
- [93] Lei, Y.; Yuan, S.; Dou, Y.; Wang, Y.; Wen, Z. *J. Phys. Chem. A* **2008**, *112*, 8497–8504.
- [94] Lan, Z.; Fabiano, E.; Thiel, W. *J. Phys. Chem. B* **2009**, *113*, 3548–3555.
- [95] Lan, Z.; Fabiano, E.; Thiel, W. *ChemPhysChem* **2009**, *10*, 1225–1229.
- [96] Gustavsson, T.; Banyasz, A.; Lazzarotto, E.; Markovitsi, D.; Scalmani, G.; Frisch, M. J.; Barone, V.; Improta, R. *J. Am. Chem. Soc.* **2006**, *128*, 607–619.
- [97] Fleig, T.; Knecht, S.; Haettig, C. *J. Phys. Chem. A* **2007**, *111*, 5482–5491.
- [98] Epifanovsky, E.; Kowalski, K.; Fan, P.-D.; Valiev, M.; Matsika, S.; Krylov, A. I. *J. Phys. Chem. A* **2008**, *112*, 9983–9992.

- [99] Hudock, H. R.; Levine, B. G.; Thompson, A. L.; Satzger, H.; Townsend, D.; Gador, N.; Ullrich, S.; Stolow, A.; Martinez, T. J. *J. Phys. Chem. A* **2007**, *111*, 8500–8508.
- [100] Barbatti, M.; Lischka, H. *J. Am. Chem. Soc.* **2008**, *130*, 6831–6839.
- [101] Serrano-Andrés, L.; Merchán, M.; Borin, A. C. *J. Am. Chem. Soc.* **2008**, *130*, 2473–2484.
- [102] Okamoto, A.; Saito, Y.; Saito, I. *J. Photochem. Photobiol., C* **2005**, *6*, 108–122.
- [103] Jean, J. M.; Hall, K. B. *Biochemistry* **2002**, *41*, 13152–13161.
- [104] Serrano-Andres, L.; Merchan, M.; Borin, A. C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8691–8696.
- [105] Broo, A. *J. Phys. Chem. A* **1998**, *102*, 526–531.
- [106] Perun, S.; Sobolewski, A. L.; Domcke, W. *Mol. Phys.* **2006**, *104*, 1113–1121.
- [107] Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 37–41.
- [108] Kelley, J. K., Shana O. Barton *Science* **1999**, *283*, 375.
- [109] Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133*, 14912–14915.
- [110] McLean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639–5648.
- [111] Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.
- [112] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark,

M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. GAUSSIAN 09, Revision B.01, Gaussian, Inc.: Wallingford CT, 2010.

- [113] Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2010**, *132*, 114110.
- [114] Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.
- [115] Scalmani, G.; Frisch, M. J.; Mennucci, B.; Tomasi, J.; Cammi, R.; Barone, V. *J. Chem. Phys.* **2006**, *124*, 094107.
- [116] Improta, R.; Barone, V.; Scalmani, G.; Frisch, M. J. *J. Chem. Phys.* **2006**, *125*, 054103.
- [117] Improta, R.; Scalmani, G.; Frisch, M. J.; Barone, V. *J. Chem. Phys.* **2007**, *127*, 074504.
- [118] Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. *J. Chem. Phys.* **2008**, *128*, 044118.
- [119] Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- [120] Gordon, M. S.; Schmidt, M. W. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier, Amsterdam, 2005; pp 1167–1189.

- [121] Samanta, P. K.; Manna, A. K.; Pati, S. K. *J. Phys. Chem. B* **2012**, *116*, 7618–7626.
- [122] Dennington, R.; Keith, T.; Millam, J. Gauss View Version 5. Semichem Inc. Shawnee Mission KS 2009.
- [123] Li, D.; Fedeles, B. I.; Singh, V.; Peng, C. S.; Silvestre, K. J.; Simi, A. K.; Simpson, J. H.; Tokmakoff, A.; Essigmann, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E3252–E3259.
- [124] Singh, V.; Fedeles, B. I.; Essigmann, J. M. *RNA* **2014**, *21*, 1–13.
- [125] Mons, M.; Piuze, F.; Dimicoli, I.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2006**, *110*, 10921–10924.
- [126] Choi, M. Y.; Miller, R. E. *J. Am. Chem. Soc.* **2006**, *128*, 7320–7328.
- [127] Choi, M. Y.; Dong, F.; Miller, R. E. *Phil. Trans. R. Soc. A* **2005**, *363*, 393–413.
- [128] Laxer, A.; Major, D. T.; Gottlieb, H. E.; Fischer, B. *J. Org. Chem.* **2001**, *66*, 5463–5481.
- [129] Shukla, M. K.; Leszczynski, J. *WIREs Comput. Mol. Sci.* **2013**, *3*, 637–649.
- [130] Gedik, M.; Brown, A. *J. Photochem. Photobiol. A* **2013**, *259*, 25–32.
- [131] Sholokh, M.; Improta, R.; Mori, M.; Sharma, R.; Kenfack, C.; Shin, D.; Voltz, K.; Stote, R. H.; Zaporozhets, O. A.; Botta, M.; Tor, Y.; Mély, Y. *Angew. Chem., Int. Ed.* **2016**, *55*, 7974–7978.
- [132] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [133] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- [134] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.
- [135] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

- [136] Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910–1918.
- [137] Pople, J. A.; Head-Gordon, M.; ; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968–5975.
- [138] Dunning Jr., T. H. *J. Chem Phys.* **1989**, *90*, 1007–23.
- [139] Frisch, M. J. et al. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.
- [140] Leszczynski, J. In *Advances in Molecular Structure and Research*; Hargittai, M., Ed.; JAI Press: Stamford, CT, 2000; Vol. 6; p 209.
- [141] Gorb, L.; Kaczmarek, A.; Gorb, A.; Sadlej, A. J.; Leszczynski, J. *J. Phys. Chem. B* **2005**, *109*, 13770–13776.
- [142] Hanus, M.; Ryjáček, F.; Kabeláč, M.; Kubař, T.; Bogdan, T. V.; Trygubenko, S. A.; Hobza, P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192–4203.
- [143] Fazaeli, R.; Monajjemi, M.; Ataherian, F.; Zare, K. *J. Mol. Struct. (Theochem)* **2002**, *581*, 51 – 58.
- [144] Katilius, E.; Woodbury, N. W. *J. Biomed. Opt.* **2006**, *11*, 044004–044004–7.
- [145] Stanley, R.; Hou, Z.; Yang, A.; Hawkins, M. *J. Phys. Chem. B* **2005**, *109*, 3690–3695.
- [146] Lane, R. S. K.; Magennis, S. W. *RSC Adv.* **2012**, *2*, 11397–11403.
- [147] Lane, R. S. K.; Jones, R.; Sinkeldam, R. W.; Tor, Y.; Magennis, S. W. *ChemPhysChem* **2014**, *15*, 867–871.
- [148] Rovira, A. R.; Fin, A.; Tor, Y. *J. Am. Chem. Soc.* **2015**, *137*, 14602–14605.
- [149] Chawla, M.; Poater, A.; Oliva, R.; Cavallo, L. *Phys. Chem. Chem. Phys.* **2016**, *18*, 18045–18053.

- [150] Zhang, L.; Kong, X.; Wang, M.; Zheng, M. *International Journal of Quantum Chemistry* **2017**, e25377–n/a, e25377.
- [151] Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117 – 129.
- [152] Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.
- [153] Cammi, R.; Tomasi, J. *J. Comput. Chem.* **1995**, *16*, 1449–1458.
- [154] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem Phys.* **1972**, *56*, 2257–2261.
- [155] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4*, 294–301.
- [156] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem Phys.* **1980**, *72*, 650–654.
- [157] Gill, P. M.; Johnson, B. G.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1992**, *197*, 499 – 505.
- [158] Wang, Y.; Li, H. *J. Chem Phys.* **2010**, *133*, 034108.
- [159] Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- [160] Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.
- [161] Cossi, M.; Barone, V. *J. Chem Phys.* **2001**, *115*, 4708–4717.
- [162] Wang, Y.; Li, H. *J. Chem Phys.* **2009**, *131*, 206101.
- [163] Peticolas, W. L. *Ann. Rev. Phys. Chem.* **1967**, *18*, 233–260.
- [164] Beerepoot, M. T. P.; Friese, D. H.; List, N. H.; Kongsted, J.; Ruud, K. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19306–19314.
- [165] Monson, P. R.; McClain, W. M. *J. Chem Phys.* **1970**, *53*, 29–37.

- [166] McClain, W. M. *J. Chem Phys.* **1971**, *55*, 2789–2796.
- [167] Andrews, D. L.; Thirunamachandran, T. *J. Chem. Phys.* **1977**, *67*, 5026–5033.
- [168] Cronstrand, P.; Luo, Y.; Ågren, H. In *Response Theory and Molecular Properties (A Tribute to Jan Lindenberg and Poul Jørgensen)*; Jensen, H., Ed.; Advances in Quantum Chemistry; Amsterdam Academic Press, 2005; Vol. 50; pp 1 – 21.
- [169] Rizzo, A.; Coriani, S.; Ruud, K. In *Computational Strategies for Spectroscopy*; Barone, V., Ed.; John Wiley & Sons, Inc., Hoboken, NJ, USA, 2011; Chapter 2, pp 77–135.
- [170] Matczyszyn, K.; Olesiak-Banska, J.; Nakatani, K.; Yu, P.; Murugan, N. A.; Zaleśny, R.; Roztoczyńska, A.; Bednarska, J.; Bartkowiak, W.; Kongsted, J.; Ågren, H.; Samoć, M. *J. Phys. Chem. B* **2015**, *119*, 1515–1522.
- [171] Sinkeldam, R. W.; Hopkins, P. A.; Tor, Y. *ChemPhysChem* **2012**, *13*, 3350–3356.
- [172] Samanta, P. K.; Pati, S. K. *Phys. Chem. Chem. Phys.* **2015**, *17*, 10053–10058.
- [173] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekstrm, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernndez, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Httig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertens, S. Hst, I.-M. Hyvik, M. F. Iozzi, B. Jansik, H. J. Aa. Jensen, D. Jonsson, P. Jrgensen, J. Kauczor, S. Kirpekar, T. Kjrgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutns, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. Rybkin, P. Salek, C. C. M. Samson, A. Snchez de Mers, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K.

- O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, and H. Ågren, “The Dalton quantum chemistry program system”, *WIREs Comput. Mol. Sci.* **2014**, *4*:269-284.
- [174] Friese, D. H.; Hattig, C.; Ruud, K. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1175–1184.
- [175] Friese, D. H.; Beerepoot, M. T. P.; Ringholm, M.; Ruud, K. *J. Chem. Theory Comput.* **2015**, *11*, 1129–1144.
- [176] Samanta, P. K.; Pati, S. K. *New J. Chem.* **2013**, *37*, 3640–3646.
- [177] Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- [178] Jansen, H.; Ros, P. *Chem Phys. Lett.* **1969**, *3*, 140 – 143.
- [179] Liu, B.; McLean, A. D. *J. Chem Phys.* **1973**, *59*, 4557–4558.
- [180] van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873–1885.
- [181] Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024–11031.
- [182] Hobza, P.; Bludsky, O.; Suhai, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3073–3078.
- [183] Šponer, J. E.; Réblová, K.; Mokdad, A.; Sychrovský, V.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2007**, *111*, 9153–9164.
- [184] Šponer, J. E.; Špačková, N.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399–11410.
- [185] Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146–10159.
- [186] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem Phys.* **2010**, *132*, 154104.

- [187] Bader, R. F. W. *Atoms in Molecules - A Quantum Theory*; Oxford University Press, 1990.
- [188] Keith, T. A. AIMAll (Version 13.01.27). TK Gristmill Software, Overland Park KS, USA, 2017 (aim.tkgristmill.com).
- [189] Seeman, N. C.; Rosenberg, J. M.; Suddath, F.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109 – 144.
- [190] Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145–167.
- [191] Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, *99*, 9747–9754.
- [192] Popelier, P. L. A. *Atoms in Molecules : An Introduction*; Harlow: Prentice Hall, 2000.
- [193] Cavaluzzi, M. J.; Borer, P. N. *Nucleic Acids Res.* **2004**, *32*, e13.
- [194] Bode, B. M.; Gordon, M. S. *J. Mol. Graph. Model.* **1998**, *16*, 133–138.

# Appendix A

## Appendix to Chapter 2

## A1 Vertical Excitation Energies

**Table A1:** Vertical excitation energies of the modified nucleobases computed with TD-CAM-B3LYP/6-311++G(2df,2p) in dioxane.

	<b>S1</b>	<b>S2</b>	<b>S3</b>
<b>Nucleobase</b>	$\Delta E$ , eV	$\Delta E$ , eV	$\Delta E$ , eV
<sup>th</sup> <b>A</b> denine	4.06 (0.182)	4.58 <sup>a</sup> (0.002)	5.03 <sup>b</sup> (0.113)
<sup>th</sup> <b>C</b> ytosine	4.27 (0.153)	5.09 <sup>a</sup> (0.325)	5.10 <sup>b</sup> (0.002)
<sup>th</sup> <b>G</b> uanine	4.31 (0.130)	5.16 <sup>b</sup> (0.000)	5.18 <sup>a</sup> (0.022)
<sup>th</sup> <b>U</b> racil	4.62 (0.126)	5.04 <sup>b</sup> (0.000)	5.34 <sup>a</sup> (0.002)

All S<sub>1</sub> states are  $\pi\pi^*$  type. <sup>a</sup>  $\pi\pi^*$  transition and <sup>b</sup>  $n\pi^*$  transition

**Table A2:** Vertical excitation energies of the modified nucleobases computed with TD-PBE0/6-311++G(2df,2p) in dioxane.

	<b>S1</b>	<b>S2</b>	<b>S3</b>
<b>Nucleobase</b>	$\Delta E$ , eV	$\Delta E$ , eV	$\Delta E$ , eV
<sup>th</sup> <b>A</b> denine	3.85 (0.154)	4.21 <sup>a</sup> (0.002)	4.90 <sup>b</sup> (0.081)
<sup>th</sup> <b>C</b> ytosine	3.95 (0.117)	4.71 <sup>b</sup> (0.000)	4.84 <sup>a</sup> (0.259)
<sup>th</sup> <b>G</b> uanine	3.99 (0.098)	4.90 <sup>b</sup> (0.000)	5.04 <sup>a</sup> (0.003)
<sup>th</sup> <b>U</b> racil	4.29 (0.093)	4.75 <sup>b</sup> (0.000)	5.23 <sup>a</sup> (0.161)

All S<sub>1</sub> states are  $\pi\pi^*$  type. <sup>a</sup>  $\pi\pi^*$  transition and <sup>b</sup>  $n\pi^*$  transition

**Table A3:** Vertical excitation energies of the modified nucleobases computed with TD-B3LYP/6-311++G(2df,2p) in dioxane.

Nucleobase	S1	S2	S3
	$\Delta E$ , eV	$\Delta E$ , eV	$\Delta E$ , eV
<sup>th</sup> Adenine	3.72 (0.144)	4.03 <sup>a</sup> (0.002)	4.75 <sup>b</sup> (0.079)
<sup>th</sup> Cytosine	3.79 (0.107)	4.50 <sup>b</sup> (0.000)	4.67 <sup>a</sup> (0.240)
<sup>th</sup> Guanine	3.83 (0.088)	4.76 <sup>b</sup> (0.000)	4.85 <sup>a</sup> (0.004)
<sup>th</sup> Uracil	4.12 (0.083)	4.60 <sup>b</sup> (0.000)	5.07 <sup>a</sup> (0.160)

All S<sub>1</sub> states are  $\pi\pi^*$  type. <sup>a</sup>  $\pi\pi^*$  transition and <sup>b</sup>  $n\pi^*$  transition.

### A1.1 $\Lambda$ Diagnostic Parameter

**Table A4:** Vertical excitation energies and  $\Lambda$  diagnostic of the modified nucleobases computed with GAMESS. Oscillator strengths are given in parentheses.

Nucleobase	S <sub>1</sub>		S <sub>2</sub>		S <sub>3</sub>	
	$\Delta E$ , eV	$\Lambda$	$\Delta E$ , eV	$\Lambda$	$\Delta E$ , eV	$\Lambda$
<sup>th</sup> Adenine	4.08 <sup>a</sup> (0.180)	0.799	4.75 <sup>b</sup> (0.003)	0.371	5.01 <sup>a</sup> (0.114)	0.742
<sup>th</sup> Cytosine	4.30 <sup>a</sup> (0.140)	0.666	5.08 <sup>a</sup> (0.319)	0.764	5.34 <sup>b</sup> (0.017)	0.393
<sup>th</sup> Guanine	4.18 <sup>a</sup> (0.124)	0.683	5.15 <sup>a</sup> (0.036)	0.777	5.27 <sup>b</sup> (0.000)	0.378
<sup>th</sup> Uracil	4.55 <sup>a</sup> (0.112)	0.646	5.16 <sup>b</sup> (0.000)	0.374	5.27 <sup>a</sup> (0.189)	0.777

<sup>a</sup>  $\pi\pi^*$  transition

<sup>b</sup>  $n\pi^*$  transition

## A2 Coordinates

### A2.1 Ground State Geometries of the Modified Nucleobases

*th Adenine*

*CAM-B3LYP/6-311++G(2df,2p)*

*Water*

C	0.305883	-0.955480	-0.000023
C	-0.045559	0.430163	0.000159
C	-1.451576	0.732538	0.000171
C	-1.872133	-1.523323	-0.000097
N	-0.646480	-1.945973	-0.000072
C	1.048322	1.257697	-0.000239
C	1.665528	-1.141097	0.000231
S	2.478296	0.351713	-0.000056
H	1.089707	2.333245	-0.000284
H	2.205456	-2.071489	0.000276
N	-1.888063	1.996453	-0.000241
H	-2.877366	2.173823	-0.000524
H	-1.259934	2.777976	0.001876
H	-2.650464	-2.278569	-0.000591
N	-2.331590	-0.245535	0.000160

*th Adenine*

*CAM-B3LYP/6-311++G(2df,2p)*

*Dioxane*

C	-0.305655	-0.953594	0.000072
C	0.046120	0.431927	-0.000005
C	1.453136	0.726562	-0.000021
C	1.866939	-1.527143	-0.000058
N	0.641813	-1.944796	0.000043
C	-1.047619	1.260013	-0.000100
C	-1.666052	-1.137861	0.000106
S	-2.478360	0.353348	-0.000029
H	-1.090026	2.335549	-0.000225
H	-2.201752	-2.070341	0.000151
N	1.895787	1.995499	-0.000154
H	2.886526	2.160662	0.000657
H	1.270470	2.777550	0.001199
H	2.644403	-2.282505	-0.000166
N	2.328532	-0.246974	-0.000048

*th Adenine*  
*B3LYP/6-311++G(2df,2p)*  
*Water*

C	0.305681	-0.959843	0.000091
C	-0.048094	0.434943	-0.000039
C	-1.456868	0.734484	-0.000003
C	-1.879476	-1.528905	-0.000165
N	-0.643896	-1.953217	0.000050
C	1.050589	1.266311	-0.000251
C	1.672804	-1.146462	0.000330
S	2.491997	0.354809	-0.000086
H	1.096645	2.341982	-0.000570
H	2.215940	-2.075467	0.000668
N	-1.903027	2.001061	0.000268
H	-2.893827	2.173741	0.000096
H	-1.280287	2.787676	0.000698
H	-2.655521	-2.287125	-0.000446
N	-2.342037	-0.250833	-0.000155

*th Adenine*  
*PBE0/6-311++G(2df,2p)*  
*Water*

C	0.303318	-0.956396	-0.000136
C	-0.047436	0.432355	0.000263
C	-1.451635	0.732320	0.000338
C	-1.871230	-1.521184	-0.000073
N	-0.639841	-1.947835	-0.000164
C	1.050420	1.262127	-0.000142
C	1.669816	-1.139861	0.000091
S	2.477715	0.352691	-0.000019
H	1.099168	2.339639	0.000006
H	2.216441	-2.068905	-0.000023
N	-1.893569	1.993269	-0.000408
H	-2.883340	2.166151	-0.001218
H	-1.268816	2.776974	0.001554
H	-2.647917	-2.281732	-0.000190
N	-2.334946	-0.249911	0.000304

*th Adenine*  
*B3LYP/6-311++G(2df,2p)*  
*Dioxane*

C	0.305604	-0.958168	0.003200
C	-0.048801	0.436642	-0.001545
C	-1.458172	0.728553	-0.001325
C	-1.874116	-1.532627	-0.002577
N	-0.638965	-1.952053	0.002403
C	1.049657	1.268690	-0.005830
C	1.673566	-1.143146	0.005390
S	2.491996	0.356781	-0.000880
H	1.095977	2.344322	-0.015588
H	2.213007	-2.073971	0.009000
N	-1.911179	2.000744	-0.015506
H	-2.902208	2.156221	0.052988
H	-1.294051	2.782865	0.095842
H	-2.649165	-2.291168	-0.004918
H	-2.338701	-0.252469	-0.002202

*th Adenine*  
*PBE0/6-311++G(2df,2p)*  
*Dioxane*

C	0.302986	-0.954619	0.002394
C	-0.048013	0.434186	-0.001234
C	-1.452641	0.726532	-0.001020
C	-1.866255	-1.524535	-0.002030
N	-0.635355	-1.946532	0.001857
C	1.049873	1.264312	-0.004451
C	1.670253	-1.136726	0.004109
S	2.477669	0.354219	-0.000666
H	1.099501	2.341729	-0.011828
H	2.212751	-2.067791	0.006951
N	-1.901084	1.992586	-0.009965
H	-2.891723	2.150616	0.036873
H	-1.280308	2.774457	0.066024
H	-2.642033	-2.285224	-0.003900
N	-2.331863	-0.251225	-0.001903

*th* Cytosine  
CAM-B3LYP/6-311++G(2df,2p)  
Water

C	0.765915	-0.511762	0.000000
C	-0.001641	0.683728	0.000000
C	-1.444819	0.539085	0.000000
C	-1.281418	-1.793198	0.000000
N	-2.018013	-0.636811	0.000000
N	0.092608	-1.710697	0.000000
O	-1.812800	-2.899133	0.000000
C	0.783463	1.803461	0.000000
C	2.108166	-0.287623	0.000000
S	2.427490	1.395952	0.000000
H	0.485049	2.837651	0.000000
H	0.594417	-2.583615	0.000000
H	2.908701	-1.005928	0.000000
N	-2.220022	1.624619	0.000000
H	-3.219041	1.511047	0.000000
H	-1.843376	2.554074	0.000000

*th* Cytosine  
B3LYP/6-311++G(2df,2p)  
Water

C	0.768010	-0.514579	0.000000
C	-0.002237	0.687201	0.000000
C	-1.446794	0.541463	0.000000
C	-1.288800	-1.800516	0.000000
N	-2.025046	-0.640726	0.000000
N	0.093650	-1.714919	0.000000
O	-1.820362	-2.912669	0.000000
C	0.788391	1.812098	0.000000
C	2.118041	-0.292234	0.000000
S	2.443030	1.400847	0.000000
H	0.494569	2.847679	0.000000
H	0.598307	-2.587275	0.000000
H	2.919513	-1.009869	0.000000
N	-2.229504	1.629412	0.000000
H	-3.228814	1.513777	0.000000
H	-1.857273	2.561159	0.000000

*th* Cytosine  
CAM-B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.765915	-0.511762	0.000000
C	-0.001641	0.683728	0.000000
C	-1.444819	0.539085	0.000000
C	-1.281418	-1.793198	0.000000
N	-2.018013	-0.636811	0.000000
N	0.092608	-1.710697	0.000000
O	-1.812800	-2.899133	0.000000
C	0.783463	1.803461	0.000000
C	2.108166	-0.287623	0.000000
S	2.427490	1.395952	0.000000
H	0.485049	2.837651	0.000000
H	0.594417	-2.583615	0.000000
H	2.908701	-1.005928	0.000000
N	-2.220022	1.624619	0.000000
H	-3.219041	1.511047	0.000000
H	-1.843376	2.554074	0.000000

*th* Cytosine  
B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.769902	-0.516427	0.000000
C	-0.000429	0.685864	0.000000
C	-1.446874	0.535979	0.000000
C	-1.297540	-1.805725	0.000000
N	-2.027148	-0.635002	0.000000
N	0.095355	-1.713071	0.000000
O	-1.819538	-2.910032	0.000000
C	0.789038	1.810231	0.000000
C	2.120362	-0.289546	0.000000
S	2.446412	1.403992	0.000000
H	0.492623	2.845165	0.000000
H	0.592641	-2.589143	0.000000
H	2.923672	-1.005209	0.000000
N	-2.230031	1.632561	0.000000
H	-3.228057	1.509140	0.000000
H	-1.855709	2.562075	0.000000

*th* Cytosine  
PBE0/6-311++G(2df,2p)  
Water

C	0.761486	-0.513388	0.000000
C	-0.004399	0.684036	0.000000
C	-1.444787	0.538689	0.000000
C	-1.285039	-1.793200	0.000000
N	-2.021391	-0.639716	0.000000
N	0.091100	-1.707700	0.000000
O	-1.812485	-2.902247	0.000000
C	0.786460	1.806517	0.000000
C	2.109975	-0.288307	0.000000
S	2.426919	1.392158	0.000000
H	0.495448	2.845051	0.000000
H	0.594331	-2.579804	0.000000
H	2.914836	-1.004884	0.000000
N	-2.221718	1.623062	0.000000
H	-3.220170	1.508213	0.000000
H	-1.845886	2.552370	0.000000

*th* Guanine  
CAM-B3LYP/6-311++G(2df,2p)  
Water

C	0.768595	-0.453154	0.037298
C	-0.003158	0.750820	0.004098
C	-1.450946	0.667554	0.016457
C	-1.090474	-1.753349	0.095872
N	0.196691	-1.707496	0.091414
C	0.759329	-1.881749	0.041655
C	2.109966	-0.191699	0.015469
S	2.408439	-1.492319	0.044484
H	0.423889	-2.903663	0.073590
H	2.920156	-0.899258	0.032144
H	-2.906582	-0.751950	0.045490
N	-1.903712	-0.635525	0.061602
O	-2.233550	-1.600946	0.011853
N	-1.744452	-2.941120	0.092683
H	-2.700742	-2.987396	0.398758
H	-1.181378	-3.752714	0.280298

*th* Cytosine  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.763267	-0.515026	0.000000
C	-0.002617	0.682898	0.000000
C	-1.444574	0.533399	0.000000
C	-1.293337	-1.797916	0.000000
N	-2.023409	-0.634132	0.000000
N	0.092343	-1.705543	0.000000
O	-1.811421	-2.899776	0.000000
C	0.787191	1.804700	0.000000
C	2.112080	-0.285656	0.000000
S	2.430192	1.395134	0.000000
H	0.493767	2.842524	0.000000
H	0.587926	-2.581373	0.000000
H	2.918717	-1.000171	0.000000
N	-2.222032	1.625938	0.000000
H	-3.219060	1.502863	0.000000
H	-1.844355	2.552986	0.000000

*th* Guanine  
CAM-B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.765289	-0.453566	0.036725
C	-0.003815	0.752016	0.007861
C	-1.453626	0.676357	0.014693
C	-1.091470	-1.749533	0.090146
N	0.190890	-1.707900	0.090761
C	0.760553	-1.880521	0.034578
C	2.106392	-0.196244	0.015469
S	2.409393	-1.486699	0.039228
H	0.421658	-2.901285	0.063849
H	2.911526	-0.909157	0.029848
H	-2.907965	-0.747303	0.006321
N	-1.906694	-0.634211	0.055637
O	-2.235332	-1.603610	0.017123
N	-1.751068	-2.944019	0.080261
H	-2.676013	-2.990796	0.471197
H	-1.167649	-3.744371	0.255860

*th* Guanine  
 B3LYP/6-311++G(2df,2p)  
 Water

C	0.767854	-0.455311	0.036634
C	-0.005906	0.756587	0.003669
C	-1.456005	0.674756	0.015108
C	-1.097566	-1.758094	0.093687
N	0.197355	-1.710146	0.090522
C	0.762660	-1.891907	0.040752
C	2.117940	-0.195445	0.015915
S	2.422142	-1.497470	0.042725
H	0.434427	-2.916445	0.072129
H	2.929537	-0.902051	0.032472
H	-2.915240	-0.750912	0.041016
N	-1.911295	-0.636371	0.059216
O	-2.245325	-1.610659	0.012930
N	-1.750110	-2.953899	0.081307
H	-2.697881	-3.003071	0.416330
H	-1.180515	-3.759134	0.282660

*th* Guanine  
 PBE0/6-311++G(2df,2p)  
 Water

C	0.762182	-0.457811	0.037043
C	-0.007369	0.749207	0.004350
C	-1.453589	0.668930	0.015984
C	-1.095030	-1.752067	0.093961
N	0.197849	-1.708630	0.090762
C	0.760267	-1.882984	0.040485
C	2.110552	-0.193943	0.015585
S	2.406191	-1.486737	0.042943
H	0.431896	-2.909478	0.071856
H	2.925773	-0.899136	0.031710
H	-2.907471	-0.751550	0.042088
N	-1.904464	-0.636252	0.059743
O	-2.239556	-1.601620	0.011766
N	-1.744838	-2.941252	0.083305
H	-2.694627	-2.989408	0.408836
H	-1.175696	-3.745515	0.283682

*th* Guanine  
 B3LYP/6-311++G(2df,2p)  
 Dioxane

C	0.764562	-0.455742	0.036179
C	-0.006468	0.757750	0.007343
C	-1.458643	0.683907	0.013187
C	-1.098779	-1.754137	0.088920
N	0.191616	-1.710562	0.090689
C	0.763846	-1.890651	0.033757
C	2.114369	-0.199994	0.015810
S	2.423076	-1.491917	0.037691
H	0.432060	-2.914046	0.062473
H	2.921114	-0.911787	0.029976
H	-2.916617	-0.746408	0.002269
N	-1.914255	-0.635005	0.053291
O	-2.246796	-1.613570	0.018550
N	-1.757134	-2.956335	0.071641
H	-2.672863	-3.007122	0.487004
H	-1.167019	-3.751360	0.256163

*th* Guanine  
 PBE0/6-311++G(2df,2p)  
 Dioxane

C	0.758778	-0.458392	0.035969
C	-0.008105	0.750077	0.007792
C	-1.456108	0.677729	0.013626
C	-1.096342	-1.748068	0.087793
N	0.191953	-1.709371	0.089728
C	0.761394	-1.881422	0.032948
C	2.106800	-0.198652	0.015598
S	2.406902	-1.480963	0.036872
H	0.429213	-2.906634	0.061351
H	2.916897	-0.909276	0.029355
H	-2.908690	-0.746675	0.002859
N	-1.907244	-0.634793	0.052698
O	-2.241486	-1.604282	0.017408
N	-1.752313	-2.943793	0.071018
H	-2.668101	-2.992470	0.483611
H	-1.161477	-3.736228	0.258530

*th* Uracil  
CAM-B3LYP/6-311++G(2df,2p)  
Water

C	0.330460	-0.827760	0.000000
C	0.431020	0.589830	0.000000
C	-0.770520	1.411980	0.000000
C	-2.081180	-0.706730	0.000000
N	-1.938320	0.672420	0.000000
N	-0.923650	-1.417810	0.000000
O	-3.181620	-1.224590	0.000000
C	1.725060	1.021290	0.000000
C	1.539730	-1.446150	0.000000
S	2.802410	-0.284810	0.000000
O	-0.800140	2.625980	0.000000
H	2.074040	2.039110	0.000000
H	-1.020710	-2.420450	0.000000
H	-2.805180	1.190650	0.000000
H	1.753670	-2.500370	0.000000

*th* Uracil  
B3LYP/6-311++G(2df,2p)  
Water

C	0.330016	-0.829244	0.000000
C	0.431299	0.594539	0.000000
C	-0.771793	1.418455	0.000000
C	-2.091723	-0.710286	0.000000
N	-1.945302	0.673495	0.000000
N	-0.925645	-1.420315	0.000000
O	-3.195297	-1.233632	0.000000
C	1.732754	1.027237	0.000000
C	1.544488	-1.453667	0.000000
S	2.816480	-0.287471	0.000000
O	-0.805593	2.638634	0.000000
H	2.087404	2.043266	0.000000
H	-1.020175	-2.423870	0.000000
H	-2.811989	1.193354	0.000000
H	1.760144	-2.507904	0.000000

*th* Uracil  
CAM-B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.332260	-0.829804	0.000000
C	0.432307	0.587627	0.000000
C	-0.769724	1.415526	0.000000
C	-2.086865	-0.705979	0.000000
N	-1.939919	0.673608	0.000000
N	-0.922073	-1.418074	0.000000
O	-3.179743	-1.227174	0.000000
C	1.724360	1.019967	0.000000
C	1.542616	-1.446996	0.000000
S	2.805652	-0.284252	0.000000
O	-0.794683	2.624934	0.000000
H	2.065696	2.040478	0.000000
H	-1.026972	-2.418963	0.000000
H	-2.805949	1.192709	0.000000
H	1.758106	-2.501018	0.000000

*th* Uracil  
B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.331515	-0.831404	0.000000
C	0.432562	0.592015	0.000000
C	-0.770711	1.422232	0.000000
C	-2.097840	-0.709441	0.000000
N	-1.946910	0.674547	0.000000
N	-0.924341	-1.420753	0.000000
O	-3.193848	-1.236312	0.000000
C	1.732089	1.025676	0.000000
C	1.547407	-1.454186	0.000000
S	2.819924	-0.286775	0.000000
O	-0.799805	2.637633	0.000000
H	2.078765	2.044532	0.000000
H	-1.026354	-2.422809	0.000000
H	-2.812548	1.195754	0.000000
H	1.765166	-2.508120	0.000000

*th* Uracil  
PBE0/6-311++G(2df,2p)  
Water

C	0.325732	-0.826966	0.000000
C	0.427384	0.590800	0.000000
C	-0.772111	1.413743	0.000000
C	-2.083211	-0.706451	0.000000
N	-1.937704	0.670908	0.000000
N	-0.923055	-1.415277	0.000000
O	-3.183560	-1.226874	0.000000
C	1.727179	1.021850	0.000000
C	1.540540	-1.448064	0.000000
S	2.799234	-0.286831	0.000000
O	-0.803733	2.629524	0.000000
H	2.083264	2.039520	0.000000
H	-1.019249	-2.417720	0.000000
H	-2.805091	1.187936	0.000000
H	1.759450	-2.503509	0.000000

*th* Uracil  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.327224	-0.828841	0.000000
C	0.428395	0.588570	0.000000
C	-0.771359	1.417253	0.000000
C	-2.088742	-0.705688	0.000000
N	-1.938963	0.671917	0.000000
N	-0.921901	-1.415356	0.000000
O	-3.182122	-1.229477	0.000000
C	1.726539	1.020257	0.000000
C	1.543405	-1.448689	0.000000
S	2.802109	-0.286307	0.000000
O	-0.798256	2.628736	0.000000
H	2.075276	2.040606	0.000000
H	-1.025438	-2.416166	0.000000
H	-2.805603	1.189616	0.000000
H	1.764505	-2.503842	0.000000

## A2.2 Excited State Geometries of the Modified Nucleobases

*th* Adenine

*CAM-B3LYP/6-311++G(2df,2p)*

*Water*

C	0.193750	-0.973043	-0.000254
C	-0.001829	0.452111	-0.027551
C	-1.334919	0.891964	-0.040549
C	-2.059935	-1.278108	0.014569
N	-0.812212	-1.835851	0.020477
C	1.192206	1.132842	-0.033666
C	1.557104	-1.314593	0.004705
S	2.590936	0.050938	-0.021943
H	1.360080	2.196787	-0.039601
H	1.952553	-2.316419	0.020037
N	-1.695286	2.189618	-0.064780
H	-2.670924	2.416946	-0.152162
H	-1.024475	2.923860	-0.203194
H	-2.881448	-1.983500	0.034578
N	-2.364861	-0.016304	-0.010666

*th* Adenine

*B3LYP/6-311++G(2df,2p)*

*Water*

C	0.288401	-0.947375	0.001470
C	-0.070452	0.449494	0.012107
C	-1.450934	0.738353	0.030620
C	-1.927895	-1.509926	0.003948
N	-0.622104	-1.928251	-0.004294
C	1.046030	1.265580	-0.001567
C	1.681147	-1.137124	-0.006617
S	2.564964	0.353300	-0.002821
H	1.092579	2.342076	-0.020095
H	2.192544	-2.085422	-0.007737
N	-1.963029	1.989495	0.045919
H	-2.957026	2.097923	0.162601
H	-1.384013	2.795767	0.204138
H	-2.664645	-2.304797	-0.001795
N	-2.380901	-0.285416	0.013554

*th* Adenine

*CAM-B3LYP/6-311++G(2df,2p)*

*Dioxane*

C	0.194037	-0.970851	-0.002969
C	-0.003209	0.453462	-0.009584
C	-1.332564	0.890331	-0.040793
C	-2.056726	-1.276433	-0.019914
N	-0.812812	-1.834513	-0.006001
C	1.195290	1.132451	0.022869
C	1.552962	-1.311291	0.014379
S	2.592246	0.056710	0.024027
H	1.360674	2.196193	0.063682
H	1.945484	-2.314015	0.014800
N	-1.697906	2.190300	-0.067071
H	-2.673922	2.397623	-0.191120
H	-1.031505	2.917026	-0.253707
H	-2.878328	-1.982141	-0.018987
N	-2.365693	-0.014051	-0.029611

*th* Adenine

*B3LYP/6-311++G(2df,2p)*

*Dioxane*

C	0.289505	-0.945871	0.000011
C	-0.068654	0.450100	-0.000028
C	-1.446986	0.735611	-0.000022
C	-1.925505	-1.507738	-0.000017
N	-0.623002	-1.926493	0.000002
C	1.052535	1.265267	-0.000074
C	1.677475	-1.137714	0.000035
S	2.567777	0.353716	0.000025
H	1.102278	2.341886	-0.000199
H	2.184247	-2.088059	0.000089
N	-1.964979	1.983989	0.000020
H	-2.965183	2.087812	0.000154
H	-1.391326	2.806896	0.000298
H	-2.662078	-2.302817	0.000001
N	-2.381454	-0.282905	-0.000058

*th* Adenine  
PBE0/6-311++G(2df,2p)  
Water

C	-0.287796	-0.942725	0.004396
C	0.070565	0.448212	-0.006779
C	1.448330	0.732906	-0.005794
C	1.915194	-1.507236	-0.005681
N	0.614031	-1.923324	0.003798
C	-1.044660	1.261106	-0.020019
C	-1.680351	-1.126503	0.011553
S	-2.551074	0.354853	0.003916
H	-1.092041	2.339362	-0.050139
H	-2.195077	-2.075072	0.027025
N	1.959396	1.977672	-0.012017
H	2.954604	2.086218	0.082593
H	1.381759	2.787433	0.124551
H	2.652335	-2.304387	-0.007538
N	2.371235	-0.288104	-0.016325

*th* Cytosine  
CAM-B3LYP/6-311++G(2df,2p)  
Water

C	0.749202	-0.502028	-0.003599
C	-0.034150	0.701346	-0.002460
C	-1.432411	0.551310	-0.027575
C	-1.336204	-1.769202	-0.016457
N	-2.044314	-0.657799	-0.025033
N	0.112000	-1.662971	-0.013601
O	-1.755474	-2.923494	-0.013246
C	0.788100	1.803472	0.033154
C	2.130115	-0.286552	0.010361
S	2.515376	1.366208	0.026750
H	0.521661	2.845358	0.066877
H	0.609153	-2.543010	-0.004402
H	2.891928	-1.047593	0.005269
N	-2.251322	1.641094	-0.004839
H	-3.218293	1.471339	-0.226397
H	-1.889547	2.530141	-0.304802

*th* Adenine  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.194685	-0.967946	-0.004450
C	-0.000001	0.454432	-0.009873
C	-1.332213	0.894602	-0.014328
C	-2.058181	-1.270230	0.032690
N	-0.818328	-1.834387	0.030883
C	1.206604	1.130111	0.008019
C	1.551454	-1.312563	-0.024386
S	2.593029	0.059590	-0.051895
H	1.377581	2.193137	0.091396
H	1.949162	-2.315282	-0.046551
N	-1.697105	2.193695	-0.064485
H	-2.674270	2.388912	-0.202631
H	-1.037442	2.903077	-0.328787
H	-2.883647	-1.975290	0.046579
N	-2.370588	-0.004607	0.037820

*th* Cytosine  
CAM-B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.756693	-0.501345	0.008035
C	-0.034275	0.699215	0.002530
C	-1.434025	0.545177	0.029475
C	-1.349287	-1.771587	0.013039
N	-2.050845	-0.654787	0.024049
N	0.116856	-1.660742	0.017205
O	-1.745929	-2.922939	0.004253
C	0.786159	1.802347	-0.043862
C	2.132747	-0.281757	-0.004419
S	2.515329	1.378319	-0.024615
H	0.514419	2.841683	-0.103635
H	0.603532	-2.546056	0.002742
H	2.89747	-1.039729	0.004868
N	-2.249496	1.639119	0.014262
H	-3.219831	1.455403	0.207700
H	-1.894106	2.517728	0.348372

*th* Cytosine  
B3LYP/6-311++G(2df,2p)  
Water

C	-0.337965	-0.837448	0.011887
C	-0.360090	0.603684	-0.011024
C	0.889225	1.275900	-0.004679
C	2.119032	-0.706098	-0.009304
N	2.083994	0.613644	-0.020120
N	0.851585	-1.444537	0.023081
O	3.114127	-1.434359	-0.02056
C	-1.659216	1.054113	-0.040552
C	-1.602066	-1.432525	0.017629
S	-2.863219	-0.264654	-0.007902
H	-2.017693	2.068786	-0.082713
H	0.936272	-2.452079	0.026038
H	-1.813022	-2.488495	0.039680
N	0.955846	2.646741	-0.052886
H	1.855047	3.038765	0.180710
H	0.173563	3.173412	0.302233

*th* Cytosine  
PBE0/6-311++G(2df,2p)  
Water

C	0.752354	-0.502859	0.005322
C	-0.029150	0.701190	-0.002538
C	-1.435248	0.560204	0.022679
C	-1.343112	-1.760139	0.012567
N	-2.053905	-0.650191	0.015692
N	0.107045	-1.664845	0.020239
O	-1.760750	-2.916194	0.006533
C	0.801564	1.795796	-0.040585
C	2.132324	-0.291729	-0.007021
S	2.520101	1.368193	-0.031762
H	0.536076	2.840903	-0.077838
H	0.598247	-2.547819	0.013632
H	2.898202	-1.051651	0.002295
N	-2.250626	1.655811	-0.010215
H	-3.211255	1.479572	0.236119
H	-1.886907	2.528520	0.334881

*th* Cytosine  
B3LYP/6-311++G(2df,2p)  
Dioxane

C	-0.344184	-0.840748	0.020551
C	-0.361510	0.601538	-0.013490
C	0.888913	1.273628	-0.002496
C	2.130677	-0.695163	-0.019030
N	2.085007	0.624424	-0.031107
N	0.849192	-1.442260	0.049426
O	3.106686	-1.432975	-0.054864
C	-1.660843	1.051491	-0.059401
C	-1.604299	-1.433070	0.021424
S	-2.870460	-0.260004	-0.017543
H	-2.014762	2.066349	-0.128420
H	0.946725	-2.448380	0.047750
H	-1.817040	-2.488557	0.048821
N	0.947144	2.645861	-0.035150
H	1.860304	3.02825	0.155980
H	0.183870	3.164464	0.369068

*th* Cytosine  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.758956	-0.502075	0.012148
C	-0.027158	0.70019	-0.004805
C	-1.434053	0.558408	0.023679
C	-1.357248	-1.757396	0.005501
N	-2.060768	-0.642356	0.008176
N	0.107541	-1.661625	0.037356
O	-1.755109	-2.911408	-0.017246
C	0.804002	1.794491	-0.056960
C	2.134275	-0.290445	-0.001888
S	2.522937	1.375717	-0.036897
H	0.534796	2.837372	-0.120309
H	0.586472	-2.550837	0.026450
H	2.901493	-1.048839	0.012977
N	-2.24359	1.659639	0.002867
H	-3.209258	1.467832	0.215372
H	-1.888326	2.516092	0.393578

*th* Guanine

*CAM-B3LYP/6-311++G(2df,2p)*

*Water*

C	0.726507	-0.434766	-0.015660
C	-0.046664	0.783869	-0.002972
C	-1.451354	0.697595	-0.025987
C	-1.150405	-1.725641	-0.069204
N	0.194693	-1.652348	-0.047920
C	0.771020	1.893228	0.031323
C	2.107739	-0.198337	0.009875
S	2.499488	1.459337	0.049156
H	0.485849	2.930143	0.047282
H	2.867539	-0.961437	0.006272
H	-2.941402	-0.745813	-0.077141
N	-1.937167	-0.648273	-0.060663
O	-2.291233	1.613512	-0.019845
N	-1.683252	-2.939688	-0.100351
H	-2.675570	-3.097816	-0.117861
H	-1.063356	-3.731225	-0.106303

*th* Guanine

*B3LYP/6-311++G(2df,2p)*

*Water*

C	0.317318	-0.791886	-0.000104
C	0.445824	0.652192	-0.000110
C	-0.722997	1.452588	-0.000070
C	-1.972746	-0.665108	0.000071
N	-0.862643	-1.436058	-0.000004
C	1.773944	1.030075	-0.000200
C	1.548084	-1.450630	-0.000129
S	2.891143	-0.363735	-0.000214
H	2.180708	2.026715	-0.000296
H	1.688973	-2.518642	-0.000069
H	-2.801057	1.209333	0.000095
N	-1.944523	0.67586	-0.000315
O	-0.829619	2.69696	0.000025
N	-3.137253	-1.310448	0.000520
H	-4.024407	-0.83621	0.000836
H	-3.124589	-2.316798	0.000704

*th* Guanine

*CAM-B3LYP/6-311++G(2df,2p)*

*Dioxane*

C	0.725753	-0.432015	-0.015600
C	-0.046284	0.785815	-0.002926
C	-1.452410	0.712342	-0.025366
C	-1.149059	-1.721949	-0.069072
N	0.191220	-1.653851	-0.047869
C	0.776242	1.890841	0.031268
C	2.103628	-0.199965	0.009650
S	2.504279	1.459635	0.048829
H	0.484502	2.925967	0.047192
H	2.857916	-0.968353	0.005662
H	-2.942943	-0.736934	-0.075307
N	-1.938744	-0.650838	-0.060073
O	-2.296365	1.611209	-0.019681
N	-1.681210	-2.944736	-0.100845
H	-2.671720	-3.106488	-0.118900
H	-1.052375	-3.728340	-0.106961

*th* Guanine

*B3LYP/6-311++G(2df,2p)*

*Dioxane*

C	0.317787	-0.788593	-0.000074
C	0.446851	0.653502	-0.00006
C	-0.714433	1.465339	-0.000032
C	-1.971069	-0.662534	0.000069
N	-0.866638	-1.434343	0.000001
C	1.776480	1.025350	-0.000105
C	1.543464	-1.449636	-0.000162
S	2.893813	-0.364741	-0.000143
H	2.176501	2.024763	-0.000116
H	1.678632	-2.518229	-0.000172
H	-2.798569	1.217369	0.000464
N	-1.949092	0.674365	-0.000249
O	-0.83483	2.698068	0.000143
N	-3.138865	-1.318166	0.000393
H	-4.027984	-0.850127	0.0002
H	-3.111778	-2.323596	0.000464

*th* Guanine  
PBE0/6-311++G(2df,2p)  
Water

C	0.728221	-0.434809	-0.015895
C	-0.046873	0.782585	-0.002647
C	-1.45711	0.703574	-0.025743
C	-1.152931	-1.722707	-0.069436
N	0.193512	-1.660155	-0.048666
C	0.777626	1.888353	0.031864
C	2.104873	-0.202889	0.009222
S	2.496514	1.464258	0.049064
H	0.49254	2.928264	0.04834
H	2.870963	-0.962876	0.005321
H	-2.947176	-0.746457	-0.077441
N	-1.943876	-0.646766	-0.061188
O	-2.297399	1.619891	-0.019622
N	-1.68029	-2.938790	-0.100153
H	-2.671486	-3.103325	-0.117474
H	-1.054679	-3.725811	-0.105547

*th* Uracil  
CAM-B3LYP/6-311++G(2df,2p)  
Water

C	0.709769	-0.466910	-0.052011
C	-0.051965	0.743141	-0.006556
C	-1.46246	0.686330	0.010398
C	-1.317306	-1.79271	-0.064328
N	-1.999265	-0.634235	-0.021384
N	0.086424	-1.647608	-0.078954
O	-1.80133	-2.903434	-0.090314
C	0.793345	1.835447	0.01459
C	2.097286	-0.276415	-0.064083
S	2.520882	1.362128	-0.021367
O	-2.254742	1.639265	0.048995
H	0.531756	2.877702	0.048759
H	0.601748	-2.516904	-0.111441
H	-3.005690	-0.705147	-0.010321
H	2.840698	-1.055308	-0.096123

*th* Guanine  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.311999	-0.787554	-0.000148
C	0.441236	0.648061	-0.000071
C	-0.718216	1.456383	-0.000053
C	-1.962759	-0.660796	0.000041
N	-0.864264	-1.433759	-0.000035
C	1.770518	1.015032	-0.000076
C	1.539913	-1.445404	-0.000218
S	2.874935	-0.366494	-0.000067
H	2.17234	2.015765	-0.000084
H	1.675904	-2.515751	-0.000216
H	-2.78734	1.213983	0.000022
N	-1.938797	0.671103	-0.000581
O	-0.837894	2.684844	0.00031
N	-3.127438	-1.310784	0.000776
H	-4.014443	-0.841337	0.001402
H	-3.098802	-2.315091	0.001107

*th* Uracil  
CAM-B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.712858	-0.464677	-0.051751
C	-0.050279	0.742674	-0.006427
C	-1.464811	0.693088	0.010494
C	-1.324402	-1.795306	-0.064567
N	-2.003273	-0.637102	-0.021422
N	0.087114	-1.645639	-0.078598
O	-1.792164	-2.909053	-0.091357
C	0.795861	1.834358	0.014564
C	2.099135	-0.27591	-0.063977
S	2.526595	1.366808	-0.021442
O	-2.250981	1.637782	0.048415
H	0.527186	2.874729	0.048968
H	0.594352	-2.518547	-0.11085
H	-3.009821	-0.702188	-0.010363
H	2.841780	-1.055677	-0.095827

*th* Uracil  
B3LYP/6-311++G(2df,2p)  
Water

C	0.720367	-0.467980	-0.052491
C	-0.048889	0.747850	-0.006853
C	-1.470107	0.698555	0.010090
C	-1.327127	-1.797952	-0.064622
N	-2.011129	-0.638808	-0.023401
N	0.087813	-1.654747	-0.079986
O	-1.803014	-2.918594	-0.088428
C	0.799181	1.835033	0.014892
C	2.106635	-0.279746	-0.064107
S	2.534321	1.378978	-0.020407
O	-2.269392	1.658136	0.049456
H	0.534432	2.877621	0.049383
H	0.599586	-2.526658	-0.111787
H	-3.017658	-0.711283	-0.009665
H	2.854132	-1.055066	-0.096212

*th* Uracil  
PBE0/6-311++G(2df,2p)  
Water

C	0.713994	-0.467321	-0.052335
C	-0.051857	0.743242	-0.006887
C	-1.468503	0.69252	0.010050
C	-1.320031	-1.790308	-0.064559
N	-2.00211	-0.634554	-0.023048
N	0.085974	-1.64875	-0.080226
O	-1.796483	-2.906255	-0.088514
C	0.799772	1.827215	0.014565
C	2.099202	-0.2775	-0.063707
S	2.518148	1.369652	-0.020054
O	-2.263944	1.647721	0.049404
H	0.536444	2.872171	0.048681
H	0.596878	-2.520311	-0.112126
H	-3.007653	-0.709233	-0.009720
H	2.849319	-1.052948	-0.095665

*th* Uracil  
B3LYP/6-311++G(2df,2p)  
Dioxane

C	0.723647	-0.465507	-0.052015
C	-0.047457	0.747768	-0.006518
C	-1.471392	0.705769	0.010381
C	-1.335081	-1.800495	-0.064474
N	-2.015306	-0.6425	-0.022450
N	0.088047	-1.652646	-0.078869
O	-1.792851	-2.925427	-0.090776
C	0.80191	1.834332	0.014979
C	2.108432	-0.279486	-0.064124
S	2.539348	1.38386	-0.020649
O	-2.265036	1.656826	0.048685
H	0.529954	2.875037	0.049323
H	0.591749	-2.528404	-0.110707
H	-3.022112	-0.708124	-0.01053
H	2.855297	-1.055663	-0.096396

*th* Uracil  
PBE0/6-311++G(2df,2p)  
Dioxane

C	0.717089	-0.464905	-0.052033
C	-0.05048	0.742912	-0.006743
C	-1.469913	0.699119	0.010094
C	-1.327125	-1.792396	-0.064382
N	-2.005775	-0.637665	-0.022298
N	0.086066	-1.646734	-0.079206
O	-1.786605	-2.912543	-0.090306
C	0.802679	1.825979	0.014721
C	2.100947	-0.27698	-0.063793
S	2.523006	1.374341	-0.020436
O	-2.260185	1.646337	0.04848
H	0.531741	2.868908	0.049146
H	0.588716	-2.522093	-0.111111
H	-3.011509	-0.705807	-0.010499
H	2.850498	-1.053133	-0.095775

# Appendix B

## Appendix to Chapter 3

## B1 Thermochemistry

**Table B1:** Relative thermodynamic parameters (kcal/mol) for <sup>th</sup>Guanine nucleobase tautomers in the gas phase and water. Relative energies were computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	Gas phase			Water			
	$\Delta E$	$\Delta H$	$\Delta G$	$\Delta E$	$\Delta H$	$\Delta G$	$\Delta\Delta G$
<sup>th</sup> A (i, 1H)	6.73	6.53	6.85	6.47	6.14	6.79	-0.06
(i, 3H)	14.73	14.63	14.60	10.03	9.74	10.22	-4.38
<sup>th</sup> G (k-i, 1H, 3H)	5.17	5.15	5.02	5.71	5.61	5.63	0.61
(e-a)	5.98	5.98	6.05	9.59	9.55	9.66	3.62
(k-a, 3H)	7.99	8.12	7.81	1.79	2.03	1.39	-6.42
(e-i, 3H)	15.81	15.73	15.73	17.87	17.74	17.85	2.12
<sup>th</sup> U (k-i, 3H)	-0.97	-1.26	-0.76	3.24	3.03	3.35	4.10
(e-i)	27.37	27.36	27.33	24.37	24.29	24.39	-2.94

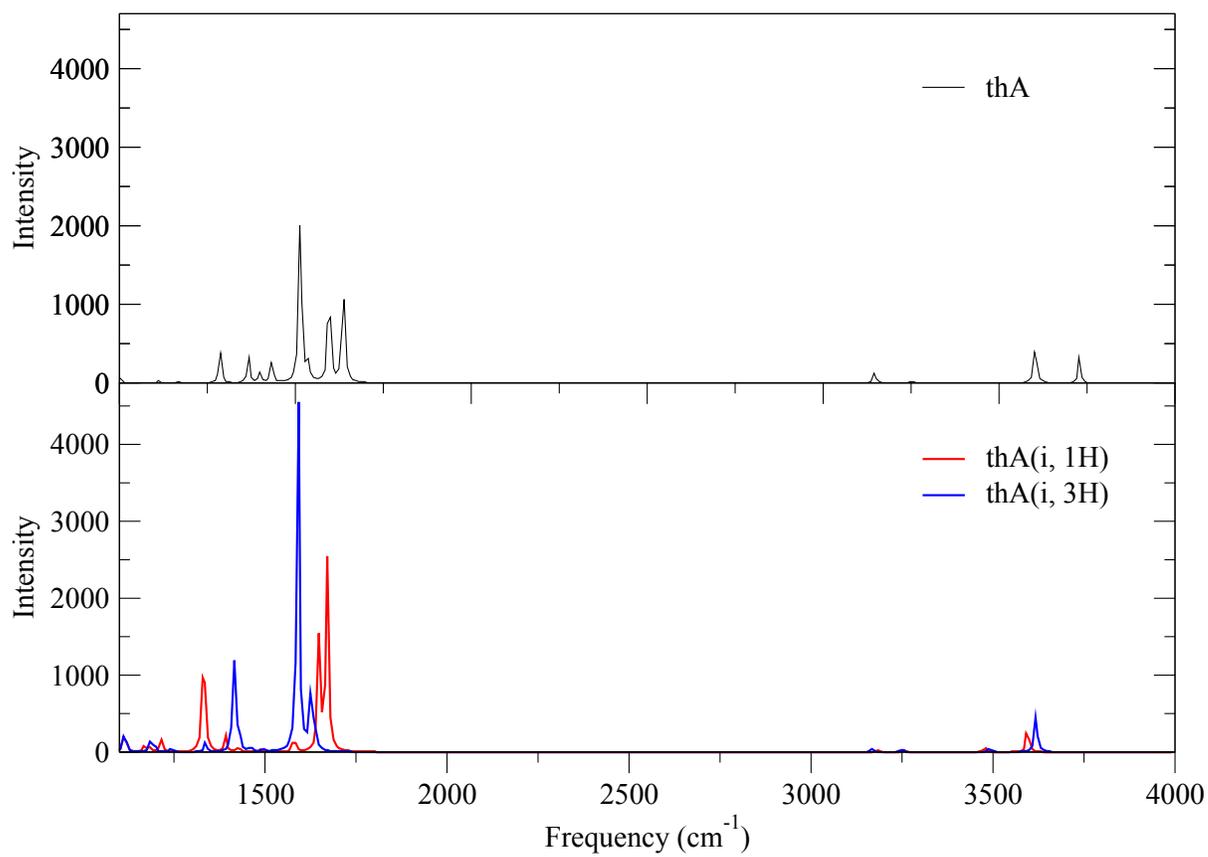
Positive values reflect the difference in energies of tautomers compared to the lowest energy (canonical) tautomer. Thermochemistry was calculated at 298.15K.

### B1.1 CCSD(T) Energies

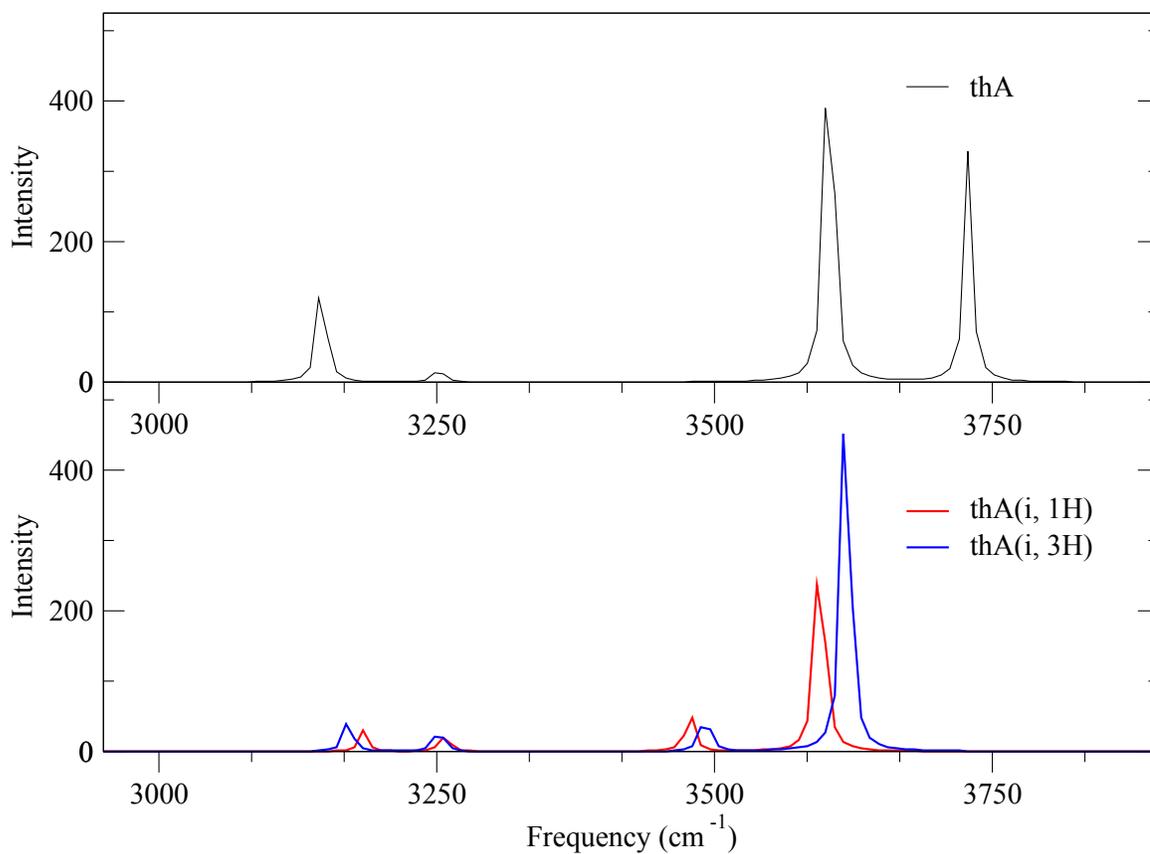
**Table B2:** Relative ground state energies (kcal/mol) for the abundant tautomers of the modified nucleobases in the gas phase and solution computed at CCSD(T)/cc-pVDZ level of theory in comparison with the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	Gas phase		Water	
	B3LYP	CCSD(T)	B3LYP	CCSD(T)
<sup>th</sup> Adenine (i, 1H)	6.73	4.96	6.47	4.20
<sup>th</sup> Guanine (k-a, 3H)	7.99	8.44	1.79	2.87
(k-i, 1H, 3H)	5.17	5.37	5.17	6.02
(e-a)	5.98	8.27	9.59	11.20
<sup>th</sup> Cytosine (k-i, 1H)	-0.97	-4.01	3.24	0.64

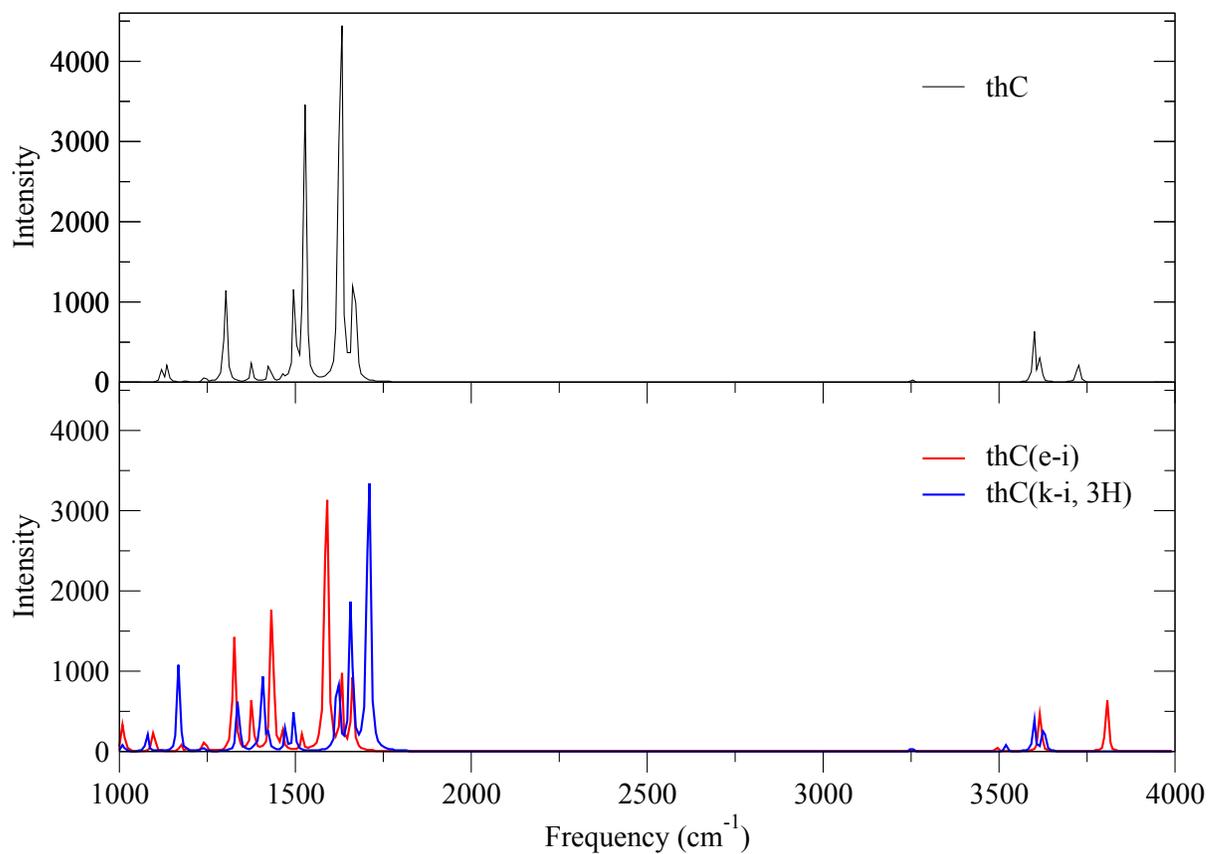
## B2 IR Spectra



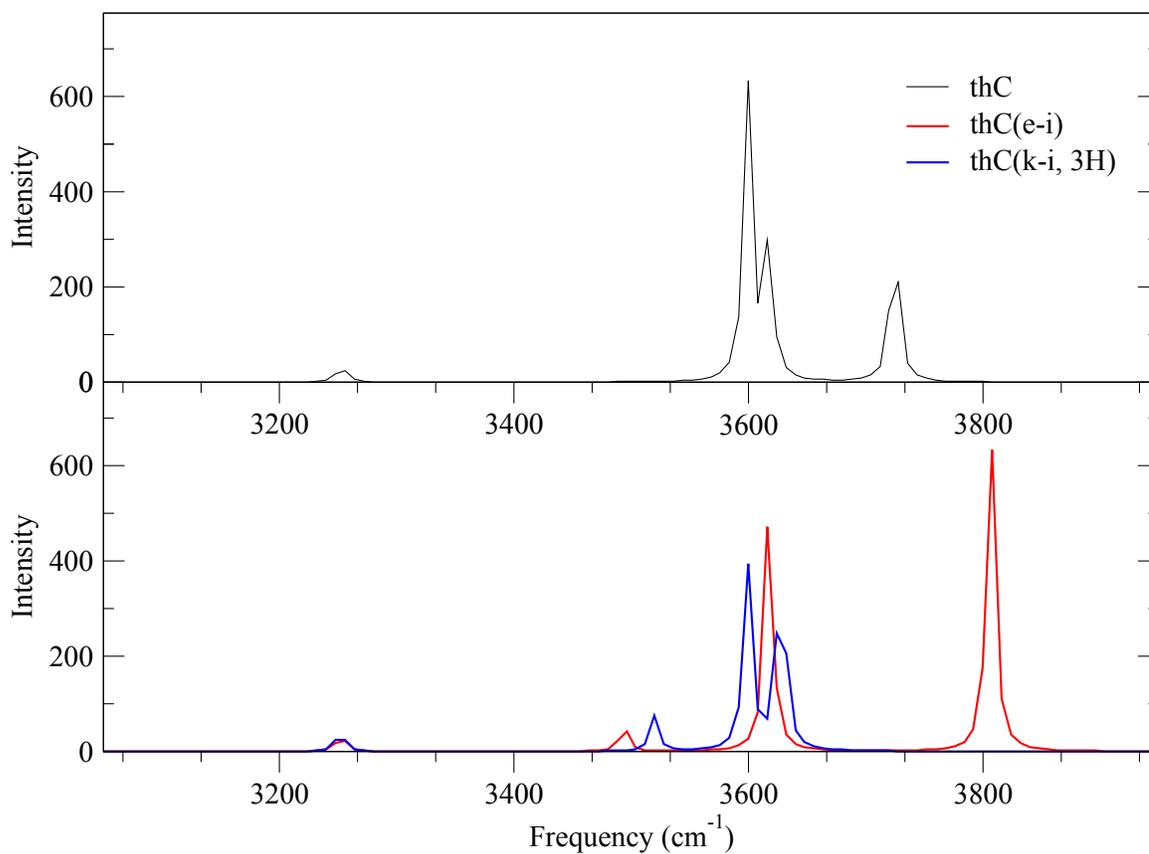
**Figure B1:** Computed IR spectra of the *thA* tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).



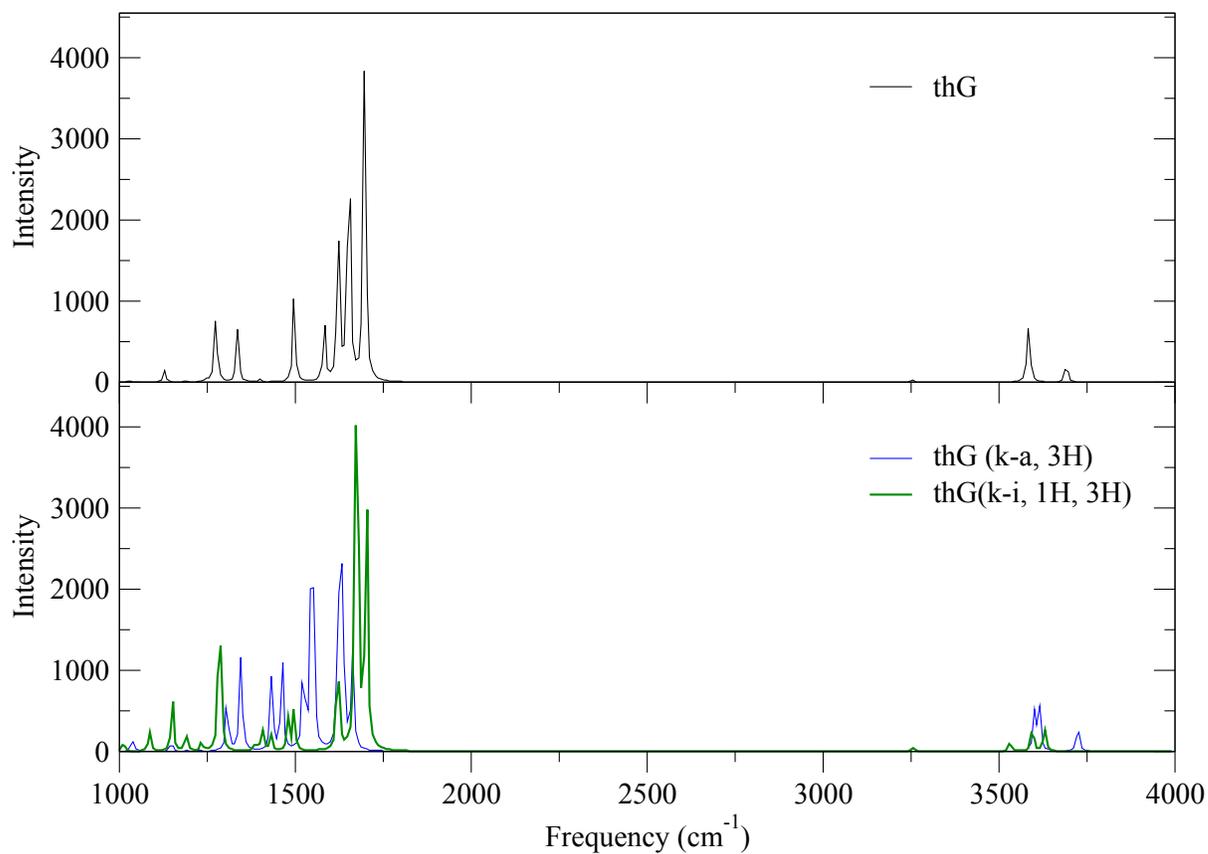
**Figure B2:** *Computed IR spectra in the OH region of the <sup>th</sup>A tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).*



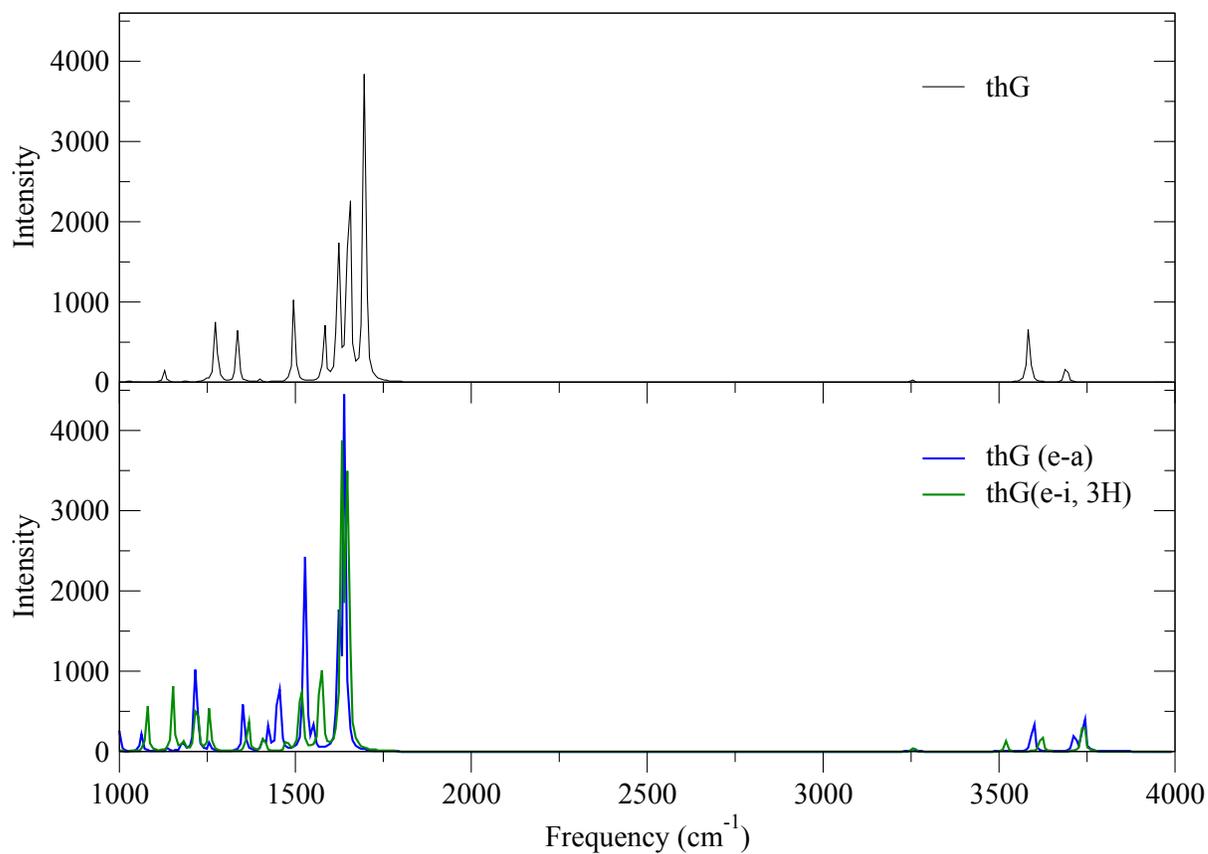
**Figure B3:** Computed IR spectra of the *thC* tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).



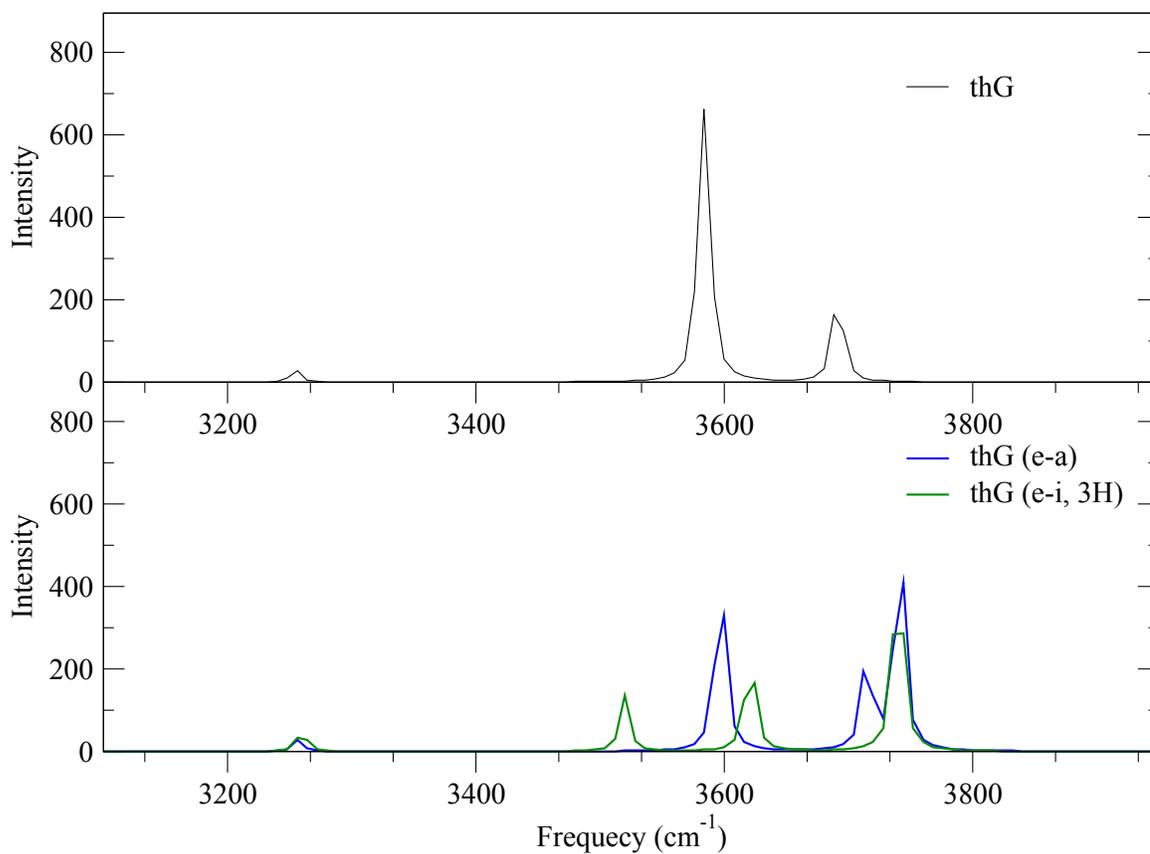
**Figure B4:** *Computed IR spectra in the OH region of the <sup>th</sup>C tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).*



**Figure B5:** *Computed IR spectra of the <sup>th</sup>G tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).*



**Figure B6:** *Computed IR spectra of the <sup>th</sup>G enol tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).*



**Figure B7:** *Computed IR spectra in the NH region of the <sup>th</sup>G enol tautomers at the B3LYP/6-311++G(2df,2p) level of theory in water (PCM).*

## B3 Vertical Excitation Energies of Low-Lying Singlet States

**Table B3:** Vertical excitation energies of the singlet state  $S_1$  of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
<sup>th</sup> A (a)	3.7564	330.06	0.1046	39 → 40 0.69730
(i, 1H)	4.1716	297.21	0.1416	39 → 40 0.68510
(i, 3H)	4.0357	307.22	0.1678	39 → 40 0.69030
<sup>th</sup> G (k-a,1H)	3.9219	316.13	0.0633	43 → 44 0.69329
(e-a)	3.4915	355.10	0.0607	43 → 44 0.69879
(k-i,1H,3H)	3.8427	322.65	0.0598	43 → 44 0.69590
(k-a, 3H)	3.9248	315.90	0.0002	42 → 44 0.68875
<sup>th</sup> C (k-a)	3.7868	327.41	0.0793	43 → 44 0.69722
(k-i, 3H)	4.1909	295.84	0.0933	43 → 44 0.69293

**Table B4:** Vertical excitation energies of the singlet state  $S_2$  of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
<sup>th</sup> A (a)	3.9143	316.74	0.0020	38 → 40 0.69972
(i, 1H)	4.7139	263.02	0.0137	38 → 40 0.64034
				39 → 42 0.25815
(i, 3H)	4.1167	301.17	0.0004	37 → 40 0.69728
<sup>th</sup> G (k-a,1H)	4.6693	265.53	0.0094	43 → 45 0.69377
(e-a)	4.2564	291.29	0.0012	41 → 44 0.69820
(k-i,1H,3H)	4.5650	271.60	0.0001	41 → 44 0.66608
(k-a, 3H)	4.3715	283.62	0.0973	43 → 44 0.68997
<sup>th</sup> C (k-a)	4.2799	289.69	0.0004	41 → 44 0.66866
(k-i, 3H)	4.8559	255.33	0.0484	42 → 44 0.68271

**Table B5:** Vertical excitation energies of the singlet state  $S_3$  of the modified nucleobase tautomers in the gas phase. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$ (a)	4.7963	258.50	0.0471	37 $\rightarrow$ 40	0.58602	
				39 $\rightarrow$ 42	-0.32549	
	(i, 1H)	4.8372	256.31	0.0008	39 $\rightarrow$ 41	0.69805
	(i, 3H)	4.7473	261.17	0.0194	38 $\rightarrow$ 40	0.67868
${}^{th}G$	(k-a,1H)	4.7069	263.41	0.0002	41 $\rightarrow$ 44	0.68579
	(e-a)	4.6618	265.96	0.0802	42 $\rightarrow$ 44	0.62711
	(k-i,1H,3H)	4.9375	251.11	0.0036	43 $\rightarrow$ 45	0.68695
	(k-a, 3H)	4.7625	260.34	0.0006	43 $\rightarrow$ 45	0.69528
${}^{th}C$	(k-a)	4.5696	271.32	0.0025	39 $\rightarrow$ 44	0.64324
					40 $\rightarrow$ 44	0.21895
	(k-i, 3H)	4.9134	252.34	0.0016	40 $\rightarrow$ 44	0.65793
					39 $\rightarrow$ 44	0.23094

**Table B6:** Vertical excitation energies of the singlet state  $S_1$  of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition					
${}^{th}A$ (a)	3.7507	330.56	0.1367	39 $\rightarrow$ 40	0.70012				
				(i, 1H)	4.0826	303.69	0.1928	39 $\rightarrow$ 40	0.69276
				(i, 3H)	3.9745	311.95	0.2229	39 $\rightarrow$ 40	0.69391
${}^{th}G$	(k-a, 1H)	3.7602	329.73	0.0810	43 $\rightarrow$ 44	0.69758			
	(e-a)	3.5027	353.96	0.0804	43 $\rightarrow$ 44	0.70012			
	(k-i,1H,3H)	3.7511	330.52	0.0726	43 $\rightarrow$ 44	0.69884			
	(k-a, 3H)	4.2143	294.20	0.1153	43 $\rightarrow$ 44	0.69620			
${}^{th}C$	(k-a)	3.8398	322.90	0.0926	43 $\rightarrow$ 44	0.69969			
	(k-i, 3H)	4.1456	299.08	0.1136	43 $\rightarrow$ 44	0.69620			

**Table B7:** Vertical excitation energies of the singlet state  $S_2$  of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$ (a)	4.2108	294.44	0.0014	38 $\rightarrow$ 40	0.70135	
	4.6985	263.88	0.0309	38 $\rightarrow$ 40	0.65897	
				39 $\rightarrow$ 41	0.21819	
(i, 3H)	4.4891	276.19	0.0008	37 $\rightarrow$ 40	0.69790	
${}^{th}G$	(k-a, 1H)	4.8302	256.69	0.0000	41 $\rightarrow$ 44	0.67219
	(e-a)	4.3890	282.49	0.0016	41 $\rightarrow$ 44	0.70107
	(k-i,1H,3H)	4.7321	262.01	0.0001	40 $\rightarrow$ 44	0.66626
	(k-a, 3H)	4.3196	287.03	0.0001	41 $\rightarrow$ 44	0.68779
${}^{th}C$	(k-a)	4.7339	261.91	0.2533	42 $\rightarrow$ 44	0.67684
	(k-i, 3H)	4.9021	252.92	0.1312	42 $\rightarrow$ 44	0.69496

**Table B8:** Vertical excitation energies of the singlet state  $S_3$  of the modified tautomers in water. Computed at the B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$ (a)	4.7343	261.88	0.0751	37 $\rightarrow$ 40	0.62783	
				39 $\rightarrow$ 41	-0.29563	
	4.9854	248.69	0.1303	39 $\rightarrow$ 41	0.63708	
				38 $\rightarrow$ 40	-0.20542	
(i, 3H)	4.6185	268.45	0.0559	38 $\rightarrow$ 40	0.68053	
${}^{th}G$	(k-a, 1H)	4.8894	253.58	0.0135	42 $\rightarrow$ 44	0.57954
					43 $\rightarrow$ 45	-0.28212
					43 $\rightarrow$ 46	0.24212
	(e-a)	4.6272	267.95	0.1359	42 $\rightarrow$ 44	0.64811
					43 $\rightarrow$ 47	0.21531
	(k-i,1H,3H)	4.9058	252.73	0.1180	42 $\rightarrow$ 44	0.65044
					43 $\rightarrow$ 46	-0.20277
	(k-a, 3H)	4.9071	252.67	0.2421	42 $\rightarrow$ 44	0.62322
40 $\rightarrow$ 44					0.23790	
${}^{th}C$	(k-a)	4.8192	257.27	0.0002	40 $\rightarrow$ 44	0.62405
					39 $\rightarrow$ 44	0.30743
	(k-i, 3H)	5.1814	239.29	0.0027	40 $\rightarrow$ 44	0.66148
				39 $\rightarrow$ 44	-0.21868	

**Table B9:** Vertical excitation energies of the singlet state  $S_1$  of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.0125	309.00	0.1319	39 $\rightarrow$ 40 0.69966
(i, 1H)	4.5294	273.73	0.1879	39 $\rightarrow$ 40 0.68783
(i, 3H)	4.3971	281.97	0.2321	39 $\rightarrow$ 40 0.69358
${}^{th}G$ (k-a,1H)	4.3271	286.53	0.0922	43 $\rightarrow$ 44 0.69404
(e-a)	3.8422	322.69	0.0840	43 $\rightarrow$ 44 0.69903
(k-i, 1H, 3H)	4.3751	283.39	0.0923	43 $\rightarrow$ 44 0.69188
(k-a, 3H)	4.3295	286.37	0.0002	42 $\rightarrow$ 44 0.63535
${}^{th}C$ (k-a)	4.2013	295.11	0.1104	43 $\rightarrow$ 44 0.69551
(k-i, 3H)	4.6081	269.06	0.1377	43 $\rightarrow$ 44 0.69380

**Table B10:** Vertical excitation energies of the singlet state  $S_2$  of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.4103	281.13	0.0019	38 $\rightarrow$ 40 0.67540
(i, 1H)	4.9947	248.23	0.0205	38 $\rightarrow$ 40 0.59525
				39 $\rightarrow$ 42 0.35366
(i, 3H)	4.5746	271.03	0.0004	37 $\rightarrow$ 40 0.65449
${}^{th}G$ (k-a,1H)	5.0570	245.17	0.0001	41 $\rightarrow$ 44 0.63435
(e-a)	4.7902	258.83	0.0022	41 $\rightarrow$ 44 0.67846
(k-i, 1H, 3H)	4.9699	249.47	0.0001	40 $\rightarrow$ 44 0.57739
				39 $\rightarrow$ 44 -0.29257
(k-a, 3H)	4.7420	261.46	0.1349	43 $\rightarrow$ 44 0.68647
${}^{th}C$ (k-a)	4.8924	253.42	0.0025	39 $\rightarrow$ 44 0.53154
				41 $\rightarrow$ 44 -0.38207
(k-i, 3H)	5.2300	237.06	0.0748	42 $\rightarrow$ 44 0.66647

**Table B11:** Vertical excitation energies of the singlet state  $S_3$  of the modified nucleobase tautomers in the gas phase. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$ (a)	4.9965	248.14	0.0667	37 $\rightarrow$ 40	0.59225	
				39 $\rightarrow$ 45	0.34166	
	(i, 1H)	5.3185	233.12	0.0009	39 $\rightarrow$ 41	0.65148
	(i, 3H)	5.0781	244.15	0.0347	38 $\rightarrow$ 40	0.67188
${}^{th}G$	(k-a,1H)	5.1542	240.55	0.0032	42 $\rightarrow$ 44	0.54345
					43 $\rightarrow$ 48	-0.20059
					43 $\rightarrow$ 49	0.30548
	(e-a)	4.9110	252.46	0.1031	42 $\rightarrow$ 44	0.62864
					43 $\rightarrow$ 50	0.25588
	(k-i, 1H, 3H)	5.3088	233.55	0.0729	42 $\rightarrow$ 44	0.62961
				43 $\rightarrow$ 49	-0.29351	
	(k-a, 3H)	5.2357	236.81	0.0012	43 $\rightarrow$ 45	0.63458
${}^{th}C$	(k-a)	5.0692	244.58	0.2196	42 $\rightarrow$ 44	0.64603
					(k-i, 3H)	5.4164
					40 $\rightarrow$ 44	0.57030

**Table B12:** Vertical excitation energies of the singlet state  $S_1$  of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$	(a)	4.0053	309.55	0.1716	39 $\rightarrow$ 40	0.70004
	(i, 1H)	4.4248	280.20	0.2560	39 $\rightarrow$ 40	0.69494
	(i, 3H)	4.2997	288.36	0.2999	39 $\rightarrow$ 40	0.69636
${}^{th}G$	(k-a,1H)	4.1756	296.93	0.1193	43 $\rightarrow$ 44	0.69734
	(e-a)	3.8574	321.42	0.1103	43 $\rightarrow$ 44	0.69996
	(k-i, 1H, 3H)	4.2683	290.47	0.1118	43 $\rightarrow$ 44	0.69513
	(k-a, 3H)	4.5705	271.27	0.1552	43 $\rightarrow$ 44	0.69538
${}^{th}C$	(k-a)	4.2380	292.55	0.1293	43 $\rightarrow$ 44	0.69891
	(k-i, 3H)	4.5623	271.76	0.1704	43 $\rightarrow$ 44	0.69550

**Table B13:** Vertical excitation energies of the singlet state  $S_2$  of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.6960	264.02	0.0021	37 $\rightarrow$ 40 0.67432
	4.9658	249.68	0.0386	38 $\rightarrow$ 40 0.62947
				39 $\rightarrow$ 42 0.29296
	4.9166	252.17	0.0008	37 $\rightarrow$ 40 0.66221
${}^{th}G$ (k-a,1H)	5.1054	242.85	0.0259	42 $\rightarrow$ 44 0.59977
	4.8715	254.51	0.1706	43 $\rightarrow$ 48 0.34213
				42 $\rightarrow$ 44 0.64631
	5.1190	242.21	0.0001	43 $\rightarrow$ 49 0.23119
				40 $\rightarrow$ 44 0.58416
	4.7180	262.79	0.0000	39 $\rightarrow$ 44 -0.29666
40 $\rightarrow$ 44 0.65062				
${}^{th}C$ (k-a)	5.0204	246.96	0.3068	42 $\rightarrow$ 44 0.66451
	5.1499	240.75	0.1348	42 $\rightarrow$ 44 0.67724

**Table B14:** Vertical excitation energies of the singlet state  $S_3$  of the modified nucleobase tautomers in water. Computed at the CAM-B3LYP/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.9536	250.29	0.1006	38 $\rightarrow$ 40 0.61094
				39 $\rightarrow$ 44 0.30722
	5.3877	230.13	0.2357	39 $\rightarrow$ 42 0.60302
				38 $\rightarrow$ 40 -0.30122
4.9460	250.68	0.0766	38 $\rightarrow$ 40 0.67922	
${}^{th}G$ (k-a,1H)	5.1788	239.40	0.0000	41 $\rightarrow$ 44 0.50275
	4.9141	252.30	0.0033	40 $\rightarrow$ 44 0.41286
				41 $\rightarrow$ 44 0.67671
	5.1915	238.82	0.1467	42 $\rightarrow$ 44 0.65529
				43 $\rightarrow$ 48 0.23462
	5.1922	238.79	0.2622	42 $\rightarrow$ 44 0.67061
${}^{th}C$ (k-a)	5.2759	235.00	0.0015	39 $\rightarrow$ 44 0.63484
				40 $\rightarrow$ 44 0.20453
	5.6331	220.10	0.0044	40 $\rightarrow$ 44 0.59214
				39 $\rightarrow$ 44 -0.28206

**Table B15:** Vertical excitation energies of the singlet state  $S_1$  of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
$^{th}A$ (a)	3.8483	322.18	0.1104	39 $\rightarrow$ 40 0.69881
(i, 1H)	4.2857	289.3	0.1518	39 $\rightarrow$ 40 0.68712
(i, 3H)	4.1503	298.74	0.1808	39 $\rightarrow$ 40 0.69223
$^{th}G$ (k-a,1H)	4.0452	306.5	0.0690	43 $\rightarrow$ 44 0.69451
(e-a)	3.6004	344.37	0.0653	43 $\rightarrow$ 44 0.69960
(k-i, 1H, 3H)	3.9885	310.85	0.0656	43 $\rightarrow$ 44 0.69664
(k-a, 3H)	4.0314	307.55	0.0002	42 $\rightarrow$ 44 0.68056
$^{th}C$ (k-a)	3.9076	317.29	0.0849	43 $\rightarrow$ 44 0.69789
(k-i, 3H)	4.3188	287.08	0.1015	43 $\rightarrow$ 44 0.69433

**Table B16:** Vertical excitation energies of the singlet state  $S_2$  of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
$^{th}A$ (a)	4.0353	307.25	0.0018	38 $\rightarrow$ 40 0.69763
(i, 1H)	4.8349	256.44	0.0143	38 $\rightarrow$ 40 0.62952
				39 $\rightarrow$ 41 0.29295
(i, 3H)	4.2324	292.94	0.0004	37 $\rightarrow$ 40 0.69207
$^{th}G$ (k-a,1H)	4.8081	257.86	0.0000	41 $\rightarrow$ 44 0.67971
(e-a)	4.3870	282.62	0.0013	41 $\rightarrow$ 44 0.69706
(k-i, 1H, 3H)	4.6787	265.00	0.0001	41 $\rightarrow$ 44 0.65105
				39 $\rightarrow$ 44 -0.20965
(k-a, 3H)	4.4930	275.95	0.1053	43 $\rightarrow$ 44 0.69182
$^{th}C$ (k-a)	4.4656	277.64	0.0007	41 $\rightarrow$ 44 0.64957
				39 $\rightarrow$ 44 -0.20622
(k-i, 3H)	5.0095	247.50	0.0564	42 $\rightarrow$ 44 0.69464

**Table B17:** Vertical excitation energies of the singlet state  $S_3$  of the modified nucleobase tautomers in the gas phase. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$ (a)	4.9012	252.97	0.0525	37 $\rightarrow$ 40	0.59399	
				39 $\rightarrow$ 42	-0.34566	
	5.0971	243.24	0.1056	39 $\rightarrow$ 41	0.60494	
				38 $\rightarrow$ 40	-0.27565	
(i, 3H)	4.8785	254.15	0.0229	38 $\rightarrow$ 40	0.68081	
${}^{th}G$	(k-a,1H)	4.9501	250.47	0.0087	43 $\rightarrow$ 45	0.66598
	(e-a)	4.7745	259.68	0.0881	42 $\rightarrow$ 44	0.63257
					43 $\rightarrow$ 48	0.20306
	(k-i, 1H, 3H)	5.1467	240.90	0.0634	42 $\rightarrow$ 44	0.62459
					43 $\rightarrow$ 46	-0.27101
(k-a, 3H)	5.0329	246.35	0.0008	43 $\rightarrow$ 45	0.68874	
${}^{th}C$	(k-a)	4.7295	262.15	0.0021	39 $\rightarrow$ 44	0.64408
					41 $\rightarrow$ 44	0.21192
	(k-i, 3H)	5.0587	245.09	0.0018	40 $\rightarrow$ 44	0.64291
					39 $\rightarrow$ 44	-0.25304

**Table B18:** Vertical excitation energies of the singlet state  $S_1$  of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition		
${}^{th}A$	(a)	3.8422	322.69	0.1445	39 $\rightarrow$ 40	0.70080
	(i, 1H)	4.1931	295.69	0.2062	39 $\rightarrow$ 40	0.69437
	(i, 3H)	4.0812	303.80	0.2380	39 $\rightarrow$ 40	0.69560
${}^{th}G$	(k-a,1H)	3.8848	319.15	0.0883	43 $\rightarrow$ 44	0.69850
	(e-a)	3.6116	343.29	0.0864	43 $\rightarrow$ 44	0.70085
	(k-i, 1H, 3H)	3.8941	318.39	0.0798	43 $\rightarrow$ 44	0.69946
	(k-a, 3H)	4.3342	286.06	0.1237	43 $\rightarrow$ 44	0.69697
${}^{th}C$	(k-a)	3.9592	313.16	0.0996	43 $\rightarrow$ 44	0.70038
	(k-i, 3H)	4.2740	290.09	0.1241	43 $\rightarrow$ 44	0.69726

**Table B19:** Vertical excitation energies of the singlet state  $S_2$  of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.3222	286.86	0.0015	38 $\rightarrow$ 40 0.69837
	4.8209	257.18	0.0313	38 $\rightarrow$ 40 0.65288
				39 $\rightarrow$ 41 0.24365
(i, 3H)	4.5893	270.16	0.0008	37 $\rightarrow$ 40 0.69367
${}^{th}G$ (k-a,1H)	4.9211	251.94	0.0000	41 $\rightarrow$ 44 0.65361
	4.5098	274.92	0.0017	40 $\rightarrow$ 44 -0.20546
				41 $\rightarrow$ 44 0.69863
	4.8313	256.63	0.0001	39 $\rightarrow$ 44 -0.21592
				40 $\rightarrow$ 44 0.65189
(k-a, 3H)	4.4145	280.86	0.0001	41 $\rightarrow$ 44 0.67664
${}^{th}C$ (k-a)	4.8556	255.34	0.2699	42 $\rightarrow$ 44 0.67839
	(k-i, 3H)	5.0217	246.90	0.1349

**Table B20:** Vertical excitation energies of the singlet state  $S_3$  of the modified nucleobase tautomers in water. Computed at the PBE0/6-311++G(2df,2p) level of theory.

Tautomer	$\Delta E$ , eV	$\Delta E$ , nm	f	Transition
${}^{th}A$ (a)	4.8457	255.86	0.0808	37 $\rightarrow$ 40 0.62251
				39 $\rightarrow$ 41 -0.30246
	5.1202	242.15	0.1505	39 $\rightarrow$ 41 0.62934
				38 $\rightarrow$ 40 -0.23861
(i, 3H)	4.7524	260.89	0.0602	38 $\rightarrow$ 40 0.68335
${}^{th}G$ (k-a,1H)	5.0031	247.81	0.0159	42 $\rightarrow$ 44 0.58594
	4.7398	261.58	0.1457	43 $\rightarrow$ 45 0.34902
				42 $\rightarrow$ 44 0.65016
	5.0365	246.17	0.1262	42 $\rightarrow$ 44 0.65528
				43 $\rightarrow$ 46 -0.21497
(k-a, 3H)	5.0375	246.12	0.2560	42 $\rightarrow$ 44 0.64555
${}^{th}C$ (k-a)	4.9642	249.76	0.0008	39 $\rightarrow$ 44 0.49966
				40 $\rightarrow$ 44 0.47909
	(k-i, 3H)	5.3018	233.85	0.0029
				40 $\rightarrow$ 44 0.64888

## B4 Ground State Geometries of the Modified Nucleobases and their Tautomers

### B4.1 Gas-Phase

	<sup>th</sup> A (a)			(i, 3H)			
C	-0.305537	-0.956817	0.005770	C	1.197212	1.155222	0.000000
C	0.049197	0.438042	-0.003004	C	1.509840	-1.307190	0.000000
C	1.458580	0.724513	-0.002754	C	0.000000	0.490538	0.000000
C	1.870713	-1.535082	-0.004647	C	0.196997	-0.926399	0.000000
C	-1.049391	1.270298	-0.010327	C	-1.373976	1.046878	0.000000
C	-1.674031	-1.140934	0.009553	C	-2.168583	-1.146515	0.000000
N	0.635736	-1.950884	0.004246	N	-0.929778	-1.731611	0.000000
N	1.917247	2.001866	-0.027551	N	-2.419976	0.108007	0.000000
N	2.336166	-0.253385	-0.003640	N	-1.686518	2.285862	0.000000
H	-1.095612	2.345797	-0.030345	H	1.367103	2.218432	0.000000
H	-2.210631	-2.073255	0.015962	H	1.925384	-2.300045	0.000000
H	2.906911	2.136695	0.091110	H	-0.855495	2.872817	0.000000
H	1.308429	2.771722	0.176652	H	-2.993460	-1.852512	0.000000
H	2.645006	-2.293978	-0.008841	H	-0.849338	-2.733688	0.000000
S	-2.492208	0.358108	-0.001467	S	2.530673	0.080249	0.000000

	(i, 1H)			<sup>th</sup> G (k-a, 1H)			
C	0.255213	-0.938439	0.000000	C	0.691878	1.455236	0.003017
C	-1.373820	0.967888	0.000000	C	-0.480210	0.592413	0.009153
C	0.000000	0.475546	0.000000	C	-0.312239	-0.835382	-0.000223
C	1.144583	1.222816	0.000000	C	1.950911	-0.668838	-0.001558
C	1.597011	-1.223364	0.000000	C	-1.792508	0.985130	0.009627
C	-1.949345	-1.427502	0.000000	C	-1.520122	-1.488890	-0.005978
N	-0.755762	-1.891307	0.000000	N	0.925290	-1.448426	0.010266
N	-2.294209	-0.093689	0.000000	N	1.881184	0.713602	-0.002642
N	-1.676353	2.207874	0.000000	N	3.223039	-1.190837	-0.067256
H	-2.792143	-2.109704	0.000000	H	2.724000	1.262998	-0.083102
H	2.060343	-2.194310	0.000000	H	3.962412	-0.686110	0.394023
H	1.228120	2.295453	0.000000	H	3.249502	-2.187388	0.078621
H	-2.681022	2.370772	0.000000	H	-2.177575	1.990112	0.013641
H	-3.274035	0.141192	0.000000	H	-1.693288	-2.550630	-0.012382
S	2.531322	0.217174	0.000000	O	0.719850	2.669966	-0.006278
				S	-2.828544	-0.371694	-0.000461

(e-a)			(k-i, 1H,3H)				
C	0.778931	1.289455	-0.000225	C	-1.461559	0.686654	0.000000
C	-0.448380	0.564603	-0.002573	C	0.000000	0.739807	0.000000
C	-0.312349	-0.869813	0.002186	C	0.768482	-0.463492	0.000000
C	1.942760	-0.682708	-0.004217	C	-1.255505	-1.821929	0.000000
C	-1.755400	0.995833	-0.004328	C	0.778852	1.864771	0.000000
C	-1.544255	-1.491406	0.004555	C	2.116156	-0.230757	0.000000
N	0.907950	-1.481524	0.006251	N	0.116983	-1.680830	0.000000
N	1.932668	0.699199	-0.001321	N	-1.960895	-0.615291	0.000000
N	3.194764	-1.223241	-0.050245	N	-1.762259	-2.992013	0.000000
H	3.971973	-0.627250	0.167708	H	0.440175	2.886303	0.000000
H	3.278295	-2.207569	0.129925	H	2.922935	-0.942748	0.000000
H	-2.127690	2.005163	-0.007823	H	-2.775706	-2.983944	0.000000
H	-1.750596	-2.546926	0.006743	H	0.627396	-2.547955	0.000000
H	1.638424	2.952286	0.003151	H	-2.967743	-0.678169	0.000000
O	0.726097	2.629665	0.006027	O	-2.210944	1.642092	0.000000
S	-2.814669	-0.338115	-0.000193	S	2.437821	1.468525	0.000000

(e-i, 3H)			(k-a, 3H)				
C	-1.429247	0.510629	0.000000	C	-1.774133	0.982703	-0.007586
C	0.000000	0.694130	0.000000	C	-1.521892	-1.493038	0.001391
C	0.790883	-0.497065	0.000000	C	-0.462876	0.593573	-0.006669
C	-1.256594	-1.816635	0.000000	C	-0.330257	-0.823993	-0.002450
C	0.754000	1.838719	0.000000	C	0.738374	1.459671	0.000156
C	2.134527	-0.235301	0.000000	C	2.029414	-0.494689	0.000125
N	0.135125	-1.701684	0.000000	N	0.955478	-1.349380	-0.004078
N	-2.005968	-0.637824	0.000000	N	1.963560	0.799484	-0.014956
N	-1.760102	-2.989136	0.000000	N	3.259537	-1.098378	0.061860
H	0.405848	2.856723	0.000000	H	-2.148397	1.991846	-0.008483
H	2.957299	-0.928348	0.000000	H	-1.706311	-2.553432	0.002699
H	-3.102634	1.348682	0.000000	H	1.097487	-2.333853	0.145134
H	-2.772636	-2.926352	0.000000	H	4.025994	-0.459606	-0.074746
H	0.638263	-2.573632	0.000000	H	3.377729	-2.007415	-0.353136
O	-2.175097	1.625983	0.000000	O	0.660629	2.672959	0.009740
S	2.420615	1.471795	0.000000	S	-2.828334	-0.364291	0.000064

	${}^{th}C$	(k-a)			(e-i)		
C	0.326398	-0.870716	0.009343	C	-0.450345	0.601954	0.010794
C	0.397968	0.556042	-0.001177	C	-0.315498	-0.818581	0.004348
C	-0.867956	1.273206	0.000293	C	2.040708	-0.488337	0.002422
C	-2.112946	-0.715498	-0.003298	C	-1.768624	0.972021	0.011416
C	1.692746	1.013049	-0.012681	C	-1.502237	-1.494039	-0.002877
C	1.559505	-1.467916	0.010901	C	0.761014	1.456818	0.000667
N	-2.018029	0.665448	0.000730	N	0.968829	-1.356492	0.012410
N	-0.916941	-1.449965	0.012408	N	1.985033	0.781447	0.020875
N	-0.863850	2.627933	-0.019464	N	0.779213	2.733699	-0.018944
H	2.049950	2.028501	-0.032786	H	1.098388	-2.323332	-0.234031
H	-1.028508	-2.450851	0.011367	H	-1.675324	-2.556386	-0.002485
H	1.799482	-2.516614	0.017896	H	-2.177439	1.968119	0.013760
H	-1.751347	3.093429	0.067267	H	-0.165041	3.111508	-0.028173
H	-0.027543	3.154631	0.145920	H	3.215853	-1.996338	0.244846
O	-3.171144	-1.310469	-0.013238	O	3.258788	-1.074468	-0.028768
S	2.809035	-0.278641	-0.004985	S	-2.818023	-0.381088	-0.001549

	(k-i, 3H)				${}^{th}U$ (2k-6k, 1H)		
C	0.000000	0.730847	0.000000	C	-1.282504	-1.812469	0.000000
C	0.755862	-0.482027	0.000000	C	0.000000	0.730803	0.000000
C	-1.285150	-1.815320	0.000000	C	0.757497	-0.478143	0.000000
C	0.808029	1.836057	0.000000	C	-1.463379	0.696059	0.000000
C	2.106739	-0.275773	0.000000	C	0.792196	1.846841	0.000000
C	-1.467212	0.684874	0.000000	C	2.106606	-0.262563	0.000000
N	0.089623	-1.696850	0.000000	N	-1.973579	-0.603516	0.000000
N	-1.972736	-0.616361	0.000000	N	0.089550	-1.693148	0.000000
N	-2.305427	1.644958	0.000000	H	0.591507	-2.565732	0.000000
H	0.591518	-2.568993	0.000000	H	-2.979886	-0.691450	0.000000
H	2.897457	-1.005691	0.000000	H	0.466154	2.872437	0.000000
H	0.514174	2.871742	0.000000	H	2.905221	-0.983718	0.000000
H	-1.845752	2.549245	0.000000	O	-2.198613	1.660285	0.000000
H	-2.978523	-0.701668	0.000000	O	-1.851505	-2.882992	0.000000
O	-1.847038	-2.891538	0.000000	S	2.446478	1.431475	0.000000
S	2.462975	1.412094	0.000000				

<b>(2k-6e)</b>			<b>(2e-6k)</b>				
C	-1.344962	-1.785412	0.000000	C	-1.081248	-1.755118	0.000000
C	0.000000	0.684431	0.000000	C	0.000000	0.753831	0.000000
C	0.751964	-0.530948	0.000000	C	0.767723	-0.460874	0.000000
C	-1.443529	0.546343	0.000000	C	-1.451967	0.685979	0.000000
C	0.806077	1.797793	0.000000	C	0.774218	1.884389	0.000000
C	2.105298	-0.320977	0.000000	C	2.117239	-0.211652	0.000000
N	-2.049521	-0.583664	0.000000	N	-1.907863	-0.646348	0.000000
N	0.054826	-1.712064	0.000000	N	0.193290	-1.718634	0.000000
H	0.530216	-2.600248	0.000000	H	-2.913196	-0.739996	0.000000
H	0.527391	2.838502	0.000000	H	0.443387	2.908466	0.000000
H	2.899955	-1.046175	0.000000	H	2.918092	-0.929885	0.000000
H	-1.686945	2.444158	0.000000	H	-2.647169	-2.883180	0.000000
O	-2.221353	1.643979	0.000000	O	-2.245020	1.604612	0.000000
O	-1.886623	-2.868109	0.000000	O	-1.688501	-2.959273	0.000000
S	2.456686	1.367469	0.000000	S	2.432080	1.478590	0.000000

## B4.2 Water (PCM)

${}^{th}A$			$(a)$			$(i, 3H)$		
C	0.305681	-0.959843	0.000091	C	1.198997	1.158402	0.000000	
C	-0.048094	0.434943	-0.000039	C	1.503567	-1.310002	0.000000	
C	-1.456868	0.734484	-0.000003	C	0.000000	0.492103	0.000000	
C	-1.879476	-1.528905	-0.000165	C	0.192000	-0.925130	0.000000	
C	1.050589	1.266311	-0.000251	C	-1.367751	1.043722	0.000000	
C	1.672804	-1.146462	0.000330	C	-2.161094	-1.153399	0.000000	
N	-0.643896	-1.953217	0.000050	N	-0.935720	-1.728106	0.000000	
N	-1.903027	2.001061	0.000268	N	-2.414711	0.114407	0.000000	
N	-2.342037	-0.250833	-0.000155	N	-1.675074	2.292886	0.000000	
H	1.096645	2.341982	-0.000570	H	1.375532	2.220577	0.000000	
H	2.215940	-2.075467	0.000668	H	1.915265	-2.304155	0.000000	
H	-2.893827	2.173741	0.000096	H	-0.838304	2.870424	0.000000	
H	-1.280287	2.787676	0.000698	H	-2.986625	-1.855327	0.000000	
H	-2.655521	-2.287125	-0.000446	H	-0.856914	-2.732520	0.000000	
S	2.491997	0.354809	-0.000086	S	2.523455	0.075783	0.000000	

$(i, 1H)$			${}^{th}G$			$(k-a, 1H)$		
C	0.259192	-0.937144	0.000000	C	0.767762	-0.456873	-0.028677	
C	-1.377509	0.959203	0.000000	C	0.768438	1.890799	0.033765	
C	0.000000	0.476793	0.000000	C	2.118489	-0.200270	-0.009034	
C	1.145860	1.225612	0.000000	C	-0.002962	0.757152	-0.003744	
C	1.602773	-1.219221	0.000000	C	-1.100903	-1.755306	-0.078209	
C	-1.950155	-1.425625	0.000000	C	-1.453256	0.678889	-0.015261	
N	-0.751516	-1.891249	0.000000	H	0.442767	2.916335	0.058477	
N	-2.292200	-0.097840	0.000000	H	2.928315	-0.909002	-0.020746	
N	-1.694295	2.200183	0.000000	H	-1.188791	-3.757280	-0.254507	
H	-2.793616	-2.104575	0.000000	H	-2.704198	-2.998283	-0.393594	
H	2.074666	-2.186542	0.000000	H	-2.916059	-0.743243	-0.032712	
H	1.240340	2.297436	0.000000	N	0.194136	-1.710587	-0.074809	
H	-2.702805	2.337148	0.000000	N	-1.756449	-2.949367	-0.058485	
H	-3.275407	0.130702	0.000000	N	-1.911821	-0.631345	-0.051222	
S	2.533871	0.223154	0.000000	O	-2.240234	1.616929	0.006503	
				S	2.426921	1.492215	0.038956	

(e-a)			(k-i, 1H,3H)				
C	0.777898	1.291240	-0.000316	C	0.759723	-0.448727	-0.015220
C	-0.448608	0.562446	-0.001591	C	0.802245	1.866917	0.032038
C	-0.312878	-0.871269	0.000967	C	2.111638	-0.251613	0.009306
C	1.949595	-0.681159	-0.004068	C	-0.000422	0.756394	-0.002583
C	-1.756001	0.993221	-0.002397	C	-1.291971	-1.646782	-0.064696
C	-1.544065	-1.494091	0.002865	C	-1.472058	0.690457	-0.026509
N	0.911487	-1.482827	0.004347	H	0.498493	2.899118	0.048242
N	1.933141	0.700365	-0.003451	H	0.575757	-2.520437	-0.053801
N	3.201427	-1.216477	-0.051626	H	2.902617	-0.981123	0.007239
H	3.976189	-0.620004	0.181081	H	-1.401512	-3.708427	-0.181660
H	3.298299	-2.197200	0.148071	H	-2.904503	-2.861265	-0.164352
H	-2.132684	2.001154	-0.004537	N	0.069009	-1.649288	-0.049541
H	-1.757147	-2.548821	0.004709	N	-1.904512	-2.845868	-0.070836
H	1.620674	2.975153	0.004484	N	-2.037417	-0.558294	-0.057469
O	0.716461	2.627076	0.005077	O	-2.168386	1.709825	-0.020012
S	-2.815940	-0.339791	0.000496	S	2.456670	1.437755	0.048854

(e-i, 3H)			(k-a, 3H)				
C	0.751574	1.868986	0.023952	C	1.778830	0.985014	0.001773
C	0.786374	-0.488119	-0.017435	C	1.516115	-1.491708	-0.001914
C	2.137146	-0.207245	-0.006193	C	0.463751	0.598641	-0.000538
C	-0.001756	0.717091	-0.000144	C	0.326406	-0.819608	-0.002689
C	-1.100840	-1.750121	-0.045206	C	-0.740298	1.447720	-0.001076
C	-1.414982	0.522189	-0.013513	C	-2.029746	-0.510306	0.000868
H	0.417462	2.891690	0.040427	N	-0.955314	-1.345712	-0.004875
H	2.961450	-0.898853	-0.014164	N	-1.952325	0.806795	-0.000509
H	-1.208699	-3.773868	-0.230282	N	-3.248116	-1.082914	0.026238
H	-2.721863	-2.962864	-0.259997	H	2.171072	1.987066	0.003686
H	-3.123871	1.313067	-0.016483	H	1.694578	-2.552834	-0.002516
N	0.210583	-1.729560	-0.044648	H	-1.088213	-2.344700	0.000099
N	-1.745742	-2.950035	-0.021240	H	-4.049954	-0.483039	-0.052381
N	-1.949498	-0.660153	-0.033603	H	-3.381224	-2.073290	-0.077453
O	-2.199950	1.604655	-0.006155	O	-0.666329	2.680051	-0.002829
S	2.411406	1.487984	0.025482	S	2.824020	-0.367456	0.001667

${}^{th}C$			$(\mathbf{k-a})$			$(\mathbf{e-i})$		
C	0.767448	-0.515839	0.000000	C	0.000000	0.752442	0.000000	
C	0.000000	0.687730	0.000000	C	0.739473	-0.467722	0.000000	
C	-1.444892	0.545354	0.000000	C	-1.326189	-1.622422	0.000000	
C	-1.292349	-1.796986	0.000000	C	0.833803	1.841043	0.000000	
C	0.793244	1.810784	0.000000	C	2.094394	-0.301501	0.000000	
C	2.117993	-0.296636	0.000000	C	-1.474307	0.706261	0.000000	
N	-2.025894	-0.635486	0.000000	N	0.031739	-1.663641	0.000000	
N	0.090297	-1.714606	0.000000	N	-2.053361	-0.557587	0.000000	
N	-2.225068	1.635122	0.000000	N	-2.255348	1.729330	0.000000	
H	0.501832	2.847046	0.000000	H	0.529582	-2.539857	0.000000	
H	0.592922	-2.588134	0.000000	H	2.866680	-1.050831	0.000000	
H	2.917793	-1.016134	0.000000	H	0.571537	2.885296	0.000000	
H	-3.224645	1.521813	0.000000	H	-1.717550	2.591805	0.000000	
H	-1.850670	2.566000	0.000000	H	-1.381381	-3.555034	0.000000	
O	-1.826497	-2.907899	0.000000	O	-1.976833	-2.795949	0.000000	
S	2.446921	1.395684	0.000000	S	2.480096	1.376930	0.000000	

$(\mathbf{k-i}, 3H)$			${}^{th}U$ ( $2\mathbf{k-6k}, 1H$ )				
C	0.000000	0.732617	0.000000	C	-1.276957	-1.802333	0.000000
C	0.752725	-0.481789	0.000000	C	0.000000	0.734420	0.000000
C	-1.277879	-1.807214	0.000000	C	0.753251	-0.477843	0.000000
C	0.810573	1.838075	0.000000	C	-1.457067	0.695398	0.000000
C	2.103540	-0.282846	0.000000	C	0.799837	1.848304	0.000000
C	-1.462896	0.685946	0.000000	C	2.102785	-0.271164	0.000000
N	0.084751	-1.698902	0.000000	N	-1.970224	-0.596364	0.000000
N	-1.969958	-0.613351	0.000000	N	0.083701	-1.693177	0.000000
N	-2.300160	1.652331	0.000000	H	0.596068	-2.561243	0.000000
H	0.597409	-2.566167	0.000000	H	-2.977070	-0.683979	0.000000
H	2.890783	-1.016455	0.000000	H	0.491697	2.879283	0.000000
H	0.524620	2.876038	0.000000	H	2.896538	-0.997549	0.000000
H	-1.827727	2.550560	0.000000	O	-2.201112	1.663275	0.000000
H	-2.975574	-0.703023	0.000000	O	-1.863522	-2.873941	0.000000
O	-1.856981	-2.886458	0.000000	S	2.449025	1.419683	0.000000
S	2.461721	1.403840	0.000000				

<b>(2k-6e)</b>			<b>(2e-6k)</b>				
C	-1.327575	-1.776843	0.000000	C	-1.290299	-1.629268	0.000000
C	0.000000	0.689921	0.000000	C	0.000000	0.744541	0.000000
C	0.749669	-0.526282	0.000000	C	0.755998	-0.464421	0.000000
C	-1.434870	0.555289	0.000000	C	-1.469554	0.693411	0.000000
C	0.809286	1.804268	0.000000	C	0.809784	1.851400	0.000000
C	2.102056	-0.324764	0.000000	C	2.107967	-0.274760	0.000000
N	-2.040814	-0.587756	0.000000	N	-2.042220	-0.567528	0.000000
N	0.052169	-1.713224	0.000000	N	0.055803	-1.661560	0.000000
H	0.542101	-2.594702	0.000000	H	0.512667	2.885567	0.000000
H	0.534818	2.845952	0.000000	H	2.895592	-1.007691	0.000000
H	2.893242	-1.053694	0.000000	H	-2.803581	-2.734844	0.000000
H	-1.718228	2.451758	0.000000	H	0.529202	-2.552267	0.000000
O	-2.227657	1.632509	0.000000	O	-2.166825	1.704988	0.000000
O	-1.892429	-2.866240	0.000000	O	-1.841762	-2.843392	0.000000
S	2.452368	1.362366	0.000000	S	2.459773	1.412167	0.000000

## B5 Ground State Geometries of the Modified Nucleosides and their Tautomers

### B5.1 Gas Phase

${}^{th}\mathbf{A}$ (a)			$(\mathbf{i}, 1\mathbf{H})$				
C	0.7822360	0.4682040	0.1105360	C	0.9386360	-0.3016980	-0.5739650
C	1.6296790	0.0501300	1.3208590	C	1.4048460	1.1560310	-0.5450770
C	2.9642830	-0.0079680	-0.6463910	C	2.9305890	0.9736200	-0.3938640
C	3.0317460	0.4137680	0.8280640	C	3.0459200	-0.3116310	0.4645980
C	3.9118240	0.7674330	-1.5427270	C	4.1334680	-1.2807390	0.0255130
C	-0.5505100	-0.1708120	-0.0478400	C	-0.4935810	-0.6261040	-0.2633500
C	-1.7736370	0.4764290	0.0459770	C	-1.6212350	0.1597980	-0.1668480
C	-2.4853570	-1.6950660	-0.4574870	C	-2.5899900	-1.9455620	0.1183700
C	-2.8843240	-0.4039180	-0.1849950	C	-2.7898990	2.1152530	-0.1373030
C	-3.1850700	2.2046550	0.3692540	C	-2.8211310	-0.6026000	0.0441840
C	-4.1869470	0.1946750	-0.1027430	C	-4.1132550	0.0654170	0.1526260
H	0.5929320	1.5425280	0.1846170	H	0.0267710	2.0771820	0.4362480
H	1.3372300	0.6115950	2.2118900	H	1.1443800	1.6737790	-1.4716810
H	2.3729710	-1.6303800	1.9230120	H	1.1521210	-0.6767350	-1.5874260
H	3.1699240	1.4984900	0.9011290	H	2.9264680	2.5124390	0.7506360
H	3.1995510	-1.0756420	-0.7155290	H	3.2033300	-0.0323600	1.5082150
H	3.6614740	1.8321110	-1.5094700	H	3.3790510	0.7990250	-1.3735180
H	3.8273820	0.4204600	-2.5752310	H	3.9696980	-1.5510130	-1.0253940
H	4.8511560	-0.1923750	1.1155460	H	4.0529130	-2.1928890	0.6245730
H	5.8779800	0.9735970	-1.5743410	H	6.0925900	-1.2689670	-0.0222950
H	-3.0851200	-2.5658280	-0.6626170	H	-2.8609180	3.1960370	-0.1756720
H	-3.3573940	3.2526960	0.5876860	H	-3.3134720	-2.7240770	0.2873220
H	-5.2715420	-1.4624280	-0.6722810	H	-4.8204630	2.0175900	0.0969760
H	-6.1866490	-0.0446810	-0.3221620	H	-6.0175400	0.0678520	0.3834130
N	-1.9441320	1.8061280	0.3262500	N	-1.6411180	1.5520570	-0.2411670
N	-4.3175920	1.4752490	0.1703970	N	-3.9817490	1.4630150	0.0297820
N	-5.3134350	-0.5414680	-0.2784160	N	-5.2130140	-0.5528860	0.3316580
O	1.5297410	-1.3420980	1.5512990	O	0.9716500	1.8864770	0.5810400
O	1.6189260	0.2203520	-1.0507750	O	1.7850280	-0.9930500	0.3547930
O	4.0132960	-0.2641140	1.5878250	O	3.5608390	2.1052900	0.1456710
O	5.2269100	0.5316500	-1.0247170	O	5.3941920	-0.6509040	0.2048490
S	-0.7874330	-1.8325810	-0.4140410	S	-0.9275660	-2.2928790	-0.0692530

(i, 3H)			<sup>th</sup> C (k-a)				
C	0.8891080	0.2160840	0.5714640	C	0.8744020	1.7103890	0.2137110
C	1.5048690	1.0165600	-0.5866640	C	1.2448790	-2.0067570	-0.1918080
C	2.9949480	1.0336040	-0.2009910	C	1.3091610	0.4271170	-0.0042670
C	3.1741940	-0.3307910	0.4713590	C	2.7343480	0.3592860	-0.1337990
C	3.7953340	-1.4014040	-0.4134240	C	3.3239350	-0.9433220	-0.3589920
C	-0.4406160	-0.4137180	0.3055740	C	3.3419570	1.5850220	-0.0105750
C	-1.6573480	0.2034610	0.1675480	C	-0.8815320	-0.8111000	-0.1448420
C	-2.3324590	-1.9902070	-0.0571130	C	-1.6141050	-0.8361390	1.2094040
C	-2.7550170	-0.6882770	-0.0358760	C	-2.6388690	0.7291620	-0.2834130
C	-3.2133900	1.9874660	0.0441210	C	-2.9875460	-0.2867410	0.8283240
C	-4.1265880	-0.1472440	-0.1803200	C	-3.6600930	0.8064890	-1.4083990
H	0.8028170	0.8982580	1.4291560	H	4.3914380	1.8212920	-0.0457360
H	0.9485220	2.6345900	-1.5636810	H	5.0150030	-1.9650570	-0.7870820
H	1.3525400	0.4804740	-1.5247840	H	5.2277530	-0.2550940	-0.7490150
H	2.7610450	2.8364170	0.4905420	H	-0.1258530	2.0641060	0.3613640
H	3.2335320	-1.4870880	-1.3509620	H	-0.1732320	-0.2473780	2.3957850
H	3.6541570	1.1420600	-1.0662850	H	-1.1168540	-1.7291280	-0.6787800
H	3.7258420	-2.3592840	0.1085720	H	-1.6888200	-1.8532030	1.5937830
H	3.7809070	-0.2041510	1.3689130	H	-2.5133230	1.7187560	0.1667660
H	5.5999780	-1.7462030	-1.1047470	H	-3.3974020	1.6200180	-2.0878740
H	-1.1626920	2.2244830	0.1964000	H	-3.4486520	0.2081710	1.6835290
H	-2.9262680	-2.8787110	-0.1898130	H	-3.6464860	-0.1321050	-1.9663450
H	-3.3327120	3.0665040	0.0672190	H	-4.6371650	-1.0148610	0.1175440
H	-4.9852510	-1.8278290	-0.3736320	H	-5.5803730	1.2485740	-1.4888910
N	-1.9238500	1.5630850	0.2117430	N	0.5807300	-0.7539570	-0.0894190
N	-4.2482410	1.2513780	-0.1261860	N	2.6116680	-2.0290520	-0.3697210
N	-5.1928750	-0.8319940	-0.3503580	N	4.6637370	-1.0561510	-0.5363940
O	0.9265800	2.3202600	-0.6555860	O	0.5877260	-3.0290380	-0.1309400
O	1.8558880	-0.7883220	0.8400820	O	-1.0581160	0.0516990	2.1670110
O	3.2515880	2.0487300	0.7532930	O	-1.3878360	0.3165570	-0.8480600
O	5.1461160	-1.0304230	-0.6538090	O	-3.7787360	-1.3710140	0.3736150
S	-0.6400500	-2.1278080	0.1734140	O	-4.9404800	1.0291310	-0.8074060
				S	2.1931950	2.8140070	0.2577140

(k-i, 3H)			(e-i)				
C	0.7751700	-1.0448470	0.2348460	C	0.6557460	1.6694340	-0.0914110
C	1.5737500	-1.0177590	-1.1070690	C	1.2961280	0.4668770	0.0308430
C	2.7612820	-0.1134970	0.7947860	C	1.6034200	-1.8275650	0.4689940
C	2.8530640	-0.2607300	-0.7351830	C	2.7216600	0.5792870	-0.0390040
C	3.4298350	1.1294590	1.3456930	C	3.1185260	1.8752180	-0.2248620
C	-0.6386780	1.6179000	-0.2319190	C	3.5704280	-0.6204070	0.1141260
C	-1.2594630	0.4213910	-0.0030380	C	-0.7061910	-1.0330230	0.1178980
C	-1.4430970	-1.9809080	0.3497470	C	-1.3925900	-0.6468310	-1.2257120
C	-2.6849110	0.5539720	0.0643080	C	-2.6081280	0.1926550	-0.7691890
C	-3.0995420	1.8463200	-0.1160840	C	-2.7352200	-0.1727030	0.7175270
C	-3.5380550	-0.6124620	0.2874210	C	-3.4343460	0.8711460	1.5637390
H	0.2436250	-0.9113240	-2.5404690	H	1.7271240	-3.6254860	0.8923510
H	0.4084080	1.8178430	-0.3367580	H	4.1233900	2.2522370	-0.3087390
H	0.9038740	-2.0287390	0.6831120	H	5.2250000	0.2330880	-0.2483860
H	1.8282090	-2.0391080	-1.3948490	H	-0.3921570	1.8856830	-0.0208770
H	2.7972630	0.7283610	-1.2030900	H	-0.7239770	-0.0689450	-1.8654880
H	2.9405490	2.0194870	0.9375950	H	-0.8638360	-2.0986340	0.2480840
H	3.2098920	-1.0037300	1.2545950	H	-2.3700590	1.2575520	-0.8567860
H	3.3449200	1.1538490	2.4348190	H	-2.6537160	-1.6884480	-2.2632250
H	4.7725580	-0.4713780	-0.8625270	H	-2.8596480	1.8022840	1.5491150
H	5.2803170	1.8093660	1.2884000	H	-3.2714190	-1.1270410	0.7976410
H	-3.3399770	-2.6379430	0.5250260	H	-3.5189970	0.5299950	2.5978360
H	-4.1080130	2.2222440	-0.1154970	H	-4.5141790	0.2521940	-1.1251640
H	-5.2500830	0.2150520	0.2955430	H	-5.2425250	1.6653220	1.4977950
N	-0.6609510	-0.8406820	0.1594310	N	0.7185430	-0.7981120	0.2196200
N	-2.8065220	-1.7917830	0.3883500	N	2.8817060	-1.7857900	0.4384800
N	-4.8062880	-0.6924160	0.3863690	N	4.8412450	-0.6772210	-0.0076800
O	0.9284660	-0.3455370	-2.1732650	O	1.0053200	-2.9928320	0.7717070
O	1.3679090	-0.0586150	1.0757680	O	-1.3892550	-0.3112650	1.1494600
O	3.9998980	-0.9579180	-1.1721020	O	-1.7793530	-1.8365230	-1.8800600
O	4.7957330	1.0608070	0.9329400	O	-3.7360260	-0.1188990	-1.5578980
O	-0.9669160	-3.0957980	0.4631490	O	-4.7232820	1.0538370	0.9704850
S	-1.7799850	2.8967640	-0.3661790	S	1.7883650	2.9462870	-0.3108990

${}^{th}G$ (k-a, 1H)			(k-a, 3H)				
C	0.3066620	-0.4271180	-0.3020710	C	0.1413650	-0.4129020	0.4753540
C	1.5992190	0.0414630	-0.2089510	C	1.4552630	-0.0729670	0.2645150
C	1.9500890	-2.2391270	0.1835820	C	1.5760230	-2.3172340	-0.2721930
C	2.5376620	-1.0067140	0.0732380	C	2.2806860	-1.1517450	-0.1594930
C	3.2259200	1.6256500	-0.2675740	C	3.3931150	1.3056510	0.1697580
C	3.9532420	-0.6913460	0.1877760	C	3.7241690	-0.9368800	-0.4129750
C	-0.9240570	0.3659300	-0.5856100	C	-0.9771180	0.4173930	1.0475380
C	-1.5521300	1.0891210	0.6192400	C	-1.4872680	1.5678610	0.1417520
C	-2.9989340	1.3040750	0.1365060	C	-2.7105800	0.9628890	-0.5524040
C	-3.2200020	0.1607530	-0.9069160	C	-3.2706560	0.0458180	0.5334000
C	-4.2578910	-0.8808020	-0.5039980	C	-3.9941290	-1.2071420	0.0676000
H	2.4348390	-3.1800940	0.3798090	H	1.4591700	1.9917170	0.3874700
H	2.9632370	3.5947330	-0.5342800	H	1.9589510	-3.2757890	-0.5779660
H	4.5226770	3.1949380	-0.0012380	H	3.4437310	3.2209400	0.9132110
H	5.1724610	0.9378140	-0.0046130	H	4.8821750	2.6428550	0.2011130
H	-0.6743450	1.1313740	-1.3334780	H	-0.2204190	1.4629840	-1.3408160
H	-0.7545450	-0.0290420	2.0260560	H	-0.6459830	0.8274370	2.0078570
H	-1.0343240	2.0245710	0.8380170	H	-1.8326910	2.3905400	0.7675830
H	-3.0728460	2.2688880	-0.3633420	H	-2.3833770	0.3622010	-1.4122040
H	-3.5193600	0.5954740	-1.8656010	H	-3.2788160	2.4880970	-1.6225660
H	-3.5666480	0.8134140	1.9225740	H	-3.3606910	-1.7469660	-0.6449040
H	-3.9845200	-1.3221490	0.4592510	H	-3.9407590	0.6300590	1.1718330
H	-4.2602770	-1.6786910	-1.2453890	H	-4.1498150	-1.8511690	0.9384150
H	-5.5712440	0.3335100	0.2203100	H	-5.7220840	-1.6042760	-0.7609700
N	1.9647450	1.3656920	-0.3594350	N	2.0499880	1.1743330	0.4368630
N	3.6903450	2.8999510	-0.4841890	N	3.8798020	2.5812280	0.2708190
N	4.1969240	0.6784590	-0.0019860	N	4.1961230	0.3517190	-0.1862580
O	4.8703330	-1.4555540	0.4087520	O	4.4537120	-1.8345300	-0.7882190
O	-1.6332710	0.2776110	1.7819920	O	-0.5267440	2.1424670	-0.7276830
O	-1.9527170	-0.4901360	-1.0682350	O	-2.1261900	-0.3959950	1.2879170
O	-3.9380500	1.3321140	1.1969030	O	-3.6645450	1.9231930	-0.9460600
O	-5.5631610	-0.3310070	-0.4792000	O	-5.2255740	-0.8186200	-0.5203560
S	0.2624150	-2.1406160	-0.0515090	S	-0.0700740	-2.1057360	0.1431190

(e-a)			(e-i, 3H)				
C	0.2845180	0.7258090	-0.3141390	C	0.1343100	-0.3775800	0.5306280
C	1.4876000	0.0593800	-0.1757880	C	1.4296470	0.0083890	0.2791550
C	2.2183250	2.2851530	0.0455120	C	1.6009980	-2.2708500	-0.2017370
C	2.5892220	0.9614710	0.0262450	C	2.2644560	-1.0751800	-0.1347370
C	2.9097500	-1.7182730	-0.0557320	C	3.3334900	1.5073740	0.0739620
C	3.8655570	0.3465520	0.1762810	C	3.6466460	-0.7564390	-0.3873460
C	-1.0783560	0.2148740	-0.6573930	C	-1.0106590	0.3868450	1.1470470
C	-1.4852750	-1.2152080	-0.2019010	C	-1.5851590	1.5682870	0.3353660
C	-2.9932790	-1.0662340	0.0700040	C	-2.6771170	0.9070120	-0.4901020
C	-3.3024910	0.4131620	-0.2347060	C	-3.2508800	-0.0764750	0.5309350
C	-4.3318700	1.0424690	0.6876600	C	-3.9157820	-1.2947320	-0.0821050
H	0.0766920	-1.8098710	0.7486180	H	1.4054580	2.0635370	0.4561750
H	2.3896660	-3.6900320	-0.0946800	H	1.9981660	-3.2282420	-0.4905590
H	2.8396940	3.1499180	0.1996480	H	4.7234570	2.7491360	-0.1441370
H	4.0657550	-3.3763880	0.1948060	H	5.3266570	-1.4135450	-0.8828600
H	5.7046640	0.5603670	0.4501130	H	-0.7047890	0.7376010	2.1372980
H	-1.2043470	0.2411610	-1.7513620	H	-0.9879460	2.9863490	-0.8580320
H	-1.3306600	-1.9209150	-1.0214250	H	-2.0709410	2.2598120	1.0324290
H	-3.1494790	-1.2707710	1.1316220	H	-2.2188730	0.3641660	-1.3248090
H	-3.6506420	0.4778000	-1.2772120	H	-3.1827360	-1.8554760	-0.6687710
H	-4.0072210	0.9418050	1.7226410	H	-3.9770670	0.4524510	1.1575840
H	-4.4445170	2.1073170	0.4707740	H	-4.3089510	-1.9462440	0.7023550
H	-4.6595000	-1.9210840	-0.4371460	H	-4.3245590	1.4159150	-1.3772780
H	-6.0503460	0.6702590	-0.1923070	H	-5.4451000	-1.5268580	-1.3075810
N	1.6751400	-1.3001280	-0.1980920	N	2.0011440	1.2514650	0.3988620
N	3.1655120	-3.0531630	-0.1076150	N	3.7423750	2.7164280	0.1122470
N	4.0308620	-0.9399350	0.1291690	N	4.1435470	0.4223600	-0.2776100
O	4.9297560	1.1348240	0.3713820	O	4.4394790	-1.7779460	-0.7519850
O	-0.8585770	-1.6659760	0.9761120	O	-0.5855360	2.2429870	-0.3979240
O	-2.0632450	1.0918670	-0.0923160	O	-2.1335720	-0.4948350	1.3270160
O	-3.7425060	-1.9706370	-0.7261530	O	-3.5878570	1.8777060	-0.9608410
O	-5.5805680	0.3544320	0.5841170	O	-4.9661400	-0.7938000	-0.9141000
S	0.5388370	2.4324150	-0.1805170	S	-0.0338750	-2.0904190	0.2500150

(k-i, 1H,3H)			<sup>th</sup> U (2k-6k, 1H)				
C	1.0455230	-0.0420900	0.5056360	C	0.6361320	1.6258600	0.2139100
C	1.5060280	1.0316680	-0.5056560	C	1.2537990	0.4271700	-0.0107900
C	3.0410680	1.0367550	-0.3111960	C	1.4471810	-1.9724950	-0.3700790
C	3.3262730	-0.3292630	0.3549150	C	2.6780220	0.5597130	-0.0592870
C	4.3123950	-1.2041630	-0.3941880	C	3.1053860	1.8463940	0.1292460
C	-0.2760960	-0.6887480	0.2409970	C	3.5434940	-0.5935930	-0.2738780
C	-1.5110720	-0.0993240	0.1226240	C	-0.7738640	-1.0489950	-0.2576470
C	-2.1238260	-2.3398450	-0.0426190	C	-1.5634280	-1.0394730	1.0907360
C	-2.5671630	-1.0466250	-0.0364050	C	-2.7636880	-0.1083810	-0.7896070
C	-3.0791850	1.7563760	0.0273120	C	-2.8401060	-0.2681370	0.7394390
C	-3.9527710	-0.5969520	-0.1602200	C	-3.4327880	1.1417330	-1.3235560
H	1.0421800	0.4216440	1.5044170	H	3.3557200	-2.6225600	-0.5347330
H	1.2628890	0.6870280	-1.5109960	H	4.1243930	2.1916900	0.1359270
H	1.5170880	2.8370880	0.1933050	H	-0.2554270	-0.9856780	2.5481610
H	3.5647690	1.1391810	-1.2614580	H	-0.4106270	1.8341280	0.3055620
H	3.7088550	-0.1376710	1.3630410	H	-0.9000800	-2.0279820	-0.7172800
H	3.9615910	-1.3659130	-1.4186900	H	-1.8238050	-2.0636670	1.3620230
H	4.3039960	2.3052730	0.5022980	H	-2.7701950	0.7163550	1.2151630
H	4.3864480	-2.1752060	0.1017930	H	-2.9355610	2.0265130	-0.9136460
H	6.2275540	-1.0395500	-0.8148370	H	-3.2193650	-0.9930260	-1.2529970
H	-1.0550090	1.9340580	0.1331250	H	-3.3593050	1.1745550	-2.4132610
H	-2.7216350	-3.2274420	-0.1578310	H	-4.7595780	-0.4645330	0.8883430
H	-4.2392840	3.2712050	-0.0742340	H	-5.2805520	1.8266050	-1.2426870
H	-5.0300120	1.1242510	-0.1787100	N	0.6631230	-0.8371960	-0.1854670
N	-1.8036080	1.2513520	0.1528570	N	2.8205550	-1.7763250	-0.4007270
N	-3.2632840	3.0191530	0.0331600	O	0.9833850	-3.0903990	-0.4882500
N	-4.0828080	0.7870840	-0.0975620	O	4.7551730	-0.5862810	-0.3331750
O	0.9057380	2.2928500	-0.3225260	O	-0.9047520	-0.3911120	2.1628470
O	2.0779870	-1.0187900	0.4307180	O	-1.3721080	-0.0556170	-1.0832900
O	3.3585680	2.1367020	0.5359120	O	-3.9869340	-0.9599990	1.1835990
O	5.5575600	-0.5103530	-0.3758140	O	-4.7941010	1.0748680	-0.8965100
O	-4.9262560	-1.3120890	-0.2980570	S	1.7876900	2.8975780	0.3662790
S	-0.4309510	-2.4193940	0.1470090				

(2k-6e)			(2k-6k)				
C	0.7703830	-1.0617150	-0.3649390	C	0.3861010	1.5905070	0.0759610
C	1.5207660	-1.1173960	1.0024090	C	1.1141720	0.4452620	-0.0896410
C	2.7719910	-0.2620920	0.7528490	C	1.6238770	-1.8455240	-0.2973870
C	2.7941350	-0.1095000	-0.7752720	C	2.5228570	0.6802780	-0.1519950
C	3.4905060	1.1399880	-1.2753400	C	2.8352280	2.0069550	-0.0324270
C	-0.5914320	1.6342340	0.1230450	C	3.4787610	-0.4269190	-0.3161880
C	-1.2093530	0.4275850	-0.0800860	C	-0.7560830	-1.2544580	-0.2478940
C	-1.4778370	-1.9739290	-0.4137950	C	-1.5118060	-1.2945740	1.1313360
C	-2.6355950	0.5416860	-0.0293990	C	-2.7645070	-0.4188350	0.8857330
C	-3.0627410	1.8241890	0.2208080	C	-2.8368790	-0.4071870	-0.6438790
C	-3.3961820	-0.6646020	-0.2491010	C	-3.5933930	0.7419220	-1.2792860
H	0.4560710	1.8579650	0.1307660	H	1.9156090	-3.6718280	-0.4121540
H	0.8835080	-2.0265890	-0.8534420	H	3.8203320	2.4393590	-0.0502070
H	1.0393870	-0.9866840	2.8879710	H	-0.1653870	-1.4249910	2.5599390
H	1.8264000	-2.1474940	1.1912840	H	-0.6762570	1.7096630	0.1682430
H	2.6030630	0.7203490	1.2096380	H	-0.7938230	-2.2501200	-0.6838700
H	2.9622920	2.0254240	-0.9074370	H	-1.8145330	-2.3252650	1.3251000
H	3.2768430	-0.9947190	-1.2099480	H	-2.0470630	0.8459910	2.1717160
H	3.4889410	1.1630000	-2.3676800	H	-3.2842830	1.6843440	-0.8268440
H	4.6877770	-0.3647740	1.0240540	H	-3.2976080	-1.3467840	-0.9777060
H	5.3334800	1.8239120	-1.1058450	H	-3.3575730	0.7657300	-2.3469330
H	-4.0685980	2.1967790	0.3192100	H	-3.6606420	-0.8621320	1.3237560
H	-5.0558990	0.2782970	-0.1613760	H	-5.4754040	1.2935810	-1.2021320
N	-0.6555610	-0.8209870	-0.3176520	N	0.6548530	-0.8833970	-0.2019450
N	-2.8548290	-1.8106150	-0.4296580	N	2.9009970	-1.6959650	-0.3465870
O	0.7003990	-0.6466890	2.0558450	O	1.1372650	-3.1045420	-0.3142810
O	1.4209090	-0.0558010	-1.1430020	O	4.6790340	-0.2662670	-0.4052020
O	3.9193640	-0.8789790	1.2982530	O	-0.7980650	-0.7795140	2.2350830
O	4.8221830	1.0890240	-0.7594480	O	-1.4720430	-0.3619110	-1.0732810
O	-0.9693770	-3.0733870	-0.4879880	O	-2.5795160	0.8965200	1.3683120
O	-4.7408390	-0.6234460	-0.2771400	O	-4.9772280	0.4856010	-1.0612190
S	-1.7414860	2.8847880	0.3889560	S	1.4302480	2.9560010	0.1576980

## B5.2 Water

${}^{th}\mathbf{A}$ (a)			(i, 1H)				
C	0.7862600	0.4669080	0.1042560	C	0.8289720	0.3483950	0.0534150
C	1.6361120	0.0933810	1.3271460	C	1.5958280	-0.1013640	1.3062580
C	2.9730890	-0.0315280	-0.6406180	C	3.0430140	0.2455480	0.9149990
C	3.0349320	0.4423750	0.8176450	C	3.0622400	-0.0039760	-0.6132800
C	3.9185310	0.7183100	-1.5601350	C	3.8031420	1.0510970	-1.4210350
C	-0.5525070	-0.1665250	-0.0327520	C	-0.5323410	-0.2127760	-0.1549130
C	-1.7758310	0.4828890	0.0285610	C	-1.7334410	0.4371820	0.0345650
C	-2.4891110	-1.7057110	-0.4030210	C	-2.5101520	-1.6496320	-0.6628580
C	-2.8857850	-0.4066760	-0.1753750	C	-2.8676370	-0.3947560	-0.2521940
C	-3.1992830	2.2211040	0.2662340	C	-3.0571500	2.1937300	0.5701950
C	-4.1973680	0.1841310	-0.1190620	C	-4.2165600	0.1402350	-0.0978760
H	0.6139930	1.5455310	0.1409780	H	0.7235250	1.4371290	0.1046620
H	1.3453120	0.6866190	2.1966930	H	0.8023560	-1.7411860	2.0548700
H	2.3953480	-1.5707800	1.9665680	H	1.2630210	0.4300300	2.1969870
H	3.1692170	1.5278430	0.8532550	H	3.2267330	1.3003490	1.1183110
H	3.2127200	-1.0995410	-0.6748070	H	3.3813480	2.0379260	-1.2044430
H	3.6849270	1.7861410	-1.5421670	H	3.4958520	-0.9871530	-0.8124830
H	3.8247410	0.3556650	-2.5862760	H	3.6242150	-1.3577600	1.8133730
H	4.8598910	-0.1408780	1.1192760	H	3.6652110	0.8464660	-2.4858530
H	5.8720620	0.9930950	-1.5436840	H	5.6573990	1.6646640	-1.5725080
H	-3.0879090	-2.5805190	-0.5927670	H	-3.1541380	-2.4630920	-0.9486570
H	-3.3732540	3.2786690	0.4351060	H	-3.2321830	3.2145670	0.8859550
H	-5.2748410	-1.5434690	-0.4437150	H	-5.0816230	1.9366680	0.4639370
H	-6.2063100	-0.1055450	-0.2345510	H	-6.1230670	0.0061630	-0.1624550
N	-1.9537080	1.8263860	0.2500670	N	-1.8533580	1.7568340	0.4508080
N	-4.3282410	1.4843810	0.1008030	N	-4.1962010	1.4715740	0.3284920
N	-5.3071580	-0.5532230	-0.2857950	N	-5.2786740	-0.5377470	-0.3270430
O	1.5391270	-1.2926480	1.6153980	O	1.5601650	-1.5088020	1.5105180
O	1.6182250	0.1679030	-1.0499100	O	1.6907050	-0.0144800	-1.0425060
O	4.0237280	-0.2060350	1.5980910	O	4.0055960	-0.4854630	1.6442530
O	5.2332620	0.4727090	-1.0481580	O	5.1835530	0.9978140	-1.0672330
S	-0.7923910	-1.8431520	-0.3474430	S	-0.8096190	-1.8335450	-0.6920210

(i, 3H)			<sup>th</sup> C (k-a)				
C	0.4784630	-0.4884150	-0.3144710	C	0.7430150	1.6869910	0.1779800
C	1.6551310	0.1905330	-0.1214110	C	1.2622930	0.4345230	-0.0246690
C	2.4692830	-1.9730120	-0.0544610	C	1.3417740	-1.9760210	-0.2962240
C	2.8041280	-0.6462930	0.0254690	C	2.6925990	0.4520640	-0.0951430
C	3.0812690	2.0517020	0.1413960	C	3.2217030	1.7113590	0.0547510
C	4.1296180	-0.0326190	0.2290970	C	3.3735110	-0.8063340	-0.2996940
C	-0.8880700	0.0591630	-0.5861640	C	-0.8507640	-0.9245810	-0.2051400
C	-1.4310200	1.0862990	0.4300600	C	-1.5729570	-1.0280280	1.1490180
C	-2.9499850	1.0064860	0.1762920	C	-2.6429280	0.5906380	-0.2203610
C	-3.1540460	-0.4396980	-0.3134810	C	-2.9810100	-0.5335860	0.7611290
C	-3.8803080	-1.3501370	0.6715260	C	-3.6908290	0.9288360	-1.2636410
H	1.0312190	2.1785360	-0.0677150	H	4.2544650	2.0136530	0.0449600
H	3.1108420	-2.8335370	0.0318650	H	5.1647950	-1.7311720	-0.5429160
H	3.1361860	3.1328460	0.1927020	H	5.2778410	-0.0198620	-0.3483600
H	5.0843630	-1.6750310	0.3198960	H	-0.2783310	1.9903370	0.2857690
H	-0.9012980	0.5288790	-1.5802890	H	-1.0352670	-1.8430010	-0.7535820
H	-1.1996820	0.7495290	1.4395520	H	-1.3269460	-0.3586860	2.9676440
H	-1.4899930	2.8495630	-0.3500160	H	-1.5900770	-2.0613580	1.4976530
H	-3.4073480	-1.2942700	1.6516520	H	-2.4268310	1.4928190	0.3604920
H	-3.5250360	1.2154370	1.0756780	H	-3.2487640	1.6148780	-1.9907200
H	-3.6985870	-0.4129710	-1.2615840	H	-3.5295890	-0.1619090	1.6273920
H	-3.8110850	-2.3829020	0.3230970	H	-3.9649920	-2.2256890	0.7229880
H	-4.1250830	2.3429480	-0.6745570	H	-4.0083590	0.0240380	-1.7827570
H	-5.7387980	-1.1818930	0.0651370	H	-5.4806720	1.7260530	-1.2180270
N	1.8387380	1.5623150	-0.0593500	N	0.6051550	-0.7855820	-0.1658390
N	4.1711820	1.3641140	0.2722430	N	2.7084110	-1.9341560	-0.3815500
N	5.2387950	-0.6712250	0.3682440	N	4.7095240	-0.8447540	-0.4056020
O	-0.9089010	2.3902220	0.2759340	O	0.7424140	-3.0523110	-0.3340520
O	-1.8417750	-0.9949840	-0.5335990	O	-0.9546240	-0.1767430	2.0987610
O	-3.2539820	1.9691990	-0.8344470	O	-1.4466920	0.1656430	-0.8994850
O	-5.2349580	-0.9526760	0.8526970	O	-3.7097800	-1.5510510	0.0857440
S	0.7912860	-2.1899260	-0.3067320	O	-4.7832020	1.5441110	-0.5818040
				S	1.9926600	2.8650400	0.2819650

(k-i, 3H)			(e-i)				
C	0.7495190	-1.1335010	0.2583360	C	0.7607390	1.7155870	-0.2114100
C	1.5243160	-1.0635390	-1.1044120	C	1.3078230	0.4817440	0.0128700
C	2.7430300	-0.2065560	-0.7687920	C	1.4337990	-1.7967230	0.6015790
C	2.7722330	-0.2121210	0.7662340	C	2.7381220	0.4851650	-0.0396200
C	3.4583630	0.9887250	1.3854980	C	3.2363670	1.7309650	-0.3170240
C	-0.5343980	1.6000360	-0.1146550	C	3.4888310	-0.7519190	0.2213830
C	-1.2149180	0.4254400	0.0489190	C	-0.8081400	-0.8549530	0.2074940
C	-1.5131900	-1.9586690	0.3985350	C	-1.4488500	-0.5757100	-1.1647500
C	-2.6351480	0.6162500	0.0389720	C	-2.8158680	0.1560690	0.7189020
C	-2.9883820	1.9278430	-0.1403840	C	-2.9166720	-0.2614690	-0.7668000
C	-3.5415150	-0.5139230	0.2108380	C	-3.3057670	1.5583490	1.0345940
H	0.1921400	-1.1414000	-2.5356170	H	1.4130470	-3.5839520	1.1351330
H	0.5228700	1.7702790	-0.1395890	H	4.2664830	2.0289000	-0.4117480
H	0.8428450	-2.1427170	0.6501700	H	5.2112520	-0.0250920	-0.1202770
H	1.8545810	-2.0710060	-1.3624380	H	-0.2680470	2.0166340	-0.1908040
H	2.5503950	0.8115670	-1.1240050	H	-0.9938350	0.3031280	-1.6158220
H	2.9422240	1.9054130	1.0889270	H	-1.0457410	-1.8739720	0.5029320
H	3.2667260	-1.1310080	1.1053320	H	-2.1440000	-2.1571110	-2.0144430
H	3.4434610	0.9162690	2.4751640	H	-2.8326990	2.2811100	0.3641130
H	4.6574850	-0.1969060	-1.0363630	H	-3.0290100	1.8081690	2.0618600
H	5.2581390	1.7635430	1.1888970	H	-3.3271930	0.5365020	-1.3796120
H	-3.4372370	-2.5408300	0.5021770	H	-3.3870160	-0.5554140	1.3270620
H	-3.9782860	2.3482050	-0.1869990	H	-3.7852740	-1.9537570	-0.2118830
H	-5.2069580	0.3914940	0.0903450	H	-5.0490830	2.4501610	0.9838910
N	-0.6770650	-0.8640210	0.2201690	N	0.6379970	-0.7234220	0.2920970
N	-2.8669910	-1.7160820	0.3819460	N	2.7218300	-1.8487140	0.5851090
N	-4.8195580	-0.5368250	0.2238250	N	4.7647160	-0.8955980	0.1550910
O	0.8112930	-0.4933150	-2.1850290	O	0.7559320	-2.8975730	0.9511700
O	1.3967100	-0.2227690	1.1448650	O	-1.3254130	-1.6418210	-2.0728730
O	3.9176550	-0.7222400	-1.3679650	O	-1.4284480	0.0785860	1.0808150
O	4.7986680	0.9744410	0.8875910	O	-3.7543660	-1.3914720	-0.9938690
O	-1.0967850	-3.0990620	0.5537770	O	-4.7190280	1.5546000	0.8663220
S	-1.6161450	2.9259180	-0.2921480	S	1.9898450	2.8814930	-0.5061730

${}^{th}G$ (k-a, 1H)			(k-a, 3H)				
C	0.9466310	0.7290170	0.2578930	C	0.1914800	-0.3160790	0.3666530
C	1.6619320	1.0681970	-1.0971380	C	1.5265230	-0.0189770	0.2566820
C	3.1183740	0.5865730	-0.8847540	C	1.5922680	-2.2325700	-0.4174010
C	3.2389310	0.6415630	0.6353480	C	2.3336740	-1.1046320	-0.1888880
C	4.3142290	-0.2184850	1.2685410	C	3.5106080	1.2885410	0.3773780
C	-0.2106220	-0.2055900	0.2025450	C	3.7882840	-0.9249800	-0.3359440
C	-1.5477660	0.1283070	0.1874490	C	-0.9381240	0.5611210	0.8158960
C	-1.6992780	-2.2086120	0.0801070	C	-1.5464210	1.4954040	-0.2752100
C	-2.3961350	-1.0290470	0.1146550	C	-2.8387520	0.7774060	-0.6783800
C	-3.3279360	1.5478590	0.2205990	C	-3.2661770	0.1348910	0.6382060
C	-3.8376490	-0.8552690	0.0948500	C	-4.1199910	-1.1196270	0.5600830
H	0.5826240	1.6676800	0.6832700	H	1.6229470	1.9954280	0.7881520
H	1.1690900	0.9681750	-2.9905550	H	1.9442860	-3.1942500	-0.7485160
H	1.6475580	2.1468930	-1.2410470	H	3.5448670	3.2294180	1.0828650
H	2.6075780	-0.8697170	-2.0462900	H	5.0615130	2.5590470	0.6144670
H	3.4017210	1.6863030	0.9311300	H	-0.3708220	1.0443510	-1.7752690
H	3.8377550	1.2478080	-1.3717340	H	-0.5921280	1.1705240	1.6561830
H	4.1710060	-0.2182480	2.3522640	H	-1.8224340	2.4409440	0.1905450
H	4.2305250	-1.2428520	0.9038300	H	-2.6135230	-0.0028260	-1.4144530
H	6.2737240	-0.2098580	1.2759510	H	-3.5261700	2.0280640	-2.0055620
H	-2.0909530	-3.2105770	0.0420370	H	-3.6964580	-1.8092080	-0.1756520
H	-3.2815410	3.5593930	0.2037980	H	-3.7823900	0.8835110	1.2497050
H	-4.8435480	2.9166810	0.0017350	H	-4.1002480	-1.6074890	1.5383020
H	-5.2057850	0.6570560	0.1604800	H	-5.9921910	-1.5379750	0.1805380
N	-2.0387560	1.4166100	0.2298520	N	2.1605390	1.1826070	0.5328340
N	-3.9027770	2.7775730	0.3307970	N	4.0593030	2.4909390	0.6361280
N	-4.2106280	0.4814390	0.1506160	N	4.2968010	0.3095290	-0.0233470
O	1.0457560	0.4263240	-2.2045870	O	4.5197030	-1.8420900	-0.7187610
O	1.9681720	0.1919190	1.1124730	O	-0.6908420	1.8445730	-1.3415730
O	3.2744840	-0.7417080	-1.3567210	O	-2.0306010	-0.2538040	1.2628450
O	5.5771860	0.3504300	0.9219950	O	-3.8360170	1.6510780	-1.1750940
O	-4.6840730	-1.7385270	0.0377060	O	-5.4483390	-0.7453680	0.2073200
S	-0.0180380	-1.9259700	0.1319260	S	-0.0636280	-1.9740010	-0.0879430

(e-a)			(e-i, 3H)				
C	0.2922860	0.4098760	0.0910000	C	0.2140380	-0.2319910	0.2859710
C	1.5689500	-0.1203870	-0.0177560	C	1.5479160	0.0842820	0.2543880
C	2.0682160	2.1775340	0.0172170	C	1.6617650	-2.1353760	-0.4567890
C	2.5794010	0.9019770	-0.0505580	C	2.3751940	-1.0008260	-0.1675130
C	3.1647200	-1.7279560	-0.1926490	C	3.5388410	1.4526170	0.5105800
C	3.9226900	0.4346430	-0.1515720	C	3.7908990	-0.7501850	-0.2120000
C	-0.9746640	-0.3645220	0.1061550	C	-0.9444240	0.6369140	0.6552720
C	-1.9163990	-0.2357670	1.3075420	C	-1.6361380	1.4043610	-0.5091320
C	-3.1480040	-0.4560540	-0.7619130	C	-2.9611250	0.6558000	-0.7052850
C	-3.2222200	-0.8195610	0.7323150	C	-3.2636490	0.1540180	0.7043940
C	-4.1003450	0.6534310	-1.1978970	C	-4.1057560	-1.1054300	0.8281210
H	2.5990610	3.1138930	0.0102690	H	1.6277020	2.0812930	0.8434640
H	2.8724570	-3.7412240	-0.1134630	H	2.0370330	-3.0918010	-0.7767880
H	4.5184670	-3.2493770	-0.1689640	H	5.0322920	2.5704310	0.7378330
H	5.7509200	0.8737090	-0.2452350	H	5.4882540	-1.4694080	-0.5925740
H	-0.6950710	-1.4193540	0.0301430	H	-0.6020380	1.3627890	1.3990050
H	-1.5582070	-0.8144850	2.1609860	H	-1.0312230	2.2549930	-2.1614100
H	-2.6030980	1.2091100	2.4390900	H	-1.8723280	2.4147330	-0.1696240
H	-3.3497320	-1.3615490	-1.3435200	H	-2.7986010	-0.1928790	-1.3758920
H	-3.3924030	-2.4802170	1.7640790	H	-3.7260170	-1.8685900	0.1425330
H	-3.8521420	0.9558360	-2.2180110	H	-3.7336120	0.9601830	1.2801440
H	-3.9798200	1.5192280	-0.5487710	H	-3.9331370	1.5535800	-2.1475260
H	-4.0953330	-0.3756600	1.2130940	H	-4.0070570	-1.4832030	1.8493090
H	-5.6556950	-0.3936940	-1.7794020	H	-5.9965690	-1.5739540	0.6643660
N	1.8868580	-1.4485120	-0.0898440	N	2.1642600	1.2757080	0.5654110
N	3.5499560	-3.0275470	-0.3189220	N	4.0223170	2.5991230	0.8342490
N	4.2122460	-0.8288980	-0.2206980	N	4.3279140	0.3800570	0.1027850
O	4.9053590	1.3412120	-0.1730940	O	4.5656440	-1.7655410	-0.6053790
O	-1.8087360	-0.0094350	-1.0190180	O	-0.8148020	1.4623820	-1.6617570
O	-2.0691130	1.1369000	1.6416740	O	-1.9765350	-0.1689560	1.2480000
O	-3.2622430	-2.2410140	0.8403650	O	-4.0071460	1.4819930	-1.1910690
O	-5.4656450	0.2579840	-1.0969440	O	-5.4627660	-0.7843880	0.5368580
S	0.3729450	2.1390210	0.1429120	S	-0.0064200	-1.8861200	-0.2114740

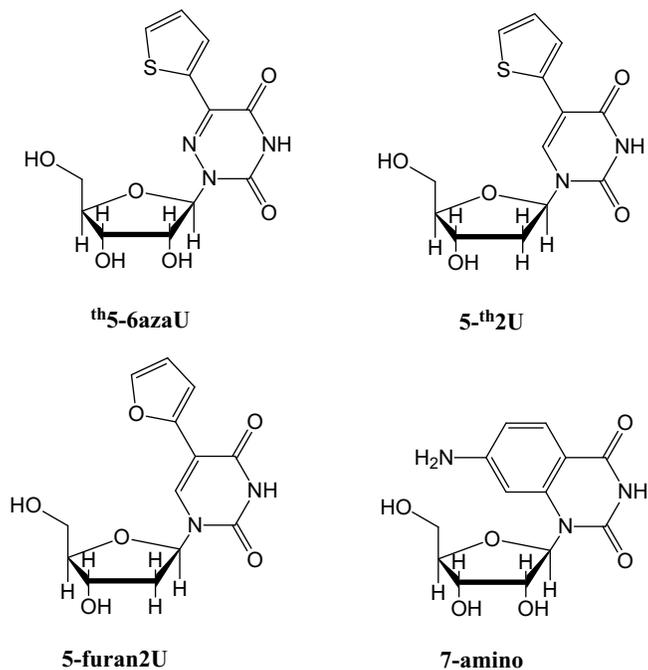
(k-i, 1H,3H)			<sup>th</sup> U (2k-6k, 1H)				
C	1.0100790	0.3094880	0.2302000	C	0.7215190	1.6705120	0.0941480
C	1.8052260	0.6123390	-1.0612820	C	1.2717210	0.4256570	-0.0396680
C	3.2102900	0.8977340	-0.5046200	C	1.3369460	-1.9877300	-0.2663320
C	3.2798420	0.0039420	0.7466190	C	2.7038390	0.4728670	-0.0469830
C	4.1607380	-1.2258090	0.5992220	C	3.2015960	1.7435900	0.0812200
C	-0.2572600	-0.4522160	0.0663680	C	3.4985430	-0.7377030	-0.1640920
C	-1.5513180	-0.0028830	0.1374140	C	-0.8375240	-0.9670810	-0.2459220
C	-1.9480430	-2.2396110	-0.3730740	C	-1.6263250	-0.9747920	1.0674420
C	-2.5189340	-1.0232120	-0.1140200	C	-2.8135910	0.1269260	-0.7416980
C	-3.3118250	1.6404130	0.4836770	C	-3.0502600	-0.7041560	0.5361930
C	-3.9441150	-0.7176550	-0.0642990	C	-3.2193780	1.5879290	-0.6436590
H	0.7923360	1.2644460	0.7251670	H	3.2064780	-2.7605790	-0.3269000
H	1.6568950	2.4773670	-1.5071590	H	4.2357420	2.0393670	0.1046430
H	1.8284640	-0.2868000	-1.6765320	H	-0.3119670	1.9453200	0.1432240
H	3.6346030	0.6009330	1.5904200	H	-1.0019060	-1.9279460	-0.7297230
H	3.9102910	-1.7620900	-0.3206630	H	-1.5637090	-1.9478610	1.5568860
H	3.9786200	-1.8938250	1.4448130	H	-1.6041150	-0.0119040	2.7669790
H	3.9951150	0.6576250	-1.2197970	H	-2.7851070	2.0390220	0.2524130
H	4.1534060	2.6015560	-0.1944570	H	-2.8400000	2.1207450	-1.5193690
H	6.0881010	-1.5325030	0.4093710	H	-3.3598230	-0.3448420	-1.5620300
H	-1.3325880	2.0255080	0.5795520	H	-3.6378310	-0.1518240	1.2704080
H	-2.4507860	-3.1691320	-0.5772910	H	-3.9493400	-2.3968130	0.9524890
H	-4.6354870	2.9848760	0.7713050	H	-4.9228880	2.5369730	-0.4508250
H	-5.1973770	0.8458980	0.2794340	N	0.6146740	-0.8122430	-0.1688730
N	-1.9879340	1.2791530	0.4154910	N	2.7198870	-1.8773730	-0.2575020
N	-3.6305970	2.8501960	0.7545780	O	0.8179710	-3.0906920	-0.3525360
N	-4.2170850	0.6060410	0.2375870	O	4.7180630	-0.8044530	-0.1775830
O	1.2770050	1.6535710	-1.8441680	O	-1.1630520	0.0619890	1.9144040
O	1.9368710	-0.4491680	1.0127210	O	-1.4023700	0.0766350	-1.0212080
O	3.2440360	2.2895380	-0.1845770	O	-3.7141340	-1.9057040	0.1580400
O	5.5145730	-0.7795410	0.5789290	O	-4.6427590	1.6285850	-0.5956980
O	-4.8508730	-1.5142910	-0.2603590	S	1.9425390	2.8788980	0.2104130
S	-0.2476380	-2.1544190	-0.3122590				

(2k-6e)			(2k-6k)				
C	0.8282380	-0.9394670	-0.2805120	C	0.7220540	-1.0089300	-0.1478360
C	1.5874420	-1.0443600	1.0537360	C	1.4148920	-0.6813240	1.2124310
C	2.8098680	0.1019460	-0.7374970	C	2.6624750	0.1227270	0.7654400
C	2.9971750	-0.5145580	0.6730920	C	2.7287930	-0.1265390	-0.7503250
C	3.3659900	1.5020380	-0.9265530	C	3.4060290	0.9757500	-1.5385010
C	-0.7455150	1.6647460	0.1934340	C	-0.6629280	1.6694980	0.1964080
C	-1.2703170	0.4189610	-0.0279510	C	-1.2940190	0.4710430	0.0028280
C	-1.3566900	-1.9851960	-0.3543530	C	-1.5831230	-1.8195570	-0.4584660
C	-2.7028560	0.4368490	-0.0536220	C	-2.7178320	0.5808830	0.0283180
C	-3.2292810	1.6920750	0.1522110	C	-3.1379960	1.8656650	0.2544480
C	-3.3701230	-0.8135760	-0.2799090	C	-3.5572080	-0.5981700	-0.1989290
H	0.2817470	1.9581780	0.2702960	H	0.3840370	1.9023830	0.1902120
H	1.0100140	-1.8553380	-0.8416080	H	0.7740230	-0.0736570	1.8479990
H	1.3605060	-0.4127660	2.8875880	H	0.8955340	-2.0634170	-0.3394700
H	1.6504360	-2.0835330	1.3789060	H	1.0481150	-2.1511290	2.4429360
H	2.9640560	2.1635840	-0.1539020	H	2.4555490	1.1852620	0.9289570
H	3.0534370	1.8820950	-1.9026440	H	2.8644470	1.9150280	-1.4003140
H	3.2692960	-0.5702860	-1.4733090	H	3.2425950	-1.0786680	-0.9293200
H	3.2821490	0.2519190	1.3911390	H	3.4197310	0.7328100	-2.6031810
H	3.8941790	-2.1513120	0.0479980	H	4.5570790	0.2336930	1.0525760
H	5.1515920	2.3037820	-0.9198290	H	5.1792950	1.8237970	-1.4141790
H	-4.2616410	1.9962940	0.1876890	H	-1.6975780	-3.6357300	-0.8646390
H	-5.1146950	-0.0185550	-0.2744010	H	-4.1517230	2.2204470	0.3192790
N	-0.6209510	-0.7993630	-0.2209760	N	-0.7182810	-0.7922190	-0.2123490
N	-2.7358430	-1.9295710	-0.4134600	N	-2.8814000	-1.7725290	-0.4618370
O	0.9437990	-0.2487850	2.0354460	O	1.3594010	-0.2219560	-1.1445360
O	1.3994260	0.1657750	-0.9635450	O	1.7807440	-1.8698050	1.8867610
O	4.0252280	-1.4891540	0.7353560	O	3.8291860	-0.2530870	1.4611250
O	4.7847610	1.4184860	-0.8396210	O	4.7337250	1.0706610	-1.0155210
O	-0.7750690	-3.0632230	-0.4229740	O	-0.9935180	-2.9883820	-0.7135710
O	-4.7030530	-0.8864390	-0.3637690	O	-4.7841370	-0.5622410	-0.1707940
S	-1.9926590	2.8347110	0.3723230	S	-1.8155500	2.9261660	0.4273610

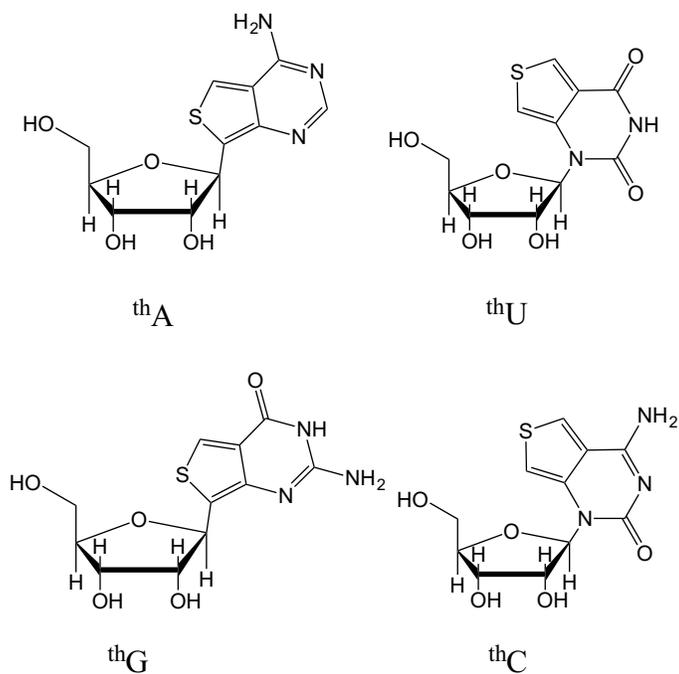
# Appendix C

## Appendix to Chapter 4

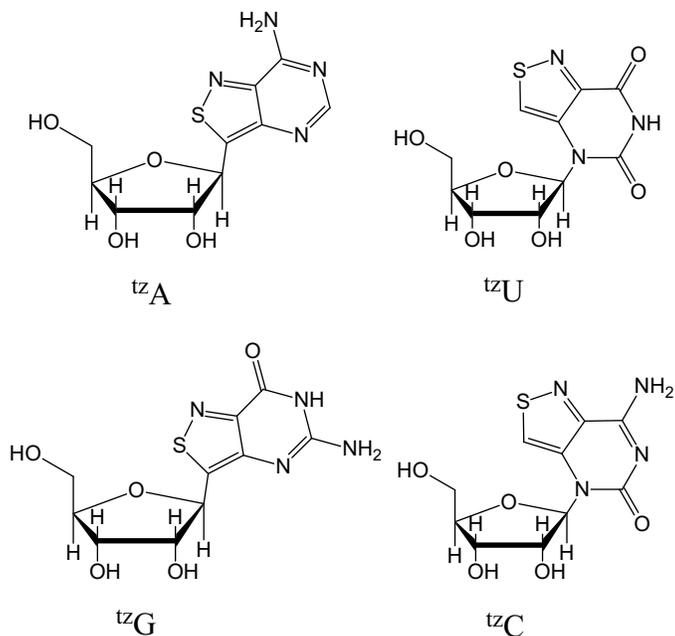
## C1 Nucleoside Structures



**Figure C1:** Structures of the modified nucleosides and ribonucleosides. *th* **5-6azaU**: 5-(thiophen-2-yl)-6-aza-uridine, **5-th2U**: 5-(thiophen-2-yl)-2'-deoxyuridine, **5-furan2U**: 5-(furan-2-yl)-2'-deoxyuridine, **7-amino**: 7-amino-1-ribose-quinazoline-2,4(1H,3H)-dione.



**Figure C2:** Structures of the modified thiophene nucleosides. <sup>th</sup>**A**: thieno[3,4-d]adenosine, <sup>th</sup>**U**:thieno[3,4-d]uridine, <sup>th</sup>**G**: thieno[3,4-d]guanosine, <sup>th</sup>**C**: thieno[3,4-d]cytidine.



**Figure C3:** Structures of the modified isothiazole nucleosides. <sup>tz</sup>**A**: isothiazolo[4,3-d]adenosine, <sup>tz</sup>**U**:isothiazolo[4,3-d]uridine, <sup>tz</sup>**G**: isothiazolo[4,3-d]guanosine, <sup>tz</sup>**C**: isothiazolo[4,3-d]cytidine.

## C2 Photophysical Properties

### C2.1 Charge Transfer Character of Low-Lying States

**Table C1:** *Excitation energies, oscillator strengths ( $f$ ),  $\Lambda$  diagnostic and TPA cross sections ( $\sigma^{TPA}$  in GM) of low-lying singlet states of the thiophene nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water (PCM). Also provided is the dominant character of the transition.*

Nucleoside	$S_n$	Energy (eV)	$f$ (a.u.)	$\Lambda$ Diagnostic	Transition	Type	$\sigma^{TPA}$ (GM)
<i>th</i> <b>A</b>	<b>S</b> <sub>1</sub>	3.92	0.254	0.79	H $\rightarrow$ L	L	0.3
	<b>S</b> <sub>2</sub>	4.74	0.004	0.46	H-2 $\rightarrow$ L	L	0.1
	<b>S</b> <sub>3</sub>	4.95	0.088	0.67	H-1 $\rightarrow$ L	L	0.5
	<b>S</b> <sub>4</sub>	5.40	0.112	0.58	H-3 $\rightarrow$ L	L	13.9
	<b>S</b> <sub>5</sub>	5.52	0.010	0.23	H $\rightarrow$ L+2	R	0.1
<i>th</i> <b>C</b>	<b>S</b> <sub>1</sub>	4.34	0.154	0.66	H $\rightarrow$ L	L	1.5
	<b>S</b> <sub>2</sub>	5.05	0.342	0.74	H-1 $\rightarrow$ L	L	0.7
	<b>S</b> <sub>3</sub>	5.29	0.006	0.34	H-5 $\rightarrow$ L	CT	0.1
	<b>S</b> <sub>4</sub>	5.48	0.154	0.61	H-2 $\rightarrow$ L	L	2.2
	<b>S</b> <sub>5</sub>	5.57	0.000	0.34	H-4 $\rightarrow$ L	CT	0.0
<i>th</i> <b>G</b>	<b>S</b> <sub>1</sub>	4.12	0.195	0.69	H $\rightarrow$ L	L	1.8
	<b>S</b> <sub>2</sub>	5.04	0.033	0.71	H-1 $\rightarrow$ L	L	0.3
	<b>S</b> <sub>3</sub>	5.19	0.000	0.37	H-5 $\rightarrow$ L	CT	0.0
	<b>S</b> <sub>4</sub>	5.45	0.009	0.26	H $\rightarrow$ L+1	CT	0.9
	<b>S</b> <sub>5</sub>	5.60	0.802	0.58	H-1 $\rightarrow$ L	L	3.3
<i>th</i> <b>U</b>	<b>S</b> <sub>1</sub>	4.61	0.124	0.63	H $\rightarrow$ L	L	1.7
	<b>S</b> <sub>2</sub>	5.11	0.004	0.31	H-3 $\rightarrow$ L	CT	0.0
	<b>S</b> <sub>3</sub>	5.28	0.201	0.71	H-1 $\rightarrow$ L	L	0.7
	<b>S</b> <sub>4</sub>	6.06	0.030	0.29	H $\rightarrow$ L+5	R	0.2
	<b>S</b> <sub>5</sub>	6.14	0.348	0.26	H $\rightarrow$ L+6	CT	0.9

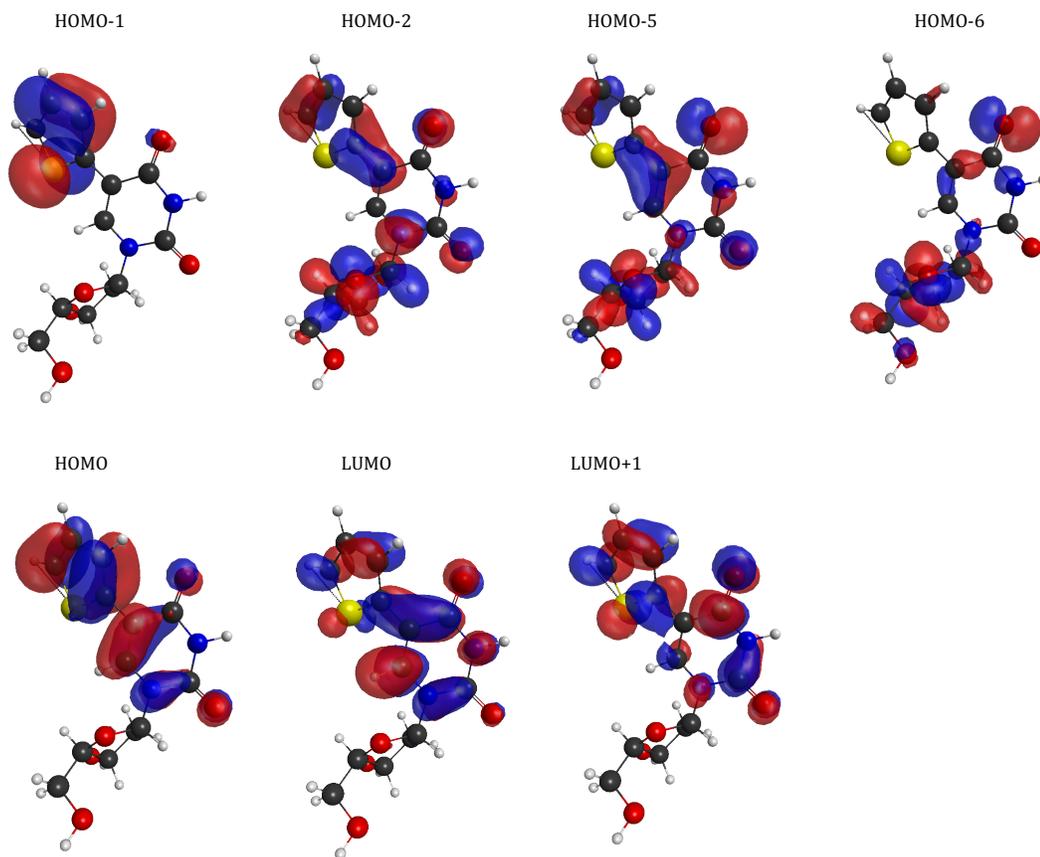
Types of transitions: R=Rydberg, CT= Charge Transfer, L=Local Orbitals: H= HOMO, L=LUMO

**Table C2:** *Excitation energies, oscillator strengths ( $f$ ),  $\Lambda$  diagnostic and TPA cross sections ( $\sigma^{TPA}$  in GM) of low-lying singlet states of the isothiazole nucleosides computed with CAM-B3LYP/6-31++G(d,p) in water (PCM). Also provided is the dominant character of the transition.*

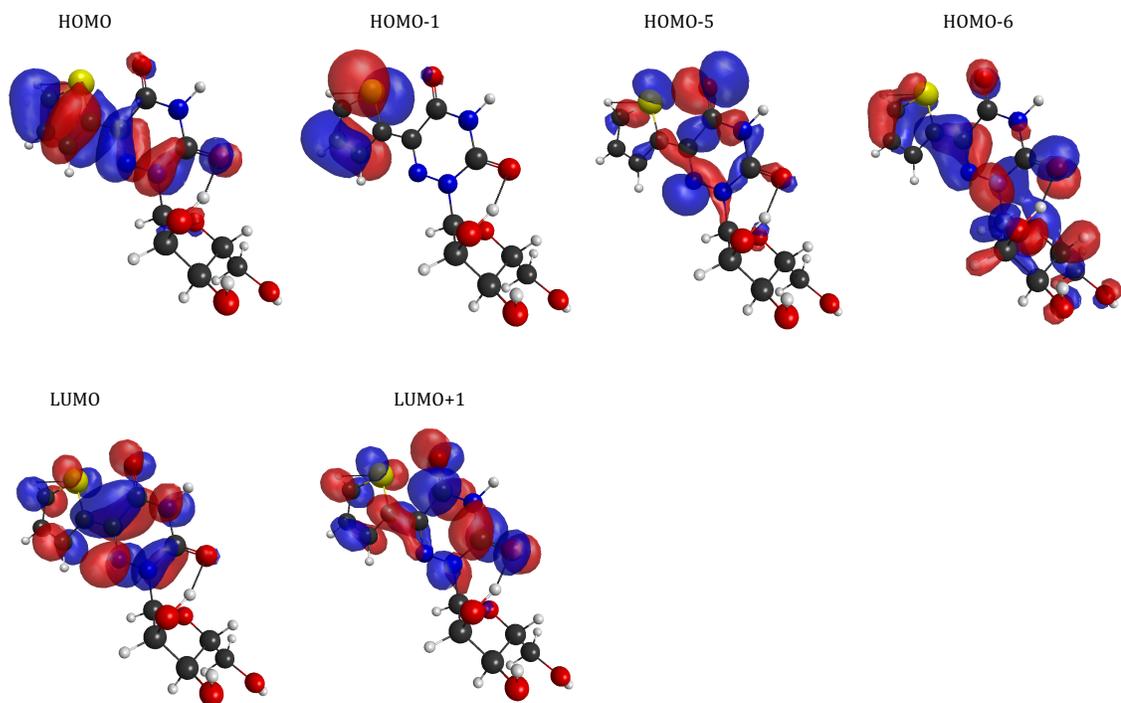
Nucleoside	$S_n$	Energy (eV)	$f$ (a.u.)	$\Lambda$ Diagnostic	Transition	Type	$\sigma^{TPA}$ (GM)
${}^{tz}\mathbf{A}$	$S_1$	4.04	0.317	0.81	H $\rightarrow$ L	L	0.4
	$S_2$	4.53	0.002	0.46	H-2 $\rightarrow$ L	L	0.4
	$S_3$	5.08	0.096	0.67	H-3 $\rightarrow$ L	L	0.1
	$S_4$	5.34	0.004	0.58	H-1 $\rightarrow$ L	L	17.1
	$S_5$	5.56	0.003	0.45	H-8 $\rightarrow$ L	L	0.3
${}^{tz}\mathbf{C}$	$S_1$	4.22	0.199	0.69	H $\rightarrow$ L	L	0.9
	$S_2$	5.03	0.008	0.38	H-6 $\rightarrow$ L	CT	0.1
	$S_3$	5.11	0.305	0.57	H-2 $\rightarrow$ L	L	1.6
	$S_4$	5.43	0.001	0.34	H-3 $\rightarrow$ L	CT	0.1
	$S_5$	5.55	0.064	0.68	H-4 $\rightarrow$ L	L	2.6
${}^{tz}\mathbf{G}$	$S_1$	4.10	0.234	0.72	H $\rightarrow$ L	L	1.7
	$S_2$	4.94	0.001	0.43	H-5 $\rightarrow$ L	CT	0.1
	$S_3$	5.21	0.015	0.51	H-1 $\rightarrow$ L	L	0.4
	$S_4$	5.33	0.006	0.44	H-4 $\rightarrow$ L+1	CT	0.1
	$S_5$	5.72	0.621	0.50	H $\rightarrow$ L+3	CT	5.8
${}^{tz}\mathbf{U}$	$S_1$	4.50	0.191	0.67	H $\rightarrow$ L	L	1.2
	$S_2$	4.89	0.000	0.35	H-5 $\rightarrow$ L	CT	0.0
	$S_3$	5.54	0.032	0.47	H-8 $\rightarrow$ L	CT	0.2
	$S_4$	5.56	0.091	0.58	H-3 $\rightarrow$ L	L	0.5
	$S_5$	6.01	0.011	0.29	H-1 $\rightarrow$ L	CT	0.1

Types of transitions: R=Rydberg, CT= Charge Transfer, L=Local  
Orbitals: H= HOMO, L=LUMO

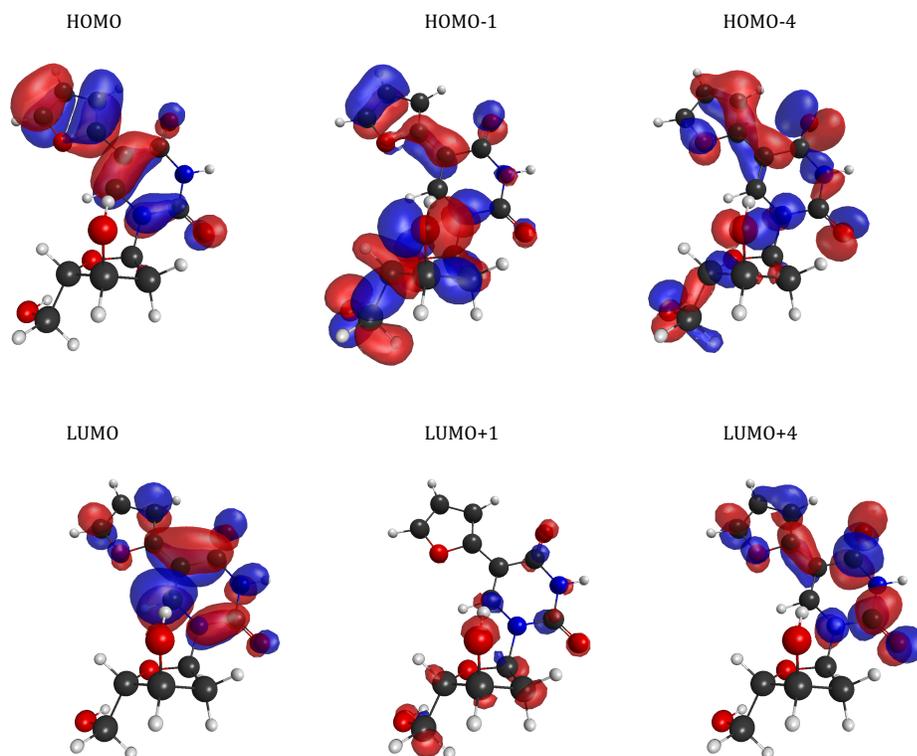
## Valence Molecular Orbitals



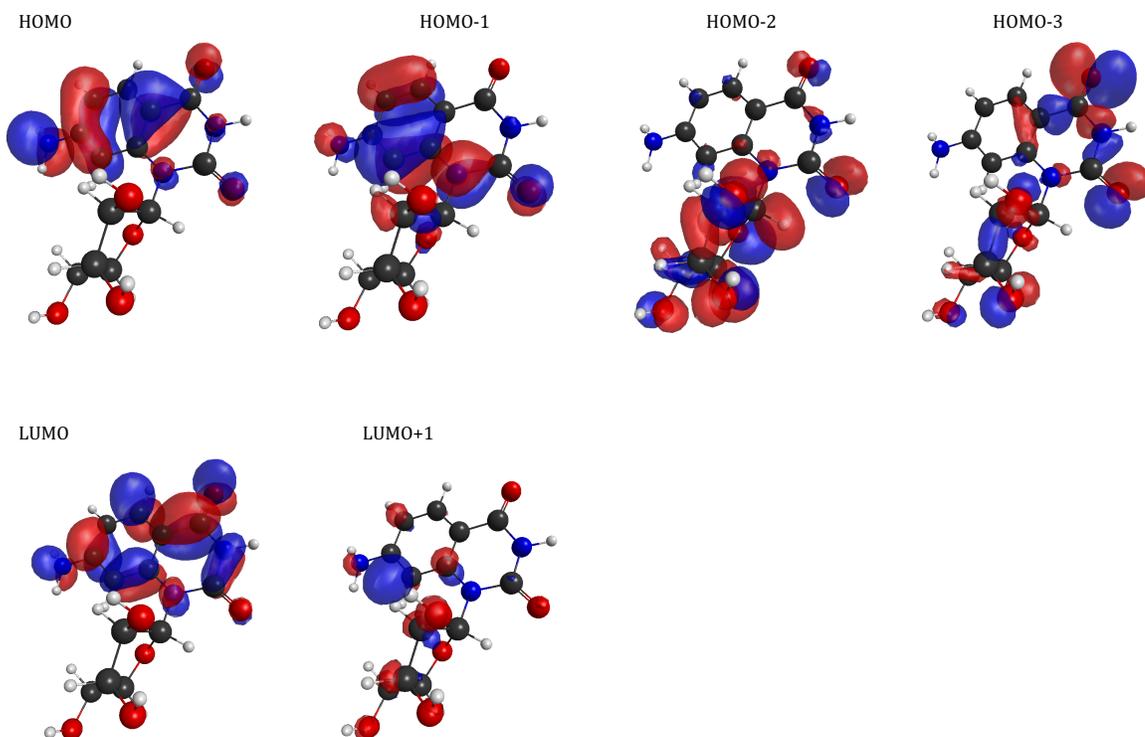
**Figure C4:** Valence molecular orbitals of  $^{th}5-2U$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



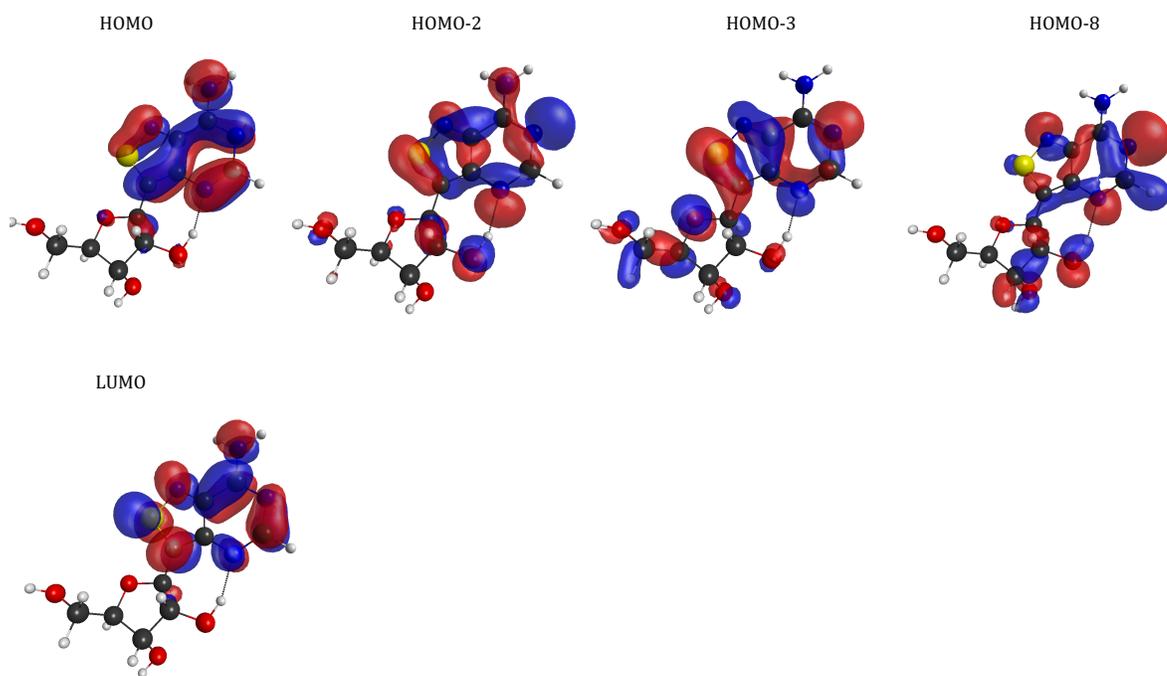
**Figure C5:** Valence molecular orbitals of <sup>th</sup>5-6azaU computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



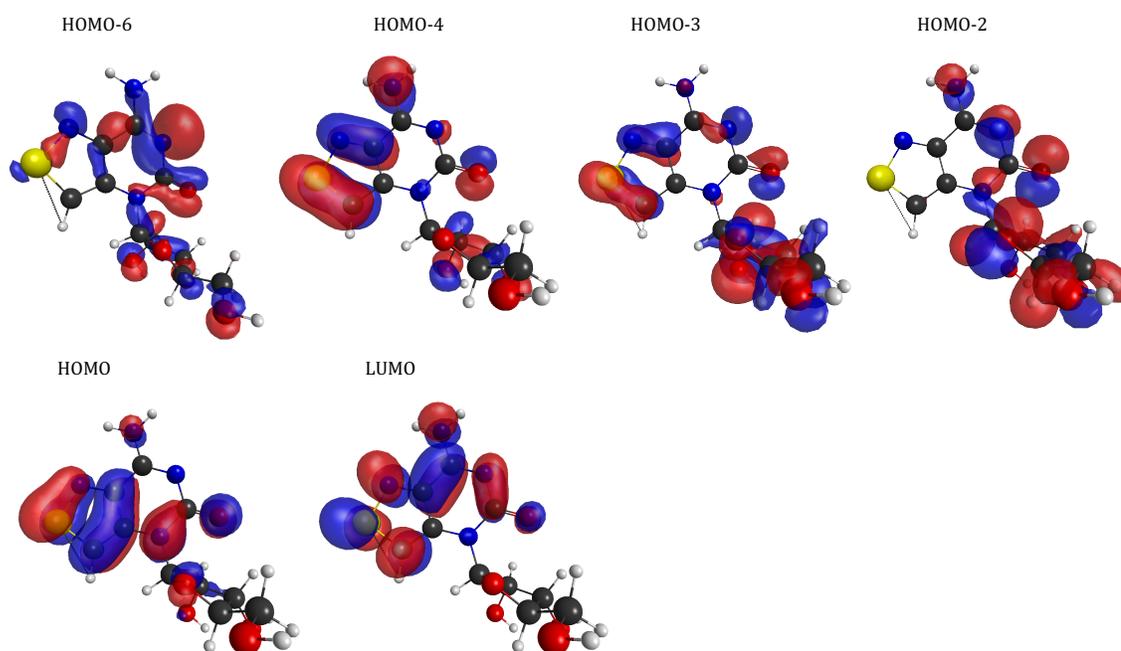
**Figure C6:** Valence molecular orbitals of 5-furan2U computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



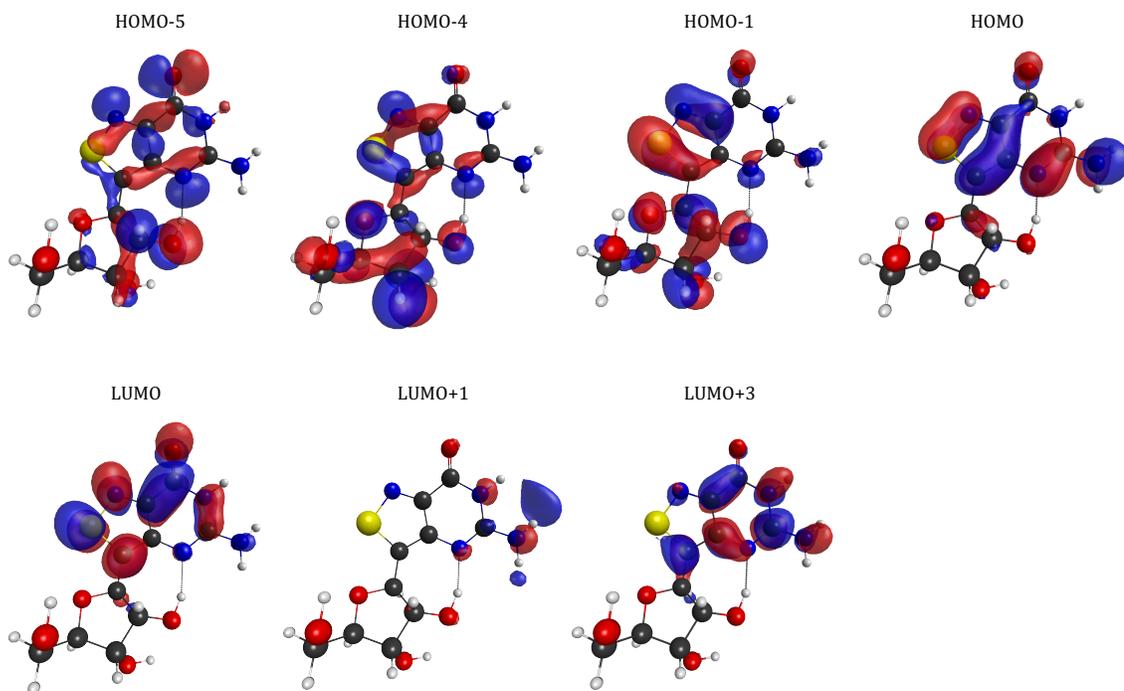
**Figure C7:** Valence molecular orbitals of 7-amino computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



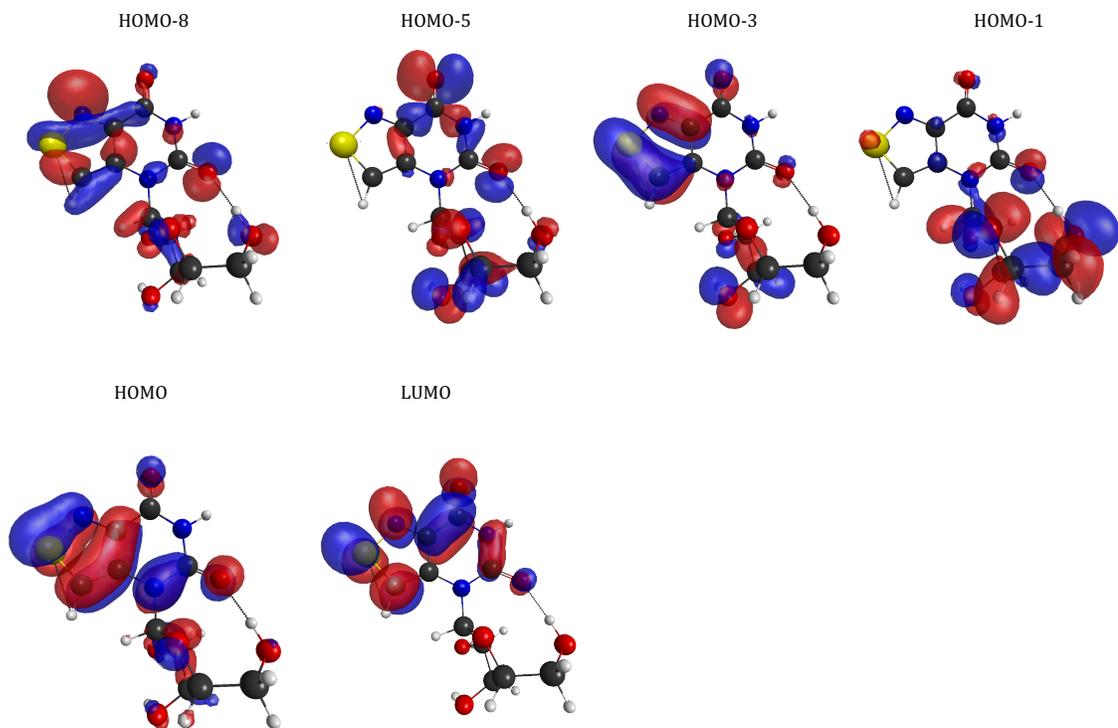
**Figure C8:** Valence molecular orbitals of  $t^2A$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



**Figure C9:** Valence molecular orbitals of  ${}^{tz}C$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



**Figure C10:** Valence molecular orbitals of  $t^z G$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.



**Figure C11:** Valence molecular orbitals of  $t^zU$  computed at  $S_0$  minimum at the TD-DFT CAM-B3LYP/6-31++G(d,p) level of theory with solvation in water. Visualization of orbitals was performed with MacMolPlt<sup>194</sup> v7.7, with isovalue=0.03 a.u.

## C2.2 Two-Photon Absorption Properties for $S_1$

**Table C3:** Two-photon absorption properties of the first excited state,  $S_1$  for the modified nucleobases computed with B3LYP/6-31++g(d,p) in water (PCM)

Nucleobase	Energy (eV)	f	Energy(nm)	$\langle\delta^{TPA}\rangle$	$\sigma^{TPA}(\text{GM})$
<sup>th</sup> 5-6azaU	3.480	0.291	713	572	3.7
<sup>th</sup> 5-2U	3.822	0.272	649	230	2.4
5-furan2U	3.759	0.279	660	517	4.0
amino-7	4.288	0.264	578	230	2.3
<sup>th</sup> A	3.782	0.162	656	31	0.2
<sup>th</sup> C	3.877	0.105	640	67	0.5
<sup>th</sup> G	3.784	0.094	655	131	1.0
<sup>th</sup> U	4.100	0.082	605	75	0.7
<sup>tz</sup> A	3.782	0.174	652	79	0.6
<sup>tz</sup> C	3.877	0.175	642	196	0.7
<sup>tz</sup> G	3.784	0.123	671	172	1.3
<sup>tz</sup> U	4.081	0.149	608	84	0.8

**Table C4:** Two-photon absorption properties of the first excited state,  $S_1$  for the modified nucleosides computed with B3LYP/6-31++g(d,p) in water (PCM)

Nucleoside	Energy (eV)	f	Energy(nm)	$\langle\delta^{TPA}\rangle$	$\sigma^{TPA}(\text{GM})$
<sup>th</sup> 5-6azaU	3.362	0.380	738	564	3.5
<sup>th</sup> 5-2U	3.757	0.308	660	283	2.2
5-furan2U	3.665	0.344	677	682	4.9
amino-7	4.265	0.293	581	310	3.1
<sup>th</sup> A	3.649	0.202	680	41	0.3
<sup>th</sup> C	3.887	0.110	638	157	1.3
<sup>th</sup> G	3.664	0.133	677	231	1.7
<sup>th</sup> U	4.121	0.082	602	157	1.4
<sup>tz</sup> A	3.750	0.248	326	56	0.4
<sup>tz</sup> C	3.835	0.141	321	130	1.0
<sup>tz</sup> G	3.700	0.176	336	250	1.9
<sup>tz</sup> U	4.070	0.116	304	177	1.6

**Table C5:** Two-photon absorption properties of the first excited state,  $S_1$  for the modified nucleobases computed with CAM-B3LYP/6-31++g(d,p) in water (PCM)

Nucleobase	Energy (eV)	f	Energy(nm)	$\langle\delta^{TPA}\rangle$	$\sigma^{TPA}(\text{GM})$
<sup>th</sup> 5-6azaU	3.848	0.391	644	410	3.3
<sup>th</sup> 5-2U	4.210	0.380	589	109	1.0
<b>5-furan2U</b>	4.174	0.409	594	319	3.0
<b>amino-7</b>	4.548	0.289	545	203	2.3
<sup>th</sup> A	4.045	0.205	613	29	0.3
<sup>th</sup> C	4.291	0.149	578	61	0.6
<sup>th</sup> G	4.223	0.141	587	124	1.2
<sup>th</sup> U	4.555	0.122	544	77	0.9
<sup>tz</sup> A	4.088	0.235	607	58	0.5
<sup>tz</sup> C	4.208	0.230	589	62	0.6
<sup>tz</sup> G	4.090	0.183	606	142	1.3
<sup>tz</sup> U	4.446	0.208	588	72	0.8

**Table C6:** Two-photon absorption properties of the first excited state,  $S_1$  for the modified nucleosides computed with CAM-B3LYP/6-31++g(d,p) in water (PCM)

Nucleoside	Energy (eV)	f	Energy(nm)	$\langle\delta^{TPA}\rangle$	$\sigma^{TPA}(\text{GM})$
<b>th5-6azaU</b>	3.733	0.485	664	521	3.9
<b>th5-2U</b>	4.147	0.418	598	108	1.0
<b>5-furan2U</b>	4.064	0.473	610	413	3.7
<b>amino-7</b>	4.543	0.324	546	273	3.1
<sup>th</sup> A	3.921	0.254	632	39	0.3
<sup>th</sup> C	4.339	0.154	571	142	1.4
<sup>th</sup> G	4.115	0.195	603	196	1.8
<sup>th</sup> U	4.610	0.124	538	145	1.7
<sup>tz</sup> A	4.038	0.317	614	41	0.4
<sup>tz</sup> C	4.219	0.199	588	95	0.9
<sup>tz</sup> G	4.098	0.234	605	191	1.7
<sup>tz</sup> U	4.504	0.191	551	111	1.2

### C3 Two-photon Absorption Properties for all Low-Lying States

**Table C7:** Two-photon absorption properties of the low-lying states of <sup>th</sup>5-6azaU, <sup>th</sup>5-2U, 5-furan2U and 7-amino computed with B3LYP/6-31++g(d,p) in water (PCM)

Nucleoside	S <sub>n</sub>	Energy (eV)	f (a.u.)	Energy (nm)	$\langle\delta^{TPA}\rangle$ (a.u.)	$\sigma^{TPA}$ (GM)
<sup>th</sup> 5-6azaU	S <sub>1</sub>	3.362	0.380	738	564	3.5
	S <sub>2</sub>	3.953	0.019	627	298	2.5
	S <sub>3</sub>	4.215	0.001	588	224	2.1
	S <sub>4</sub>	4.421	0.000	561	162	1.7
	S <sub>5</sub>	4.520	0.064	549	4848	53.6
<sup>th</sup> 5-2U	S <sub>1</sub>	3.757	0.308	660	283	2.2
	S <sub>2</sub>	4.546	0.098	545	164	1.8
	S <sub>3</sub>	4.612	0.222	538	386	4.4
	S <sub>4</sub>	4.798	0.010	517	336	4.2
	S <sub>5</sub>	4.998	0.019	496	1765	23.8
5-furan2U	S <sub>1</sub>	3.665	0.344	677	682	5.0
	S <sub>2</sub>	4.640	0.376	534	86	1.0
	S <sub>3</sub>	4.853	0.000	511	21	0.3
	S <sub>4</sub>	4.911	0.020	505	102	1.3
	S <sub>5</sub>	5.146	0.019	482	1449	20.8
amino-7	S <sub>1</sub>	4.265	0.293	581	310	3.1
	S <sub>2</sub>	4.521	0.082	548	248	2.7
	S <sub>3</sub>	4.911	0.207	505	31	0.4
	S <sub>4</sub>	4.971	0.002	499	4	0.1
	S <sub>5</sub>	5.109	0.001	485	37	0.5

**Table C8:** Two-photon absorption properties of the low-lying states of  ${}^{th}A$ ,  ${}^{th}C$ ,  ${}^{th}G$  and  ${}^{th}U$  computed with  $B3LYP/6-31++g(d,p)$  in water (PCM)

Nucleoside	$S_n$	Energy (eV)	$f$ (a.u.)	Energy (nm)	$\langle\delta^{TPA}\rangle$ (a.u.)	$\sigma^{TPA}$ (GM)
${}^{th}A$	$S_1$	3.649	0.202	680	41	0.3
	$S_2$	4.233	0.004	586	11	0.1
	$S_3$	4.660	0.064	532	335	3.9
	$S_4$	4.919	0.003	504	484	6.3
	$S_5$	5.029	0.029	493	390	5.3
${}^{th}C$	$S_1$	3.887	0.110	638	157	1.3
	$S_2$	4.685	0.017	529	13	0.2
	$S_3$	4.740	0.252	523	62	0.8
	$S_4$	4.906	0.011	505	10	0.1
	$S_5$	4.996	0.074	496	53	0.7
${}^{th}G$	$S_1$	3.664	0.133	677	231	1.7
	$S_2$	4.692	0.022	528	135	1.6
	$S_3$	4.831	0.001	513	18	0.2
	$S_4$	4.941	0.001	502	67	0.9
	$S_5$	5.035	0.021	492	131	1.8
${}^{th}U$	$S_1$	4.121	0.088	602	157	1.4
	$S_2$	4.732	0.002	524	2	0.0
	$S_3$	4.976	0.083	498	28	0.4
	$S_4$	5.079	0.082	488	17	0.2
	$S_5$	5.558	0.021	446	7	0.1

**Table C9:** Two-photon absorption properties of the low-lying states of  ${}^{tz}A$ ,  ${}^{tz}C$ ,  ${}^{tz}G$  and  ${}^{tz}U$  computed with  $B3LYP/6-31++g(d,p)$  in water (PCM)

Nucleoside	$S_n$	Energy (eV)	$f$ (a.u.)	Energy (nm)	$\langle\delta^{TPA}\rangle$ (a.u.)	$\sigma^{TPA}$ (GM)
${}^{tz}A$	$S_1$	3.750	0.248	661	56	0.4
	$S_2$	4.016	0.009	617	157	1.4
	$S_3$	4.378	0.001	566	885	9.2
	$S_4$	4.682	0.013	530	149	1.8
	$S_5$	4.918	0.055	504	27	0.4
${}^{tz}C$	$S_1$	3.835	0.141	647	130	1.0
	$S_2$	4.388	0.029	565	69	0.7
	$S_3$	4.560	0.124	544	130	1.5
	$S_4$	4.614	0.008	537	9	0.1
	$S_5$	4.930	0.007	503	110	1.4
${}^{tz}G$	$S_1$	3.700	0.176	670	250	1.9
	$S_2$	4.527	0.007	548	331	3.7
	$S_3$	4.645	0.008	534	290	3.4
	$S_4$	4.888	0.003	507	119	1.5
	$S_5$	5.024	0.005	494	13	0.2
${}^{tz}U$	$S_1$	4.070	0.116	609	177	1.6
	$S_2$	4.410	0.011	562	11	0.1
	$S_3$	4.585	0.009	541	21	0.2
	$S_4$	4.933	0.020	503	105	1.4
	$S_5$	5.071	0.001	489	15	0.2

**Table C10:** *Rotation of<sup>th</sup> 5-6azaU*

$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)	$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)
0	2.2	3.436	0.366	3.3	92	20.3	3.378	0.001	0.1
2	2.1	3.438	0.367	3.3	94	20.1	3.384	0.003	0.4
4	1.9	3.439	0.367	3.3	96	19.9	3.387	0.007	0.8
6	1.8	3.441	0.367	3.3	98	19.7	3.389	0.012	1.4
8	1.7	3.443	0.367	3.4	100	19.5	3.389	0.019	2.1
10	1.6	3.445	0.366	3.5	102	19.2	3.391	0.027	2.9
12	1.6	3.448	0.365	3.6	104	18.8	3.396	0.036	3.7
14	1.6	3.451	0.364	3.7	106	18.3	3.403	0.047	4.6
16	1.6	3.454	0.362	3.8	108	17.8	3.409	0.058	5.5
18	1.6	3.456	0.359	3.9	110	17.3	3.415	0.071	6.3
20	1.7	3.459	0.356	4.1	112	16.8	3.423	0.084	7.0
22	1.3	3.461	0.351	4.3	114	16.2	3.432	0.098	7.7
24	1.4	3.464	0.347	4.5	116	15.7	3.439	0.113	8.2
26	1.5	3.468	0.342	4.8	118	15.2	3.443	0.128	8.7
28	1.8	3.471	0.337	5.0	120	14.6	3.446	0.143	9.1
30	2.1	3.474	0.331	5.3	122	14.0	3.449	0.158	9.3
32	2.6	3.475	0.323	5.6	124	13.4	3.452	0.173	9.4
34	3.2	3.475	0.315	6.0	126	12.8	3.457	0.188	9.4
36	3.8	3.474	0.305	6.3	128	12.0	3.462	0.204	9.3
38	4.5	3.473	0.295	6.7	130	11.3	3.466	0.219	9.1
40	5.3	3.472	0.284	7.1	132	10.7	3.468	0.233	8.8
42	6.0	3.471	0.272	7.4	134	10.0	3.468	0.247	8.5
44	6.8	3.471	0.260	7.8	136	9.4	3.467	0.260	8.2
46	7.6	3.470	0.247	8.1	138	8.7	3.465	0.272	7.9
48	8.4	3.469	0.234	8.4	140	8.1	3.463	0.284	7.5
50	9.3	3.466	0.220	8.6	142	7.4	3.459	0.295	7.1
52	10.1	3.462	0.205	8.8	144	6.8	3.455	0.305	6.8
54	11.1	3.456	0.190	8.9	146	6.1	3.450	0.314	6.4
56	12.0	3.448	0.174	9.0	148	5.5	3.445	0.323	6.0
58	13.0	3.441	0.158	8.9	150	4.9	3.440	0.331	5.7
60	13.8	3.433	0.143	8.8	152	4.3	3.433	0.339	5.3
62	14.7	3.423	0.127	8.5	154	3.8	3.425	0.345	5.1
64	15.4	3.415	0.112	8.1	156	3.3	3.417	0.350	4.8
66	15.9	3.410	0.097	7.6	158	2.8	3.408	0.355	4.6
68	16.3	3.405	0.083	7.0	160	2.4	3.400	0.359	4.4
70	16.8	3.397	0.070	6.2	162	2.0	3.391	0.362	4.2
72	17.2	3.391	0.058	5.5	164	1.7	3.384	0.365	4.1
74	17.7	3.385	0.046	4.6	166	1.8	3.378	0.368	3.9
76	18.2	3.380	0.036	3.8	168	1.5	3.373	0.371	3.8
78	18.6	3.378	0.027	2.9	170	1.2	3.369	0.374	3.7
80	18.9	3.378	0.019	2.1	172	0.9	3.365	0.375	3.6
82	19.1	3.379	0.012	1.4	174	0.7	3.361	0.377	3.5
84	19.3	3.383	0.007	0.8	176	0.4	3.359	0.378	3.5
86	19.5	3.386	0.003	0.4	178	0.2	3.358	0.379	3.4
88	19.9	3.382	0.001	0.1	180	0.1	3.359	0.380	3.4
90	20.3	3.378	0.000	0.0					

**Table C11:** *Rotation of <sup>th</sup>5-2U*

$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)	$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)
0	2.0	3.660	0.337	1.5	92	6.6	4.163	0.003	0.1
2	2.6	3.655	0.335	1.5	94	6.5	4.163	0.007	0.3
4	2.6	3.657	0.334	1.5	96	6.3	4.159	0.014	0.6
6	2.6	3.661	0.334	1.5	98	6.2	4.153	0.023	0.9
8	2.5	3.666	0.334	1.5	100	6.0	4.146	0.033	1.3
10	2.3	3.672	0.333	1.5	102	5.6	4.141	0.045	1.7
12	2.3	3.677	0.331	1.6	104	5.3	4.137	0.059	2.1
14	2.1	3.683	0.329	1.7	106	4.9	4.132	0.074	2.4
16	2.0	3.690	0.326	1.7	108	4.5	4.125	0.089	2.8
18	2.0	3.698	0.323	1.8	110	4.1	4.116	0.105	3.1
20	1.9	3.705	0.319	1.9	112	3.8	4.103	0.120	3.4
22	1.9	3.714	0.314	2.0	114	3.5	4.089	0.135	3.7
24	1.9	3.724	0.309	2.1	116	3.2	4.074	0.149	3.9
26	1.8	3.735	0.304	2.2	118	3.1	4.058	0.163	4.1
28	1.7	3.747	0.299	2.3	120	2.9	4.043	0.177	4.3
30	1.6	3.761	0.293	2.4	122	2.7	4.027	0.191	4.4
32	1.5	3.775	0.287	2.5	124	2.4	4.012	0.204	4.4
34	1.5	3.789	0.280	2.7	126	2.1	3.996	0.217	4.5
36	1.6	3.803	0.272	2.8	128	0.9	3.993	0.232	4.3
38	1.9	3.816	0.264	2.9	130	0.8	3.974	0.243	4.3
40	2.2	3.829	0.254	3.1	132	0.7	3.954	0.254	4.2
42	2.5	3.841	0.245	3.3	134	0.6	3.934	0.264	4.2
44	2.8	3.855	0.235	3.4	136	0.6	3.914	0.273	4.1
46	3.1	3.869	0.225	3.5	138	0.7	3.893	0.281	4.1
48	3.3	3.885	0.214	3.6	140	0.8	3.872	0.289	4.0
50	3.5	3.902	0.204	3.6	142	0.8	3.854	0.297	3.9
52	3.7	3.920	0.193	3.6	144	0.7	3.837	0.305	3.8
54	3.0	3.938	0.183	3.6	146	0.6	3.819	0.312	3.6
56	3.3	3.956	0.171	3.5	148	0.6	3.799	0.317	3.5
58	3.6	3.973	0.159	3.5	150	1.5	3.776	0.322	3.4
60	4.0	3.989	0.147	3.4	152	1.4	3.758	0.326	3.3
62	5.1	3.996	0.131	3.3	154	1.4	3.740	0.330	3.2
64	5.4	4.011	0.119	3.1	156	1.4	3.723	0.333	3.1
66	5.6	4.027	0.106	2.9	158	1.4	3.706	0.335	3.0
68	5.9	4.041	0.093	2.6	160	1.5	3.689	0.336	3.0
70	6.1	4.056	0.081	2.3	162	1.6	3.674	0.337	3.0
72	6.3	4.072	0.068	2.0	164	1.7	3.660	0.338	2.9
74	6.4	4.087	0.056	1.6	166	1.0	3.651	0.341	2.8
76	6.6	4.100	0.045	1.3	168	1.0	3.641	0.342	2.8
78	6.7	4.114	0.034	1.0	170	1.0	3.632	0.343	2.7
80	6.8	4.125	0.024	0.7	172	1.2	3.625	0.344	2.7
82	6.9	4.133	0.015	0.4	174	1.4	3.617	0.343	2.7
84	6.9	4.140	0.009	0.2	176	1.5	3.611	0.342	2.7
86	7.0	4.145	0.004	0.1	178	1.7	3.607	0.342	2.8
88	6.9	4.150	0.001	0.0	180	1.8	3.605	0.341	2.8
90	6.8	4.157	0.001	0.0					

**Table C12:** *Rotation of 5-furan2U*

$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)	$\theta$ ( $^{\circ}$ )	$\Delta E$ (kJ/mol)	$\omega_0$ (eV)	$f$ (a.u.)	$\sigma^{TPA}$ (GM)
0	0.0	3.663	0.343	5.0	92	16.8	3.918	0.004	0.4
2	0.2	3.662	0.342	5.0	94	16.5	3.922	0.008	0.8
4	0.3	3.661	0.340	5.1	96	16.2	3.922	0.013	1.2
6	0.5	3.661	0.338	5.1	98	16.1	3.919	0.020	1.8
8	0.7	3.664	0.336	5.2	100	15.9	3.918	0.028	2.4
10	0.8	3.669	0.335	5.2	102	15.7	3.915	0.038	3.0
12	1.0	3.675	0.333	5.3	104	15.4	3.909	0.048	3.7
14	1.3	3.680	0.330	5.3	106	15.1	3.905	0.059	4.3
16	1.6	3.683	0.325	5.4	108	14.7	3.904	0.071	5.0
18	2.1	3.686	0.320	5.6	110	14.2	3.904	0.084	5.5
20	2.5	3.689	0.314	5.7	112	13.9	3.898	0.097	6.1
22	3.0	3.692	0.308	5.9	114	13.6	3.891	0.110	6.6
24	3.6	3.694	0.301	6.1	116	13.3	3.883	0.122	7.0
26	4.2	3.698	0.294	6.3	118	13.0	3.875	0.135	7.3
28	4.7	3.702	0.286	6.4	120	12.8	3.866	0.148	7.6
30	5.3	3.708	0.278	6.6	122	12.6	3.856	0.160	7.8
32	5.9	3.714	0.269	6.8	124	12.4	3.846	0.173	8.0
34	6.5	3.720	0.261	6.9	126	12.3	3.834	0.184	8.1
36	7.2	3.727	0.252	7.1	128	12.2	3.822	0.196	8.2
38	7.9	3.732	0.242	7.2	130	12.3	3.809	0.206	8.2
40	8.7	3.737	0.232	7.4	132	12.4	3.796	0.217	8.2
42	9.5	3.741	0.221	7.5	134	12.6	3.781	0.227	8.2
44	10.3	3.747	0.211	7.5	136	12.9	3.764	0.236	8.1
46	11.0	3.754	0.200	7.6	138	13.2	3.748	0.244	8.1
48	11.7	3.760	0.188	7.6	140	13.6	3.733	0.252	8.0
50	12.4	3.765	0.176	7.5	142	14.1	3.716	0.260	7.9
52	13.0	3.772	0.164	7.4	144	14.5	3.699	0.266	7.8
54	13.6	3.778	0.152	7.2	146	15.0	3.685	0.273	7.6
56	14.2	3.785	0.139	6.9	148	15.4	3.672	0.280	7.5
58	14.7	3.792	0.127	6.6	150	15.9	3.659	0.287	7.3
60	15.2	3.801	0.115	6.2	152	16.4	3.647	0.293	7.2
62	15.7	3.809	0.102	5.8	154	16.9	3.636	0.298	7.0
64	16.1	3.816	0.090	5.3	156	17.4	3.625	0.304	6.9
66	16.5	3.823	0.078	4.8	158	18.0	3.614	0.308	6.8
68	16.8	3.832	0.067	4.2	160	18.6	3.604	0.312	6.7
70	17.0	3.842	0.056	3.5	162	19.3	3.594	0.316	6.6
72	17.3	3.851	0.045	2.9	164	19.9	3.585	0.319	6.5
74	17.4	3.858	0.036	2.3	166	20.5	3.577	0.322	6.4
76	17.6	3.865	0.027	1.7	168	20.9	3.570	0.324	6.3
78	17.6	3.873	0.019	1.2	170	21.5	3.564	0.326	6.3
80	17.6	3.882	0.013	0.8	172	21.8	3.558	0.327	6.2
82	17.5	3.892	0.007	0.4	174	22.1	3.554	0.328	6.2
84	17.4	3.900	0.003	0.2	176	22.3	3.551	0.328	6.2
86	17.3	3.907	0.001	0.0	178	22.4	3.549	0.328	6.2
88	17.2	3.911	0.000	0.0	180	22.4	3.549	0.328	6.2
90	17.0	3.915	0.001	0.1					

## C4 Coordinates

### C4.1 Nucleobases

All geometries were optimized at the PBE0/6-31++g(d,p) level of theory with PCM solvation in water.

th5-6azaU			th5-2U				
C	-1.0776	1.2208	0.0817	C	1.0465	1.1498	-0.1342
C	-0.2214	0.0077	0.0240	C	0.2112	-0.0324	0.0827
C	-3.0005	-0.3167	-0.0018	C	0.8479	-1.2222	0.2717
N	-2.0711	-1.3229	-0.0364	C	3.0612	-0.3023	0.0166
O	-4.2063	-0.5102	-0.0196	N	2.2035	-1.3524	0.2324
O	-0.6774	2.3716	0.1478	O	4.2790	-0.4239	-0.0144
N	-2.4271	0.9352	0.0562	H	2.6294	-2.2580	0.3806
H	-3.0659	1.7228	0.1028	O	0.6457	2.2977	-0.2909
C	1.2275	0.1129	0.0267	N	2.4160	0.8976	-0.1571
C	2.0281	1.2353	-0.0115	H	3.0131	1.7008	-0.3207
C	3.4132	0.9235	-0.0122	C	-1.2433	0.0854	0.0840
C	3.6481	-0.4247	0.0291	C	-2.0224	1.1773	0.3947
H	1.6250	2.2387	-0.0393	C	-3.4173	0.9116	0.3087
H	4.2015	1.6671	-0.0408	C	-3.6910	-0.3727	-0.0733
H	4.6003	-0.9402	0.0378	H	0.3094	-2.1421	0.4716
S	2.1911	-1.3350	0.0634	H	-1.5985	2.1329	0.6751
N	-0.7440	-1.1811	-0.0261	H	-4.1870	1.6448	0.5239
H	-2.4084	-2.2759	-0.0697	H	-4.6538	-0.8453	-0.2197
				S	-2.2484	-1.2703	-0.3465

<b>5-furan2U</b>				<b>thA</b>			
C	0.3962	-1.2987	0.0361	C	0.3038	-0.9578	0.0134
C	-0.0004	0.1051	0.0050	C	-0.0458	0.4385	-0.0181
C	0.9840	1.0473	-0.0202	C	-1.4551	0.7386	-0.0219
C	2.7733	-0.5691	0.0002	C	-1.8789	-1.5236	-0.0178
N	2.3069	0.7218	-0.0198	N	-0.6427	-1.9525	0.0151
O	3.9642	-0.8544	-0.0109	C	1.0572	1.2677	-0.0361
H	3.0060	1.4517	-0.0660	C	1.6734	-1.1457	0.0304
O	-0.3646	-2.2611	0.0625	S	2.4946	0.3519	-0.0055
N	1.7717	-1.5094	0.0340	H	1.1151	2.3472	-0.0840
H	2.0750	-2.4769	0.0480	H	2.2168	-2.0806	0.0530
C	-1.4042	0.4636	0.0046	N	-1.9068	2.0025	-0.0322
C	-2.5709	-0.2507	-0.0511	H	-2.9016	2.1485	0.0488
C	-3.6305	0.7081	-0.0230	H	-1.2960	2.7902	0.1137
C	-3.0384	1.9303	0.0451	H	-2.6579	-2.2845	-0.0281
O	-1.6856	1.7966	0.0615	N	-2.3428	-0.2470	-0.0309
H	0.7641	2.1078	-0.0382				
H	-2.6473	-1.3257	-0.1030				
H	-4.6932	0.5113	-0.0511				
H	-3.4103	2.9430	0.0863				

<b>7-amino</b>				<b>thC</b>			
C	1.3497	1.3175	0.0018	C	0.7651	-0.5188	-0.0106
C	0.0089	0.7643	0.0058	C	-0.0005	0.6866	-0.0045
C	-0.1748	-0.6314	0.0099	C	-1.4456	0.5456	-0.0075
C	-1.1266	1.5904	0.0031	C	-1.2922	-1.7935	0.0017
C	2.2323	-1.0017	-0.0036	N	-2.0283	-0.6369	-0.0068
N	0.9434	-1.4496	0.0123	N	0.0876	-1.7165	-0.0176
C	-1.4500	-1.1857	0.0126	O	-1.8308	-2.9076	0.0272
H	-0.9784	2.6658	-0.0018	C	0.7934	1.8116	0.0063
O	3.2014	-1.7548	-0.0183	C	2.1181	-0.2954	-0.0048
H	0.8317	-2.4547	-0.0113	S	2.4446	1.3947	0.0090
H	-1.5785	-2.2636	0.0379	H	0.5013	2.8532	0.0207
O	1.6349	2.5142	-0.0014	H	0.5903	-2.5931	0.0090
N	2.3672	0.3699	0.0018	H	2.9284	-1.0112	-0.0093
H	3.3186	0.7183	-0.0131	N	-2.2218	1.6346	-0.0095
C	-2.3952	1.0575	0.0035	H	-3.2247	1.5236	-0.0028
C	-2.5758	-0.3489	0.0054	H	-1.8432	2.5672	-0.0006
N	-3.8338	-0.8767	0.0445				
H	-3.2655	1.7072	0.0002				
H	-4.6112	-0.2807	-0.1948				
H	-3.9602	-1.8489	-0.1931				

	<b>thG</b>		
S	-2.8300	-0.3600	0.0009
O	0.7259	2.6658	-0.0063
N	0.9168	-1.4622	0.0073
N	1.8773	0.7056	-0.0065
N	3.2117	-1.1860	-0.0546
C	-0.3133	-0.8409	-0.0027
C	-0.4759	0.5889	0.0050
C	0.7023	1.4368	-0.0007
C	1.9525	-0.6775	0.0012
C	-1.7904	0.9903	0.0077
C	-1.5313	-1.4853	-0.0054
H	-2.1829	1.9980	0.0142
H	-1.7235	-2.5497	-0.0130
H	2.7319	1.2496	-0.0324
H	3.9834	-0.6248	0.2728
H	3.2896	-2.1746	0.1335

	<b>tzA</b>		
S	2.4029	-0.3849	-0.0019
N	1.0050	-1.2624	0.0044
N	-0.6607	1.9638	0.0032
N	-2.3158	0.2078	0.0015
N	-1.7765	-2.0400	0.0182
H	-1.0848	-2.7735	-0.0124
H	-2.7551	-2.2844	-0.0165
C	1.6811	1.1470	0.0003
C	0.3016	0.9899	0.0050
C	-0.0071	-0.4034	0.0069
C	-1.4067	-0.7571	0.0073
C	-1.8864	1.4960	-0.0003
H	-2.6853	2.2353	-0.0045
H	2.2559	2.0642	-0.0113

	<b>thU</b>		
C	-0.7696	-1.4086	0.0043
C	0.4313	-0.5824	0.0007
C	0.3257	0.8421	0.0046
C	1.7363	-1.0131	-0.0038
C	-2.0890	0.7131	0.0011
N	-0.9294	1.4285	0.0106
C	1.5433	1.4679	0.0032
S	2.8135	0.3045	-0.0041
H	2.0958	-2.0330	-0.0113
O	-3.1989	1.2342	-0.0062
H	-1.0287	2.4344	-0.0140
H	1.7666	2.5261	0.0086
O	-0.7997	-2.6331	0.0097
N	-1.9413	-0.6677	-0.0002
H	-2.8107	-1.1890	-0.0031

	<b>tzC</b>		
S	2.7217	-0.2671	-0.0095
O	-3.2048	-1.2007	-0.0169
N	-0.9605	-1.4457	0.0163
N	1.6624	0.9945	-0.0083
N	-0.6870	2.6211	0.0050
H	0.2156	3.0719	-0.0079
H	-1.5214	3.1884	0.0008
N	-1.9697	0.7143	0.0083
C	1.5434	-1.5030	0.0025
H	1.8129	-2.5510	0.0070
C	0.3018	-0.9094	0.0071
C	0.4355	0.5022	0.0007
C	-0.7879	1.2933	0.0060
C	-2.0983	-0.6532	0.0024
H	-1.1019	-2.4466	-0.0135

	<b>tzG</b>				<b>tzU</b>		
S	-2.7629	-0.3142	0.0058	S	2.7457	-0.2415	-0.0106
O	0.6395	2.6624	-0.0130	O	-0.7191	2.6401	0.0123
N	-1.7610	0.9938	0.0046	O	-3.1984	-1.1805	-0.0077
N	0.9254	-1.4647	-0.0094	N	-0.9335	-1.4259	0.0142
N	1.8351	0.7255	0.0080	N	1.7022	1.0281	-0.0140
H	2.6769	1.2910	0.0349	N	-1.9007	0.6946	0.0007
N	3.2099	-1.1380	0.0565	H	-2.7614	1.2317	-0.0034
H	3.3119	-2.1272	-0.1168	C	1.5569	-1.4644	0.0038
H	3.9811	-0.5629	-0.2481	H	1.8117	-2.5163	0.0156
C	-1.5405	-1.4948	0.0005	C	0.3244	-0.8572	0.0038
C	-0.3084	-0.8641	-0.0003	C	0.4686	0.5567	-0.0053
C	-0.5139	0.5513	0.0011	C	-0.7222	1.4195	0.0051
C	0.6494	1.4377	-0.0047	C	-2.0799	-0.6847	0.0017
C	1.9454	-0.6551	-0.0002	H	-1.0492	-2.4303	-0.0162
H	-1.7566	-2.5555	-0.0012				

## C4.2 Nucleosides

All geometries were optimized at the PBE0/6-31++g(d,p) level of theory with PCM solvation in water.

<b>th5-6azaU</b>				<b>th5-2U</b>			
C	2.0596	-1.6612	-0.2165	C	-2.2930	1.9178	0.0115
C	2.0682	-0.1940	-0.0356	C	-1.8515	0.5242	0.0446
N	0.9636	0.4726	0.1076	C	-0.5098	0.2937	0.0234
C	-0.3771	-1.5180	-0.0025	C	0.0965	2.6141	-0.0741
N	-0.2234	-0.1605	0.1455	N	0.4394	1.2752	-0.0496
O	-1.4645	-2.0906	0.0468	O	0.9133	3.5243	-0.1325
O	3.0534	-2.3494	-0.3798	O	-3.4555	2.3088	0.0194
N	0.7945	-2.2010	-0.1972	N	-1.2573	2.8426	-0.0304
H	0.7104	-3.2077	-0.3033	C	-2.8307	-0.5568	0.0968
C	3.3226	0.5420	-0.0258	C	-4.1081	-0.5382	0.6099
C	3.4594	1.8998	0.1867	C	-4.7671	-1.7953	0.5182
C	4.8034	2.3365	0.1373	C	-3.9950	-2.7562	-0.0747
C	5.6756	1.3065	-0.1116	H	-0.1042	-0.7101	0.0945
H	2.6109	2.5478	0.3724	H	-4.5478	0.3505	1.0437
H	5.1136	3.3651	0.2811	H	-5.7738	-1.9784	0.8774
H	6.7542	1.3478	-0.2002	H	-4.2380	-3.7914	-0.2773
S	4.8740	-0.1985	-0.2856	S	-2.4582	-2.1352	-0.5358
O	-5.7214	0.4663	-1.5313	H	-1.5307	3.8191	-0.0491
C	-4.3768	0.8444	-1.7605	C	1.8619	0.9113	-0.0847
C	-3.5244	0.2535	-0.6547	H	2.3979	1.8342	0.1482
H	-3.6686	-0.8353	-0.6106	C	2.8955	-1.1380	0.3900
O	-2.1647	0.5559	-0.9914	H	2.2517	-2.0102	0.2001
C	-1.3860	0.7437	0.1640	C	3.4322	-0.6335	-0.9504
C	-3.7065	0.8663	0.7403	O	2.1007	-0.0634	0.9115
H	-3.8994	1.9410	0.6356	C	2.3124	0.2983	-1.4041
C	-2.3214	0.6404	1.3864	H	2.6228	1.0539	-2.1289
H	-2.0788	1.4172	2.1149	H	1.5077	-0.3009	-1.8466
O	-2.2991	-0.5784	2.0922	C	3.9381	-1.5484	1.4021
H	-2.2121	-1.3124	1.4518	H	3.4353	-1.8145	2.3412
O	-4.7385	0.3133	1.5093	H	4.4410	-2.4456	1.0157
H	-4.3363	-0.4008	2.0306	O	4.8573	-0.4883	1.5980
H	-0.9230	1.7331	0.1292	H	5.5599	-0.7956	2.1813
H	-6.2582	0.7955	-2.2602	O	3.6630	-1.7405	-1.7946
H	-4.0135	0.4630	-2.7241	H	4.2139	-1.4622	-2.5347
H	-4.2617	1.9379	-1.7581	H	4.3606	-0.0739	-0.7843

5-furan2U				7-amino			
C	2.8770	-0.0429	1.3698	C	2.7896	2.6471	0.4296
C	2.7426	-1.1287	0.2982	O	-1.5177	-0.1488	-1.1324
O	2.0464	-0.5196	-0.8174	O	-1.2404	-1.6093	2.0951
C	1.8552	0.8475	-0.6204	H	-1.1564	-1.2277	2.9767
C	2.7441	1.2436	0.5539	O	-3.5938	-1.6977	0.8980
C	4.0737	-1.6504	-0.2142	H	-3.1680	-2.1509	1.6435
C	-2.3433	1.6587	-0.1095	O	-4.7817	1.4527	-0.9631
C	-1.8496	0.2888	-0.1312	H	-5.0435	1.5193	-0.0380
C	0.0052	2.4263	-0.3894	O	4.6342	-1.1475	-0.0983
N	0.4008	1.1038	-0.3618	O	0.6471	-3.0986	-0.9293
O	0.7961	3.3500	-0.5378	N	0.5858	-0.8725	-0.3314
O	-3.5156	2.0051	-0.0008	C	3.4247	1.4365	0.2854
N	-1.3466	2.6185	-0.2400	N	2.6045	-2.0458	-0.5437
H	-1.6512	3.5862	-0.2456	H	3.0572	-2.9302	-0.7429
C	-2.7773	-0.8194	-0.0325	C	-0.8520	-0.9728	-0.1957
C	-4.1333	-0.9136	0.1334	H	-1.0825	-2.0267	-0.3768
C	-4.4385	-2.3093	0.1590	C	-1.4575	-0.5847	1.1541
C	-3.2552	-2.9621	0.0102	H	-1.0634	0.3693	1.5145
O	-2.2368	-2.0692	-0.1067	C	-2.9456	-0.4517	0.7628
H	-4.8143	-0.0813	0.2218	H	-3.4539	0.3092	1.3724
H	-5.4117	-2.7659	0.2743	C	-2.8929	-0.0581	-0.7248
H	-2.9728	-4.0033	-0.0318	H	-3.4939	-0.7813	-1.2906
C	-0.5078	0.0894	-0.2741	C	-3.3745	1.3495	-1.0489
H	2.0887	1.3895	-1.5391	H	-2.8818	2.0842	-0.3967
H	3.7216	1.5374	0.1593	H	-3.1001	1.5823	-2.0815
H	2.3569	2.0778	1.1400	C	0.6603	1.5474	-0.0020
H	3.8505	-0.0997	1.8636	C	1.3040	0.3159	-0.1014
O	1.9335	-0.1653	2.4198	C	2.7090	0.2595	0.0188
H	2.1323	-1.9674	0.6525	C	3.4188	-0.9853	-0.1875
H	1.0343	-0.1042	2.0764	C	1.2341	-2.0661	-0.6269
H	4.7159	-0.8092	-0.5135	H	4.5046	1.3657	0.3698
O	3.8996	-2.5532	-1.2879	C	1.3865	2.7185	0.2737
H	4.5774	-2.1997	0.5864	H	-0.4055	1.6154	-0.1746
H	3.4316	-2.0825	-1.9895	N	0.7395	3.9107	0.4200
H	-0.0856	-0.9060	-0.3407	H	3.3563	3.5487	0.6415
				H	-0.2130	3.9895	0.0992
				H	1.2766	4.7631	0.3670

thA				thC			
S	-0.3690	-1.7639	0.2141	S	1.9005	2.8340	0.5555
O	4.2644	-2.0831	-0.9802	O	0.7051	-2.9289	-0.9541
H	5.0336	-1.6181	-1.3289	O	-1.7425	-1.8098	1.9839
O	1.8623	-0.3880	-0.9809	H	-1.4865	-1.7610	2.9115
O	3.5919	1.5183	1.4661	O	-3.8358	-0.2399	1.7392
H	3.4398	1.6944	2.4019	H	-3.6467	-1.1613	1.9841
O	1.9410	2.6806	-0.2937	O	-1.5068	-0.0288	-0.9600
H	2.7030	2.9618	0.2368	O	-3.7605	0.8790	-2.5495
N	-2.0908	1.6098	-0.6897	H	-4.2297	1.6107	-2.9664
N	-4.3702	0.9940	-0.2065	N	0.5760	-0.7791	-0.1990
N	-4.9964	-1.0791	0.5579	N	2.6745	-1.8571	-0.6491
H	-5.9496	-0.7528	0.6088	N	4.6736	-0.8116	-0.3322
H	-4.7827	-1.9943	0.9166	H	5.1311	-1.6782	-0.5728
C	3.7578	-1.3679	0.1325	H	5.2432	-0.0270	-0.0619
H	4.5558	-1.1210	0.8458	C	-0.8503	-0.9007	-0.0539
H	3.0513	-2.0395	0.6312	H	-1.0879	-1.9438	-0.2835
C	3.0426	-0.0883	-0.2449	C	1.3080	-1.9035	-0.6270
H	3.7098	0.5518	-0.8439	C	-1.4262	-0.5883	1.3427
C	2.5507	0.7114	0.9588	H	-0.7338	-0.0061	1.9555
H	2.1568	0.0365	1.7292	C	-2.7208	0.2089	1.0182
C	1.4102	1.5381	0.3471	H	-2.5758	1.2691	1.2592
H	0.6501	1.8264	1.0829	C	3.3389	-0.7667	-0.3264
C	0.8415	0.5863	-0.7169	C	-2.8545	0.0476	-0.5052
H	0.6411	1.1706	-1.6236	H	-3.3783	-0.8981	-0.7225
C	-0.4082	-0.1314	-0.3156	C	2.6516	0.4638	0.0094
C	-1.7087	0.3475	-0.3096	C	-3.5697	1.1867	-1.1793
C	-2.6497	-0.6403	0.1413	H	-4.5320	1.3287	-0.6684
C	-2.0530	-1.8408	0.4658	H	-2.9742	2.1034	-1.0591
H	-2.5002	-2.7642	0.8100	C	1.2236	0.4248	0.0444
C	-3.3791	1.8315	-0.6095	C	3.1537	1.7178	0.2708
H	-3.7213	2.8224	-0.9046	H	4.1819	2.0552	0.2951
C	-4.0322	-0.2331	0.1685	C	0.6735	1.6516	0.3243
				H	-0.3654	1.9384	0.3624

thG				thU			
S	0.1258	-1.9699	-0.5325	S	1.9092	2.9066	0.4376
O	-4.4460	-0.8506	2.3416	O	0.8137	-2.9410	-0.9032
H	-5.0880	-1.5213	2.5990	O	-1.2917	-1.7224	1.9514
O	-1.9533	-0.2188	0.9491	H	-1.2250	-1.4300	2.8676
O	-4.1406	0.9053	-1.6663	O	-3.5842	-1.6880	0.6094
H	-3.9047	1.8473	-1.6383	H	-3.2275	-2.2230	1.3361
O	-2.0156	2.3446	-1.1178	O	4.6994	-0.7632	-0.1027
H	-1.4557	2.7863	-1.7648	O	-1.4501	0.0792	-1.1067
O	4.7725	-1.5698	-0.3065	O	-4.6519	1.7721	-0.6506
N	1.9462	1.3240	0.5217	H	-5.0975	1.4494	-1.4425
N	4.1704	0.5258	0.3351	N	0.6068	-0.7272	-0.3153
H	5.1564	0.7498	0.3958	N	2.7072	-1.7723	-0.5140
C	-4.2454	-0.9365	0.9398	H	3.1992	-2.6358	-0.7139
H	-5.1852	-0.7659	0.3974	C	-0.8308	-0.8604	-0.2529
H	-3.8542	-1.9243	0.6571	H	-1.0465	-1.8875	-0.5628
C	-3.2630	0.1202	0.5104	C	1.3290	-1.8726	-0.6009
H	-3.5696	1.0955	0.9226	C	-1.4789	-0.6108	1.1090
C	-3.1020	0.2334	-1.0080	H	-1.0883	0.3047	1.5645
H	-3.0274	-0.7751	-1.4326	C	-2.9491	-0.4287	0.6867
C	-1.7455	0.9505	-1.1247	H	-3.4978	0.2376	1.3648
H	-1.2011	0.6739	-2.0306	C	3.4802	-0.6781	-0.1749
C	-0.9997	0.5157	0.1643	C	-2.8329	0.1729	-0.7219
H	-0.6962	1.4291	0.6925	H	-3.4395	-0.4325	-1.4094
C	0.2180	-0.3156	-0.0325	C	2.6868	0.5195	0.0489
C	1.5354	0.0737	0.1183	C	-3.2552	1.6323	-0.8198
C	2.4469	-0.9948	-0.1886	H	-2.7897	2.2158	-0.0182
C	1.8196	-2.1585	-0.5592	H	-2.9149	2.0483	-1.7759
H	2.2696	-3.1066	-0.8226	C	1.2597	0.4813	-0.0471
C	3.8756	-0.7627	-0.0761	C	3.1725	1.7758	0.3125
C	3.2305	1.5040	0.6124	H	4.2065	2.0726	0.4290
N	3.7373	2.7190	0.9479	C	0.6992	1.7200	0.1350
H	4.6668	2.7766	1.3361	H	-0.3345	2.0178	0.0563
H	3.0701	3.3718	1.3327				

tzA				tzC			
S	-0.6905	-1.9992	-0.6383	S	-3.5952	-2.1325	0.0367
O	1.9875	-0.8088	-0.5566	O	0.6271	2.1270	-0.2771
O	0.9286	2.5564	0.1555	O	3.5461	-0.4786	2.8868
H	-0.0540	2.4617	0.1831	O	4.1979	0.0344	-1.3843
O	3.3382	2.0346	-0.9067	O	2.1061	-1.2806	-2.2723
H	4.2992	2.0335	-0.9839	O	1.4858	-0.5128	0.8352
O	3.9708	-2.3653	0.7506	N	-1.6151	2.2230	0.0444
H	4.2044	-2.8742	1.5348	N	-3.8894	2.2966	0.3792
N	-2.3071	-1.8904	-0.3222	N	-3.9662	-0.5364	0.1934
N	-1.6865	1.6567	0.0933	N	-0.4585	0.1369	-0.2882
N	-4.0180	1.2166	0.4817	C	-0.4476	1.5364	-0.1712
N	-4.9394	-0.8983	0.3254	C	-2.7688	1.6093	0.1676
H	-4.8346	-1.8817	0.1284	C	-2.8500	0.1605	0.0768
H	-5.8491	-0.5353	0.5679	C	-1.6464	-0.5583	-0.1454
C	1.0283	0.2278	-0.6799	C	-1.9194	-1.9087	-0.1915
H	1.0811	0.6896	-1.6799	C	3.5391	0.3539	1.7385
C	1.4573	1.2813	0.3555	C	2.8831	-0.3756	0.5951
H	1.2099	0.8948	1.3573	C	2.9866	0.3715	-0.7348
C	2.9772	1.2315	0.1990	C	1.7585	-0.1400	-1.5105
H	3.4811	1.5769	1.1102	C	0.7929	-0.5889	-0.3993
C	3.2282	-0.2621	-0.0701	H	-1.2653	-2.7553	-0.3494
H	3.9785	-0.3744	-0.8631	H	2.9881	1.2849	1.9273
C	3.6899	-1.0342	1.1471	H	4.5634	0.6055	1.4301
H	4.5916	-0.5441	1.5408	H	3.3469	-1.3679	0.4804
H	2.9126	-1.0112	1.9238	H	2.8990	1.4483	-0.5773
C	-0.3346	-0.3415	-0.4757	H	1.3178	0.6300	-2.1505
C	-1.5169	0.3198	-0.1574	H	0.5279	-1.6286	-0.6121
C	-2.5870	-0.6141	-0.0899	H	-3.8609	3.3024	0.4550
C	-3.8877	-0.0847	0.2474	H	-4.7668	1.8109	0.4870
C	-2.9185	1.9999	0.3957	H	3.9773	-0.0020	3.6046
H	-3.0889	3.0532	0.6074	H	4.5178	0.7979	-1.8785
				H	3.0564	-1.1973	-2.4578

	tzG				tzU		
S	-0.3449	-2.3449	-0.6751	S	3.5037	-2.0394	-0.5191
O	-4.6717	-1.4212	0.2916	O	-0.7066	2.1755	0.4013
O	0.9766	2.3506	0.0089	O	3.8122	2.3563	0.3676
O	3.4008	2.0785	-0.9934	O	-3.4228	2.0009	-0.2341
O	2.2494	-0.9452	-0.5934	O	-3.2971	-2.2852	0.2432
O	3.5490	-1.1007	1.9333	O	-1.4785	-1.4768	2.0267
N	-3.1218	2.8841	0.5035	O	-1.6080	-0.0717	-1.2255
N	-3.8200	0.6863	0.3324	N	1.5465	2.1692	0.3697
N	-1.5657	1.2728	-0.0643	N	3.9138	-0.4817	-0.2070
N	-1.9634	-2.3331	-0.3686	N	0.3523	0.1704	0.0318
C	-2.7944	1.6035	0.2337	C	0.3253	1.5256	0.2682
C	-3.7051	-0.6846	0.1598	C	2.8225	1.6543	0.2499
C	-2.3347	-1.0786	-0.1706	C	2.8071	0.2161	-0.0326
C	-1.3307	-0.0703	-0.2572	C	1.5730	-0.4855	-0.1448
C	-0.1075	-0.6568	-0.5453	C	1.8185	-1.8116	-0.4240
C	1.2111	0.0054	-0.7536	C	-3.7311	1.0528	-1.2337
C	1.5691	1.1069	0.2583	C	-3.0273	-0.2738	-1.0277
C	3.0863	1.1919	0.0651	C	-3.1861	-0.8800	0.3676
C	3.4798	-0.2364	-0.3417	C	-1.8652	-0.5222	1.0643
C	4.2464	-1.0210	0.7033	C	-0.8968	-0.5649	-0.1207
H	1.2441	0.4446	-1.7653	H	1.1339	-2.6346	-0.5745
H	1.3489	0.7453	1.2726	H	-4.8133	0.8888	-1.2025
H	3.5905	1.5048	0.9877	H	-3.4777	1.4320	-2.2343
H	4.0691	-0.1861	-1.2668	H	-3.3827	-0.9938	-1.7749
H	5.1972	-0.5199	0.9056	H	-4.0532	-0.4653	0.8954
H	4.4656	-2.0234	0.3117	H	-1.9154	0.4833	1.4854
H	-4.0869	3.1776	0.4770	H	-0.6114	-1.6104	-0.2769
H	-2.4400	3.5911	0.2715	H	1.4716	3.1627	0.5605
H	-4.7425	1.0037	0.6076	H	-2.4546	2.0872	-0.1758
H	0.0019	2.1975	-0.0169	H	-2.9740	-2.6643	1.0751
H	2.7446	2.7916	-0.9603	H	-1.7046	-1.1546	2.9068
H	2.7264	-1.5778	1.7674				

# Appendix D

## Appendix to Chapter 5

## D1 Binding Energies

**Table D1:** Binding energies in (kcal/mol) of the modified base pairs computed with  $B3LYP/6-31++G(d,p)$  in the gas phase.

Basepair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$ +BSSE	ZPE	$\Delta E_{bind}$ +BSSE+ZPE	$\Delta E_{natural}$
<b>A-U</b>	-12.91	0.70	-12.22	1.40	-10.81	
<b>A-<sup>th</sup>U</b>	-12.41	0.70	-11.71	1.29	-10.43	+0.50
<sup>th</sup> <b>A-U</b>	-13.52	0.73	-12.80	0.93	-11.87	-0.61
<sup>th</sup> <b>A-<sup>th</sup>U</b>	-12.89	0.73	-12.16	0.83	-11.33	+0.02
<b>G-C</b>	-26.16	0.98	-25.19	1.73	-23.46	
<b>G-<sup>th</sup>C</b>	-26.75	0.98	-25.77	1.51	-24.26	-0.59
<sup>th</sup> <b>G-C</b>	-23.76	0.96	-22.81	1.54	-21.26	+2.40
<sup>th</sup> <b>G-<sup>th</sup>C</b>	-24.21	0.96	-23.25	1.34	-21.90	+1.95

**Table D2:** Binding energies in (kcal/mol) of the modified base pairs computed with dispersion  $B3LYP-D3/6-31++G(d,p)$  in the gas phase

Basepair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$ +BSSE	ZPE	$\Delta E_{bind}$ +BSSE+ZPE	$\Delta E_{natural}$
<b>A-U</b>	-17.13	0.73	-16.40	1.37	-15.02	
<b>A-<sup>th</sup>U</b>	-16.72	0.74	-15.99	1.30	-14.69	+0.41
<sup>th</sup> <b>A-U</b>	-17.78	0.76	-17.02	0.94	-16.08	-0.65
<sup>th</sup> <b>A-<sup>th</sup>U</b>	-17.25	0.77	-16.49	0.86	-15.63	-0.12
<b>G-C</b>	-31.14	1.01	-30.13	1.79	-28.35	
<b>G-<sup>th</sup>C</b>	-31.80	1.01	-30.79	1.54	-29.25	-0.66
<sup>th</sup> <b>G-C</b>	-28.73	1.00	-27.73	1.61	-26.12	+2.41
<sup>th</sup> <b>G-<sup>th</sup>C</b>	-29.25	1.00	-28.25	1.38	-26.87	+1.89

**Table D3:** Binding energies in (kcal/mol) of the modified base pairs computed with M06-2X/6-31++G(d,p) in the gas phase

Basepair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$	ZPE	$\Delta E_{bind}$	$\Delta E_{natural}$
			+BSSE		+BSSE+ZPE	
A-U	-15.16	0.79	-14.38	1.12	-13.25	
A- <sup>th</sup> U	-14.70	0.79	-13.91	0.96	-12.95	+0.46
<sup>th</sup> A-U	-15.77	0.81	-14.96	0.80	-14.16	-0.61
<sup>th</sup> A- <sup>th</sup> U	-15.17	0.81	-14.36	0.75	-13.61	-0.01
G-C	-28.55	1.06	-27.49	1.56	-25.93	
G- <sup>th</sup> C	-29.13	1.06	-28.07	1.26	-26.81	-0.58
<sup>th</sup> G-C	-26.06	1.03	-25.03	1.42	-23.61	+2.49
<sup>th</sup> G- <sup>th</sup> C	-26.49	1.03	-25.46	1.17	-24.29	+2.06

**Table D4:** Binding energies in (kcal/mol) of the modified base pairs computed with dispersion M062X-D3/6-31++G(d,p) in the gas phase

Basepair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$	ZPE	$\Delta E_{bind}$	$\Delta E_{natural}$
			+BSSE		+BSSE+ZPE	
A-U	-15.63	0.79	-14.85	1.12	-13.72	
A- <sup>th</sup> U	-15.45	0.79	-14.66	0.96	-13.70	+0.18
<sup>th</sup> A-U	-16.49	0.81	-15.68	0.80	-14.88	-0.86
<sup>th</sup> A- <sup>th</sup> U	-16.18	0.81	-15.37	0.73	-14.64	-0.55
G-C	-29.10	1.06	-28.04	1.56	-26.47	
G- <sup>th</sup> C	-29.72	1.06	-28.66	1.26	-27.40	-0.62
<sup>th</sup> G-C	-25.65	1.03	-24.62	1.42	-23.19	+3.45
<sup>th</sup> G- <sup>th</sup> C	-27.10	1.03	-26.07	1.17	-24.90	+2.00

**Table D5:** Binding energies in (kcal/mol) of the modified base pairs computed with  $\omega$ B97XD/6-31++G(d,p) in the gas phase

Basepair	$\Delta E_{bind}$	BSSE	$\Delta E_{bind}$	ZPE	$\Delta E_{bind}$	$\Delta E_{natural}$
			+BSSE		+BSSE+ZPE	
A-U	-16.36	0.71	-15.65	1.37	-14.28	
A- <sup>th</sup> U	-15.89	0.72	-15.17	1.27	-13.89	+0.47
<sup>th</sup> A-U	-17.09	0.74	-16.35	1.02	-15.33	-0.73
<sup>th</sup> A- <sup>th</sup> U	-16.47	0.75	-15.72	0.72	-14.75	-0.11
G-C	-30.47	0.98	-29.48	1.90	-27.58	
G- <sup>th</sup> C	-31.21	0.99	-30.22	1.66	-28.55	-0.74
<sup>th</sup> G-C	-27.89	0.97	-26.92	1.71	-25.21	+2.58
<sup>th</sup> G- <sup>th</sup> C	-28.46	0.97	-27.49	1.41	-26.07	+2.01

## D2 Intermolecular Distances

**Table D6:** *Intermolecular distances of the natural and modified nucleobases computed with B3LYP/6-31++G(d,p)*

Basepair	$d_{NHO}$	$d_{NHN}$	$d_{OHN}$	$\Delta d$		
<b>A-U</b>	2.956	2.885		(0.026)	(0.035)	
<b>A-<sup>th</sup>U</b>	2.979	2.887		0.023	0.002	
<sup>th</sup> <b>A-U</b>	2.919	2.884		-0.037	-0.001	
<sup>th</sup> <b>A-<sup>th</sup>U</b>	2.943	2.887		-0.013	0.002	
<b>G-C</b>	2.807	2.958	2.951	(-0.053)	(0.008)	(0.041)
<b>G-<sup>th</sup>C</b>	2.781	2.967	2.934	-0.026	0.009	-0.017
<sup>th</sup> <b>G-C</b>	2.847	2.966	2.976	0.040	0.008	0.025
<sup>th</sup> <b>G-<sup>th</sup>C</b>	2.821	2.976	2.959	0.014	0.018	0.008

All distances are in angstrom. The last three columns represent the differences in distances between the natural-natural base pairs and the modified pairs. Values in brackets are differences in computed and experimental distances for the natural-natural pairs.

**Table D7:** *Intermolecular distances of the natural and modified nucleobases computed with B3LYP-D3/6-31++G(d,p)*

Basepair	$d_{NHO}$	$d_{NHN}$	$d_{OHN}$	$\Delta d$		
<b>A-U</b>	2.943	2.885		(0.013)	(-0.030)	
<b>A-<sup>th</sup>U</b>	2.963	2.887		0.025	0.002	
<sup>th</sup> <b>A-U</b>	2.906	2.884		-0.037	-0.001	
<sup>th</sup> <b>A-<sup>th</sup>U</b>	2.928	2.821		0.015	0.001	
<b>G-C</b>	2.780	2.925	2.919	(-0.080)	(-0.025)	(0.009)
<b>G-<sup>th</sup>C</b>	2.755	2.934	2.901	-0.025	0.009	-0.018
<sup>th</sup> <b>G-C</b>	2.815	2.930	2.942	0.035	0.005	-0.023
<sup>th</sup> <b>G-<sup>th</sup>C</b>	2.791	2.939	2.924	0.011	0.014	0.005

All distances are in angstrom. The last three columns represent the differences in distances between the natural-natural base pairs and the modified pairs. Values in brackets are differences in computed and experimental distances for the natural-natural pairs.

**Table D8:** Intermolecular distances of the natural and modified nucleobases computed with  $M062X/6-31++G(d,p)$

Basepair	$d_{NHO}$	$d_{NHN}$	$d_{OHN}$	$\Delta d$		
<b>A-U</b>	2.973	2.812		(0.043)	(-0.038)	
<b>A-<sup>th</sup>U</b>	2.996	2.816		0.023	0.004	
<sup>th</sup> <b>A-U</b>	2.939	2.811		-0.034	-0.001	
<sup>th</sup> <b>A-<sup>th</sup>U</b>	2.962	2.814		-0.011	0.002	
<b>G-C</b>	2.812	2.952	2.940	(-0.048)	(0.002)	(0.030)
<b>G-<sup>th</sup>C</b>	2.779	2.957	2.928	-0.033	0.005	-0.012
<sup>th</sup> <b>G-C</b>	2.854	2.958	2.964	0.042	0.006	0.024
<sup>th</sup> <b>G-<sup>th</sup>C</b>	2.827	2.966	2.951	0.015	0.014	0.011

All distances are in angstrom. The last three columns represent the differences in distances between the natural-natural base pairs and the modified pairs. Values in brackets are differences in computed and experimental distances for the natural-natural pairs.

**Table D9:** Intermolecular distances of the natural and modified base pairs computed with  $M062X-D3/6-31++G(d,p)$

Basepair	$d_{NHO}$	$d_{NHN}$	$d_{OHN}$	$\Delta d$		
<b>A-U</b>	2.973	2.810		(0.043)	(-0.040)	
<b>A-<sup>th</sup>U</b>	2.994	2.813		0.021	0.003	
<sup>th</sup> <b>A-U</b>	2.938	2.809		-0.035	-0.001	
<sup>th</sup> <b>A-<sup>th</sup>U</b>	2.961	2.812		-0.012	0.002	
<b>G-C</b>	2.811	2.951	2.939	(-0.049)	(0.001)	(0.029)
<b>G-<sup>th</sup>C</b>	2.777	2.955	2.927	-0.034	0.004	-0.012
<sup>th</sup> <b>G-C</b>	2.852	2.957	2.963	0.041	0.006	0.024
<sup>th</sup> <b>G-<sup>th</sup>C</b>	2.825	2.964	2.949	0.014	0.013	0.010

All distances are in angstrom. The last three columns represent the differences in distances between the natural-natural base pairs and the modified pairs. Values in brackets are differences in computed and experimental distances for the natural-natural pairs.

## D3 Excitation Properties of the Modified Nucleobase Pairs

**Table D10:** Vertical excitation energies, oscillator strengths and dominant transitions of low-lying states of the modified base pairs computed with B3LYP/6-31++G(d,p) in gas phase. H=HOMO, L=LUMO

Basepair	S <sub>n</sub>	Transition	Orbitals	ΔE (eV)	ΔE (nm)	f
<b>A</b> - <sup>th</sup> <b>U</b>	1	78→79	H→L	3.78	328	0.002
	2	77→79	H-1→L	4.13	300	0.083
	3	73→79	H-5→L	4.51	275	0.000
	4	76→79	H-2→L	4.80	258	0.000
	5	78→80	H→L+1	4.95	251	0.290
	6	75→80	H-3→L	5.10	243	0.111
<sup>th</sup> <b>A</b> - <b>U</b>	1	68→69	H→L	3.73	332	0.105
	2	68→70	H→L+1	3.98	312	0.015
	3	67→69	H-1→L	4.07	305	0.001
	4	63→70	H-5→L+1	4.72	263	0.000
	5	66→69	H-2→L	4.78	260	0.065
	6	65→69	H-3→L	4.84	256	0.006
<sup>th</sup> <b>A</b> - <sup>th</sup> <b>U</b>	1	82→83	H→L	3.57	348	0.008
	2	82→84	H→L+1	3.77	329	0.116
	3	80→84	H-2→L+1	4.06	305	0.001
	4	81→83	H-1→L	4.10	302	0.076
	5	81→84	H-1→L+1	4.39	283	0.006
	6	80→83	H-2→L	4.46	278	0.000
<b>G</b> - <sup>th</sup> <b>C</b>	1	82→83	H→L	2.80	442	0.005
	2	81→83	H-1→L	3.92	316	0.105
	3	80→83	H-2→L	4.26	291	0.000
	4	82→84	H→L+1	4.50	275	0.000
	5	79→83	H-3→L	4.50	275	0.023
	6	78→83	H-4→L	4.57	271	0.000
<sup>th</sup> <b>G</b> - <b>C</b>	1	72→73	H→L	3.20	388	0.001
	2	72→74	H→L+1	3.59	345	0.071
	3	72→75	H→L+2	4.19	296	0.000
	4	72→76	H→L+3	4.44	279	0.004
	5	71→73	H-1→L	4.55	272	0.015
	6	72→83	H→L+4	4.66	266	0.001
<sup>th</sup> <b>G</b> - <sup>th</sup> <b>C</b>	1	86→87	H→L	2.64	470	0.001
	2	86→88	H→L+1	3.56	345	0.075
	3	84→87	H-2→L	3.89	296	0.107
	4	85→87	H-1→L	4.00	279	0.001
	5	86→89	H→L+2	4.31	272	0.000
	6	83→87	H-3→L	4.37	266	0.001

**Table D11:** Vertical excitation energies, oscillator strengths and dominant transitions of low-lying states of the modified base pairs computed with B3LYP/6-31++G(d,p) in water. H=HOMO, L=LUMO

Basepair	$S_n$	Transition	Orbitals	$\Delta E$ (eV)	$\Delta E$ (nm)	$f$
<b>A-<sup>th</sup>U</b>	1	77→79	H-1→L	4.00	310	0.092
	2	78→79	H→L	4.04	307	0.002
	3	73→79	H-5→L	4.73	262	0.000
	4	78→80	H→L+1	4.98	252	0.448
	5	76→79	H-2→L	4.99	249	0.050
	6	77→80	H-1→L+1	5.02	247	0.057
<b><sup>th</sup>A-U</b>	1	68→69	H→L	3.76	329	0.149
	2	68→70	H→L+1	4.30	288	0.006
	3	65→69	H-3→L	4.34	286	0.001
	4	67→69	H-1→L	4.59	270	0.018
	5	66→69	H-2→L	4.73	262	0.084
	6	63→70	H-5→L+1	5.02	247	0.000
<b><sup>th</sup>A-<sup>th</sup>U</b>	1	82→83	H→L	3.75	331	0.125
	2	82→84	H→L+1	3.88	320	0.037
	3	81→83	H-1→L	3.98	311	0.068
	4	81→84	H-1→L+1	4.13	300	0.023
	5	78→83	H-4→L	4.34	286	0.001
	6	76→84	H-6→L+1	4.72	263	0.000
<b>G-<sup>th</sup>C</b>	1	82→83	H→L	3.57	347	0.015
	2	81→83	H-1→L	3.85	322	0.107
	3	82→84	H→L+1	4.68	265	0.101
	4	80→83	H-2→L	4.77	260	0.302
	5	78→83	H-4→L	4.93	251	0.000
	6	79→83	H-3→L	4.96	250	0.089
<b><sup>th</sup>G-C</b>	1	72→73	H→L	3.59	346	0.0874
	2	72→74	H→L+1	4.11	302	0.0072
	3	71→73	H-1→L	4.68	265	0.0397
	4	68→73	H-4→L	4.87	255	0.0000
	5	70→73	H-2→L	4.87	255	0.0356
	6	71→74	H-1→L+1	4.95	250	0.0820
<b><sup>th</sup>G-<sup>th</sup>C</b>	1	86→87	H→L	3.43	361	0.0226
	2	86→88	H→L+1	3.62	342	0.0772
	3	85→87	H-1→L	3.84	323	0.1079
	4	85→88	H-1→L+1	4.28	290	0.0030
	5	84→87	H-2→L	4.72	263	0.1053
	6	83→87	H-3→L	4.77	260	0.1683

## D4 Coordinates

### D4.1 Base Pairs

#### B3LYP

All geometries were optimized at the B3LYP/6-31++g(d,p) level of theory in the gas phase.

	$\mathbf{A}^{th}\mathbf{U}$ (Cs)				$\mathbf{A}^{th}\mathbf{U}$ (C1)		
C	1.7602330	1.5411770	0.0000000	C	1.7601650	1.5410670	0.0000030
C	1.8494780	-0.9486130	0.0000000	C	1.8495210	-0.9487300	0.0000020
C	3.3128180	-0.8430520	0.0000000	C	3.3128600	-0.8430960	0.0000010
C	3.9450720	0.4439280	0.0000000	C	3.9450410	0.4439120	0.0000000
C	4.2181660	-1.8765490	0.0000000	C	4.2182790	-1.8765350	-0.0000060
C	5.3150240	0.3692360	0.0000000	C	5.3149990	0.3693130	-0.0000050
C	-2.3478540	1.3496730	0.0000000	C	-2.3478480	1.3497090	0.0000010
C	-2.3975990	-0.9859480	0.0000000	C	-2.3976070	-0.9859130	0.0000000
C	-3.8088360	-0.8903490	0.0000000	C	-3.8088430	-0.8903070	0.0000000
C	-4.3448880	0.4007300	0.0000000	C	-4.3448900	0.4007760	0.0000000
C	-5.9199070	-1.1556670	0.0000000	C	-5.9199130	-1.1556160	0.0000090
H	0.1457770	0.2460270	0.0000000	H	0.1457630	0.2458710	0.0000010
H	3.5451670	2.5065370	0.0000000	H	3.5450520	2.5065040	-0.0000040
H	3.9991960	-2.9348560	0.0000000	H	3.9993720	-2.9348550	-0.0000100
H	6.0391280	1.1713250	0.0000000	H	6.0390490	1.1714500	-0.0000080
H	-0.7239010	-2.1889500	0.0000000	H	-0.7239030	-2.1888930	-0.0000010
H	-1.7092490	2.2292040	0.0000000	H	-1.7092380	2.2292370	0.0000010
H	-2.2750550	-3.0176310	0.0000000	H	-2.2750310	-3.0176100	-0.0000030
H	-6.4076370	0.9384240	0.0000000	H	-6.4076390	0.9384720	-0.0000080
H	-6.9166970	-1.5750070	0.0000000	H	-6.9167040	-1.5749520	0.0000130
N	1.1920660	0.2686660	0.0000000	N	1.1920530	0.2685280	0.0000020
N	3.1442820	1.5792130	0.0000000	N	3.1442100	1.5791610	-0.0000010
N	-1.6930970	0.1700420	0.0000000	N	-1.6930980	0.1700730	0.0000010
N	-1.7433680	-2.1613650	0.0000000	N	-1.7433700	-2.1613280	-0.0000010
N	-3.6648760	1.5575980	0.0000000	N	-3.6648680	1.5576370	0.0000000
N	-4.8068620	-1.8505070	0.0000000	N	-4.8068710	-1.8504620	-0.0000050
N	-5.7088550	0.2094980	0.0000000	N	-5.7088560	0.2095480	-0.0000040
O	1.0824840	2.5596560	0.0000000	O	1.0823640	2.5595090	-0.0000010
O	1.2318940	-2.0164180	0.0000000	O	1.2319830	-2.0165570	0.0000000
S	5.8339540	-1.2917670	0.0000000	S	5.8340280	-1.2916590	0.0000200

<b>A-<sup>th</sup>U</b>			<b>G-C</b>				
		(Cs)			(Cs)		
C	1.7602330	1.5411770	0.0000000	C	2.7795160	-1.2810330	0.0000000
C	1.8494780	-0.9486130	0.0000000	C	2.9495840	1.0628740	0.0000000
C	3.3128180	-0.8430520	0.0000000	C	4.2158840	-1.4370940	0.0000000
C	3.9450720	0.4439280	0.0000000	C	4.9599590	-0.3015320	0.0000000
C	4.2181660	-1.8765490	0.0000000	C	-1.3005020	1.5250560	0.0000000
C	5.3150240	0.3692360	0.0000000	C	-1.4392140	-0.9508310	0.0000000
C	-2.3478540	1.3496730	0.0000000	C	-2.8538490	-0.7394480	0.0000000
C	-2.3975990	-0.9859480	0.0000000	C	-3.3169450	0.5820680	0.0000000
C	-3.8088360	-0.8903490	0.0000000	C	-4.9799720	-0.8832110	0.0000000
C	-4.3448880	0.4007300	0.0000000	H	0.2934310	0.1850780	0.0000000
C	-5.9199070	-1.1556670	0.0000000	H	0.5659440	2.4689920	0.0000000
H	0.1457770	0.2460270	0.0000000	H	0.9475600	-2.2398090	0.0000000
H	3.5451670	2.5065370	0.0000000	H	2.3700800	-3.2804570	0.0000000
H	3.9991960	-2.9348560	0.0000000	H	4.6815470	-2.4142560	0.0000000
H	6.0391280	1.1713250	0.0000000	H	4.8842230	1.7766230	0.0000000
H	-0.7239010	-2.1889500	0.0000000	H	6.0443350	-0.3033910	0.0000000
H	-1.7092490	2.2292040	0.0000000	H	-0.8601480	3.4954510	0.0000000
H	-2.2750550	-3.0176310	0.0000000	H	-5.3341220	1.2457300	0.0000000
H	-6.4076370	0.9384240	0.0000000	H	-6.0005540	-1.2403090	0.0000000
H	-6.9166970	-1.5750070	0.0000000	N	1.9778260	-2.3523190	0.0000000
N	1.1920660	0.2686660	0.0000000	N	2.2025940	-0.0714840	0.0000000
N	3.1442820	1.5792130	0.0000000	N	4.3514510	0.9171630	0.0000000
N	-1.6930970	0.1700420	0.0000000	N	-0.4503560	2.5755600	0.0000000
N	-1.7433680	-2.1613650	0.0000000	N	-0.7360500	0.2703410	0.0000000
N	-3.6648760	1.5575980	0.0000000	N	-2.6104550	1.7316440	0.0000000
N	-4.8068620	-1.8505070	0.0000000	N	-3.9106340	-1.6351770	0.0000000
N	-5.7088550	0.2094980	0.0000000	N	-4.6845020	0.4730620	0.0000000
O	1.0824840	2.5596560	0.0000000	O	2.4776440	2.2075970	0.0000000
O	1.2318940	-2.0164180	0.0000000	O	-0.8099540	-2.0222020	0.0000000
S	5.8339540	-1.2917670	0.0000000				

	$G^{th}C$ (Cs)			$G^{th}C$ (C1)			
C	0.4786610	2.7847710	0.0000000	C	2.0211290	1.6586960	-0.0000090
C	2.1040550	0.9132060	0.0000000	C	2.0591220	-0.7056890	0.0000090
C	2.6504080	3.2739720	0.0000000	C	3.5222650	-0.7161720	0.0000040
C	3.0931020	1.9453150	0.0000000	C	4.1907080	0.5538170	-0.0000090
C	4.8623560	3.1327350	0.0000000	C	4.4099200	-1.7702660	0.0000080
C	-0.9995830	-1.9335530	0.0000000	C	5.5599620	0.4453430	-0.0000150
C	-1.9616370	-4.3281340	0.0000000	C	-2.1514290	-0.7951600	0.0000100
C	-2.0492980	-2.9528630	0.0000000	C	-2.2713580	1.6807700	0.0000140
C	-2.6070700	-0.1992640	0.0000000	C	-3.5796030	-0.7334570	0.0000010
C	-3.4103540	-2.4976550	0.0000000	C	-4.1785220	0.5324940	-0.0000030
C	-4.3244750	-3.5228470	0.0000000	C	-5.6787900	-1.0990310	-0.0000150
H	0.0385830	0.7493290	0.0000000	H	0.3262070	-1.8287560	0.0000200
H	0.5649570	-3.2469520	0.0000000	H	1.8370780	-2.7363290	0.0000230
H	1.0288610	-1.5466020	0.0000000	H	3.8288030	2.6175070	-0.0000260
H	3.8395740	5.0332600	0.0000000	H	4.2023510	-2.8313600	0.0000180
H	5.8880050	3.4749960	0.0000000	H	6.3021350	1.2304500	-0.0000250
H	-1.0771850	4.0711520	0.0000000	H	-0.5126170	2.8089570	0.0000080
H	-1.0779580	-4.9511240	0.0000000	H	-0.5460030	0.5146350	0.0000180
H	-1.5719720	2.3836700	0.0000000	H	-2.0369030	3.6858750	0.0000020
H	-4.5759310	-0.7566280	0.0000000	H	-6.2540870	0.9811370	-0.0000190
H	-5.4035060	-3.4689020	0.0000000	H	-6.6563900	-1.5609790	-0.0000220
N	0.2901990	-2.2781430	0.0000000	N	1.3656570	-1.8464750	0.0000200
N	0.8007090	1.4464720	0.0000000	N	1.3878880	0.4406110	0.0000040
N	1.3882550	3.7499480	0.0000000	N	3.4136010	1.6960720	-0.0000160
N	3.7985990	4.0246010	0.0000000	N	-1.5348520	2.8128990	0.0000270
N	4.4771070	1.8836010	0.0000000	N	-1.5786210	0.4912890	0.0000160
N	-0.8362250	3.0933570	0.0000000	N	-3.5958400	1.7490750	0.0000030
N	-1.3073390	-0.6413270	0.0000000	N	-4.5367480	-1.7350250	-0.0000070
N	-3.6388280	-1.1351400	0.0000000	N	-5.5270220	0.2808170	-0.0000140
O	2.2646070	-0.3199290	0.0000000	O	1.3975130	2.7284120	-0.0000170
O	-2.8962540	1.0047130	0.0000000	O	-1.4147410	-1.7970030	0.0000130
S	-3.5170220	-5.0594280	0.0000000	S	6.0391970	-1.2230130	0.0000020

	${}^{th}\mathbf{A}-{}^{th}\mathbf{U}$ (C1)			${}^{th}\mathbf{A}-{}^{th}\mathbf{U}$ (Cs)			
C	2.1075450	1.5600560	0.0000100	C	1.9970980	1.8020170	0.0000000
C	2.1286510	-0.9303350	0.0000030	C	2.0189020	-0.6883890	0.0000000
C	3.5937310	-0.8661800	0.0000000	C	3.4839680	-0.6238390	0.0000000
C	4.2613660	0.4028300	0.0000010	C	4.1512290	0.6453680	0.0000000
C	4.4701640	-1.9245510	0.0000000	C	4.3607230	-1.6819420	0.0000000
C	5.6287230	0.2899750	-0.0000020	C	5.5186220	0.5329300	0.0000000
C	-2.0352160	1.4616920	0.0000020	C	-2.1456770	1.7024070	0.0000000
C	-2.1014830	-0.8646890	0.0000000	C	-2.2112720	-0.6239830	0.0000000
C	-3.5462530	-0.7793980	-0.0000040	C	-3.6560700	-0.5390950	0.0000000
C	-4.1052540	0.5538550	-0.0000050	C	-4.2154680	0.7939980	0.0000000
C	-4.5113550	-1.7685300	-0.0000050	C	-4.6208800	-1.5285170	0.0000000
C	-5.4895640	0.5343370	-0.0000070	C	-5.5997690	0.7740560	0.0000000
H	0.4569110	0.3074180	0.0000080	H	0.3468430	0.5489050	0.0000000
H	3.9185080	2.4755040	0.0000060	H	3.8078390	2.7179400	0.0000000
H	4.2222260	-2.9764140	0.0000020	H	4.1131150	-2.7338840	0.0000000
H	6.3748840	1.0715750	-0.0000010	H	6.2645420	1.3147590	0.0000000
H	-0.4379500	-2.0798630	0.0000070	H	-0.5474590	-1.8387850	0.0000000
H	-1.3726500	2.3241220	0.0000050	H	-1.4833590	2.5650290	0.0000000
H	-1.9742920	-2.9145990	0.0000010	H	-2.0836650	-2.6738230	0.0000000
H	-4.3980240	-2.8439020	-0.0000040	H	-4.5072040	-2.6038580	0.0000000
H	-6.1600530	1.3813470	-0.0000060	H	-6.2705260	1.6208570	0.0000000
N	1.5042480	0.3024700	0.0000070	N	1.3941700	0.5442620	0.0000000
N	3.4922860	1.5594810	0.0000060	N	3.3818450	1.8018110	0.0000000
N	-1.3821590	0.2546330	0.0000030	N	-1.4922670	0.4955310	0.0000000
N	-1.4601000	-2.0499500	0.0000030	N	-1.5696120	-1.8090940	0.0000000
N	-3.3235490	1.6861560	-0.0000020	N	-3.4340710	1.9265190	0.0000000
O	1.4580010	2.5957740	0.0000090	O	1.3472580	2.8375470	0.0000000
O	1.4824390	-1.9827770	0.0000020	O	1.3729630	-1.7409960	0.0000000
S	6.1012600	-1.3847190	-0.0000200	S	5.9916570	-1.1416250	0.0000000
S	-6.0877780	-1.0798270	-0.0000180	S	-6.1975040	-0.8402890	0.0000000

	$^{th}$ <b>A-U</b> (Cs)			$^{th}$ <b>A-U</b> (C1)			
C	0.8210930	1.6388190	0.0000000	C	0.8389480	-1.3784240	-0.0000030
C	1.1519600	-0.6650940	0.0000000	C	1.1708400	0.9253630	-0.0000020
C	2.5777460	-0.4144870	0.0000000	C	2.5965230	0.6741050	-0.0000010
C	2.9811200	0.9737570	0.0000000	C	2.9992510	-0.7143240	-0.0000020
C	3.6492740	-1.2870360	0.0000000	C	3.6684680	1.5461440	-0.0000010
C	4.3585520	1.1122890	0.0000000	C	4.3766220	-0.8535090	-0.0000020
C	-3.0388940	-1.2423320	0.0000000	C	-3.0199470	1.5033680	0.0000010
C	-3.3152840	1.2212280	0.0000000	C	-3.2963190	-0.9602060	-0.0000030
C	-4.4885770	-1.3734130	0.0000000	C	-4.4696340	1.6344500	0.0000030
C	-5.2548470	-0.2580500	0.0000000	C	-5.2358810	0.5190630	0.0000030
H	0.0652610	2.4209830	0.0000000	H	0.0827360	-2.1602210	-0.0000040
H	1.2607000	-2.7145800	0.0000000	H	1.2804400	2.9747910	-0.0000010
H	3.6590970	-2.3683220	0.0000000	H	3.6787870	2.6274270	-0.0000010
H	4.9280720	2.0302370	0.0000000	H	4.9457080	-1.7717250	-0.0000030
H	-0.3631100	-2.0625330	0.0000000	H	-0.3436620	2.3234030	-0.0000020
H	-1.5189000	0.1957390	0.0000000	H	-1.4999280	0.0653250	-0.0000030
H	-4.9177240	-2.3655510	0.0000000	H	-4.8988090	2.6265720	0.0000050
H	-5.2654130	1.8287420	0.0000000	H	-5.2464590	-1.5677210	0.0000010
H	-6.3389110	-0.2880610	0.0000000	H	-6.3199460	0.5490550	0.0000050
N	0.3094640	0.3658220	0.0000000	N	0.3279200	-0.1051900	-0.0000030
N	0.6503500	-1.9147940	0.0000000	N	0.6697380	2.1752780	-0.0000020
N	2.0751930	2.0095500	0.0000000	N	2.0928770	-1.7497210	-0.0000030
N	-2.5590160	0.0661410	0.0000000	N	-2.5400790	0.1948850	-0.0000020
N	-4.6945460	0.9948520	0.0000000	N	-4.6755830	-0.7338370	0.0000000
O	-2.2549510	-2.1991630	0.0000000	O	-2.2359430	2.4601560	0.0000010
O	-2.8480050	2.3500750	0.0000000	O	-2.8290490	-2.0890600	-0.0000040
S	5.1370380	-0.4232220	0.0000000	S	5.1558230	0.6816380	0.0000040

<i>th</i> <b>A-U</b>			<i>th</i> <b>G-C</b>				
(Cs)			(Cs)				
C	0.8210930	1.6388190	0.0000000	C	0.4699810	2.7956230	0.0000000
C	1.1519600	-0.6650940	0.0000000	C	2.0511440	0.9188750	0.0000000
C	2.5777460	-0.4144870	0.0000000	C	2.6836230	3.3354400	0.0000000
C	2.9811200	0.9737570	0.0000000	C	3.0819750	1.9444800	0.0000000
C	3.6492740	-1.2870360	0.0000000	C	3.7821440	4.1728070	0.0000000
C	4.3585520	1.1122890	0.0000000	C	4.4453860	1.7595250	0.0000000
C	-3.0388940	-1.2423320	0.0000000	C	-1.0558100	-1.9609670	0.0000000
C	-3.3152840	1.2212280	0.0000000	C	-2.0935630	-2.9651680	0.0000000
C	-4.4885770	-1.3734130	0.0000000	C	-2.6251120	-0.2125930	0.0000000
C	-5.2548470	-0.2580500	0.0000000	C	-3.3778180	-2.5235930	0.0000000
H	0.0652610	2.4209830	0.0000000	H	0.0037240	0.7409770	0.0000000
H	1.2607000	-2.7145800	0.0000000	H	0.4945840	-3.2922350	0.0000000
H	3.6590970	-2.3683220	0.0000000	H	0.9775630	-1.6004440	0.0000000
H	4.9280720	2.0302370	0.0000000	H	3.7894460	5.2529660	0.0000000
H	-0.3631100	-2.0625330	0.0000000	H	4.9868930	0.8243480	0.0000000
H	-1.5189000	0.1957390	0.0000000	H	-1.1006900	4.0646810	0.0000000
H	-4.9177240	-2.3655510	0.0000000	H	-1.5843750	2.3777470	0.0000000
H	-5.2654130	1.8287420	0.0000000	H	-1.8626340	-4.0226660	0.0000000
H	-6.3389110	-0.2880610	0.0000000	H	-4.2334960	-3.1897190	0.0000000
N	0.3094640	0.3658220	0.0000000	H	-4.5910490	-0.8354400	0.0000000
N	0.6503500	-1.9147940	0.0000000	N	0.2358820	-2.3186200	0.0000000
N	2.0751930	2.0095500	0.0000000	N	0.7708960	1.4320770	0.0000000
N	-2.5590160	0.0661410	0.0000000	N	1.3674970	3.7441690	0.0000000
N	-4.6945460	0.9948520	0.0000000	N	-0.8529360	3.0884900	0.0000000
O	-2.2549510	-2.1991630	0.0000000	N	-1.3383130	-0.6528340	0.0000000
O	-2.8480050	2.3500750	0.0000000	N	-3.6432980	-1.1883250	0.0000000
S	5.1370380	-0.4232220	0.0000000	O	2.2515060	-0.3086040	0.0000000
				O	-2.9523990	0.9797240	0.0000000
				S	5.2632740	3.2713250	0.0000000

	$^{th}\mathbf{G-C}$ (C1)			$^{th}\mathbf{G-C}$ (Cs)			
C	2.0433900	1.6615420	-0.0000210	C	0.4699810	2.7956230	0.0000000
C	2.1186710	-0.6866080	0.0000210	C	2.0511440	0.9188750	0.0000000
C	3.5627300	-0.6944870	0.0000000	C	2.6836230	3.3354400	0.0000000
C	4.1851530	0.5125320	-0.0000320	C	3.0819750	1.9444800	0.0000000
C	-2.1170890	-0.7544200	0.0000270	C	3.7821440	4.1728070	0.0000000
C	-2.2726410	1.6946650	0.0000350	C	4.4453860	1.7595250	0.0000000
C	-3.5709460	-0.7263460	-0.0000010	C	-1.0558100	-1.9609670	0.0000000
C	-4.2450280	0.5539190	-0.0000090	C	-2.0935630	-2.9651680	0.0000000
C	-4.4279370	-1.8027580	-0.0000210	C	-2.6251120	-0.2125930	0.0000000
C	-5.6176090	0.3991860	-0.0000370	C	-3.3778180	-2.5235930	0.0000000
H	0.4004730	-1.8321720	0.0000580	H	0.0037240	0.7409770	0.0000000
H	1.9192000	-2.7203830	0.0000570	H	0.4945840	-3.2922350	0.0000000
H	3.8942690	2.5709640	-0.0000650	H	0.9775630	-1.6004440	0.0000000
H	4.1270940	-1.6181340	0.0000090	H	3.7894460	5.2529660	0.0000000
H	5.2639030	0.6229970	-0.0000500	H	4.9868930	0.8243480	0.0000000
H	-0.4996300	2.8133030	0.0000380	H	-1.1006900	4.0646810	0.0000000
H	-0.5150430	0.5328150	0.0000570	H	-1.5843750	2.3777470	0.0000000
H	-2.0154940	3.6975320	0.0000290	H	-1.8626340	-4.0226660	0.0000000
H	-4.1724300	-2.8527590	-0.0000200	H	-4.2334960	-3.1897190	0.0000000
H	-6.3698160	1.1744180	-0.0000470	H	-4.5910490	-0.8354400	0.0000000
N	1.4182010	0.4537240	0.0000130	N	0.2358820	-2.3186200	0.0000000
N	1.4328630	-1.8381530	0.0000530	N	0.7708960	1.4320770	0.0000000
N	3.4536170	1.6606970	-0.0000430	N	1.3674970	3.7441690	0.0000000
N	-1.5194830	2.8210030	0.0000750	N	-0.8529360	3.0884900	0.0000000
N	-1.5471290	0.5015860	0.0000430	N	-1.3383130	-0.6528340	0.0000000
N	-3.5769010	1.7592630	0.0000080	N	-3.6432980	-1.1883250	0.0000000
O	1.4553800	2.7491890	-0.0000290	O	2.2515060	-0.3086040	0.0000000
O	-1.4130500	-1.7796870	0.0000330	O	-2.9523990	0.9797240	0.0000000
S	-6.0641710	-1.2762230	-0.0000570	S	5.2632740	3.2713250	0.0000000

	${}^{th}\mathbf{G}-{}^{th}\mathbf{C}$ (Cs)			${}^{th}\mathbf{G}-{}^{th}\mathbf{C}$ (C1)			
C	2.0127630	-0.9372610	0.0000000	C	2.0127480	-0.9372670	-0.0116330
C	2.1464930	1.5126230	0.0000000	C	2.1464490	1.5126180	-0.0148080
C	3.4656820	-0.8970670	0.0000000	C	3.4656810	-0.8970620	-0.0129140
C	4.1287110	0.3888820	0.0000000	C	4.1286840	0.3889150	-0.0150820
C	4.3319160	-1.9663040	0.0000000	C	4.3319500	-1.9662650	-0.0123120
C	5.5025980	0.2458180	0.0000000	C	5.5025730	0.2458840	-0.0160650
C	-2.1877690	1.4339870	0.0000000	C	-2.1876850	1.4339220	-0.0109010
C	-2.2043170	-0.9301960	0.0000000	C	-2.2043170	-0.9302520	-0.0079620
C	-3.6666720	-0.9559690	0.0000000	C	-3.6666710	-0.9559520	-0.0066440
C	-4.3470890	0.3076790	0.0000000	C	-4.3470550	0.3077070	-0.0076100
C	-4.5442370	-2.0187370	0.0000000	C	-4.5442540	-2.0187120	-0.0045630
C	-5.7152760	0.1858210	0.0000000	C	-5.7152430	0.1858670	-0.0062610
H	0.3626960	2.6102020	0.0000000	H	0.3626510	2.6102060	-0.0146240
H	0.3989830	0.3352820	0.0000000	H	0.3989580	0.3352480	-0.0118390
H	1.8692680	3.5126610	0.0000000	H	1.8692510	3.5126420	-0.0170440
H	4.0856030	-3.0184850	0.0000000	H	4.0856580	-3.0184500	-0.0107890
H	6.2482070	1.0273850	0.0000000	H	6.2481620	1.0274720	-0.0176730
H	-0.4659510	-2.0388470	0.0000000	H	-0.4660060	-2.0389810	-0.0081120
H	-1.9635250	-2.9614980	0.0000000	H	-1.9636700	-2.9615300	-0.0056390
H	-4.0048210	2.3744540	0.0000000	H	-4.0046870	2.3744810	-0.0104660
H	-4.3269240	-3.0778760	0.0000000	H	-4.3269500	-3.0778570	-0.0034460
H	-6.4650510	0.9636650	0.0000000	H	-6.4650170	0.9637130	-0.0065660
N	1.3826350	2.6308900	0.0000000	N	1.3825920	2.6308880	-0.0155720
N	1.4310740	0.3125050	0.0000000	N	1.4310540	0.3124970	-0.0126990
N	3.4501210	1.5880980	0.0000000	N	3.4500730	1.5881170	-0.0160060
N	-1.5006060	-2.0674690	0.0000000	N	-1.5006610	-2.0675540	-0.0071740
N	-1.5429640	0.2194590	0.0000000	N	-1.5429230	0.2193660	-0.0099690
N	-3.5808530	1.4569320	0.0000000	N	-3.5807740	1.4569320	-0.0097020
O	1.3185170	-1.9702670	0.0000000	O	1.3185170	-1.9702820	-0.0097550
O	-1.5734720	2.5070580	0.0000000	O	-1.5733590	2.5069720	-0.0127650
S	5.9632610	-1.4256310	0.0000000	S	5.9632770	-1.4255530	-0.0143670
S	-6.1784100	-1.4869390	0.0000000	S	-6.1784150	-1.4868770	-0.0037890

	${}^{th}\mathbf{G}$	${}^{th}\mathbf{C}$	(Cs)
C	2.0127630	-0.9372610	0.0000000
C	2.1464930	1.5126230	0.0000000
C	3.4656820	-0.8970670	0.0000000
C	4.1287110	0.3888820	0.0000000
C	4.3319160	-1.9663040	0.0000000
C	5.5025980	0.2458180	0.0000000
C	-2.1877690	1.4339870	0.0000000
C	-2.2043170	-0.9301960	0.0000000
C	-3.6666720	-0.9559690	0.0000000
C	-4.3470890	0.3076790	0.0000000
C	-4.5442370	-2.0187370	0.0000000
C	-5.7152760	0.1858210	0.0000000
H	0.3626960	2.6102020	0.0000000
H	0.3989830	0.3352820	0.0000000
H	1.8692680	3.5126610	0.0000000
H	4.0856030	-3.0184850	0.0000000
H	6.2482070	1.0273850	0.0000000
H	-0.4659510	-2.0388470	0.0000000
H	-1.9635250	-2.9614980	0.0000000
H	-4.0048210	2.3744540	0.0000000
H	-4.3269240	-3.0778760	0.0000000
H	-6.4650510	0.9636650	0.0000000
N	1.3826350	2.6308900	0.0000000
N	1.4310740	0.3125050	0.0000000
N	3.4501210	1.5880980	0.0000000
N	-1.5006060	-2.0674690	0.0000000
N	-1.5429640	0.2194590	0.0000000
N	-3.5808530	1.4569320	0.0000000
O	1.3185170	-1.9702670	0.0000000
O	-1.5734720	2.5070580	0.0000000
S	5.9632610	-1.4256310	0.0000000
S	-6.1784100	-1.4869390	0.0000000

### B3LYP-D3

All geometries were optimized at the B3LYP-D3/6-31++g(d,p) level of theory in the gas phase.

A- <sup>th</sup> U			A-U				
C	2.4235980	1.6176080	0.0000000	C	0.0399340	-2.9771540	0.0000000
C	2.5411070	-0.7187940	0.0000000	C	0.6062320	-4.3195170	0.0000000
C	3.9486490	-0.5819150	0.0000000	C	0.9114980	1.6941780	0.0000000
C	4.4467540	0.7253030	0.0000000	C	1.9510430	-4.4713220	0.0000000
C	6.0667420	-0.7855690	0.0000000	C	2.3436580	-2.0643620	0.0000000
C	-1.5852520	1.7173810	0.0000000	C	-0.6175760	3.2909800	0.0000000
C	-1.6664960	-0.7706220	0.0000000	C	-1.3248560	1.0062430	0.0000000
C	-3.1293990	-0.6697940	0.0000000	C	-1.6760640	2.3765360	0.0000000
C	-3.7659350	0.6153500	0.0000000	C	-2.5906280	4.2979180	0.0000000
C	-4.0298180	-1.7068150	0.0000000	H	0.6013680	-0.9563390	0.0000000
C	-5.1355170	0.5353840	0.0000000	H	1.9454450	1.3582650	0.0000000
H	0.0363300	0.4322100	0.0000000	H	2.4337160	-5.4424810	0.0000000
H	0.8986640	-1.9644270	0.0000000	H	3.7991620	-3.4986910	0.0000000
H	1.7539030	2.4738250	0.0000000	H	-0.0684950	-5.1639730	0.0000000
H	2.4718970	-2.7519730	0.0000000	H	-0.7538040	5.4171360	0.0000000
H	6.4917580	1.3229220	0.0000000	H	-1.9336160	-0.9639660	0.0000000
H	7.0755640	-1.1749720	0.0000000	H	-3.2111300	0.2475640	0.0000000
H	-3.3744450	2.6792760	0.0000000	H	-3.3010380	5.1131880	0.0000000
H	-3.8010590	-2.7632590	0.0000000	N	0.6954750	3.0100540	0.0000000
H	-5.8600030	1.3373920	0.0000000	N	0.9712200	-1.9410430	0.0000000
N	1.8067510	0.4177340	0.0000000	N	2.7949380	-3.3871990	0.0000000
N	1.9176170	-1.9101700	0.0000000	N	-0.0050420	0.7045570	0.0000000
N	3.7343700	1.8632380	0.0000000	N	-1.2275260	4.5255580	0.0000000
N	4.9740450	-1.5125500	0.0000000	N	-2.2306720	0.0136950	0.0000000
N	5.8154060	0.5733210	0.0000000	N	-2.9007610	3.0226820	0.0000000
N	-1.0148140	0.4474950	0.0000000	O	3.1182450	-1.1174820	0.0000000
N	-2.9698330	1.7537690	0.0000000	O	-1.1716200	-2.7319500	0.0000000
O	-0.9072470	2.7365350	0.0000000				
O	-1.0449510	-1.8369370	0.0000000				
S	-5.6490800	-1.1285930	0.0000000				

G-C			G <sup>th</sup> C				
C	0.0470820	-1.7181480	0.0000000	C	1.6823770	4.1381280	0.0000000
C	0.6607360	3.0525740	0.0000000	C	1.9077060	1.7260390	0.0000000
C	2.5506830	1.6557020	0.0000000	C	2.5213960	3.0162540	0.0000000
C	2.8884790	4.0231420	0.0000000	C	3.8270440	4.7005370	0.0000000
C	3.4502460	2.7871990	0.0000000	C	-0.1407550	-1.9192890	0.0000000
C	-0.8871450	-2.8002710	0.0000000	C	-0.2238960	2.9882480	0.0000000
C	-1.9024300	-4.6741850	0.0000000	C	-0.3013010	-4.4933000	0.0000000
C	-1.9688640	-0.2767080	0.0000000	C	-0.8177740	-3.2159490	0.0000000
C	-2.2519080	-2.4847120	0.0000000	C	-2.2125390	-0.7786920	0.0000000
H	0.0174170	0.3544340	0.0000000	C	-2.2527810	-3.2122580	0.0000000
H	1.0903090	5.0692890	0.0000000	C	-2.7979020	-4.4728280	0.0000000
H	2.3822420	-0.4068740	0.0000000	H	0.7348400	-4.8026470	0.0000000
H	3.4701370	4.9383020	0.0000000	H	1.6654350	-0.9145410	0.0000000
H	4.0215930	0.2427350	0.0000000	H	1.7554550	-2.6755050	0.0000000
H	4.5249330	2.6579350	0.0000000	H	2.2562340	6.1818430	0.0000000
H	-1.7689050	1.8057630	0.0000000	H	4.6922670	5.3489400	0.0000000
H	-2.1470420	-5.7273740	0.0000000	H	-0.0004650	0.9158210	0.0000000
H	-3.3982360	1.1485730	0.0000000	H	-2.0443260	1.9617600	0.0000000
H	-3.8910890	-3.8335530	0.0000000	H	-2.1051330	3.7192130	0.0000000
N	1.2194560	1.8149770	0.0000000	H	-3.8402820	-4.7575230	0.0000000
N	1.5328130	4.1601490	0.0000000	H	-3.9070030	-1.9275590	0.0000000
N	3.0273580	0.4066110	0.0000000	N	0.3350140	4.1918320	0.0000000
N	-0.6063030	-0.4708340	0.0000000	N	0.5036180	1.8190480	0.0000000
N	-0.6950470	-4.1722940	0.0000000	N	1.1908480	-1.8417050	0.0000000
N	-2.4018360	1.0028750	0.0000000	N	2.5358150	5.2119970	0.0000000
N	-2.8449900	-1.2732370	0.0000000	N	3.8545590	3.3933560	0.0000000
N	-2.8908290	-3.6986450	0.0000000	N	-0.8400070	-0.7890370	0.0000000
O	1.2893990	-1.7630840	0.0000000	N	-1.5684460	2.8671430	0.0000000
O	-0.5589930	3.2661140	0.0000000	N	-2.8985230	-1.9913280	0.0000000
				O	2.4524480	0.6074370	0.0000000
				O	-2.8661560	0.2730580	0.0000000
				S	-1.5473590	-5.6781820	0.0000000

	${}^{th} \mathbf{A} - {}^{th} \mathbf{U}$			${}^{th} \mathbf{A} - \mathbf{U}$			
C	0.7026030	2.5354100	0.0000000	C	0.4577840	1.7423290	0.0000000
C	2.0724230	0.4580780	0.0000000	C	1.4968720	-2.9100700	0.0000000
C	3.1378540	2.7358560	0.0000000	C	2.5359620	-3.9299860	0.0000000
C	3.2668770	1.3075060	0.0000000	C	3.2678880	-1.1788830	0.0000000
C	4.3474340	3.3832530	0.0000000	C	3.8354610	-3.5508800	0.0000000
C	4.5766600	0.8938350	0.0000000	C	-1.3013160	0.2127560	0.0000000
C	-1.5197240	-1.7007020	0.0000000	C	-1.6119600	2.6493240	0.0000000
C	-2.6771190	0.3225630	0.0000000	C	-2.2148710	1.3345390	0.0000000
C	-2.7971330	-2.3790690	0.0000000	C	-2.5645640	3.6536470	0.0000000
C	-3.1041960	-3.7260080	0.0000000	C	-3.5957100	1.3735030	0.0000000
C	-3.9693460	-1.5316220	0.0000000	H	1.2297990	-0.8303250	0.0000000
C	-5.1399300	-2.2705610	0.0000000	H	1.5417540	1.8330800	0.0000000
H	0.5342390	-1.8642480	0.0000000	H	2.2392920	-4.9693480	0.0000000
H	1.7253090	4.2905190	0.0000000	H	4.6554330	-4.2606490	0.0000000
H	4.5461200	4.4456070	0.0000000	H	5.1643380	-1.9404790	0.0000000
H	4.9339670	-0.1262920	0.0000000	H	-1.0607140	-1.8345840	0.0000000
H	-0.0076400	0.5890780	0.0000000	H	-2.3929750	4.7204080	0.0000000
H	-0.3382780	-3.3799070	0.0000000	H	-2.7179690	-1.2714980	0.0000000
H	-2.4428790	-4.5817230	0.0000000	H	-4.3061410	0.5581080	0.0000000
H	-2.5536710	1.4031530	0.0000000	N	0.0090700	0.4458820	0.0000000
H	-6.1528790	-1.8945820	0.0000000	N	1.9533730	-1.5954800	0.0000000
N	0.8805750	1.1534910	0.0000000	N	4.1949250	-2.2256550	0.0000000
N	1.8643280	3.2899840	0.0000000	N	-0.2490370	2.8429050	0.0000000
N	-0.3529730	-2.3740640	0.0000000	N	-1.7340150	-1.0618680	0.0000000
N	-1.4927910	-0.3708740	0.0000000	O	0.2833530	-3.1537300	0.0000000
N	-3.8939320	-0.1572300	0.0000000	O	3.6167040	-0.0072690	0.0000000
O	2.1016100	-0.7773590	0.0000000	S	-4.1583650	2.9999190	0.0000000
O	-0.4076640	3.0487990	0.0000000				
S	5.6547530	2.2329490	0.0000000				
S	-4.8090080	-3.9610110	0.0000000				

${}^{th}\mathbf{G-C}$			${}^{th}\mathbf{G-}{}^{th}\mathbf{C}$				
C	1.3530850	0.2799290	0.0000000	C	0.4577810	2.7788210	0.0000000
C	1.4827900	2.7739290	0.0000000	C	2.0394910	0.9052610	0.0000000
C	2.1537370	1.4921720	0.0000000	C	2.6706620	3.3207200	0.0000000
C	2.3903900	3.8151970	0.0000000	C	3.0685010	1.9298400	0.0000000
C	3.5260540	1.5840670	0.0000000	C	3.7701920	4.1568570	0.0000000
C	-0.5747440	1.7959160	0.0000000	C	4.4312200	1.7421300	0.0000000
C	-1.0988880	-3.1323950	0.0000000	C	-1.0214450	-1.9455150	0.0000000
C	-1.9122530	-4.3264370	0.0000000	C	-1.9787840	-4.3403830	0.0000000
C	-2.9906580	-1.7389660	0.0000000	C	-2.0692400	-2.9653390	0.0000000
C	-3.2594370	-4.1548760	0.0000000	C	-2.6306320	-0.2128150	0.0000000
H	0.6876370	-4.1217340	0.0000000	C	-3.4308530	-2.5121200	0.0000000
H	0.8208380	-2.3666950	0.0000000	C	-4.3432390	-3.5387160	0.0000000
H	2.1772620	4.8743520	0.0000000	H	0.5439310	-3.2600980	0.0000000
H	4.2403230	0.7728930	0.0000000	H	1.0062200	-1.5623060	0.0000000
H	-0.6154770	-0.3131810	0.0000000	H	3.7757990	5.2372200	0.0000000
H	-1.4711300	-5.3148770	0.0000000	H	4.9668090	0.8033560	0.0000000
H	-2.3694760	2.7203240	0.0000000	H	-0.0083100	0.7210260	0.0000000
H	-2.5024760	0.9707710	0.0000000	H	-1.0919350	-4.9590700	0.0000000
H	-3.9616890	-4.9811460	0.0000000	H	-1.1145230	4.0447690	0.0000000
H	-4.7898070	-2.7466370	0.0000000	H	-1.5937860	2.3553540	0.0000000
N	0.1116550	2.9076150	0.0000000	H	-4.5991880	-0.7740440	0.0000000
N	0.2374160	-3.2205020	0.0000000	H	-5.4223410	-3.4827290	0.0000000
N	-0.0038170	0.5206890	0.0000000	N	0.2691860	-2.2915810	0.0000000
N	-1.6415700	-1.9084080	0.0000000	N	0.7594150	1.4136410	0.0000000
N	-1.9290180	1.8146490	0.0000000	N	1.3550380	3.7284910	0.0000000
N	-3.7904000	-2.9008420	0.0000000	N	-0.8643410	3.0692400	0.0000000
O	1.8024100	-0.8805120	0.0000000	N	-1.3287310	-0.6545840	0.0000000
O	-3.5534270	-0.6379850	0.0000000	N	-3.6613830	-1.1505860	0.0000000
S	4.0239790	3.2300950	0.0000000	O	2.2470510	-0.3226320	0.0000000
				O	-2.9193400	0.9895240	0.0000000
				S	5.2514520	3.2533110	0.0000000
				S	-3.5332050	-5.0749860	0.0000000

## M06-2X

All geometries were optimized at the M06-2X/6-31++g(d,p) level of theory in the gas phase.

A-U				G-C			
C	0.0427610	-2.9776580	0.0000000	C	0.0279630	-1.7052730	0.0000000
C	0.6188620	-4.3177730	0.0000000	C	0.6800650	3.0375810	0.0000000
C	0.9000090	1.6733750	0.0000000	C	2.5709610	1.6594870	0.0000000
C	1.9593420	-4.4530870	0.0000000	C	2.8932500	4.0199930	0.0000000
C	2.3323030	-2.0523800	0.0000000	C	3.4650020	2.7947810	0.0000000
C	-0.6129680	3.2733630	0.0000000	C	-0.9013080	-2.7940250	0.0000000
C	-1.3322280	1.0023420	0.0000000	C	-1.8996860	-4.6641950	0.0000000
C	-1.6739930	2.3730720	0.0000000	C	-1.9938420	-0.2864670	0.0000000
C	-2.5699400	4.2904200	0.0000000	C	-2.2593290	-2.4864270	0.0000000
H	0.5852910	-0.9596320	0.0000000	H	1.0885550	5.0492100	0.0000000
H	1.9316070	1.3285720	0.0000000	H	2.4181800	-0.3881910	0.0000000
H	2.4515740	-5.4190050	0.0000000	H	3.4653870	4.9407780	0.0000000
H	3.7983230	-3.4669370	0.0000000	H	4.0544230	0.2645100	0.0000000
H	-0.0480590	-5.1676410	0.0000000	H	4.5393520	2.6710890	0.0000000
H	-0.7307060	5.3954370	0.0000000	H	-0.0156680	0.3556210	0.0000000
H	-1.9639840	-0.9520420	0.0000000	H	-1.7898780	1.7874840	0.0000000
H	-3.2254170	0.2759870	0.0000000	H	-2.1382860	-5.7182890	0.0000000
H	-3.2738810	5.1108990	0.0000000	H	-3.4208190	1.1336230	0.0000000
N	0.6961670	2.9827460	0.0000000	H	-3.8882840	-3.8375530	0.0000000
N	0.9633110	-1.9412570	0.0000000	N	1.2464260	1.8049620	0.0000000
N	2.7935940	-3.3645020	0.0000000	N	1.5402200	4.1450640	0.0000000
N	-0.0228530	0.6918630	0.0000000	N	3.0599090	0.4203920	0.0000000
N	-1.2108330	4.5079550	0.0000000	N	-0.6338800	-0.4703180	0.0000000
N	-2.2491070	0.0266350	0.0000000	N	-0.7001410	-4.1590640	0.0000000
N	-2.8901500	3.0242310	0.0000000	N	-2.4248580	0.9897250	0.0000000
O	3.0912230	-1.1017270	0.0000000	N	-2.8622710	-1.2787060	0.0000000
O	-1.1611410	-2.7401840	0.0000000	N	-2.8892860	-3.6983780	0.0000000
				O	1.2604730	-1.7410670	0.0000000
				O	-0.5333660	3.2401060	0.0000000

${}^{th}\mathbf{A}-{}^{th}\mathbf{U}$			${}^{th}\mathbf{A}-\mathbf{U}$				
C	0.7012050	2.5195030	0.0000000	C	0.4570100	1.7199390	0.0000000
C	2.0771090	0.4551580	0.0000000	C	1.5009130	-2.9108920	0.0000000
C	3.1271980	2.7367510	0.0000000	C	2.5485470	-3.9246380	0.0000000
C	3.2667760	1.3140810	0.0000000	C	3.2536660	-1.1731470	0.0000000
C	4.3303510	3.3854940	0.0000000	C	3.8376020	-3.5316260	0.0000000
C	4.5740700	0.9117860	0.0000000	C	-1.3051630	0.2047190	0.0000000
C	-1.5142830	-1.7042120	0.0000000	C	-1.6007160	2.6354870	0.0000000
C	-2.6576970	0.3186650	0.0000000	C	-2.2110000	1.3335470	0.0000000
C	-2.7969170	-2.3741810	0.0000000	C	-2.5434040	3.6421030	0.0000000
C	-3.1137260	-3.7140500	0.0000000	C	-3.5868360	1.3841700	0.0000000
C	-3.9554350	-1.5225990	0.0000000	H	1.2169110	-0.8411090	0.0000000
C	-5.1260540	-2.2518420	0.0000000	H	1.5422510	1.8005930	0.0000000
H	0.0020700	0.5743860	0.0000000	H	2.2614940	-4.9660620	0.0000000
H	0.5280230	-1.8966080	0.0000000	H	4.6646770	-4.2325150	0.0000000
H	1.7028640	4.2807360	0.0000000	H	5.1522940	-1.9117630	0.0000000
H	4.5249390	4.4478490	0.0000000	H	-1.0941760	-1.8359510	0.0000000
H	4.9403820	-0.1051890	0.0000000	H	-2.3645880	4.7071010	0.0000000
H	-0.3652380	-3.3994270	0.0000000	H	-2.7434750	-1.2513500	0.0000000
H	-2.4618750	-4.5765830	0.0000000	H	-4.3060680	0.5769720	0.0000000
H	-2.5248670	1.3986230	0.0000000	N	1.9470940	-1.6000240	0.0000000
H	-6.1360140	-1.8694580	0.0000000	N	4.1860410	-2.2059070	0.0000000
N	0.8876500	1.1425780	0.0000000	N	-0.0018910	0.4254470	0.0000000
N	1.8514510	3.2818390	0.0000000	N	-0.2358170	2.8186320	0.0000000
N	-0.3631250	-2.3940250	0.0000000	N	-1.7568120	-1.0578450	0.0000000
N	-1.4768120	-0.3837670	0.0000000	O	0.2975910	-3.1590860	0.0000000
N	-3.8697410	-0.1482560	0.0000000	O	3.5864010	-0.0040400	0.0000000
O	2.1149070	-0.7706920	0.0000000	S	-4.1281760	3.0029390	0.0000000
O	-0.4078080	3.0179190	0.0000000				
S	5.6330000	2.2508810	0.0000000				
S	-4.8068120	-3.9305300	0.0000000				

${}^{th}\mathbf{G-C}$			${}^{th}\mathbf{G-}{}^{th}\mathbf{C}$				
C	1.3354520	0.2945590	0.0000000	C	0.4645600	2.8105830	0.0000000
C	1.4834020	2.7845980	0.0000000	C	2.0218810	0.9250780	0.0000000
C	2.1433500	1.5051480	0.0000000	C	2.6733230	3.3322510	0.0000000
C	2.3961410	3.8132190	0.0000000	C	3.0589670	1.9453710	0.0000000
C	3.5106970	1.5886240	0.0000000	C	3.7764000	4.1535160	0.0000000
C	-0.5730730	1.8247090	0.0000000	C	4.4147630	1.7485530	0.0000000
C	-1.0962660	-3.1565340	0.0000000	C	-1.0182750	-1.9686100	0.0000000
C	-1.9160230	-4.3455750	0.0000000	C	-1.9916430	-4.3527680	0.0000000
C	-2.9686930	-1.7546420	0.0000000	C	-2.0724720	-2.9828750	0.0000000
C	-3.2556970	-4.1597950	0.0000000	C	-2.6117020	-0.2338980	0.0000000
H	0.6760720	-4.1647700	0.0000000	C	-3.4256420	-2.5212830	0.0000000
H	0.8146900	-2.4115800	0.0000000	C	-4.3379620	-3.5412520	0.0000000
H	2.1905130	4.8731580	0.0000000	H	0.5253860	-3.3032350	0.0000000
H	4.2207600	0.7739290	0.0000000	H	0.9960580	-1.6079660	0.0000000
H	-0.6288540	-0.2792410	0.0000000	H	3.7908520	5.2331360	0.0000000
H	-1.4813530	-5.3357900	0.0000000	H	4.9441440	0.8064180	0.0000000
H	-2.3619480	2.7487780	0.0000000	H	-0.0180540	0.7613450	0.0000000
H	-2.4915780	0.9983760	0.0000000	H	-1.1004180	4.0774390	0.0000000
H	-3.9671150	-4.9777750	0.0000000	H	-1.1110090	-4.9802170	0.0000000
H	-4.7703710	-2.7387630	0.0000000	H	-1.5787260	2.3875750	0.0000000
N	0.1116160	2.9283070	0.0000000	H	-4.5784960	-0.7730310	0.0000000
N	0.2341290	-3.2605420	0.0000000	H	-5.4162080	-3.4822440	0.0000000
N	-0.0136920	0.5480070	0.0000000	N	0.2622370	-2.3322250	0.0000000
N	-1.6223700	-1.9344670	0.0000000	N	0.7526510	1.4459320	0.0000000
N	-1.9235070	1.8428890	0.0000000	N	1.3595420	3.7515590	0.0000000
N	-3.7731960	-2.9041480	0.0000000	N	-0.8536530	3.1017680	0.0000000
O	1.7751680	-0.8582350	0.0000000	N	-1.3119000	-0.6837470	0.0000000
O	-3.5174640	-0.6555070	0.0000000	N	-3.6451010	-1.1595590	0.0000000
S	4.0125200	3.2204700	0.0000000	O	2.2207080	-0.2935200	0.0000000
				O	-2.8860320	0.9627280	0.0000000
				S	5.2381230	3.2439810	0.0000000
				S	-3.5408290	-5.0687500	0.0000000

**M06-2X-D3**

All geometries were optimized at the M06-2X-D3/6-31++g(d,p) level of theory in the gas phase.

A- <sup>th</sup> U			A-U				
C	2.4075750	1.6015680	0.0000000	C	0.0421710	-2.9774940	0.0000000
C	2.5394800	-0.7252530	0.0000000	C	0.6189610	-4.3172940	0.0000000
C	3.9442520	-0.5793750	0.0000000	C	0.9002920	1.6720590	0.0000000
C	4.4282030	0.7253270	0.0000000	C	1.9595290	-4.4518420	0.0000000
C	6.0522330	-0.7670090	0.0000000	C	2.3311910	-2.0509830	0.0000000
C	-1.5731370	1.7043550	0.0000000	C	-0.6123050	3.2724500	0.0000000
C	-1.6667240	-0.7757030	0.0000000	C	-1.3322990	1.0016150	0.0000000
C	-3.1303210	-0.6642580	0.0000000	C	-1.6736420	2.3724460	0.0000000
C	-3.7547910	0.6217750	0.0000000	C	-2.5690390	4.2901360	0.0000000
C	-4.0346350	-1.6902030	0.0000000	H	0.5837310	-0.9590160	0.0000000
C	-5.1195800	0.5462740	0.0000000	H	1.9317360	1.3268140	0.0000000
H	0.0380870	0.4142260	0.0000000	H	2.4522850	-5.4174880	0.0000000
H	0.9198930	-1.9882860	0.0000000	H	3.7980120	-3.4647020	0.0000000
H	1.7307250	2.4528100	0.0000000	H	-0.0475240	-5.1675050	0.0000000
H	2.5036750	-2.7538160	0.0000000	H	-0.7294130	5.3945600	0.0000000
H	6.4631770	1.3389100	0.0000000	H	-1.9643550	-0.9527090	0.0000000
H	7.0636830	-1.1486290	0.0000000	H	-3.2256030	0.2755290	0.0000000
H	-3.3475190	2.6827730	0.0000000	H	-3.2727490	5.1107950	0.0000000
H	-3.8151040	-2.7486410	0.0000000	N	0.6967730	2.9814770	0.0000000
H	-5.8407090	1.3503190	0.0000000	N	0.9621760	-1.9406630	0.0000000
N	1.8003560	0.3985050	0.0000000	N	2.7932290	-3.3628170	0.0000000
N	1.9358770	-1.9214210	0.0000000	N	-0.0229730	0.6909180	0.0000000
N	3.7080590	1.8562820	0.0000000	N	-1.2098330	4.5072320	0.0000000
N	4.9714350	-1.5003220	0.0000000	N	-2.2493270	0.0260350	0.0000000
N	5.7928080	0.5846980	0.0000000	N	-2.8896030	3.0240180	0.0000000
N	-1.0128720	0.4342870	0.0000000	O	3.0895630	-1.0998570	0.0000000
N	-2.9519400	1.7536310	0.0000000	O	-1.1618780	-2.7406420	0.0000000
O	-0.8880340	2.7098590	0.0000000				
O	-1.0573230	-1.8383910	0.0000000				
S	-5.6375240	-1.1019380	0.0000000				

A-U			G-C				
C	0.0421710	-2.9774940	0.0000000	C	0.0284290	-1.7050490	0.0000000
C	0.6189610	-4.3172940	0.0000000	C	0.6795490	3.0374130	0.0000000
C	0.9002920	1.6720590	0.0000000	C	2.5703260	1.6590380	0.0000000
C	1.9595290	-4.4518420	0.0000000	C	2.8928150	4.0195280	0.0000000
C	2.3311910	-2.0509830	0.0000000	C	3.4644510	2.7942570	0.0000000
C	-0.6123050	3.2724500	0.0000000	C	-0.9008580	-2.7937220	0.0000000
C	-1.3322990	1.0016150	0.0000000	C	-1.8993790	-4.6639050	0.0000000
C	-1.6736420	2.3724460	0.0000000	C	-1.9932600	-0.2860830	0.0000000
C	-2.5690390	4.2901360	0.0000000	C	-2.2588930	-2.4860460	0.0000000
H	0.5837310	-0.9590160	0.0000000	H	1.0881690	5.0489180	0.0000000
H	1.9317360	1.3268140	0.0000000	H	2.4172330	-0.3886970	0.0000000
H	2.4522850	-5.4174880	0.0000000	H	3.4650560	4.9402480	0.0000000
H	3.7980120	-3.4647020	0.0000000	H	4.0535930	0.2639570	0.0000000
H	-0.0475240	-5.1675050	0.0000000	H	4.5387940	2.6704590	0.0000000
H	-0.7294130	5.3945600	0.0000000	H	-0.0150420	0.3558800	0.0000000
H	-1.9643550	-0.9527090	0.0000000	H	-1.7892670	1.7878720	0.0000000
H	-3.2256030	0.2755290	0.0000000	H	-2.1380210	-5.7179780	0.0000000
H	-3.2727490	5.1107950	0.0000000	H	-3.4202230	1.1338970	0.0000000
N	0.6967730	2.9814770	0.0000000	H	-3.8879460	-3.8370350	0.0000000
N	0.9621760	-1.9406630	0.0000000	N	1.2457700	1.8047340	0.0000000
N	2.7932290	-3.3628170	0.0000000	N	1.5398030	4.1447610	0.0000000
N	-0.0229730	0.6909180	0.0000000	N	3.0590850	0.4199080	0.0000000
N	-1.2098330	4.5072320	0.0000000	N	-0.6333110	-0.4700640	0.0000000
N	-2.2493270	0.0260350	0.0000000	N	-0.6997950	-4.1588020	0.0000000
N	-2.8896030	3.0240180	0.0000000	N	-2.4242450	0.9900850	0.0000000
O	3.0895630	-1.0998570	0.0000000	N	-2.8617530	-1.2782890	0.0000000
O	-1.1618780	-2.7406420	0.0000000	N	-2.8889290	-3.6979810	0.0000000
				O	1.2609680	-1.7409930	0.0000000
				O	-0.5338520	3.2401400	0.0000000

	${}^{th}\mathbf{A}-{}^{th}\mathbf{U}$			${}^{th}\mathbf{A}-\mathbf{U}$			
C	0.7001170	2.5181310	0.0000000	C	0.4579770	1.7185600	0.0000000
C	2.0766390	0.4542840	0.0000000	C	1.5001950	-2.9109330	0.0000000
C	3.1260120	2.7361940	0.0000000	C	2.5484230	-3.9240410	0.0000000
C	3.2660350	1.3135250	0.0000000	C	3.2518250	-1.1721130	0.0000000
C	4.3290260	3.3852370	0.0000000	C	3.8372280	-3.5301730	0.0000000
C	4.5734270	0.9115680	0.0000000	C	-1.3049150	0.2038640	0.0000000
C	-1.5132650	-1.7038280	0.0000000	C	-1.5994730	2.6347930	0.0000000
C	-2.6563150	0.3194860	0.0000000	C	-2.2103130	1.3330450	0.0000000
C	-2.7960770	-2.3734670	0.0000000	C	-2.5418000	3.6417650	0.0000000
C	-3.1132520	-3.7132570	0.0000000	C	-3.5861350	1.3842400	0.0000000
C	-3.9544220	-1.5215520	0.0000000	H	1.2149000	-0.8411060	0.0000000
C	-5.1252150	-2.2505340	0.0000000	H	1.5432610	1.7986010	0.0000000
H	0.0013930	0.5727920	0.0000000	H	2.2619960	-4.9656390	0.0000000
H	0.5290690	-1.8963500	0.0000000	H	4.6647440	-4.2305350	0.0000000
H	1.7010470	4.2797410	0.0000000	H	5.1509410	-1.9094810	0.0000000
H	4.5233320	4.4476420	0.0000000	H	-1.0942680	-1.8368740	0.0000000
H	4.9399080	-0.1053450	0.0000000	H	-2.3625470	4.7066900	0.0000000
H	-0.3643430	-3.3991200	0.0000000	H	-2.7434830	-1.2519910	0.0000000
H	-2.4615900	-4.5759340	0.0000000	H	-4.3056600	0.5773030	0.0000000
H	-2.5230670	1.3993860	0.0000000	N	1.9455890	-1.5998510	0.0000000
H	-6.1350750	-1.8678820	0.0000000	N	4.1848740	-2.2042310	0.0000000
N	0.8870370	1.1413200	0.0000000	N	-0.0015910	0.4243020	0.0000000
N	1.8500700	3.2809020	0.0000000	N	-0.2344850	2.8174740	0.0000000
N	-0.3621600	-2.3937140	0.0000000	N	-1.7567860	-1.0586290	0.0000000
N	-1.4756820	-0.3833710	0.0000000	O	0.2969920	-3.1598580	0.0000000
N	-3.8684430	-0.1471990	0.0000000	O	3.5837590	-0.0027430	0.0000000
O	2.1148540	-0.7715960	0.0000000	S	-4.1268830	3.0032600	0.0000000
O	-0.4090920	3.0161650	0.0000000				
S	5.6320570	2.2509620	0.0000000				
S	-4.8064360	-3.9293590	0.0000000				

${}^{th}\mathbf{G-C}$			${}^{th}\mathbf{G-}{}^{th}\mathbf{C}$				
C	1.3351280	0.2940120	0.0000000	C	0.4640030	2.8097830	0.0000000
C	1.4831130	2.7840200	0.0000000	C	2.0213060	0.9243350	0.0000000
C	2.1430370	1.5045350	0.0000000	C	2.6727940	3.3314500	0.0000000
C	2.3959070	3.8126090	0.0000000	C	3.0583970	1.9445410	0.0000000
C	3.5103870	1.5879390	0.0000000	C	3.7759290	4.1526610	0.0000000
C	-0.5733890	1.8241320	0.0000000	C	4.4141800	1.7476310	0.0000000
C	-1.0958760	-3.1557050	0.0000000	C	-1.0177640	-1.9676140	0.0000000
C	-1.9155080	-4.3448260	0.0000000	C	-1.9907530	-4.3519190	0.0000000
C	-2.9685760	-1.7540570	0.0000000	C	-2.0718230	-2.9820310	0.0000000
C	-3.2552080	-4.1592080	0.0000000	C	-2.6116010	-0.2331650	0.0000000
H	0.6765200	-4.1636880	0.0000000	C	-3.4251200	-2.5206700	0.0000000
H	0.8150010	-2.4103950	0.0000000	C	-4.3372330	-3.5408470	0.0000000
H	2.1902830	4.8725470	0.0000000	H	0.5261220	-3.3018760	0.0000000
H	4.2203450	0.7731520	0.0000000	H	0.9965600	-1.6064380	0.0000000
H	-0.6291530	-0.2798700	0.0000000	H	3.7904050	5.2322800	0.0000000
H	-1.4807130	-5.3349910	0.0000000	H	4.9434310	0.8054220	0.0000000
H	-2.3621290	2.7482920	0.0000000	H	-0.0186480	0.7604790	0.0000000
H	-2.4918910	0.9978380	0.0000000	H	-1.1007940	4.0767190	0.0000000
H	-3.9665140	-4.9772850	0.0000000	H	-1.1099730	-4.9791610	0.0000000
H	-4.7700720	-2.7383160	0.0000000	H	-1.5792830	2.3868360	0.0000000
N	0.1113200	2.9277290	0.0000000	H	-4.5782170	-0.7724440	0.0000000
N	0.2344970	-3.2594960	0.0000000	H	-5.4154880	-3.4820240	0.0000000
N	-0.0139850	0.5474290	0.0000000	N	0.2627870	-2.3309130	0.0000000
N	-1.6222350	-1.9337000	0.0000000	N	0.7521040	1.4451200	0.0000000
N	-1.9237910	1.8423490	0.0000000	N	1.3590100	3.7507530	0.0000000
N	-3.7728920	-2.9036380	0.0000000	N	-0.8541630	3.1010110	0.0000000
O	1.7749390	-0.8587830	0.0000000	N	-1.3117290	-0.6827640	0.0000000
O	-3.5175510	-0.6550200	0.0000000	N	-3.6448180	-1.1589760	0.0000000
S	4.0123180	3.2198010	0.0000000	O	2.2202720	-0.2942890	0.0000000
				O	-2.8862450	0.9633900	0.0000000
				S	5.2376590	3.2430460	0.0000000
				S	-3.5398330	-5.0682680	0.0000000

## $\omega$ B97XD

All geometries were optimized at the  $\omega$ B97XD/6-31++g(d,p) level of theory in the gas phase.

A- <sup>th</sup> U			A-U				
C	2.3323720	1.7480090	0.0000000	C	0.0456180	-2.9689500	0.0000000
C	2.5763450	-0.5687880	0.0000000	C	0.6049560	-4.3133850	0.0000000
C	3.9707850	-0.3546650	0.0000000	C	0.9044750	1.7017430	0.0000000
C	4.3927030	0.9697990	0.0000000	C	1.9430800	-4.4679140	0.0000000
C	6.0866770	-0.4405130	0.0000000	C	2.3456690	-2.0725950	0.0000000
C	-1.6193530	-0.8573300	0.0000000	C	-0.6239120	3.2871660	0.0000000
C	-1.6917990	1.6224320	0.0000000	C	-1.3205090	1.0100860	0.0000000
C	-3.0852040	-0.8446430	0.0000000	C	-1.6737500	2.3763980	0.0000000
C	-3.7942370	0.3951690	0.0000000	C	-2.5885820	4.2862560	0.0000000
C	-3.9204550	-1.9284460	0.0000000	H	0.6123220	-0.9601200	0.0000000
C	-5.1509430	0.2318660	0.0000000	H	1.9392860	1.3690540	0.0000000
H	0.0000000	0.4412610	0.0000000	H	2.4221160	-5.4400600	0.0000000
H	1.0108330	-1.8987080	0.0000000	H	3.7913610	-3.5068290	0.0000000
H	1.6170020	2.5663350	0.0000000	H	-0.0714790	-5.1557980	0.0000000
H	2.6238440	-2.5977410	0.0000000	H	-0.7623660	5.4078350	0.0000000
H	6.3955060	1.6802340	0.0000000	H	-1.9231280	-0.9544880	0.0000000
H	7.1152250	-0.7728950	0.0000000	H	-3.2022220	0.2535360	0.0000000
H	-3.5284510	2.4755410	0.0000000	H	-3.3002300	5.0998490	0.0000000
H	-3.6311930	-2.9697100	0.0000000	N	0.6878560	3.0098790	0.0000000
H	-5.9221910	0.9882140	0.0000000	N	0.9774960	-1.9422270	0.0000000
N	1.7828640	0.5193030	0.0000000	N	2.7895720	-3.3899320	0.0000000
N	2.0239850	-1.7896380	0.0000000	N	-0.0065910	0.7113520	0.0000000
N	3.6199080	2.0650690	0.0000000	N	-1.2323020	4.5165660	0.0000000
N	5.0431360	-1.2259930	0.0000000	N	-2.2229870	0.0209630	0.0000000
N	5.7624030	0.8963840	0.0000000	N	-2.8977750	3.0173230	0.0000000
N	-1.0460090	0.3932310	0.0000000	O	3.1194540	-1.1322670	0.0000000
N	-3.0698300	1.5775890	0.0000000	O	-1.1583220	-2.7203670	0.0000000
O	-0.9397400	-1.8791170	0.0000000				
O	-1.0781800	2.6747440	0.0000000				
S	-5.5583600	-1.4479070	0.0000000				

G-C			G <sup>th</sup> C				
C	0.0424870	-1.7126940	0.0000000	C	1.6890730	4.1242900	0.0000000
C	0.6644450	3.0474160	0.0000000	C	1.9033930	1.7203860	0.0000000
C	2.5464240	1.6548270	0.0000000	C	2.5196050	3.0079830	0.0000000
C	2.8828510	4.0137960	0.0000000	C	3.8236750	4.6806120	0.0000000
C	3.4449200	2.7846770	0.0000000	C	-0.1455640	-1.9165800	0.0000000
C	-0.8906470	-2.7941610	0.0000000	C	-0.2137640	2.9856890	0.0000000
C	-1.8997400	-4.6598610	0.0000000	C	-0.3143130	-4.4857620	0.0000000
C	-1.9676080	-0.2816240	0.0000000	C	-0.8248310	-3.2115740	0.0000000
C	-2.2463040	-2.4813900	0.0000000	C	-2.2101410	-0.7796720	0.0000000
H	0.0110690	0.3519110	0.0000000	C	-2.2531710	-3.2043120	0.0000000
H	1.0886630	5.0558030	0.0000000	C	-2.7969580	-4.4596260	0.0000000
H	2.3810070	-0.3987640	0.0000000	H	0.0000000	0.9189320	0.0000000
H	3.4616710	4.9299300	0.0000000	H	0.7204090	-4.7991980	0.0000000
H	4.0178690	0.2503860	0.0000000	H	1.6554240	-0.9227740	0.0000000
H	4.5189510	2.6561090	0.0000000	H	1.7415530	-2.6813510	0.0000000
H	-1.7599010	1.7947660	0.0000000	H	2.2635870	6.1613470	0.0000000
H	-2.1438420	-5.7126720	0.0000000	H	4.6900470	5.3266500	0.0000000
H	-3.3896470	1.1421240	0.0000000	H	-2.0276950	1.9584490	0.0000000
H	-3.8818770	-3.8248120	0.0000000	H	-2.0892490	3.7143760	0.0000000
N	1.2211030	1.8114180	0.0000000	H	-3.8391400	-4.7430140	0.0000000
N	1.5305430	4.1484300	0.0000000	H	-3.9019580	-1.9177440	0.0000000
N	3.0250280	0.4117800	0.0000000	N	0.3421520	4.1831440	0.0000000
N	-0.6091670	-0.4735290	0.0000000	N	0.5076970	1.8179980	0.0000000
N	-0.6967600	-4.1624470	0.0000000	N	1.1810780	-1.8466830	0.0000000
N	-2.3950580	0.9952610	0.0000000	N	2.5399780	5.1927330	0.0000000
N	-2.8413680	-1.2711290	0.0000000	N	3.8509060	3.3792270	0.0000000
N	-2.8837430	-3.6895270	0.0000000	N	-0.8388360	-0.7902980	0.0000000
O	1.2776270	-1.7572590	0.0000000	N	-1.5539090	2.8633740	0.0000000
O	-0.5497320	3.2576870	0.0000000	N	-2.8955470	-1.9840260	0.0000000
				O	2.4425590	0.6070290	0.0000000
				O	-2.8564780	0.2686760	0.0000000
				S	-1.5561430	-5.6561370	0.0000000

	${}^{th}\mathbf{A}-{}^{th}\mathbf{U}$			${}^{th}\mathbf{A}-\mathbf{U}$			
C	0.7282670	2.5344360	0.0000000	C	0.4325360	1.7537830	0.0000000
C	2.0695360	0.4486190	0.0000000	C	1.5212160	-2.8869040	0.0000000
C	3.1566130	2.7096900	0.0000000	C	2.5629920	-3.9029630	0.0000000
C	3.2718650	1.2860870	0.0000000	C	3.2812620	-1.1579530	0.0000000
C	4.3685740	3.3409560	0.0000000	C	3.8542960	-3.5185080	0.0000000
C	4.5733610	0.8632510	0.0000000	C	-1.3008610	0.2082210	0.0000000
C	-1.5310860	-1.6882310	0.0000000	C	-1.6413950	2.6326460	0.0000000
C	-2.6810780	0.3291170	0.0000000	C	-2.2259080	1.3197320	0.0000000
C	-2.8101420	-2.3620780	0.0000000	C	-2.6015640	3.6217400	0.0000000
C	-3.1214490	-3.7038670	0.0000000	C	-3.6029340	1.3446460	0.0000000
C	-3.9724040	-1.5167630	0.0000000	H	1.2471740	-0.8181320	0.0000000
C	-5.1398640	-2.2498610	0.0000000	H	1.5152320	1.8566320	0.0000000
H	0.0000000	0.6011280	0.0000000	H	2.2722700	-4.9434770	0.0000000
H	0.5139110	-1.8613370	0.0000000	H	4.6771340	-4.2238360	0.0000000
H	1.7619860	4.2758110	0.0000000	H	5.1747180	-1.9078750	0.0000000
H	4.5779990	4.4006680	0.0000000	H	-1.0428350	-1.8287340	0.0000000
H	4.9231000	-0.1592550	0.0000000	H	-2.4420910	4.6898760	0.0000000
H	-0.3646340	-3.3717770	0.0000000	H	-2.7040020	-1.2817400	0.0000000
H	-2.4634090	-4.5616360	0.0000000	H	-4.3044910	0.5220650	0.0000000
H	-2.5546420	1.4092300	0.0000000	N	0.0000000	0.4528970	0.0000000
H	-6.1513240	-1.8713250	0.0000000	N	1.9712800	-1.5770130	0.0000000
N	0.8903460	1.1541100	0.0000000	N	4.2077650	-2.1949990	0.0000000
N	1.8910530	3.2757510	0.0000000	N	-0.2802480	2.8408280	0.0000000
N	-0.3736950	-2.3674180	0.0000000	N	-1.7233300	-1.0643230	0.0000000
N	-1.4988080	-0.3658490	0.0000000	O	0.3154330	-3.1343920	0.0000000
N	-3.8924530	-0.1422450	0.0000000	O	3.6223640	0.0101490	0.0000000
O	2.0879640	-0.7798920	0.0000000	S	-4.1751000	2.9525490	0.0000000
O	-0.3720930	3.0549380	0.0000000				
S	5.6533910	2.1849150	0.0000000				
S	-4.8132800	-3.9280680	0.0000000				

${}^{th}\mathbf{G-C}$			${}^{th}\mathbf{G-}{}^{th}\mathbf{C}$				
C	1.3506500	0.2856920	0.0000000	C	0.4884360	2.7717670	0.0000000
C	1.4787130	2.7738250	0.0000000	C	2.0430060	0.8869090	0.0000000
C	2.1484680	1.5000760	0.0000000	C	2.6995340	3.2893820	0.0000000
C	2.3809800	3.8111750	0.0000000	C	3.0815390	1.9020840	0.0000000
C	3.5158080	1.5961860	0.0000000	C	3.8028010	4.1097830	0.0000000
C	-0.5707550	1.7959860	0.0000000	C	4.4380490	1.7034710	0.0000000
C	-1.0936040	-3.1309650	0.0000000	C	-1.0388240	-1.9347160	0.0000000
C	-1.9050960	-4.3238260	0.0000000	C	-2.0223320	-4.3144650	0.0000000
C	-2.9794310	-1.7447570	0.0000000	C	-2.0965800	-2.9436700	0.0000000
C	-3.2459700	-4.1504720	0.0000000	C	-2.6277990	-0.1945390	0.0000000
H	0.6859630	-4.1220970	0.0000000	C	-3.4468930	-2.4777120	0.0000000
H	0.8179420	-2.3695740	0.0000000	C	-4.3650180	-3.4919370	0.0000000
H	2.1654230	4.8693320	0.0000000	H	0.0000000	0.7261830	0.0000000
H	4.2323390	0.7873990	0.0000000	H	0.5049010	-3.2670300	0.0000000
H	-0.6112370	-0.3061030	0.0000000	H	0.9832690	-1.5762490	0.0000000
H	-1.4644790	-5.3116830	0.0000000	H	3.8185340	5.1895480	0.0000000
H	-2.3634020	2.7101250	0.0000000	H	4.9653480	0.7603510	0.0000000
H	-2.4886430	0.9610890	0.0000000	H	-1.0694010	4.0454530	0.0000000
H	-3.9496450	-4.9746010	0.0000000	H	-1.1434750	-4.9441100	0.0000000
H	-4.7717840	-2.7454440	0.0000000	H	-1.5603300	2.3607470	0.0000000
N	0.0000000	0.5252520	0.0000000	H	-4.5949970	-0.7299760	0.0000000
N	0.1065360	2.9056590	0.0000000	H	-5.4429940	-3.4256530	0.0000000
N	0.2377330	-3.2218590	0.0000000	N	0.2419570	-2.2968840	0.0000000
N	-1.6324020	-1.9121930	0.0000000	N	0.7748340	1.4076020	0.0000000
N	-1.9207190	1.8073300	0.0000000	N	1.3873190	3.7108960	0.0000000
N	-3.7741780	-2.8991860	0.0000000	N	-0.8276640	3.0694600	0.0000000
O	1.8002160	-0.8664110	0.0000000	N	-1.3308900	-0.6471550	0.0000000
O	-3.5379480	-0.6483220	0.0000000	N	-3.6631050	-1.1159840	0.0000000
S	4.0032510	3.2316800	0.0000000	O	2.2384960	-0.3352230	0.0000000
				O	-2.9006270	1.0046340	0.0000000
				S	5.2633960	3.1968280	0.0000000
				S	-3.5744270	-5.0235920	0.0000000

## D4.2 Bases

### B3LYP

All geometries were optimized at the B3LYP/6-31++g(d,p) level of theory in the gas phase.

	<b>A</b>			<b>C</b>			
	(C1)			(C1)			
C	0.5458230	0.4427000	0.0006310	C	1.0855470	1.3365720	0.0034510
C	1.5673780	-1.5223640	-0.0025380	C	1.5404110	-0.0316190	0.0063450
C	-0.3611100	-0.6230260	-0.0057690	C	-0.2626840	1.5195990	-0.0004780
C	-1.0737130	1.9483270	-0.0015440	C	-0.6270650	-0.9027400	0.0000600
C	-1.7281970	-0.2706470	-0.0097460	H	1.7700860	2.1750670	-0.0016590
H	2.3876480	-2.2271720	-0.0022870	H	3.1736900	-1.2461270	0.0758490
H	2.6753970	0.3194000	0.0074140	H	3.5541490	0.4476910	0.1235670
H	-1.4004590	2.9849580	-0.0002920	H	-0.7275700	2.4998910	-0.0031650
H	-2.4949690	-2.1789390	-0.0347590	H	-2.0977020	0.5545860	-0.0052330
H	-3.6759920	-0.8896790	-0.0350780	N	0.7326660	-1.0773530	0.0074220
N	0.2486670	1.7487040	0.0032960	N	2.8776790	-0.2851330	-0.0116470
N	0.2951220	-1.8435560	-0.0078370	N	-1.0920610	0.4464940	-0.0021720
N	1.7860540	-0.1586540	0.0028210	O	-1.4590570	-1.8003970	-0.0018100
N	-2.0590300	1.0336550	-0.0080910				
N	-2.7164790	-1.1966330	-0.0116650				

	<b>A</b>			<b>C</b>			
	(Cs)			(Cs)			
C	0.0000000	0.5508490	0.0000000	C	1.0407910	1.1993480	0.0000000
C	0.6691970	-2.0321420	0.0000000	C	1.1344800	-0.2396670	0.0000000
C	1.3609800	0.1755130	0.0000000	C	-0.2171720	1.7175920	0.0000000
C	-0.9247520	-0.4994600	0.0000000	C	-1.1829190	-0.5336680	0.0000000
C	-1.9130580	1.4825450	0.0000000	H	1.9152070	1.8374080	0.0000000
H	0.9784500	-3.0741290	0.0000000	H	2.4117930	-1.8308010	0.0000000
H	2.1597770	2.0709090	0.0000000	H	3.2114820	-0.2839440	0.0000000
H	3.3191820	0.7619400	0.0000000	H	-0.4185970	2.7836780	0.0000000
H	-2.7213300	2.2010820	0.0000000	H	-2.2367540	1.2486470	0.0000000
H	-3.0519630	-0.3402980	0.0000000	N	0.0879260	-1.0469350	0.0000000
N	1.6697970	-1.1341860	0.0000000	N	2.3623380	-0.8231470	0.0000000
N	2.3646160	1.0847400	0.0000000	N	-1.2912550	0.8894960	0.0000000
N	-0.6355620	1.7822660	0.0000000	O	-2.2154070	-1.1910950	0.0000000
N	-0.6496460	-1.8102880	0.0000000				
N	-2.1546800	0.1227060	0.0000000				

<b>G</b>			(Cs)	<b>U</b>			(Cs)
C	0.2515760	1.0148980	0.0073880	C	1.2051150	1.1098390	0.0000000
C	0.5618340	-0.3464210	0.0014140	C	1.2909020	-0.3470260	0.0000000
C	2.3794650	0.9133010	0.0030420	C	-0.0053150	1.7116380	0.0000000
C	-1.1340940	1.4062430	0.0016150	C	-1.2206530	-0.4051320	0.0000000
C	-1.5417940	-1.0691290	-0.0000490	H	0.0458370	-2.0038650	0.0000000
H	2.4884580	-1.2398470	-0.0056850	H	2.1260200	1.6760470	0.0000000
H	3.4337290	1.1525670	0.0019690	H	-0.1250820	2.7893400	0.0000000
H	-2.1595250	-2.9775810	0.0980540	H	-2.0772930	1.4448020	0.0000000
H	-2.9568540	0.4459560	-0.0741310	N	0.0336630	-0.9898450	0.0000000
H	-3.4092780	-1.8534260	0.3611440	N	-1.1762010	0.9876220	0.0000000
N	1.4008560	1.7822310	0.0076040	O	2.3195530	-1.0080060	0.0000000
N	1.9309690	-0.3979040	-0.0003130	O	-2.2685530	-1.0303280	0.0000000
N	-0.2758560	-1.4159240	0.0107940				
N	-1.9679230	0.2350840	-0.0041170				
N	-2.5116650	-2.0435600	-0.0618350				
O	-1.6592120	2.5088260	-0.0055710				

<b>G</b>			(Cs)	<i>th</i> <b>A</b>			(C1)
C	0.9996610	0.0178100	-0.0267110	C	0.2651520	-1.7050560	-0.0126470
C	1.8374500	2.0638660	0.0228910	C	0.3038580	0.6531500	0.0032070
C	-0.0022210	0.9899970	-0.0029890	C	1.6605670	0.3732360	0.0061870
C	-0.3705210	-1.7339240	-0.0666060	C	-0.4969020	-0.5512690	-0.0043450
C	-1.3773450	0.5630290	-0.0032880	C	-1.5631510	1.9266730	-0.0095300
H	0.1509190	-3.6674560	-0.2181760	C	-1.9271400	-0.3497560	0.0004860
H	2.5911210	2.8386880	0.0417030	H	2.4790860	1.0782070	0.0114200
H	3.1019930	0.3182590	-0.0197730	H	-0.0485180	-2.7395410	-0.0360590
H	-1.5105250	-3.3989880	-0.4678950	H	-2.0424430	2.9029390	-0.0166340
H	-2.3822860	-1.2368120	0.0284630	H	-2.4741520	-2.3297940	0.2214590
N	0.5434100	2.2592240	0.0284840	H	-3.7679360	-1.1928930	0.1118130
N	0.8799100	-1.3347060	-0.0716180	N	-0.2535500	1.9095380	-0.0004120
N	2.1757820	0.7204320	-0.0107730	N	-2.4335680	0.8694850	-0.0036310
N	-0.6526470	-3.0806510	-0.0395470	N	-2.7895970	-1.4051460	-0.0230110
N	-1.4384180	-0.8729890	-0.0355620	S	1.9427320	-1.3238160	-0.0053600
O	-2.4183790	1.2009520	0.0248960				

${}^{th}\mathbf{A}$			(Cs)	${}^{th}\mathbf{C}$			(Cs)
C	0.3017630	0.9633640	0.0076830	C	0.8759510	1.2719100	0.0000000
C	1.0503170	-1.2731780	-0.0129890	C	2.1138260	-0.7287660	0.0000000
C	1.6742670	1.1510440	0.0128670	C	-0.3345590	-0.8746960	0.0000000
C	-0.0523710	-0.4389020	-0.0039210	C	-0.4001610	0.5596810	0.0000000
C	-1.4681210	-0.7250090	-0.0016010	C	-1.5740020	-1.4697390	0.0000000
C	-1.8826920	1.5427720	-0.0063000	C	-1.6974430	1.0244330	0.0000000
H	1.0989420	-2.3530160	-0.0391190	H	0.0395080	3.1768590	0.0000000
H	2.2114130	2.0882690	0.0211850	H	1.0205580	-2.4658150	0.0000000
H	-1.3250670	-2.7747730	0.2142780	H	1.7768710	3.0963870	0.0000000
H	-2.6596310	2.3038190	-0.0124180	H	-1.8203370	-2.5215890	0.0000000
H	-2.9233880	-2.1331250	0.1042800	H	-2.0540030	2.0451520	0.0000000
N	-0.6420930	1.9625440	0.0054530	N	0.8823640	2.6290260	0.0000000
N	-1.9300240	-2.0072440	-0.0289770	N	0.9118750	-1.4609770	0.0000000
N	-2.3515380	0.2561150	-0.0043000	N	2.0304220	0.6565940	0.0000000
S	2.5052920	-0.3552520	-0.0015160	O	3.1763310	-1.3351700	0.0000000
				S	-2.8302450	-0.2700590	0.0000000

${}^{th}\mathbf{C}$			(C1)	${}^{th}\mathbf{G}$			(C1)
C	0.4599750	1.6684260	0.0067130	C	1.2671410	1.4512830	-0.0015550
C	2.0791520	-0.0384470	-0.0063210	C	1.9385130	-0.9323750	-0.0019290
C	-0.2879710	-0.6795510	0.0125770	C	-0.0923090	0.9195290	-0.0121370
C	-0.6436110	0.7114320	-0.0000260	C	-0.2986230	-0.5105370	-0.0018220
C	-1.3805150	-1.5140880	0.0126000	C	-1.2626160	1.6405820	-0.0121330
C	-2.0082030	0.9031450	-0.0160570	C	-1.6385380	-0.8325960	0.0031270
H	0.9728690	3.6264000	0.1000660	H	2.7981030	-2.7388980	-0.0718160
H	1.3629240	-1.9613540	0.0127220	H	3.1862010	0.7435590	0.1005690
H	-0.7303400	3.3502650	0.2111470	H	3.8722060	-1.4781720	-0.4307830
H	-1.4081730	-2.5940230	0.0202150	H	-1.3768940	2.7153620	-0.0155460
H	-2.5632080	1.8305850	-0.0434670	H	-2.0816910	-1.8178050	0.0096090
N	0.1869860	3.0016990	-0.0133510	N	0.7413880	-1.4263050	-0.0156970
N	1.0519650	-0.9996450	0.0173150	N	2.2285750	0.4272360	0.0039090
N	1.7149000	1.3018460	0.0037900	N	3.0386240	-1.7656740	0.0650420
O	3.2429420	-0.4144730	-0.0244710	O	1.6091510	2.6268140	0.0132350
S	-2.8541560	-0.5944880	-0.0087170	S	-2.6258950	0.5914820	-0.0016020

	${}^{th}\mathbf{U}$		(Cs)
C	0.7754540	1.4256190	0.0000000
C	2.1065590	-0.7169530	0.0000000
C	-0.3366410	-0.8386410	0.0000000
C	-0.4337060	0.5922940	0.0000000
C	-1.5596770	-1.4592840	0.0000000
C	-1.7349300	1.0329700	0.0000000
H	1.0319470	-2.4342490	0.0000000
H	2.8244470	1.1943210	0.0000000
H	-1.7845250	-2.5162780	0.0000000
H	-2.0746830	2.0590460	0.0000000
N	0.9235430	-1.4297850	0.0000000
N	1.9553860	0.6723210	0.0000000
O	0.8036870	2.6469090	0.0000000
O	3.2042300	-1.2522470	0.0000000
S	-2.8406880	-0.2820660	0.0000000

### B3LYP-D3

All geometries were optimized at the B3LYP-D3/6-31++g(d,p) level of theory in the gas phase.

<b>A</b>			(C1)	<b>G</b>			(C1)
C	0.5459950	0.4434930	0.0006750	C	0.2514300	1.0151890	0.0076120
C	1.5652590	-1.5234920	-0.0025760	C	0.5616810	-0.3465360	0.0014960
C	-0.3621830	-0.6215430	-0.0057340	C	2.3796370	0.9137710	0.0030630
C	-1.0732400	1.9496310	-0.0015160	C	-1.1340500	1.4066970	0.0016850
C	-1.7290820	-0.2687430	-0.0097300	C	-1.5417800	-1.0693220	0.0000380
H	2.3851180	-2.2287280	-0.0023620	H	2.4872730	-1.2401880	-0.0057250
H	2.6746170	0.3183390	0.0073650	H	3.4340140	1.1523970	0.0019140
H	-1.3993860	2.9863250	-0.0002850	H	-2.1596440	-2.9785110	0.0968940
H	-2.4878830	-2.1786160	-0.0345570	H	-2.9570130	0.4470290	-0.0753450
H	-3.6757200	-0.8954330	-0.0352530	H	-3.4087960	-1.8548940	0.3626630
N	0.2497400	1.7499100	0.0033380	N	1.4005930	1.7826700	0.0077690
N	0.2922460	-1.8429940	-0.0078520	N	1.9306790	-0.3978270	-0.0003580
N	1.7852980	-0.1594360	0.0028190	N	-0.2755890	-1.4161760	0.0110070
N	-2.0600060	1.0357490	-0.0080790	N	-1.9683750	0.2353130	-0.0042870
N	-2.7146340	-1.1973880	-0.0116960	N	-2.5114920	-2.0441110	-0.0614600
				O	-1.6578840	2.5098150	-0.0056420

<b>C</b>			(C1)	<b><sup>th</sup>A</b>			(C1)
C	1.0853860	1.3364790	0.0038710	C	0.2622920	-1.7045360	-0.0128650
C	1.5400730	-0.0324840	0.0070340	C	0.3044290	0.6540940	0.0033170
C	-0.2628860	1.5201100	-0.0005250	C	1.6606490	0.3724030	0.0065950
C	-0.6273010	-0.9032480	0.0001030	C	-0.4973680	-0.5496990	-0.0045350
H	1.7705440	2.1744810	-0.0008570	C	-1.5623520	1.9274290	-0.0098510
H	3.1757130	-1.2456180	0.0752250	C	-1.9272400	-0.3488820	0.0002530
H	3.5538780	0.4486220	0.1211920	H	2.4782430	1.0787500	0.0121050
H	-0.7268340	2.5008580	-0.0034240	H	-0.0571450	-2.7373970	-0.0367980
H	-2.0982510	0.5546530	-0.0057940	H	-2.0415520	2.9035950	-0.0171300
N	0.7326240	-1.0786950	0.0078260	H	-2.4673940	-2.3297760	0.2215500
N	2.8774890	-0.2850810	-0.0097470	H	-3.7661730	-1.1989320	0.1125490
N	-1.0926350	0.4469440	-0.0024050	N	-0.2522590	1.9109210	-0.0003730
O	-1.4597100	-1.8004900	-0.0019690	N	-2.4340490	0.8704030	-0.0040880
				N	-2.7868930	-1.4064280	-0.0226430
				S	1.9412490	-1.3259870	-0.0051420

${}^{th}\mathbf{C}$ (C1)			${}^{th}\mathbf{U}$ (Cs)				
C	0.4600750	1.6682950	0.0064570	C	0.7738610	1.4255900	0.0000000
C	2.0790570	-0.0389590	-0.0072290	C	2.1059480	-0.7169910	0.0000000
C	-0.2878200	-0.6799490	0.0133100	C	-0.3367240	-0.8395440	0.0000000
C	-0.6427930	0.7110830	-0.0003960	C	-0.4336010	0.5911800	0.0000000
C	-1.3806730	-1.5138880	0.0134410	C	-1.5599170	-1.4596090	0.0000000
C	-2.0069200	0.9041320	-0.0174750	C	-1.7338060	1.0333630	0.0000000
H	0.9692170	3.6279650	0.1007720	H	1.0311680	-2.4350410	0.0000000
H	1.3611050	-1.9628950	0.0136300	H	2.8233380	1.1948430	0.0000000
H	-0.7323560	3.3477310	0.2147180	H	-1.7815980	-2.5174650	0.0000000
H	-1.4051250	-2.5940770	0.0217950	H	-2.0676020	2.0616060	0.0000000
H	-2.5578110	1.8341230	-0.0465420	N	0.9230100	-1.4307200	0.0000000
N	0.1849870	3.0012560	-0.0130470	N	1.9544770	0.6728580	0.0000000
N	1.0515340	-1.0009020	0.0185550	O	0.7995100	2.6468870	0.0000000
N	1.7152000	1.3020550	0.0030470	O	3.2038260	-1.2517370	0.0000000
O	3.2427560	-0.4149990	-0.0268670	S	-2.8414870	-0.2812380	0.0000000
S	-2.8548990	-0.5932430	-0.0094350				

${}^{th}\mathbf{G}$ (C1)			$\mathbf{U}$ (Cs)				
C	0.3120110	-0.8406780	-0.0012010	C	1.2050130	1.1102500	0.0000000
C	0.4798210	0.5943160	0.0125050	C	1.2909840	-0.3470320	0.0000000
C	1.5246000	-1.4952460	-0.0079940	C	-0.0054250	1.7123610	0.0000000
C	1.7949860	0.9923120	0.0130710	C	-1.2207610	-0.4052340	0.0000000
C	-0.6956230	1.4585060	0.0036070	H	0.0452810	-2.0041690	0.0000000
C	-1.9578660	-0.6725320	-0.0005270	H	2.1268640	1.6746720	0.0000000
H	1.6974910	-2.5618000	-0.0169460	H	-0.1245870	2.7901360	0.0000000
H	2.1756410	2.0041020	0.0185820	H	-2.0781150	1.4440020	0.0000000
H	-2.7319810	1.2698870	-0.1007440	N	0.0336620	-0.9903340	0.0000000
H	-3.2538220	-2.1976710	0.0643060	N	-1.1765220	0.9879750	0.0000000
H	-3.9665850	-0.7049340	0.4307110	O	2.3201710	-1.0070230	0.0000000
N	-0.9281420	-1.4583560	0.0116770	O	-2.2685740	-1.0305210	0.0000000
N	-1.8887450	0.7167980	-0.0037240				
N	-3.2354420	-1.1947500	-0.0682400				
O	-0.7210910	2.6825030	-0.0088040				
S	2.8450930	-0.3712680	-0.0009400				

## M06-2X

All geometries were optimized at the M06-2X/6-31++g(d,p) level of theory in the gas phase.

A (C1)			G (C1)				
C	1.2290150	-0.6040030	0.0038910	C	0.5283220	-0.8502890	-0.0012960
C	1.2895510	1.6989610	0.0003960	C	0.8538810	0.4991330	0.0094480
C	-0.1781810	-0.5176690	0.0024260	C	2.7069180	-0.5294130	0.0027620
C	-0.7074610	0.7703120	-0.0000620	C	-0.2160180	1.4652240	0.0034410
C	-2.2787590	-0.7777580	0.0008380	C	-1.6675720	-0.5675020	-0.0046470
H	1.3447720	-2.6501980	0.0071510	H	1.8513640	-2.5027730	-0.0125360
H	1.9221170	2.5824630	-0.0003690	H	3.7573110	-0.7830660	0.0014410
H	2.8767930	-1.8033730	0.0072920	H	-2.2687070	1.4107060	-0.0839300
H	-2.7650800	1.3111270	-0.0028330	H	-3.0668020	-1.9927070	0.0889400
H	-3.2764820	-1.1942840	0.0005350	H	-3.6758870	-0.4330970	0.3793340
N	1.8698770	-1.7910850	0.0063270	N	1.7271010	-1.5014830	-0.0041500
N	1.9442730	0.5268230	0.0028230	N	2.2181750	0.6787820	0.0107350
N	-0.0216930	1.9189000	-0.0011840	N	-0.6971200	-1.4395910	0.0055970
N	-1.1737730	-1.4741500	0.0029590	N	-1.4735410	0.7876280	-0.0062400
N	-2.0676610	0.5820590	-0.0010620	N	-2.9697970	-0.9993010	-0.0681450
				O	-0.1937600	2.6776040	-0.0020510

C <sup>th</sup> (C1)			U (Cs)				
C	1.0490500	1.1864600	0.0042570	C	0.0540730	-1.7062600	0.0000000
C	1.1279070	-0.2539720	0.0067890	C	1.2395880	-1.0695190	0.0000000
C	-0.2004350	1.7097120	0.0005140	C	1.2728220	0.3915600	0.0000000
C	-1.1822800	-0.5248430	0.0007980	C	-1.2291600	0.3605050	0.0000000
H	1.9298270	1.8138160	0.0015990	H	2.1799990	-1.6012270	0.0000000
H	2.3898360	-1.8454300	0.0510370	H	-0.0242170	2.0006210	0.0000000
H	3.1959910	-0.3028100	0.0785910	H	-0.0297690	-2.7869640	0.0000000
H	-0.3963250	2.7764400	-0.0018730	H	-2.0248700	-1.5093410	0.0000000
H	-2.2210220	1.2543860	-0.0040670	N	0.0004550	0.9870110	0.0000000
N	0.0828420	-1.0503630	0.0066190	N	-1.1391770	-1.0240420	0.0000000
N	2.3497360	-0.8392080	-0.0021840	O	2.2719120	1.0813370	0.0000000
N	-1.2778410	0.8909670	-0.0010480	O	-2.2900020	0.9485360	0.0000000
O	-2.2129250	-1.1705440	-0.0012290				

${}^{th}\mathbf{A}$ (C1)			${}^{th}\mathbf{G}$ (C1)				
C	0.0514390	0.4381090	-0.0035990	C	0.3106760	-0.8387480	-0.0022510
C	1.4681750	0.7223670	-0.0021850	C	0.4772550	0.5893190	0.0148660
C	1.8719930	-1.5393270	-0.0067260	C	1.5201180	-1.4870760	-0.0085080
C	-0.3028510	-0.9551460	0.0082820	C	1.7873110	0.9857760	0.0162590
C	-1.0452260	1.2717750	-0.0135230	C	-0.6996020	1.4548010	0.0033680
C	-1.6709880	-1.1368240	0.0127660	C	-1.9504370	-0.6728000	-0.0035010
H	1.3028700	2.7640690	0.2113650	H	1.6940150	-2.5528260	-0.0189850
H	2.6494650	-2.2992790	-0.0131210	H	2.1689600	1.9970770	0.0231690
H	2.9087210	2.1374970	0.1011190	H	-2.7314020	1.2590590	-0.1124210
H	-1.0899510	2.3516040	-0.0396690	H	-3.2445250	-2.1871780	0.0606100
H	-2.2085620	-2.0733820	0.0211010	H	-3.9445850	-0.6967560	0.4502620
N	0.6398440	-1.9565300	0.0059330	N	-0.9303210	-1.4586440	0.0091480
N	1.9167590	2.0037130	-0.0286100	N	-1.8848910	0.7133970	-0.0064250
N	2.3452820	-0.2523100	-0.0058440	N	-3.2293160	-1.1846580	-0.0715100
S	-2.4894970	0.3618340	-0.0022680	O	-0.7209890	2.6698510	-0.0098400
				S	2.8280780	-0.3694040	0.0011000

${}^{th}\mathbf{C}$ (C1)			${}^{th}\mathbf{U}$ (Cs)				
C	0.8786630	1.2733060	0.0028740	C	0.0027100	0.7287500	0.0000000
C	2.1128800	-0.7197880	-0.0070490	C	0.7597160	-0.4834680	0.0000000
C	-0.3270750	-0.8715400	0.0167140	C	0.7936400	1.8444140	0.0000000
C	-0.3957360	0.5568740	-0.0018720	C	2.1083790	-0.2640640	0.0000000
C	-1.5624250	-1.4626820	0.0176280	C	-1.2858520	-1.8102960	0.0000000
C	-1.6876700	1.0176700	-0.0214230	C	-1.4658290	0.6971190	0.0000000
H	0.0284760	3.1442150	0.1984490	H	0.4625910	2.8735700	0.0000000
H	1.0334900	-2.4566830	0.0203370	H	0.5843180	-2.5775870	0.0000000
H	1.7552560	3.0916030	0.0887710	H	2.9101780	-0.9877730	0.0000000
H	-1.8071060	-2.5143050	0.0294360	H	-2.9868620	-0.6894780	0.0000000
H	-2.0437250	2.0380940	-0.0536620	N	0.0862610	-1.6990260	0.0000000
N	0.8645160	2.6283870	-0.0174370	N	-1.9772350	-0.5994400	0.0000000
N	0.9196810	-1.4529840	0.0247840	O	-1.8610480	-2.8788150	0.0000000
N	2.0281840	0.6683100	-0.0005700	O	-2.1936750	1.6672390	0.0000000
O	3.1712380	-1.3149980	-0.0279610	S	2.4494730	1.4292010	0.0000000
S	-2.8088700	-0.2710290	-0.0098250				

### M06-2X-D3

All geometries were optimized at the M06-2X-D3/6-31++g(d,p) level of theory in the gas phase.

<b>A</b>			(Cs)	<b>G</b>			(Cs)
C	1.2290370	-0.6040360	0.0038890	C	0.5283250	-0.8502730	-0.0013090
C	1.2895820	1.6989820	0.0003960	C	0.8538870	0.4991730	0.0094380
C	-0.1781990	-0.5176860	0.0024250	C	2.7069660	-0.5294630	0.0027730
C	-0.7074840	0.7703220	-0.0000620	C	-0.2160260	1.4652980	0.0034290
C	-2.2788580	-0.7777490	0.0008390	C	-1.6676010	-0.5674750	-0.0046710
H	1.3449050	-2.6502680	0.0071510	H	1.8513050	-2.5028060	-0.0125430
H	1.9221420	2.5824820	-0.0003680	H	3.7573480	-0.7831260	0.0014630
H	2.8768720	-1.8033530	0.0072930	H	-2.2687140	1.4107850	-0.0839460
H	-2.7651170	1.3111610	-0.0028350	H	-3.0667100	-1.9927680	0.0889720
H	-3.2765690	-1.1942780	0.0005360	H	-3.6759200	-0.4331910	0.3793630
N	1.8699530	-1.7911150	0.0063280	N	1.7271010	-1.5015050	-0.0041550
N	1.9442920	0.5268210	0.0028230	N	2.2182170	0.6787510	0.0107480
N	-0.0216760	1.9189060	-0.0011840	N	-0.6971350	-1.4395670	0.0055770
N	-1.1738590	-1.4741410	0.0029580	N	-1.4735660	0.7876810	-0.0062660
N	-2.0677110	0.5820750	-0.0010600	N	-2.9698030	-0.9993500	-0.0681360
				O	-0.1938040	2.6776880	-0.0020340

<b>C</b>			(Cs)	<b>U</b>			(Cs)
C	1.0490610	1.1865010	0.0042580	C	0.0540760	-1.7062650	0.0000000
C	1.1279360	-0.2539640	0.0067920	C	1.2396080	-1.0695240	0.0000000
C	-0.2004510	1.7097350	0.0005120	C	1.2728390	0.3915740	0.0000000
C	-1.1822910	-0.5248530	0.0007980	C	-1.2291810	0.3605150	0.0000000
H	1.9298180	1.8139110	0.0016010	H	2.1800170	-1.6012560	0.0000000
H	2.3898100	-1.8454770	0.0510320	H	-0.0242100	2.0006560	0.0000000
H	3.1960440	-0.3028910	0.0785930	H	-0.0297710	-2.7869680	0.0000000
H	-0.3963550	2.7764670	-0.0018750	H	-2.0248880	-1.5093570	0.0000000
H	-2.2210420	1.2543910	-0.0040680	N	0.0004530	0.9870370	0.0000000
N	0.0828530	-1.0503580	0.0066180	N	-1.1391940	-1.0240460	0.0000000
N	2.3497570	-0.8392440	-0.0021810	O	2.2719520	1.0813330	0.0000000
N	-1.2778590	0.8909700	-0.0010530	O	-2.2900470	0.9485190	0.0000000
O	-2.2129220	-1.1705790	-0.0012230				

${}^{th}\mathbf{A}$ (C1)			${}^{th}\mathbf{G}$ (Cs)				
C	-1.0522028	1.2684543	-0.0137144	C	0.3106860	-0.8387200	-0.0022670
C	0.0466898	0.4377263	-0.0037903	C	0.4772770	0.5893800	0.0148530
C	-1.6715151	-1.1418109	0.0125741	C	1.5201060	-1.4871000	-0.0085100
C	-0.3038691	-0.9564713	0.0080905	C	1.7873770	0.9857580	0.0162650
C	1.4626607	0.7257758	-0.0023763	C	-0.6996310	1.4548790	0.0033690
C	1.8725309	-1.5348286	-0.0069175	C	-1.9504690	-0.6727810	-0.0035220
H	-1.0998184	2.3481592	-0.0398609	H	1.6939050	-2.5528650	-0.0189840
H	1.2918902	2.7670288	0.2111732	H	2.1690920	1.9970350	0.0231860
H	2.8994134	2.1447575	0.1009270	H	-2.7314210	1.2591360	-0.1124430
H	2.6520342	-2.2926970	-0.0133128	H	-3.2444030	-2.1872400	0.0606580
H	-2.2065800	-2.0798054	0.0209098	H	-3.9445970	-0.6968550	0.4502850
N	0.6415029	-1.9553289	0.0057410	N	-0.9303300	-1.4586150	0.0091320
N	2.3423731	-0.2465493	-0.0060356	N	-1.8849310	0.7134350	-0.0064670
N	1.9078126	2.0083181	-0.0288020	N	-3.2293160	-1.1847170	-0.0715110
S	-2.4940329	0.3546503	-0.0024598	O	-0.7211290	2.6699360	-0.0098220
				S	2.8281330	-0.3694790	0.0011160

${}^{th}\mathbf{C}$ (Cs)			${}^{th}\mathbf{U}$ (Cs)				
C	0.8786260	1.2733680	0.0028820	C	0.0015490	0.7294550	0.0000000
C	2.1128840	-0.7197610	-0.0070530	C	0.7584270	-0.4818460	0.0000000
C	-0.3271200	-0.8715600	0.0167170	C	0.7944230	1.8445970	0.0000000
C	-0.3958350	0.5569030	-0.0018650	C	2.1071880	-0.2657890	0.0000000
C	-1.5624670	-1.4627480	0.0176180	C	-1.2846780	-1.8094680	0.0000000
C	-1.6878300	1.0176050	-0.0214310	C	-1.4653680	0.6963440	0.0000000
H	0.0287510	3.1445040	0.1984940	H	0.4642110	2.8737430	0.0000000
H	1.0335700	-2.4566820	0.0203740	H	0.5867130	-2.5729680	0.0000000
H	1.7555150	3.0914760	0.0887470	H	2.9070720	-0.9918450	0.0000000
H	-1.8070850	-2.5143850	0.0294180	H	-2.9846350	-0.6887610	0.0000000
H	-2.0439650	2.0380020	-0.0536700	N	0.0863910	-1.6974740	0.0000000
N	0.8646580	2.6284690	-0.0174270	N	-1.9765340	-0.5997910	0.0000000
N	0.9196830	-1.4529780	0.0247950	O	-1.8599980	-2.8802610	0.0000000
N	2.0281550	0.6683430	-0.0005800	O	-2.1976150	1.6662470	0.0000000
O	3.1712520	-1.3149450	-0.0279720	S	2.4496180	1.4281640	0.0000000
S	-2.8090150	-0.2711610	-0.0098480				

$\omega$ B97XD

All geometries were optimized at the  $\omega$ B97XD/6-31++g(d,p) level of theory in the gas phase.

<b>A</b>			(C1)	<b>G</b>			(C1)
C	1.2266030	-0.6062960	0.0038950	C	0.5292770	-0.8497180	-0.0006160
C	1.2896080	1.6975800	0.0004050	C	0.8525850	0.4991190	0.0085830
C	-0.1787860	-0.5173550	0.0024230	C	2.7067450	-0.5299700	0.0013610
C	-0.7079850	0.7689410	-0.0000620	C	-0.2154640	1.4632750	0.0035380
C	-2.2808910	-0.7754910	0.0008250	C	-1.6680690	-0.5667220	-0.0027890
H	1.3534480	-2.6524730	0.0071590	H	1.8530580	-2.5015970	-0.0115830
H	1.9221160	2.5808300	-0.0003550	H	3.7571110	-0.7835010	-0.0005530
H	2.8785940	-1.7997410	0.0072940	H	-2.2659680	1.4122350	-0.0772440
H	-2.7644830	1.3115670	-0.0028390	H	-3.0692500	-1.9941890	0.0919200
H	-3.2788700	-1.1909080	0.0005170	H	-3.6828800	-0.4320570	0.3626060
N	1.8726310	-1.7910120	0.0063350	N	1.7275080	-1.5019030	-0.0040940
N	1.9430640	0.5257110	0.0028310	N	2.2191920	0.6787520	0.0088230
N	-0.0223920	1.9174580	-0.0011800	N	-0.6952900	-1.4381770	0.0067540
N	-1.1771600	-1.4740800	0.0029500	N	-1.4730820	0.7880010	-0.0045930
N	-2.0681880	0.5833910	-0.0010700	N	-2.9673490	-1.0020730	-0.0614100
				O	-0.1942540	2.6783790	-0.0020010

<b>C</b>			(C1)	<b>U</b>			(Cs)
C	1.0477700	1.1860530	0.0046790	C	0.0535190	-1.7062070	0.0000000
C	1.1297540	-0.2523430	0.0074410	C	1.2380120	-1.0677580	0.0000000
C	-0.2013550	1.7097290	0.0004010	C	1.2720880	0.3911580	0.0000000
C	-1.1815710	-0.5237800	0.0009440	C	-1.2283930	0.3599500	0.0000000
H	1.9268530	1.8164250	0.0024160	H	2.1773240	-1.6016300	0.0000000
H	2.3901700	-1.8458840	0.0499070	H	-0.0234980	1.9992050	0.0000000
H	3.1978160	-0.3067240	0.0763860	H	-0.0292790	-2.7866690	0.0000000
H	-0.3967600	2.7761440	-0.0023120	H	-2.0223450	-1.5107030	0.0000000
H	-2.2195250	1.2561010	-0.0047490	N	0.0006550	0.9871770	0.0000000
N	0.0825000	-1.0488550	0.0070950	N	-1.1390380	-1.0238920	0.0000000
N	2.3504530	-0.8404940	0.0001230	O	2.2733970	1.0821930	0.0000000
N	-1.2786010	0.8905830	-0.0013530	O	-2.2907900	0.9493910	0.0000000
O	-2.2131430	-1.1723440	-0.0011730				

${}^{th}\mathbf{A}$ (C1)			${}^{th}\mathbf{G}$ (C1)				
C	0.0502110	0.4375580	-0.0017850	C	0.3115470	-0.8378840	-0.0014130
C	1.4650330	0.7237310	0.0004280	C	0.4772730	0.5899530	0.0136300
C	1.8719510	-1.5376840	-0.0057300	C	1.5196960	-1.4874930	-0.0085390
C	-0.3038760	-0.9551280	0.0071090	C	1.7882900	0.9857920	0.0143860
C	-1.0476310	1.2707250	-0.0102230	C	-0.6985590	1.4532030	0.0040590
C	-1.6711830	-1.1384770	0.0104650	C	-1.9509100	-0.6720350	-0.0020960
H	1.3090570	2.7705540	0.1962400	H	1.6931130	-2.5534440	-0.0180790
H	2.6492350	-2.2975840	-0.0119670	H	2.1690950	1.9971770	0.0204000
H	2.9115870	2.1354270	0.0981080	H	-2.7281660	1.2608470	-0.1052070
H	-1.0903010	2.3507130	-0.0327500	H	-3.2455380	-2.1890780	0.0642730
H	-2.2076420	-2.0757880	0.0167220	H	-3.9521120	-0.6960780	0.4338940
N	0.6391000	-1.9560730	0.0044340	N	-0.9288190	-1.4572140	0.0101970
N	1.9194600	2.0021980	-0.0210330	N	-1.8838880	0.7134640	-0.0047550
N	2.3436120	-0.2523590	-0.0032250	N	-3.2262980	-1.1877940	-0.0666900
S	-2.4911420	0.3603590	-0.0017740	O	-0.7229640	2.6708340	-0.0082070
				S	2.8285860	-0.3690620	-0.0005130

${}^{th}\mathbf{C}$ (C1)			${}^{th}\mathbf{U}$ (Cs)				
C	0.8747420	1.2746040	0.0046050	C	0.0015490	0.7294550	0.0000000
C	2.1101920	-0.7186990	-0.0055980	C	0.7584270	-0.4818460	0.0000000
C	-0.3275670	-0.8699600	0.0160180	C	0.7944230	1.8445970	0.0000000
C	-0.3978340	0.5572560	-0.0003300	C	2.1071880	-0.2657890	0.0000000
C	-1.5606810	-1.4650160	0.0160700	C	-1.2846780	-1.8094680	0.0000000
C	-1.6913720	1.0153650	-0.0188260	C	-1.4653680	0.6963440	0.0000000
H	0.0363810	3.1546710	0.1876710	H	0.4642110	2.8737430	0.0000000
H	1.0293680	-2.4544030	0.0186940	H	0.5867130	-2.5729680	0.0000000
H	1.7614230	3.0888460	0.0859360	H	2.9070720	-0.9918450	0.0000000
H	-1.8008260	-2.5179020	0.0261150	H	-2.9846350	-0.6887610	0.0000000
H	-2.0477300	2.0356450	-0.0481930	N	0.0863910	-1.6974740	0.0000000
N	0.8691540	2.6293980	-0.0124960	N	-1.9765340	-0.5997910	0.0000000
N	0.9185240	-1.4519070	0.0234310	O	-1.8599980	-2.8802610	0.0000000
N	2.0249520	0.6669680	0.0015040	O	-2.1976150	1.6662470	0.0000000
O	3.1708920	-1.3148120	-0.0259640	S	2.4496180	1.4281640	0.0000000
S	-2.8098410	-0.2756030	-0.0094380				