A COMPUTATIONAL STUDY

ON BIOCHEMICAL REACTIONS

OF SULFUR-CONTAINING COMPOUNDS

XI YU

(B. Appl. Sc. in Applied Chemistry (Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

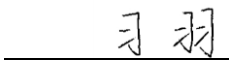
DEPARTMENT OF CHEMISTRY
NATIONAL UNIVERSITY OF SINGAPORE

2016

Thesis Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety, under the supervision of Professor Wong Ming Wah, Richard, Department of Chemistry, National University of Singapore. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Xi Yu		04-Aug-2016
Name	Signature	Date

Acknowledgements

I would like to express my sincere appreciations to my supervisor Prof. Wong Ming Wah, Richard, for his great help and financial support during my graduate study. During the hard times of research, his encouragement, understanding and patience really helped me a lot to overcome all the difficulties and move forward.

I thank NUS and the Department of Chemistry for the financial support and their help in various aspects of my graduate study. I also thank all my seniors, Dr. Yang Hui, Dr. Xie Huifang, Dr. Vu Viet Cuong, Dr. Cao Ye, Dr. Kee Choon Wee, Dr. Zhou Yujing from my laboratory, as well as Dr. Wang Chunyan, Dr. Xu Gangqin, Dr. Xie Huihua, Dr. Zhu Liang, Dr. Tan Shu Fen from NUS. I learnt a lot from them through group meetings and discussions.

I thank all my other laboratory colleagues, Chen Weiyan, Yu Zongrong, Guo Jinlong, Ang Shi Jun, and all my church friends, for their help in my research and various aspects of my life.

I thank my collaborators, Assoc. Prof. Huang Dejian and his student Liang Dong, for their experimental support and feedback to one of my projects.

I thank my beloved parents and grandmother, for their great love and continuous support in my research and life throughout my graduate study.

I thank everyone else who has been kind to me or cared about me in my life.

Last but not least, I would like to thank and praise the God, the Almighty, for without His graces and blessings, the completion of this thesis would have never been possible. Thank you, God.

Table of Content

Thesis Declaration ..... **i**

Acknowledgements ..... **iii**

Table of Content ..... **v**

Summary ..... **xi**

List of Tables ..... **xv**

List of Figures ..... **xvii**

Lists of Abbreviations ..... **xxiii**

Chapter 1 Introduction **1**

1.1 General Introduction 1

1.2 DNA Methylation and DNA Methyltransferases 2

1.2.1 Epigenetics, DNA Methylation and DNA Methyltransferases *2*

1.2.2 Molecular Modeling of Human DNMT1 Structures and Objectives of the Study *6*

1.2.3 Mechanistic Study on DNA Methylation and Objectives of the Study *9*

1.3 H_2S Release from Organosulfur Compounds 17

1.3.1 H_2S in Biological Systems *17*

1.3.2 Organosulfur Compounds as H_2S Donors *18*

1.3.3 Objectives of the Study *23*

1.4 References 24

Chapter 2 Theoretical Methodology **33**

2.1 The Schrödinger Equation 33

2.2 Born−Oppenheimer Approximation 34

2.3 Potential Energy Surface and Thermochemistry Corrections.....35

2.4 Hartree−Fock Theory.....37

2.4.1 Variational Method.....*37*

2.4.2 Hartree−Fock Approximation.....*38*

2.4.3 Basis Set.....*39*

2.4.4 The Hartree−Fock Equation.....*41*

2.5 Perturbation Theory.....43

2.6 Coupled-Cluster Theory.....44

2.7 Density Functional Theory.....46

2.7.1 Hohenberg−Kohn Theorems.....*46*

2.7.2 Kohn−Sham Theorem.....*47*

2.7.3 The Exchange-Correlation Functionals.....*49*

2.8 Solvation Methods.....52

2.9 Homology Modeling.....54

2.10 Classical Molecular Dynamics.....56

2.10.1 System Setup and Initialization.....*57*

2.10.2 Force Fields and Force Calculation.....*59*

2.10.3 Integration Algorithms.....*60*

2.11 References.....62

Chapter 3 Homology Modeling of hDNMT1 with Molecular Dynamics Simulations.....69

3.1 Introduction.....69

3.2 Computational Methods.....70

3.2.1 Homology Modeling.....*70*

3.2.2 Molecular Dynamics *71*

3.3 Results and Discussions 74

3.3.1 Selection and Quality of the Unrefined Model *74*

3.3.2 Evaluation of the Overall MD Performance and Analysis of the MD Trajectory *76*

3.3.3 Examination of the Active Site Interactions and the Two Critical Distances *82*

3.3.4 Evaluation of the Final Homology Model *85*

3.4 Conclusion 88

3.5 References 89

Chapter 4 Mechanism Study on DNA Methylation Based on Simple Model Systems 93

4.1 Introduction 93

4.2 Computational Methods 95

4.3 Results and Discussion 96

4.3.1 Intermediates from the Nucleophilic Addition of Cys-S to Cytosine-C6 96

4.3.2 Transition States from the Methylation of Cytosine-C5 by SAM *107*

4.3.3 Performance of the Small Model System and Possible Improvements . *122*

4.4 Conclusion 124

4.5 References 126

Chapter 5 Conformational Study on Sulfur-Containing Compounds in H_2S Releasing Reactions 129

5.1 Introduction 129

5.1.1 Modeling of GSH by MeSH and Assumptions on the Nucleophilic Substitutions.....*129*

5.1.2 Classification of the Reactions between MeSH and DADS/DATS.....*131*

5.1.3 Classification of the Equilibrium Structures from the Reactions of MeSH and DADS/DATS and Definition of Torsional Angles.....*132*

5.1.4 Classification of the Equilibrium Structures in the Hypothetical Cα Nucleophilic Substitution of DMDS/DPDS by MeSH and Definition of Torsional Angles.....*134*

5.2 Computational Methods.....135

5.3 Results & Discussion.....136

5.3.1 Conformational Analysis of Reactants and Products in the Reaction of MeSH and DADS/DATS.....*136*

5.3.2 Conformational Study on Reactants and Products from the Hypothetical Cα Nucleophilic Substitution of DMDS/DPDS by MeSH.....*158*

5.4 Conclusion.....164

5.5 References.....165

Chapter 6 Computational Study of Reaction Mechanisms on H_2S Releasing Reactions from Organosulfur Compounds.....169

6.1 Introduction.....169

6.2 Computational Methods.....170

6.3 Results & Discussion.....172

6.3.1 Transition State Study in Reaction of MeSH and DADS/DATS.....*172*

6.3.2 Energy Profiles from the Reaction of MeSH and DADS/DATS and MeSH

.....*204*

6.3.3 Effects of Temperature and Basis Sets Tested on the Cα Nucleophilic Substitution of DADS by MeSH *207*

6.3.4 Computational Study on the Cα Nucleophilic Substitution of DADS by GSH *209*

6.3.5 Transition State Study on the Hypothetical Cα Nucleophilic Substitution of DMDS or DPDS by MeSH *222*

6.4 Conclusion 228

6.5 References 231

Chapter 7 Conclusions and Outlook 235

7.1 Study on the Structure of Human DNMT1 and DNA Methylation Mechanism 235

7.2 Study on H_2S Releasing Reactions from the Reaction of GSH and Organosulfur Compounds 237

7.3 References 241

Chapter 8 Appendix 243

List of Tables 243

List of Figures 245

8.1 References 273

Summary

This thesis reports on the computational studies of biochemical reactions in two sulfur-containing systems. In the first system (Chapters 3 and 4), DNA is methylated by DNA methyltransferases (DNMTs) via the action of the Cys-thiolate and transfers a methyl group from S-adenosyl-$_L$-methionine (SAM) to the C5 of cytosine. In the second system (Chapters 5 and 7), H_2S is released from the reaction of glutathione (GSH) and garlic-derived organosulfur compounds.

Chapter 1 introduces the thesis with the objectives of the study on these two systems. Chapter 2 gives an overview of various theoretical methods and models used in this thesis.

Chapter 3 presents the establishment of a 3D model of the human DNMT1 (hDNMT1)-DNA-SAH complex via homology modeling and MD simulations based on the structure of the mouse DNMT1 (mDNMT1)-DNA-SAH complex. The MD simulations showed that the important H-bonds in this homology model at the active site are similar to that observed in the mDNMT1-DNA complex. The overall structure of this final hDNMT1-DNA-SAH complex resembles that of the mDNMT1-DNA-SAH complex.

Chapter 4 presents a mechanistic study on the first two steps in the DNA methylation reaction—the Cys nucleophilic attack and the methyl transfer—at the (PCM/)M06-2X/6-31G* level based on small model systems. The results showed that the intermediate 1 (I1) is stable with the Glu side chain through H-bonds with cytosine-N4-H and via direct protonation of cytosine-N3, or the

Arg side chain through the interaction with cytosine-O2 or N3. The I1 is able to react with SAM with or without the surrounding Cys side chain to increase the nucleophilicity of cytosine-C5, the Glu side chain to stabilize the TS2 through the H-bond with cytosine-N3 or via direct protonation of it, or the Arg side chain that possibly destabilizes the TS2 through the interaction with cytosine-O2. The solvent effects are crucial for calculating the TS1 and TS2 structures, but not for the I1 structures.

Chapter 5 reports on the conformational analysis of the reactants and products in H_2S release at the SMD/M06-2X/6-31+G* level. GSH was simplified to MeSH and the S_N2 mechanism was assumed. Some similarities in the conformations of the allyl-containing compounds were identified. They are due to the resonance stabilization by hyperconjugations. Some similarities in the conformations of the alkyl-containing compounds were also presented.

Chapter 6 presents a computational study on the transition states (TSs) and the energy profiles of the H_2S releasing reactions.

The TSs from the reaction of GSH (modeled as MeSH) and DADS/DATS were studied. The conformation of the forming products in all TS usually resembles the most stable forms of products. In the S nucleophilic substitutions, the TS conformations with the lowest ΔH_{298} have the two ending substituent groups pointing to the same direction. Moreover, the C-H(MeS^- or AS^-)$\cdots\pi$(-CH=CH_2), LP(S)(DADS or MeSSA)$\cdots\pi$(-CH=CH_2) and C-S(DATS)$\cdots\pi$(-CH=CH_2) interactions were observed in the TSs.

The Cα nucleophilic substitutions are always slower than the S nucleophilic substitutions due to the higher coordination number at Cα than S

in the TS, and the additional requirement for re-hybridization of Cα from sp^3 to sp^2. These results agree with the experimental results from Liang *et al.* The TSs from the Cα nucleophilic substitution of DADS by the full reactant GSH were also studied.

The Cα nucleophilic substitution of DMDS/DPDS by GSH (modeled as MeSH) was studied. The Cα nucleophilic substitutions on DMDS or DPDS are slower than that on DADS possibly due to the increased stabilization of the S_N2 transition states by π-conjugation in DADS. These results are consistent with the unpublished experimental results from Liang *et al.* The Cα nucleophilic substitution on DPDS is slower than that on DADS because the primary carbon in the methyl is more reactive than the secondary carbon in the propyl group.

Chapter 7 summarizes the whole thesis and suggests the possible future work. Chapter 8 is the Appendix.

List of Tables

Table 1.1 Reported molecular modeling studies of hDNMT1 catalytic domain with the substrate and SAM or SAH[a]

Table 1.2 Summary of key findings from computational studies on DNA methylation

Table 4.1 Key geometric parameters for TS from methyl transfer[a]

Table 5.1 Reaction steps[1] studied in H_2S release by reaction of MeSH and (a) DADS or (b) DATS, and categorization[a]

Table 5.2 List of reactants and products from the reaction of MeSH and DADS/DATS and categorization[a]

Table 5.3 List of reactants and products from the Cα Nucleophilic Substitution of DMDS/DPDS by MeSH and categorization

Table 5.4 Top 16 of the calculated DATS conformations and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

Table 5.5 Geometric information on the crystal structures[10d] of DAS and DADS[a]

Table 5.6 Donor-Acceptor interactions in the optimized conformers of MeSSA from the NBO analysis[a]

Table 5.7 Top 27 of the calculated DPDS conformations and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

Table 6.1 Calculated TS conformations from the reaction of MeS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.2 Calculated TS conformations from the reaction of MeS^- and (a) DADS or (b) MeSSA via S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.3 Calculated TS conformations from the reaction of MeS^- and (a) DATS or (b) MeS_3A via side-S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.4 Calculated TS conformations from the reaction of MeS^- and DATS via mid-S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.5 Calculated TS conformations from the reaction of MeS^- and (a) ASSH or (b) MeSSH to release H_2S and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.6 Calculated TS conformations from the reaction of AS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.7 Interaction analysis for (a) C-H···π interaction, (b) LP(S)···π interaction

Table 6.8 Activation barriers ($\Delta H^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔH_{298}, kJ mol^{-1}) from the reaction of MeSH and (a) DADS or (b) DATS[a]

Table 6.9 Temperature effects (in K) on the activation barriers ($\Delta H^{\ddagger}$ or $\Delta G^{\ddagger}$, in kJ mol^{-1}) tested on the Cα nucleophilic substitution of DADS by MeSH from single point energy calculations on the SMD/M06-2X/6-31+G* optimized geometries

Table 6.10 Basis set effects on the activation barriers ($\Delta H^{\ddagger}_{298}$ or $\Delta G^{\ddagger}_{298}$, in kJ mol^{-1}) tested on the Cα nucleophilic substitution of DADS by MeSH from single point energy calculations on the SMD/M06-2X/6-31+G* optimized geometries

Table 6.11 The calculated GSH conformations in this study and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

Table 6.12 Calculated TS conformations from the reaction of GSH and DADS via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.13 The comparison between the full TS and the small TS models in terms of their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.14 Interaction analysis for (a) C-H···O interaction, (b) C-H···N interaction, and (c) π···π stacking[a]

Table 6.15 Calculated TS conformations from the reaction of MeS^- and (a) DMDS or (b) DPDS via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

Table 6.16 Activation barriers ($\Delta H^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔH_{298}, kJ mol^{-1}) comparisons for the reaction of MeS^- and (a) diallyl/dialkyl disulfides or (b) S-allyl/alkyl-methyl disulfides.

List of Figures

Figure 1.1 (a) DNA methylation occurs at the C5 of cytosine. B stands for base. (b) Structure of SAM. (c) Structure of SAH.[13]

Figure 1.2 (a) Structure of the M.HhaI with SAM and the modified dsDNA (6MHT)[18]. The substrate 4'-thio-2'-deoxycytidine is boxed. The protein is shown by the cartoon model, SAM is shown by the CPK model and colored by atom, and the nucleotides are shown by the stick model and colored by atom. (b) Arrangement of Cys81, Glu119, Arg163, Arg165, Phe79, SAM and the substrate 4'-thio-2'-deoxycytidine (sC) in the active center. Two alternative positions of the Cys-S and the transferring methyl-C are shown. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

Figure 1.3 Interaction of the flipped-out cytosine nucleotide (fC or fC' with 5-methyl) with the active site residues of (a) M.HhaI (1MHT)[27a] and (b) mDNMT1 (4DA4)[28c]. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

Figure 1.4 Proposed DNA methylation mechanism[36a, 37c, 38].

Figure 1.5 H_2S production from garlic-derived polysulfides[65c]. (a) Proposed mechanism of H_2S production from the reaction of GSH and DADS via Cα nucleophilic attack (red) but not by thiol/disulfide exchange (blue). (b)&(c) H_2S production by thiol/disulfide exchange of organic trisulfides.

Figure 1.6 H_2S production from DATS[65a]. Proposed mechanism of H_2S production from the reaction of GSH and DATS via the allylic sulfur nucleophilic attack (pink) or the central sulfur nucleophilic attack (blue).

Figure 1.7 DAS formation via α carbon nucleophilic attack[65a]. (a) DAS formation from the reaction of DADS and ASH. (b) DAS formation from the reaction of GSSA and ASH.

Figure 3.1 The 3D structures of (a) mDNMT1 (4DA4)[1] (light blue) and (b) homology model of hDNMT1 (4DAX-i) (red). The proteins are shown by the cartoon model, the Zn(II) ion and SAH are shown by the CPK model and colored by atom. The nucleotides are shown by the stick model and colored by atom.

Figure 3.2 Interaction of the flipped-out methylated cytosine nucleotide with the active site residues of (a) mDNMT1 (4DA4)[1] and (b) homology model of hDNMT1 (4DAX-i). The heavy atom distances are indicated by the dotted lines. Distances are in Å.

Figure 3.3 Plots of (a) pressure (in bar); (b) temperature (in K); and (c) density (in g cm^{-3}); (d) energies (in kcal mol^{-1}) against time (in ps).

Figure 3.4 Structural drift during the production phase for (a) protein backbone; (b) DNA backbone; (c) SAH with reference to the model structure right after the equilibration. Time is in ps, and RMSD is in Å.

Figure 3.5 RMSF of hDNMT1 based on protein Cα. RMSF is in Å.

Figure 3.6 Possible hydrogen bonds between the cytosine nucleotide and residues in the active site of hDNMT1.

Figure 3.7 The structures of SAH in (a) mDNMT1 (4DA4)[1] and (b) homology model of hDNMT1 (from the MD production phase). SAH is shown by the stick model and colored by atom. The major differences are boxed.

Figure 3.8 The 3D structures of (a) mDNMT1 (4DA4)[1] (light blue) and (b) final homology model of hDNMT1 (4DAX-f) (cyan). The proteins are shown by the cartoon model, the Zn(II) ion and SAH are shown by the CPK model and colored by atom. The nucleotides are shown by the stick model and colored by atom.

Figure 3.9 Interaction of the flipped-out cytosine nucleotide with the active site residues of (a) mDNMT1 (4DA4)[1] and (b) final homology model of hDNMT1 (4DAX-f). Hydrogens were omitted for comparison purpose. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

Figure 4.1 A small-truncated model to represent the active site of M.HhaI.

Figure 4.2 Relative orientation of A1)–2) Cys-thiolate and B1)–2) SAM-methyl to cytosine in the crystal structures of the M.HhaI-DNA complex (6MHT)[5] and the mDNMT1-DNA complex (4DA4)[6].

Figure 4.3 Intermediate 1 (I1) from the Cys-thiolate nucleophilic addition to the cytosine-C6. A1)–D2) I1 formed from cytosine, Cys with or without Glu or Arg. E) Atom numbering for reference. The methods (PCM/)M06-2X, DF-MP2 and CCSD are indicated in the brackets for each intermediate. Except for A3), all other methods used the 6-31G* basis set. The interactions are indicated by the dotted lines. Distances are in Å.

Figure 4.4 TS1 from the Cys-thiolate nucleophilic addition to the cytosine-C6. A1)–E) TS1 formed from cytosine, Cys with or without Glu or Arg. The methods M06-2X and PCM/M06-2X are indicated in the brackets for each TS. The breaking/forming bonds are indicted by

the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

Figure 4.5 TS of the methyl transfer from the sulfonium ion of the model SAM to cytosine-C5. A1)–G2) TS formed from cytosine, SAM with or without Cys, Glu or Arg. H) atom numbering for reference. The methods M06-2X and PCM/M06-2X are indicated in the brackets for each transition state. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

Figure 4.6 Constrained and optimized geometries of (a) preTS2-CysCytosine-GluH-Arg-SAM and (b) TS2-CysCytosine-GluH-Arg-SAM at the M06-2X/6-31G* level.

Figure 5.1 Molecular structure of GSH. (a) GSH in the fully protonated form, consisting of three amino acid residues: Glu (red), Cys (black) and Gly (blue). (b) The expected most abundant form of GSH at pH=7.4.

Figure 5.2 Torsional angles in DATS.

Figure 5.3 Torsional angles in DPDS.

Figure 5.4 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers for: A1)−3) diallyl (poly)sulfides; B1)−3) diallyl (poly)sulfides; C1)−3-2) S-allyl-methyl (poly)sulfides; D) bisulfide; E1)−3) methyl (poly)sulfides; F) dimethyl disulfide. Interactions are indicated by the dotted lines and distances are in Å. Torsional angles (in ^{o}) are given in the brackets. The conformations shown in C2-2) and C3-2) correspond to the respective enantiomers of the optimized conformers for comparison purpose.

Figure 5.5 Crystal structures[10d] of (a) DAS and (b) DADS. Torsional angles (in ^{o}) are given in the brackets.

Figure 5.6 Optimized geometry of DADS37. Torsional angles (in ^{o}) are given in the brackets.

Figure 5.7 The lowest energy conformations of the -SS-allyl group.

Figure 5.8 Structures of MeSSA. (a)−(f) SMD/M06-2X/6-31+G* optimized conformers of MeSSA and their relative energies (ΔE). (g) Chemical structure of MeSSA with labeling of atoms and torsional angles. Interactions are indicated by the dotted lines and distance are in Å. Torsional angles (in ^{o}) are given in the brackets.

Figure 5.9 Visualization of NBO donor-acceptor interactions for: A1)−2) τ_3~120^{o}, B1)−2) τ_2~-60^{o}, C1)−2) τ_2~180^{o}, D) τ_1~90^{o}.

Figure 5.10 The stable torsional angles presented by Newman projection. (a)−(b) τ_3 close to 120^o or -120^o. (c)−(d) τ_2 close to 60^o or -60^o. (e) τ_2 close to 180^o. (f)−(g) τ_1 close to 90^o or -90^o.

Figure 5.11 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers for: A1)−3-2) dialkyl (poly)sulfides; B) alkyl polysulfides. Torsional angles (in ^{o}) are given in the brackets.

Figure 5.12 Optimized geometries of (a) DPDS8, (b) DPDS37, and (c) DPDS3. Torsional angles (in ^{o}) are given in the brackets.

Figure 5.13 The lowest energy conformations of the -SS-propyl group.

Figure 6.1 Chemical structure of the TS in reaction of MeS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.2 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DADS, or (b) MeSSA via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in ^{o}.

Figure 6.3 Chemical structure of the TS in reaction of MeS^- and (a) DADS or (b) MeSSA via S nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.4 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DADS, or (b) MeSSA via S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in ^{o}.

Figure 6.5 Relative orientation of different substituent groups in transition states from the S nucleophilic substitution.

Figure 6.6 Chemical structure of the TS in reaction of MeS^- and (a) DATS or (b) MeS_3A via side-S nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.7 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DATS, or (b) MeS_3A via side-S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in ^{o}.

Figure 6.8 Chemical structure of the TS in reaction of MeS^- and DATS via mid-S nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.9 SMD/M06-2X/6-31+G* optimized geometry of the lowest energy TS in reaction of MeS^- and DATS via mid-S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

Figure 6.10 Chemical structure of the TS in reaction of MeS^- and (a) ASSH or (b) MeSSH to release H_2S with labeling of atoms and torsional angles.

Figure 6.11 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) ASSH, or (b) MeSSH to release H_2S. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in °.

Figure 6.12 Chemical structure of the TS in reaction of AS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.13 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of AS^- and (a) DADS, or (b) MeSSA via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

Figure 6.14 The ionization state of GSH in the reaction.

Figure 6.15 SMD/M06-2X/6-31+G* optimized geometries of GSH conformations. Interactions are indicated by the dotted lines and distances are in Å.

Figure 6.16 SMD/M06-2X/6-31+G* optimized geometries of TS in reaction of GSH and DADS via Cα nucleophilic substitution and the superposition of GSH in these TS with different colorings in carbons. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

Figure 6.17 Intramolecular interactions in the SMD/M06-2X/6-31+G* optimized geometries of TS in reaction of GSH and DADS via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

Figure 6.18 Chemical structure of the TS in reaction of MeS^- with (a) DMDS or (b) DPDS via Cα nucleophilic substitution with labeling of atoms and torsional angles.

Figure 6.19 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DMDS, or (b) DPDS via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in ^{o}.

Lists of Abbreviations

OS	Oxidation state
3-MST/CAT	3-Mercaptopyruvate sulfurtransferase/Cysteine aminotransferase
AIM	Atoms in Molecules
Ala	Alanine
AMBER/Amber	Assisted Model Building with Energy Refinement
Arg	Arginine
ASH	Allyl thiol
Asp	Aspartic acid
ASSH	Allyl perthiol
B3LYP	Becke, three-parameter, Lee−Yang−Parr
BCP	Bond critical point
Bn	Benzyl
C (atom)	Carbon
C (DNA/RNA)	Cytosine nucleotide
CBS (enzyme)	Cystathionine β-synthase
CBS (basis set)	Complete basis set
CC	Coupled-cluster
CP	Critical point
CSE	Cystathionine γ-lyase
Cys	Cysteine
Cyt	Cytosine
DADS	Diallyl disulfide
DAS	Diallyl sulfide
DATS	Diallyl trisulfide

DF-	Density fitting-
DFT	Density Functional Theory
DMDS	Dimethyl disulfide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DOPE	Discrete Optimized Protein Energy
DPDS	Dipropyl disulfide
dsDNA	Double stranded DNA
E	Electronic energy/energies
EHT	Extended Hückel Theory
G (DNA/RNA)	Guanine nucleotide
G	Gibbs free energy/energies
GAFF	General Amber Force Field
GGA	Generalized gradient approximation
Gln	Glutamine
Glu	Glutamic acid
GluH	Protonated glutamic acid
Gly	Glycine
GSA	S-allylglutathione
GSH	Glutathione
GSSA	Allyl-glutathione disulfide
GSSG	Glutathione disulfide
GSSH	Glutathione perthiol
GSSSA	Allyl-glutathione trisulfide
GTF	Gaussian-type function
H	Enthalpy/enthalpies
H_2S	Hydrogen sulfide

H-bond	Hydrogen bond
hDNMT1	Human DNMT1
HF	Hartree−Fock
His	Histidine
IEFPCM	Integral equation formalism of the polarizable continuum model
IPCM	Isodensity surface polarizable continuum model
IRC	Intrinsic Reaction Coordinate
KIE	Kinetic isotope effect
lg	Leaving group
LP	Lone pair
LSDA	Local spin density approximation
LYP	Lee−Yang−Parr
M.EcoRII	EcoRII methylase
M.HhaI	Modification methylase HhaI
M06	Minnesota 06
MCPB	Metal center parameter builder
MD	Molecular dynamics
mDNMT1	Mouse DNMT1
MeSH	Methylmercaptan
Met	Methionine
MFEP	Minimum free energy path
MM	Molecular mechanics
MMFFs	Merck Molecular Force Field, static
MOE	Molecular Operating Environment
MOLPDF	Modeller objective function
MP	Møller−Plesset

NA	Not applicable
NBO	Natural Bond Orbital
ND	Not determined
NPT	Isobaric-isothermal ensemble
nuc	Nucleophile
NVE	Microcanical ensemble
NVT	Canonical ensemble
OPLS_2005	Optimized Potentials for Liquid Simulations, 2005
PBC	Periodic boundary condition
PCM	Polarizable continuum model
PES	Potential energy surface
Phe	Phenylalanine
PME	Particle-Mesh-Ewald
PMF	Potential of mean force
Pro	Proline
QM	Quantum mechanics
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
RNA	Ribonucleic acid
S (thermochemistry)	Entropy/entropies
S (atom)	Sulfur
SAH	S-adenosyl-$_L$-homocysteine
SAM	S-adenosyl-$_L$-methionine
SCC-DFTB	Self-consistent-charge density functional tight-binding
SCF	Self-consistent field
SCRF	Self-consistent reaction field

Ser	Serine
SMD	Solvation model based on density
S_N2	Biomolecular nucleophilic substitution
STO	Slater-type orbital
TPP	Thiamine pyrophosphate
Trp	Tryptophan
TS	Transition state
Val	Valine
vdW	Van der Waals
VMN	Vosko−Wilk−Nusair
WHAM	Weighted histogram analysis method
ZPVE	Zero-point vibrational energy
γ-Glu-Cys-Gly	$_L$-γ-glutamyl-$_L$-cysteinylglycine, GSH

Chapter 1 Introduction

1.1 General Introduction

Sulfur (S with the atomic number of 16) is required by all living organisms[1] and can be found in compounds with various oxidation states (OS), such as thiols, sulfides and sulfonium ions (OS: -2), perthiols and disulfides (OS: -1), elemental sulfurs, sulfoxides and sulfenates (OS: 0), sulfones and sulfinates (OS: +2), sulfites and sulfonates (OS: +4), sulfates (OS: +6), and compounds with combinations of these oxidation states[2]. It is present in all major classes of biomolecules[1b], including proteins with amino acid residues cysteine (Cys) or methionine (Met), thiosugars and sulfur-containing nucleic acids with oxygen atoms replaced by sulfur atoms, cofactors such as thiamine pyrophosphate (TPP)[3], biotin[4], alpha-lipoic acids[5], coenzyme A[6], S-adenosyl-$_L$-methionine (SAM)[7], glutathione (GSH)[8], and iron-sulfur ([Fe-S]) clusters[9], and the related metabolites such as S-adenosyl-$_L$-homocysteine (SAH)[10], hydrogen sulfide (H_2S)[11] and glutathione disulfide (GSSG)[12]. These sulfur-containing compounds can play critical roles in biological systems as cofactors for enzyme catalysis (e.g. biotin, SAM), as antioxidants (e.g. GSH), as metal chelators (e.g. in iron-sulfur clusters), and as signaling agents (e.g. H_2S).

This thesis mainly reports on the computational studies of biochemical reactions in two different sulfur-containing systems. In the first system, the deoxyribonucleic acid (DNA) is methylated by an enzyme DNA methyltransferase (DNMT) that covalently binds to DNA via the action of a Cys-S(H) in the active site and transfers a methyl group from the cofactor

SAM to the 5-carbon (C5) of the DNA base cytosine (see section 1.2 for more details). In the second system, H_2S is released from the reaction of a free -SH in cells (e.g. GSH) and garlic-derived organosulfur compounds (see section 1.3 for more details).

1.2 DNA Methylation and DNA Methyltransferases

1.2.1 Epigenetics, DNA Methylation and DNA Methyltransferases

Epigenetics is the study of any reversible and heritable changes in gene expression that influence cellular phenotype without changes in genomic DNA sequence[13]. Epigenetic regulations include DNA methylation, covalent histone modification or incorporation of histone variants, ribonucleic acid (RNA) interference and chromatin remodeling[13]. All of these changes function synergistically to regulate the chromatin structure, and therefore determine the transcription of genome in an organism[14].

DNA methylation is a key epigenetic regulatory mechanism and has many biological functions. It plays a major role in long-term gene silencing, including X-chromosome inactivation and parental imprinting[15], as well as the suppression of transposons and other parasitic elements for genome integrity[16]. Aberrant DNA methylation patterns are associated with many human diseases, and are frequently identified in various cancers[13]. These changes involve global hypomethylation accompanied by hypermethylation at specific loci[13]. Hypomethylation leads to genomic instability and possible stimulation of oncogenes and satellite DNA; whereas hypermethylation found in the

promoter regions of tumor-suppressor genes leads to the inactivation of these protective genes[17]. Therefore, DNA methylation is a promising anti-cancer target to reverse the methylation defects in cancer cells.

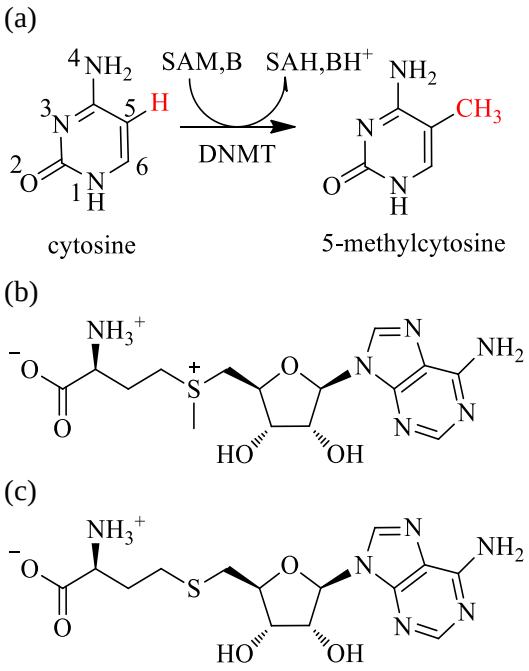


Figure 1.1 (a) DNA methylation occurs at the C5 of cytosine. B stands for base. (b) Structure of SAM. (c) Structure of SAH.[13]

In vertebrates, DNMTs catalyze the transfer of a methyl group to the C5 of cytosine to form 5-methylcytosine[13] (Figure 1.1(a)). SAM is an electrophilic methyl donor[7, 13] (Figure 1.1(b)), and it changes to SAH after the reaction[10, 13] (Figure 1.1(c)). The methylation occurs primarily in CpG dinucleotide $\begin{matrix}5'\text{-CpG-}3'\\3'\text{-GpC-}5'\end{matrix}$ doublets, where 5'-CpG-3' means that the C3' on the deoxyribose of a cytosine nucleotide (C) is linked to the C5' on the deoxyribose of a guanine nucleotide (G) via a phosphodiester bond (p)[13]. The cytosine in the complementary 3'-GpC-5' base pairs is also methylated symmetrically.

Approximately 70% of all CpG sites are methylated[13]. However, "CpG islands", the CpG-rich regions of the genome found in the promoter regions of many genes, are usually unmethylated[14].

Active mammalian DNMTs include DNMT1, DNMT3a, and DNMT3b[14]. DNMT1 is mainly responsible for the maintenance of the established DNA methylation patterns by targeting hemi-methylated DNA, whereas DNMT3a and 3b are responsible for the establishment of de novo DNA methylation patterns by targeting unmethylated DNA[14].

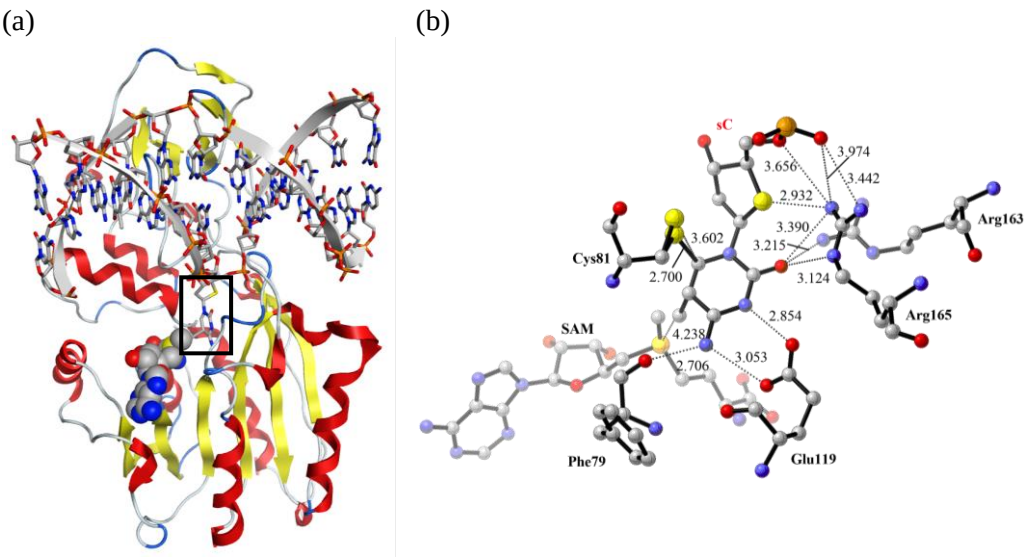


Figure 1.2 (a) Structure of the M.HhaI with SAM and the modified dsDNA (6MHT)[18]. The substrate 4'-thio-2'-deoxycytidine is boxed. The protein is shown by the cartoon model, SAM is shown by the CPK model and colored by atom, and the nucleotides are shown by the stick model and colored by atom. (b) Arrangement of Cys81, Glu119, Arg163, Arg165, Phe79, SAM and the substrate 4'-thio-2'-deoxycytidine (sC) in the active center. Two alternative positions of the Cys-S and the transferring methyl-C are shown. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

The structures of DNMT1, DNMT3a and DNMT3b can be divided into a large N-terminal regulatory domain and a smaller C-terminal catalytic domain[14]. The N-terminal domain is unique to the active eukaryotic cytosine-

C5 DNMTs and has several functions, such as nuclear localization, targeting the replication foci, binding to unmethylated CpGs, and interacting with other proteins, DNA and chromatins[14]. The C-terminal domain is conserved in both prokaryotic and eukaryotic cytosine-C5 DNMTs[14]. This domain contains the active center of the protein and has ten structural motifs characteristic to all cytosine-C5 DNMTs[14]. It binds to the substrate DNA and the cofactor SAM or SAH, and has the common "SAM-dependent methyltransferase fold"[14]. The crystal structure of a well-studied bacterial DNMT, modification methylase HhaI (M.HhaI), in complex with DNA (6MHT)[18] is shown in Figure 1.2(a). M.HhaI only has the catalytic domain that consists of 327 amino acid residues and does not bind to any Zn^{2+} (Zn(II)) unlike DNMT1[18]. In addition, M.HhaI methylates the 5'-C on both strands in a 5'-GpCpGpC-3' tetranucleotide doublets[19]. The "SAM-dependent methyltransferase fold" is the formed by 6 parallel β-strands and a 7^{th} antiparallel strand, inserted into the sheet between strands 5 and 6[14]. Six helices are folded around the β-sheets[14]. The double stranded DNA (dsDNA) is embedded in the cleft of the catalytic domain, with the modified substrate 4'-thio-2'-deoxycytidine of the target strand looped out of the dsDNA and inserted into the active site[18]. The substrate is surrounded by four strictly conserved residues from the catalytic core—a cysteine (Cys81), a glutamic acid (Glu119) and two arginines (Arg163/165), and is also close to the cofactor SAM[18] (numbering based on M.HhaI) (Figure 1.2(b)). The substrate 4'-thio-2'-deoxycytidine (sC) has the O4' of the deoxyribose replaced by a sulfur that is incorporated to slow down the methylation reaction

to capture this covalent adduct between the substrate and M.HhaI by crystallization[18].

Because DNMT1 is the most abundant type of DNMTs and it is also responsible for gene silencing in cancer, the human DNMT1 (hDNMT1) was chosen as the protein target in my research and the main focus is on the substrate-binding active site and the SAM-binding site in the C-terminal catalytic domain.

1.2.2 Molecular Modeling of Human DNMT1 Structures and Objectives of the Study

Table 1.1 Reported molecular modeling studies of hDNMT1 catalytic domain with the substrate and SAM or SAH[a]

Types	Templates	Results and Findings	References
Homology modeling	M.HhaI (6MHT) M.HaeIII (1DCT) hDNMT2 (1G55)	Modeled protein residues are from 1135 to 1602, cytosine nucleotide and SAM included; the model has several minor but potentially important structural differences c.f. the template proteins; model validation by the predicted best novel ligand in silico and its experimental testing	Siedlecki *et al.* (2003)[20]
	M.HhaI (5MHT)	Modeled residues are from the catalytic domain[b]; the model shows substantial interactions between the hemimethylated dsDNA and the active site	Fang *et al.* (2003)[21]
	M.HhaI (5MHT)	Modeled protein residues are from 1139 to 1616, cytosine and SAM included	Liu *et al.* (2009)[22]
	M.HhaI (6MHT) M.HaeIII (1DCT) hDNMT2 (1G55)	Modeled protein residues are from 1133 to 1601[b], dsDNA and SAH included; the model agrees with the proposed mechanism of DNA methylation	Yoo *et al.* (2011)[23]
	M.HhaI (4MHT)	Modeled protein residues are from 1139 to 1616[b], cytosine and SAH included	Weng *et al.* (2014)[24]
	mDNMT1 (4DA4)	Modeled protein residues are from 729 to 1598, SAM and two Zn(II) ions included; the initial homology model is very similar to mDNMT1 with reasonable Ramachandran plot and Verify3D score	Joshi *et al.* (2016)[25]
Molecular dynamics	Homology model from hDNNT1 (4DA4)	The first and the second MD simulations showed the change in the catalytic loop and the helix following it which is similar to that in the structure in the inactive form of hDNMT1(3PTA)	Same as above

	hDNMT1 (3PTA) M.HhaI (1MHT)	Modeled protein residues are from 1135 to 1600; the catalytic loop is different from the crystal structure but is similar to that in the homology model reported by the same group in 2011	Yoo *et al.* (2013)[26]

[a] Only one modeling study by Joshi *et al.* (2016)[25] did not include the substrate in the 3D structure.
[b] The reported residue numbering deviates by -1 from other references and is found to be incorrect.

The reported crystal structures of M.HhaI (from *Haemophilus haemolyticus* or *Haemophilus parahaemolyticus*)[18-19, 27], mouse DNMT1 (mDNMT1)[28] and hDNMT1[28a, 29] involving the catalytic domains of the native proteins are summarized in Table S 8.1. If the dsDNA is present in the crystal structure, only the structures without major modification of the cytosine base (e.g. absence of the aromatic 6-membered ring) or the deoxyribose (e.g. use of a non-sugar or something other than a furanose) are shown.

Until now, no crystal structure of hDNMT1 bound to the dsDNA at the active site has been reported, but such structures have been reported for other DNMTs including M.HhaI (1MHT, 3MHT–6MHT, 2HR1, 2Z6A)[18, 27b, 27c, 27f, 27g, 30] and mDNMT1 (4DA4)[28b] (Table S 8.1). Therefore, the 3D structures of hDNMT1 in complex with its substrate (cytosine, cytosine nucleotide or dsDNA) at the active site were derived by homology modeling based on related DNMTs[20-21, 23-24, 31] or from classical molecular dynamics (MD) simulations of an inactive hDNMT1[26] (Table 1.1). The detailed computational methods can be found in Table S 8.2. Homology modeling and classical MD methods will be described in Chapter 2.

The first homology model of hDNMT1 was established by Siedlecki *et al.*[20] based on the overall structural conservation of the catalytic domain of 24 homologous methyltransferases, and the crystal structures of bacterial DNMTs M.HhaI (6MHT)[18], M.HaeIII (1DCT)[32], and the tRNA cytosine

methyltransferase hDNMT2 (1G55)[33] were selected for model construction due to more than 50% of sequence similarity. This model has been used in many other computational studies[34] and facilitated the discovery or study of various substrate-competitive ligands. Later on, three homology models of hDNMT1, each derived from a single crystal structure of M.HhaI (4MHT[27b] or 5MHT[27c]), were reported independently[21-22, 24]. None of these models were described or evaluated in detail and no other studies were reported based on those models. In 2011, another homology model of hDNMT1 was reported by Yoo *et al.*[23] based on the three crystal structures used to generate the first homology model but using different computer programs and approaches. Based on this model, the authors reported an extensive molecular modeling study on various substrate-competitive ligands from different sources[23] and discovered a novel ligand in a separate study[35], which is beyond the scope of this thesis. In the same year, the first crystal structure of hDNMT1 in complex with DNA at the N-terminal domain (3PTA) was published (Table S 8.1), with the catalytic loop (involving the catalytic Cys) in an open conformation far away from the active site[28a]. This crystal structure was used to model the active hDNMT1-DNA complex using MD simulations in 2012[26] and the model is comparable to the homology model reported by the same group in 2011[23].

In 2012, the first crystal structure of mDNMT1 bound to DNA at the active site (4DA4) was reported by Song *et al.*[28c] (Table S 8.1), and this is the first eukaryotic DNM1-DNA co-crystal in the active form. Its structure retains all the key binding features between the substrate and the active site as

previously observed for the M.HhaI-DNA complex with the same modified substrate 5-fluoro-2'-deoxycytidine (1MHT[27a]). In 2016, a homology model of hDNMT1 was published by Joshi *et al.*[25] based on this crystal structure, but the DNA was not included in the model (Table 1.1), because the authors aimed to study another binding site rather than the active site. MD simulations were also reported on their homology model.

In this thesis, a novel 3D model of the hDNMT1-DNA complex was established via homology modeling and refined by MD simulations based on the known active mDNMT1-DNA complex (4DA4)[28c] that may provide a basis for other molecular modeling studies in the future. This study will be presented in Chapter 3.

1.2.3 Mechanistic Study on DNA Methylation and Objectives of the Study

Studies on the DNA methylation mechanism have been an important area of research over the years experimentally and computationally. Biochemical experiments[36] and crystallographic studies of the complexes between DNMTs such as M.HhaI and mDNMT1, dsDNA and SAM or SAH[18, 27a-c, 27f, 27g, 28c] (Table S 8.1) have revealed important features of the methylation reaction. For instance, when the modified substrate 5-fluoro-2'-deoxycytidine in the hemi-methylated dsDNA binds to the active site of M.HhaI (1MHT)[27a] or mDNMT1 (4DA4)[28c], it is flipped out of the dsDNA and methylated at C5 through the nucleophilic attack of the catalytic Cys-S(H) (Cys81 in M.HhaI or Cys1229 in mDNMT1) at C6, forming a covalent

intermediate (Figure 1.3(a)–(b)). Because C5 is linked to a fluorine instead of a hydrogen, this covalent adduct cannot be resolved by deprotonation and its crystal structure was obtained. As shown in Figure 1.3, the flipped-out cytosine nucleotide (fC or fC') is covalently linked to Cys81 or Cys1229 while hydrogen-bonded to residues in the active site of M.HhaI or mDNMT1, including a glutamic acid (Glu119 or Glu1269), a phenylalanine (Phe79) or a proline (Pro1227), and two arginines (Arg163/165, or Arg1313/1315).

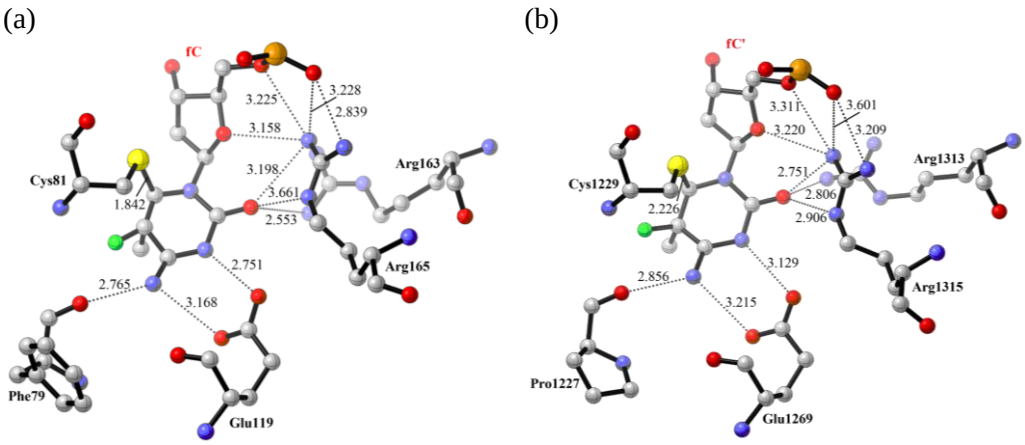


Figure 1.3 Interaction of the flipped-out cytosine nucleotide (fC or fC' with 5-methyl) with the active site residues of (a) M.HhaI (1MHT)[27a] and (b) mDNMT1 (4DA4)[28c]. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

Four of these residues—a Cys, a Glu and two Arg's—are completely conserved in the active site of the cytosine-C5 DNMT family[30], suggesting the important roles of them in the methylation reaction. The individual roles of Cys, Glu and Arg have been further studied by site-directed mutagenesis in various DNMTs[27f, 31, 37]. First, the mutation of Cys186 in EcoRII methylase (M.EcoRII) to alanine (Ala), glycine (Gly), valine (Val), tryptophan (Trp) or serine (Ser) leads to the significant reduction or loss of the methylase activity[37a, 37c]. The mutation of Cys1229 in mDNMT1 to Trp or Cys1226 in

hDNMT1 to Ala also leads to the loss of the methylase activity[31, 37d]. These results suggest the critical role of Cys as an enzyme nucleophile. Second, the mutation of Glu119 in M.HhaI to Ala, aspartic acid (Asp) or glutamine (Gln) largely lowers the rate of methyl transfer and results in the loss of enzyme trapping by the 5-fluorocytosine that is characteristic in the native DNMTs (Figure 1.3)[37b]. The mutation of Glu119 to Asp also decreases the DNA binding affinity possibly due to the water-assisted binding between Asp and cytosine[37b]. These results support the key role of Glu in tight DNA binding and proper positioning of the substrate for the nucleophilic attack by Cys81, as well as the critical role for the methylase activity. Third, the mutation of Arg165 in M.HhaI to Ala causes reduced base flipping and restacking transitions, greatly decreased catalytic rate, and altered cytosine orientation relative to the DNA backbone[27f]. These results imply the important role of this Arg for base flipping, cytosine positioning and catalysis. In short, Glu and Arg can facilitate the nucleophilic attack of Cys via protonation of N3 of cytosine[27b] or electrostatic interaction with O2 of cytosine[37c] (Figure 1.3(a)).

The kinetic studies on M.HhaI[36] showed that in the absence of SAM, M.HhaI can catalyze the exchange of C5-tritium (^{3}H) of cytosine with water protons faster than the rate of methylation, thus suggesting that the formation of a transient covalent intermediate is through the nucleophilic attack at the C6 of cytosine alone. Unlike the relatively stable methylated covalent adduct, this intermediate was never observed directly with or without SAM as it is faster than the methyl transfer or the proton transfer at C5[36].

A stepwise reaction mechanism has been proposed and revised by different groups[36a, 37c, 38] based on the biochemical studies and crystallographic studies as well as analogy to the methylation reaction of 2'-deoxyuridine monophosphate by thymidylate synthase[36a, 39]. The general scheme is shown in Figure 1.4 (numbering based on M.HhaI). The overall chemical reaction consists of three steps. In the step 1, the thiolate of Cys81 acts as a nucleophile to attack the C6 of cytosine, forming the covalent intermediate 1 (I1) via the transition state 1 (TS1). This intermediate is stabilized via transient protonation of N3 on cytosine by Glu119 or electrostatic interaction of O2 with Arg163/165. In the step 2, the C5 of cytosine is activated and performs a nucleophilic attack on the methyl group of SAM to form the 5-methyl covalent intermediate 2 (I2) and SAH via the transition state 2 (TS2). In the step 3, the covalent complex is resolved by β-elimination of the C5-proton and the C6-thiolate to generate 5-methylcytosine and the free enzyme. The base involved in the deprotonation step was not clearly identified in the experimental studies.

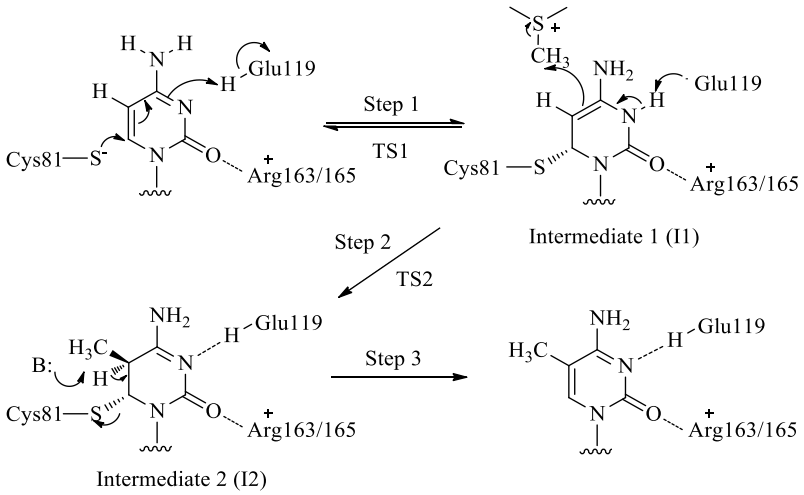


Figure 1.4 Proposed DNA methylation mechanism[36a, 37c, 38].

To improve the understanding of the DNA methylation mechanism and to solve some of the questions including the protonation state of Cys and the identity of the extracting base, computational calculations on the reaction profile were reported in six separate studies[40]. The detailed computational methods, the corresponding references and the numbering of these studies are summarized in Table S 8.3. The methods will be described in more detail in Chapter 2. Quantum mechanics (QM) calculations were reported in the studies by Peräkylä (Study 1)[40a], Zangi *et al.* (Study 2)[40c] and Du *et al.* (Study 5)[40f] on smaller model systems with implicit solvation models, while the hybrid quantum mechanics/quantum mechanics or molecular mechanics (QM/QM or QM/MM) approaches allowed the study of the methylation reaction in the enzyme environment with explicit solvent molecules in the studies by Zhang *et al.* (Study 3)[40b], Yang *et al.* (Study 4)[40d], Du *et al.* (Study 5)[40f] and Aranda *et al.* (Study 6)[40e]. The Study 5[40f] covered both types of calculations. Studies 4[40d] and 6[40e] even adopted the QM/MM-MD approaches to account for the enzyme active site dynamics in the reaction. The structure of M.HhaI (2HR1[27f] or 6MHT[18]) was used to model the C-terminal catalytic domain in Studies 1–4[40a-d] and 6[40e] because it is the most well characterized DNMTs and shares the similar methylation mechanism with mammalian enzymes[17]. In the Study 5[40f], the structure of mDNMT1 (4DA4[28c]) was used to model the active site environment of hDNMT1 because they possess identical residues in the active site.

The key insights from these six studies for the first two steps in the DNA methylation (Figure 1.4) are summarized in Table 1.2 because the major

controversies were identified from these two reaction steps. The numbering of these studies is shown in Table 1.2 and will always be used to refer to these studies in this chapter and Chapter 4.

Table 1.2 Summary of key findings from computational studies on DNA methylation

(a) Step 1: Nucleophilic attack on C6

Study	Key findings from Step 1
1	(QM) Deprotonated thiolate used as the enzyme nucleophile N3 protonation needed for nucleophilic attack Proton may come from Glu119 Nonaqueous active site preferred for faster reaction
2	(QM) Deprotonated thiolate proved to be the enzyme nucleophile Phosphate on DNA abstracts the proton from Cys81-SH via a bridging water Sufficient activation of C6 by Arg165 and either Arg163 or Glu119 protonation
3	(QM/MM) Deprotonated thiolate used as the enzyme nucleophile Stable intermediate I1 with N3 protonation Arg163 and Arg165 not provide stabilization of the TS1
4	(QM/MM only) Deprotonated thiolate proved to be the enzyme nucleophile Unstable intermediate I1 with/without N3 protonation The S-C6 bond breaks while N3 remains protonated when minimized
5	(Simple Model) Deprotonated thiolate used as enzyme nucleophile, N3 protonation not shown in the TS1 Predicted KIEs not agree with the experimental KIEs, so this is not the rate-limiting step
6	(QM/MM-MD) After DNA and SAM binding, Cys81 deprotonation by a DNA phosphate through Ser85 in a concerted process, this step is endergonic Favorable proton release from the DNA phosphate to the bulk water Fast and reversible Cys81 nucleophilic addition to cytosine without N3 protonation/deprotonation Easy protonation of N3 by Glu119 after the intermediate I1 formation

(b) Step 2: Methyl transfer to C5

Study	Key findings from Step 2
1	(QM) More favorable for neutral cytosine Highly exothermic in the gas phase and solution
2	(QM) Methyl transfer with stabilization from the protonated Glu119 and Arg165

3	(QM/MM) Not undergo methylation Methyl transfer concerted with Cys81 addition without N3 protonation/deprotonation
4	(QM/MM only) Methylation of the unstable intermediate I1 with/without N3 protonation showed the breakage of the S-C6 bond, though the N3-proton returns to Glu119 in the former case (QM/MM-MD) Methyl transfer concerted with Cys81 addition during MD but the process is asynchronous with the Cys81 addition first supported by spontaneous and reversible N3 protonation/deprotonation, followed by methyl transfer according to the MD trajectory
5	(Simple Model) N3 protonation shown in the TS2 Predicted KIEs agree with the experimental KIEs, so this is the rate-limiting step (QM/MM) Cluster TS2 model contains cytosine nucleotide, SAM, eight near protein residues (including three of the four completely conserved residues: Cys1226, Glu1266, Arg1312 (2^{nd} Arg)) and three water molecules N3 protonation shown in the TS2 TS2 with a nearly complete covalent bond to Cys1226, and near-symmetrical S_N2 feature for methyl transfer with noncompressed reaction coordinates Predicted KIEs agree well with the experimental KIEs, so this is the rate-limiting step (barrier is 12.0 kcal mol^{-1})
6	(QM/MM-MD) Methyl transfer without N3 protonation is preferred over methyl transfer with N3 protonation Rate-limiting step (19.1 kcal mol^{-1}) Potential energy surface (PES) tested by the B3LYP/MM single point energy calculations: methyl transfer concerted with Cys81 addition but asynchronous with the Cys81 addition first followed by methyl transfer

In general, the results from most of these studies agree well on the role of Glu119 in the reaction steps 1 and 2 via protonation or deprotonation of the N3 of cytosine. The results also supported the roles of Arg163 and Arg165 in the stabilization of the reaction step 1 via electrostatic interactions or H-bonds with the O2 of cytosine.

For the nucleophilic attack of C6 on cytosine by the Cys-S(H) (Table 1.2 (a)), the enzyme nucleophile was found to be the deprotonated thiolate although the Cys residue is usually protonated at the physiological pH. The

base that deprotonates the Cys-SH was suggested to be a DNA phosphate group via a bridging water according to Studies 2[40c] and 6[40e].

The concerted mechanism that combines the nucleophilic attack of the Cys-thiolate on C6 and methyl transfer from SAM to C5 was proposed and tested with the QM/MM or QM/MM-MD approaches using the B3LYP method in Studies 3[40b] and 4[40d] to resolve problems in the calculation of the stepwise mechanism (Table 1.2(a)–(b)). According to the Study 4, the transient covalent intermediate I1 with or without N3 protonation dissociated upon minimization in the QM/MM calculation. Such findings seem to contradict the experimental findings mentioned earlier. In Studies 5[40f] and 6[40e], the stepwise mechanism was supported with the QM/QM or QM/MM-MD approaches using the M06-2X method. In particular, when the authors in the Study 5 compared the potential energy surfaces (PESs) computed at the B3LYP/6-31G**//AM1/MM level with that computed at the M06-2X/6-31G**//AM1/MM level, they found that the PESs from the former method supported the concerted mechanism while the PESs from the latter method supported the stepwise mechanism. This was explained by the lack of dispersion contributions in the B3LYP method to stabilize the unbounded Cys in order to describe the rapid equilibrium between the reaction complex before the Cys nucleophilic attack and the transient covalent intermediate I1 after the nucleophilic attack.

The study of the DNA methylation mechanism in this thesis was proposed when only Studies 1–4[40a-d] were reported. At that time, the existence and stability of the covalent intermediate I1 (Figure 1.4) from the Cys-thiolate

addition was disputed, and the subsequent methyl transfer to the I1 was not supported in the protein environment according to the two QM/MM calculations. The breakage of the C6-S bond is uncommon in the Study 4. Beside, most of these studies chose the B3LYP[41] method to study the system, but this method cannot provide the correct London dispersion interaction energies as described in the Study 6[40e]. In theory, the M06-2X[42] method can provide the correction to this dispersion term in the system and may be more suitable to study this system.

In this thesis, the first two steps in the DNA methylation reaction were studied with the B3LYP or the M06-2X method based on small model systems to probe the key TSs and intermediates with the objectives to resolve the above-mentioned controversies based on Studies 1–4[40a-d]. This study will be presented in Chapter 4.

1.3 H_2S Release from Organosulfur Compounds

1.3.1 H_2S in Biological Systems

H_2S is a small gaseous molecule that can be both toxic and beneficial to biological systems depending on its concentrations[43]. In mammals, the biosynthesis of H_2S usually starts from sulfur amino acids and involves the action of four enzymes: cystathionine γ-lyase (CSE), cystathionine β-synthase (CBS), and 3-mercaptopyruvate sulfurtransferase together with cysteine aminotransferase (3-MST/CAT)[43a, 43c, 44].

With the most reduced form of sulfur (OS: -2), H_2S can act as a good reducing agent in several chemical reactions[43b]. In the body, H_2S can readily

react with various oxidants in the redox systems such as oxygen (O_2)[43b, 45], superoxide radical anion ($O_2^{\cdot -}$)[43b, 46], hydrogen peroxide (H_2O_2)[43b, 46b, 47], peroxynitrite (ONOOH/$ONOO^-$)[43b, 48], and hypochlorite (HOCl/^-OCl)[43b, 49]. It can also react with various sulfur-containing compounds such as disulfides (RSSR)[43b, 50], S^0-containing compounds[43b, 51], and thiol derivatives like S-nitrosothiol (RSNO)[43b, 52]. Because the reaction of H_2S with disulfides is reversible, the reverse reaction can be utilized to generate H_2S.

Moreover, H_2S is a common sulfur nucleophile especially when deprotonated and can react with many metal systems in the body. Typical examples are heme irons in the mitochondrial cytochrome *c* oxidase[53], and nonheme irons in iron-sulfur ([Fe-S]) clusters[54].

Because H_2S can interact with various biological targets in the body, it has been discovered as a gaseous mediating molecule or the so-called "gasomediator" that regulates multiple physiological functions in the body[11]. These functions include but not limited to vasodilation and anti-hypertension[55], anti-inflammation[56], reducing oxidative stress[57], cytoprotection against apoptosis[58], increasing fibrinolytic activity[11, 59], anti-platelet activation and aggression[11, 59], promoting angiogenesis[60], cardioprotection[58b, 58c, 61], suppressing metabolism[62], and preventing atherogenesis[11, 63].

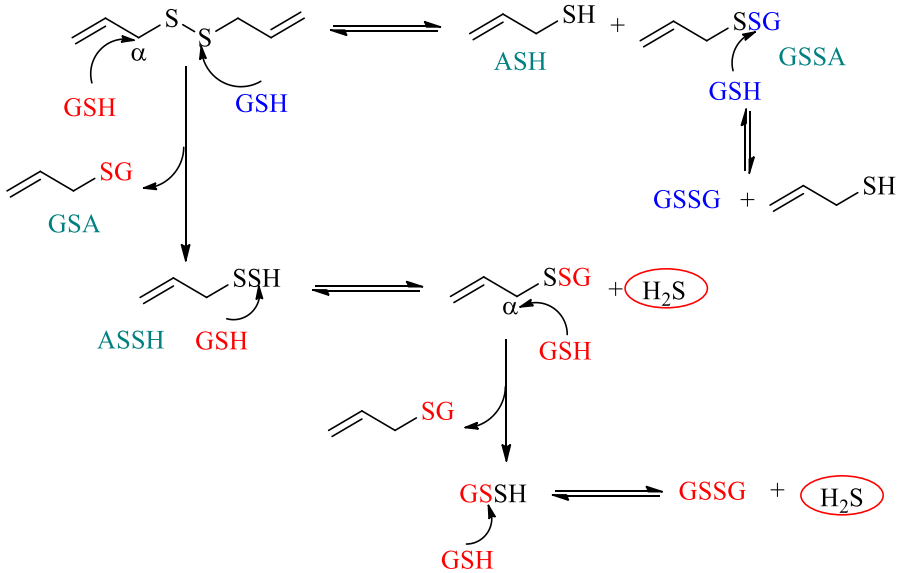
1.3.2 Organosulfur Compounds as H_2S Donors

Direct use of H_2S to study or modulate the biological effects as therapeutic agents is not preferred because it is a highly reactive gas under normal conditions. Several H_2S donors have reported so far[64], of which, garlic-

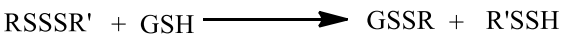
derived organosulfur compounds have been studied as H_2S-releasing agents recently with strong interests[43a, 65]. The most abundant organosulfur compound in the intact garlic is a sulfur amino acid S-allyl-cysteine sulfoxide(alliin), which is rapidly converted to diallyl thiosulfinate (allicin) when the garlic is processed[43a, 66]. Allicin is highly unstable and decomposes to form a variety of lipid-soluble organosulfur compounds such as diallyl sulfide (DAS), and diallyl disulfide (DADS) and diallyl trisulfide (DATS)[43a, 66].

Benavides *et al.*[65c] first reported the rapid H_2S release within a 10-minute reaction by the garlic-derived DADS and DATS, in the presence of the reduced thiols such as GSH, by correlating the vasodilative activity of garlic compounds with H_2S production. Based on these results, the authors proposed the mechanism of H_2S production from the reaction of GSH and DADS/DATS (Figure 1.5).

(a)



(b)



(c)

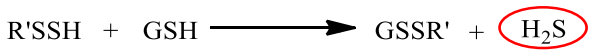


Figure 1.5 H_2S production from garlic-derived polysulfides[65c]. (a) Proposed mechanism of H_2S production from the reaction of GSH and DADS via Cα nucleophilic attack (red) but not by thiol/disulfide exchange (blue). (b)&(c) H_2S production by thiol/disulfide exchange of organic trisulfides.

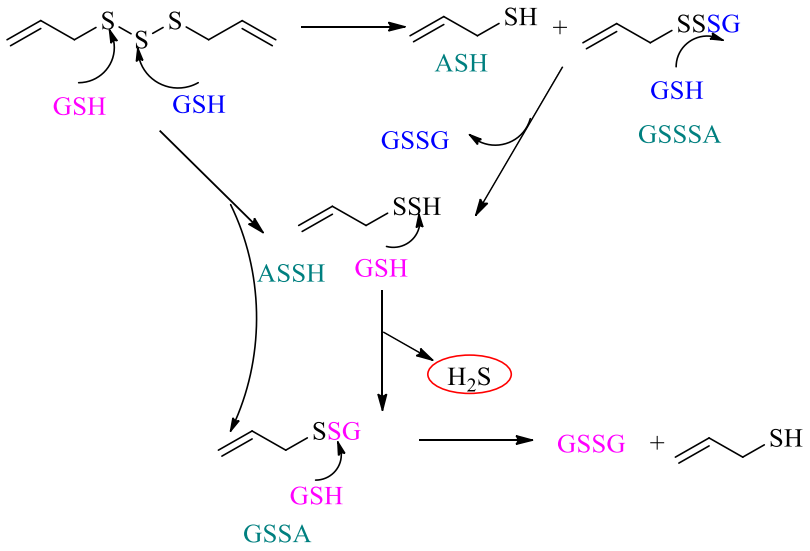


Figure 1.6 H_2S production from DATS[65a]. Proposed mechanism of H_2S production from the reaction of GSH and DATS via the allylic sulfur nucleophilic attack (pink) or the central sulfur nucleophilic attack (blue).

The proposed mechanism suggested that the rapid formation of H_2S from the reaction of GSH and DADS proceeds via Cα nucleophilic attack on DADS to form S-allylglutathione (GSA) and the key intermediate allyl perthiol (ASSH) (Figure 1.5(a), red) instead of the more common thiol/disulfide exchange that is the nucleophilic attack on a disulfide linkage by a thiol. This is because the latter reaction does not generate H_2S from the formation of allyl thiol (ASH) or allyl-glutathione disulfide (GSSA) (Figure 1.5(a), blue). ASSH further undergoes the S nucleophilic substitution, which is similar to thiol/disulfide exchange, forming GSSA and H_2S. GSSA can also undergo the

Cα nucleophilic substitution to form another key intermediate glutathione perthiol (GSSH), leading to more H_2S production. On the contrary, organic trisulfides such as DATS may directly undergo the S nucleophilic substitution, forming ASSH that then reacts with GSH to form H_2S (Figure 1.5(b)&(c)).

Later on, Truong *et al.*[65b] observed that DADS was able to induce cytotoxicity towards hepatocytes and inhibit cytochrome *c* oxidase dependent mitochondrial respiration, both of which can be prevented by the H_2S scavenger hydroxocobalamin. Their results linked the hepatocyte cytotoxicity of DADS to H_2S inhibition of the mitochondrial cytochrome *c* oxidase[53]. Moreover, the authors observed the depletion of the intracellular GSH towards DADS in hepatocytes, supporting the reaction of GSH and DADS in the production of H_2S.

In 2015, Liang *et al.*[65a] further investigated the H_2S production from the reaction of GSH and DADS/DATS at the physiological pH of 7.4 and a temperature of 37ºC, and characterized the reaction products by HPLC, LC-MS and fluorescence for better mechanistic understanding. They observed that the rapid H_2S release (within 25 minutes) occurred when reacting GSH with DATS but not with the purified DADS. They attributed the rapid H_2S release from DADS reported by Benavides *et al.*[65c] to the presence of DATS impurity in the commercial source of DADS identified by HPLC.

By examining the reaction products from the short duration reaction, the authors found that DADS reacted rapidly with GSH via thiol/disulfide exchange that cannot produce any H_2S as indicated by the blue pathway in Figure 1.5(a). However, when the reaction time was prolonged to 1.5 hour or

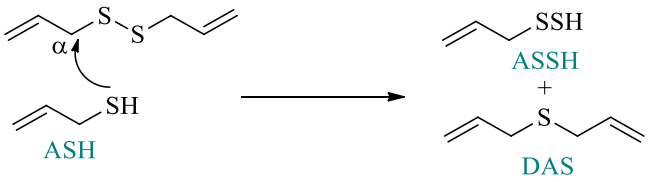
more, the Cα nucleophilic attack on DADS occurred slowly by measuring the reaction products like GSA, ASSH and H_2S (Figure 1.5(a), red).

From the reaction of GSH and DATS, the authors identified two possible thiol/disulfide exchange pathways. The first pathway is the nucleophilic attack of GSH on the allylic sulfur of DATS to form GSSA and ASSH which leads to H_2S production (Figure 1.6, pink). The second pathway is the nucleophilic attack of GSH on the central sulfur of DATS to form ASH and allyl-glutathione trisulfide (GSSSA) (Figure 1.6, blue). GSSSA can further react with GSH to form ASSH for H_2S production (Figure 1.6, pink).

Hence, this study showed that DADS is a slow H_2S donor via Cα nucleophilic attack, while DATS is rapid H_2S donor via thiol/disulfide exchange.

In addition, the authors discovered the formation of DAS alongside the formation of GSA from the Cα nucleophilic attack on DADS. They proposed that DAS may come from the reaction of ASH with either DADS or GSSA via Cα nucleophilic attack to form ASSH (Figure 1.7(a)) or GSSH (Figure 1.7(b)), leading to more H_2S production (Figure 1.5(a)).

(a)



(b)

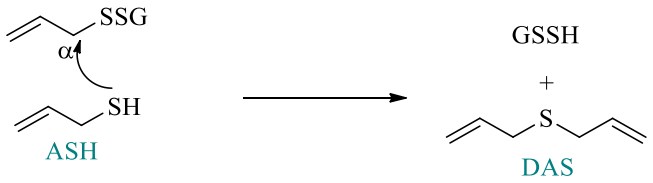


Figure 1.7 DAS formation via α carbon nucleophilic attack[65a]. (a) DAS formation from the reaction of DADS and ASH. (b) DAS formation from the reaction of GSSA and ASH.

In an unpublished study by Liang *et al.*, the authors tested the H_2S releasing ability of dialkyl disulfides. They found that dimethyl disulfide (DMDS) and dipropyl disulfide (DPDS) did not generate any detectable H_2S or products from the Cα nucleophilic substitution even when reacted with GSH for 12 hours. These results suggested that the Cα nucleophilic substitution of these dialkyl disulfides by GSH did not occur.

1.3.3 Objectives of the Study

The study by Liang *et al.*[65a] explained the misunderstanding of DADS as a rapid H_2S first reported by Benavides *et al.*[65c] and provided some experimental evidence to prove that the Cα nucleophilic substitution of DADS is much slower than the thiol/disulfide exchange, but they did not provide any explanations on the relative rates. Therefore, these two competing steps in reaction of GSH and DADS (Figure 1.5(a)) were studied computationally to calculate the respective activation barriers to compare with the experimental results and to elucidate the chemical reasons behind this.

Moreover, the overall reaction profile leading to H_2S production in reaction of GSH and DADS/DATS (Figures 1.5(a), 1.6 and 1.7) was studied computationally to improve the understanding of the reaction mechanism.

Finally, the hypothetical Cα nucleophilic substitution of dialkyl disulfides DMDS/DPDS by GSH and was studied computationally to calculate the activation barriers and to understand how the change of the allyl group to the alkyl group affects this reaction.

These studies will be presented in two separate chapters: Chapter 5 and Chapter 6. Chapter 5 focuses on the conformational analysis of the equilibrium structures (reactants and products) in the H_2S releasing reactions, while Chapter 6 presents the computational study on the reaction mechanisms and the energy profiles of the H_2S releasing reactions.

1.4 References

1. (a) In *Van Nostrand's Scientific Encyclopedia*, John Wiley & Sons, Inc.: 2005; (b) *Nat. Chem. Biol.* **2006,** *2* (4), 169.

2. Thompson, J. D.; Plewniak, F.; Poch, O., *Nucleic Acids Res.* **1999,** *27* (13), 2682.

3. Sauder, J. M.; Arthur, J. W.; Dunbrack Jr, R. L., *Proteins* **2000,** *40* (1), 6.

4. Jones, D. T.; Taylort, W. R.; Thornton, J. M., *Nat.* **1992,** *358* (6381), 86.

5. Bowie, J.; Luthy, R.; Eisenberg, D., *Sci.* **1991,** *253* (5016), 164.

6. Saqi, M. A.; Russell, R. B.; Sternberg, M. J., *Protein Eng.* **1998,** *11* (8), 627.

7. Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M., *Nat.* **1987,** *326* (6111), 347.

8. Greer, J., *Proteins* **1990,** *7* (4), 317.

9. Johnson, D. C.; Dean, D. R.; Smith, A. D.; Johnson, M. K., *Annu. Rev. Biochem.* **2005,** *74*, 247.

10. Unger, R.; Harel, D.; Wherland, S.; Sussman, J. L., *Proteins* **1989,** *5* (4), 355.

11. Kolluru, G. K.; Shen, X.; Bir, S. C.; Kevil, C. G., *Nitric Oxide* **2013,** *35*, 5.

12. Levitt, M., *J. Mol. Biol.* **1992,** *226* (2), 507.

13. Weber, W. W., In *Comprehensive Medicinal Chemistry II*, Triggle, D. J., Ed. Elsevier: Oxford, 2007; pp 251.

14. Jurkowska, R. Z.; Jurkowski, T. P.; Jeltsch, A., *ChemBioChem* **2011,** *12* (2), 206.

15. Miranda, T. B.; Jones, P. A., *J. Cell. Physiol.* **2007,** *213* (2), 384.

16. Howard, G.; Eiges, R.; Gaudet, F.; Jaenisch, R.; Eden, A., *Oncogene* **2008,** *27* (3), 404.

17. Jeltsch, A., *ChemBioChem* **2002,** *3* (4), 274.

18. Kumar, S.; Horton, J. R.; Jones, G. D.; Walker, R. T.; Roberts, R. J.; Cheng, X., *Nucleic Acids Res.* **1997,** *25* (14), 2773.

19. Cheng, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J., *Cell* **1993,** *74* (2), 299.

20. Siedlecki, P.; Boy, R. G.; Comagic, S.; Schirrmacher, R.; Wiessler, M.; Zielenkiewicz, P.; Suhai, S.; Lyko, F., *Biochem. Biophys. Res. Commun.* **2003,** *306* (2), 558.

21. Fang, M. Z.; Wang, Y.; Ai, N.; Hou, Z.; Sun, Y.; Lu, H.; Welsh, W.; Yang, C. S., *Cancer Res.* **2003,** *63* (22), 7563.

22. Liu, Z.; Liu, S.; Xie, Z.; Pavlovicz, R. E.; Wu, J.; Chen, P.; Aimiuwu, J.; Pang, J.; Bhasin, D.; Neviani, P.; Fuchs, J. R.; Plass, C.; Li, P. K.; Li, C.; Huang, T. H.; Wu, L. C.; Rush, L.; Wang, H.; Perrotti, D.; Marcucci, G.; Chan, K. K., *J. Pharmacol. Exp. Ther.* **2009,** *329* (2), 505.

23. Yoo, J.; Medina-Franco, J. L., *J. Comput. Aided Mol. Des.* **2011,** *25* (6), 555.

24. Weng, J.-R.; Lai, I. L.; Yang, H.-C.; Lin, C.-N.; Bai, L.-Y., *Phytother. Res.* **2014,** *28* (1), 49.

25. Joshi, M.; Rajpathak, S. N.; Narwade, S. C.; Deobagkar, D., *Chem. Biol. Drug Des.* **2016**, n/a.

26. Yoo, J.; Kim, J. H.; Robertson, K. D.; Medina-Franco, J. L., *Adv. Protein Chem. Struct. Biol.* **2012,** *87*, 10.1016/B978.

27. (a) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X., *Cell* **1994,** *76* (2), 357; (b) O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1996,** *261* (5), 634; (c) O'Gara, M.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1996,** *263* (4), 597; (d) O'Gara, M.; Zhang, X.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1999,** *287* (2), 201; (e) Zhou, L.; Cheng, X.; Connolly, B. A.; Dickman, M. J.; Hurd, P. J.; Hornby, D. P., *J. Mol. Biol.* **2002,** *321* (4), 591; (f) Shieh, F. K.; Youngblood, B.; Reich, N. O., *J. Mol. Biol.* **2006,** *362*

(3), 516; (g) Youngblood, B.; Shieh, F. K.; Buller, F.; Bullock, T.; Reich, N. O., *Biochemistry* **2007,** *46* (30), 8766.

28. (a) Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J., *Sci.* **2011,** *331* (6020), 1036; (b) Takeshita, K.; Suetake, I.; Yamashita, E.; Suga, M.; Narita, H.; Nakagawa, A.; Tajima, S., *Proc. Natl. Acad. Sci. U.S.A.* **2011**; (c) Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J., *Sci.* **2012,** *335* (6069), 709.

29. Zhang, Z. M.; Liu, S.; Lin, K.; Luo, Y.; Perry, J. J.; Wang, Y.; Song, J., *J. Mol. Biol.* **2015**.

30. Kumar, S.; Cheng, X.; Klimasauskas, S.; Mi, S.; Posfai, J.; Roberts, R. J.; Wilson, G. G., *Nucleic Acids Res.* **1994,** *22* (1), 1.

31. Lee, B.; Muller, Mark T., *Biochem. J.* **2009,** *421* (3), 449.

32. Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N., *Cell* **1995,** *82* (1), 143.

33. Dong, A.; Yoder, J. A.; Zhang, X.; Zhou, L.; Bestor, T. H.; Cheng, X., *Nucleic Acids Res.* **2001,** *29* (2), 439.

34. (a) Singh, N.; Duenas-Gonzalez, A.; Lyko, F.; Medina-Franco, J. L., *ChemMedChem* **2009,** *4* (5), 792; (b) Liu, Z.; Xie, Z.; Jones, W.; Pavlovicz, R. E.; Liu, S.; Yu, J.; Li, P. K.; Lin, J.; Fuchs, J. R.; Marcucci, G.; Li, C.; Chan, K. K., *Bioorg. Med. Chem. Lett.* **2009,** *19* (3), 706; (c) Kuck, D.; Singh, N.; Lyko, F.; Medina-Franco, J. L., *Bioorg. Med. Chem.* **2010,** *18* (2), 822; (d) Siedlecki, P.; Garcia Boy, R.; Musch, T.; Brueckner, B.; Suhai, S.; Lyko, F.; Zielenkiewicz, P., *J. Med. Chem.* **2006,** *49* (2), 678.

35. Yoo, J.; Medina-Franco, J. L., *J. Mol. Model.* **2012,** *18* (4), 1583.

36. (a) Wu, J. C.; Santi, D. V., *J. Biol. Chem.* **1987,** *262* (10), 4778; (b) Malygin, E. G.; Hattman, S., *Crit. Rev. Biochem. Mol. Biol.* **2012,** *47* (2), 97.

37. (a) Wyszynski, M. W.; Gabbara, S.; Bhagwat, A. S., *Nucleic Acids Res.* **1992,** *20* (2), 319; (b) Shieh, F. K.; Reich, N. O., *J. Mol. Biol.* **2007,** *373* (5), 1157; (c) Gabbara, S.; Sheluho, D.; Bhagwat, A. S., *Biochemistry* **1995,** *34* (27), 8914; (d) Schermelleh, L.; Haemmer, A.; Spada, F.; Rösing, N.; Meilinger, D.; Rothbauer, U.; Cardoso, M. C.; Leonhardt, H., *Nucleic Acids Res.* **2007,** *35* (13), 4301.

38. Chen, L.; MacMillan, A. M.; Verdine, G. L., *J. Am. Chem. Soc.* **1993,** *115* (12), 5318.

39. Santi, D. V.; Norment, A.; Garrett, C. E., *Proc. Natl. Acad. Sci. U.S.A.* **1984,** *81* (22), 6993.

40. (a) Peräkylä, M., *J. Am. Chem. Soc.* **1998,** *120* (49), 12895; (b) Zhang, X.; Bruice, T. C., *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (16), 6148; (c) Zangi, R.; Arrieta, A.; Cossío, F. P., *J. Mol. Biol.* **2010,** *400* (3), 632; (d) Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S., *Biochemistry* **2013,** *52* (16), 2828; (e) Aranda, J.; Zinovjev, K.; Świderek, K.; Roca, M.; Tuñón, I., *ACS Catal.* **2016,** *6* (5), 3262; (f) Du, Q.; Wang, Z.; Schramm, V. L., *Proc. Natl. Acad. Sci.* **2016,** *113* (11), 2916.

41. Becke, A. D., *J. Chem. Phys.* **1993,** *98* (7), 5648.

42. (a) Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1-3), 215; (b) Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008,** *41* (2), 157.

43. (a) Gu, X.; Zhu, Y. Z., *Expert Rev. Clin. Pharm.* **2011,** *4* (1), 123; (b) Li, Q.; Lancaster Jr, J. R., *Nitric Oxide* **2013,** *35*, 21; (c) Szabo, C., *Nat. Rev. Drug Discov.* **2007,** *6* (11), 917.

44. (a) Erickson, P. F.; Maxwell, I. H.; Su, L. J.; Baumann, M.; Glode, L. M., *Biochem. J.* **1990,** *269* (2), 335; (b) Jhee, K. H.; Kruger, W. D., *Antioxid. Redox Signal.* **2005,** *7* (5-6), 813; (c) Kery, V.; Bukovska, G.; Kraus, J. P., *J. Biol. Chem.* **1994,** *269* (41), 25283; (d) Kimura, H., *Antioxid. Redox Signal.* **2010,** *12* (9), 1111; (e) Meier, M.; Janosik, M.; Kery, V.; Kraus, J. P.; Burkhard, P., *EMBO J.* **2001,** *20* (15), 3910.

45. (a) Chen, K. Y.; Morris, J. C., *Environ. Sci. Technol.* **1972,** *6* (6), 529; (b) Kuhn, A. T.; Chana, M. S.; Kelsall, G. H., *J. Chem. Technol. Biotechnol.* **1983,** *33A* (8), 406; (c) O'Brien, D. J.; Birkner, F. B., *Environ. Sci. Technol.* **1977,** *11* (12), 1114.

46. (a) Asada, K. K., S., *Agri. Biol. Chem.* **1976,** *40* (9), 1891; (b) Searcy, D. G.; Whitehead, J. P.; Maroney, M. J., *Arch. Biochem. Biophys.* **1995,** *318* (2), 251.

47. (a) Classen, A. B., O., *J. Chem. Soc. Abstr.* **1883,** *44*, 934; (b) Hoffmann, M. R., *Environ. Sci. Technol.* **1977,** *11* (1), 61.

48. Carballal, S.; Trujillo, M.; Cuevasanta, E.; Bartesaghi, S.; Möller, M. N.; Folkes, L. K.; García-Bereguiaín, M. A.; Gutiérrez-Merino, C.; Wardman, P.; Denicola, A.; Radi, R.; Alvarez, B., *Free Radic. Biol. Med.* **2011,** *50* (1), 196.

49. (a) Laggner, H.; Muellner, M. K.; Schreier, S.; Sturm, B.; Hermann, M.; Exner, M.; Gmeiner, B. M.; Kapiotis, S., *Free Radic. Res.* **2007,** *41* (7),

741; (b) Nagy, P.; Winterbourn, C. C., *Chem. Res. Toxicol.* **2010,** *23* (10), 1541.

50. (a) Liu, D. K.; Chang, S. G., *Can. J. Chem.* **1987,** *65* (4), 770; (b) Rao, G. S.; Gorin, G., *J. Org. Chem.* **1959,** *24* (6), 749; (c) Schneider, J. F.; Westley, J., *J. Biol. Chem.* **1969,** *244* (20), 5735.

51. (a) Hylin, J. W.; Wood, J. L., *J. Biol. Chem.* **1959,** *234* (8), 2141; (b) Kutney, G. W. T., K., *Chem. Rev.* **1982,** *82* (4), 333; (c) Toohey, J. I., *Anal. Biochem.* **2011,** *413* (1), 1.

52. Munro, A. P.; Williams, D. L. H., *J. Chem. Soc., Perkin Trans. 2* **2000,** (9), 1794.

53. Cooper, C. E.; Brown, G. C., *J. Bioenerg. Biomembr.* **2008,** *40* (5), 533.

54. (a) Xu, X. M.; Moller, S. G., *Antioxid. Redox Signal.* **2011,** *15* (1), 271; (b) Qi, W.; Cowan, J. A., *Coord. Chem. Rev.* **2011,** *255* (7-8), 688.

55. (a) Yang, G.; Wu, L.; Jiang, B.; Yang, W.; Qi, J.; Cao, K.; Meng, Q.; Mustafa, A. K.; Mu, W.; Zhang, S.; Snyder, S. H.; Wang, R., *Sci.* **2008,** *322* (5901), 587; (b) Wang, R., *Kidney Int.* **2009,** *76* (7), 700.

56. Fiorucci, S.; Antonelli, E.; Distrutti, E.; Rizzo, G.; Mencarelli, A.; Orlandi, S.; Zanardo, R.; Renga, B.; Di Sante, M.; Morelli, A.; Cirino, G.; Wallace, J. L., *Gastroenterol.* **2005,** *129* (4), 1210.

57. (a) Kimura, Y.; Kimura, H., *FASEB J.* **2004,** *18* (10), 1165; (b) Jha, S.; Calvert, J. W.; Duranski, M. R.; Ramachandran, A.; Lefer, D. J., *Am. J. Physiol. Heart Cir. Physiol.* **2008,** *295* (2), H801; (c) Wei, H. L.; Zhang, C. Y.; Jin, H. F.; Tang, C. S.; Du, J. B., *Acta. Pharmacol. Sin.* **2008,** *29* (6), 670.

58. (a) Calvert, J. W.; Jha, S.; Gundewar, S.; Elrod, J. W.; Ramachandran, A.; Pattillo, C. B.; Kevil, C. G.; Lefer, D. J., *Circ. Res.* **2009,** *105* (4), 365; (b) Sodha, N. R.; Clements, R. T.; Feng, J.; Liu, Y.; Bianchi, C.; Horvath, E. M.; Szabo, C.; Sellke, F. W., *Eur. J. Cardiothorac. Surg.* **2008,** *33* (5), 906; (c) Sivarajah, A.; Collino, M.; Yasin, M.; Benetti, E.; Gallicchio, M.; Mazzon, E.; Cuzzocrea, S.; Fantozzi, R.; Thiemermann, C., *Shock* **2009,** *31* (3), 267; (d) Biermann, J.; Lagreze, W. A.; Schallner, N.; Schwer, C. I.; Goebel, U., *Mol. Vis.* **2011,** *17*, 1275; (e) Shi, S.; Li, Q. S.; Li, H.; Zhang, L.; Xu, M.; Cheng, J. L.; Peng, C. H.; Xu, C. Q.; Tian, Y., *Cell Biol. Int.* **2009,** *33* (10), 1095.

59. Banerjee, S. K.; Maulik, S. K., *Nutr. J.* **2002,** *1*, 4.

60. (a) Bir, S. C.; Kolluru, G. K.; McCarthy, P.; Shen, X.; Pardue, S.; Pattillo, C. B.; Kevil, C. G., *J. Am. Heart Assoc.* **2012,** *1* (5), e004093; (b) Cai, W. J.; Wang, M. J.; Moore, P. K.; Jin, H. M.; Yao, T.; Zhu, Y. C., *Cardiovas. Res.* **2007,** *76* (1), 29; (c) Papapetropoulos, A.; Pyriochou, A.; Altaany, Z.; Yang, G.; Marazioti, A.; Zhou, Z.; Jeschke, M. G.; Branski, L. K.; Herndon, D. N.; Wang, R.; Szabó, C., *Proc. Natl. Acad. Sci.* **2009,** *106* (51), 21972.

61. (a) Lavu, M.; Bhushan, S.; Lefer, D. J., *Clin. Sci. (Lond.)* **2011,** *120* (6), 219; (b) Zhang, Z.; Huang, H.; Liu, P.; Tang, C.; Wang, J., *Can. J. Physiol. Pharmacol.* **2007,** *85* (12), 1248.

62. (a) Li, R. Q.; McKinstry, A. R.; Moore, J. T.; Caltagarone, B. M.; Eckenhoff, M. F.; Eckenhoff, R. G.; Kelz, M. B., *J. Pharmacol. Exp. Ther.* **2012,** *341* (3), 735; (b) Blackstone, E.; Roth, M. B., *Shock* **2007,** *27* (4), 370; (c) Blackstone, E.; Morrison, M.; Roth, M. B., *Sci.* **2005,** *308* (5721), 518; (d)

Volpato, G. P.; Searles, R.; Yu, B.; Scherrer-Crosbie, M.; Bloch, K. D.; Ichinose, F.; Zapol, W. M., *Anesthesiology* **2008,** *108* (4), 659.

63. (a) Distrutti, E.; Mencarelli, A.; Santucci, L.; Renga, B.; Orlandi, S.; Donini, A.; Shah, V.; Fiorucci, S., *Hepatology* **2008,** *47* (2), 659; (b) Sen, U.; Sathnur, P. B.; Kundu, S.; Givvimani, S.; Coley, D. M.; Mishra, P. K.; Qipshidze, N.; Tyagi, N.; Metreveli, N.; Tyagi, S. C., *Am. J. Physiol. Cell Physiol.* **2012,** *303* (1), C41.

64. (a) Zhao, Y.; Biggs, T. D.; Xian, M., *Chem. Commun.* **2014,** *50* (80), 11788; (b) Song, Z. J.; Ng, M. Y.; Lee, Z.-W.; Dai, W.; Hagen, T.; Moore, P. K.; Huang, D.; Deng, L.-W.; Tan, C.-H., *Med. Chem. Comm.* **2014,** *5* (5), 557.

65. (a) Liang, D.; Wu, H.; Wong, M. W.; Huang, D., *Org. Lett.* **2015,** *17* (17), 4196; (b) Truong, D.; Hindmarsh, W.; O'Brien, P. J., *Chem. Biol. Interact.* **2009,** *180* (1), 79; (c) Benavides, G. A.; Squadrito, G. L.; Mills, R. W.; Patel, H. D.; Isbell, T. S.; Patel, R. P.; Darley-Usmar, V. M.; Doeller, J. E.; Kraus, D. W., *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (46), 17977.

66. (a) Amagase, H., *J. Nutr.* **2006,** *136* (3 Suppl), 716S; (b) Amagase, H.; Petesch, B. L.; Matsuura, H.; Kasuga, S.; Itakura, Y., *J. Nutr.* **2001,** *131* (3s), 955S.

Chapter 2 Theoretical Methodology

This chapter gives an overview of the theoretical methodology in the computational chemistry relevant to this thesis. Sections 2.1–2.7 introduce some fundamental topics in quantum mechanics, including the Schrödinger equation and the approximations to solve it, the *ab initio* methods and the Density Functional Theory methods. Section 2.8 deals with the modeling of solvated systems. The more detailed information on quantum chemistry and solvation can be found in several textbooks[1]. Sections 2.9 and 2.10 describe two types of molecular modeling methods—homology modeling and classical molecular dynamics—that can be used in conjunction to study biomolecules.

2.1 The Schrödinger Equation

In quantum mechanics (QM), the microscopic properties of a system are described by the wavefunction and the Schrödinger equation[2]. In chemistry, the stationary states of a system are described by the time-independent Schrödinger equation:

$$\boldsymbol{H\psi = E\psi} \tag{2.1}$$

$$\boldsymbol{H = T + V} \tag{2.2}$$

where H is the Hamiltonian operator of the system, E is the energy (eigenvalue) of the system and ψ is the wavefunction (eigenvector) of H. The Hamiltonian operator H is the sum of the kinetic energy term T and the potential energy term V.

For a system with N electrons and M nuclei, the Hamiltonian operator[1c] is:

$$\boldsymbol{H} = -\frac{\mathbf{1}}{\mathbf{2}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\boldsymbol{\nabla_i^2} - \frac{\mathbf{1}}{\mathbf{2}}\sum_{\boldsymbol{A=1}}^{\boldsymbol{M}}\frac{\mathbf{1}}{\boldsymbol{M_A}}\boldsymbol{\nabla_A^2} - \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\sum_{\boldsymbol{A=1}}^{\boldsymbol{M}}\frac{\boldsymbol{Z_A}}{\boldsymbol{r_{iA}}} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\sum_{\boldsymbol{j>i}}^{\boldsymbol{N}}\frac{\mathbf{1}}{\boldsymbol{r_{ij}}} + \sum_{\boldsymbol{A=1}}^{\boldsymbol{M}}\sum_{\boldsymbol{B>A}}^{\boldsymbol{M}}\frac{\boldsymbol{Z_A Z_B}}{\boldsymbol{R_{AB}}} \tag{2.3}$$

$$\boldsymbol{\nabla^2} = \frac{\boldsymbol{d^2}}{\boldsymbol{dx^2}} + \frac{\boldsymbol{d^2}}{\boldsymbol{dy^2}} + \frac{\boldsymbol{d^2}}{\boldsymbol{dz^2}} \tag{2.4}$$

where i and j denote electrons, A and B denote nuclei, M_A and Z_A are the mass and the charge of the nucleus A, and $r_{iA}/r_{ij}/R_{AB}$ are the distances between different particles. The first two terms in (2.3) are kinetic energy terms of electrons and nuclei respectively, and the last three terms are potential energy terms to account for the electron−nuclear attraction, electron−electron repulsion and nuclear−nuclear repulsion respectively.

2.2 Born−Oppenheimer Approximation

Because nuclei are much heavier than electrons, they move much more slowly than electrons. Thus, the nuclear and electronic motions can be decoupled, considering the motions of electrons in the presence of fixed nuclei. This is the basic idea of the Born−Oppenheimer approximation[3].

Based on this approximation, an electronic Hamiltonian[1c] is constructed by neglecting the kinetic energy term of nuclei and considering the nuclear−nuclear repulsion to be constant:

$$\boldsymbol{H_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}} \quad \textbf{(2.5)}$$

The electronic wavefunction and the pure electronic energy E_{elec} are solved from the electronic Schrödinger equation:

$$\boldsymbol{H_{elec}\psi_{elec}(r;R) = E_{elec}(R)\psi_{elec}(r;R)} \quad \textbf{(2.6)}$$

The electronic wavefunction ψ_{elec} depends on the electronic coordinates r at a specific set of nuclear coordinates R. Similarly, the pure electronic energy E_{elec} also depends on the nuclear coordinates R.

The total electronic energy with fixed nuclei should also include the constant nuclear−nuclear repulsion:

$$\boldsymbol{E_{tot}(R) = E_{elec}(R) + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}} \quad \textbf{(2.7)}$$

2.3 Potential Energy Surface and Thermochemistry Corrections

The plot of total electronic energy E of a molecule against nuclear coordinates defines the potential energy surface (PES) that describes the nuclear motions. The PES is multi-dimensional, though in practice, one usually examines 1−2 specific dimensions of interest.

Stationary points with zero gradient on a molecular PES are highly important in chemistry studies. Stationary points with positive curvatures in all

dimensions are equilibrium structures that correspond to reactants, products or intermediates. Stationary points with one and only one negative curvature in one dimension but positive curvatures in all other dimensions are first-order saddle points that correspond to transition states.

The total electronic energy describes the electronic energy of a motionless molecule at zero Kelvin. However, according to the Heisenberg's uncertainty principle[4], molecules vibrate even at zero Kelvin, so the lowest vibrational energy level is higher than the total electronic energy. This energy correction is the zero-point vibrational energy (ZPVE)[1a] and can be determined by the harmonic oscillator approximation.

Thermodynamic properties such as enthalpies *H* and Gibbs free energies *G* can be calculated with corrections to *H* and entropy *S*. These terms are computed from statistical mechanics for an N-particle canonical ensemble at the fixed volume V and temperature T [5]:

$$\boldsymbol{H = U + PV = k_B T^2 (\frac{\partial lnQ}{\partial T})_{N,V} + k_B TV (\frac{\partial lnQ}{\partial V})_{T,N}} \quad \textbf{(2.8)}$$

$$\begin{aligned} \boldsymbol{G = H - TS} &= \boldsymbol{H - [k_B T lnQ + k_B T^2 \left(\frac{\partial lnQ}{\partial T}\right)_{N,V}]} \\ &= \boldsymbol{k_B TV (\frac{\partial lnQ}{\partial V})_{T,N} - k_B T lnQ} \end{aligned} \quad \textbf{(2.9)}$$

where Q represents the total partition functions, which is a function of T, N and V, and is expressed as the product of the electronic, translational, rotational and vibrational partition functions[5]. The electronic partition function is the spin multiplicity of the molecule. The translational and rotational partition functions are easy to calculate from certain physical properties of the

molecule. The vibrational partition function is more complicated to compute that requires the calculation of all vibrational modes from frequency calculations.

2.4 Hartree−Fock Theory

The Hartree−Fock (HF) method[6] is an approximation method to solve the time-independent Schrödinger Equation in many-electron systems based on the Born−Oppenheimer approximation and the idea of separating the wavefunction ψ into individual one-electron spin orbitals. By implementing the variational method[1b], the modified equation is solved iteratively to give the HF wavefunction and energies. The HF method is the basis for most wavefunction-based *ab initio* methods.

2.4.1 Variational Method

The variational method[1b] is one of the most useful methods to find an approximate solution to the time-independent Schrödinger equation. This method is based on the Variation Theorem.

For a system with a normalized wavefunction Ψ which satisfies the necessary boundary conditions, the expectation value E of the Hamiltonian H is an upper bound to the (ground-state) lowest energy of the system E_0:

$$\boldsymbol{E_{trial} = \frac{\langle \Psi_{trial} | H | \Psi_{trial} \rangle}{\langle \Psi_{trial} | \Psi_{trial} \rangle} = \langle \Psi_{trial} | H | \Psi_{trial} \rangle \geq E_0} \qquad \textbf{(2.10)}$$

Thus, for any normalized trial wavefunction Ψ_{trial}, the calculated energy E_{trial} is never lower than the true energy E_0. As E_{trial} gets closer to E_0, Ψ_{trial} gets closer to the exact wavefunction Ψ_o.

In the variational method, a given Ψ_{trial} which depends on certain parameters is used as the starting point for further optimization by evaluating and optimizing E_{trial} iteratively from the change of parameters until Ψ_{trial} that gives the lowest energy is found.

2.4.2 Hartree−Fock Approximation

For an N-electron system occupying N spin orbitals, assuming that each electron interacts with the an average field of other electrons (Hartree Approximation[6d]), the wavefunction of the system can be described by a determinant of individual one-electron spin orbitals, which is termed as a Slater Determinant[7]:

$$\boldsymbol{\Psi}(\mathbf{1},\ldots,\boldsymbol{N}) = \frac{\mathbf{1}}{\sqrt{\boldsymbol{N}!}}\begin{vmatrix} \boldsymbol{\phi_1}(\mathbf{1}) & \cdots & \boldsymbol{\phi_N}(\mathbf{1}) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi_1}(\boldsymbol{N}) & \cdots & \boldsymbol{\phi_N}(\boldsymbol{N}) \end{vmatrix} \qquad \textbf{(2.11)}$$

where $1/\sqrt{N!}$ is the normalization factor, and $\phi_i(j)$ describes the i^{th} spin orbital occupied by the j^{th} electron. A spin orbital ϕ is a one-electron molecular orbital, which is the product of the spatial orbital and the spin function (spin up or spin down).

The Slater determinant in (2.11) satisfies the antisymmetry principle as the sign of the determinant changes when any two electrons are exchanged. It

also satisfies the Pauli Exclusion Principle as the determinant vanishes when any two electrons occupy the same spin orbital.

2.4.3 Basis Set

The molecular orbital can be constructed as a linear combination of N predefined one-electron basis functions:

$$\boldsymbol{\phi_i = \sum_{\mu=1}^{N} c_{\mu i} \chi_\mu} \tag{2.12}$$

where $c_{\mu i}$ denotes the coefficient associated with the basis function χ_μ.

Basis functions are mathematical descriptions of atomic orbitals and the collection of basis functions is called a basis set.

Both Slater-type orbitals[8] (STOs) and Gaussian-type functions[9] (GTFs) can be used to express basis functions. An STO is characterized by $\exp(-\xi r)$ while a GTF is characterized by $\exp(-\alpha r^2)$, where ξ and α are the Slater orbital exponent and Gaussian orbital exponent respectively, and r is distance from the nucleus. The STOs, though provide a more accurate description of atomic orbitals than GTFs especially close to or far away from the nucleus, is more computationally expensive in the molecular integral evaluations. The GTF is much more efficient in the integral evaluations and the linear combination of several primitive GTFs resembles an STO in the atomic orbital description. Such basis functions are contracted Gaussians. Pople's basis sets are some of the most widely used Gaussian-type basis sets[1].

A minimal basis set is the simplest type of basis sets where each atomic orbital of the atom is described by a single contracted GTF. Take STO-3G[10] as an example, each atomic orbital in each atom is described by a contracted GTF consisting of three primitive GTFs to mimic an STO. Such basis sets are seldom used in calculations as they do not allow the change of orbital size that is usually desirable to fit different molecular environments.

Larger basis sets use more than one contracted GTF for each atomic orbital. A split-valence basis set uses one contracted GTF for each core orbital but two or more contracted GTFs for each valence orbital in each atom. The additional contracted GTFs for the valence orbitals are more diffuse and help improve the chemical bonding descriptions. The double-split valence basis set 6-31G[11] uses a contracted GTF with 6 primitives for the core orbitals and two contracted GTFs with 3 and 1 primitives for the valence orbitals. The triple-split valence basis set 6-311G[12] uses one more GTF with 1 primitive for the valence orbitals.

Polarization functions[13] and diffuse functions[13b, 14] are useful additions to the basis sets. Polarization functions are contracted GTFs whose angular momentum are larger than the valence orbitals to allow the orbitals to change shapes. 6-31G(d,p) or 6-31G** is formed from 6-31G with d functions added to heavy atoms (first *) and p functions added to hydrogen and helium (second *). Diffuse functions are valence orbitals with a larger size that allows the orbital to expand into a larger space, and they are important for systems with lone pairs (LPs) and anions. 6-31++G* is formed from 6-31G* with larger s

and p functions added to heavy atoms (first +) and diffuse p functions added to hydrogen and helium (second +).

The choice of basis sets is important for the accuracy and the computational cost of a calculation. The more accurate results usually require a larger basis set at a higher computational cost. When the number of basis functions in a basis set is close to infinity, the calculation is said to be close to the complete basis set (CBS) limit.

2.4.4 The Hartree−Fock Equation

Based on the HF approximation and the Variation Theorem, the electronic Hamiltonian is simplified to the Fock operator[6a, 6b], which is a one-electron Hamiltonian:

$$\boldsymbol{F}[\{\boldsymbol{\phi_i}\}](\mathbf{1}) = \boldsymbol{H^{core}}(\mathbf{1}) + \sum_{\boldsymbol{j=1}}^{\frac{\boldsymbol{N}}{\mathbf{2}}}[\mathbf{2}\boldsymbol{J_j}(\mathbf{1}) - \boldsymbol{K_j}(\mathbf{1})] \tag{2.13}$$

where $H^{core}(1)$ is the core Hamiltonian operator which is the sum of the kinetic energy term and the nuclear−electron attraction term, $J_j(1)$ is the Coulomb operator for the repulsion between the electron 1 and the rest of electrons, and $K_j(1)$ is the exchange operator arising from the antisymmetry principle for electrons with the same spin. The new HF equation[6d] based on (2.13) is as follows:

$$\boldsymbol{F}[\{\boldsymbol{\phi_i}\}](\mathbf{1})\boldsymbol{\phi_i}(\mathbf{1}) = \boldsymbol{\varepsilon_i}\boldsymbol{\phi_i}(\mathbf{1}) \tag{2.14}$$

The variational method is used to solve for the HF equation. An initial guess for all the orbitals ϕ_i is used to construct the Fock operator, and the new set of ϕ_i is calculated from the HF equation and compared with the old set of ϕ_i. This process is repeated until ϕ_i no longer varies.

By substituting the basis sets into the HF equation, we have the Roothaan−Hall equation in a matrix form[15]:

$$\boldsymbol{FC_i = \varepsilon_i SC_i} \qquad \textbf{(2.15)}$$

where F is the Fock matrix, C_i is the coefficient matrix and S is the overlap matrix.

The HF method is also known as the self-consistent field (SCF) method because the final orbitals calculated from the HF equation are identical to the orbitals used in the Fock operator.

Because of the implementation of the Hartree Approximation, the electron correlation between electrons of the opposite spin is not accounted for in the HF method. The difference between the exact energy of the system with full electron correlation and the HF energy is the correlation energy. The electron correlation is treated in some higher-level *ab initio* methods, which are called the post-HF methods. One of the approaches is based on the Perturbation Theory.

2.5 Perturbation Theory

The Perturbation Theory[1b, 16] is another important approximation to solve the Schrödinger equation and can be applied to any electronic states including excited states.

To find the solution to the time-independent Schrödinger equation with the Hamiltonian H, we assume that we already know the exact solution the simpler system with the Hamiltonian $H^{(0)}$ that is slightly different from H:

$$\boldsymbol{H\psi_i = (H^{(0)} + \lambda V)\psi_i = E_i\psi_i} \quad \textbf{(2.16)}$$

$$\boldsymbol{H^{(0)}\psi_i{}^{(0)} = E_i{}^{(0)}\psi_i{}^{(0)}} \quad \textbf{(2.17)}$$

where V is a perturbation on $H^{(0)}$ with an order parameter λ that ranges from 0 to 1, $H^{(0)}$ and H are the unperturbed and perturbed Hamiltonians respectively. If V is small, then E_i and ψ_i can be expressed as a Taylor series based on $E_i{}^{(0)}$ and $\psi_i{}^{(0)}$:

$$\boldsymbol{E_i = E_i{}^{(0)} + \lambda E_i{}^{(1)} + \lambda^2 E_i{}^{(2)} + \cdots} \quad \textbf{(2.18)}$$

$$\boldsymbol{\psi_i = \psi_i{}^{(0)} + \lambda\psi_i{}^{(1)} + \lambda^2\psi_i{}^{(2)} + \cdots} \quad \textbf{(2.19)}$$

where $E_i{}^{(j)}$ and $\psi_i{}^{(j)}$ are the j^{th} order corrections to the energy and wavefunction respectively.

By substituting (2.18) and (2.19) to (2.16), we have:

$$\boldsymbol{\left(H^{(0)} + \lambda V\right)\left(\psi_i{}^{(0)} + \lambda\psi_i{}^{(1)} + \cdots\right)} \quad \textbf{(2.20)}$$
$$\boldsymbol{= (E_i{}^{(0)} + \lambda E_i{}^{(1)} + \cdots)(\psi_i{}^{(0)} + \lambda\psi_i{}^{(1)} + \cdots)}$$

After the rearrangement of (2.20), we get a set of perturbation equations for the j^{th} order corrections:

$$\boldsymbol{H^{(0)}\psi_i^{(0)} = E_i^{(0)}\psi_i^{(0)}} \quad \textbf{(2.21)}$$

$$\boldsymbol{H^{(0)}\psi_i^{(j)} + V\psi_i^{(j-1)} = E_i^{(0)}\psi_i^{(j)} + E_i^{(1)}\psi_i^{(j-1)} + \cdots + E_i^{(j)}\psi_i^{(0)}, \quad j = 1, \ldots, n} \quad \textbf{(2.22)}$$

Timing $\psi_i^{(0)}$ to both sides of the perturbation equations and assuming that any order correction is orthogonal to the unperturbed wavefunction $\psi_i^{(0)}$, we then have the solution to the j^{th} order energies:

$$\boldsymbol{E_i^{(0)} = \langle \psi_i^{(0)} | H | \psi_i^{(0)} \rangle} \quad \textbf{(2.23)}$$

$$\boldsymbol{E_i^{(j)} = \langle \psi_i^{(0)} | V | \psi_i^{(j)} \rangle, j = 1, \ldots, n} \quad \textbf{(2.24)}$$

The first and higher order corrections to the wavefunction can be solved from (2.22).

The Perturbation Theory is the basis for the Møller−Plesset (MP) methods[17], which improved the HF methods by adding the electron correlation as perturbations. The name MPn means the incorporation of the n^{th} order perturbation terms to the energy and the wavefunction.

2.6 Coupled-Cluster Theory

The coupled-cluster (CC) theory[1b, 18] is one of the most accurate *ab initio* methods to estimate the electron correlation energy based on the HF theory.

The exact ground-state electronic wavefunction in the CC theory is similar to that in the full configuration interaction (CI) theory and is expressed as:

$$\boldsymbol{\psi = e^{T}\psi_{HF}} \tag{2.25}$$

where e^{T} for an n-electron system is defined by a Taylor series based on the cluster operator T:

$$\boldsymbol{e^{T} = 1 + T + \frac{T^{2}}{2!} + \frac{T^{3}}{3!} + \cdots} \tag{2.26}$$

The cluster operator T is:

$$\boldsymbol{T = T_{1} + T_{2} + \cdots + T_{n}} \tag{2.27}$$

where T_1 is the operator that describes all single electron excitations from occupied spin orbitals to virtual spin orbitals, T_2 is the operator that describes all double electron excitations from occupied spin orbitals to virtual spin orbitals and so on. Among these operators, T_2 is the most important contributor to T.

Two approximations are made in the CC calculations. The first one is the use of a finite basis set instead of the complete basis set (CBS) in the full CI calculations. The second one is the inclusion of some of the excitation operators T_i instead of all the operators to approximate the cluster operator T.

The CCD, CCSD and CCSDT methods are acronyms for the CC methods that include double (D) excitations, single and double (SD) excitations, and single, double and triple (SDT) excitations, respectively.

2.7 Density Functional Theory

The accuracy of the wavefunction-based *ab initio* methods (HF and post-HF methods) can be improved systematically by expanding the basis set and using higher-level electron correlated methods. However, these methods are computationally expensive and the system size is highly restricted with the more accurate approaches.

The Density Functional Theory (DFT) methods[1a, 1c, 19] are based on the electron density of the system that contains much fewer variables than the wavefunction, and the electron exchange and correlation are explicitly considered in the formulation. Therefore, the DFT methods address the high computational cost of the post-HF methods and the lack of electron correlation in the HF method. These methods have become very popular in the study of the geometries and other molecular properties with accuracy comparable to the wavefunction-based *ab initio* methods.

However, due to the approximations in the exchange-correlation terms, the DFT methods cannot be improved systematically and the calculated energies may be lower than the true energies of the system.

2.7.1 Hohenberg−Kohn Theorems

In 1964, Hohenberg and Kohn[20] proved two theorems that are fundamental to the establishment of DFT methods. The first Hohenberg−Kohn

Theorem proved that the ground-state energy and all other electronic properties of a system are uniquely determined by the ground-state electron density $\rho_0(\vec{r})$. In other words, the ground-state electronic energy E_0 is a functional of ρ_0:

$$\begin{aligned} \boldsymbol{E_0} = \boldsymbol{E}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] &= \boldsymbol{T}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] + \boldsymbol{V_{Ne}}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] + \boldsymbol{V_{ee}}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] \\ &= \boldsymbol{T}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] + \int \boldsymbol{\rho_0}(\vec{\boldsymbol{r}})\boldsymbol{v}(\vec{\boldsymbol{r}})\boldsymbol{d}\vec{\boldsymbol{r}} + \boldsymbol{V_{ee}}[\boldsymbol{\rho_0}(\vec{\boldsymbol{r}})] \end{aligned} \quad \textbf{(2.28)}$$

where T is the electronic kinetic energy term, V_{Ne} and V_{ee} are the nuclear−electron attraction and electron−electron repulsion terms respectively, and $v(\vec{r})$ is the external potential on the electron *i* and depends on its xyz coordinates. However, the V_{ee} and T functionals are unknown.

The second Hohenberg−Kohn Theorem proved that the true ground-state electron density $\rho_0(\vec{r})$ minimizes the energy functional $E[\rho_{trial}(\vec{r})]$ of any trial density function $\rho_{trial}(\vec{r})$, just like what is stated in the Variation Theorem.

2.7.2 Kohn−Sham Theorem

To find out the unknown functionals and the expression of electron density in the Hohenberg−Kohn equation, Kohn and Sham[21] considered a fictitious reference system *s* with the same number of non-interacting electrons *n* as the actual system and described its ground-state wave function as a Slater determinant of the lowest-energy Kohn−Sham spin orbitals. The new Kohn−Sham energy is:

$$\begin{aligned} \boldsymbol{E[\rho(\vec{r})] = T_s[\rho(\vec{r})] + V_{Ne}[\rho(\vec{r})] + J[\rho(\vec{r})]} \\ \boldsymbol{+\Delta T_s[\rho(\vec{r})] + \Delta V_{ee}[\rho(\vec{r})]} \\ \boldsymbol{= T_s[\rho(\vec{r})] + V_{Ne}[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})]} \end{aligned} \quad \textbf{(2.29)}$$

where $T_s[\rho(\vec{r})]$ is the electronic kinetic energy for the reference system, $J[\rho(\vec{r})]$ is the repulsion between electrons in the average field, $\Delta T_s[\rho(\vec{r})]$ and $\Delta V_{ee}[\rho(\vec{r})]$ are the energy differences between the actual system and reference system in the kinetic energy and electron-electron repulsion respectively. $E_{XC}[\rho(\vec{r})]$ is the sum of the last two terms and is called the exchange-correlation energy functional.

The ground-state electron density of the system is the same as that of the fictitious system *s* and is given by the Kohn−Sham spin orbitals ψ_i:

$$\boldsymbol{\rho = \rho_s = \sum_i |\psi_i|^2} \quad \textbf{(2.30)}$$

Thus, the full energy expression is:

$$\begin{aligned} \boldsymbol{E[\rho(\vec{r})] = T_s[\rho(\vec{r})] + V_{Ne}[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})]} \\ \boldsymbol{= -\frac{1}{2}\sum_i \langle \psi_i(1)|\nabla_1{}^2|\psi_i(1)\rangle - \sum_A \sum_i Z_A \int \frac{|\psi_i(\vec{r_1})|^2}{\vec{r_{1A}}} d\vec{r_1}} \\ \boldsymbol{+\frac{1}{2}\iint \frac{|\psi_i(\vec{r_1})|^2 |\psi_i(\vec{r_2})|^2}{\vec{r_{12}}} d\vec{r_1} d\vec{r_2} + E_{XC}[\rho(\vec{r})]} \end{aligned} \quad \textbf{(2.31)}$$

The Kohn−Sham orbitals can be solved from the Kohn−Sham equations using the variational method:

$$\boldsymbol{\left[-\frac{1}{2}\nabla_1{}^2 + \sum_A \frac{Z_A}{\vec{r_{1A}}} + \int \frac{\rho(\vec{r_2})}{\vec{r_{12}}} d\vec{r_2} + v_{XC}(1)\right]\psi_i(1) = \varepsilon_i \psi_i(1)} \quad \textbf{(2.32)}$$

where v_{XC} is called the exchange-correlation potential, which is the functional derivative $\delta E_{XC}/\delta\rho$.

2.7.3 The Exchange-Correlation Functionals

Although the Kohn−Sham equation is exact in theory, the exact functional for E_{XC} and the derivative v_{XC} are unknown, and these terms are approximated in various ways. The DFT methods with increased complexity in the approximate exchange-correlation potentials are typically classified into the Local (Spin) Density Approximation (L(S)DA) methods, the Generalized Gradient Approximation (GGA) methods, the meta-GGA methods and the hybrid DFT methods[1b].

The LDA is based on an electrically neutral system with a homogeneous electron gas moving in a uniformly distributed and positively charged background. The exchange-correlation potential is evaluated at the 'local' value of electron density and can be split into the exchange and correlation parts. The exchange part has a simple analytical solution and is commonly expressed by the Slater exchange[22]. The correlation part can be described the complex Vosko−Wilk−Nusair (VWN) functional[23], which was derived from the density fitting to some accurate numerical results. While the LDA assigns the same Kohn−Sham orbital to paired electrons, the LSDA separates the Kohn−Sham orbitals for electrons with different spins to give better description of open-shell molecules and dissociating molecules[1b].

The GGA includes the reduced gradient correction to the LSDA to account for the non-homogeneity of electron distribution. Some exchange or

correlation functionals include one or more empirical parameters fitting to the known values of different molecular properties. Some of the popular exchange functionals are Becke88[24], PW91[25], and the parameter-free PBE[26]. Some of the popular correlation functionals are Lee−Yang−Parr (LYP)[27], P86[28], and the parameter-free PW91[25] and PBE[26].

The meta-GGA further includes the second derivative of electron density Laplacian or the kinetic energy density to the GGA. Such functionals include TPSS[29] and M06L[30].

The hybrid DFT methods incorporate some HF exact exchange with the DFT exchange-correlation functionals.

In principle, one can construct a DFT exchange-correlation functional by mixing any exchange functional with any correlation functional. Some of the popular DFT methods are the B3LYP functional and the M06 family of functionals.

B3LYP (Becke, three-parameter, Lee−Yang−Parr)[31] is a hybrid-GGA functional that dominates in computational chemistry. It combines some HF exchange, the Becke88 exchange functional and the LYP correlation functional. The B3LYP exchange-correlation functional[1b] is as follows:

$$\begin{gathered} \boldsymbol{E_{xc}^{B3LYP} = (1 - a_o - a_x)E_x^{LSDA} + a_o E_x^{HF} + a_x E_x^{B88}} \\ \boldsymbol{+(1 - a_c)E_c^{VWN} + a_c E_c^{LYP},} \\ \boldsymbol{a_o = 0.20, a_x = 0.72, a_c = 0.81} \end{gathered} \qquad \textbf{(2.33)}$$

where E_x^{LSDA} is the LSDA exchange functional, E_x^{HF} is the HF exact exchange, E_x^{B88} is the Becke's 1988 exchange functional, E_c^{VWN} is the VWN LSDA correlation functional and E_c^{LYP} is the LYP correlation functional. The three

empirical parameter values of a_o, a_x, a_c were optimized to fit the experimental molecular atomization energies. B3LYP generally performs well in the calculation of geometries and some thermochemical quantities like atomization energies for small organic molecules, but it cannot give reliable results to systems with certain non-covalent interactions, such as weak and π-π interactions.

The M06 (Minnesota 06)[32] family of functionals are a set of meta-GGA functionals developed by Zhao and Truhlar at the University of Minnesota. The general exchange-correlation functional is:

$$\boldsymbol{E_{xc} = \frac{X}{100} E_x^{HF} + \left(1 - \frac{X}{100}\right) E_x^{M06} + E_c^{M06}} \qquad \textbf{(2.34)}$$

where E_x^{HF} is the HF exchange, X is the percentage of the HF exchange, and E_x^{M06} and E_c^{M06} are the meta-GGA exchange and correlation functionals respectively. The M06 functionals are highly parameterized, and each functional, i.e. E_x^{M06} and E_c^{M06}, has about 35 parameters. Depending on the percentage of the HF exchange, the M06 family consists of four functionals: M06-L with 0% E_x^{HF}, M06 with 27% E_x^{HF}, M06-2X with 54% E_x^{HF} and M06-HF with 100% E_x^{HF}. The M06 family of functionals has been shown to perform well in describing the weak and π-π interactions. The M06-2X functional is particularly good in studying the thermochemistry, kinetics and non-covalent interactions for organic systems, and this functional gives comparable results to some of the later series of the Minnesota functionals, i.e. the M08[33] family and the M11[34] family.

2.8 Solvation Methods

While computational studies of molecular systems are usually carried out in the gas phase by treating molecules as isolated, non-interacting species, most molecular systems in chemical processes are in the solution phase. Because the solute and the solvent molecules interact in solutions, the geometries and properties of the solvated systems may differ significantly from the gas phase systems especially when a polar solvent is present.

The effects of solvation are modeled either by explicitly incorporating individual solvent molecules (i.e. explicit solvent) or by implicitly treating the solvent as a continuous medium (i.e. implicit solvent).

The explicit solvent models are highly computationally expensive with *ab initio* and DFT methods, thus one may only include very few important solvent molecules in the model if desired. However, explicit solvent models are common in molecular dynamics simulation with non-quantum mechanical methods, which will be described in section 2.10.

The implicit solvent models treat a solvent as a polarizable medium and place the solute molecule in a specific cavity inside the medium. The dipole of the solute induces an electric 'reaction field' in the cavity surface that polarizes the surrounding medium, and the polarized solvent in turn polarizes the solute. This mutual polarization gives rise to the self-consistency in the system, thus implicit solvent models are generally known as self-consistent reaction field (SCRF) models. The solvation energy contributors include the cavitation energy needed in the cavity creation, the electrostatic interaction energy, the dispersion interaction energy, the exchange repulsion energy due

to the Pauli repulsion, and the energy due to change in the solvent structure. Some examples of the implicit solvent models are PCM and SMD.

The PCM (polarizable continuum model)[35] is one of the commonly used implicit solvent models. The solute cavity is defined by the interlocking atom-centered spheres with van der Waals radii. Each sphere surface is divided into small regions known as the surface element and each surface element has one point charge. The reaction field is described by these point charges. The solute-solvent interaction potential is added to the gas phase solute Hamiltonian to construct the solution phase wavefunction. The PCM has many variants, such as the integral equation formalism variant (IEFPCM)[36] and the isodensity surface variant (IPCM)[37]. The PCM primarily accounts for the electrostatic contribution to the free energy of solvation and is not suitable for non-polar solvents.

The SMD (solvation model based on density)[38] is one of the universal solvation models developed by Cramer and Truhlar. The density means that the full solute electron density is considered. The solvation energy contributors are the bulk-electrostatic energy from the IEFPCM algorithm, and the cavity-dispersion-solvent structure term from short-range interactions between the solute and the molecules in the first solvation shell. The cavities are defined by superposition of spheres centered on nuclei. The cavity for the bulk-electrostatic term is defined by the interlocking of atom-centered spheres with intrinsic atomic Coulomb radii. The SMD is highly parameterized and has been shown to give good results for free energies of solvation for both polar and non-polar solvents.

2.9 Homology Modeling

Homology modeling[39] is a useful computational method for protein structure prediction. The aim is to model a protein with unknown structure (the target) based on its amino acid sequence and one or more related proteins with known 3D structures (the templates). It relies on the identification of the similarity between the target sequence and the template sequences and the proper sequence alignment. The basis of this method is that the 3D structures of proteins in the same family are more conserved than the corresponding amino acid sequences[40]. It is known that proteins with at least 25% sequence identity have similar structures[41].

Homology modeling usually consists of the following five steps: template identification, template selection, target-template sequence alignment, model construction, and model evaluation. Some or all of the steps may be repeated until the model is deemed satisfactory.

The first three steps are usually carried out together. The template identification starts by searching the database of known protein structures against the target sequence. This is done via sequence alignment methods[42] that compare target sequence with other protein sequences or fold-recognition methods known as threading[43]. After the identification of potential templates, it is necessary to select the appropriate templates for model construction. The basic idea is to choose the templates with the higher overall sequence similarity (i.e. higher sequence identity and no large insertions or deletions) and higher quality of the 3D structure. If a protein−ligand model is the target, then the chosen template should have a similar ligand instead of the protein

alone. Once the templates are selected, the target and the template residues will be compared for the optimal target−template alignment. It is usually much more difficult to obtain an accurate alignment when the target−template sequence identity is less than 40% and the alignment accuracy in this case largely affects the quality of the final model[44].

Given the target−template alignment and the 3D structures of the templates, a 3D model of the target can be constructed by different methods, including rigid-body assembly methods, segment matching methods and methods based on 'satisfaction of spatial restraints'. The rigid-body assembly methods[45] construct the model by connecting the conserved core regions and rebuilding the loops and side chains. The segment matching methods[46] construct the model by combining short segments of the target protein, each built from the matching segments in the template structures. Both the loops and the side chains can be modeled in this way. The methods based on 'satisfaction of spatial constraints'[47] construct the model by minimizing the violations of spatial restraints derived from the template structures in the template−target alignment. Such method is used in the program MODELLER[47b]. The complete model from the homology modeling programs can be refined further by geometry optimization and other molecular modeling techniques such as molecular dynamics simulations that will be described in section 2.10 to improve the model quality.

Finally, the complete model will be assessed for potential problems in various ways[39b]. For instance, the model can be compared with the template structures in terms of the Root-Mean-Square Deviation (RMSD) and some

other structural features such as the Ramachandran plot, atom clashes, bond lengths, bond angles and dihedrals. Moreover, the model can be evaluated by some energy scores.

2.10 Classical Molecular Dynamics

Classical molecular dynamics (MD)[48] is a computational method that simulates the time dependent behavior of a system by integrating Newton's equation of motion. It is useful to study the conformations, dynamics and thermodynamics of a system, and has been widely applied in the simulations of materials and biomolecules. The more detailed information on the classical MD can be found in textbooks[49].

A typical MD simulation consists of the following three stages. First, the system is energy minimized to remove excess energy. Second, an equilibration phase is carried out on the minimized system to equilibrate the kinetic and potential energies until the properties of the system no longer change. Third, a production phase is carried out on the equilibrated system to collect thermodynamic data and the trajectory (i.e. atomic coordinates and velocities). Heating or cooling can be included in the MD simulation, and it aims to sample the conformational space by overcoming the conformational barriers when the temperature rises and further stabilizing the system when the temperature lowers.

In practice, an MD simulation requires the preparation of the molecular system and a set of parameters, conditions and algorithms to follow in the simulation. Energy of the system is usually calculated by molecular mechanics

(MM) or force fields, and statistical mechanics is used to convert the microscopic information such as atom positions and velocities to macroscopic properties such as pressure, temperature and energy. The actual MD starts with the initialization of the system by setting initial atom positions and assigning initial velocities, then the calculation of the force on each atom from energies and atom positions, followed by the integration of Newton's equation of motion to generate the new atom positions and velocities after a short duration (i.e. timestep). The steps after the initialization are repeated until the simulation time is reached. The simulation time should be longer than the type of molecular event studied yet the calculation is able to finish within a reasonable timeframe.

2.10.1 System Setup and Initialization

An MD simulation considers statistical mechanical ensembles, which are collections of all possible systems with different microscopic states but an identical macroscopic or thermodynamics state. The microcanical ensemble (NVE) describes an isolated system with N particles and a fixed energy E in a fixed volume V. In typical experiments, pressure or temperature control is usually required and a canonical ensemble or an isobaric-isothermal ensemble is to be used. The canonical ensemble (NVT) describes a system with N particles at a fixed temperature T in a fixed volume V, and allows temperature control. The isobaric-isothermal ensemble (NPT) describes a system with N particles at fixed pressure P and temperature T, and allows pressure and temperature control. Thermostats are needed for temperature control, such as

Berendsen[50], Andersen[51], Langevin[52] and Nosé-Hoover[53]. Similarly, barostats are required in pressure control, such as Berendsen[50], Andersen[51] and Parrinello-Rahman[54].

The selection of the timestep is crucial to the accuracy, stability, efficiency and search space of an MD simulation. The timestep must be smaller than the fastest bond vibrations in the system that usually involves the motion of hydrogens with typical values of 0.5–1 fs. By introducing bond restraints such as SHAKE[55], these fastest bond vibrations can be removed and the timestep can be extended to 2 fs.

The solvation effects can be modeled by an implicit solvent or an explicit solvent. The implicit model treats the solvent as a continuum, while the explicit model adds solvent molecules in the system and is much more computationally expensive. The explicit model allows the study of direct solvent-solute non-covalent interactions and viscosity effects of the solvent, so it is commonly used in MD simulations. Examples of water models are SPC[56], SPC/E[57] and TIP3P[58]. Moreover, adding counterions is commonly used to neutralize the system.

The number of molecules in the simulated system is usually limited and many molecules are near the surface that have different properties than the rest of the system, making such system inappropriate to study bulk properties. To avoid the surface effects, periodic boundary conditions (PBCs) need to be used. In PBCs, a small simulation box is replicated in all directions throughout space and when a molecule leaves this box, its image will enter through the opposite side of the same box, thus keeping the number of particles constant

throughout the simulation. To properly account for electrostatic interactions under PBCs, the Particle-Mesh-Ewald (PME) method[59] is commonly used with a cut-off value to split the short- and long-range parts.

After the system setup and the input of the starting structure, charges and other parameters needed in the force field calculation, the system is initialized with initial atom positions and velocities. The initial atom positions are from the starting structure of the system, and the initial velocities of atoms are related to the kinetic energy and are assigned according to the Maxwell-Boltzmann distribution.

2.10.2 Force Fields and Force Calculation

Force fields are usually used to calculate the potential energy of the system based on the atomic coordinates. A force field consists of a set of equations for potential energy calculation, and a set of parameters used in the equations mostly derived from empirical data. Some popular force fields for studying biomolecules are CHARMM[60], AMBER[61], GROMOS[62] and OPLS[63]. Force fields are much faster than quantum mechanical methods but usually cannot describe bond forming or breaking processes.

The potential energy function in a typical force field has bonded potentials to describe bond stretch, angle bending and torsion changes, and non-bonded potentials that include van der Waals and electrostatic interactions. The bond stretch and angle bending terms are usually modeled by quadratic functions based on Hooke's law. The torsion term can be described by cosine or sine functions. The van der Waals interaction is usually described by the

Lennard-Jones potential and the electrostatic interaction is based on Coulomb's law.

The force acting on each atom F_i is derived from Newton's equation of motion:

$$\boldsymbol{F_i = -\frac{\partial V(r_1, \cdots, r_N)}{\partial r_i}} \quad \textbf{(2.35)}$$

$$\boldsymbol{F_i = m_i a_i(t) = m_i \frac{\partial^2 r_i(t)}{\partial t^2}} \quad \textbf{(2.36)}$$

where $V(r_1, \cdots, r_N)$ is the potential energy of the system, r_i is the atomic coordinate, m_i is the mass of the atom *i*, and $a_i(t)$ is the acceleration at time t.

2.10.3 Integration Algorithms

To solve the equation of motion, several numerical integration methods have been developed based on the timestep δt. Some of the popular methods are the Verlet algorithm[64] and its related methods such as the Leapfrog algorithm[65].

The Verlet algorithm is the most basic and commonly used integration method. The derivation starts from the expression of positions $r(t \pm \delta t)$ from $r(t)$ by a Taylor series:

$$\boldsymbol{r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{3!}\frac{d^3 r(t)}{dt^3}\delta t^3 + O(\delta t^4)} \quad \textbf{(2.37)}$$

$$\boldsymbol{r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \frac{1}{3!}\frac{d^3 r(t)}{dt^3}\delta t^3 + O(\delta t^4)} \quad \textbf{(2.38)}$$

where $v(t)$ is the velocity which is the first derivative of r with respect to t, $a(t)$ is the acceleration which is the second derivative of r with respect to t and $O(\delta t^4)$ is the 4th order expansion term.

From the sum of (2.37) and (2.38), we have:

$$\boldsymbol{r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 + O(\delta t^4)} \quad \textbf{(2.39)}$$

The Verlet algorithm uses the current position $r(t)$, the acceleration $a(t)$ and the previous position $r(t-\delta t)$ to predict positions after the timestep $r(t+\delta t)$, and the associated error term is of order δt^4. The explicit velocities are not used in the calculation. This method is simple and computationally efficient but with moderate precision.

The velocities can be computed from the difference of (2.37) and (2.38):

$$\boldsymbol{v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} + O(\delta t^2)} \quad \textbf{(2.40)}$$

The Leapfrog algorithm is improved from the Verlet algorithm that uses velocities at the half timestep. The new expressions for the positions and velocities are as follows:

$$\boldsymbol{r(t+\delta t) = r(t) + v\left(t+\frac{1}{2}\delta t\right)\delta t} \quad \textbf{(2.41)}$$

$$\boldsymbol{v\left(t+\frac{1}{2}\delta t\right) = v\left(t-\frac{1}{2}\delta t\right) + a(t)\delta t} \quad \textbf{(2.42)}$$

In this way, the velocities and the positions are computed at different times. The velocities at time t can be approximated by the average of the velocities at $t - \delta t$ and $t + \delta t$.

2.11 References

1. (a) Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*. Wiley: 2013; (b) Levine, I. N., *Quantum Chemistry*. Pearson Education: 2013; (c) Szabo, A.; Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications: 2012.

2. Schrödinger, E., *Phys. Rev.* **1926,** *28* (6), 1049.

3. Born, M.; Oppenheimer, R., *Ann. Phys.* **1927,** *389* (20), 457.

4. Heisenberg, W., *Z. Physik* **1927,** *43* (3-4), 172.

5. Ochterski, J. W. Thermochemistry in Gaussian 2000. http://www.gaussian.com/g_whitepap/thermo.htm (accessed 09 December 2015).

6. (a) Fock, V. Z., *Z. Physik* **1933,** *81* (3-4), 195; (b) Fock, V. Z., *Z. Physik* **1930,** *61*, 126; (c) Hartree, D. R., *Math. Proc. Cambridge Philos. Soc.* **1928,** *24* (01), 111; (d) Hartree, D. R., *Math. Proc. Cambridge Philos. Soc.* **1928,** *24* (01), 89.

7. Slater, J. C., *Phys. Rev.* **1929,** *34* (10), 1293.

8. Slater, J. C., *Phys. Rev.* **1930,** *36* (1), 57.

9. Boys, S. F., *Proc. R. Soc. London A* **1950,** *200* (1063), 542.

10. Hehre, W. J.; Stewart, R. F.; Pople, J. A., *J. Chem. Phys.* **1969,** *51* (6), 2657.

11. Ditchfield, R.; Hehre, W. J.; Pople, J. A., *J. Chem. Phys.* **1971,** *54* (2), 724.

12. (a) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., *J. Chem. Phys.* **1980,** *72* (1), 650; (b) McLean, A. D.; Chandler, G. S., *J. Chem. Phys.* **1980,** *72* (10), 5639.

13. (a) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A., *J. Chem. Phys.* **1982,** *77* (7), 3654; (b) Frisch, M. J.; Pople, J. A.; Binkley, J. S., *J. Chem. Phys.* **1984,** *80* (7), 3265.

14. Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R., *J. Comput. Chem.* **1983,** *4* (3), 294.

15. (a) Hall, G. G., *Proc. R. Soc. London A* **1951,** *205* (1083), 541; (b) Roothaan, C. C. J., *Rev. Mod. Phys.* **1951,** *23* (2), 69.

16. (a) Schrödinger, E., *Ann. Phys.* **1926,** *385* (13), 437; (b) Peter W. Atkins, R. S. F., *Molecular Quantum Mechanics*. 5th ed.; Oxford University Press: 2010.

17. Møller, C.; Plesset, M. S., *Phys. Rev.* **1934,** *46* (7), 618.

18. Bartlett, R. J., *J. Phys. Chem.* **1989,** *93* (5), 1697.

19. Orio, M.; Pantazis, D. A.; Neese, F., *Photosynth. Res.* **2009,** *102* (2-3), 443.

20. Hohenberg, P.; Kohn, W., *Phys. Rev.* **1964,** *136* (3B), B864.

21. Kohn, W.; Sham, L. J., *Phys. Rev.* **1965,** *140* (4A), A1133.

22. Slater, J. C., *Phys. Rev.* **1951,** *81* (3), 385.

23. Vosko, S. H.; Wilk, L.; Nusair, M., *Can. J. Phys.* **1980,** *58* (8), 1200.

24. Becke, A. D., *Phys. Rev. A* **1988,** *38* (6), 3098.

25. Perdew, J. P., In *Electronic Structure of Solids '91*, Ziesche P., E. H., Ed. Akademie Verlag: Berlin, 1991; pp 11.

26. (a) Perdew, J. P.; Burke, K.; Ernzerhof, M., *Phys. Rev. Lett.* **1996,** *78*, 1396; (b) Perdew, J. P.; Burke, K.; Ernzerhof, M., *Phys. Rev. Lett.* **1996,** *77* (18), 3865.

27. Lee, C.; Yang, W.; Parr, R. G., *Phys. Rev. B* **1988,** *37* (2), 785.

28. Perdew, J. P., *Phys. Rev. B* **1986,** *33* (12), 8822.

29. Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E., *Phys. Rev. Lett.* **2003,** *91* (14), 146401.

30. Zhao, Y.; Truhlar, D. G., *J. Chem. Phys.* **2006,** *125* (19), 194101.

31. Becke, A. D., *J. Chem. Phys.* **1993,** *98* (7), 5648.

32. (a) Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1-3), 215; (b) Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008,** *41* (2), 157.

33. Zhao, Y.; Truhlar, D. G., *J. Chem. Theory Comput.* **2008,** *4* (11), 1849.

34. (a) Peverati, R.; Truhlar, D. G., *J. Phys. Chem. Lett.* **2011,** *2* (21), 2810; (b) Peverati, R.; Truhlar, D. G., *J. Phys. Chem. Lett.* **2012,** *3* (1), 117.

35. (a) Miertuš, S.; Scrocco, E.; Tomasi, J., *Chem. Phys.* **1981,** *55* (1), 117; (b) Miertuš, S.; Tomasi, J., *Chem. Phys.* **1982,** *65* (2), 239; (c) Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005,** *105* (8), 2999.

36. (a) Mennucci, B.; Cancès, E.; Tomasi, J., *J. Phys. Chem. B* **1997,** *101* (49), 10506; (b) Cancès, E.; Mennucci, B.; Tomasi, J., *J. Chem. Phys.*

1997, *107* (8), 3032; (c) Cancès, E.; Mennucci, B., *J. Math. Chem.* **1998,** *23* (3-4), 309.

37. Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J., *J. Phys. Chem.* **1996,** *100* (40), 16098.

38. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B* **2009,** *113* (18), 6378.

39. (a) Martí-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sánchez, R.; and, F. M.; Šali, A., *Annu. Rev. Biophys. Biomol. Struct.* **2000,** *29* (1), 291; (b) Sánchez, R.; Šali, A., In *Protein Structure Prediction*, Webster, D., Ed. Humana Press: 2000; Vol. 143, pp 97.

40. Chothia, C.; Lesk, A. M., *EMBO J.* **1986,** *5* (4), 823.

41. Müller, A.; MacCallum, R. M.; Sternberg, M. J. E., *J. Mol. Biol.* **1999,** *293* (5), 1257.

42. (a) Doolittle, R. F., In *Methods in Enzymology*, Academic Press: 1990; Vol. Volume 183, pp 99; (b) Sauder, J. M.; Arthur, J. W.; Dunbrack Jr, R. L., *Proteins* **2000,** *40* (1), 6; (c) Thompson, J. D.; Plewniak, F.; Poch, O., *Nucleic Acids Res.* **1999,** *27* (13), 2682.

43. (a) Bowie, J.; Luthy, R.; Eisenberg, D., *Sci.* **1991,** *253* (5016), 164; (b) Jones, D. T.; Taylort, W. R.; Thornton, J. M., *Nat.* **1992,** *358* (6381), 86.

44. Saqi, M. A.; Russell, R. B.; Sternberg, M. J., *Protein Eng.* **1998,** *11* (8), 627.

45. (a) Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., and Hill, R. C., *J. Mol. Biol.* **1969,** *42* (1), 65; (b) Greer, J.,

Proteins **1990,** *7* (4), 317; (c) Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M., *Nat.* **1987,** *326* (6111), 347.

46. (a) Jones, T. A.; Thirup, S., *EMBO J.* **1986,** *5* (4), 819; (b) Levitt, M., *J. Mol. Biol.* **1992,** *226* (2), 507; (c) Unger, R.; Harel, D.; Wherland, S.; Sussman, J. L., *Proteins* **1989,** *5* (4), 355.

47. (a) Havel, T. F.; Snow, M. E., *J. Mol. Biol.* **1991,** *217* (1), 1; (b) Šali, A.; Blundell, T. L., *J. Mol. Biol.* **1993,** *234* (3), 779.

48. (a) Rahman, A., *Phys. Rev.* **1964,** *136* (2A), A405; (b) Alder, B. J.; Wainwright, T. E., *J. Chem. Phys.* **1959,** *31* (2), 459.

49. (a) Kukol, A., *Molecular Modeling of Proteins*. Springer New York: 2015; (b) Frenkel, D.; Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press: San Diego, 2002.

50. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *J. Chem. Phys.* **1984,** *81* (8), 3684.

51. Andersen, H. C., *J. Chem. Phys.* **1980,** *72* (4), 2384.

52. Grest, G. S.; Kremer, K., *Phys. Rev. A* **1986,** *33* (5), 3628.

53. (a) Nosé, S., *Mol. Phys.* **1984,** *52* (2), 255; (b) Hoover, W. G., *Phys. Rev. A* **1985,** *31* (3), 1695; (c) Nosé, S., *J. Chem. Phys.* **1984,** *81* (1), 511.

54. (a) Parrinello, M.; Rahman, A., *J. Appl. Phys.* **1981,** *52* (12), 7182; (b) Parrinello, M.; Rahman, A., *Phys. Rev. Lett.* **1980,** *45* (14), 1196.

55. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., *J. Comput. Phys.* **1977,** *23* (3), 327.

56. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., In *Intermolecular Forces*, Pullman, B., Ed. Springer Netherlands: 1981; Vol. 14, pp 331.

57. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *J. Phys. Chem.* **1987,** *91* (24), 6269.

58. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** *79* (2), 926.

59. (a) Darden, T.; York, D.; Pedersen, L., *J. Chem. Phys.* **1993,** *98* (12), 10089; (b) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., *J. Chem. Phys.* **1995,** *103* (19), 8577.

60. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., *J. Phys. Chem. B* **1998,** *102* (18), 3586.

61. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. Soc.* **1995,** *117* (19), 5179.

62. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F., *J. Comput. Chem.* **2004,** *25* (13), 1656.

63. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., *J. Am. Chem. Soc.* **1996,** *118* (45), 11225.

64. Verlet, L., *Phys. Rev.* **1967,** *159* (1), 98.

65. Hockney, R. W. E., J. W., *Computer simulation using particles*. New York: McGraw-Hill: 1981.

Chapter 3 Homology Modeling of hDNMT1 with Molecular Dynamics Simulations

3.1 Introduction

As discussed in section 1.2.2, the crystal structure of hDNMT1 bound to the hemi-methylated DNA at the active site is not yet available, so the crystal structure of the productive mDNMT1-DNA complex (4DA4, residues 731-1602, resolution 2.6 Å)[1] was used to generate a novel 3D model of the hDNMT1-DNA complex only involving the C-terminal domain via homology modeling. The rationale behind is that mDNMT1 shares nearly 90% sequence identity in the C-terminal domain as hDNMT1, and these two DNMTs are almost completely identical in the active site[2]. Moreover, the structure of this mDNMT1-DNA complex retains all the key binding features between the protein and the substrate in the active site as that predicted by the previously established models of the hDNMT1-substrate complex[3] and observed for the M.HhaI-substrate complexes[4].

The homology model was further refined by MD simulations because such simulations allow the whole system to relax on the PES and are able to study the stability and dynamic properties of the hDNMT1-DNA complex.

3.2 Computational Methods

3.2.1 Homology Modeling

The amino acid sequence of the C-terminal domain of hDNMT1 was retrieved from the NCBI protein database (Accession number NP_001370.1). The crystal structure of mDNMT1 (4DA4)[1] with selected residues from 1137 to 1600 from the monomer A was chosen as the template for generating the 3D structure of residues from 1134 to 1598 of hDNMT1. It is worth noting that the crystal structure from 4DA4 is a dimer, and the protein monomer A has fewer missing residues than the protein monomer B, so the monomer A was selected for homology modeling. Sequence alignment was done using the Clustal Omega server[5]. Five models were built using Modeller 9.11[6]. Coordinates of the dsDNA, SAH and Zn(II) bound to the monomer A were copied to the model. The Zn(II) ion is about 30 Å away from the substrate cytosine in the active site. The best model, model 5, was chosen based on the Modeller objective function (MOLPDF)[7] energy score and the Discrete Optimized Protein Energy (DOPE)[8] score. This model was validated by PROCHECK[9] and Molecular Operating Environment (MOE) release 2014.09[10]. The RMSD for the protein backbone was calculated by selecting the following three options to improve the superposition in MOE2014.09: "Optimize Gap Penalties for Superposition", "Accent Secondary Structure Matches" and "Refine with Gaussian Distance Weights".

The MODELLER[6] program was chosen for the model construction because it is capable to build the protein model together with other non-protein

ligands while some programs cannot deal with some of the non-protein ligands such as SWISS-MODEL workspace 7[11], and it is one of the most widely used and easily accessible software for homology modeling. It has been shown that MODELLER is able to produce similar results as many other commonly used homology modeling programs such as Prime and SWISS-MODEL when sequence identities between the target and the templates are greater than 30%[12]. Moreover, the MODELLER program has been used to generate the first high quality homology model of hDNMT1 by Siedlecki *et al.*[3a] (Table 1.1).

3.2.2 Molecular Dynamics

The best model was used to build the initial structure for the MD simulations. The methylated 5-fluoro-2'-deoxycytidine and the 5-methyl-2'-deoxycytidine were remodeled to 2'-deoxycytidine. Missing residues were added using the Structure Preparation application in MOE2014.09[10]. Hydrogen atoms were added to this model by the LEaP module of the AMBER 12 (Assisted Model Building with Energy Refinement, 12)[13] simulation package. The hydrogen positions were adjusted based on the hydrogen bonding networks predicted from protonation and tautomeric states of Cys and histidine (His) residues at the pH of 7 and the temperature of 300 K using the Protonate3D[14] in MOE2014.09[10]. The force field used is Amber12:EHT[10]. This force field is an all-atom force field that combines parameters from the AMBER 12[13] force field for proteins and DNA, and parameters from the 2D Extended Hückel Theory (EHT)[15] for small molecules.

The catalytic Cys1226 is deprotonated while Glu1266 is protonated according to the DNA reaction mechanism. The three Cys residues—Cys1476, Cys1478 and Cys1485—bound to Zn(II) are deprotonated, and the His residue—His1502—bound to Zn(II) is protonated at the δ-N as the HID tautomer. The protonation states of other His residues are summarized in Table S 8.4.

For the energy minimizations and MD simulations carried out by the AMBER 12[13], the Amber99SB[16] force field with modification for DNA by parmbsc0[17] (ff99bsc0) was used for proteins and DNA. The metal center parameter builder (MCPB)[18] was used to generate the charges and parameters for the Zn(II)-binding region (Zn(II), Cys1476, Cys1478, Cys1485 and His1502). MCPB describes the Zn(II)-binding regions by a bonded model, so that it treats the interactions between metal ions and its ligating residues via bond, angle, torsional, electrostatics and van der Waals (vdW) terms. Partial atomic charges for SAH were calculated at the HF/6-31G* level of theory with optimization using the Gaussian 09[19] suite of programs. The charges were fitted to each atom of SAH with the RESP algorithm[20]. Bond length, bond angle, torsional angle and vdW parameters of SAH were generated by the Antechamber module using the General Amber Force Field (GAFF)[21]. The hydrogen positions in the model were minimized with 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization. A distance dependent dielectric function was used with the dielectric constant of 4.0. Furthermore, the model was solvated in a periodic octahedral box of TIP3P water[22] with 10 Å buffer around the complex. No ions were added because the system is already neutral after the treatment by

the Protonate3D[14]. The total number of atoms in the system was 62925, of which 54810 were water molecules.

The subsequent energy minimizations and MD simulations were performed with the PME method[23] with 12 Å cut-off for the non-bonded interactions and long range electrostatic interactions. In the MD simulations, all bonds to hydrogen atoms were constrained with the SHAKE algorithm[24], so a timestep of 2 fs was used. The temperature was controlled by the Langevin dynamics[25] approach with a collision frequency of 2 ps^{-1}, and the pressure was maintained using isotropic position scaling. The solvent was minimized with 2500 steps of steepest descent minimization followed by 2500 cycles of conjugate gradient minimization, while the solute was restrained with a force constant of 50 kcal mol^{-1} $Å^{-2}$. The 20 ps NVT MD simulation was first performed to equilibrate the system with the restraint of the solute at 10 kcal mol^{-1} $Å^{-2}$ at a constant temperature of 300 K. Then the 30 ps NPT MD simulation was performed to further equilibrate the system with the restraint of the solute at 10 kcal mol^{-1} $Å^{-2}$ at a constant pressure of 1 atm and a constant temperature of 300 K. The force constant on the solute was reduced in two steps: first to 1 kcal mol^{-1} $Å^{-2}$ with the 40 ps NPT MD simulation, and then to 0.1 kcal mol^{-1} $Å^{-2}$ with the 50 ps NPT MD simulation. Finally, a 9 ns NPT production run was carried out at a constant pressure of 1 atm and a constant temperature of 300 K, and was recorded every 1 ps. Analysis was performed using standard AMBER 12[13] and VMD[26] tools, as well as PROCHECK[9] and MOE2014.09[10].

3.3 Results and Discussions

3.3.1 Selection and Quality of the Unrefined Model

A homology model of the hDNMT1-DNA complex at the C-terminal domain (residues 1134–1598) was generated using the active mDNMT1-DNA complex at the C-terminal domain (residues 1137–1600) as described in section 3.2.1. The result from the sequence alignment with 1-letter abbreviations of amino acids is shown in Figure S 8.1, where there is one additional Ala (Ala1481) between Glu1480 and Gly1482 in hDNMT1 as compared with mDNMT1.

Five similar homology models were generated automatically by 'satisfaction of spatial constraints' that is described by MOLPDFs[7]. The MOLPDF score is the sum of the restraint violations; hence, the lower the MOLPDF score, the less the violation of restraints is, and the better the model is. Models 4 and 5 have the lower MOLPDF scores (3825.18750 and 3826.16528) as compared with other three models (3838.14111 to 3904.82886).

These two models were further evaluated with DOPE[8] scores. DOPE is a pairwise atomic-distance dependent statistical potential optimized for the assessment of protein models that is able to separate native-like structures from the decoys. The lower the DOPE score, the more native-like the model is, and the better the model is. The DOPE score for the model 5 (-48567.378906) is much lower than that for the model 4 (-48201.914063), so the model 5 was selected as the best model for later studies. By examining the DOPE per-

residue score (results not shown), it was observed that the per-residue score for the model 4 is much worse than that for the model 5 between residues 1454 and 1474, which is close to the missing Ala1481 region.

This initial homology model, named as 4DAX-i based on the PDB ID of the template mDNMT1, was validated using PROCHECK[9]. The Ramachandran plot of Phi and Psi angles of the protein in this model indicated that 100.0% of residues (465) are present in the allowed regions (with 90.9% in the most favored regions) just like mDNMT1 (4DA4)[1] (100.0% in the allowed regions with 90.8% in the most favored regions).

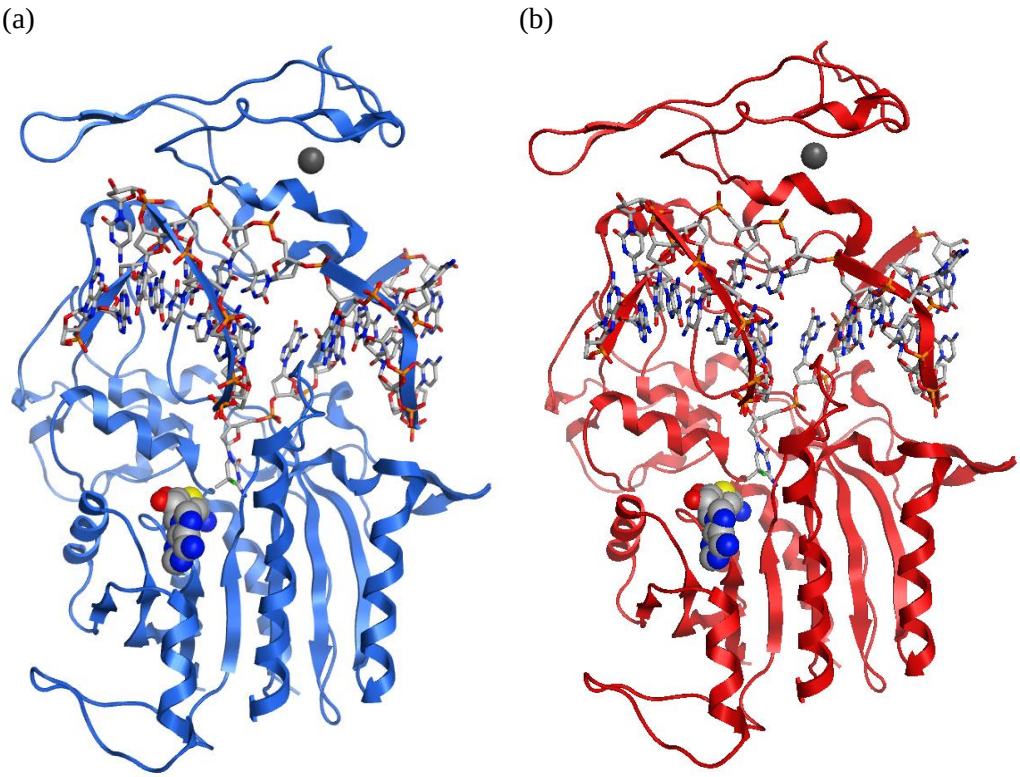


Figure 3.1 The 3D structures of (a) mDNMT1 (4DA4)[1] (light blue) and (b) homology model of hDNMT1 (4DAX-i) (red). The proteins are shown by the cartoon model, the Zn(II) ion and SAH are shown by the CPK model and colored by atom. The nucleotides are shown by the stick model and colored by atom.

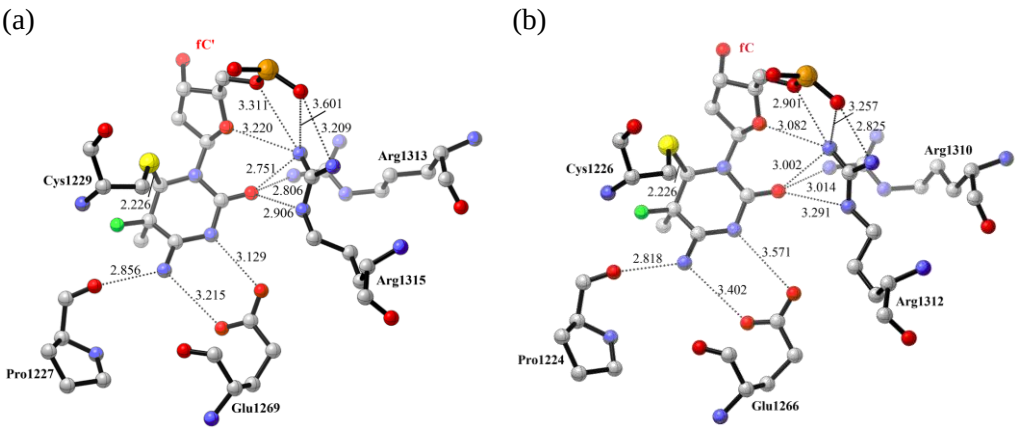


Figure 3.2 Interaction of the flipped-out methylated cytosine nucleotide with the active site residues of (a) mDNMT1 (4DA4)[1] and (b) homology model of hDNMT1 (4DAX-i). The heavy atom distances are indicated by the dotted lines. Distances are in Å.

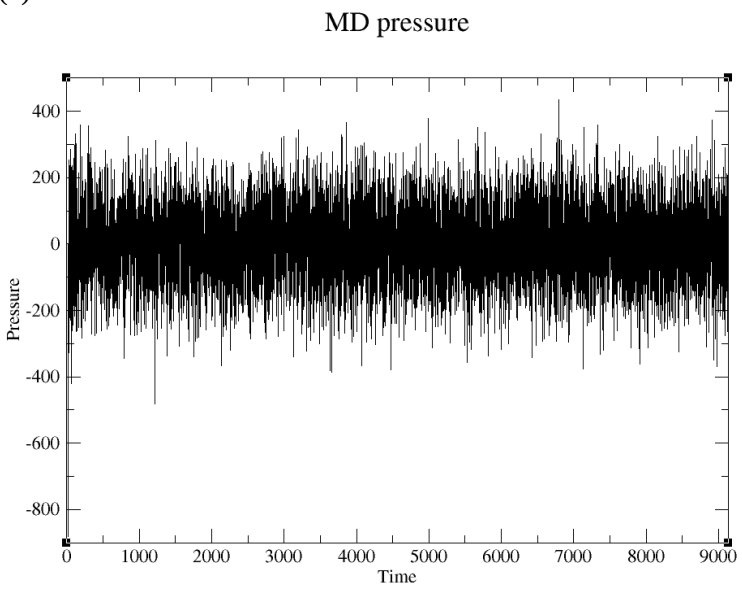
The overall structure of this hDNMT1-DNA complex is very similar to that of the mDNMT1-DNA complex with a protein backbone RMSD of 0.196 Å computed by MOE2014.09[10] (Figure 3.1). Its active site is nearly identical to that of the mDNMT1-DNA complex (Figure 3.2).

3.3.2 Evaluation of the Overall MD Performance and Analysis of the MD Trajectory

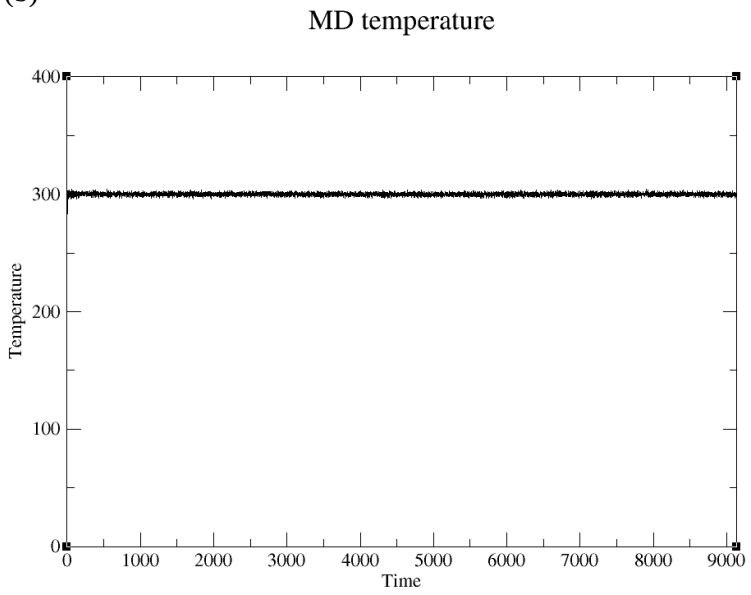
After the MD simulations, several physical properties from the simulations were examined to evaluate the MD performance. First, the pressure or temperature was plotted against time through the equilibration and production phases from 0–9140 ps (Figure 3.3(a)–(b)). The plots show that the pressure has relatively stable fluctuations around 1 atm (~1 bar) and the temperature has relatively stable fluctuations around 300 K. These results suggested the proper control of the desired pressure and temperature during the MD simulations as the actual pressure and temperature generally agree

with the setup values. Second, the density of the system was plotted against time from 20–9140 ps (Figure 3.3(c)). The plot shows that the density had already been equilibrated to ~1.03 g cm^{-3} during the NPT equilibration phase (20–140 ps) and this density value was maintained throughout the production phase (140–9140 ps). Third, the total energy, total potential energy and total kinetic energy were plotted against time from 0–9140 ps (Figure 3.3(d)). The plot shows that the total system energy is the sum of the total kinetic and potential energies. Moreover, the total kinetic energy is almost stable throughout the constant temperature MD simulations, whereas the total energy mirrors the total kinetic energy and both energies were reduced quickly from 0–90 ps during the equilibration phase and was almost stabilized after 90 ps throughout the MD simulations.

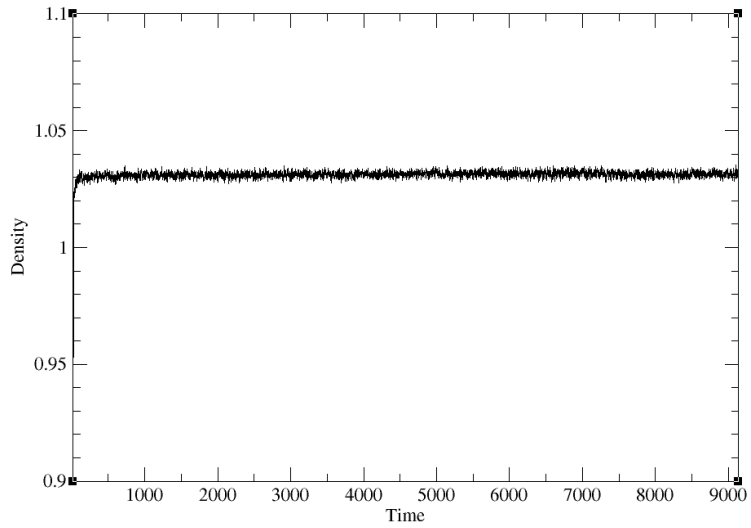
(a)



(b)


MD temperature
Temperature
Time
0
100
200
300
400
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(c)


MD density
Density
Time
0.9
0.95
1
1.05
1.1
1000
2000
3000
4000
5000
6000
7000
8000
9000

(d)

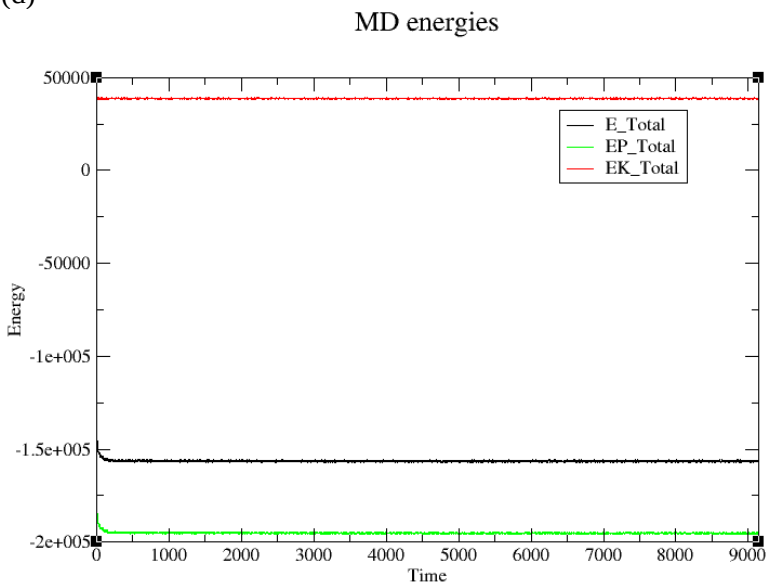
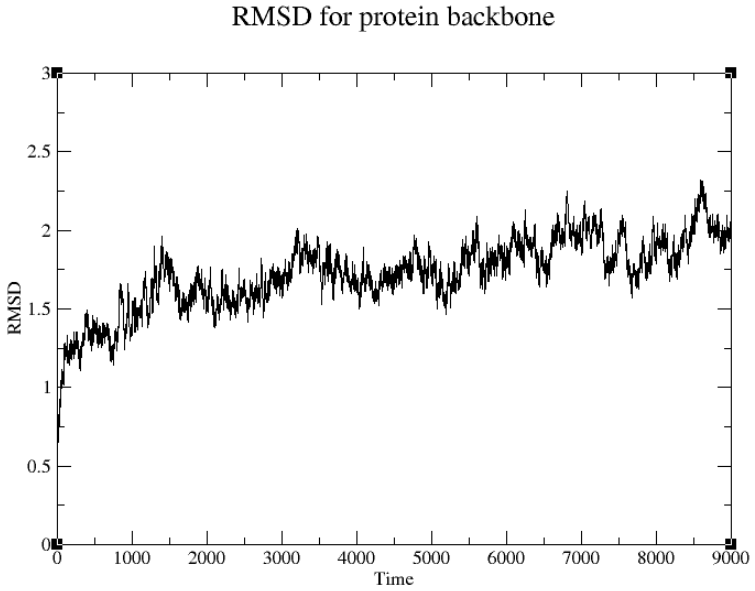
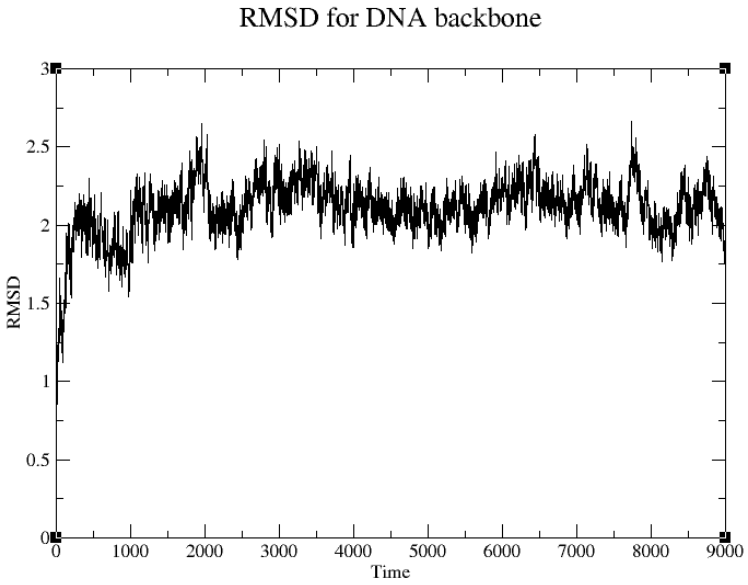


Figure 3.3 Plots of (a) pressure (in bar); (b) temperature (in K); and (c) density (in g cm^{-3}); (d) energies (in kcal mol^{-1}) against time (in ps).

(a)



(b)



(c)

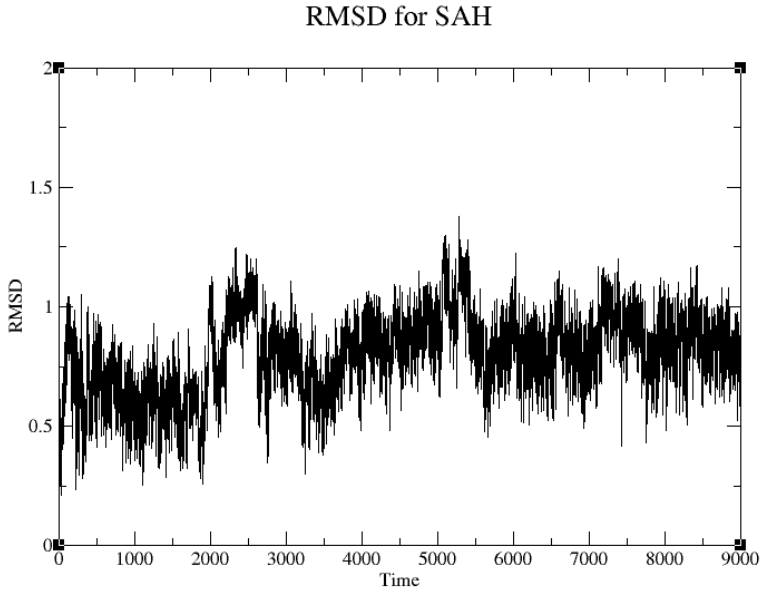


Figure 3.4 Structural drift during the production phase for (a) protein backbone; (b) DNA backbone; (c) SAH with reference to the model structure right after the equilibration. Time is in ps, and RMSD is in Å.

Structural drifts of the protein backbone, DNA backbone and SAH were measured from plots of the corresponding RMSD against time during the production phase with reference to the model generated right from the

equilibration phase (Figure 3.4(A)–(C)). The RMSD for the protein backbone is relatively stable around 1.75 Å between ~3 ns and ~5.5 ns and has a slight upward trend after ~5.5 ns. The RMSD for the DNA backbone is relatively stable around 2.10 Å after ~3.5 ns, and the RMSD for SAH is relatively stable around 0.75 Å after ~5.5 ns. Moreover, from manual inspection, no significant conformational changes of the protein backbone or the DNA backbone were observed during the 9 ns production phase, suggesting the relative stability of this homology model of the hDNMT1-DNA complex throughout the simulation.

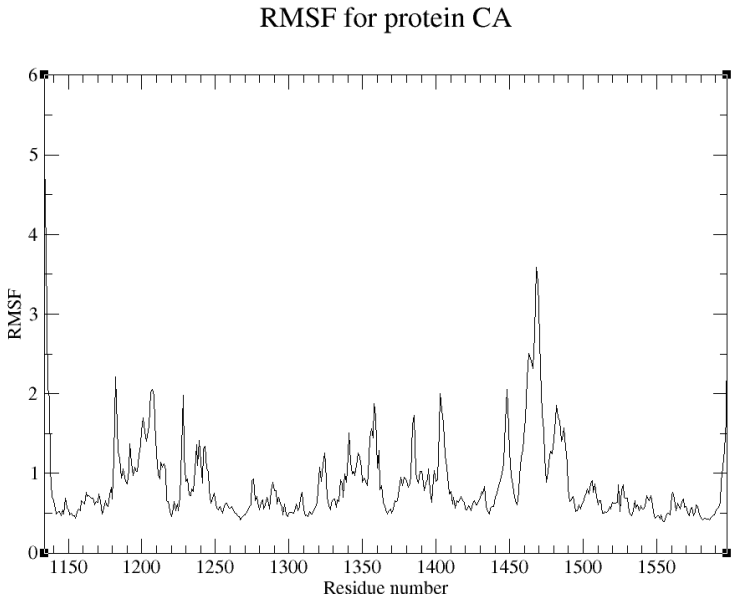


Figure 3.5 RMSF of hDNMT1 based on protein Cα. RMSF is in Å.

To understand the contribution of each protein residue to the conformational sampling of the hDNMT1 protein during the MD production phase, the per-residue fluctuations described by the root-mean-square fluctuations (RMSFs) of the protein Cα atoms were calculated and plotted in

Figure 3.5. Overall, the RMSF values for most of the residues are small (< 3 Å), suggesting the relative stability of the protein secondary structure over the simulations. The only two short regions that have the slightly higher RMSFs (~3 Å) are formed by residues 1134–1135 and 1468–1470. The first region is the free end of the protein that is solvent exposed, so it is highly flexible. The second region features the turn of a loop region that is also solvent exposed, so it is highly flexible, too. The RMSF values at the catalytic loop region in this study are totally different when compared with that of the homology model of hDNMT1 without the dsDNA generated from the same crystal structure of mDNMT1[27] because the dsDNA binds to the catalytic loop region to prevent its movement over the simulations.

3.3.3 Examination of the Active Site Interactions and the Two Critical Distances

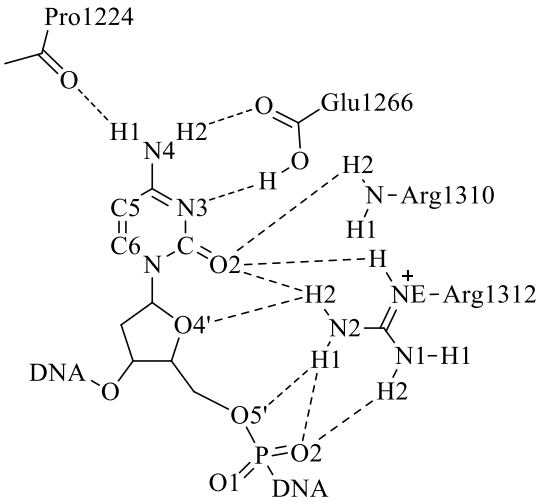


Figure 3.6 Possible hydrogen bonds between the cytosine nucleotide and residues in the active site of hDNMT1.

To examine the important hydrogen bonds (H-bonds) formed between the substrate cytosine nucleotide and the protein residues in the active site including Pro1224, Glu1266, Arg1312 and Arg1310, and their stabilities during the MD production phase, the distances of nine possible H-bonds (Figure 3.6) initially identified from the crystal structure of the mDNMT1-DNA complex (4DA4)[1] were measured and plotted against time (Figure S 8.2(a)–(j)). The cytosine base ring forms six possible H-bonds with Pro1224, Glu1266, Arg1312 and Arg1310, including the N4-H1···O interaction with the Pro1224 backbone C=O, the O-H···N3 and N4-H2···O interactions with the Glu1266 COOH, two N-H···O2 interactions with the Arg1312 NH's, and one N-H···O2 interaction with the Arg1310 NH. The sugar and phosphate also form four possible N-H···O interactions with the Arg1312 NH's. The five H-bonds formed between the cytosine base and Pro1224 (N4-H1···O: ~2.05 Å), Glu1266 (O-H···N3: ~1.95 Å, N4-H2···O: ~1.90 Å) or Arg1312 (NE-H···O2: ~2.20 Å, N2-H2···O2: ~1.90 Å) are highly stable during the MD simulations (Figure S 8.2(a)–(e)). The H-bond between the N1-H2 of Arg1312 and the O2 of the phosphate is also highly stable (~1.75–1.80 Å) (Figure S 8.2(i)). The H-bond between the N2-H2 of Arg1312 and the O4' of the sugar is mostly stable (~2.25 Å) with a breakage during ~0.7–1.2 ns only (Figure S 8.2(f)). Interestingly, the N2-H1 of Arg1312 may form H-bonds with either the O5' of the sugar (~1.90 Å for about 5.5 ns) or the O2 of the phosphate (min. average distance over every 50 ps ~1.90–1.95 Å for about 3.5 ns) but not both, so the plots of these two distances are mirror images of each other (Figure S 8.2(g) and (h)). The N-H···O2 interaction between the NH of Arg1310 and the

cytosine base is not stable and appears < 1 ns (min. average distance over every 50 ps ~1.95 Å) during the simulations, but the average distance is always < 4 Å such that electrostatic interactions can occur between them (Figure S 8.2(j)). The D-H-A (D/A=N or O) angles were also checked to confirm the H-bonds (results are not shown). Overall, the majority of the important H-bonds observed in the crystal structure of the mDNMT1-DNA complex in the active site (4DA4)[1] are present in the homology model of the hDNMT1-DNA complex and remained stable during the MD simulations with small fluctuations.

Two distances—C6(cytosine)-S(Cys1226) and C5(cytosine)-S(SAH) — that are important in the DNA methylation mechanism were measured and plotted against time (Figure S 8.2(k)–(l)). The C6(cytosine)-S(Cys1226) distance increased from 3.5 Å to 5–6 Å after 0–3 ns because the side chain of Cys1226 turned away from the C6 of cytosine and remained relatively stable afterwards. The C5(cytosine)-S(SAH) distance remained relatively stable around 5 Å with random decreases to ~4 Å. From manual inspection, it was observed that the homocysteine arm (-S-CH_2-CH_2-CH(-N)-C) of SAH tends to move a lot during the simulations. The most preferred conformation of SAH in this study (Figure 3.7(b)) was also observed in the other reported models of the hDNMT1-substrate-SAH complex[3b, 3c], but it is slightly different from that present in the crystal structures of M.HhaI, mDNMT1 and non-productive hDNMT1 in complex with SAH listed in Table S 8.1 (Figure 3.7(a)). Because the previously reported 3D structures of the hDNMT1-substrate-SAH complex were generated using different computational methods and algorithms, the

change in the SAH conformation may not a result of technical errors in the simulations. The homocysteine arm of SAH has at least four rotatable bonds and is also exposed to solvent, thus this arm is expected to be highly dynamic during the MD simulations.

(a) SAH (mDNMT1) (b) SAH (hDNMT1)

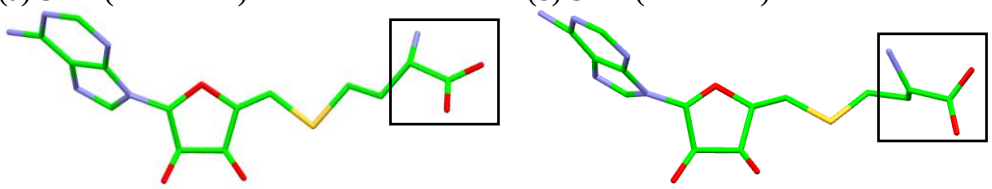


Figure 3.7 The structures of SAH in (a) mDNMT1 (4DA4)[1] and (b) homology model of hDNMT1 (from the MD production phase). SAH is shown by the stick model and colored by atom. The major differences are boxed.

3.3.4 Evaluation of the Final Homology Model

The structure with the lowest total potential energy at 3.987 ns from the MD production phase was chosen as the final homology model and named as 4DAX-f. The ligand-free hDNMT1 in this model was validated using PROCHECK[9]. The Ramachandran plot of Phi and Psi angles indicated that 100.0% of residues (465) are present in the allowed regions (with 83.3% in the most favored regions) similar to mDNMT1 (4DA4)[1]. No atom clashes were detected (interatomic distance < 2.2 Å) just like mDNMT1 and the initial model of hDNMT1 (4DAX-i). No outliers of bond lengths or bond angles were identified with reference to the standard values of these terms from experimental studies[28], although few outliers were detected in mDNMT1 or the initial model of hDNMT1. Therefore, this final homology model of hDNMT1 (4DAX-f) meets the general requirements of a quality experimental structure of a protein.

(a) (b)

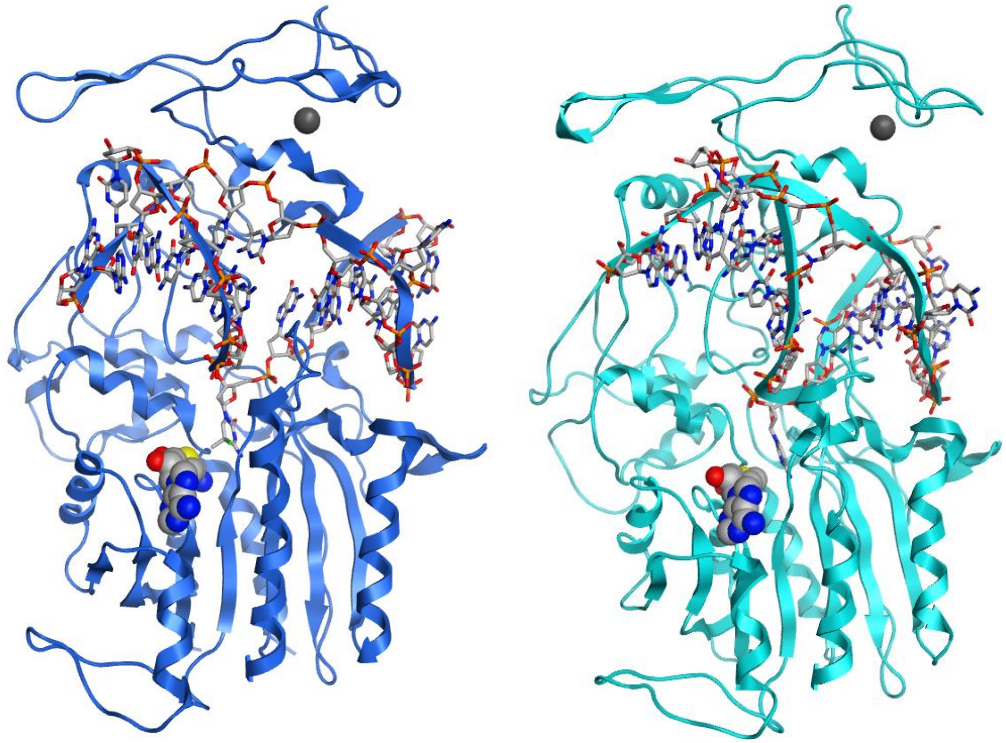


Figure 3.8 The 3D structures of (a) mDNMT1 (4DA4)[1] (light blue) and (b) final homology model of hDNMT1 (4DAX-f) (cyan). The proteins are shown by the cartoon model, the Zn(II) ion and SAH are shown by the CPK model and colored by atom. The nucleotides are shown by the stick model and colored by atom.

(a) (b)

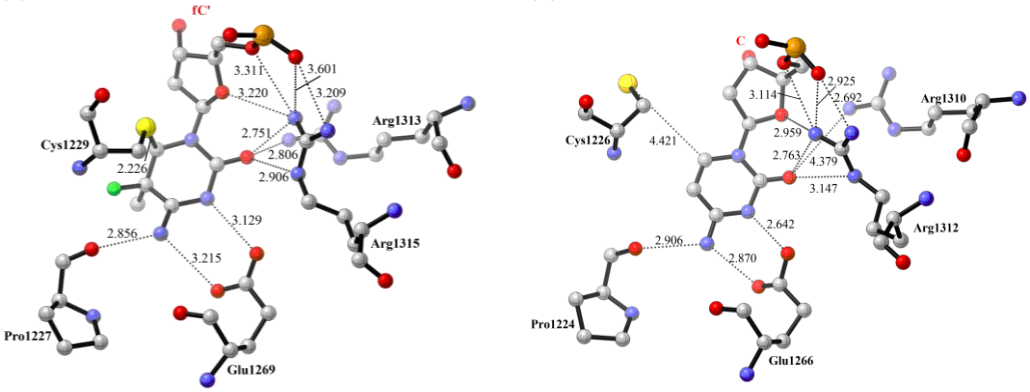


Figure 3.9 Interaction of the flipped-out cytosine nucleotide with the active site residues of (a) mDNMT1 (4DA4)[1] and (b) final homology model of hDNMT1 (4DAX-f). Hydrogens were omitted for comparison purpose. The heavy atom distances are indicated by the dotted lines. Distances are in Å.

The overall structure of the final hDNMT1-DNA complex is very similar to that of the mDNMT1-DNA complex with a protein backbone RMSD of 1.555 Å computed by MOE2014.09[10] (Figure 3.8). Its active site is similar to that of the mDNMT1-DNA with slight conformational changes of the side chains of Cys1226, Arg1312 and Arg1310 (Figure 3.9). The cytosine ring is almost sandwiched between the thiolate of Cys1226 and the cofactor SAH, similar to that in the mDNMT1-DNA-SAH complex. The C6(cytosine)-S(Cys1226) distance is 4.421 Å, which is much longer than that of the C6-S bond in the mDNMT1-DNA complex (2.226 Å) mainly due to the rotation of the Cys1226 side chain since the C6-S covalent bond is absent in the homology model and this homology model includes SAH instead of the functional cofactor SAM that may help bring the thiolate of Cys1226 closer to the C6 of cytosine via electronic effects. In addition, the C5(cytosine)-S(SAH) distance is 4.993 Å, which is longer than that in the mDNMT1-DNA complex (4.548 Å). In the two models of the hDNMT1-DNA-SAH complex reported by Yoo *et al.*[3b, 3c], the C6(cytosine)-S(Cys1226) and the C5(cytosine)-S(SAH) distances were measured to be 3.3/3.4 Å and 3.0/3.5 Å respectively, both of which are shorter than that observed in the final model in our study, possibly due to the use of different computational algorithms. It is worth noting that no strong attractive interactions can be identified between S(Cys1226) and C6(cytosine) or between C5(cytosine) and S(SAH), and these two distances are determined by the dynamics of the protein, DNA and homocysteine arm of SAH.

3.4 Conclusion

A novel 3D model of the hDNMT1-DNA complex at the C-terminal domain was generated and validated in this study. The model generation is based on the crystal structure of a productive covalent complex of mDNMT1-DNA (4DA4)[1] using Modeller 9.11[6]. The initial homology model (4DAX-i) was further refined by MD simulations (a 140 ps equilibration phase followed by a 9 ns production phase) using AMBER 12[13] to generate the final homology model as the lowest energy structure at 3.987 ns from the production phase.

From the analysis of some physical parameters of the MD simulations and some structural properties measured over time in the production phase, it was observed that the protein backbone remained relatively stable throughout the simulations, and the important H-bonds in the homology model of the hDNMT1-DNA complex at the active site are similar to that observed in the crystal structure of the mDNMT1-DNA complex (4DA4)[1] and remained stable during the MD simulations with small fluctuations.

The evaluation of the final homology model (4DAX-f) demonstrated the correctness of the protein structure because it meets the general requirements of a quality experimentally-determined protein structure. Furthermore, the overall structure of the final hDNMT1-DNA-SAH complex (4DAX-f) resembles that of the mDNMT1-DNA-SAH complex with minor differences in the key active site interactions.

3.5 References

1. Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J., *Sci.* **2012,** *335* (6069), 709.

2. Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J., *Sci.* **2011,** *331* (6020), 1036.

3. (a) Siedlecki, P.; Boy, R. G.; Comagic, S.; Schirrmacher, R.; Wiessler, M.; Zielenkiewicz, P.; Suhai, S.; Lyko, F., *Biochem. Biophys. Res. Commun.* **2003,** *306* (2), 558; (b) Yoo, J.; Medina-Franco, J. L., *J. Comput. Aided Mol. Des.* **2011,** *25* (6), 555; (c) Yoo, J.; Kim, J. H.; Robertson, K. D.; Medina-Franco, J. L., *Adv. Protein Chem. Struct. Biol.* **2012,** *87*, 10.1016/B978.

4. Sankpal, U. T.; Rao, D. N., *Crit. Rev. Biochem. Mol. Biol.* **2002,** *37* (3), 167.

5. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G., *Mol. Syst. Biol.* **2011,** *7* (1).

6. Šali, A.; Blundell, T. L., *J. Mol. Biol.* **1993,** *234* (3), 779.

7. Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A., *Curr. Protoc. Bioinformatics* **2006,** *0 5*, Unit.

8. Shen, M. Y.; Sali, A., *Protein Sci.* **2006,** *15* (11), 2507.

9. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., *J. Appl. Crystallogr.* **1993,** *26* (2), 283.

10. *Molecular Operating Environment (MOE), 2014.09*, Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal QC, Canada, H3A 2R7, 2014.

11. (a) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T., *Nat. Protoc.* **2008,** *4* (1), 1; (b) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., *Bioinformatics* **2006,** *22* (2), 195.

12. Dolan, M. A.; Noah, J. W.; Hurt, D., *Methods Mol. Biol.* **2012,** *857*, 399.

13. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 12. University of California, San Francisco: 2012.

14. Labute, P., *Proteins* **2009,** *75* (1), 187.

15. Gerber, P. R.; Muller, K., *J. Comput. Aided Mol. Des.* **1995,** *9* (3), 251.

16. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., *Proteins* **2006,** *65* (3), 712.

17. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M., *Biophys. J.* **2007,** *92* (11), 3817.

18. Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M., *J. Chem. Theory Comput.* **2010,** *6* (9), 2935.

19. 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. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

20. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., *J. Phys. Chem.* **1993,** *97* (40), 10269.

21. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., *J. Comput. Chem.* **2004,** *25* (9), 1157.

22. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** *79* (2), 926.

23. (a) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., *J. Chem. Phys.* **1995,** *103* (19), 8577; (b) Darden, T.; York, D.; Pedersen, L., *J. Chem. Phys.* **1993,** *98* (12), 10089.

24. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., *J. Comput. Phys.* **1977,** *23* (3), 327.

25. Grest, G. S.; Kremer, K., *Phys. Rev. A* **1986,** *33* (5), 3628.

26. Humphrey, W.; Dalke, A.; Schulten, K., *J. Mol. Graph.* **1996,** *14* (1), 33.

27. Joshi, M.; Rajpathak, S. N.; Narwade, S. C.; Deobagkar, D., *Chem. Biol. Drug Des.* **2016**, n/a.

28. Engh, R. A.; Huber, R., *Acta Crystallogr. Sect. A* **1991,** *47* (4), 392.

Chapter 4 Mechanism Study on DNA Methylation Based on Simple Model Systems

4.1 Introduction

As already stated in section 1.2.3, the first two steps in the DNA methylation reaction—the Cys nucleophilic addition to the C6 of cytosine and the methyl transfer from SAM to the C5 of cytosine—were studied using the hybrid functionals B3LYP[1] or M06-2X[2] based on small model systems to probe the key transition sates (TS) and intermediates with or without the conserved residues—Cys, Glu or Arg—in the active site surrounding them. This study aims to resolve some controversies in the DNA methylation mechanism (Figure 1.4) over the existence and stability of the covalent DNMT-cytosine intermediate via the catalytic Cys-thiolate, and the ability of this intermediate to undergo methylation in the protein environment. In addition, it was also hoped to improve the understanding of the roles of these conserved residues from the calculations.

A truncated model of the active site was proposed based on Studies 1–2[3] (refer to section 1.2.3 for the naming of Studies 1–6) and is depicted in Figure 4.1. This model captures the key features of the protein active site: the cofactor SAM (as trimethylsulfonium ion), the target 2'-deoxycytidine (as 1-methylcytosine), and the side chains of the catalytic Cys81-thiolate (as methylthiolate), Glu119 (as acetic acid) and Arg165 (as 1-methylguanidinium) (numbering based on M.HhaI). The binding role of Arg1312 (Arg165 in

M.HhaI) is more important than that of Arg 1310 (Arg163 in M.HhaI) in the active site of hDNMT1 as discussed in section 3.3 and it has been shown that the inclusion of Arg165 is enough to stabilize the transient species in the active site (Study 2)[3b]. This strategy has also been practiced in the study of the hDNMT1 transition states (Study 5)[4].

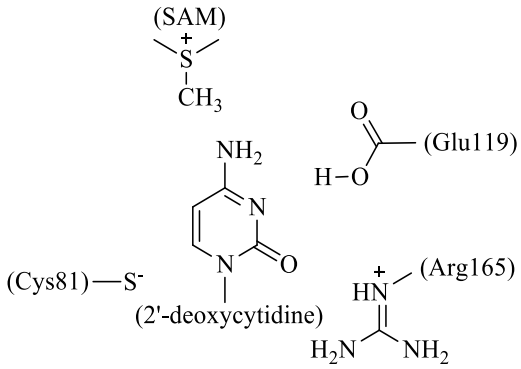


Figure 4.1 A small-truncated model to represent the active site of M.HhaI.

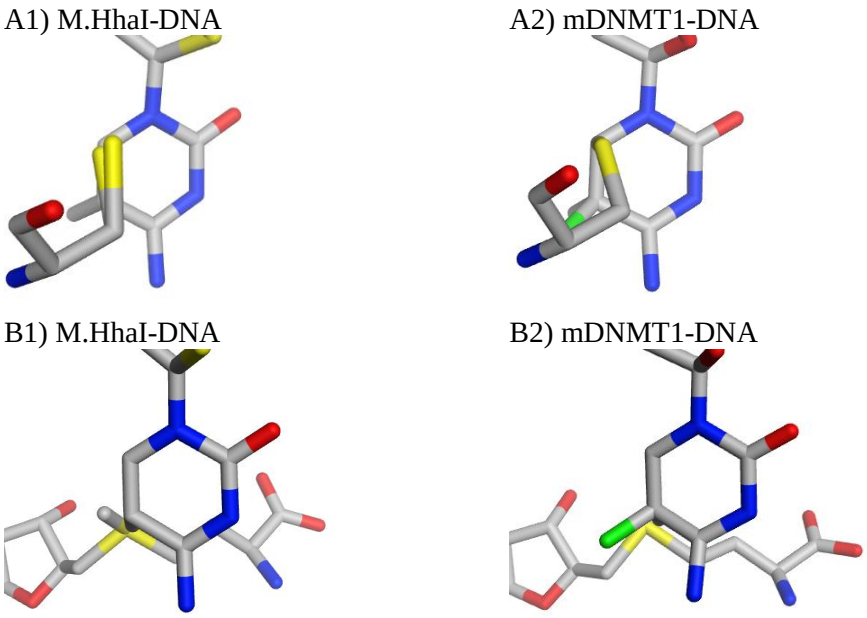


Figure 4.2 Relative orientation of A1)–2) Cys-thiolate and B1)–2) SAM-methyl to cytosine in the crystal structures of the M.HhaI-DNA complex (6MHT)[5] and the mDNMT1-DNA complex (4DA4)[6].

The relative orientations of the Cys-thiolate and SAM-methyl to cytosine in the active M.HhaI-DNA complex (6MHT)[5] and mDNMT1-DNA complex (4DA4)[6] were considered in this study and are shown in Figure 4.2.

4.2 Computational Methods

As mentioned in sections 1.2.3 and 2.7.3, the B3LYP[1] functional was used in most of the studies on the DNA methylation mechanism prior to 2016[3, 7], and it is the most widely used functional with high efficiency in calculation. However, the B3LYP[1] functional is not capable to describe the London dispersion interaction, whereas the M06-2X[2] functional is able to capture this interaction in the systems studied. To investigate the reliability of the B3LYP/6-31G* and M06-2X/6-31G* calculations in predicting the C-S bond formation, the simplest intermediate 1 was calculated by both methods and the results were compared with that from the higher *ab initio* levels of theory—MP2/aug-cc-pVTZ with density fitting (DF)[8] and the more demanding CCSD/6-31G*. aug-cc-pVTZ[9] is a Dunning's correlation-consistent, polarizable valence, triple-zeta basis set augmented with diffuse functions. The more reliable DFT method was chosen to report on the calculations afterwards. The B3LYP and M06-2X level calculations were performed using the *Gaussian* 09[10] suite of programs, and the DF-MP2 and CCSD level calculations were performed using the MOLPRO2015.1[11] program.

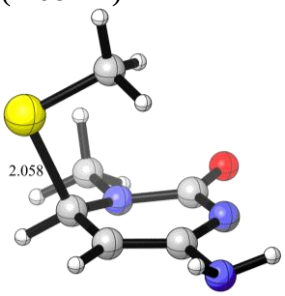
Equilibrium structures and transition states were fully optimized using the chosen DFT method together with the 6-31G* basis set in the gas phase or using the PCM[12] implicit solvent model. The solvent investigated is acetone

according to the Study 2[3b] so as to mimic the dielectric constant of the active site environment. For the fully optimized geometries, vibrational frequency analysis was performed at the same level of theory to confirm the nature of the stationary points as equilibrium structures (with all real frequencies) or transition states (with one and only one imaginary frequency). For some critical transition states, Intrinsic Reaction Coordinate (IRC) calculations in the forward and reverse directions were performed to follow the reaction path. The Natural Bond Orbital (NBO)[13] analysis was performed based on the DFT/6-31G* wavefunction, and the atomic charges were obtained directly from the NBO analysis. All of these calculations were performed using the *Gaussian* 09[10] suite of programs.

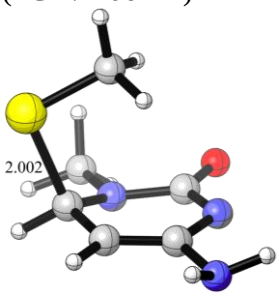
4.3 Results and Discussion

4.3.1 Intermediates from the Nucleophilic Addition of Cys-S to Cytosine-C6

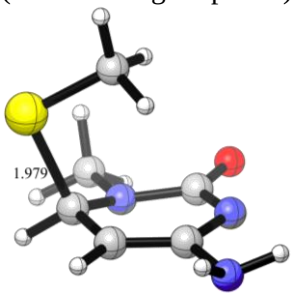
A1) I1-CysCytosine (M06-2X)



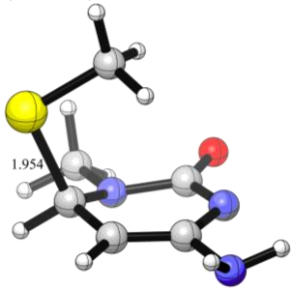
A2) I1-CysCytosine (PCM/M06-2X)



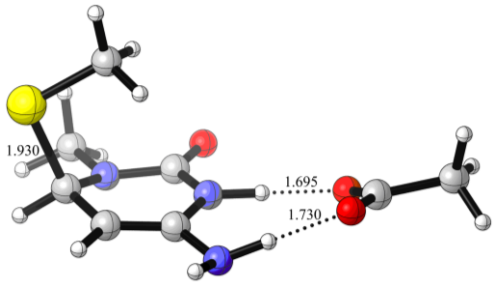
A3) I1-CysCytosine
(DF-MP2/aug-cc-pTVZ)



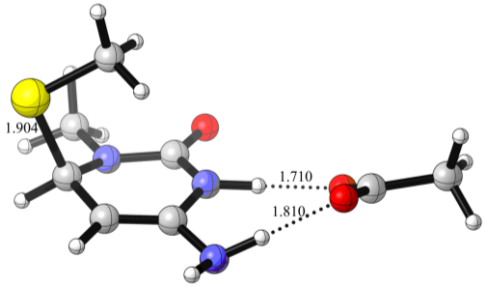
A4) I1-CysCytosine
(CCSD)



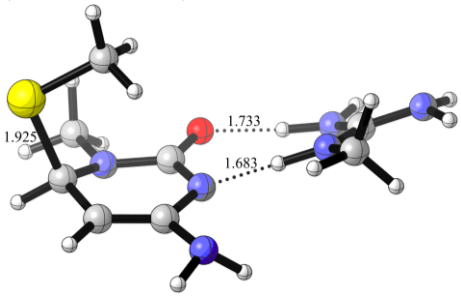
B1) I1-CysCytosineH-Glu
(M06-2X)



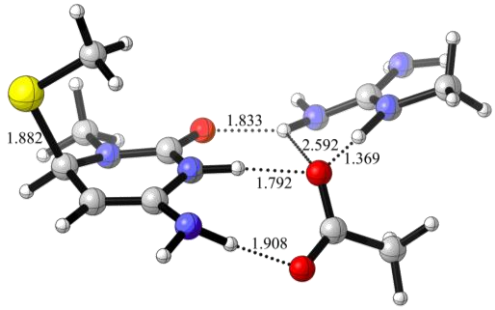
B2) I1-CysCytosineH-Glu
(PCM/M06-2X)



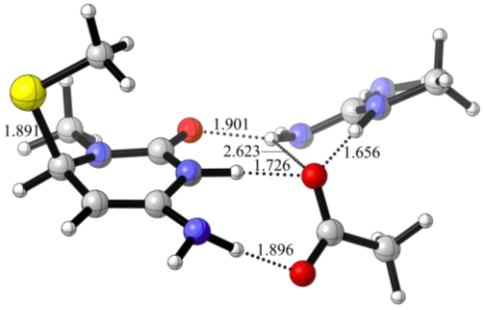
C) I1-CysCytosine-Arg
(PCM/M06-2X)



D1) I1-CysCytosineH-Glu-Arg
(M06-2X)



D2) I1-CysCytosineH-Glu-Arg
(PCM/M06-2X)



E)

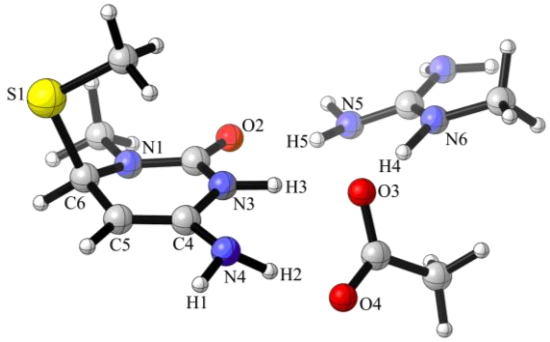
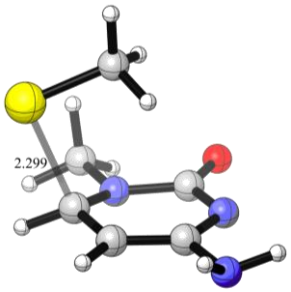
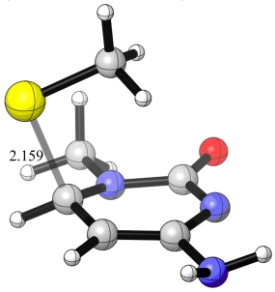


Figure 4.3 Intermediate 1 (I1) from the Cys-thiolate nucleophilic addition to the cytosine-C6. A1)–D2) I1 formed from cytosine, Cys with or without Glu or Arg. E) Atom numbering for reference. The methods (PCM/)M06-2X, DF-MP2 and CCSD are indicated in the brackets for each intermediate. Except for A3), all other methods used the 6-31G* basis set. The interactions are indicated by the dotted lines. Distances are in Å.

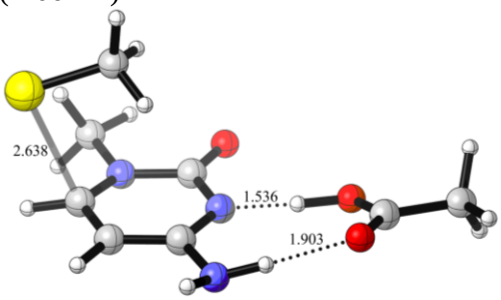
A1) TS1-CysCytosine (M06-2X)



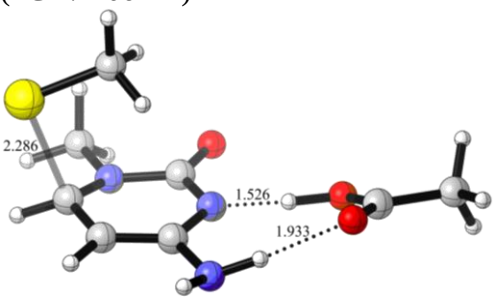
A2) TS1-CysCytosine (PCM/M06-2X)



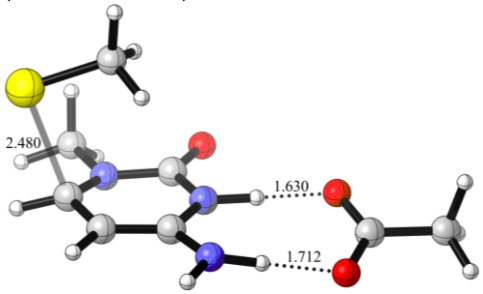
B1) TS1-CysCytosine-GluH (M06-2X)



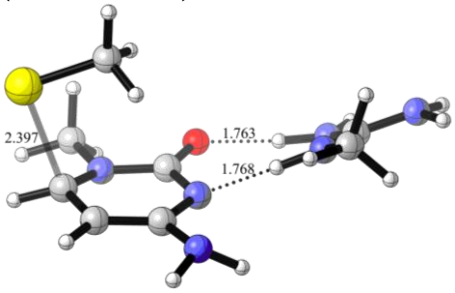
B2) TS1-CysCytosine-GluH (PCM/M06-2X)



C) TS1-CysCytosineH-Glu (PCM/M06-2X)



D) TS1-CysCytosine-Arg (PCM/M06-2X)



E) TS1-CysCytosineH-Glu-Arg (PCM/M06-2X)

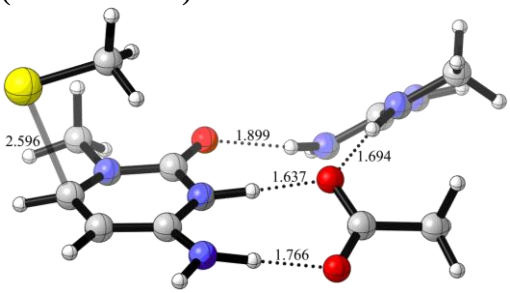


Figure 4.4 TS1 from the Cys-thiolate nucleophilic addition to the cytosine-C6. A1)–E) TS1 formed from cytosine, Cys with or without Glu or Arg. The methods M06-2X and PCM/M06-2X are indicated in the brackets for each TS. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

To study the intermediate 1 (I1) formed from the Cys-thiolate nucleophilic addition to the C6 of cytosine (Cyt), the co-factor SAM (as trimethylsulfonium ion) was excluded from the small model system depicted in Figure 4.1 because it is not directly involved in this reaction step. The intermediate 1 was constructed by the addition of the catalytic Cys81 (as methylthiolate) to the target 2'-deoxycytidine (as 1-methylcytosine) first, followed by the addition of Glu119 (as acetic acid) and/or Arg165 (as 1-methylguanidinium). The starting geometries were built based on the crystal structure of M.HhaI (6MHT)[5]. These intermediates were fully optimized in the gas phase or a solvent environment modeled by PCM and the optimized

geometries together with the atom numbering for the biggest system are shown in Figure 4.3. The similar approach was adopted in the study of the transition state 2 from the methylation step in section 4.3.2.

The simplest intermediate I1-CysCytosine, formed from the addition of Cys-thiolate (S1) to Cyt-C6, was optimized at the (PCM/)B3LYP/6-31G*, (PCM/)M06-2X/6-31G*, DF-MP2/aug-cc-pVTZ and CCSD/6-31G* levels of theory. The C6-S1 bond formed between Cys-S and Cyt-C6 is always broken after the optimization at the (PCM/)B3LYP/6-31G* level. Such results were supported by the potential energy surface (PES) scan along the C6-S1 bond varied from 1.8 Å to 3.6 Å in the increment of 0.1 Å (Figure S 8.3A1-1)–1-2)). The energy of the B3LYP/6-31G* optimized system decreases monotonically as the C6-S1 distance increases from 1.8 Å to 3.0 Å before the complete bond breakage. The similar trend was reported in a PES study at the B3LYP/6-31+G* level in the Study 1[3a] and a similar study at the B3LYP/6-31G** level in the Study 2[3b]. Similarly, the energy of the PCM/B3LYP/6-31G* optimized system almost decreases monotonically as the C6-S1 distance increases from 1.8 Å to 3.1 Å before the complete bond breakage with a small kink around 2.8 Å that corresponds to a rotation around the N1-CH_3 bond in Cyt. Re-optimization with other basis sets such as 6-31+G* and 6-31G(2d) generated the similar results using the B3LYP functional. In contrast, I1-CysCytosine is stable when optimized at the (PCM/)M06-2X/6-31G*, DF-MP2/aug-cc-pVTZ and CCSD/6-31G* levels of theory as shown in Figure 4.3A1)–4) and the optimized structures are highly similar. The relative orientation of the -S1-CH_3 group in Cys to the cytosine ring is different from that in the crystal structure

of M.HhaI (Figure 4.2(a)). The -CH_3 group in the optimized geometries is on top of the cytosine ring to avoid steric repulsion between this group and the cytosine ring in the input structure. The same relative orientation was reported in Studies 1[3a] and 5[4]. In addition, this -CH_3 group forms the possible CH···π interactions with the cytosine ring (distance < 3 Å) that may not be captured by the B3LYP functional. The C6-S1 distances are 2.058 Å and 2.002 Å at the M06-2X/6-31G* and PCM/M06-2X/6-31G* levels respectively. The C6-S1 distance at the M06-2X/6-31G* level is slightly larger than that at the DF-MP2/aug-cc-pVTZ (+3.99%) and CCSD/6-31G* (+5.32%) levels. The (PCM/)M06-2X/6-31G* level results were also supported by the PES scan along the C6-S1 bond (Figure S 8.3A2-1)–2-2)). Therefore, the M06-2X functional slightly underestimates the C-S bond stability but it is better than the B3LYP functional at describing the C-S bond formation in the I1 regarding the molecular geometry when compared to the DF-MP2 and CCSD methods, so this functional was chosen to report all the other calculations in this chapter.

From the PES scan along the C6-S1 bond (Figure S 8.3A2-1)–2-2)), transition states were located at the M06-2X/6-31G* and PCM/M06-2X/6-31G* levels of theory and the optimized geometries are shown in Figure 4.4A2-1)–2-2). The C6-S1 distances in TS1-CysCytosine are 2.299 Å and 2.159 Å in the gas phase and solution phase respectively. These distances are slightly larger than that in the corresponding I1-CysCytosine by 6.48% and 2.80% respectively. Thus, TS1-CysCytosine is a late transition state that closely resembles the product I1-CysCytosine.

All attempts with full optimization failed to obtain the preTS1-CysCytosine geometries that resemble the crystal structure of M.HhaI (6MHT)[5] with the S1(Cys)-C6(Cyt) distance close to 3.602 Å before the nucleophilic addition. Such results are understandable because there is no force to hold Cys-S1 close to Cyt-C6 in the small system, unlike in the active site environment. Similar results were obtained for all the other preTS1 geometries; consequently, energy profiles from this step cannot be computed.

For the intermediate 1 formed from the addition of Cys-S1 to Cyt-C6 in the presence of the protonated Glu (GluH), the Glu-carboxylic proton (H3) is always transferred to Cyt-N3 after the optimization at the (PCM/)M06-2X/6-31G* level of theory. The optimized geometries of I1-CysCytosineH-Glu are shown in Figure 4.3B1)–2). The C6-S1 distances are 1.930 Å and 1.904 Å in the gas phase and solution phase respectively. The N3-H3 and N4-H2 in Cyt each forms a single H-bond with one of the carboxylate O's in Glu as indicated by the short N3-H3⋯O3 (1.695/1.710 Å) and N4-H2⋯O4 (1.730/1.810 Å) distances.

The nucleophilic addition was also examined by the PES scan along the C6-S1 bond (Figure S 8.3B1)–2)). The energy of the M06-2X/6-31G* optimized system decreases and then increases monotonically as the C6-S1 distance increases from 1.8 Å to 3.0 Å before the complete bond breakage. Thus, no transition state was located at the M06-2X/6-31G* level. In contradiction, a transition state was located at the PCM/M06-2X/6-31G* level and the optimized geometry is shown in Figure 4.4C). The C6-S1 distance in TS1-CysCytosineH-Glu is 2.480 Å and it is larger than that in the

corresponding I1-CysCytosineH-Glu by 30.25%. However, the H-bonds between the N3-H3 or N4-H2 in Cyt and one of the carboxylate O's in Glu are both strengthened in TS1-CysCytosineH-Glu when compared with I1-CysCytosineH-Glu as indicated by the shorter N3-H3···O3 (1.630 Å) and N4-H2···O4 (1.712 Å) distances, possibly to provide more stabilization of the TS1.

Two special transition states TS1-CysCytosine-GluH without the proton transfer from GluH to Cyt-N3 were located at the M06-2X/6-31G* and PCM/M06-2X/6-31G* levels from direct optimization of the input structures, and the optimized geometries are shown in Figure 4.4B1)–2). The C6-S1 distances are 2.638 Å and 2.286 Å in the gas phase and solution phase respectively. In addition, two H-bonds are formed between Cyt and GluH as indicated by the short O3-H3···N3 (1.536/1.526 Å) and N4-H2···O4 (1.903/1.933 Å) distances. The PCM/M06-2X/6-31G* optimized geometry of TS1-CysCytosine-GluH (Figure 4.4B2)) is similar to that optimized at the PCM/M06-2X/6-31+G** level reported in the Study 5[4] with the C6-S1 distance of 2.35 Å.

For the intermediate 1 formed from the addition of Cys-S1 to Cyt-C6 in the presence of Arg, the N6-H4 proton in Arg is transferred to Cyt-N3 after the optimization at the M06-2X/6-31G* level of theory possibly due to the greater stability of this structure over the input structure. The PCM/M06-2X/6-31G* optimized I1-CysCytosine-Arg is shown in Figure 4.3C) and it does not resemble that in the crystal structure of M.HhaI (6MHT)[5]. The C6-S1 distance is 1.925 Å. Instead of forming two H-bonds between either one of the NH's in Arg and Cyt-O2, one of the NH's forms a H-bond with Cyt-O2 as indicated by

the short N5-H5···O2 (1.733 Å) distance, while the other one of the NH's forms a H-bond with Cyt-N3 as indicated by the short N6-H4···N3 (1.683 Å) distance.

The nucleophilic addition was examined by the PES scan along the C6-S1 bond (Figure S 8.3C)). A transition state was located at the PCM/M06-2X/6-31G* level and the optimized geometry is shown in Figure 4.4C). The C6-S1 distance in TS1-CysCytosine-Arg is 2.397 Å, and it is larger than that in the corresponding I1-CysCytosine-Arg by 24.52%. The H-bonds between the N5-H5 in Arg and Cyt-O2 and between the N6-H4 in Arg and Cyt-N3 are both weakened in TS1-CysCytosine-Arg when compared with I1-CysCytosine-Arg as indicated by the larger N5-H5···O2 (1.763 Å) and N6-H4···N3 (1.768 Å) distances, suggesting the better stabilization of the I1-CysCytosine over the TS1-CysCytosine by Arg.

For the intermediate 1 formed from the addition of Cys-S1 to Cyt-C6 in the presence of GluH and Arg, the Glu-carboxylic proton (H3) is always transferred to Cyt-N3 after the optimization at the (PCM/)M06-2X/6-31G* level of theory, which is similar to the optimization of the intermediate 1 in the presence of GluH. The optimized geometries of I1-CysCytosineH-Glu-Arg are shown in Figure 4.3D1)–2) and they do not resemble the crystal structure of M.HhaI (6MHT)[5]. The C6-S1 distances are 1.882 Å and 1.891 Å in the gas phase and solution phase respectively. Instead of forming two H-bonds between either one of the NH's in Arg and Cyt-O2, one of the NH's forms a H-bond with Cyt-O2 as indicated by the short N5-H5···O2 (1.833/1.901 Å) distance, and the other NH forms a H-bond with Glu-O3 as indicated by the

short N6-H4···O3 (1.369/1.656 Å) distance. In addition, one additional H-bond is possibly formed between the NH in Arg and Glu-O3 as indicated the moderate N5-H5···O3 (2.592/2.623 Å) distance. The deviation of the interaction patterns from that in the crystal structure is due to the shifting of Arg towards Glu. Similar to I1-CysCytosineH-Glu, the N3-H3 and N4-H2 in Cyt each forms a single H-bond with one of the carboxylate O's in Glu as indicated by the short N3-H3···O3 (1.792/1.726 Å) and N4-H2···O4 (1.908/1.896 Å) distances.

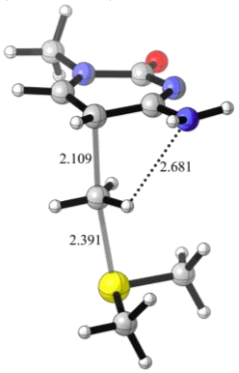
Again, this nucleophilic addition reaction was examined by the PES scan along the C6-S1 bond (Figure S 8.3C1)–2)). As the C6-S1 distance increases, the oppositely charged Cys-thiolate and Arg gets closer due to the collapsing of charges when optimized at the M06-2X/6-31G* level, so no transition state could be located. However, a transition state was located at the PCM/M06-2X/6-31G* level and the optimized geometry is shown in Figure 4.4E). The C6-S1 distance in TS1-CysCytosineH-Glu-Arg is 2.596 Å and it is larger than that in the corresponding I1-CysCytosineH-Glu-Arg by 37.28%. The H-bond between the N3-H3 in Arg and Glu-O3 is slightly weakened in TS1-CysCytosineH-Glu-Arg when compared with I1-CysCytosineH-Glu-Arg as indicated by the larger N3-H3···O3 (1.694 Å) distance. On the contrary, the other three H-bonds between the N5-H5 in Arg and Cyt-O2, between the N3-H3 in Cyt and Glu-O3, and between the N4-H2 in Cyt and Glu-O4 are all strengthened in TS1-CysCytosineH-Glu-Arg as indicated by the shorter N5-H5···O2 (1.899 Å), N4-H2···O4 (1.637 Å) and N6-H4···O3 (1.766 Å) distances.

All of these results showed that the intermediate 1 (I1) is a stable structure with the surrounding side chains of Glu or Arg. The side chain of Glu stabilizes the I1 not only through hydrogen bonding between the N4-H2 in Cyt and Glu-O4 but also via direct protonation of Cyt-N3 by Glu-OH as shown in I1-CysCytosineH-Glu and I1-CysCytosineH-Glu-Arg. The C6-S1 distances in I1-CysCytosineH-Glu are slightly shortened when compared with I1-CysCytosine by 6.22% and 2.45% in the gas phase and solution phase respectively. These results suggested that the role of Glu in N3 protonation is critical to the reaction step 1 in the gas phase and solution phase, which are consistent with the Study 1[3a] and experimental results[14]. Although none of the optimized geometries of the I1 involving the side chain of Arg closely resemble that in the crystal structure of M.HhaI (6MHT)[5], the roles of the N5-H5 and N6-H4 in Arg in the hydrogen bonding with Cyt-O2 or Cyt-N3 were identified and the position of the Arg side chain could be easily changed in the I1 depending on its surroundings. In our biggest simple model system calculations, the N6-H4 in Arg even interacts with the Glu-O3 in the I1, but this interaction was never observed in the protein active site environment as the two N5-H's Arg tend to interact with Cyt-O2, Sugar-O4' and the phosphate O's from the DNA backbone, thus shifting the side chain of Arg away from the side chains of Glu. The N6-O3 and N5-O3 distances in the crystal structure of M.HhaI (6MHT)[5] are 3.557 Å and 5.529 Å respectively. By comparing the H-bond distances in the I1 and the corresponding TS1, it was noticed that the side chain of Glu tends to stabilize the TS1 more than the I1, whereas the side chain of Arg tends to stabilize the I1 more than the TS1.

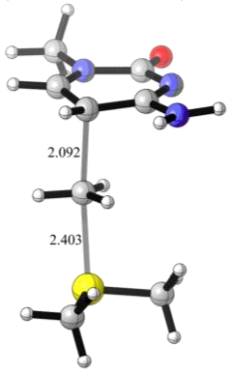
By comparing the same I1 optimized in the gas phase and solution phase (Figure 4.3), it was found that the C6-S1 bond is shorter in the solution phase by -1.37–2.80% in I1-CysCytosine and I1-CysCytosineH-Glu, but it is slightly longer in the solution phase by 0.48% in I1-CysCytosineH-Glu-Arg. Nonetheless, the I1 geometries in these two phases resemble each other closely. However, the PES scan in the gas phase can deviate a lot from that in the solution phase (Figure S 8.3), and most of the TS1 was only located in the solution phase (Figure 4.4). Hence, it is better to include the solvent effects for calculating the PES and TS1 structures.

4.3.2 Transition States from the Methylation of Cytosine-C5 by SAM

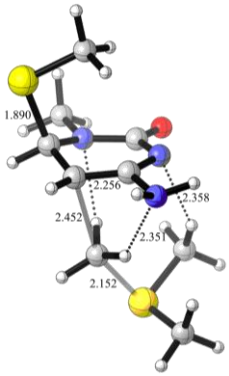
A1) TS2-Cytosine-SAM (M06-2X)



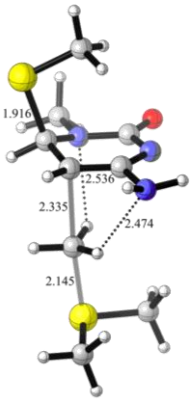
A2) TS2-Cytosine-SAM (PCM/M06-2X)



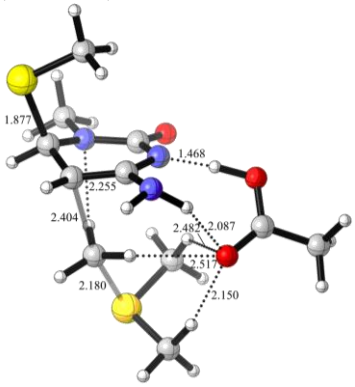
B1) TS2-CysCytosine-SAM (M06-2X)



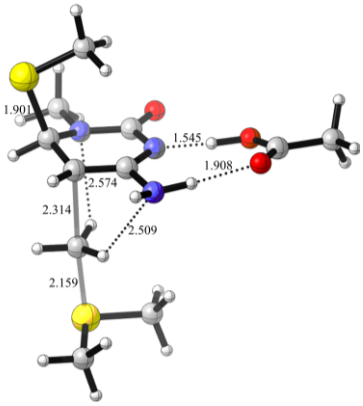
B2) TS2-CysCytosine-SAM (PCM/M06-2X)



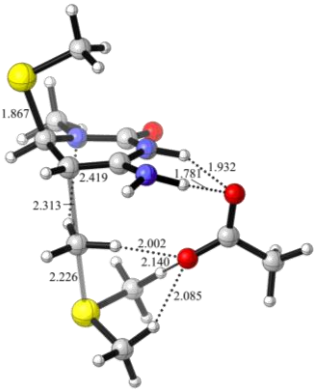
C1) TS2-CysCytosine-GluH-SAM (M06-2X)



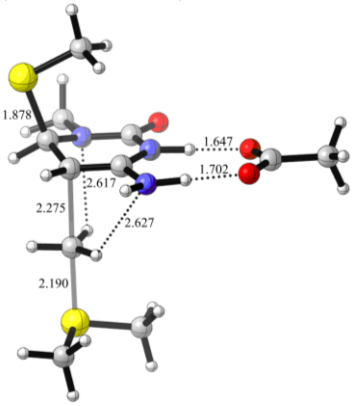
C2) TS2-CysCytosine-GluH-SAM (PCM/M06-2X)



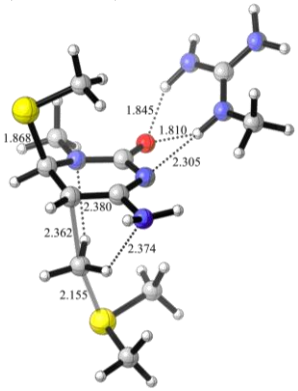
D1) TS2-CysCytosineH-Glu-SAM (M06-2X)



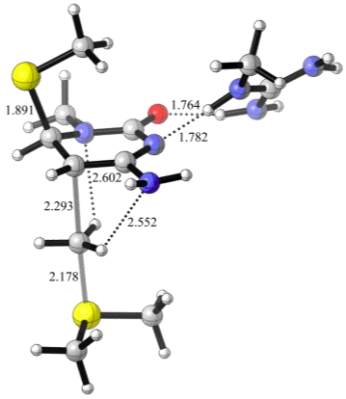
D2) TS2-CysCytosineH-Glu-SAM (PCM/M06-2X)



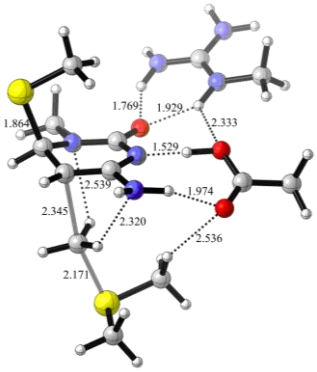
E1) TS2-CysCytosine-Arg-SAM (M06-2X)



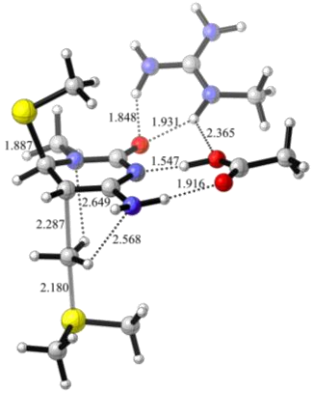
E2) TS2-CysCytosine-Arg-SAM (PCM/M06-2X)



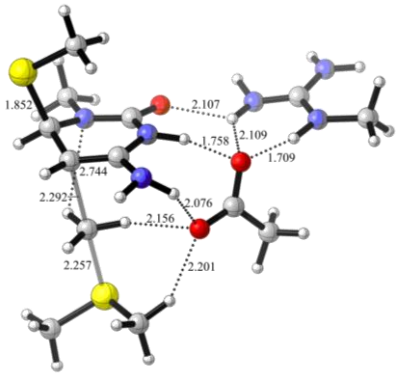
F1) TS2-CysCytosine-GluH-Arg-SAM (M06-2X)



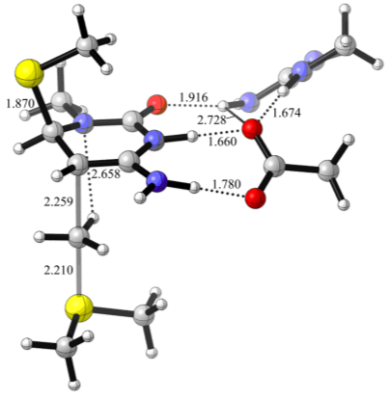
F2) TS2-CysCytosine-GluH-Arg-SAM (PCM/M06-2X)



G1) TS2-CysCytosineH-Glu-Arg-SAM (M06-2X)



G2) TS2-CysCytosineH-Glu-Arg-SAM (PCM/M06-2X)



H)

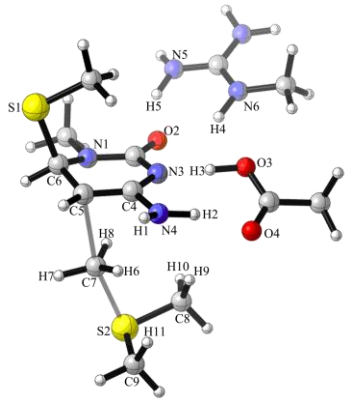


Figure 4.5 TS of the methyl transfer from the sulfonium ion of the model SAM to cytosine-C5. A1)–G2) TS formed from cytosine, SAM with or without Cys, Glu or Arg. H) atom numbering for reference. The methods M06-2X and PCM/M06-2X are indicated in the brackets for each transition state. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

Table 4.1 Key geometric parameters for TS from methyl transfer[a]

(a)

No.	A1)	A2)	B1)	B2)
Name	TS2-Cytosine-SAM		TS2-CysCytosine-SAM	
Method	M06-2X	PCM/M06-2X	M06-2X	PCM/M06-2X
Distance				
C5-S2	4.490	4.488	4.508	4.475
Bond Angles				
C5-C7-S2	172.4	173.5	156.5	174.8
C4-C5-S2	92.3	96.6	72.8	87.8
Torsional Angles				
C8-S2-C5-C4	-9.1	5.3	-41.8	-11.0
C9-S2-C5-C4	91.3	106.1	56.5	90.0
Improper Torsional Angle				
H6-C7-H7-H8	-174.3	-172.5	156.3	157.9

(b)

No.	C1)	C2)	D1)	D2)
Name	TS2-CysCytosine-GluH-SAM		TS2-CysCytosineH-Glu-SAM	
Method	M06-2X	PCM/M06-2X	M06-2X	PCM/M06-2X
Distance				
C5-S2	4.537	4.469	4.533	4.464
Bond Angles				
C5-C7-S2	163.5	175.8	174.0	178.2
C4-C5-S2	80.6	89.3	91.1	93.4
Torsional Angles				
C8-S2-C5-C4	-45.0	-8.6	-47.1	-7.0
C9-S2-C5-C4	54.0	92.3	53.8	94.6
Improper Torsional Angle				
H6-C7-H7-H8	161.5	160.0	167.6	164.1

(c)

No.	E1)	E2)
Name	TS2-CysCytosine-Arg-SAM	
Method	M06-2X	PCM/M06-2X
Distance		
C5-S2	4.491	4.469
Bond Angles		
C5-C7-S2	167.6	177.1
C4-C5-S2	81.2	91.5
Torsional Angles		
C8-S2-C5-C4	-32.7	-4.4
C9-S2-C5-C4	66.9	96.9
Improper Torsional Angle		
H6-C7-H7-H8	158.3	162.1

(d)

No.	F1)	F2)	F1)	F2)
Name	TS2-CysCytosine-GluH-Arg-SAM		TS2-CysCytosineH-Glu-Arg-SAM	
Method	M06-2X	PCM/M06-2X	M06-2X	PCM/M06-2X
Distance				
C5-S2	4.481	4.466	4.548	4.469
Bond Angles				
C5-C7-S2	165.7	177.2	178.1	178.2
C4-C5-S2	78.7	91.9	87.5	94.7
Torsional Angles				
C8-S2-C5-C4	-15.2	1.6	50.8	-2.1
C9-S2-C5-C4	83.9	103.2	152.9	99.4
Improper Torsional Angle				
H6-C7-H7-H8	161.2	162.7	171.9	166.5

[a] Distances are in Å, bond angles and torsional angles are in °.

To study the transition state 2 (TS2) formed from the electrophilic methyl transfer from the co-factor SAM to the C5 of cytosine, the small model system depicted in Figure 4.1 was used with the C6-S bond already formed between Cys81 and the C5 of cytosine from the reaction step 1. The transition state 2 was constructed by the addition of Glu119 and/or Arg165 to the Cys-Cytosine covalent adduct and SAM. The transferring methyl (-CH_3) group adopted the trigonal planar geometry and was placed right in between Cyt-C5 and SAM-S2 in the input structure based on the S_N2 reaction mechanism. The uncatalyzed reaction was also studied. The relative orientations of these

molecules were built based on the crystal structure of M.HhaI (6MHT)[5] as mentioned earlier. All of the (PCM/)M06-2X/6-31G* optimized transition states 2 and the atom numbering for the biggest system are shown in Figure 4.5. Some of the key geometric parameters are summarized in Table 4.1.

In the uncatalyzed TS2-Cytosine-SAM optimized at the (PCM/)M06-2X/6-31G* level (Figure 4.5A1)–2) and Table 4.1(a)), the C5-S2 distances are 4.490/4.488 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å), but close to that in the crystal structure of mDNMT1 (4DA4)[6] (4.548 Å). The methyl group being transferred is almost planar as measured by the H6-C7-H7-H8 angle (-174.3/-172.5^{o}), and the top of the methyl umbrella points towards Cyt-C5. The methyl group is transferred almost linearly from the donor S2 to the acceptor C5 as measured by the C5-C7-S2 angle (172.4/173.5^{o}). The angles C4-C5-S2 (92.3/96.6^{o}), C8-S2-C5-C4 (-9.1/5.3^{o}) and C9-S2-C5-C4 (91.3/106.1^{o}) did not deviate a lot from that in the crystal structure of M.HhaI (6MHT)[5] (98.1^{o}, -4.1^{o} and 109.6^{o}). In addition, the CH from the transferring methyl may form the C7-H6···N4 interaction with the cytosine ring in the gas phase. The (PCM/)M06-2X/6-31G* optimized geometries of preTS2-Cytosine-SAM showed the side-by-side arrangement of the SAM and cytosine due to the absence of attractive interactions to hold them together to resemble TS2-Cytosine-SAM. Similar observation was reported in the Study 1[3a] for the HF/3-21+G* optimized preTS2-CysCytosine-SAM. The IRC calculations were performed on TS2-Cytosine-SAM to confirm the reaction pathway.

In TS2-CysCytosine-SAM optimized at the (PCM/)M06-2X/6-31G* level (Figure 4.5B1)–2) and Table 4.1(a)), the C6-S1 distances are 1.890/1.916 Å, and the relative orientations of the -S1-CH_3 group in Cys to the cytosine ring are highly similar to that in I1-CysCytosine. The C5-S2 distances are 4.508/4.475 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is slightly non-planar as measured by the H6-C7-H7-H8 angle (156.3/157.9^o), and the top of the methyl umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 in the solution phase as measured by the C5-C7-S2 angle (174.8^o), but not in the gas phase (156.5^o) as the two methyl groups in the leaving SAH were tilted up towards the cytosine ring. Thus, the angles C4-C5-S2 (87.8^o), C8-S2-C5-C4 (-11.0^o) and C9-S2-C5-C4 (90.0^o) in the solution phase deviated less from that in the crystal structure of M.HhaI (6MHT)[5] (98.1^o, -4.1^o, 109.6^o) than the corresponding angles in the gas phase (72.8^o, -41.8^o, 56.5^o). The CH's from the transferring methyl may form the C7-H7⋯N1 or C7-H6⋯N4 interactions with the cytosine ring in the gas phase and solution phase. Moreover, the CH from the leaving SAH form the possible C8-H9⋯N3 interaction with the cytosine ring in the gas phase due to the tilting of SAH (Figure 4.5B1)). The IRC calculations were performed on TS2-CysCytosine-SAM to confirm the reaction pathway.

Similar to section 4.3.1, all attempts with full optimization failed to obtain the preTS2-CysCytosine-SAM geometries with the similar relative orientation of SAM to the cytosine ring as the crystal structure of M.HhaI (6MHT)[5]. Such results are understandable because no strong attractive force is

present to hold SAM right underneath the cytosine ring with the methyl-C7 directly pointing towards Cyt-C5 without the intact active site environment. Similar results were obtained for all preTS2 geometries. In some cases, the interaction patterns between other parts of the system were also changed in the fully optimized preTS2. Hence, energy profiles from this step cannot be computed.

Transition states 2 with the deprotonated or protonated Cyt-N3 were obtained in the presence of GluH or Glu. The (PCM/)M06-2X/6-31G* optimized TS2-CysCytosine-GluH-SAM (Figure 4.5C1)–2) and Table 4.1(b)) is similar to the TS2-CysCytosine-SAM if GluH were absent. The C6-S1 distances are 1.877/1.901 Å, and the C5-S2 distances are 4.537/4.469 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is slightly non-planar as measured by the H6-C7-H7-H8 angle (161.5/160.0^o), and the top of the methyl umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 in the solution phase as measured by the C5-C7-S2 angle (175.8^o), but not in the gas phase (163.5^o). The angles C4-C5-S2 (89.3^o), C8-S2-C5-C4 (-8.6^o) and C9-S2-C5-C4 (92.3^o) in the solution phase deviated less from that in the crystal structure of M.HhaI (6MHT)[5] (98.1^o, -4.1^o, 109.6^o) than the corresponding angles in the gas phase (80.6^o, -45.0^o, 54.0^o). The CH's from the transferring methyl may form the C7-H7⋯N1 or C7-H6⋯N4 interactions with the cytosine ring in the gas phase and solution phase. With the addition of GluH, at least two H-bonds are formed between the OH in GluH and Cyt-N3, and between the NH in Cyt and GluH-O4 as indicated by the short O3-

H3⋯N3 (1.468/1.545 Å) and N4-H2⋯O4 (2.087/1.908 Å) distances. In the gas phase (Figure 4.5C1)), the carboxylic group of GluH is almost perpendicular to the cytosine ring, so three additional H-bonds may form as indicated by the moderate C7-H6⋯O4 (2.517 Å), C8-H9⋯O4 (2.482 Å) and C9-H11⋯O4 (2.150 Å) distances. TS2-CysCytosine-GluH-SAM optimized at the PCM/M06-2X/6-31G* level (Figure 4.5C2-2)) resembles that optimized at the PCM/M06-2X/6-31+G** level reported in the Study 5[4] with the C5-C7 distance of 2.21 Å and the C7-S2 distance of 2.25 Å. The IRC calculations were performed on TS2-CysCytosine-GluH-SAM to confirm the reaction pathway. This time, the IRC from the gas phase showed proton transfer from Cyt-N3 to Glu-O3 with almost concurrent methyl transfer, whereas the IRC from the solution phase showed proton transfer from Cyt-N3 to Glu-O3 prior to the TS2 as a hidden step.

In TS2-CysCytosineH-Glu-SAM optimized at the (PCM/)M06-2X/6-31G* level (Figure 4.5D1-1)–2-2) and Table 4.1(b)), the C6-S1 distances are 1.867/1.878 Å that are smaller than that in TS2-CysCytosine-GluH-SAM. The C5-S2 distances are 4.533/4.464 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is closer to planarity than that in TS2-CysCytosine-GluH-SAM as measured by the H6-C7-H7-H8 angle (167.6/164.1^{o}), and the top of the methyl umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 as measured by the C5-C7-S2 angle (174.0/178.2^{o}). However, the angles C4-C5-S2 (93.4^{o}), C8-S2-C5-C4 (-7.0^{o}) and C9-S2-C5-C4 (94.6^{o}) in the solution phase deviated less from that in the crystal structure of M.HhaI

(6MHT)[5] (98.1°, -4.1°, 109.6°) than the corresponding angles in the gas phase (91.1°, -47.1°, 53.8°). The CH's from the transferring methyl may form the C7-H7···N1 or C7-H6···N4 interactions with the cytosine ring in the gas phase and solution phase. In the gas phase, the carboxylate group of Glu is almost perpendicular to the cytosine ring similar, resulting in five possible H-bonds as indicated by the short N3-H3···O3 (1.932 Å) and N4-H2···O3 (1.781 Å) distances, and the moderate C7-H6···O4 (2.002 Å), C8-H9···O4 (2.140 Å) and C9-H11···O4 (2.085 Å) distances. In the solution phase, the carboxylate group of Glu is in the plane of the cytosine ring similar to I1-CysCytosineH-Glu, resulting in two possible H-bonds as indicated by the short N3-H3···O3 (1.647 Å) and N4-H2···O4 (1.702 Å) distances. The IRC calculations were performed on TS2-CysCytosineH-Glu-SAM to confirm the reaction pathway.

In TS2-CysCytosine-Arg-SAM optimized at the (PCM/)M06-2X/6-31G* level (Figure 4.5E1-1)–2-2) and Table 4.1(c)), the C6-S1 distances are 1.868/1.891 Å, and the C5-S2 distances are 4.491/4.469 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is slightly non-planar as measured by the H6-C7-H7-H8 angle (158.3/162.1°), and the top of the methyl umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 in the solution phase as measured by the C5-C7-S2 angle (177.1°), but not in the gas phase (167.6°). The angles C4-C5-S2 (91.5°), C8-S2-C5-C4 (-4.4°) and C9-S2-C5-C4 (96.9°) in the solution phase deviated less from that in the crystal structure of M.HhaI (6MHT)[5] (98.1°, -4.1°, 109.6°) than the corresponding angles in the gas phase (81.2°, -32.7°, 66.9°). The CH's from the transferring methyl may

form the C7-H7⋯N1 or C7-H6⋯N4 interactions with the cytosine ring in the gas phase and solution phase. In the gas phase, three H-bonds are formed between Arg and the cytosine ring as indicated by the short N5-H5⋯O2 (1.845 Å) and N6-H4⋯O2 (1.810 Å) distances, and the moderate N6-H4⋯N3 (2.305 Å) distance. In the solution phase, the Arg was shifted towards Cyt-N3, resulting in only two possible H-bonds as indicated by the short N5-H5⋯O2 (1.764 Å) and N6-H4⋯N3 (1.782 Å) distances. The IRC calculations were performed on TS2-CysCytosine-Arg-SAM to confirm the reaction pathway.

Transition states 2 with the deprotonated or protonated Cyt-N3 were obtained in the presence of GluH or Glu and Arg. In TS2-CysCytosine-GluH-Arg-SAM optimized at the (PCM/)M06-2X/6-31G* level (Figure 4.5F1)–2) and Table 4.1(d)), the C6-S1 distances are 1.864/1.887 Å, and the C5-S2 distances are 4.481/4.466 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is slightly non-planar as measured by the H6-C7-H7-H8 angle (161.2/162.7^{o}), and the top of the Me umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 in the solution phase as measured by the C5-C7-S2 angle (177.2^{o}), but not in the gas phase (165.7^{o}). The angles C4-C5-S2 (91.9^{o}), C8-S2-C5-C4 (1.6^{o}) and C9-S2-C5-C4 (103.2^{o}) in the solution phase deviated less from that in the crystal structure of M.HhaI (6MHT)[5] (98.1^{o}, -4.1^{o}, 109.6^{o}) than the corresponding angles in the gas phase (78.7^{o}, -15.2^{o}, 83.9^{o}). The CH's from the transferring methyl may form the C7-H7⋯N1 or C7-H6⋯N4 interactions with the cytosine ring as shown in the gas phase and solution phase. Two H-bonds are formed between Arg and the cytosine ring as

indicated by the short N5-H5···O2 (1.769/1.848 Å) and N6-H4···O2 (1.929/1.931 Å) distances. Similar to TS2-CysCytosine-GluH-SAM, at least two H-bonds are formed between the OH in GluH and Cyt-N3 and between the NH in Cyt and the GluH-O4 as indicated by the short O3-H3···N3 (1.529/1.547 Å) and N4-H2···O4 (1.974/1.916 Å) distances. In the gas phase (Figure 4.5F2-1)), the carboxylic group of GluH was slightly off the plane of the cytosine ring, such that one additional H-bond may form as indicated by the moderate C7-H6···O4 (2.536 Å) distance. The IRC calculations were performed on TS2-CysCytosine-GluH-Arg-SAM to confirm the reaction pathway.

TS2-CysCytosineH-Glu-Arg-SAM optimized at the (PCM/)M06-2X/6-31G* level of theory (Figure 4.5G1)–2) and Table 4.1(d)) is similar to I1-CysCytosineH-Glu-Arg if SAM were taken out. The C6-S1 bond distances are 1.852/1.870 Å that are smaller than that in TS2-CysCytosine-GluH-Arg-SAM. The C5-S2 distances are 4.548/4.469 Å, longer than that in the crystal structure of M.HhaI (6MHT)[5] (4.238 Å). The methyl group being transferred is slightly non-planar as measured by the H6-C7-H7-H8 angle (171.9/166.5^{o}), and the top of the methyl umbrella points towards SAM-S2. The methyl group is transferred almost linearly from S2 to C5 as measured by the C5-C7-S2 angle (178.1/178.2^{o}). However, the angles C4-C5-S2 (94.7^{o}), C8-S2-C5-C4 (-2.1^{o}) and C9-S2-C5-C4 (99.4^{o}) in the solution phase deviated less from that in the crystal structure of M.HhaI (6MHT)[5] (98.1^{o}, -4.1^{o}, 109.6^{o}) than the corresponding angles in the gas phase (87.5^{o}, 50.8^{o}, 152.9^{o}) due to the interactions of the leaving SAH with Glu-O4. The CH's from the transferring

methyl may form the C7-H7···N1 interaction with the cytosine ring as shown in the gas phase and solution phase. Instead of forming two H-bonds between either one of the NH's in Arg and Cyt-O2 like TS2-CysCytosine-GluH-Arg-SAM, one of the NH's forms a H-bond with Cyt-O2 as indicated by the short N5-H5···O2 (2.107/1.916 Å) distance, and a H-bond with Glu-O3 as indicated by the moderate N5-H5···O3 (2.109/2.728 Å) distance; while the other NH forms a H-bond with Glu-O3 as indicated by the short N6-H4···O3 (1.709/1.674 Å) distance. Similar to I1-CysCytosineH-Glu and I1-CysCytosineH-Glu-Arg, the N3-H3 and N4-H2 in Cyt each forms a H-bond with one of the carboxylate O's in Glu as indicated by the short N3-H3···O3 (1.758/1.1.660 Å) and N4-H2···O4 (1.758/1.660 Å) distances. Two additional H-bonds are formed between Glu and SAM in the gas phase as indicated by the moderate C7-H6···O4 (2.156 Å) and C8-H10···O4 (2.201 Å) distances. The IRC calculations were performed on TS2-CysCytosineH-Glu-Arg-SAM to confirm the reaction pathway.

To investigate the role of Cys in the methylation reaction, the geometry of TS2-Cytosine-SAM was compared with that of the other TS2 with the covalent bond formed between Cys-S1 and Cyt-C6. The C6-C7 distance is smaller than the C7-S2 distance with the top of the transferring methyl umbrella points towards Cyt-C6 in the uncatalyzed reaction, whereas it is larger than the C7-S2 distance with the top of the transferring methyl umbrella points towards SAM-S2 in the catalyzed reactions. These results suggested that the uncatalyzed methylation reaction involves a late TS, while the catalyzed methylation reactions all involve an early TS, thus such reaction is

more favorable with the help of the side chains of Cys. More importantly, the addition of Cys-S1 to Cyt-C6 helps to increase the nucleophilicity of Cyt-C5 as evidenced by the more negative NBO charge on C5 in the PCM/M06-2X/6-31G* optimized TS1-CysCytosine-SAM (-0.451 e) than that in TS1-Cytosine-SAM (-0.484 e).

All of the results in this section showed that the complex of I1 and SAM is able to undergo methylation with or without the surrounding side chains of Glu or Arg. The roles of Glu and Arg were discussed separately below.

The direct protonation of Cyt-N3 by Glu-OH is not always required in the TS2, different from the I1. The side chain of Glu can stabilize the TS2 through indirect hydrogen bonding between Cyt-N3 and Glu-OH or via direct protonation of Cyt-N3 by Glu-OH. Similar results were reported in the Study 1[3a] though the N3 protonation was modeled by the addition of a proton to Cyt-N3 in the absence of Glu. The energies of TS2-CysCysotineH-Glu-SAM are lower than that of TS2-CysCytosine-GluH-SAM in the gas phase and solution phase, and the C6-S1 bonds in TS2-CysCytosineH-Glu-SAM are stronger. But the geometries of TS2-CysCytosine-GluH-Arg-SAM resemble the crystal structure of M.HhaI (6MHT)[5] more closely than that of TS2-CysCytosineH-Glu-SAM. Additionally, the proton transfer from N3-H3 to Glu-O3 prior to or at the TS2 was observed in the IRC calculations for TS2-CysCytosine-GluH in the gas phase or solution phase, which seems to suggest that the proton on N3 is not preferred in the TS2-CysCytosineH-Glu-SAM. Overall, those results conveyed contradictory messages regarding the role of Glu for N3-protonation

in the TS2 and energy profiles of this reaction step may be necessary to address the controversies.

The (PCM/)M06-2X/6-31G* optimized geometries of TS2-CysCytosine-Arg-SAM and TS2-CysCytosine-GluH-Arg-SAM closely resemble that in the crystal structure of M.HhaI (6MHT)[5] with one exception unlike I1-CysCytosine-Arg, I1-CysCytosineH-Glu-Arg and TS2-CysCytosineH-Glu-Arg-SAM with the shifting of the Arg side chain. The roles of the two NH's in Arg in the hydrogen bonding with Cyt-O2 were demonstrated. By comparing TS2-CysCytosine-SAM and TS2-CysCytosine-Arg-SAM, it was found that the C5-C7 distance decreases and the C7-S2 distance increases when the side chain of Arg is added to the TS2, suggesting that the latter TS2 is later than the former TS2, thus Arg may slightly destabilize the TS2 possibly due to the electron-withdrawing effects of the NH's via hydrogen bonding.

By comparing the same TS2 optimized in the gas phase and solution phase (Figure 4.5), it was found that the C6-S1 bond is always longer in the solution phase. Most of the TS2 geometries in the solution phase resemble the crystal structure of M.HhaI (6MHT)[5] more closely than that in the gas phase. In the gas phase, the tilting of the methyl groups in the leaving SAH towards the cytosine ring and the change of the relative orientation of Glu to the cytosine ring are more likely to form more intramolecular interactions due to the lack of solvent stabilization. Hence, the solvent effects should be included for calculating the TS2 structures.

4.3.3 Performance of the Small Model System and Possible Improvements

The small model system is useful to study the TS1, I1 and TS2 from the first two steps in the DNA methylation mechanism as well as the individual roles of Cys, GluH and Arg, but it failed to calculate the preTS1 and preTS2 that resemble the preTS in the intact protein active site. In the protein active site, the species involved in the reaction are held together by the folded conformation of that protein. However, in the small model system, there is no such force to hold the species in a certain way other than the bonded and non-bonded interactions present in that system. Consequently, the energy profiles from these two reaction steps could not be computed to confirm the roles of Cys, Glu and Arg directly. The concerted one-step reaction involving the nucleophilic attack and concurrent methyl transfer cannot be studied by such systems, either. One important point learnt from the small model system study is that the solvation should be included in the optimization stage of the calculation in this system due to the prominent solvent effects on the PES profiles and the TS geometries.

To probe the preTS1 and preTS2 and to study the one-step concerted process, constrained optimization was attempted based on the B3LYP functional with different approaches, including constraining the selected atomic coordinates and/or constraining certain distances/angles, but all of these approaches failed due to technical errors or undesired final geometries. It is impossible to locate the Cys-S1 position in the preTS1 and to maintain the relative orientation of SAM to the cytosine ring in the preTS2. In the four

(PCM/)B3LYP/6-31G* constrained and optimized geometries of the I1 including SAM, the breakage of the C6-S1 bonds was observed.

Based on these results, we think that the only way to improve the calculation of the preTS1 and preTS2 is to expand our small model system which may require the implementation of the state-of-the-art methods such as QM/QM or QM/MM methods.

One possible improvement to probe the preTS2 is to include two more C's at each end of the SAM model and restrain the atomic coordinates of these two C's together with the other three ending C's in Cys, Arg and Glu during the optimization. The geometries of the M06-2X/6-31G* optimized preTS2-CysCytosine-GluH-Arg-SAM and TS2-CysCytosine-GluH-Arg-SAM are shown in Figure 4.6.

(a) PreTS2-CysCytosine-GluH-Arg-SAM (b) TS2-CysCytosine-GluH-Arg-SAM

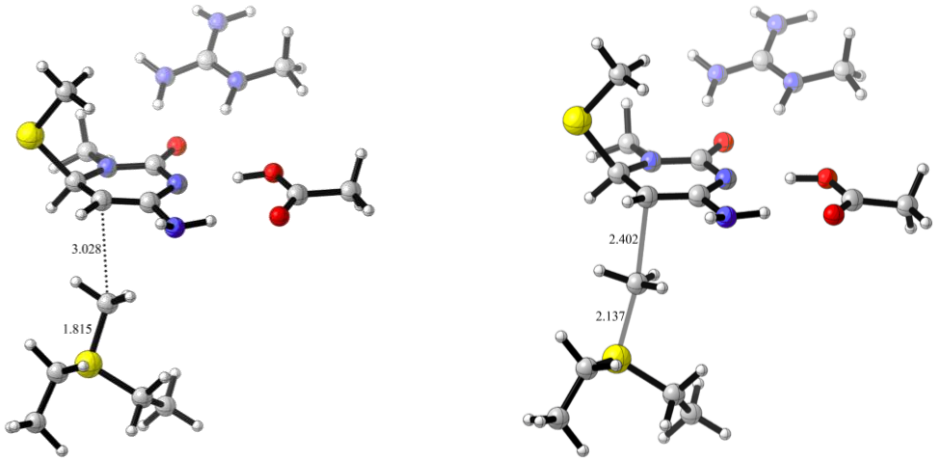


Figure 4.6 Constrained and optimized geometries of (a) preTS2-CysCytosine-GluH-Arg-SAM and (b) TS2-CysCytosine-GluH-Arg-SAM at the M06-2X/6-31G* level.

4.4 Conclusion

In this chapter, the first two steps in the DNA methylation reaction—the Cys nucleophilic addition to the C6 of cytosine and the methyl transfer from SAM to the C5 of cytosine—were mainly studied at the (PCM/)M06-2X/6-31G* level of theory based on a truncated small model system (Figure 4.1) to probe the key transition sates (TS1 and TS2) and intermediates (I1) with or without the side chains of Cys, Glu or Arg surrounding them, to study the existence and stability of the I1, and the ability of this intermediate to undergo methylation. The roles of the side chains of Cys, Glu and Arg were also discussed.

The results from the reaction step 1 showed that the I1 is a stable structure with the surrounding Glu or Arg side chains. The role of the Glu side chain is to stabilize the TS1 and I1 through hydrogen bonding with Cyt-N4-H2 and via direct protonation of Cyt-N3 by Glu-OH in the gas phase and solution phase. The role of the Arg side chain is to stabilize the TS1 and I1 through the interaction of its two NH's with Cyt-O2 or Cyt-N3 though the later interaction is not observed in the protein active site. Moreover, the position of the Arg side chain could be changed relatively easily depending on its outside environment. The geometries of the I1 in the gas phase and solution phase resemble each other closely. However, the solvent effects were important for calculating the PES and the TS1 structures. Furthermore, although the M06-2X functional slightly underestimates the C-S bond stability, it is better than the B3LYP functional at describing the C-S bond formation when benchmarked against the DF-MP2 and CCSD methods.

The results from the reaction step 2 showed that the complex of I1 and SAM is able to undergo methylation with or without the surrounding Cys, Glu or Arg side chains. The nucleophilic addition of the Cys-S1 to Cyt-C6 is important to increase the nucleophilicity of Cyt-C5 for this reaction step. The Glu side chain can stabilize the TS2 through hydrogen bonding between Cyt-N3 and Glu-OH or via direct protonation of Cyt-N3 by Glu-OH, but it is still uncertain whether the hydrogen bonding or the direct protonation is more important to the TS2. Energy profiles from this reaction step are probably necessary to resolve the uncertainty. The Arg side chain possibly destabilizes the TS2 through the interaction of its two NH's with Cyt-O2 due to the electron-withdrawing effects of the NH's via H-bonds. Most of the TS2 geometries in the solution phase resemble the crystal structure of M.HhaI (6MHT)[5] more closely than that in the gas phase, suggesting that the solvent effects should be included for calculating the TS2 structures.

Some disadvantages of using this small model system to study the first two reaction steps in the DNA methylation mechanism include the problems in finding the preTS1 and preTS2 that resemble the protein active site and in locating the TS for the one-step concerted reaction that combines the first two reaction steps into one. Hence, the energy profiles from these two reaction steps could not be computed and compared to confirm the roles of Cys, Glu and Arg directly. The problems encountered in the constrained optimization and one possible example to improve the small model system were also presented.

4.5 References

1. Becke, A. D., *J. Chem. Phys.* **1993,** *98* (7), 5648.

2. (a) Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008,** *41* (2), 157; (b) Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1-3), 215.

3. (a) Peräkylä, M., *J. Am. Chem. Soc.* **1998,** *120* (49), 12895; (b) Zangi, R.; Arrieta, A.; Cossío, F. P., *J. Mol. Biol.* **2010,** *400* (3), 632.

4. Du, Q.; Wang, Z.; Schramm, V. L., *Proc. Natl. Acad. Sci.* **2016,** *113* (11), 2916.

5. Kumar, S.; Horton, J. R.; Jones, G. D.; Walker, R. T.; Roberts, R. J.; Cheng, X., *Nucleic Acids Res.* **1997,** *25* (14), 2773.

6. Takeshita, K.; Suetake, I.; Yamashita, E.; Suga, M.; Narita, H.; Nakagawa, A.; Tajima, S., *Proc. Natl. Acad. Sci. U.S.A.* **2011**.

7. Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S., *Biochemistry* **2013,** *52* (16), 2828.

8. Werner, H.-J.; Manby, F. R.; Knowles, P. J., *J. Chem. Phys.* **2003,** *118* (18), 8149.

9. (a) Kendall, R. A.; Dunning, T. H.; Harrison, R. J., *J. Chem. Phys.* **1992,** *96* (9), 6796; (b) Woon, D. E.; Dunning, T. H., *J. Chem. Phys.* **1993,** *98* (2), 1358; (c) Dunning, T. H., *J. Chem. Phys.* **1989,** *90* (2), 1007.

10. 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. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

11. Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. *MOLPRO*, 2015.1; Cardiff, UK, 2015.

12. (a) Miertuš, S.; Scrocco, E.; Tomasi, J., *Chem. Phys.* **1981,** *55* (1), 117; (b) Miertuš, S.; Tomasi, J., *Chem. Phys.* **1982,** *65* (2), 239; (c) Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.* **2005,** *105* (8), 2999.

13. Weinhold, F.; Landis., C. R., *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*. Cambridge University Press: Cambridge, UK, 2005.

14. Shieh, F. K.; Reich, N. O., *J. Mol. Biol.* **2007,** *373* (5), 1157.

Chapter 5 Conformational Study on Sulfur-Containing Compounds in H_2S Releasing Reactions

5.1 Introduction

To study the H_2S releasing reactions from the reaction of GSH and various polysulfides described in section 1.3[1], a computational study was carried out on the equilibrium structures—reactants and products—in the reactions as the starting points for studying the transition states and as part of the energy profile studies. The geometric features of the reactants and products were summarized and investigated.

5.1.1 Modeling of GSH by MeSH and Assumptions on the Nucleophilic Substitutions

(a) (b)

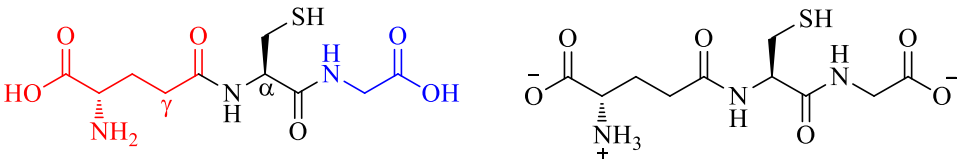


Figure 5.1 Molecular structure of GSH. (a) GSH in the fully protonated form, consisting of three amino acid residues: Glu (red), Cys (black) and Gly (blue). (b) The expected most abundant form of GSH at pH=7.4.

The structure of the nucleophile GSH is shown in Figure 5.1(a). It is a tripeptide L-γ-glutamyl-L-cysteinylglycine (γ-Glu-Cys-Gly) with a free thiol group and a special γ-peptide bond between the γ-carboxyl group of Glu (in red) and α-amino group of Cys (in black)[2].

GSH is a highly flexible molecule with nine rotatable bonds by counting the number of non-terminal single bonds excluding the peptide bonds (Figure 5.1(a)). Its conformations have been studied experimentally[2-3] and theoretically[3b, 4]. These studies showed that GSH does not adopt a strongly preferred conformation at any pH, and the experimentally determined crystal structures are not preferred in solutions so they are not suitable as starting points for computational studies. In addition, only the thiol group in GSH is known to undergo the greatest amount of chemical change during the proposed nucleophilic substitution reactions. Therefore, a simple thiol methylmercaptan (MeSH) was chosen as the model compound for GSH for all reactions studied in this chapter and Chapter 6. It has been used as the model compound for GSH in other computational studies, too[5]. The full reactant GSH was used to study the Cα nucleophilic substitution of DADS in Chapter 6 based on the transition state conformations generated from an automated conformational search approach.

At the physiological pH of 7.4[1a], the thiol group in GSH is mostly protonated as it is weakly acidic with a pKa around 9−10, the two carboxyl groups are deprotonated, and the amino group is protonated with a pKa around 8−9[2b, 4c, 6] (Figure 5.1(b)). However, the free thiol is a relatively weak nucleophile and not preferred in nucleophilic substitutions as reported earlier[5, 7]. Therefore, the thiolate anion was used as the attacking nucleophile throughout the reaction.

To perform the calculation, the nucleophilic substitution was assumed to be a bimolecular reaction (S_N2) because the thiolate anion is a strong

nucleophile and the nucleophilic attack occurs at the unhindered primary alkyl carbon, primary allyl carbon or disulfide sulfur positions. In particular, the reaction mechanism of the thiol/disulfide exchange has been studied extensively and most of the studies supported the S_N2 mechanism from the reaction of the disulfide linkage and thiolate anion[5a, 7-8]. Therefore, each nucleophilic substitution is a one-step process without any intermediates. This assumption was also used in the transition state study in Chapter 6.

5.1.2 Classification of the Reactions between MeSH and DADS/DATS

Table 5.1 Reaction steps[1] studied in H_2S release by reaction of MeSH and (a) DADS or (b) DATS, and categorization[a]

(a)

No.	Reaction of MeSH and DADS	Reaction Type
1	MeSH deprotonation	Protonation/deprotonation
2	DADS + MeS^- → MeSA + ASS^-	Cα nucleophilic substitution
3	DADS + MeS^- → MeSSA + AS^-	S nucleophilic substitution
4	DADS + AS^- → DAS + ASS^-	Cα nucleophilic substitution
5	MeSSA + MeS^- → MeSA + $MeSS^-$	Cα nucleophilic substitution
6	MeSSA + MeS^- → DMDS + AS^-	S nucleophilic substitution
7	MeSSA + AS^- → DAS + $MeSS^-$	Cα nucleophilic substitution
8	ASS^- protonation	Protonation/deprotonation
9	ASSH + MeS^- → MeSSA + HS^-	H_2S release
10	$MeSS^-$ protonation	Protonation/deprotonation
11	MeSSH + MeS^- → DMDS + HS^-	H_2S release

(b)

No.	Reaction of MeSH and DATS[a]	Reaction Type
12	DATS + MeS^- → MeSSA + ASS^-	Side-S nucleophilic substitution
13	DATS + MeS^- → MeS_3A + AS^-	Mid-S nucleophilic substitution
14	MeS_3A + MeS^- → DMDS + ASS^-	Side-S nucleophilic substitution

[a] Reaction steps that already appeared in the reaction of MeSH and DADS are omitted.

The reaction steps considered in the study from the reaction of GSH (modeled as MeSH) and DADS/DATS are adapted from Figures 1.5A, 1.6 and

1.7 in section 1.3.2 and summarized in Table 5.1. The reaction steps are classified based on reaction mechanisms (e.g. nucleophilic substitution) or roles of the reaction (e.g. H_2S release) and these terms were used throughout this chapter and Chapter 6.

The protonation/deprotonation steps, though not mentioned in the studies by Benavides *et al.*[1b] or Liang *et al.*[1a], are critical to generate the thiolate form of MeSH or to form back the deprotonated ASSH or MeSSH for reactions to occur. These steps were discussed in the energy studies in Chapter 6.

5.1.3 Classification of the Equilibrium Structures from the Reactions of MeSH and DADS/DATS and Definition of Torsional Angles

Table 5.2 List of reactants and products from the reaction of MeSH and DADS/DATS and categorization[a]

No.	Compound Type	Compounds and Reaction Step No.
A	Diallyl (poly)sulfides	DAS (4,7), DADS (2−4), DATS (12,13)
B	Allyl (poly)sulfides	AS^- (3,4,6,7,13), ASS^- (2,4,12,14), ASSH (9)
C	S-allyl-methyl (poly)sulfides	MeSA (2,5), MeSSA (3,5−7,9,12), MeS_3A (13,14)
D	Bisulfide	HS^- (9,11)
E	Methyl (poly)sulfides	MeS^- (2,3,5,6,9,11,12−14), $MeSS^-$ (5,7), MeSSH (11)
F	Dimethyl disulfide	DMDS (6,11,14)

[a] Reaction step number is in the brackets.

To understand the reactants and products better, these compounds were categorized by chemical similarities (Table 5.2).

To locate the stable structures of these compounds, three torsional angles were considered in the conformational studies. These angles are depicted in Figure 5.2 using the symmetrical molecule DATS as an example.

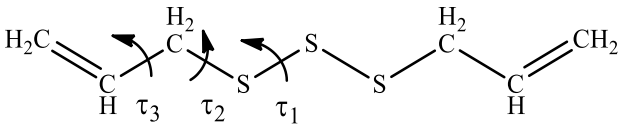


Figure 5.2 Torsional angles in DATS.

The torsional angle τ_1 describes the rotation around the disulfide bond, i.e. τ(C-S-S-X, X=H, C or S). The disulfide bonds are present in various polysulfides. τ_1 has the optimal values close to ±90^o in linear compounds[9]; therefore, compounds containing a disulfide bond have both right-handed and left-handed forms. These two forms are mirror images of each other, i.e. enantiomers. Thus, they have opposite torsional configurations but the same energy and other geometric properties. To simplify the study, only one of the enantiomeric pair was calculated explicitly. However, the geometry of its enantiomer will be presented and discussed if needed. The symbol " | " will be used to separate the torsional configuration of the enantiomers.

The torsional angle τ_2 is the rotation around the C-S bond, i.e. τ(C-C-S-X, X=C or S); while τ_3 is the rotation around the C-C bond next to the C=C bond, i.e. τ(C=C-C-S). These two torsional angles are present in the allyl-containing compounds. Several experimental and computational studies[10] suggested that the optimal τ_2 is close to ±60^o or 180^o and the optimal τ_3 is close to ±120^o.

5.1.4 Classification of the Equilibrium Structures in the Hypothetical Cα Nucleophilic Substitution of DMDS/DPDS by MeSH and Definition of Torsional Angles

Table 5.3 List of reactants and products from the Cα Nucleophilic Substitution of DMDS/DPDS by MeSH and categorization

No.	Compound Type	Compounds
A	Dialkyl (poly)sulfides	MeSMe, MeSPr, DMDS[a], DPDS
B	Alkyl (poly)sulfides	MeS^{-}[a], $MeSS^{-}$[a], $PrSS^{-}$

[a] DMDS, MeS^{-} and $MeSS^{-}$ were included in Table 5.2 and are excluded this section.

To understand the reactants and products better, these compounds were categorized by chemical similarities (Table 5.3) as described in section 5.1.3.

To locate the stable structures of these compounds, torsional angles of τ_1, τ_2 and τ_3 were considered in the conformational studies. These angles are depicted in Figure 5.3 using DPDS as an example.

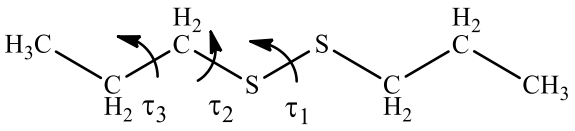


Figure 5.3 Torsional angles in DPDS.

The torsional angles of τ_1, τ_2 and τ_3 are defined similarly to that in section 5.1.3. τ_1 or τ(C-S-S-C) is the rotation around the disulfide bond, and it has the optimal values close to $\pm 90^{o}$ as presented earlier. Again, only one of the enantiomeric pair resulted from the right-handed or left-handed twist around τ_1 was calculated explicitly. However, the geometry of its enantiomer will be discussed if needed. τ_2 or τ(C-C-S-C) is the rotation around the C-S bond, and τ_3 or τ(C-C-C-S) is the rotation around the C-C bond. Several experimental

and computational studies[11] suggested that the optimal τ_2 is close to ±60° or 180°, similar to τ_2 in allyl-containing polysulfides in section 5.1.3. τ_3 is the typical torsional angle in unbranched alkanes and its optimal values are known to be close to ±60° or 180°.

5.2 Computational Methods

All equilibrium structures and transition states were fully optimized using the hybrid density functional M06-2X[12] together with the 6-31+G* basis set using the SMD[13] implicit solvent model. The addition of a diffuse function to the 6-31G* basis set provided a better description of the sulfur anions, because anions have a more diffuse lone pair. The solvent investigated is water in order to account for the aqueous reaction environment. For the fully optimized geometries, vibrational frequency analysis was performed at the same level of theory to confirm the nature of the stationary points as equilibrium structures or transition states. Unless otherwise stated, the relative energies reported correspond to relative enthalpies or Gibbs free energies at 298.13 K (ΔH_{298} or ΔG_{298}), computed at the SMD/M06-2X/6-31+G* level. The NBO[14] analysis was performed based on the SMD/M06-2X/6-31+G* wavefunction, and the atomic charges and donor-acceptor interactions were obtained directly from the NBO analysis. All calculations were performed using the *Gaussian* 09[15] suite of programs.

The accuracy of the M06-2X density functional in describing the S nucleophilic substitution reaction has been reported in a DFT benchmarking study for the thiol/disulfide exchange by Neves *et al.*[5b], in which 92 density

functionals were investigated. When benchmarked against the MP2/aug-cc-pVTZ optimized geometry in vacuum, M06-2X showed good performance at describing the optimized geometries with the 6-31G* and 6-31+G* basis sets. When benchmarked against the reaction energy and activation barrier computed at CCSD(T)/CBS//MP2/aug-cc-pVDZ level of theory, M06-2X also performed well in computing these energies from single point energy calculations with 10 Pople's basis sets (e.g. 6-311+G(2d,2p)), based on the M06-2X/6-31G* optimized geometry.

5.3 Results & Discussion

5.3.1 Conformational Analysis of Reactants and Products in the Reaction of MeSH and DADS/DATS

The systematic conformational search for all compounds in Table 5.2 was carried out manually by combining the different optimal angles of τ_1, τ_2 and τ_3 and optimizing each corresponding conformation. Symbols are used to describe a specific conformation of a compound. For τ_1, the optimal angles of 90^o and -90^o are denoted by "+" and "–". For τ_2, the optimal angles of 60^o, -60^o and 180^o are denoted by "G(+)", "G(–)" and "anti". For τ_3, the optimal angles of 120^o and -120^o are denoted by "G(+)" and "G(–)". For instance, by considering the (τ_3, τ_2, τ_1, τ_1, τ_2, τ_3) combination for DATS (Figure 5.2), our conformational search generated 42 distinct conformers as shown in Table S 8.5 in Appendix with their numbering and the corresponding conformations. The conformers were named according to the ranking in ΔH_{298}. Leaving out the three sterically unfavorable conformers (DATS40−42), 39 conformers

were fully optimized at the SMD/M06-2X/6-31+G* level of theory and are reported in Table S 8.5(c). Table 5.4 shows the top 16 conformers of DATS with either ΔH_{298} or $\Delta G_{298} \leq 5$ kJ mol^{-1}. Out of these 39 conformers, DATS1, denoted by G(+)G(–)++G(–)G(+), is the lowest energy conformer according to ΔH_{298} and ΔG_{298}. This conformer was chosen as the starting point to construct the relevant transition states in section 6.3.1. The similar process was repeated for other compounds in Table 5.2 and the results are summarized in Table S 8.5.

Table 5.4 Top 16 of the calculated DATS conformations and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

No.	Conformation	ΔH_{298}	ΔG_{298}
1	G(+)G(–)++G(–)G(+)	0.00	0.00
2	G(+)G(–)++G(+)G(–)	1.73	0.51
3	G(–)G(+)++G(+)G(–)	2.04	2.39
4	G(+)G(–)–+G(+)G(–)	2.12	4.00
5	G(–)G(+)–+G(+)G(–)	4.13	6.15
6	G(+)G(+)+–G(–)G(+)	4.83	8.15
7	G(+)G(+)+–G(+)G(–)	5.41	8.43
8	G(–)G(–)++G(–)G(+)	6.55	6.69
9	G(+)G(+)++G(–)G(+)	6.64	7.42
10	G(+)G(+)++G(+)G(–)	7.15	5.07
11	G(+)anti—G(–)G(+)	7.32	1.81
12	G(+)anti—G(+)G(–)	7.33	5.93
13	G(+)anti+–G(–)G(+)	7.38	8.87
14	G(+)anti++G(–)G(+)	7.67	7.23
15	G(–)G(–)++G(+)G(–)	7.70	3.56
16	G(+)anti++G(+)G(–)	7.95	4.45

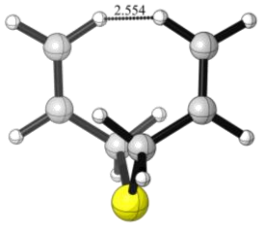
In general, the ranking in ΔH_{298} and in ΔG_{298} for all compounds agree well with each other (Table 5.4 and Table S 8.5(a)–(h)). However, the ranking in ΔH_{298} and in ΔG_{298} deviate the most for the low-lying conformers of DATS (ΔH_{298} and $\Delta G_{298} \leq 9$ kJ mol^{-1}). Because DATS is the most complex

compound studied in the series, it has a lot more low-lying conformers with nearly identical values in ΔH_{298}, such that a small difference in S may easily alter the ranking in ΔG_{298}. For MeSSA and MeS_3A (Table S 8.5(g)−(h)), the ranking in ΔH_{298} and in ΔG_{298} resulted in two different lowest energy conformers with the energy difference smaller than 1 kJ mol^{-1}, thus both conformers were chosen for the transition state study later. Labels of (H) or (G) were added to distinguish these two lowest energy conformers.

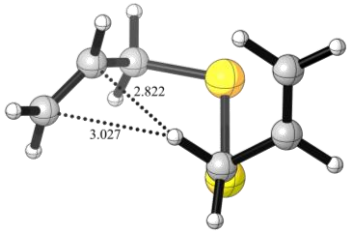
The systems studied in Chapters 5 and 6 almost all contain relatively flexible molecules and the calculations of them resulted in several low frequency vibrational modes that can easily change the entropy corrections to the Gibbs free energies. Therefore, most of the results and discussions will be based on enthalpy corrected energies instead of free energies in this thesis later. Free energies can be found in the Appendix.

The geometries of the lowest energy conformers for all compounds are summarized in Figure 5.4 and Table S 8.6.

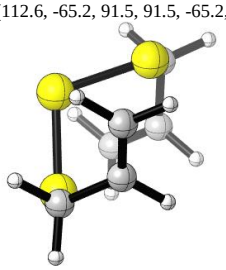
A1) DAS1
G(+)G(–)G(–)G(+):
(117.4, -66.9, -66.9, 117.5)



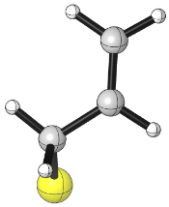
A2) DADS1
G(+)G(–)–G(+)G(–):
(113.1, -66.8, -92.3, 63.6, -113.4)



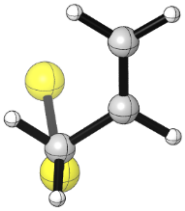
A3) DATS1
G(+)G(–)++G(–)G(+):
(112.6, -65.2, 91.5, 91.5, -65.2, 112.6)



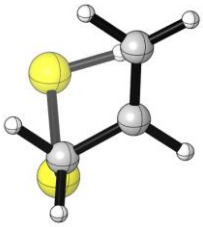
B1) AS^-
G(+):
(115.7)



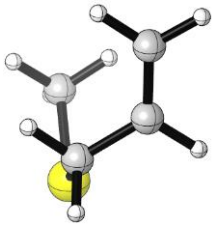
B2) ASS^-1
G(+)G(–):
(114.4, -70.8)



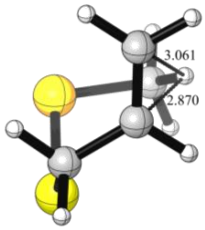
B3) ASSH1
G(+)G(–)+:
(109.7, -67.9, 79.3)



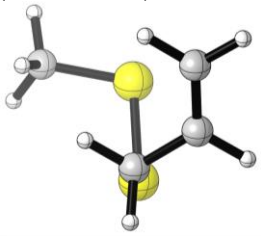
C) MeSA1
G(+)G(–):
(108.6, -66.1)



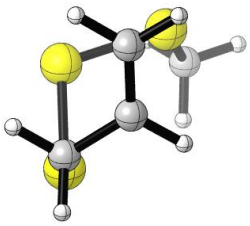
C2-1) MeSSA1(H)
G(+)G(–)+:
(113.1, -63.4, 93.4)



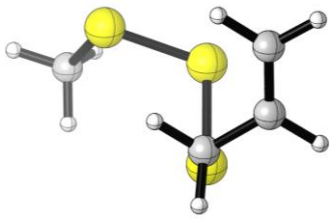
C2-2) MeSSA2(G)
G(+)G(–)–:
(112.1, -66.5, -80.0)



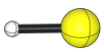
C3-1) $MeS_3A1(H)$
G(+)G(–)++:
(110.7, -63.1, 90.3, 86.5)



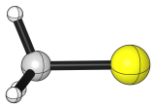
C3-2) $MeS_3A2(G)$
G(+)G(–)—:
(111.4, -63.9, -78.0, -80.3)



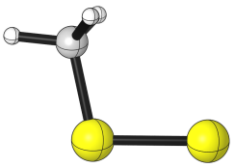
D) HS^-



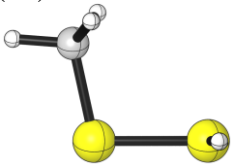
E1) MeS^-



E2) $MeSS^-$



E3) MeSSH
G(+):
(84.7)



F) DMDS
G(+):
(81.4)

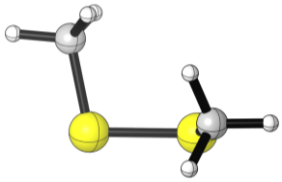
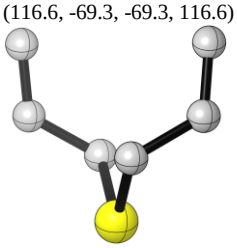


Figure 5.4 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers for: A1)−3) diallyl (poly)sulfides; B1)−3) diallyl (poly)sulfides; C1)−3-2) S-allyl-methyl (poly)sulfides; D) bisulfide; E1)−3) methyl (poly)sulfides; F) dimethyl disulfide. Interactions are indicated by the dotted lines and distances are in Å. Torsional angles (in °) are given in the brackets. The conformations shown in C2-2) and C3-2) correspond to the respective enantiomers of the optimized conformers for comparison purpose.

(a) DAS-crystal
G(+)G(–)G(–)G(+):
(116.6, -69.3, -69.3, 116.6)



(b) DADS-crystal
G(+)G(–)–G(+)G(–):
(114.6, -57.2, -89.1, 63.6, -112.4)

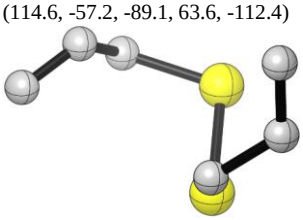


Figure 5.5 Crystal structures[10d] of (a) DAS and (b) DADS. Torsional angles (in °) are given in the brackets.

Table 5.5 Geometric information on the crystal structures[10d] of DAS and DADS[a]

No.	(a)	(b)
Name	DAS-crystal	DADS-crystal
Conformation	G(+)G(–)G(–)G(+)	G(+)G(–)–G(+)G(–)
Bond Lengths		
C=C	1.312	1.313
C-C	1.490	1.474
C-S	1.814	1.838
S-S	NA[b]	2.030
S-C	1.815	1.830
C-C	1.490	1.483
C=C	1.312	1.294
Bond Angles		
C=C-C	124.9	123.8
C-C-S	112.1	112.8
C-S-C/S	100.6	105.2
S-S-C	NA	104.4
S-C-C	112.1	113.0
C-C=C	124.9	124.2

[a] Bond lengths are in Å, and bond angles are in °.
[b] NA, not applicable.

All of the optimized conformers for reactants and products were manually inspected and compared to their input structures. The majority of the optimized geometries match the corresponding conformations well (not all results are shown). In general, the values of τ_3 deviate the least from the optimal values (usually < 10°), followed by τ_2 (denoted by G(+) or G(–)). The values of τ_1 can easily deviate from the optimal values by up to ~20°, followed by the values of τ_2 (denoted by anti). However, the change in τ_2 (denoted by G(+) or G(–)) can be as large as ~50° in some high energy conformers. For example, DADS37 has the conformation of G(+)G(+)+–G(+)G(+) (Table S 8.5(c)). However, its optimized geometry (Figure 5.6) shows that the one of the τ_2 (red) deviates from the optimal value of 60° by +48° possibly to avoid the steric repulsion between the two circled hydrogens.

DAS37
G(+)G(+)+–G(+)G(+)
(115.0, 108.0, -84.5, 91.3, 69.9, 115.6)

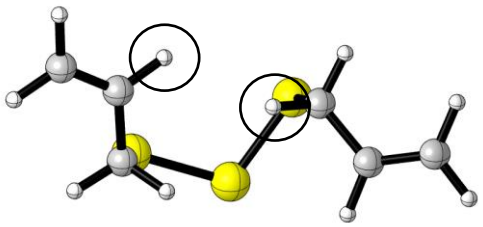


Figure 5.6 Optimized geometry of DADS37. Torsional angles (in °) are given in the brackets.

(a) G(+)G(–)+ (b) G(–)G(+)– (c) G(+)G(–)– (d) G(–)G(+)+

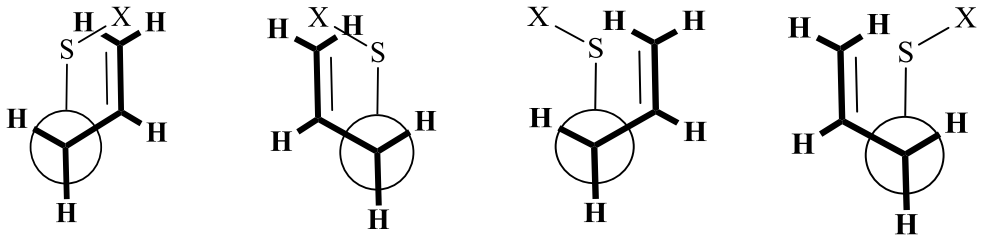


Figure 5.7 The lowest energy conformations of the -SS-allyl group.

For diallyl (poly)sulfides, including DAS, DADS and DATS, two allyl groups (-CH_2-CH=CH_2) are joined by one, two or three sulfur atoms. The SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers—DAS1, DADS1 and DATS1—are shown in Figure 5.4A1)−3) and Table S 8.6(a). The geometries of DAS1 and DADS1 are highly similar to their reported crystal structures[10d] (Figure 5.5 and Table 5.5). The geometry of DAS1 agrees well with that of DAS-crystal: the deviations in bond lengths, bond angles and torsional angles are 0.40%−1.68%, -0.96%−0.18% and -3.46%−0.72% respectively. The geometry of DAS1 mostly matches that of DADS-crystal: the deviations in bond lengths, bond angles and torsional angles are 0.27%−3.09%, -1.71%−-0.12%, and -1.29%−3.61% respectively. The only outlier is τ_2 denoted by G(−) with a deviation of -16.7% and it is probably due to potential distortions in crystal packing. In addition, the preference for the G(+)G(–)–G(+)G(–) conformation being the most stable form of DADS is consistent with a study by Suzuki *et al.*[10a] experimentally and computationally. Moreover, the geometry of DATS1 is close to that in the reported crystal structure of the 2CuCl·DATS complex[10b]. In these conformers, the -S-allyl groups all adopt the G(+)G(–)|G(–)G(+) conformation with (τ_3, τ_2) ≈ (±120^o, ±300^o) (Figure 5.7(a)–(d)). With at least one disulfide bond in DADS or DATS, the -SS-allyl groups may adopt the G(+)G(–)+|G(–)G(+)– conformation with (τ_3, τ_2, τ_1) ≈ (±120^o, ±300^o, ±90^o) (Figure 5.7(a)–(b)), or the G(+)G(–)–|G(–)G(+)+ conformation with (τ_3, τ_2, τ_1) ≈ (±120^o, ±300^o, ±270^o) (Figure 5.7(c)–(d)). With two adjacent disulfide bonds in DATS, the trisulfide linkage (-S-S-S-) adopts the ++|— conformation with (τ_1, τ_1) ≈ (±90^o, ±90^o)

possibly to avoid the steric repulsion between the two ending groups of the molecule in the +– conformation (Figure 5.6). In addition, DAS1 has one set of close dihydrogen contact (2.554 Å) as the two ending -CH=CH_2 groups are close and almost parallel to each other. DADS1 has one allyl-hydrogen in close proximity to the two sp^2 carbons of the other allyl group (2.822 Å and 3.027 Å) that suggests the possible C-H···π interaction.

For S-allyl (poly)sulfides, such as the deprotonated ASH and ASSH and the protonated ASSH, one allyl group is linked to one or two sulfur atoms and such compounds have a sulfur-containing ionizable group such as a thiol group (-SH) or a perthiol group (-SSH). The SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers—AS^-, ASS^-1 and ASSH1—are shown in Figure 5.4B1)−3) and Table S 8.6(b). The -S-allyl groups in the deprotonated and protonated ASSH adopt the identical conformations to that in diallyl (poly)sulfides, but the -SS-allyl group in the protonated ASSH only adopts the G(+)G(–)+|G(–)G(+)– conformation.

For S-allyl-methyl (poly)sulfides, which are MeSA, MeSSA and MeS_3A, one allyl group and one methyl group are joined by one, two or three sulfur atoms. The SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers or the corresponding enantiomers—MeSA1, MeSSA1(H), MeSSA2(G), MeS_3A1(H) and MeS_3A2(G)—are shown in Figure 5.4C1)−3-2) and Table S 8.6(c). The preference for the G(+)G(–)+|G(–)G(+)– conformation over the G(+)G(–)–|G(–)G(+)+ conformation in the most stable form of MeSSA by considering ΔE (difference is 0.86 kJ mol^{-1}, Figure 5.8(a)−(b)) or ΔH_{298} agrees well with the conformational study by Lin *et al*[10c].

However, the difference in ΔE in their study and our study cannot be compared because their results were obtained at different levels of theory, i.e. HF/6-31G* or B3LYP/6-31G* in the gas phase. The -S-allyl groups in all these conformers adopt the same conformation as that in diallyl (poly)sulfides and S-allyl (poly)sulfides. The -SS-allyl groups in MeSSA1(H) and MeS_3A1(H) adopt the G(+)G(–)+|G(–)G(+)– conformation, while such groups in MeSSA2(G) and MeS_3A2(G) adopt the G(+)G(–)–|G(–)G(+)+ conformation. Similar to DATS, the trisulfide linkages (-S-S-S-) in MeS_3A1(H) and MeS_3A2(G) all adopt the ++|— conformation. Moreover, MeSSA1(H) has one methyl-hydrogen close to the two sp^2 carbons of the allyl group (2.870 Å and 3.061 Å) that possibly implies the C-H···π interaction.

For bisulfide, methyl (poly)sulfides and dimethyl disulfide, the SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers—HS^-, MeS^-, $MeSS^-$, MeSSH, DMDS—are shown in Figure 5.4D)−F) and Table S 8.6(d). For methyl (poly)sulfides, one methyl group is bound to one or two sulfur atoms and such compounds have the ionizable -SH group or -SSH group. The deprotonated H_2S, MeSH and MeSSH all have only one most stable conformer due to the lack of torsional angles. The protonated MeSSH and DMDS can adopt the +|– conformation around the disulfide bond. The geometries of DMDS is similar to the gas-phase structure determined by electron diffraction[16] and the enantiomer of the reported crystal structure[17]. The deviations in bond lengths, bond angles and torsional angles from the gas-phase structure are -1.18%−1.77%, -1.07%−-0.54% and -4.57% respectively;

while the deviations from the enantiomer of the crystal structure are 0.75%−1.60%, -1.73% and -5.35% respectively.

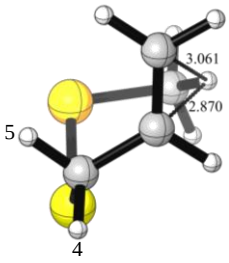
To summarize, the lowest energy conformers of the allyl-containing compounds, including diallyl (poly)sulfides, allyl (poly)sulfides and S-allyl-methyl (poly)sulfides, share some similar structural features. First, the -S-allyl groups all adopt the G(+)G(–)|G(–)G(+) conformation in the most stable conformers. From the conformational analysis of ASS^- and MeSA (Table S 8.5(d) and (f)), the preferred conformation of the -S-allyl group has the order of G(+)G(–)|G(–)G(+) > G(+)G(+)|G(–)G(–) > G(+)anti|G(–)anti based on ΔH_{298}, whereas it has the order of G(+)G(–)|G(–)G(+) > > G(+)anti|G(–)anti > G(+)G(+)|G(–)G(–) based on ΔG_{298}. The similar orders in the preferred conformation of the -S-allyl group are also observed in other compounds in Table S 8.5(a)−(c), (e), and (g)–(h)). Second, the -SS-allyl groups may adopt the G(+)G(–)+|G(–)G(+)– or the G(+)G(–)–|G(–)G(+)+ conformation in the most stable conformers (Table S 8.5(b)–(c), (e) and (g)–(h)). The G(+)G(–)–|G(–)G(+)+ conformation of the -SS-allyl group is preferred over the G(+)G(–)+|G(–)G(+)– conformation based on ΔG_{298} when the substituent groups on the disulfide are bulkier in size such as the methyl (in MeSSA) and the S-methyl (in MeS_3A). In general, the conformation of the -SS-allyl group has the order of G(+)G(–)+|G(–)G(+)–/G(+)G(–)–|G(–)G(+)+ > G(+)G(+)+|G(–)G(–)–/G(+)G(+)–|G(–)G(–)+/G(+)anti+|G(–)anti–/G(+)anti–|G(–)anti+ based on ΔH_{298} and ΔG_{298}. Third, the trisulfide linkages (-S-S-S-) all adopt the ++|— conformation rather than the +– conformation in the most stable conformers,

as demonstrated from the conformational analysis of DATS (Table S 8.5(c)) and MeS_3A (Table S 8.5(h)).

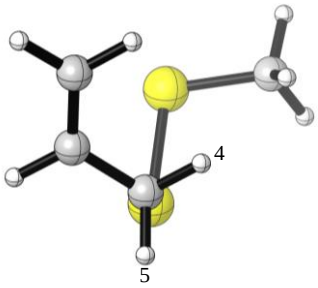
From Figure 5.4 and the discussions above, one can see that the preferred conformations of the -S(S)-allyl groups in the lowest energy conformers studied are not resulted from the presence of highly favorable intramolecular interactions such as strong hydrogen bonds (H-bonds) or ionic bonds. Therefore, the NBO analysis was performed on MeSSA to study the resonance stabilization by hyperconjugation or negative hyperconjugation that may contribute to the optimal values of τ_3 (~$\pm120^o$), τ_2 (~$\pm60^o$ or ~180^o), τ_1 (~$\pm90^o$) and even their combinations in the -S(S)-allyl groups. Hyperconjugation is the resonance stabilization resulted from the interaction of a σ bonding orbital to an adjacent empty or half-filled non-bonding orbital, σ or π antibonding orbital; whereas negative hyperconjugation is the resonance stabilization resulted from the interaction of a π bonding orbital or a nonbonding orbital to an adjacent σ antibonding orbital. To simplify our discussion, "hyperconjugation(s)" or "hyperconjugative interaction(s)" will be used to describe both the hyperconjugation and negative hyperconjugation in this thesis.

The SMD/M06-2X/6-31+G* optimized conformers of MeSSA with the relative electronic energies (ΔE) are shown in Figure 5.8(a)−(f) and Table S 8.7. The two α-H's of MeSSA are also numbered in Figure 5.8(a)−(f). The chemical structure of MeSSA with atom numbering is depicted in Figure 5.8(g).

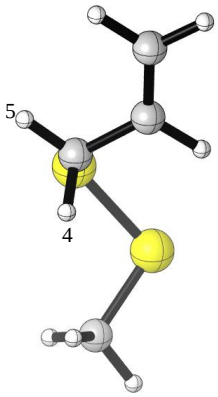
(a) MeSSA1 (H)
G(+)G(–)+: (113.1, -63.4, 93.4)
ΔE: 0.00 kJ mol^{-1}



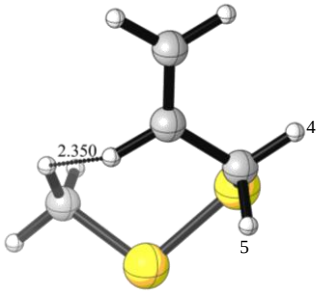
(b) MeSSA2 (G)
G(–)G(+)+: (-112.1, 66.5, 80.0)
ΔE: 0.86 kJ mol^{-1}



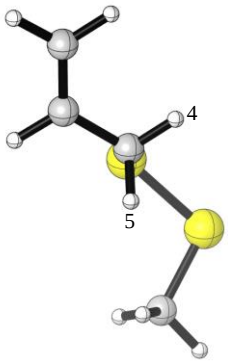
(c) MeSSA3
G(+)G(+)+: (116.3, 71.8, 84.0)
ΔE: 5.49 kJ mol^{-1}



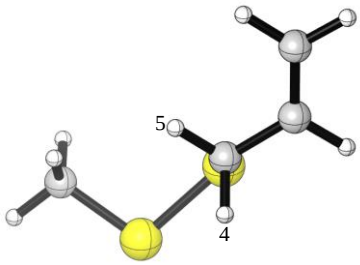
(d) MeSSA5
G(–)G(–)+: (-112.3, -67.3, 94.2)
ΔE: 5.97 kJ mol^{-1}



(e) MeSSA4
G(–)anti+: (-109.4, -167.4, 83.7)
ΔE: 6.03 kJ mol^{-1}



(f) MeSSA6
G(+)anti+: (107.8, 168.5, 81.1)
ΔE: 6.16 kJ mol^{-1}



(g)

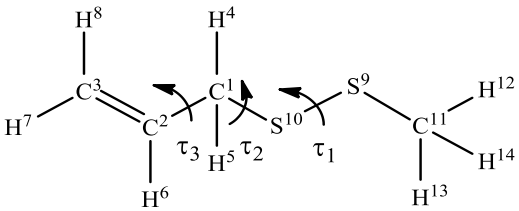


Figure 5.8 Structures of MeSSA. (a)−(f) SMD/M06-2X/6-31+G* optimized conformers of MeSSA and their relative energies (ΔE). (g) Chemical structure of MeSSA with labeling of atoms and torsional angles. Interactions are indicated by the dotted lines and distance are in Å. Torsional angles (in ^{o}) are given in the brackets.

Table 5.6 Donor-Acceptor interactions in the optimized conformers of MeSSA from the NBO analysis[a]

(a)

Name				MeSSA1 G(+)G(–)+		MeSSA2 G(–)G(+)+	
τ_3				~120°		~-120°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
6	$\pi(C^2=C^3)$	128	$\sigma^*(C^1-S^{10})$	NA[b]	27.82	NA	27.78
4	$\sigma(C^1-S^{10})$	130	$\pi^*(C^2=C^3)$	NA	19.46	NA	19.25
3/2	$\sigma(C^1-H^n)$	131	$\sigma^*(C^2-H^6)$	5	16.15	4	15.82
2/3	$\sigma(C^1-H^n)$	129	$\sigma^*(C^2=C^3)$	4	11.92	5	12.13
7	$\sigma(C^2-H^6)$	127/126	$\sigma^*(C^1-H^n)$	5	10.59	4	10.63
2/3	$\sigma(C^1-H^n)$	130	$\pi^*(C^2=C^3)$	4	9.67	5	9.46
6	$\pi(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	8.24	5	7.95
5	$\sigma(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	3.81	5	3.72
4	$\sigma(C^1-S^{10})$	129	$\sigma^*(C^2=C^3)$	NA	3.51	NA	3.39
Total Energies					111.17		110.12
τ_2				~-60°		~60°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
32	$LP(2)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	17.61	NA	16.07
32	$LP(2)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	5	10.33	4	11.67
2/3	$\sigma(C^1-H^n)$	134	$\sigma^*(S^9-S^{10})$	4	10.67	5	10.54
10	$\sigma(S^9-S^{10})$	126/127	$\sigma^*(C^1-H^n)$	4	4.73	4	4.85
31	$LP(1)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	2.76	NA	3.77
31	$LP(1)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	5	4.02	4	3.10
Total Energies					50.12		50.00
τ_1				~90°		~90°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
30	$LP(2)(S^9)$	128	$\sigma^*(S^{10}-C^1)$	NA	28.79	NA	29.29
32	$LP(2)(S^{10})$	135	$\sigma^*(S^9-C^{11})$	NA	25.02	NA	23.85
Total Energies					53.81		53.14
Sum of Total Energies					215.10		213.26
Relative Sum of Total Energies					51.71		49.87

(b)

Name				MeSSA3 G(+)G(+)+		MeSSA5 G(–)G(–)+	
τ_3				~120°		~-120°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
6	$\pi(C^2=C^3)$	128	$\sigma^*(C^1-S^{10})$	NA	23.30	NA	25.82
4	$\sigma(C^1-S^{10})$	130	$\pi^*(C^2=C^3)$	NA	17.57	NA	17.66

3/2	$\sigma(C^1\text{-}H^n)$	131	$\sigma^*(C^2\text{-}H^6)$	5	16.61	4	16.48
2/3	$\sigma(C^1\text{-}H^n)$	129	$\sigma^*(C^2\text{=}C^3)$	4	10.88	5	11.84
7	$\sigma(C^2\text{-}H^6)$	127/126	$\sigma^*(C^1\text{-}H^n)$	5	10.54	4	10.63
2/3	$\sigma(C^1\text{-}H^n)$	130	$\pi^*(C^2\text{=}C^3)$	4	10.38	5	9.92
6	$\pi(C^2\text{=}C^3)$	126/127	$\sigma^*(C^1\text{-}H^n)$	4	9.33	5	8.49
5	$\sigma(C^2\text{=}C^3)$	126/127	$\sigma^*(C^1\text{-}H^n)$	4	3.64	5	3.64
4	$\sigma(C^1\text{-}S^{10})$	129	$\sigma^*(C^2\text{=}C^3)$	NA	4.31	NA	3.77
Total Energies					106.57		108.24
τ_2				~60°		~-60°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
32	$LP(2)(S^{10})$	125	$\sigma^*(C^1\text{-}C^2)$	NA	17.49	NA	19.20
32	$LP(2)(S^{10})$	127/126	$\sigma^*(C^1\text{-}H^n)$	4	9.62	5	8.87
2/3	$\sigma(C^1\text{-}H^n)$	134	$\sigma^*(S^9\text{-}S^{10})$	5	9.54	4	10.33
10	$\sigma(S^9\text{-}S^{10})$	126/127	$\sigma^*(C^1\text{-}H^n)$	5	5.23	4	4.85
31	$LP(1)(S^{10})$	125	$\sigma^*(C^1\text{-}C^2)$	NA	3.26	NA	2.64
31	$LP(1)(S^{10})$	127/126	$\sigma^*(C^1\text{-}H^n)$	4	3.26	5	4.10
Total Energies					48.41		50.00
τ_1				~90°		~90°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
30	$LP(2)(S^9)$	128	$\sigma^*(S^{10}\text{-}C^1)$	NA	27.78	NA	28.37
32	$LP(2)(S^{10})$	135	$\sigma^*(S^9\text{-}C^{11})$	NA	23.97	NA	25.61
Total Energies					51.76		53.97
Sum of Total Energies					206.73		212.21
Relative Sum of Total Energies					43.34		48.82

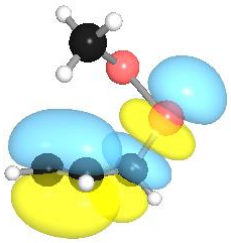
(c)

Name				MeSSA4 G(–)anti+		MeSSA6 G(+)anti+	
τ_3				~-120°		~120°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
6	$\pi(C^2\text{=}C^3)$	128	$\sigma^*(C^1\text{-}S^{10})$	NA	28.41	NA	28.37
4	$\sigma(C^1\text{-}S^{10})$	130	$\pi^*(C^2\text{=}C^3)$	NA	23.64	NA	23.60
2/3	$\sigma(C^1\text{-}H^n)$	131	$\sigma^*(C^2\text{-}H^6)$	4	14.39	5	14.31
3/2	$\sigma(C^1\text{-}H^n)$	129	$\sigma^*(C^2\text{=}C^3)$	5	12.64	4	12.64
7	$\sigma(C^2\text{-}H^6)$	126/127	$\sigma^*(C^1\text{-}H^n)$	4	9.96	5	9.92
3/2	$\sigma(C1\text{-}H^n)$	130	$\pi^*(C^2\text{=}C^3)$	5	6.90	4	6.95
6	$\pi(C^2\text{=}C^3)$	127/126	$\sigma^*(C^1\text{-}H^n)$	5	5.94	4	5.94
5	$\sigma(C^2\text{=}C^3)$	127/126	$\sigma^*(C^1\text{-}H^n)$	5	3.56	4	3.51
4	$\sigma(C^1\text{-}S^{10})$	129	$\sigma^*(C^2\text{=}C^3)$	NA	2.80	NA	2.80
Total Energies					108.24		108.03

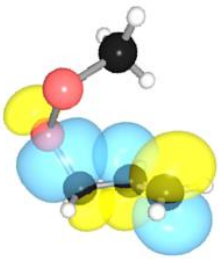
τ_2				~180°		~180°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
32	LP(2)(S^{10})	126/127	σ*(C^1-H^n)	4	9.16	5	10.54
32	LP(2)(S^{10})	127/126	σ*(C^1-H^n)	5	7.91	4	6.44
10	σ(S^9-S^{10})	125	σ*(C^1-C^2)	NA	5.15	NA	5.06
1	σ(C^1-C^2)	134	σ*(S^9-S^{10})	NA	4.18	NA	4.18
Total Energies					26.40		26.23
τ_1				~90°		~90°	
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	n	E(2)
30	LP(2)(S^9)	128	σ*(S^{10}-C^1)	NA	15.40	NA	15.36
32	LP(2)(S^{10})	135	σ*(S^9-C^{11})	NA	13.72	NA	13.77
Total Energies					29.12		29.12
Sum of Total Energies					163.76		163.39
Relative Sum of Total Energies					0.40		0.00

[a] E(2) is the two-electron stabilizing energy (in kJ mol^{-1}).
[b] NA, not applicable.

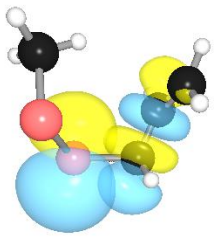
A1) MeSSA1, τ_3~120°
π(C^2=C^3)→σ*(C^1-S^{10})



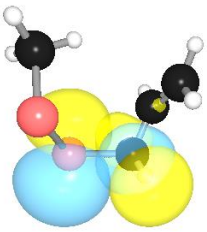
A2) MeSSA1, τ_3~120°
σ(C^1-S^{10})→π*(C^2=C^3)



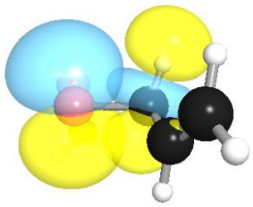
B1) MeSSA1, τ_2~-60°
LP(2)(S^{10})→σ*(C^1-C^2)



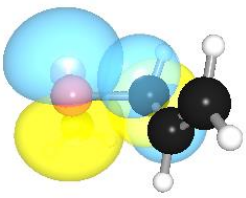
B2) MeSSA1, τ_2~-60°
LP(2)(S^{10})→σ*(C^1-$H^{4/5}$)



C1) MeSSA4, τ_2~180°
LP(2)(S^{10})→σ*(C^1-H^5)



C2) MeSSA4, τ_2~180°
LP(2)(S^{10})→σ*(C^1-H^4)



D) MeSSA1, τ_1~90°
LP(2)(S^9)→σ*(S^{10}-C^1)

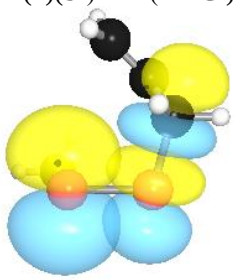


Figure 5.9 Visualization of NBO donor-acceptor interactions for: A1)−2) τ_3~120°, B1)−2) τ_2~-60°, C1)−2) τ_2~180°, D) τ_1~90°.

As shown in Figure 5.8(a)−(f), the conformation of the -SS-allyl groups in the SMD/M06-2X/6-31+G* optimized MeSSA has the order of G(+)G(–)+|G(–)G(+)– > G(+)G(–)–|G(–)G(+)+ > G(+)G(+)+|G(–)G(–)– > G(+)G(+)–|G(–)G(–)+ > G(+)anti–|G(–)anti+ > G(+)anti+|G(–)anti– based on ΔE. This order is slightly different from the order based on ΔH_{298} or ΔG_{298} (Table S 8.5(f)). It is also different from the order reported by Lin *et al.*[10c] for the last four conformations optimized at the HF/6-31G* or B3LYP/6-31G* level in the gas phase. Their order is G(+)G(–)+|G(–)G(+)– > G(+)G(–)–|G(–)G(+)+ > G(+)anti+|G(–)anti– > G(+)G(+)+|G(–)G(–)– > G(+)G(+)–|G(–)G(–)+ and the ΔE values from the HF and B3LYP methods are comparable. Because the optimized geometries in our study are similar to that in their study, the difference in the order is possibly due to the solvent effects in ΔE calculation, considering the ΔE difference for the last four conformations is very small (within 0.7 kJ mol^{-1}). More importantly, the six conformations considered in our study were identified as the six low-lying conformations of MeSSA, and such results demonstrated that our simplified conformational analysis is able to generate all the low-lying energy conformers for MeSSA. In addition,

MeSSA5 has one set of close dihydrogen contact (2.350 Å) between the allyl-hydrogen and methyl-hydrogen behind it.

Several hyperconjugative interactions that may attribute to the optimal values of τ_1, τ_2 and τ_3 were identified from the NBO analysis (Table 5.6 and Table S 8.8). Some of the donor-acceptor interactions are pictured in Figure 5.9. To illustrate the torsional angles more clearly, the optimal τ_3, τ_2, and τ_1 are depicted in Figure 5.10 by Newman projection.

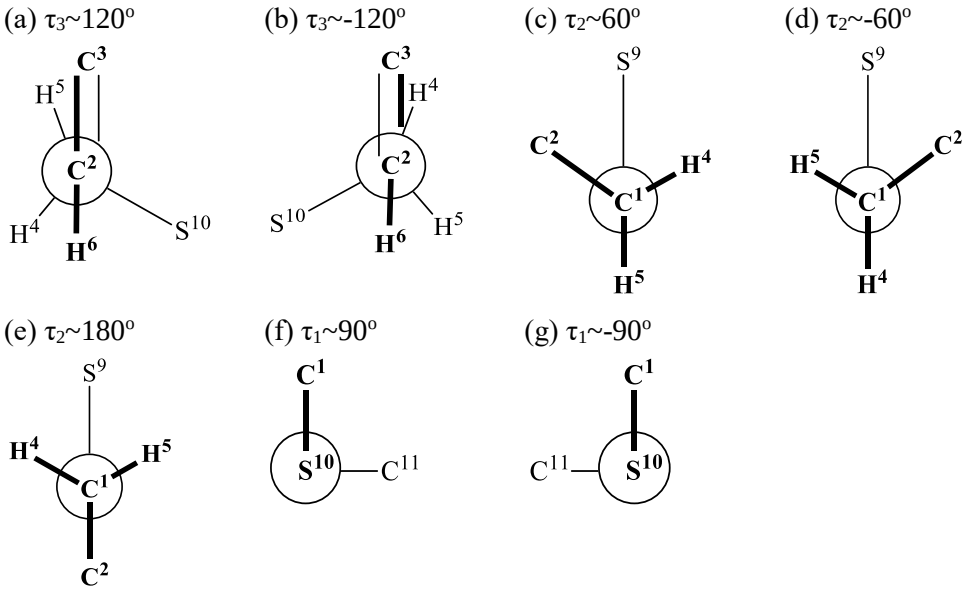


Figure 5.10 The stable torsional angles presented by Newman projection. (a)−(b) τ_3 close to 120° or -120°. (c)−(d) τ_2 close to 60° or -60°. (e) τ_2 close to 180°. (f)−(g) τ_1 close to 90° or -90°.

When τ_3 is close to ±120° (denoted by G(+)|G(−)) (Figure 5.10(a)−(b)), the C^2=C^3 bond is anticlinal (~±120°) to the C^1-$H^{4/5}$ and C^1-S^{10} bonds, while the C^2-H^6 bond is synclinal or gauche (~±60°) to these two bonds. Meanwhile, the C^2=C^3 bond is eclipsed with the C^1-$H^{5/4}$ bond, and the C^2-H^6 bond is anti to this bond. Thus, hyperconjugations may occur between the σ or π bonding orbital of the C^2=C^3 bond and the antibonding orbital of the C^1-$H^{4/5}$ or C^1-S^{10}

bond, and vice versa. Similarly, hyperconjugations may occur between the bonding orbital of the C^2-C^6 bond and the antibonding orbital of the C^1-$H^{5/4}$ bond, and vice versa. The NBO analysis of the donor-acceptor interactions (Table 5.6(a)−(c) and Table S 8.8(a)−(c)) confirms these orbital interactions. The two-electron stabilizing energy E(2) is the measure of the resonance stabilization due to the donor-acceptor interactions. This energy depends on the energy gap between the donor and acceptor orbitals, i.e. E(j)-E(i), as well as the level of orbital overlap F(i,j). Out of the nine donor-acceptor interactions, $\pi(C^2=C^3)\rightarrow\sigma^*(C^1\text{-}S^{10})$, $\sigma(C^1\text{-}S^{10})\rightarrow\pi^*(C^2=C^3)$, $\sigma(C^1\text{-}H^{5/4})\rightarrow\sigma^*(C^2\text{-}H^6)$, $\sigma(C^1\text{-}H^{5/4})\rightarrow\sigma^*(C^2=C^3)$ and $\sigma(C^2\text{-}H^6)\rightarrow\sigma^*(C^1\text{-}H^{5/4})$ are the more important stabilizing interactions that contribute to 73−82% of the total stabilizing energies around τ_3 (total: 106.57–111.17 kJ mol^{-1}). These interactions are presented in the descending order with stabilizing energies larger than or close to 10 kJ mol^{-1}. The donation from $\pi(C^2=C^3)$ to the adjacent $\sigma^*(C^1\text{-}S^{10})$ and the back donation from $\sigma(C^1\text{-}S^{10})$ to $\pi^*(C^2=C^3)$ are the two most important interactions caused by the small energy gap between the two orbitals (< 0.8 Hartree) and the favorable orbital overlap at an angle of ~30^o (Figure 5.9A1)−2)). The hyperconjugation between the C=C and C-S bonds has been reported in a spectroscopic study on DADS by Koch[10e]. The energy gaps between the orbitals of the C^1-$H^{5/4}$ bond and that of the C^2-H^6 bond are larger, but the orbital overlap is the largest due to the almost eclipsed bonding-antibonding orbital alignment, thus leading to the relatively large stabilizing interactions. The orbital overlap between $\sigma(C^1\text{-}S^{4/5})$ and $\sigma^*(C^2=C^3)$ is similar to that between $\pi(C^2=C^3)$ and $\sigma^*(C^1\text{-}S^{10})$ at an angle of ~30^o, but the energy

gap between these two orbitals are increased, so the interaction energy of $\sigma(C^1\text{-}S^{4/5})\rightarrow\sigma^*(C^2\text{=}C^3)$ is smaller than that of $\sigma(C^1\text{-}H^{5/4})\rightarrow\sigma^*(C^2\text{-}H^6)$ but larger than that of $\sigma(C^2\text{-}H^6)\rightarrow\sigma^*(C^1\text{-}H^{5/4})$.

When τ_2 is close to ±60° (denoted by G(+)|G(−)) (Figure 5.10(c)−(d)), the $S^9\text{-}S^{10}$ bond is anti to the $C^1\text{-}H^{5/4}$ bond, and it is gauche (~±60°) to the $C^1\text{-}C^2$ and $C^1\text{-}H^{4/5}$ bonds. Therefore, hyperconjugations may occur between the bonding orbital of the $S^9\text{-}S^{10}$ bond and the antibonding orbital of the $C^1\text{-}H^{5/4}$ bond, and vice versa. Moreover, hyperconjugations may occur between the either one of the two nonbonding orbitals or lone pairs on S^{10} and the antibonding orbital of the $C^1\text{-}C^2$ or $C^1\text{-}H^{4/5}$ bond. Again, the NBO analysis of the donor-acceptor interactions (Table 5.6(a)−(b) and Table S 8.8(a)−(b)) confirms these orbital interactions. $LP(2)(S^{10})\rightarrow\sigma^*(C^1\text{-}C^2)$, $LP(2)(S^{10})\rightarrow\sigma^*(C^1\text{-}H^{4/5})$ and $\sigma(C^1\text{-}H^{5/4})\rightarrow\sigma^*(S^9\text{-}S^{10})$ are the more important stabilizing interactions that contribute to 76−77% of the total stabilizing energies around τ_2 (total: 48.41–50.12 kJ mol^{-1}), and they are generally presented in the descending order with stabilizing energies larger or close to 10 kJ mol^{-1}. The donation from the 3p-like $LP(2)(S^{10})$ (Table S 8.9(a)−(b)) to $\sigma^*(C^1\text{-}C^2)$ or $\sigma^*(C^1\text{-}H^{4/5})$ are highly favorable because of the small energy gap between the two orbitals (~0.8 Hartree) and the relatively large orbital overlap (Figure 5.9B1)−2)). The energy gap between $\sigma(C^1\text{-}H^{5/4})$ and $\sigma^*(S^9\text{-}S^{10})$ is smaller but the orbital overlap is decreased possibly due to the mismatch in the orbital size, thus leading to reduced stabilizing interaction.

When τ_2 is close to 180° (denoted by anti) (Figure 5.10(e)), the $S^9\text{-}S^{10}$ bond is anti to the $C^1\text{-}C^2$ bond and gauche (~±60°) to the $C^1\text{-}H^4$ and $C^1\text{-}H^5$

bonds. Thus, hyperconjugations may occur between the bonding orbital of the S^9-S^{10} bond and the antibonding orbital of the C^1-C^2 bond, and vice versa. In addition, hyperconjugations may occur between the either one of the lone pairs on S^{10} and the antibonding orbital of the C^1-H^4 or C^1-H^5 bond. The NBO analysis of the donor-acceptor interactions (Table 5.6(c) and Table S 8.8(c)) confirms most of these orbital interactions. Two donor-acceptor interactions between LP(1)(S^{10}) and the antibonding orbital of the C^1-H^4 or C^1-H^5 bond are missing. The LP(1)(S^{10}) has more 3s character and less 3p character (Table S 8.9(c)) when compared with that in the other four conformers (Table S 8.9(a)−(b)), thus the energy lowering in this lone pair will further increase the energy gap between the interacting orbitals and even diminish the two proposed orbital stabilizations. LP(2)(S^{10})→σ*(C^1-$H^{5/4}$), LP(2)(S^{10})→σ*(C^1-$H^{4/5}$), σ(S^9-S^{10})→σ*(C^1-C^2) and σ(C^1-H^2)→σ*(S^9-S^{10}) are the four actual stabilizing interactions around τ_2, and they are presented in the descending order with the largest interaction close to 10 kJ mol^{-1}. The total stabilizing energies around τ_2 are 26.23–26.40 kJ mol^{-1}. The donation from the 3p-like LP(2)(S^{10}) (Table S 8.9(c)) to the neighboring σ*(C^1-$H^{5/4}$) or σ*(C^1-$H^{4/5}$) is highly favorable due to the small energy gap between the two orbitals (~0.84 Hartree) and the relatively large orbital overlap (Figure 5.9C1)−2)).

When τ_1 is close to ±90^o (denoted by G(+)|G(−)) (Figure 5.10(f)−(g)), the S^{10}-C^1 bond is almost perpendicular to the S^9-C^{11} bond. Consequently, hyperconjugations may occur between one of the two lone pairs on S^{10} and the antibonding orbital of the S^9-C^{11} bond, and between one of the two lone pairs on S^9 and the antibonding orbital of the S^{10}-C^1 bond. The NBO analysis of the

donor-acceptor interactions (Table 5.6(a)−(c) and Table S 8.8(a)−(c)) confirms these orbital interactions. LP(2)(S^9)→σ*(S^{10}-C^1) (Figure 5.9D)) is larger than LP(2)(S^{10})→σ*(S^9-C^{11}), and both interactions have stabilizing energies more than 10 kJ mol^{-1}. The total stabilizing energies around τ_1 are 29.12–53.97 kJ mol^{-1}. The energy gap and orbital overlap are very similar to each other for these two interactions. The LP(S)→σ*(S-C) interaction has been reported in an NMR study on other organosulfur compounds by Bass *et al*.[18]

The total stabilizing energies for each torsional angle are compared separately. For τ_3, only one type of configurations was studied (G(+)|G(−)) (Figure 5.10(a)−(b)), thus the total stabilizing energies are comparable in all conformers and the energy difference is within 4.6 kJ mol^{-1} (Table 5.6(a)−(c)). For τ_2, two types of configurations were studied (denoted by G(+)|G(−) and anti) (Figure 5.10(c)−(d) and (e)). The total stabilizing energies for the G(+)|G(−) configuration in the first four conformers (Table 5.6(a)−(b)) are similar and almost twice larger than that for the anti configuration in the last two conformers (Table 5.6(c)) because of the two missing donor-acceptor interactions discussed earlier. This difference can account for the preference of the G(+)|G(−) configuration over the anti configuration. For τ_1, one type of configurations was studied (+|−) (Figure 5.10(f)−(g)), the total stabilizing energies in the first four conformers (Table 5.6(a)−(b)) are very close and about 1.7 times larger than that in the last two conformers (Table 5.6(c)). This is due to the largely decreased orbital overlap (from ~0.5 a.u. to 0.3 a.u.) in the last two conformers, though the energy gap between the two interacting orbitals is slightly reduced.

The sum of the total stabilizing energies from the donor-acceptors interactions associated with the three torsional angles (τ_3, τ_2, τ_1) and the energies relative to that of the G(+)anti+|G(–)anti– conformation are shown in Table 5.6(a)−(c). The order in the sum of total stabilizing energies is G(+)G(–)+|G(–)G(+)– > G(+)G(–)–|G(–)G(+)+ > G(+)G(+)–|G(–)G(–)+ > G(+)G(+)+|G(–)G(–)– > G(+)anti–|G(–)anti+ > G(+)anti+|G(–)anti–, and it is almost inversely correlated with the order in ΔE in a qualitative way (Figure 5.8). This correlation is reasonable because the resonance stabilization of a conformer will lower its electronic energy. The only outlier is the order between the G(+)G(+)–|G(–)G(–)+ and G(+)G(+)+|G(–)G(–)– conformations, as ΔE of the G(+)G(+)–|G(–)G(–)+ conformation is slightly higher than that of the G(+)G(+)+|G(–)G(–)–conformation. The close dihydrogen contact observed in the G(+)G(–)–|G(–)G(+)+ conformation (Figure 5.8(d)) may lead to some unfavorable steric repulsion, resulting in the higher electronic energy of this conformation. However, the donor-acceptor interactions presented in Table 5.6(a)−(c) cannot account for all of the hyperconjugation effects in MeSSA (e.g. geminal hyperconjugations), and such interactions are also known to overestimate the hyperconjugation effects. As a result, ΔE and the sum of total stabilizing energies from the three torsional angles cannot be compared quantitatively and accurately.

5.3.2 Conformational Study on Reactants and Products from the Hypothetical Cα Nucleophilic Substitution of DMDS/DPDS by MeSH

Apart from the previously studied compounds (DMDS, MeS^- and $MeSS^-$), all other reactants and products from Table 5.3 were studied computationally as follows.

The systematic conformational search for these compounds was carried out manually by the same approach described in section 5.3.1. For τ_1, the optimal angles of 90° and -90° are denoted by "+" and "–". For τ_2 and τ_3, the optimal angles of 60°, -60° and 180° are denoted by "G(+)", "G(–)" and "anti". For example, by considering the (τ_3, τ_2, τ_1, τ_2, τ_3) combination for DPDS (Figure 5.3), the conformational search generated 45 distinct conformers as shown in Table S 8.10 with their numbering and the corresponding conformations. All of these conformers were fully optimized at the SMD/M06-2X/6-31+G* level of theory and are reported in Table S 8.10(b). Table 5.7 shows the top 27 conformers of DPDS with either ΔH_{298} or $\Delta G_{298} \leq$ 5 kJ mol^{-1}. DPDS1, denoted by antiG(+)+G(−)G(−), is the lowest energy conformer according to ΔH_{298}, while DPDS7, denoted by G(+)G(+)+G(+)G(+), is the lowest energy conformer according to ΔG_{298}. These two conformers were chosen as the starting points to construct the relevant transition states in section 6.3.5. Labels of (H) or (G) were added to distinguish these two lowest energy conformers. The similar process was repeated for other compounds in Table 5.3 and the results are summarized in Table S 8.10.

Table 5.7 Top 27 of the calculated DPDS conformations and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

No.	Conformation	ΔH_{298}	ΔG_{298}
1	antiG(+)+G(−)G(−)	0.00	0.29
2	G(+)G(+)+G(−)G(−)	1.14	3.14
3	antiG(+)−G(+)anti	2.55	0.58
4	antiG(+)+G(+)anti	2.99	0.64
5	antiG(+)+G(+)G(+)	3.12	1.37
6	antiG(+)−G(−)G(−)	3.41	2.27
7	G(+)G(+)+G(+)G(+)	3.91	0.00
8	G(+)G(+)−G(−)G(+)	4.07	5.39
9	G(+)G(+)−G(+)G(+)	4.24	7.02
10	antianti+G(−)G(−)	4.38	4.87
11	antiG(+)−G(+)G(+)	5.05	3.79
12	antianti+G(−)anti	5.20	1.38
13	antiG(+)−G(−)G(+)	5.30	5.55
14	G(+)anti+G(−)G(−)	5.32	2.33
15	antianti+G(+)anti	5.75	2.48
16	antiG(+)+G(−)anti	6.09	4.73
17	antianti+G(+)G(+)	6.19	3.71
18	antiG(+)−antiG(−)	6.48	4.92
19	G(+)anti−G(+)G(+)	6.51	6.67
20	antiG(+)+G(+)G(−)	6.55	6.24
21	G(+)G(+)+G(+)G(−)	6.97	2.69
22	antiG(+)+antiG(−)	6.98	3.89
23	antiG(+)−antiG(+)	7.07	3.37
24	G(+)anti−G(−)G(−)	7.19	4.06
25	antiG(+)+antiG(+)	7.23	4.96
26	G(+)anti+G(+)G(+)	7.66	3.76
27	antianti+antianti	8.64	4.41

For PrSS$^-$, the ranking in ΔH_{298} and in ΔG_{298} agree well with each other (Table S 8.10(c)). For MeSPr, the ranking in ΔH_{298} and in ΔG_{298} deviate for the first four conformers out of the five conformers. These four conformers have very close values in ΔH_{298} (within 3 kJ mol^{-1}), so a small difference in S can easily change the ranking in ΔG_{298}. For DPDS, the top 6 conformers have the similar ranking in ΔH_{298} and in ΔG_{298} ($\leq$ 3.5 kJ mol^{-1}), except DPDS2. However, most of the other low-lying conformers (from the top 7 to the top 27) deviate a lot based on the ranking in ΔH_{298} (3.9−8.7 kJ mol^{-1}) and in ΔG_{298} (0.0−7.1 kJ mol^{-1}). The reasoning is similar to that for DATS in section 5.3.1.

The geometries of the lowest energy conformers for all compounds are summarized in Figure 5.11 and Table S 8.11.

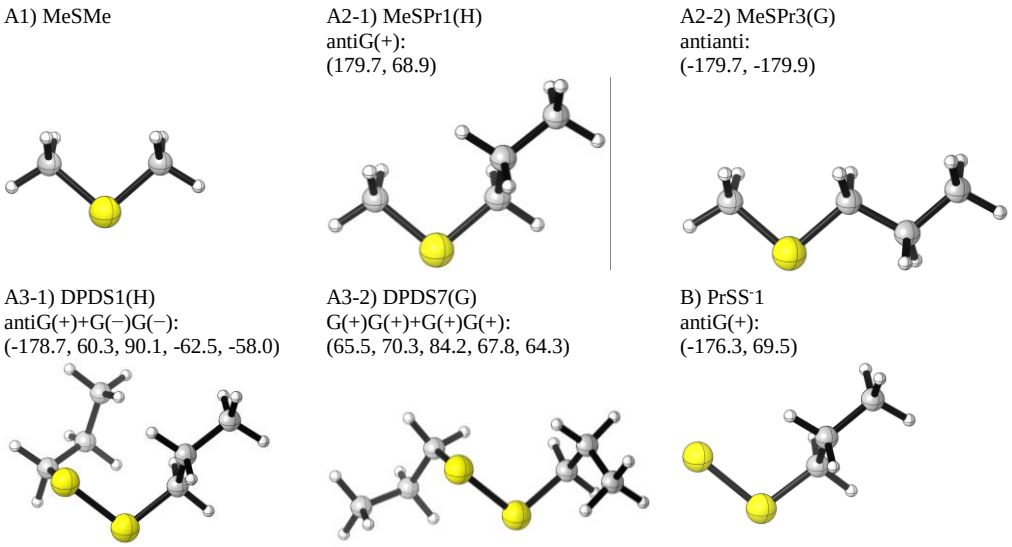
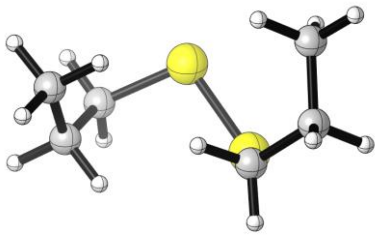


Figure 5.11 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy conformers for: A1)−3-2) dialkyl (poly)sulfides; B) alkyl polysulfides. Torsional angles (in °) are given in the brackets.

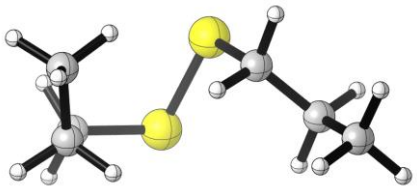
All optimized conformers for all reactants and products were manually inspected and compared to the input structures. The majority of the optimized geometries match the corresponding conformations well (not all results are shown). In general, the values of τ_3 deviate the least ($\leq 6^o$ for G(+) or G(−), $\leq 3^o$ for anti) from the optimal values, followed by τ_2 (denoted by anti) and τ_1 (all $\leq 10^o$). The values of τ_2 (denoted by G(+) or G(−)) can easily deviate from the optimal values by up to ~20^o. However, the change in τ_3 or τ_2 for the (τ_3, τ_2) combination denoted by G(+)G(–)|G(−)G(+) can be as large as ~34^o (Figure 5.12(a)−(b), red). Moreover, the change in τ_1 can be as large as ~26^o for the (τ_2, τ_1, τ_2) combination denoted by G(+)–G(+)|G(−)+G(−) to avoid the steric

repulsion between the two alkyl groups next to the disulfide bridge (Figure 5.12(c), red).

(a) DPDS8
G(+)G(+)−G(−)G(+):
(62.8, 66.2, -93.4, -91.0, 65.3)



(b) DPDS37
antiG(+)+G(−)G(+):
(179.5, 68.2, 91.2, -66.7, 94.0)



(c) DPDS3
antiG(+)−G(+)anti:
(177.1, 69.7, -109.5, 69.7, 177.1)

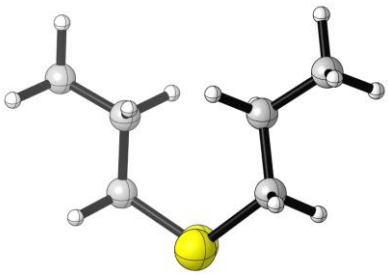
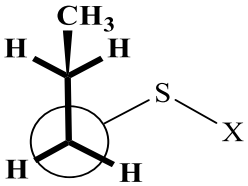
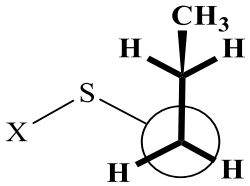


Figure 5.12 Optimized geometries of (a) DPDS8, (b) DPDS37, and (c) DPDS3. Torsional angles (in °) are given in the brackets.

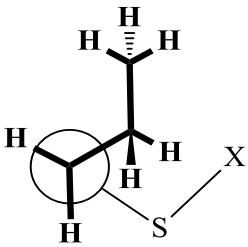
(a) antiG(+)+



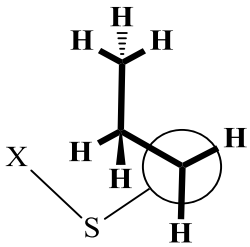
(b) antiG(–)–



(c) G(+)G(+)–



(d) G(–)G(–)+



(e) G(+)G(+)+ (f) G(–)G(–)–

Figure 5.13 The lowest energy conformations of the -SS-propyl group.

For dialkyl (poly)sulfides, including MeSMe, MeSPr and DPDS, two alkyl groups (-CH_3 or -$CH_2CH_2CH_3$) are joined by one or two sulfur atoms. The SMD/M06-2X/6-31+G* optimized geometries—MeSMe, MeSPr1(H), MeSPr3(G), DPDS1(H) and DPDS7(G)—are shown in Figure 5.11A1)−3-2) and Table S 8.11(a)−(b). The preferred conformations of MeSPr1(H) and MeSPr3(G), i.e. antiG(+) and antianti, are consistent with a spectroscopic study reported by Sugeta *et al.*[11f]. The bond lengths, bond angles and torsional angle around the disulfide linkage in DPDS1(H) and DPDS7(G) (Table S 8.11(a)−(b)) agree well with the typical values in the chain-like organic polysulfides[9]. In addition, the preference for the G(+)+G(+)|G(−)−G(−) conformation around the central C-S-S-C group has been reported by studies on the conformations of diethyl disulfide experimentally and computationally[11b, 11d]. In these conformers, the -S-propyl groups may adopt the antiG(+)|antiG(–), G(+)G(+)|G(–)G(–) or antianti conformations with (τ_3, τ_2) ≈ (180^o, $\pm60^o$), ($\pm60^o$, $\pm60^o$) or (180^o, 180^o) respectively. With one disulfide bond in DPDS, the -SS-propyl groups may adopt the antiG(+)+|antiG(–)–, G(+)G(+)–|G(–)G(–)+ or G(+)G(+)+|G(–)G(–)– conformations with (τ_3, τ_2, τ_1)

≈ (180°, ±60°, ±90°), (±60°, ±60°, ±270°) or (±60°, ±60°, ±90°) respectively (Figure 5.13).

For S-alkyl polysulfide, such as the deprotonated PrSSH, one alkyl group (-$CH_2CH_2CH_3$) is linked to two sulfur atoms and such compound has an ionizable -SSH group. The SMD/M06-2X/6-31+G* optimized geometry of the lowest energy conformer—$PrSS^-1$—is shown in Figure 5.11B) and Table S 8.11(b). The -S-propyl group in the deprotonated PrSSH adopts the identical conformation to that in MeSPr1(H) and DPDS1(H).

From the conformational analysis of $PrSS^-$ and MeSPr (Table S 8.10(c) and Table S 8.11(a)), the conformation of the -S-propyl groups has the order of antiG(+)|antiG(–) > G(+)G(+)|G(–)G(–) > antianti > G(+)anti|G(–)anti > G(+)G(–)|G(–)G(+) based on ΔH_{298}, whereas it has the order of antianti > G(+)anti|G(–)anti > antiG(+)|antiG(–) > G(+)G(+)|G(–)G(–) > G(+)G(–)|G(–)G(+) based on ΔG_{298}. From the conformational analysis of DPDS (Table 5.7), the conformation of the -SS-propyl groups usually has the order of antiG(+)+|antiG(–)– > G(+)G(+)+|G(–)G(–)–/G(+)G(+)–|G(–)G(–)+/antiG(+)–|antiG(–)+ > G(+)G(–)+|G(–)G(+)–/antianti+|antianti− > G(+)anti+|G(–)anti– > G(+)anti–|G(–)anti+ based on ΔH_{298}, and the order of G(+)G(+)+|G(–)G(–)–> antiG(+)+|antiG(–)–/antiG(+)–|antiG(–)+/G(+)G(+)–|G(–)G(–)+ > antianti+|antianti−/G(+)anti+|G(–)anti–> G(+)G(–)+|G(–)G(+)–/G(+)anti–|G(–)anti+ based on ΔG_{298}.

5.4 Conclusion

By simplifying the nucleophile GSH to MeSH and assuming the uncomplicated S_N2 mechanism in all of the nucleophilic substitutions, the conformations of reactants and products, from the reaction of MeSH and DADS/DATS as well as from the hypothetical Cα nucleophilic substitution of DMDS/DPDS by MeSH, were studied computationally.

The three torsional angles considered in the conformation analysis are τ(C-S-S-X, X=H, C or S) (τ_1), τ(C-S-S-X, X=H, C or S) (τ_2), and τ(C-C-S-X, X=C or S) (τ_3). τ_1 was denoted by "+" or "–" for values close to ±90°, and τ_2 was denoted by "G(+)" or "G(–)" for values close to ±60°. τ_3 was denoted by "G(+)" or "G(–)" for values close to ±120° if the C-S bond is linked to an allyl group, whereas τ_3 was denoted by "G(+)", "G(–)" or "anti" for values close to ±60° or 180° if the C-S bond is linked to an alkyl group.

From the analysis of the lowest energy conformers of the allyl-containing compounds from the reaction of MeSH and DADS/DATS, it was found that the -S-allyl groups all adopt the G(+)G(–)|G(–)G(+) conformation for the (τ_3, τ_2) combination, while the -SS-allyl groups may adopt the G(+)G(–)+|G(–)G(+)– or the G(+)G(–)–|G(–)G(+)+ conformation for the (τ_3, τ_2, τ_1) combination. The G(+)G(–)–|G(–)G(+)+ conformation of the -SS-allyl groups is preferred over the G(+)G(–)+|G(–)G(+)– conformation in ΔG_{298} when the substituent groups on the disulfide are bulkier in size to reduce the steric repulsion. It was demonstrated that the preferences for these conformations are due to the resonance stabilization by hyperconjugations based on the NBO

analysis. In addition, the trisulfide linkages (-S-S-S-) all adopt the ++|— conformation to avoid the steric repulsion in the +–|–+ conformation.

From the analysis of the lowest energy conformers of the propyl-containing compounds from the Cα nucleophilic substitution of DPDS by MeSH, it was found that the -S-propyl groups may adopt the antiG(+)|antiG(–), G(+)G(+)|G(–)G(–) or antianti conformation for the (τ_3, τ_2) combination, while the -SS-propyl groups may adopt antiG(+)+|antiG(–)–, G(+)G(+)–|G(–)G(–)+ or G(+)G(+)+|G(–)G(–)– conformation for the (τ_3, τ_2, τ_1) combination. The conformations of the -S-propyl groups partially agrees with results from the conformational studies on MeSPr[11f]. The central C-S-S-C group prefers to adopt the G(+)+G(+)|G(−)−G(−) conformation, which is consistent with experimental studies[11b, 11d].

The study of these reactants and products is necessary for studying the transition states and energy profile studies in Chapter 6. Furthermore, to our knowledge, this is probably the first comprehensive computational study on the stable conformers of the allyl-containing organosulfur compounds in an aqueous environment.

5.5 References

1. (a) Liang, D.; Wu, H.; Wong, M. W.; Huang, D., *Org. Lett.* **2015,** *17* (17), 4196; (b) Benavides, G. A.; Squadrito, G. L.; Mills, R. W.; Patel, H. D.; Isbell, T. S.; Patel, R. P.; Darley-Usmar, V. M.; Doeller, J. E.; Kraus, D. W., *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (46), 17977.

2. (a) Wright, W., *Acta Crystallogr.* **1958,** *11* (9), 632; (b) Krezel, A.; Bal, W., *Org. Biomol. Chem.* **2003,** *1* (22), 3885.

3. (a) Görbitz, C. H., *Acta Chem. Scand.* **1987,** *41b*, 362; (b) Zhang, R.; Wu, W.; Luo, S., *J. Solution Chem.* **2011,** *40* (10), 1784; (c) York, M. J.; Beilharz, G. R.; Kuchel, P. W., *Int. J. Pept. Protein Res.* **1987,** *29* (5), 638; (d) Fujiwara, S.; Formicka-Kozlowska, G.; Kozlowski, H., *Bull. Chem. Soc. Jpn.* **1977,** *50* (12), 3131.

4. (a) Laurence, P. R.; Thomson, C., *Theor. Chim. Acta* **1980,** *57* (1), 25; (b) Lampela, O.; Juffer, A. H.; Rauk, A., *J. Phys. Chem. A* **2003,** *107* (43), 9208; (c) Kurian, R. Computational and Experimental Investigations of Metalprotein Thiol/Disulfide Exchange Reactions. Electronic Theses and Dissertations, University of Maine, 2013.

5. (a) Aida, M.; Nagata, C., *Chem. Phys. Lett.* **1984,** *112* (2), 129; (b) Neves, R. P. P.; Fernandes, P. A.; Varandas, A. J. C.; Ramos, M. J., *J. Chem. Theory Comput.* **2014,** *10* (11), 4842.

6. (a) Tajc, S. G.; Tolbert, B. S.; Basavappa, R.; Miller, B. L., *J. Am. Chem. Soc.* **2004,** *126* (34), 10508; (b) Rabenstein, D. L., *J. Am. Chem. Soc.* **1973,** *95* (9), 2797; (c) Huckerby, T. N.; Tudor, A. J.; Dawber, J. G., *J. Chem. Soc., Perkin Trans. 2* **1985,** (6), 759.

7. (a) Nagy, P., *Antioxid. Redox Signal.* **2013,** *18* (13), 1623; (b) Singh, R.; Whitesides, G. M., In *Sulphur-Containing Functional Groups (1993)*, John Wiley & Sons, Inc.: 2010; pp 633.

8. (a) Fava, A.; Iliceto, A.; Camera, E., *J. Am. Chem. Soc.* **1957,** *79* (4), 833; (b) Liang, J.; Fernandez, J. M., *ACS nano* **2009,** *3* (7), 1628; (c) Fernandes, P.

A.; Ramos, M. J., *Chemistry* **2004,** *10* (1), 257; (d) Bach, R. D.; Dmitrenko, O.; Thorpe, C., *J. Org. Chem.* **2008,** *73* (1), 12.

9. Steudel, R., *Chem. Rev.* **2002,** *102* (11), 3905.

10. (a) Suzuki, H.; Fukushi, K.; Ikawa, S.-I.; Konaka, S., *J. Mol. Struc.* **1990,** *221*, 141; (b) Salivon, N. F.; Olijnik, V. V.; Shkurenko, A. A., *Russ J. Coord. Chem.* **2007,** *33* (12), 908; (c) Lin, A. C.; Salpietro, S. J.; Deretey, E.; Csizmadia, I. G., *Can. J. Chem.* **2000,** *78* (3), 362; (d) Bartkowska, B.; Kruger, C., *Acta Crystallogr. Sect. C* **1997,** *53* (8), 1064; (e) Koch, H. P., *J. Chem. Soc.* **1949,** (0), 394.

11. (a) Ohsaku, M.; Allinger, N. L., *J. Phys. Chem.* **1988,** *92* (16), 4591; (b) Ackermann, K. R.; Koster, J.; Schlücker, S., *Chem. Phys.* **2009,** *355* (1), 81; (c) Allinger, N. L.; Kao, J.; Chang, H. M.; Boyd, D. B., *Tetrahedron* **1976,** *32* (23), 2867; (d) Sugeta, H.; Go, A.; Miyazawa, T., *Bull. Chem. Soc. Jpn.* **1973,** *46* (11), 3407; (e) Brandt, N. N.; Chikishev, A. Y.; Kargovsky, A. V.; Nazarov, M. M.; Parashchuk, O. D.; Sapozhnikov, D. A.; Smirnova, I. N.; Shkurinov, A. P.; Sumbatyan, N. V., *Vib. Spectrosc.* **2008,** *47* (1), 53; (f) Nogami, N.; Sugeta, H.; Miyazawa, T., *Chem. Lett.* **1975,** (2), 147; (g) Van Wart, H. E.; Scheraga, H. A., *Proc. Natl. Acad. Sci. U.S.A.* **1986,** *83* (10), 3064.

12. (a) Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008,** *41* (2), 157; (b) Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1-3), 215.

13. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B* **2009,** *113* (18), 6378.

14. Weinhold, F.; Landis., C. R., *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*. Cambridge University Press: Cambridge, UK, 2005.

15. 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. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

16. Yokozeki, A.; Bauer, S. H., *J. Phys. Chem.* **1976,** *80* (6), 618.

17. Mundt, O.; Becker, G.; Baumgarten, J.; Riffel, H.; Simon, A., *Z. Anorg. Allg. Chem.* **2006,** *632* (10-11), 1687.

18. Bass, S. W.; Evans, S. A., *J. Org. Chem.* **1980,** *45* (4), 710.

Chapter 6 Computational Study of Reaction Mechanisms on H_2S Releasing Reactions from Organosulfur Compounds

6.1 Introduction

In Chapter 5, a computational study on the conformations of reactants and products from the reaction of GSH and various polysulfides was reported. This chapter reported on the mechanistic study of these reactions.

First, the mechanisms of the H_2S releasing reactions from GSH (modeled as MeSH) and DADS/DATS were studied computationally. The reaction steps considered are shown in Table 5.1 in section 5.1.2. There are two main objectives of the study: (1) to confirm the experimental results from Liang *et al.*[1] by comparing the activation barriers between the Cα nucleophilic substitution and S nucleophilic substitution of DADS by GSH and to elucidate the chemical reasons; (2) to examine the overall reaction profile leading to H_2S production.

Second, the hypothetical Cα nucleophilic substitution of DMDS/DPDS by GSH (modeled as MeSH) was studied computationally. The aim of the study is to compute the activation barriers in these reactions and to understand how the allyl group or the alkyl group next to the disulfide linkage affects the Cα nucleophilic substitution.

To construct the transition states in this chapter, κ_1/κ_1', κ_2 and κ_3 may be defined for torsional angles close to the reaction center in the dissociating

reactant that share the identical definition and notation as τ_1, τ_2 and τ_3 respectively. χ_1/χ_1', χ_2/χ_2' and χ_3/χ_3' may be defined for torsional angles in the forming product just like κ_1/κ_1', κ_2 and κ_3. For torsional angles around the breaking S_{lg}-Cα/S bond, their values are difficult to predict and the corresponding notations will be modified according to the actual angles. Moreover, for torsional angles in the group being transferred that are adjacent to the breaking/forming S_{lg}-Cα/S bond, their values are expected to deviate a lot from their optimal values due to the increased steric bulk around the Cα/S center, but their notations will not be changed.

6.2 Computational Methods

Because this chapter is a continuation of the study presented in Chapter 5, the computational methods in this chapter mostly follow that in section 5.2. All equilibrium structures and transition state were fully optimized at the M06-2X[2]/6-31+G* level of theory using the SMD[3] implicit solvent model. For the fully optimized geometries, vibrational frequency analysis was performed at the same level to confirm the nature of the stationary points as equilibrium structures or transition states. Unless otherwise stated, the relative energies reported correspond to the relative enthalpies ΔH_{298}, computed at the SMD/M06-2X/6-31+G* level. Single point energy calculations at the reaction temperature of 310 K (37^{o}C) or with larger basis sets such as 6-311+G(2d,p) and 6-311++G(2d,2p) were performed for some selected SMD/M06-2X/6-31+G* optimized geometries. The NBO[4] analysis was performed based on the SMD/M06-2X/6-31+G* wavefunction, and the atomic charges and donor-

acceptor interactions were obtained directly from the NBO analysis. All calculations were performed using the *Gaussian* 09[5] suite of programs.

Charge density analysis, based on AIM[6] was carried out using the MORPHY98[7] program. In the AIM theory, a critical points (CP) is defined by a stationary point with electron density ρ such that the Laplacian of electron density ($\nabla^2\rho$) is zero. Bond critical points (BCP) are one type of critical points with ρ at minimum along the direction of the bond path and at maximum in the other two directions that are perpendicular to the bond path. Therefore, BCP is a second-order saddle point. It is known that the electron density at BCP is correlated with the bond strengths for non-covalent interactions, especially hydrogen bonds[8]. For non-covalent interactions described by BCPs, the values of ρ are usually small and positive and the values of $\nabla^2\rho$ are positive.

For the Cα nucleophilic substitution of DADS by the full reactant GSH, the transition state structures were first obtained from the conformational search of a manually built input structure using the Amber12:EHT[9] force field and the LowModeMD[10] search method implemented in MOE2014.09[9]. The breaking and forming C-S bonds were fixed at 2.45 Å and 2.65 Å respectively, while the S-C-S angle was fixed at 180°. The Amber12:EHT[9] force field was explained in Chapter 3. This force field was chosen because it can describe the conformations of GSH reasonably well while being able to generate conformations for DADS. The LowModeMD[10] search method generates conformations based on a ~1 ps MD simulation at a constant temperature followed by geometry optimization. It is expected to efficiently locate most of the local minima of a complex system and is the recommended method for

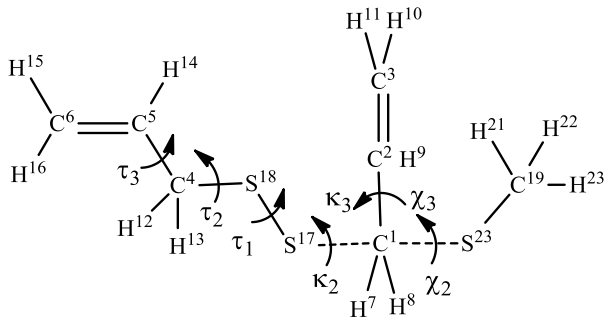
conformation generation in MOE[9]. The solvation model was used with a distance dependent dielectric of 78.4 for water. 200 conformations were generated from the conformational search, and the process was repeated five times. Other parameters set in the LowModeMD search were as follows: Rejection Limit (100 attempts), Iteration Limit (10000 attempts), RMS Gradient (0.05 kcal mol^{-1} $Å^{-2}$), MM Iteration Limit (1000 steps), RMSD Limit (0.5 Å), Energy window (20 kcal mol^{-1}). Only the lowest energy conformer from each conformational search was selected for the full DFT optimization described earlier.

6.3 Results & Discussion

6.3.1 Transition State Study in Reaction of MeSH and DADS/DATS

The transition state study for reactions in Table 5.1 was performed based on the most stable conformation(s) of the corresponding reactants to study the transition state features and to approximate reaction energies and activation barriers. This decision was made to simplify the transition state study due to the presence of many low-lying conformations for most of the reactants and the relatively large number of reactions studied.

(a)



(b)

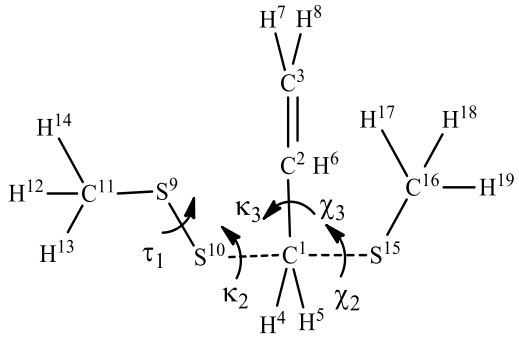


Figure 6.1 Chemical structure of the TS in reaction of MeS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution with labeling of atoms and torsional angles.

To construct the transition states in reaction of MeS^- and DADS (Table 5.1, reaction 2) or MeSSA (Table 5.1, reaction 5) via Cα nucleophilic substitution, χ_2, χ_3, κ_2 and κ_3 were defined as described in section 6.1. The torsional angles and atom numbering in the transition state (TS) models are depicted in Figure 6.1(a)−(b). As mentioned in section 5.1.1, the nucleophilic substitution was assumed to take place in the S_N2 manner at all times. For the reaction of MeS^- and DADS, the atom numbering in Figure 6.1(a) is based on the nucleophile attack on the Cα in the -SS-allyl group denoted by G(+)G(−)−, and it will be changed accordingly for the nucleophile attack on the Cα in the -SS-allyl group denoted by G(−)G(+)−. “o” will be added to the end of DADS1 to indicate the TS from the latter case.

Table 6.1 Calculated TS conformations from the reaction of MeS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

(a)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔH_{298}
1	TS-C-MeS-A-DADS1	G(+)anti; G(−)G(+)	7.13
2	TS-C-MeS-G-DADS1	G(+)G(+); G(−)G(+)	8.87
3	TS-C-MeS-G'-DADS1[a]	G(+)G(−); G(−)G(+)	0.00
4	TS-C-MeS-A-DADS1o	G(−)anti; G(+)G(−)	5.42
5	TS-C-MeS-G-DADS1o	G(−)G(+); G(+)G(−)	0.27
6	TS-C-MeS-G'-DADS1o	G(−)G(−); G(+)G(−)	9.76

(b)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔH_{298}
1	TS-C-MeS-A-MeSSA1	G(−)anti; G(+)G(−)	5.51
2	TS-C-MeS-G-MeSSA1	G(−)G(+); G(+)G(−)	0.00
3	TS-C-MeS-G'-MeSSA1	G(−)G(−); G(+)G(−)	7.09
4	TS-C-MeS-G'-MeSSA2	G(+)G(−); G(−)G(+)	1.70

[a] TS-C-MeS-G'-DADS1 has two imaginary frequencies at 501.91 cm^{-1} and 8.87 cm^{-1} respectively.

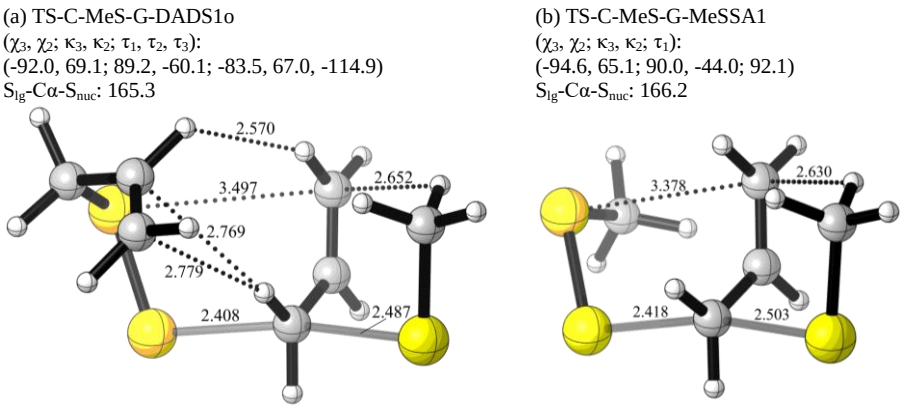


Figure 6.2 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DADS, or (b) MeSSA via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

By considering the three possible values of χ_2 (±60° and 180°) and the nucleophilic attack on the either side of DADS1 (G(+)G(–)–G(+)G(–)), six TS conformations were studied from the reaction of MeS^- and DADS1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.1(a) with

the TS names, conformations, and their relative enthalpies. The transition states are named after the reaction type Cα nucleophilic substitution (C), the nucleophile MeS^- (MeS), the electrophile DADS1 (DADS1 or DADS1o), and the simplified χ_2 notation A, G and G' that correspond to the "anti", "G(+)" and "G(−)" configurations respectively. The similar names of transition states are given to all other TS involving MeS^- as the nucleophile in this section. The names of the electrophiles will be replaced by the actual electrophile, and the reaction type C will be replaced by S, SS or MS to describe the S, side-S and mid-S nucleophilic substitutions respectively. Moreover, the simplified χ_1 notation G and G' will be used to for the "+" and "–" configurations respectively.

Out of these six SMD/M06-2X/6-31+G* optimized conformations, TS-C-MeS-G'-DADS1 is not a true TS because it has two imaginary frequencies. One imaginary frequency is around 500 cm^{-1} that describes the vibration along the reaction coordinate $S_{lg}\cdots C\alpha\cdots S_{nuc}$ and is characteristic in all transition states from the Cα nucleophilic substitution. "lg" stands for the leaving group, and "nuc" stands for the nucleophile. The other imaginary frequency is at 8.87 cm^{-1} that describes the rotation around κ_2. Our calculation failed to eliminate this imaginary frequency after several attempts probably due to the relatively flat PES around the true TS. Therefore, the conformation and energy of TS-C-MeS-G'-DADS1 will not be analyzed and discussed.

From Table 6.1(a), the conformation of the forming MeS···A, described by (χ_3, χ_2), has the order of G(+)G(–)|G(–)G(+) > G(+)anti|G(–)anti > G(+)G(+)|G(–)G(–) in ΔH_{298}. By excluding TS-C-MeS-G'-DADS1, TS-C-

MeS-G-DADS1o is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.2(a).

From manual inspection of TS in Table 6.1(a), τ_3 deviates the least (-1.1−3.1^o) from that in DADS1, followed by τ_2 (-5−8.4^o) and τ_1 (-0.8−9.1^o). Nevertheless, the change in these torsional angles did not vary their torsional configurations as compared to DADS1.

Similarly, by considering the three possible values of χ_2 (±60^o and 180^o), three TS conformations were first studied from the reaction of MeS^- and MeSSA1 at the SMD/M06-2X/6-31+G* level, and the results are reported in Table 6.1(b) with the TS names, conformations, and their relative energies. The conformation of the forming MeS···A, described by (χ_3, χ_2), has the order of G(+)G(–)|G(–)G(+) > G(+)anti|G(–)anti > G(+)G(+)|G(–)G(–) in ΔH_{298}, which is identical to that in the reaction of MeS^- and DADS. To simplify the calculation, only one TS conformation with the most preferred conformation of the forming MeS···A denoted by G(+)G(−) was studied for the reaction of MeS^- and MeSSA2 (Table 6.1(b)). Overall, TS-C-MeS-G-MeSSA1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.2(b).

From manual inspection of TS in Table 6.1(b), τ_1 in the first three transition states deviates by -1.3−1.6^o from that in MeSSA1, while τ_1 in the last transition state deviates by -10.7^o from that in MeSSA2. Again, such changes did not vary their torsional configurations as compared to MeSSA1 or MeSSA2.

From Figure 6.2, TS-C-MeS-G-DADS1o and TS-C-MeS-G-MeSSA1 have the identical conformation of the forming MeS···A—G(+)G(–)|G(–)G(+)—such that the forming MeS···A resembles the most stable form of MeSA, i.e. MeSA1, in Figure 5.4. The geometric features of TS-C-MeS-G-DADS1o and TS-C-MeS-G-MeSSA1 agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking Cα-S_{lg} bond (2.408−2.418 Å) that is longer than the C-S bond in DADS1 or MeSSA1 (1.841−1.843 Å), and the other corresponds to the forming Cα-S_{nuc} bond (2.487−2.503 Å) that is longer than the C-S bond in MeSA1 (1.831 Å). The Cα is pentacoordinate, and almost adopts a trigonal planar geometry considering the three covalent bonds around it, thus it is approximately sp^2-hybridized instead of sp^3-hybridized in DADS1 and MeSSA1. The nucleophile MeS^- attacks the backside of Cα at an angle of 165.3−166.2° from the leaving group and causes an inversion at Cα.

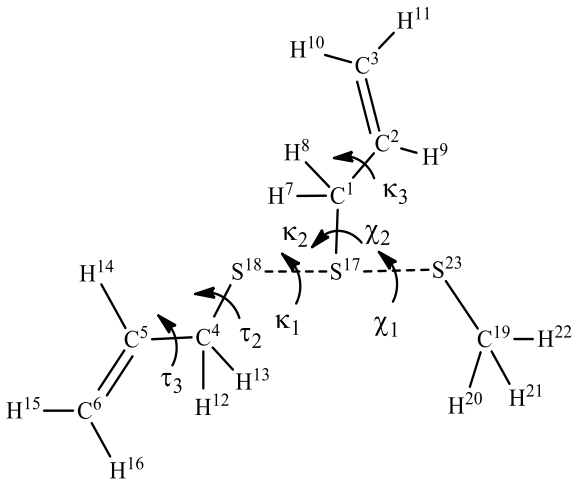
By examining the torsional angles, χ_3 and κ_3 are different from τ_3 and have values close to ±90° with the opposite signs due to the backside attack. χ_2 is similar to τ_2 and has values close to 60°. κ_2 is increased by 6.7° in TS-C-MeS-G-DADS1o from the corresponding τ_2 (-66.8°) in DADS1, while it is increased by 19.4° in TS-C-MeS-G-MeSSA1 from the corresponding τ_2 (-63.4°) in MeSSA1.

TS-C-MeS-G-DADS1o has one set of close dihydrogen contact (2.570 Å) between the two hydrogens on the two -CH=CH_2 groups. It also has two possible C-H···π interactions as indicated by the short distance (2.652 Å) between the methyl-hydrogen and the sp^2 carbon of the allyl group next to it,

and the short distances (2.769 Å and 2.779 Å) between one allyl-hydrogen and the two sp^2 carbons of the other allyl group. The C-H···π interaction in the latter case was also observed in the reactant DADS1 (Figure 5.4). Moreover, it has one possible LP(S)···π interaction (3.497 Å) between S and the sp^2 carbon of the allyl group next to it.

TS-C-MeS-G-MeSSA1 has one possible C-H···π interaction (2.630 Å) between the methyl-hydrogen and the sp^2 carbon of the allyl group nearby. In addition, it has one possible LP(S)···π interaction (3.378 Å) between S and the sp^2 carbon of the allyl group next to it. These two interaction are also present in TS-C-MeS-G-DADS1o as discussed above. Such C-H···π and LP(S)···π interactions were not identified in the product MeSA and are unique to the transition states from the Cα nucleophilic substitution.

(a)



(b)

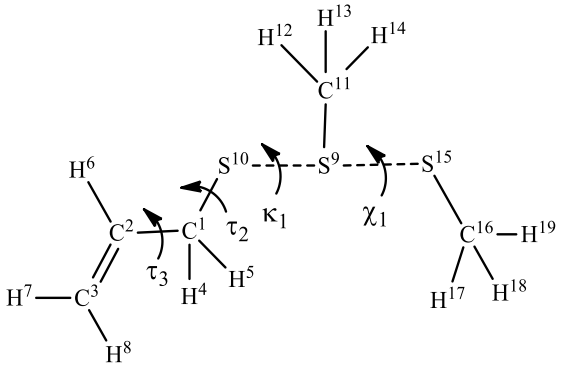


Figure 6.3 Chemical structure of the TS in reaction of MeS^- and (a) DADS or (b) MeSSA via S nucleophilic substitution with labeling of atoms and torsional angles.

To construct the transition states in reaction of MeS^- and DADS (Table 5.1, reaction 3) or MeSSA (Table 5.1, reaction 6) via S nucleophilic substitution, χ_1, χ_2, κ_1, κ_2 and κ_3 were defined as described in section 6.1. The torsional angles and atom numbering in the TS models are depicted in Figure 6.3(a)−(b). Similar as before, for the reaction of MeS^- and DADS, the atom numbering in Figure 6.3(a) is based on the nucleophile attack on the S in the -SS-allyl group denoted by G(+)G(−)−, and it will be changed accordingly for the nucleophile attack on the S in the -SS-allyl group denoted by G(−)G(+)−. Again, "o" will be added to the end of DADS1 to indicate the TS from the latter case.

Table 6.2 Calculated TS conformations from the reaction of MeS^- and (a) DADS or (b) MeSSA via S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

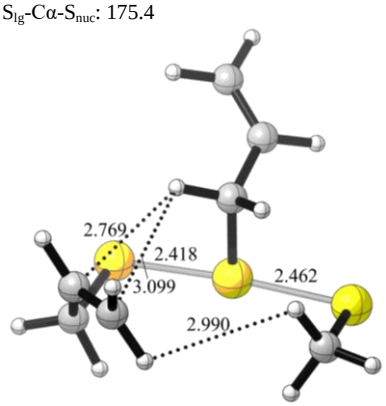
(a)

No.	Name	Conformation κ_3, χ_2, χ_1; κ_2, κ_1	ΔH_{298}
1	TS-S-MeS-G-DADS1	G(−)G(−)+; G(+)−	2.48
2	TS-S-MeS-G'-DADS1	G(−)G(−)−; G(+)−	3.77
3	TS-S-MeS-G-DADS1o	G(+)G(+)+; G(−)−	0.00
4	TS-S-MeS-G'-DADS1o	G(+)G(+)−;G(−)−	3.15

(b)

No.	Name	Conformation χ_1; κ_1	ΔH_{298}
1	TS-S-MeS-G-MeSSA1	+; +	2.66
2	TS-S-MeS-G'-MeSSA1	−; +	0.00
3	TS-S-MeS-G-MeSSA2	+; +	4.08
4	TS-S-MeS-G'-MeSSA2	−; +	2.22

(a) TS-S-MeS-G-DADS1o
(χ_2, χ_1; κ_3, κ_2, κ_1; τ_2, τ_3):
(87.1, 82.0; 118.9, -96.9, -95.1; 71.4, -112.1)
S_{lg}-Cα-S_{nuc}: 175.4



(b) TS-S-MeS-G'-MeSSA1
(χ_1; κ_1; τ_2, τ_3):
(-72.0; 100.9; -71.1, 114.6)
S_{lg}-Cα-S_{nuc}: 173.6

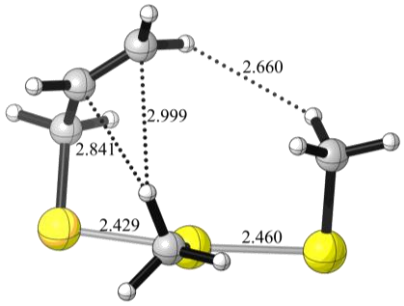


Figure 6.4 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DADS, or (b) MeSSA via S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

By considering the two possible values of χ_1 (±90°) and the nucleophilic attack on the either side of DADS1, four TS conformations were studied from the reaction of MeS^- and DADS1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.2(a) with the TS names, conformations, and their relative enthalpies. Moreover, transition states from the S/side-S/mid-S nucleophilic substitution were much more difficult to locate compared with those from the Cα nucleophilic substitution. Geometry optimization often resulted in the reactant complex or the product complex instead of the desired transition states. It is known that the S nucleophilic substitution on a disulfide linkage occurs readily and reversibly at room temperature in water[11], therefore

this reaction is expected to have a relatively small activation barrier and a small energy gap between the reactant complex and product complex. Consequently, such transition states are more difficult to locate and careful adjustments of the reaction coordinate are required. The characteristic imaginary frequency in such transition states along the reaction coordinate $S_{lg}\cdots S\cdots S_{nuc}$ is around 150−200 cm^{-1}.

From Table 6.2(a), by examining the conformation of the forming MeS···SA, described by (κ_3, χ_2, χ_1), the order between the G(+)G(+)+|G(–)G(−)− and G(+)G(+)−|G(–)G(–)+ conformations is unclear in ΔH_{298}. Despite this observation, TS-S-MeS-G-DADS1o is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.4(a).

From manual inspection of TS in Table 6.2(a), τ_3 deviates the least (-1.2−4.2°) from that in DADS1, followed by τ_2 (-5.6−7.8°). Nevertheless, the change in these torsional angles did not vary their torsional configurations as compared to DADS1.

Similarly, by considering the two possible values of χ_1 (±90°) and the two lowest energy conformers of MeSSA, four TS conformations were studied from the reaction of MeS^- and MeSSA1 or MeSSA2 at the SMD/M06-2X/6-31+G* level. The results are reported in Table 6.2(b) with the TS names, conformations, and their relative energies. TS-S-MeS-G'-MeSSA1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.4(b).

From manual inspection of TS in Table 6.2(b), τ_3 deviates the least (-3.3−2.1°) from that in the corresponding MeSSA1 or MeSSA2, followed by τ_2

(-14.1−10.2°). Nevertheless, the change in these two torsional angles did not vary their torsional configurations as compared to MeSSA1 or MeSSA2.

From Figure 6.4, the geometric features of TS-S-MeS-G-DADS1o and TS-S-MeS-G'-MeSSA1 all agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking S-S_{lg} bond (2.418−2.429 Å) that is longer than the S-S bond in DADS1 or MeSSA1 (2.059−2.063 Å), and the other corresponds to the forming S-S_{nuc} bond (2.460−2.462 Å) that is longer than the S-S bond in MeSSA1 or DMDS (2.063−2.065 Å). The S is tricoordinate, and the nucleophile MeS^- attacks the backside of the disulfide linkage at an angle of 173.6−175.4° and causes an inversion at S.

By examining the torsional angles, χ_2 and κ_2 are different from τ_2 and have values close to ±90° with the opposite signs in TS-S-MeS-G-DADS1o due to the backside attack. χ_1 does not deviate a lot from τ_1. It has values close to 80° in TS-S-MeS-G-DADS1o, and close to -70° in TS-S-MeS-G'-MeSSA1. κ_3 is increased by 5.8° in TS-S-MeS-G-DADS1o from the corresponding τ_3 (113.1°) in DADS1. κ_1 is decreased by 2.8° in TS-S-MeS-G-DADS1o from the corresponding τ_1 (-92.3°) in DADS1, while it is increased by 7.5° in TS-S-MeS-G'-MeSSA1 from the corresponding τ_1 (93.4°) in MeSSA1.

TS-S-MeS-G-DADS1o has one set of close dihydrogen contact (2.990 Å) between the methyl-hydrogen and the hydrogen on the -CH=CH_2 group. It also has one possible C-H···π interaction as indicated by the short distances (2.769 Å and 3.099 Å) between the allyl-hydrogen and the two sp^2 carbons of the

allyl group next to it. The C-H···π interaction in TS-S-MeS-G-DADS1o was also observed in the reactant DADS1 (Figure 5.4).

TS-S-MeS-G'-MeSSA1 has one set of close dihydrogen contact (2.660 Å) between the methyl-hydrogen on MeS^- and the hydrogen on the $-CH=CH_2$ group. It also has one possible C-H···π interaction as indicated by the short distances (2.841 Å and 2.999 Å) between the methyl-hydrogen and the sp^2 carbons of the allyl group in close proximity, and this interaction was also observed in the reactant MeSSA1.

The favorable transition state conformations from the S nucleophilic reaction cannot be predicted by the forming product MeSSA or DMDS unlike the Cα nucleophilic reaction. However, TS-S-MeS-G-DADS1o and TS-S-MeS-G'-MeSSA1 both have the substituent groups (R_1, R_3) at the two sides pointing to the same direction (Figure 6.5A1-1)–1-2)), possibly due to the favorable interactions between these two substituents.

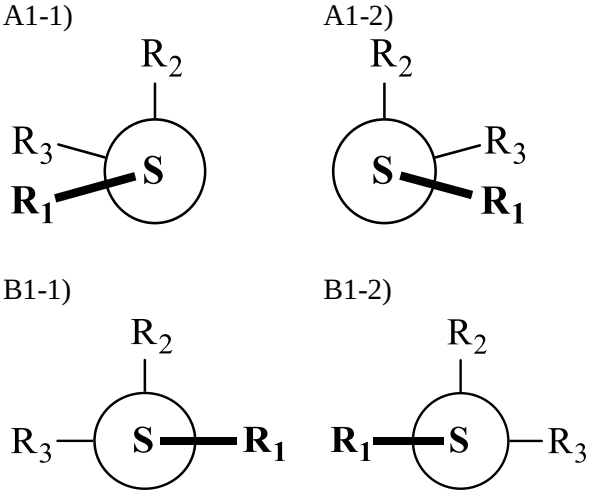
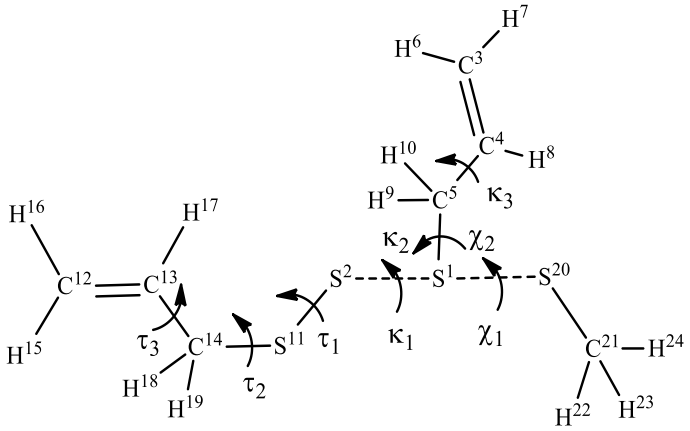


Figure 6.5 Relative orientation of different substituent groups in transition states from the S nucleophilic substitution.

(a)



(b)

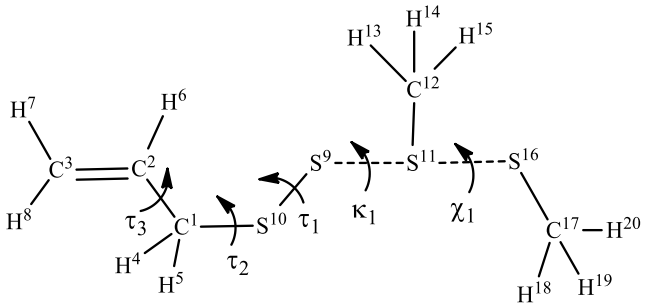


Figure 6.6 Chemical structure of the TS in reaction of MeS^- and (a) DATS or (b) MeS_3A via side-S nucleophilic substitution with labeling of atoms and torsional angles.

The transition states in reaction of MeS^- and DATS (Table 5.1, reaction 12) or MeS_3A (Table 5.1, reaction 14) via side-S nucleophilic substitution resemble that in reaction of MeS^- with DADS or MeSSA via S nucleophilic substitution. The only difference is that there are three S atoms in DATS or MeS_3A instead of two S atoms in DADS or MeSSA. As a result, τ_1 should be added in the transition states. The torsional angles and the atom numbering in the TS models are depicted in Figure 6.6(a)−(b).

Table 6.3 Calculated TS conformations from the reaction of MeS^- and (a) DATS or (b) MeS_3A via side-S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

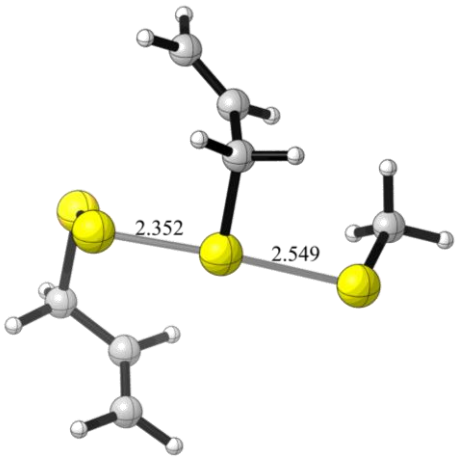
(a)

No.	Name	Conformation κ_3, χ_2, χ_1; κ_2, κ_1	ΔH_{298}
1	TS-SS-MeS-G-DATS1	G(+)G(+)+; G(−)+	0.99
2	TS-SS-MeS-G'-DATS1	G(+)G(+)−; G(−)+	0.00

(b)

No.	Name	Conformation χ_1; κ_1	ΔH_{298}
1	TS-SS-MeS-G-MeS_3A1	+; +	1.01
2	TS-SS-MeS-G'-MeS_3A1	−; +	0.08
3	TS-SS-MeS-G-MeS_3A2	+; +	0.00
4	TS-SS-MeS-G'-MeS_3A2	−; +	0.36

(a) TS-SS-MeS-G'-DATS1
(χ_2, χ_1; κ_3, κ_2, κ_1; τ_1, τ_2, τ_3):
(105.4, 64.6; 119.2, -74.2, 96.0; 89.8, -66.0, 114.8)
S_{lg}-Cα-S_{nuc}: 176.7



(b) TS-SS-MeS-G-MeS_3A2
(χ_1; κ_1; τ_1, τ_2, τ_3):
(65.8; 68.7; 75.8, 68.2, -111.7)
S_{lg}-Cα-S_{nuc}: 173.2

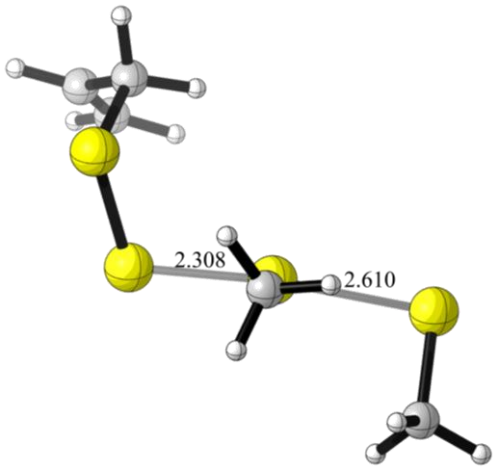


Figure 6.7 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DATS, or (b) MeS_3A via side-S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in °.

By considering the two possible values of χ_1 (±90°), two TS conformations were studied from the reaction of MeS^- and DATS1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.3(a) with the TS names, conformations, and their relative enthalpies. The conformation

of the forming MeS···SA, described by (κ_3, χ_2, χ_1), has the order of G(+)G(+)+|G(–)G(–)− > G(+)G(+)−|G(–)G(–)+ in ΔH_{298}. The same order was observed for the conformers of MeSSA based on ΔH_{298} in section 5.3.1. Overall, TS-SS-MeS-G'-DATS1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.7(a).

From manual inspection of TS in Table 6.3(a), τ_3, τ_2 and τ_1 only deviate a little from that in DATS1. The differences are 2.2° in τ_3, -0.8°–-1° in τ_2 and -1.7°–0.2° in τ_1. Thus, the change in these three torsional angles did not vary their torsional configurations as compared to DATS1.

Similarly, by considering the two possible values of χ_1 (±90°) and the two lowest energy conformers of MeS_3A, four TS state conformations were studied from the reaction of MeS^- and MeS_3A1 or MeS_3A2 at the SMD/M06-2X/6-31+G* level. The results are reported in Table 6.3(b) with the TS names, conformations, and their relative energies. TS-SS-MeS-G-MeS_3A2 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.7(b).

From manual inspection of TS in Table 6.3(b), τ_1 deviates the least (-2.2°–1.1°) from that in the corresponding MeS_3A1 or MeS_3A2, followed by τ_2 (-2.5°–5.9°) and τ_3 (-1.5°–7.4°). Nonetheless, the change in these three torsional angles did not vary their torsional configurations as compared to MeS_3A1 or MeS_3A2.

From Figure 6.7, the geometric features of TS-SS-MeS-G'-DATS1 and TS-SS-MeS-G-MeS_3A2 all agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking S_{side}-S_{lg}

bond (2.308−2.352 Å) that is longer than the S-S bond in DATS1 or MeS_3A2 (2.067−2.069 Å), and the other corresponds to the forming S_{side}-S_{nuc} bond (2.549−2.610 Å) that is longer than the S-S bond in MeSSA1 or DMDS (2.063–2.065 Å). The S_{side} is tricoordinate just like in the S nucleophilic substitution and the nucleophile MeS^- attacks the backside of the disulfide linkage at an angle of 173.2−176.7° and causes an inversion at S_{side}.

By examining the torsional angles, χ_2 and κ_2 are different from τ_2 and have values close to ±90°+15° with the opposite signs in TS-SS-MeS-G'-DATS1 due to the backside attack. χ_1 does not deviate a lot from τ_1. It has values close to -65° in TS-SS-MeS-G'-DATS1, and close to 65° in TS-SS-MeS-G-MeS_3A2. κ_3 is increased by 6.6° in TS-SS-MeS-G'-DATS1 from the corresponding τ_3 (112.6°) in DATS1. κ_1 is increased by 4.5° in TS-SS-MeS-G'-DATS1 from the corresponding τ_1 (91.5°) in DATS1, while it is decreased by 11.3° in TS-SS-MeS-G-MeS_3A2 from the corresponding τ_1 (88°) in MeSSA2.

TS-SS-MeS-G'-DATS1 has the substituent groups (R_1, R_3) at the two sides pointing to the same direction, similar to that observed in the S nucleophilic substitution reported earlier (Figure 6.5A1-1)–1-2)). However, TS-SS-MeS-G-MeS_3A2 has the substituent groups (R_1, R_3) pointing to the opposite directions (Figure 6.5A2-1)–2-2)).

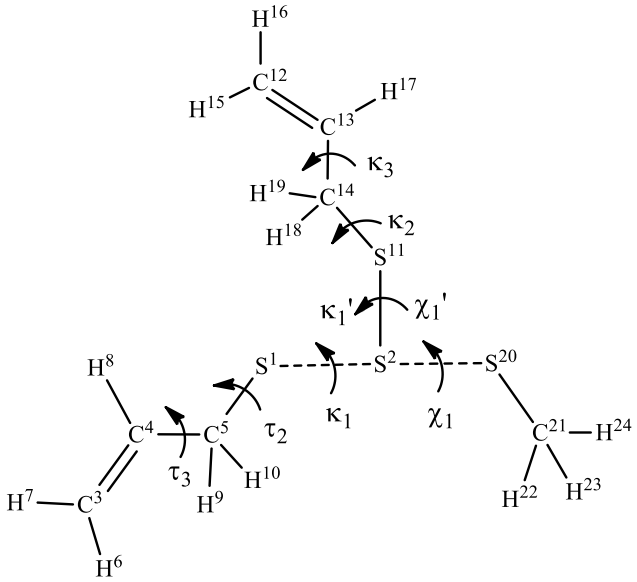


Figure 6.8 Chemical structure of the TS in reaction of MeS^- and DATS via mid-S nucleophilic substitution with labeling of atoms and torsional angles.

To build the transition states in reaction of MeS^- and DATS (Table 5.1, reaction 13) via mid-S nucleophilic substitution, χ_1, χ_1', κ_1, κ_1', κ_2 and κ_3 were defined as described in section 6.1. The torsional angles and atom numbering in the TS model are depicted in Figure 6.8.

Table 6.4 Calculated TS conformations from the reaction of MeS^- and DATS via mid-S nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

No.	Name	Conformation κ_3, κ_2, χ_1', χ_1; κ_1', κ_1	ΔH_{298}
1	TS-MS-MeS-G-DATS1	G(+)G(−)−+; ++	3.04
2	TS-MS-MeS-G'-DATS1	G(+)G(−)−−; ++	0.00

TS-MS-MeS-G'-DATS1
(χ_1, χ_1'; κ_3, κ_2, κ_1, κ_1'; τ_2, τ_3):
(-78.2, -74.7; 112.8, -69.9, 107.1, 92.8; -72.0, 114.0)
S_{lg}-Cα-S_{nuc}: 168.2

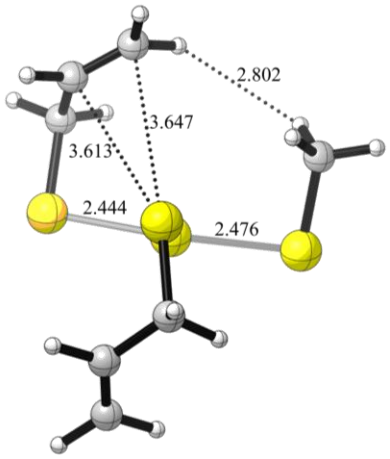


Figure 6.9 SMD/M06-2X/6-31+G* optimized geometry of the lowest energy TS in reaction of MeS^- and DATS via mid-S nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

By considering the two possible values of χ_1 (±90°), two TS conformations were studied from the reaction of MeS^- and DATS1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.4 with the TS names, conformations, and their relative enthalpies. The conformation of the forming MeS···SSA, described by (κ_3, κ_2, χ_1', χ_1), has the order of G(+)G(−)−−|G(−)G(+)++ > G(+)G(−)−+|G(−)G(+)+− in ΔH_{298}. The same order was observed for the conformers of MeS_3A in ΔH_{298} in Table S 8.5(h). Overall, TS-MS-MeS-G'-DATS1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.9.

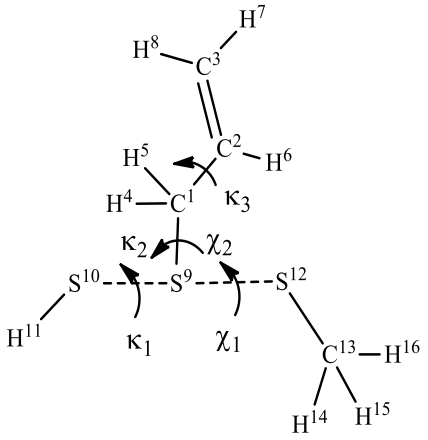
From manual inspection of TS in Table 6.4, τ_3 deviates the least (1.4−2.2°) from that in DATS1, followed by τ_2 (-8.4−1.6°). Nevertheless, the change in these two torsional angles did not vary their torsional configurations as compared to DATS1.

The geometric feature of TS-MS-MeS-G'-DATS1 agrees with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking S_{mid}-S_{lg} bond (2.444 Å) that is longer than the S-S bond in DATS1 (2.067 Å), and the other corresponds to the forming S_{mid}-S_{nuc} bond (2.476 Å) that is longer than the S-S bond in MeS_3A1 (2.066Å). The S_{mid} is tricoordinate just like in the S nucleophilic substitution, and the nucleophile MeS^- attacks the backside of the disulfide linkage at an angle of 168.2° and causes an inversion at S_{mid}.

By examining the torsional angles, χ_1' and κ_1' are different from τ_1 and have values close to ±90°+15° with the opposite signs due to the backside attack. χ_1 does not deviate a lot from τ_1 and has values close to −80°. κ_3 is increased by 0.2° from the corresponding τ_3 (112.6°) in DATS1, κ_2 is decreased by 4.7° from the corresponding τ_2 (-65.2°) in DATS1, and κ_1 is increased by 1.3° from the corresponding τ_1 (91.5°) in DATS1.

TS-MS-MeS-G'-DATS1 has one set of close dihydrogen contact (2.802 Å) between the methyl-hydrogen and the hydrogen on the -CH=CH_2 group. It also has one possible C-S···π as indicated by the moderate distances (3.613 Å and 3.647 Å) between the side-S and the two sp^2 carbons of the allyl group next to it. This interaction was not observed in the reactant DATS1 and is unique to this transition state. Furthermore, TS-MS-MeS-G'-DATS1 has the substituent groups (R_1, R_3) at the two sides pointing to the same direction (Figure 6.5A1-1)–1-2)).

(a)



(b)

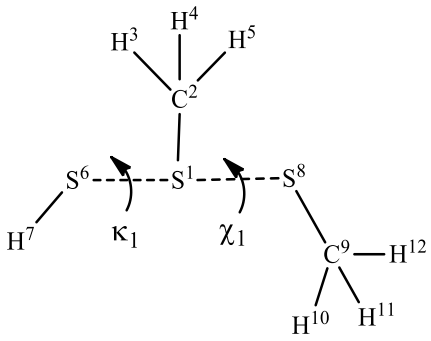


Figure 6.10 Chemical structure of the TS in reaction of MeS^- and (a) ASSH or (b) MeSSH to release H_2S with labeling of atoms and torsional angles.

The transition states in reaction of MeS^- and ASSH (Table 5.1, reaction 9) or MeSSH (Table 5.1, reaction 11) in H_2S release resemble that in reaction of MeS^- with DADS or MeSSA via S nucleophilic substitution. The only difference is that the allyl group at the unreacted side of DADS or MeSSA is replaced by a H. Therefore, τ_1 and τ_2 are absent in the transition states. The torsional angles and atom numbering in the TS models are depicted in Figure 6.10(a)−(b).

Table 6.5 Calculated TS conformations from the reaction of MeS^- and (a) ASSH or (b) MeSSH to release H_2S and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

(a)

No.	Name	Conformation $\kappa_3, \chi_2, \chi_1; \kappa_2, \kappa_1$	ΔH_{298}
1	TS-S-MeS-G-ASSH1	G(+)G(+)+; G(−)+	0.75
2	TS-S-MeS-G'-ASSH1	G(+)G(+)−; G(−)+	0.00

(b)

No.	Name	Conformation $\chi_1; \kappa_1$	ΔH_{298}
1	TS-S-MeS-G-MeSSH	+; +	0.42
2	TS-S-MeS-G'-MeSSH	−; +	0.00

(a) TS-S-MeS-G'-ASSH
($\chi_2, \chi_1; \kappa_3, \kappa_2, \kappa_1$):
(106.1, -60.4; 116.3, -74.5, 68.8)
S_{lg}-Cα-S_{nuc}: 176.8

(b) TS-S-MeS-G'-MeSSH
($\chi_1; \kappa_1$):
(-67.8; 79.7)
S_{lg}-Cα-S_{nuc}: 176.6

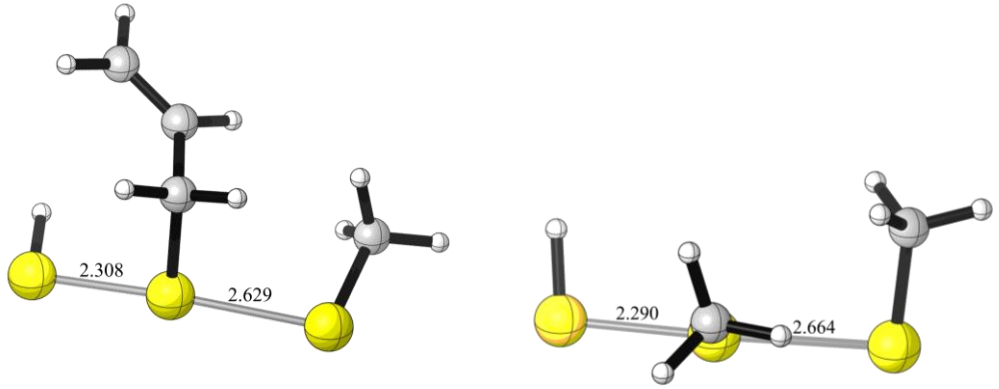


Figure 6.11 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) ASSH, or (b) MeSSH to release H_2S. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in °.

By considering the two possible values of χ_1 (±90°), two TS conformations were studied from the reaction of MeS^- and ASSH1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.5(a) with the TS names, conformations, and their relative enthalpies. The conformation of the forming MeS···SA, described by (κ_3, χ_2, χ_1), has the order of G(+)G(+)−|G(–)G(–)+ > G(+)G(+)+|G(–)G(−)− in ΔH_{298}. TS-S-MeS-G'-

ASSH1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.11(a).

Similarly, by considering the two possible values of χ_1 ($\pm 90^o$), two TS conformations were studied from the reaction of MeS^- and MeSSH at the SMD/M06-2X/6-31+G* level. The results are reported in Table 6.5(b) with the TS names, conformations, and their relative energies. TS-S-MeS-G'-MeSSH is the lowest energy transition state in ΔH_{298} and its optimized geometry are shown in Figure 6.11(b).

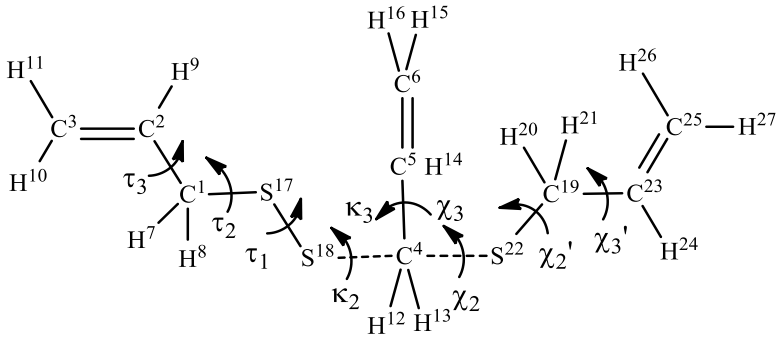
From Figure 6.11, the geometric features of TS-S-MeS-G'-ASSH1 and TS-S-MeS-G'-MeSSH all agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking $S\text{-}S_{lg}$ bond (2.290−2.308 Å) that is longer than the S-S bond in ASSH1 or DMDS (2.065−2.069 Å), and the other corresponds to the forming $S\text{-}S_{nuc}$ bond (2.629−2.664 Å) that is longer than the S-S bond in MeSSA1 or DMDS (2.063–2.065 Å). The S is tricoordinate, and the nucleophile MeS^- attacks the backside of the disulfide linkage at an angle of 176.6−176.8^o and causes an inversion at S.

By examining the torsional angles, χ_2 and κ_2 are different from τ_2 and have values close to $\pm 90^o+15^o$ with the opposite signs in TS-S-MeS-G'-ASSH1 due to the backside attack. χ_1 does not deviate a lot from τ_1. It has values close to -60^o in TS-S-MeS-G'-ASSH1, and close to -70^o in TS-S-MeS-G'-MeSSH. κ_3 is increased by 6.6^o in TS-S-MeS-G'-ASSH1 from the corresponding τ_3 (109.7^o) in ASSH1. κ_1 is decreased by 10.5^o in TS-S-MeS-

G'-ASSH1 from the corresponding τ_1 (79.3°) in ASSH1, and decreased by 5.0° in TS-S-MeS-G'-MeSSH from corresponding τ_1 (84.7°) in DMDS.

TS-S-MeS-G'-ASSH1 and TS-S-MeS-G'-MeSSH have the substituent groups (R_1, R_3) at the two sides pointing to the same direction, similar to that observed in the S nucleophilic substitution reported earlier (Figure 6.5A1-1)–1-2)).

(a)



(b)

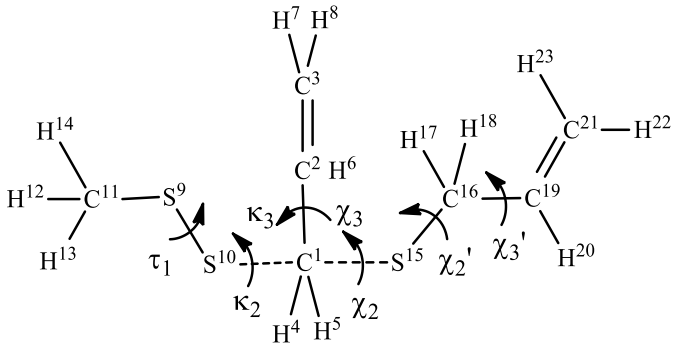


Figure 6.12 Chemical structure of the TS in reaction of AS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution with labeling of atoms and torsional angles.

To construct the transition states in reaction of AS^- and DADS (Table 5.1, reaction 4) or MeSSA (Table 5.1, reaction 7) via Cα nucleophilic substitution, χ_2, χ_2', χ_3, χ_3', κ_2 and κ_3 were defined as described in section 6.1. The torsional angles and atom numbering in the TS models are depicted in

Figure 6.12(a)−(b). Similar as before, for the reaction of MeS^- and DADS, the atom numbering in Figure 6.12(a) is based on the nucleophilic attack on the Cα in the -SS-allyl group denoted by G(+)G(−)−, and it will changed accordingly for the nucleophile attack on the Cα in the -SS-allyl group denoted by G(−)G(+)−. Again, "o" will be added to the end of the DADS1 to indicate the TS from the latter case.

Based on the analysis of TS conformations from the reaction of MeS^- and DADS or MeSSA via Cα nucleophilic substitution, it was proposed that the conformation of the forming AS···A in the transition states from the reaction of AS^- and DADS or MeSSA is more likely to resemble that in the product DAS. TS conformations with the conformations of the forming AS···A similar to the top 5 conformations of DAS with both ΔH_{298} and ΔG_{298} less than 9 kJ mol^{-1}, i.e. DAS1, DAS2, DAS3, DAS4 or DAS5 and their corresponding enantiomers (Table S 8.5(a)), were considered and optimized. In addition, only MeSSA1 was used to construct the transition states from this type of reaction similar to the transition state study in reaction of MeS^- and MeSSA via Cα nucleophilic substitution (Figure 1.5(b)).

Table 6.6 Calculated TS conformations from the reaction of AS^- and (a) DADS or (b) MeSSA via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})
(a)

No.	Name	Conformation χ_3, χ_2, χ_2', χ_3'; κ_3, κ_2	ΔH_{298}
1	TS-C-AS-GG'G'-DADS1[a]	G(+)G(−)G(−)G(+); G(−)G(+)	4.93
2	TS-C-AS-G'G'G'-DADS1	G(+)G(−)G(−)G(−);G(−)G(+)	8.46
3	TS-C-AS-G'GG-DADS1	G(+)G(+)G(+)G(−);G(−)G(+)	12.87
4	TS-C-AS-GGG'-DADS1	G(+)G(−)G(+)G(+);G(−)G(+)	10.23
5	TS-C-AS-GG'G-DADS1[b]	G(+)G(+)G(−)G(+);G(−)G(+)	13.30
6	TS-C-AS-G'GA-DADS1	G(+)antiG(+)G(−);G(−)G(+)	10.53
7	TS-C-AS-G'AG'-DADS1	G(+)G(−)antiG(−);G(−)G(+)	11.91

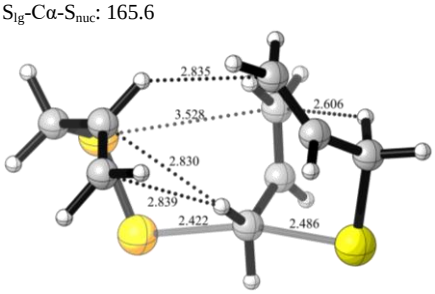
8	TS-C-AS-GG'A-DADS1[c]	G(+)antiG(−)G(+);G(−)G(+)	8.94
9	TS-C-AS-GAG'-DADS1	G(+)G(−)antiG(+);G(−)G(+)	11.50
10	TS-C-AS-G'GG-DADS1o	G(−)G(+)G(+)G(−);G(+)G(−)	0.00
11	TS-C-AS-GGG-DADS1o	G(−)G(+)G(+)G(+);G(+)G(−)	4.51
12	TS-C-AS-GG'G'-DADS1o	G(−)G(−)G(−)G(+);G(+)G(−)	12.56
13	TS-C-AS-G'G'G-DADS1o[d]	G(−)antiG(−)G(−);G(+)G(−)	12.02
14	TS-C-AS-G'GG'-DADS1o	G(−)G(−)G(+)G(−);G(+)G(−)	16.60
15	TS-C-AS-GG'A-DADS1o	G(−)antiG(−)G(+);G(+)G(−)	4.12
16	TS-C-AS-GAG-DADS1o	G(−)G(+)antiG(+);G(+)G(−)	9.19
17	TS-C-AS-G'GA-DADS1o	G(−)antiG(+)G(−);G(+)G(−)	7.95
18	TS-C-AS-G'AG-DADS1o	G(−)G(+)antiG(−);G(+)G(−)	9.24

(b)

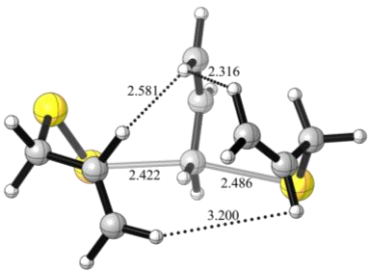
No.	Name	Conformation χ_3, χ_2, χ_2', χ_3'; κ_3, κ_2	ΔH_{298}
1	TS-C-AS-G'GG-MeSSA1	G(−)G(+)G(+)G(−); G(+)G(−)	0.00
2	TS-C-AS-GGG-MeSSA1	G(−)G(+)G(+)G(+); G(+)G(−)	1.26
3	TS-C-AS-GG'G'-MeSSA1	G(−)G(−)G(−)G(+); G(+)G(−)	4.54
4	TS-C-AS-G'G'G-MeSSA1	G(−)G(+)G(−)G(−); G(+)G(−)	2.21
5	TS-C-AS-G'GG'-MeSSA1[e]	G(−)G(−)G(+)G(−); G(+)G(−)	2.54
6	TS-C-AS-GG'A-MeSSA1	G(−)antiG(−)G(+); G(+)G(−)	3.80
7	TS-C-AS-GAG-MeSSA1	G(−)G(+)antiG(+); G(+)G(−)	4.36
8	TS-C-AS-G'AG-MeSSA1	G(−)G(+)antiG(−); G(+)G(−)	3.51
9	TS-C-AS-G'GA-MeSSA1	G(−)antiG(+)G(−); G(+)G(−)	4.49

[a] TS-C-AS-GG'G'-DADS1 has two imaginary frequencies at 497.72 cm^{-1} and 17.86 cm^{-1} respectively.
[b] χ_2 changed from 60.0° to 108.2° after optimization.
[c] TS-C-AS-GG'A-DADS1 has two imaginary frequencies at 516.99 cm^{-1} and 4.70 cm^{-1} respectively.
[d] χ_2 changed from 60.0° to 173.3° after optimization, so the G'G'G notation should have changed to G'G'A.
[e] χ_2 changed from -60.0° to -112.3° after optimization.

a) TS-C-AS-G'GG-DADS1o
(χ_3, χ_2, χ_2', χ_3'; κ_3, κ_2; τ_1, τ_2, τ_3):
(-94.7, 63.4, 72.4, -125.2; 90.8, -58.8; -84.4, 65.5, -114.6)
S_{lg}-Cα-S_{nuc}: 165.6



(View 1)



(View 2)

(b) TS-C-AS-G'GG-MeSSA1
(χ_3, χ_2, χ_2', χ_3'; κ_3, κ_2; τ_1):
(-97.4, 69.8, 75.3, -117.6; 87.5, -40.6; 91.0)
S_{lg}-Cα-S_{nuc}: 166.2

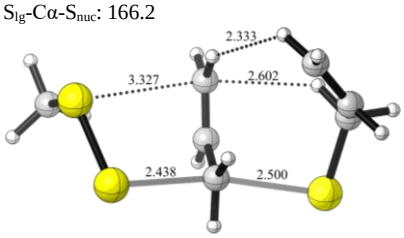


Figure 6.13 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of AS^- and (a) DADS, or (b) MeSSA via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

By considering the top 5 conformations of DAS (and their enantiomers) and the nucleophilic attack on the either side of DADS1, 18 TS conformations were studied from the reaction of AS^- and DADS1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.6(a) with the TS names, conformations, and their relative enthalpies. The transition states are named after the reaction type (C), the nucleophile (AS), the electrophile (DADS1 or DADS1o), and the simplified (χ_3', χ_2', χ_2) notation in which A, G and G' corresponds to the "anti", "G(+)" and "G(−)" configurations respectively.

Out of these 18 SMD/M06-2X/6-31+G* optimized conformations, TS-C-AS-GG'G'-DADS1 and TS-C-AS-GG'A-DADS1 are not true TS because each has two imaginary frequencies. Both transition states have one imaginary frequency around 500 cm^{-1} (497.72 cm^{-1} or 516.99 cm^{-1}) that describes the reaction coordinate $S_{lg}\cdots C\alpha\cdots S_{nuc}$ and is characteristic in all transition states from the Cα nucleophilic substitution. TS-C-AS-GG'G'-DADS1 has another imaginary frequency at 17.86 cm^{-1} that describes the rotation around κ_2. Similarly, TS-C-AS-GG'A-DADS1 has another imaginary frequency at 4.70 cm^{-1} that describes the combination of rotations around κ_2 and χ_2'. Our

calculation failed to eliminate the additional imaginary frequency in both cases even after several attempts probably due to the relatively flat PES around the true TS. Consequently, the conformations and energies of TS-C-AS-GG'G'-DADS1 and TS-C-AS-GG'A-DADS1 will not be analyzed and discussed.

From Table 6.6(a), TS conformations with ΔH_{298} or ΔG_{298} less than 9 kJ mol^{-1} were examined. The conformation of the forming AS···A, described by (χ_3, χ_2, χ_2', χ_3'), has the order of G(+)G(–)G(–)G(+)|G(–)G(+)G(+)G(–) > G(+)antiG(+)G(–)|G(–)antiG(–)G(+)/G(+)G(+)G(+)G(–)|G(–)G(–)G(–)G(+)/G(+)antiG(–)G(+)|G(–)antiG(+)G(–) in ΔH_{298}. G(+)G(–)G(–)G(+)|G(–)G(+)G(+)G(–) is most preferred conformation of the forming AS···A in ΔH_{298} and the same preference was observed for the conformers of DAS based on ΔH_{298} in section 5.3.1. For the other conformations of DAS, they have similar energies (within 1.8 kJ mol^{-1} in ΔH_{298}), and the order in the corresponding conformations of the AS···A did not differ much based on ΔH_{298}. By excluding TS-C-AS-GG'G'-DADS1 and TS-C-AS-GG'A-DADS1, TS-C-AS-G'GG-DADS1o is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.13(a).

From manual inspection of TS in Table 6.6(a), τ_3 in deviates the least (-2.5−2.6°) from that in DADS1, followed by τ_2 (-4.8−7.5°) and τ_1 (-3.7−10.1°). Nevertheless, the change in these three torsional angles did not vary their torsional configurations as compared to DADS1. Moreover, χ_2 can deviate a lot from the expected values in some of the high energy transition states such as TS-C-AS-GG'G-DADS1 and TS-C-AS-G'G'G-DADS1o possibly to avoid the steric repulsion between the allyl group on the nucleophile and the allyl

group next to it. It was also noted that TS-C-AS-GG'G'-DADS1 and TS-C-AS-G'G'G-DADS1o have the same conformation of the forming AS···A—G(+)G(+)G(−)G(+)|G(–)G(–)G(+)G(–).

Similarly, by considering the top 5 conformations of DAS (and their enantiomers), nine TS conformations were studied from the reaction of AS^- and MeSSA1 at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.6(b) with the TS names, conformations, and their relative energies. χ_2 can deviate a lot from the expected values in some of the high energy transition states such as TS-C-AS-G'GG'-MeSSA1 (ranked 4 in ΔH_{298}) possibly to avoid the steric repulsion between the allyl group on the nucleophile and allyl group next to it. It is worth noting that TS-C-AS-G'GG'-MeSSA1 has the same conformation of the forming AS···A as TS-C-AS-GG'G-DADS1 and TS-C-AS-G'G'G-DADS1o.

From Table 6.6(b), the conformation of the forming AS···A, described by (χ_3, χ_2, χ_2', χ_3'), usually has the order of G(+)G(–)G(–)G(+)|G(–)G(+)G(+)G(–) > G(+)G(+)G(+)G(–)|G(–)G(–)G(–)G(+) > G(+)G(+)G(–)G(+)|G(–)G(–)G(+)G(–) > G(+)antiG(+)G(–)|G(–)antiG(–)G(+)/G(+)antiG(–)G(+)|G(–)antiG(+)G(–) in ΔH_{298}. This order for the AS···A group in ΔH_{298} almost matches that for the conformers of DAS based on ΔH_{298}. Overall, TS-C-AS-G'GG-MeSSA1 is the lowest energy transition state in ΔH_{298}, and its optimized geometry is shown in Figure 6.13(b).

From manual inspection of TS in Table 6.6(b), τ_1 deviates by -2.4°–1.9° from that in MeSSA1. Such changes did not vary its torsional configurations as compared to MeSSA1.

From Figure 6.13, TS-C-AS-G'GG-DADS1o and TS-C-AS-G'GG-MeSSA1 have the identical conformation of the forming AS···A—G(+)G(–)G(–)G(+)|G(–)G(+)G(+)G(–)—such that the forming AS···A resembles the most stable form of DAS in Figure 5.4A1). The geometric features of TS-C-AS-G'GG-DADS1o and TS-C-AS-G'GG-MeSSA1 all agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking Cα-S_{lg} bond (2.422−2.438 Å) that is longer than the C-S bond in DADS1 or MeSSA1 (1.841−1.843 Å), and the other corresponds to the forming Cα-S_{nuc} bond (2.486−2.500 Å) that is longer than the C-S bond in MeSA1 (1.831 Å). The Cα is pentacoordinate and sp^2-hybridized. The nucleophile AS^- attacks the backside of Cα at an angle of 165.6−166.2° and causes an inversion at Cα.

By examining the torsional angles, χ_3 and κ_3 are different from τ_3 and have values close to -97−-95° or 87−91° with the opposite signs in all transition states due to the backside attack. χ_2 is similar to τ_2 and has values close to 63−70°. χ_2' is similar to τ_2 and has values close to 72−76°. χ_3' is also similar to τ_3 and has values close to -126−-117°. κ_2 is increased by 8.0° in TS-C-AS-G'GG-DADS1o from the corresponding τ_2 (-66.8°) in DADS1, and it is increased by 22.8° in TS-C-AS-G'GG-MeSSA1 from the corresponding τ_2 (-63.4°) in MeSSA1.

TS-C-AS-G'GG-DADS1o has three sets of close dihydrogen contact (2.316 Å, 2.581 Å, 3.200 Å) between the hydrogens on the three -CH=CH_2 groups. It also has three possible C-H···π interactions as indicated by the short distance (2.606 Å) between the allyl-hydrogen and the sp^2 carbon of the allyl

group next to it, the short distance (2.835 Å) between the hydrogen on the -$CH=CH_2$ group and the sp^2 carbon of the allyl group nearby, and the short distances (2.830 Å and 2.839 Å) between one allyl-hydrogen and the two sp^2 carbons of the other allyl group. The C-H···π interaction in the last case was also observed in the reactant DADS1 (Figure 5.4). Additionally, it has one possible LP(S)···π interaction (3.528 Å) between S and the sp^2 carbon of the allyl group next to it.

TS-C-AS-G'GG-MeSSA1 has one set of close dihydrogen contact (2.333 Å) between the hydrogens on the two -$CH=CH_2$ groups. It also has one possible C-H···π interaction (2.602 Å) between the allyl-hydrogen and the sp^2 carbon of the allyl group next to it; as well as one possible LP(S)···π interaction (3.327 Å) between S and the sp^2 carbon of the allyl group next to it. These two interaction are also present in TS-C-AS-G'GG-DADS1o as discussed above. Similarly, The C-H···π interaction and LP(S)···π interaction were not identified in the product DAS and are unique to the transition states from the Cα nucleophilic substitution.

To summarize the whole transition state study in this section, two general trends in geometries and non-covalent interactions can be deduced:

(1) in the Cα nucleophilic substitution, the conformation of the forming product—MeS···A or AS···A—usually resembles the most stable form of MeSA or DAS. This implies that some of the resonance stabilizations that contribute to the optimal values of τ_3, τ_2 or τ_1 may be still present in the transition states. Two unique interactions were identified in these transition states: one is the possible C-H(MeS^- or AS^-)···π(-$CH=CH_2$) interaction and the

other is the possible LP(S)(DADS or MeSSA)···π(-CH=CH_2) interaction. These two interactions further stabilize the transition states and will be discussed later.

(2) in the S/side-S/mid-S nucleophilic substitution and H_2S release, at least one of the following observations applies. First, the conformation of the forming product—MeS···SA or MeS···SSA—resembles the most stable form(s) of MeSSA or MeS_3A. This implies that some of the resonance stabilizations that contribute to the optimal values of τ_3, τ_2 or τ_1 may be still present in the transition states. Second, the transition state conformations with the lowest ΔH_{298} usually have the substituent groups (R_1, R_3) at the two sides pointing to the same direction (Figure 6.5A1-1)−1-2)). In addition, in the mid-S nucleophilic substitution, the transition state possibly has one unique C-S(DATS)···π(-CH=CH_2) interaction. It will also be discussed later.

Despite the discovery of the common features of the transition states above, it should be noted that this study is limited because many other low-lying conformations of the electrophilic reactants were not used in the transition state study and the nucleophile GSH was modeled by the simple MeSH.

Table 6.7 Interaction analysis for (a) C-H···π interaction, (b) LP(S)···π interaction and (c) C-S···π interaction in transition states by AIM and NBO[a]

(a) C-H···π

d(C-H···C)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
2.602−2.652	0.0095−0.0101	0.0305−0.0307	π(C=C)→σ*(C-H)	3.80−6.98
			σ(C-H)→π*(C=C)	1.06−1.30
			Total	5.10−8.18

(b) LP(S)···π

d(S···C)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
3.327−3.613	0.0076−0.0107	0.0201−0.0283	LP(2)(S)→π*(C=C)	2.79−4.67

(c) C-S···π

d(C-S···C)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
3.611	0.0058	0.0171	π(C=C)→σ*(S-C)	2.64

[a] Distances (d) are in Å, ρ and $\nabla^2\rho$ are from the AIM analysis (in a.u.), and E(2) is from the NBO analysis (in kJ mol^{-1}).

The unique C-H···π and LP(S)···π interactions in the TS from the Cα nucleophilic substitutions, as well as the C-S···π interaction from the mid-S nucleophilic substitution were studied by the AIM and NBO analyses. The results are summarized in Table 6.7. The more detailed results can be found in Table S 8.13. Only bond paths identified by the AIM analysis were reported.

The C-H···C distances in the C-H···π interactions are all within the sum of van der Waals (vdW) radii for C and H (2.9 Å), and the (C-)S···C distances in the LP(S)···π or the C-S···π interaction are smaller than or close to the sum of vdW radii for S and C (3.5 Å).

All of the three interactions are non-covalent interactions as evidenced by the small positive ρ and positive $\nabla^2\rho$. The strengths of these interactions from the NBO analysis have the order of C-H···π > LP(S)···π > C-S···π. Although the donor-acceptor interactions from the NBO analysis are known to overestimate the binding energies, all of the three interactions are still relatively weak as compared to normal hydrogen bonds (6−30 kJ mol^{-1})[12]. The C-H···π interaction is the result of the donation from π(C=C) to the nearby σ*(C-H) and the back donation from σ(C-H) to π*(C=C), although the back donation is less important in terms of the magnitude. For all of the C-H···π interactions, the values of ρ and $\nabla_2\rho$ are similar to that of a weak hydrogen bond, such as the C-H···O and O-H···π interactions[13]. The LP(2)(S)→π*(C=C) interaction is the result of the 3p-like LP(2)(S) donation to the nearby π*(C=C),

so the S atom acts as an electron donor in this interaction. On the contrary, the C-S···π is the result of the donation from π(C=C) to the nearby σ*(C-S). In this case, the S atom acts as an electron acceptor, so the C-S···π interaction corresponds to a chalcogen bond[14]. In addition, the bond angle C-S-C is (174.9°) is close to the ideal angle of 180° in chalcogen bonds. The C-S···π interaction is similar to the F-S···π interaction reported by Nziko *et al.*[15] based on geometries.

From Table S 8.13, one can see that the energies from the donor-acceptor interactions (E(2)) for the C-H···π or the LP(S)···π interaction are directly correlated with the corresponding electron density (ρ). This is consistent with some earlier studies on the relationship of bond strengths and electron density[8a, 8b].

6.3.2 Energy Profiles from the Reaction of MeSH and DADS/DATS and MeSH

Table 6.8 Activation barriers ($\Delta H^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔH_{298}, kJ mol^{-1}) from the reaction of MeSH and (a) DADS or (b) DATS[a]

(a)

No.	Reaction	Reaction Type	$\Delta H^{\ddagger}_{298}$	ΔH_{298}
1	MeSH deprotonation	Protonation/deprotonation	ND[b]	ND
2	DADS + MeS^- → MeSA + ASS^-	Cα nucleophilic substitution	87.72	-33.65
3	DADS + MeS^- → MeSSA + AS^-	S nucleophilic substitution	19.81	-4.49
4	DADS + AS^- → DAS + ASS^-	Cα nucleophilic substitution	83.38	-29.36
5	MeSSA + MeS^- → MeSA + $MeSS^-$	Cα nucleophilic substitution	89.35	-33.48
6	MeSSA + MeS^- → DMDS + AS^-	S nucleophilic substitution	21.07	-4.28
7	MeSSA + AS^- → DAS + $MeSS^-$	Cα nucleophilic substitution	90.50	-29.19
8	ASS^- protonation	Protonation/deprotonation	ND	ND
9	ASSH + MeS^- → MeSSA + HS^-	H_2S release	1.08	-51.16
10	$MeSS^-$ protonation	Protonation/deprotonation	ND	ND
11	MeSSH + MeS^- → DMDS + HS^-	H_2S release	2.89	-52.91

(b)

No.	Reaction	Reaction Type	$\Delta H^{\ddagger}_{298}$	ΔH_{298}
12	DATS + MeS^- → MeSSA + ASS^-	Side-S nucleophilic substitution	6.01	-3.44

13	DATS + MeS^- → MeS3A + AS^-	Mid-S nucleophilic substitution	21.58	-32.67
14	MeS_3A + MeS^- → DMDS + ASS-	Side-S nucleophilic substitution	4.89	-33.51

[a] Reaction steps that already appeared from the reaction of DADS and MeSH are omitted.
[b] ND, not determined.

The overall energy profiles, involving the activation barriers and energies of reaction, from the reaction of MeSH and DADS/DATS (Table 5.1) are summarized in Table 6.8. Excluding the deprotonation/protonation steps, all of the other reaction steps have been studied at the SMD/M06-2X/6-31+G* level of theory and reported in sections 5.3.1 and 6.3.1.

For the deprotonation/protonation steps, the pKa values of MeSH, ASSH and MeSSH are needed to determine the preferred forms of those thiols or perthiols at the reaction pH of 7.4 and to estimate how difficult the reaction steps are. The pKa of MeSH is 10.33 at 25°C[16], thus MeS^- is the less preferred form at the pH of 7.4 and must be deprotonated to act as the strong nucleophile. The activation barrier ($\Delta G^{\ddagger}_{298}$) of this deprotonation step (reaction 1) is estimated to be 58.96 kJ mol^{-1} based on the relationship of pKa to Gibbs free energy. This barrier is lower than that in the S nucleophilic substitution on DADS ($\Delta G^{\ddagger}_{298}$: 65.95 kJ mol^{-1}) and mid-S nucleophilic substitution on DATS ($\Delta G^{\ddagger}_{298}$: 63.74 kJ mol^{-1}) (Table S 8.14(a)). Therefore, it is not the rate-limiting step in the H_2S releasing reactions. Because the pKa of a perthiol is usually 1–2 units smaller than the corresponding thiol[17], the pKa of MeSSH is estimated to be 8.33–9.33, and the protonation of $MeSS^-$ (reaction 8) should be spontaneous as MeSSH is the preferred form at the pH of 7.4. Similar, the pKa of MeSSH is estimated to be 7.96–8.96 because the

pKa of ASH is 9.96 at ~25°C[18], thus the protonation of ASS^- (reaction 10) should also be spontaneous at the pH of 7.4.

In general, all of the reaction steps in Table 6.8 are exothermic as shown by the negative energies of reaction in ΔH_{298}. Among them, the S/side-S nucleophilic substitutions (reaction 3, 6 and 12) are almost thermoneutral with relatively small activation barriers. Such results are consistent with the literature findings that thiol/disulfide exchange generally occurs readily and reversibly at room temperature in water[11].

The Cα nucleophilic substitutions (reactions 2, 4, 7) are always slower than the S nucleophilic substitutions (reactions 3, 6) from the reaction of MeS^- and DADS by comparing the activation barriers in $\Delta H^{\ddagger}_{298}$. Such results are consistent with the experimental results reported by Liang *et al*.[1] One possible reasoning is that the pentacoordinate Cα centered transition states from the Cα nucleophilic substitution are more sterically crowded than the tricoordinate S centered transition states from the S nucleophilic substitution, and this results in the higher energy transition states in the former case. The Cα centered transition states also require the additional re-hybridization of Cα from sp^3 to sp^2 as compared with the S centered transition states, thus such transition states are higher in energy. Moreover, all of the Cα nucleophilic substitutions have similar energy profiles in terms of the activation barriers and energies of reaction. The same observation applies to the S nucleophilic substitutions.

The mid-S nucleophilic substitution (reaction 13) is slower than the corresponding side-S nucleophilic substitution (reaction 12), and this is consistent with a study on the reaction of a trisulfide calicheamicin γ1 with

GSH reported by Myers *et al.*[19] This is possibly due to the leaving group effect in the S_N2 reactions. In the side-S nucleophilic substitution, the leaving group is ASS^-; whereas in the mid-S nucleophilic substitution, the leaving group is AS^-. ASS^- is a better leaving group than AS^- due to its greater stability, because the negative charge is dispersed onto the two S atoms in ASS^-. Consequently, the side-S nucleophilic substitution is more exothermic than the mid-S nucleophilic substitution, it is also faster than the mid-S nucleophilic substitution. Hence, the nucleophilic substitution of DATS by MeS^- should occur mostly on the terminal S atom to form ASS^- directly in one step (Figure 1.6, pink) instead of two steps (Figure 1.6, blue). Moreover, the mid-S nucleophilic substitution has similar energy profiles as the S nucleophilic substitutions (reactions 2, 4, 7) with the identical leaving group AS^-. The two side-S nucleophilic substitutions (reactions 12 and 14) also have similar energy profiles.

The reaction steps in direct H_2S release (reaction 9 and 11) are highly exothermic unlike the thermoneutral S/side-S nucleophilic substitutions (reactions 3, 6 and 12), but they are faster than those reactions. Such results may be due to the leaving group effects mentioned above.

6.3.3 Effects of Temperature and Basis Sets Tested on the Cα Nucleophilic Substitution of DADS by MeSH

Table 6.9 Temperature effects (in K) on the activation barriers ($\Delta H^‡$ or $\Delta G^‡$, in kJ mol^{-1}) tested on the Cα nucleophilic substitution of DADS by MeSH from single point energy calculations on the SMD/M06-2X/6-31+G* optimized geometries

Reaction	Temperature	$\Delta H^‡$	$\Delta G^‡$
DADS + MeS^- → MeSA + ASS^-	298	87.46	131.64
	310	87.86	133.41

The temperature effects were tested on the Cα nucleophilic substitution of DADS by MeSH and are reported in Table 6.9. This step was chosen because it is the key yet slow step in the H_2S releasing reactions of DADS. One can see that the temperature correction from the room temperature of 25ºC (298 K) to the reaction temperature of 37ºC (310 K) will not significantly change the activation barriers or alter the barrier order (< +0.5 kJ mol^{-1} in $\Delta H^{\ddagger}$, < +1.8 kJ mol^{-1} in $\Delta G^{\ddagger}$).

Table 6.10 Basis set effects on the activation barriers ($\Delta H^{\ddagger}_{298}$ or $\Delta G^{\ddagger}_{298}$, in kJ mol^{-1}) tested on the Cα nucleophilic substitution of DADS by MeSH from single point energy calculations on the SMD/M06-2X/6-31+G* optimized geometries

Reaction	Basis Set	$\Delta H^{\ddagger}_{298}$	$\Delta G^{\ddagger}_{298}$
DADS + MeS^- → MeSA + ASS^-	6-31+G(d)	87.46	131.64
	6-311+G(2d,p)	88.98	133.38
	6-311++G(2d,2p)	88.90	133.69

The basis set effects were tested on the Cα nucleophilic substitution of DADS by MeSH and are reported in Table 6.10. The change to larger basis sets with more splitting, polarization or diffuse functions such as 6-311+G(2d,p) and 6-311++G(2d,2p) will not significantly change the activation barriers (< +0.5 kJ mol^{-1} in $\Delta H^{\ddagger}_{298}$, < +2.1 kJ mol^{-1} in $\Delta G^{\ddagger}_{298}$). This is similar to what was suggested by the benchmarking study on the S nucleophilic substitutions by Neves *et al.*[20] They found that the splitting and polarization functions affect the free energy activation barrier (1.00 kcal mol^{-1} or 4.18 kJ mol^{-1}) more than diffuse functions (0.10 kcal mol^{-1} or 0.42 kJ mol^{-1}), but overall the basis set effect is very small.

6.3.4 Computational Study on the Cα Nucleophilic Substitution of DADS by GSH

Before presenting the transition state study on the Cα nucleophilic substitution of DADS by GSH based on the conformational search described in section 6.2, we need to first examine the conformations of GSH as part of the energy profile studies based on the conformational studies on GSH reported in the literature.

As mentioned in section 5.1.1, GSH is difficult to study computationally because it has several rotatable bonds and does not adopt a strongly preferred conformation at any pH. It worth noting that GSH used in the reaction has the nucleophilic thiolate anion and its ionization state is depicted in Figure 6.14. It is the deprotonated thiolate form from the most abundant form of GSH at pH 7.4 (Figure 5.1(b)).

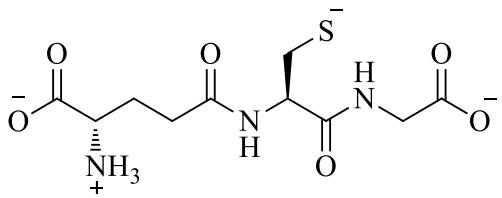


Figure 6.14 The ionization state of GSH in the reaction.

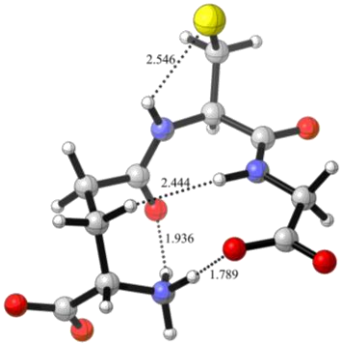
Table 6.11 The calculated GSH conformations in this study and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1})

Name	ΔH_{298}	ΔG_{298}	References
GS1	30.18	24.37	Crystal structure by Wright (1958)[21]
GS2	30.17	24.39	Crystal structure by Görbitz (1987)[22]
GS3	0.00	0.00	PCILO[a] study by Laurence *et al.* (1980)[23]
GS4 (6a)[b]	50.79	42.18	MD study in water with Gromacs all-atom force field by Laurence *et al.* (2003)[24]

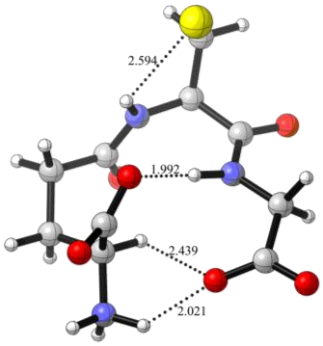
GS5	20.61	23.19	MD study in water with OPLS-AA force field by Zhang *et al.* (2011)[25]
GS6	11.80	9.14	
GS7 (Gly1)	34.81	24.73	QM study at SMD/B3LYP/6-311++G* //B3LYP/6-31+G* by Kurian (2013)[26]
GS8	11.62	13.09	From the TS study in this thesis

[a] PCILO stands for Perturbative Configuration Interaction Procedure using Localized Orbitals.
[b] The names in brackets correspond to the original name of GSH given in the references.

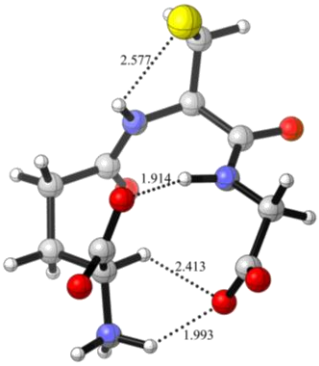
(a) GS3
ΔH_{298}: 0.00 kJ mol^{-1}



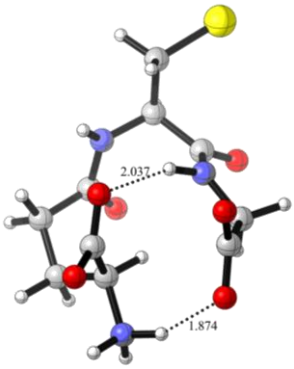
(b) GS8
ΔH_{298}: 11.62 kJ mol^{-1}



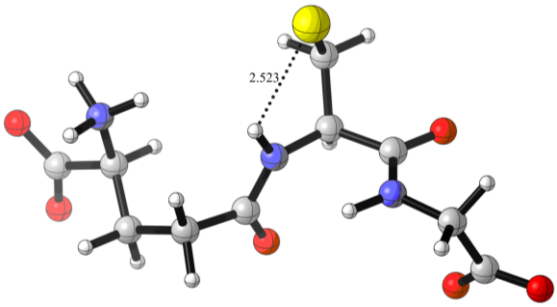
(c) GS6
ΔH_{298}: 11.80 kJ mol^{-1}



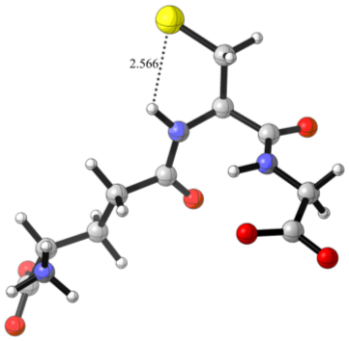
(d) GS5
ΔH_{298}: 20.61 kJ mol^{-1}



(e) GS2
ΔH_{298}: 30.17 kJ mol^{-1}



(f) GS7
ΔH_{298}: 34.81 kJ mol^{-1}



(g) GS4
ΔH_{298}: 50.79 kJ mol^{-1}

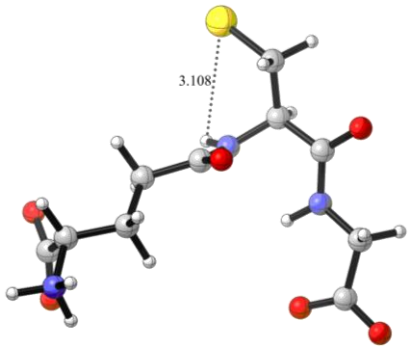


Figure 6.15 SMD/M06-2X/6-31+G* optimized geometries of GSH conformations. Interactions are indicated by the dotted lines and distances are in Å.

In our study, the seven conformations of GSH with the identical ionization state depicted in Figure 6.14, each modified from the low-lying or the lowest energy conformers of GSH from different references, are renamed as GS1–GS7 and were fully optimized at the SMD/M06-2X/6-31+G* level. The results and the corresponding references are summarized in Table 6.11. The last five conformations of GSH in Table 6.11 all have the ionization state depicted in Figure 5.1(b) with a thiol group instead of a thiolate anion before modification. The SMD/M06-2X/6-31+G* optimized geometries are shown in Figure 6.15 with the ascending order in ΔH_{298}. GS1 and GS2 are the experimentally determined crystal structures, while GS3–GS7 are structures from theoretical or computational studies. GS8 is the lowest energy conformation of GSH derived from the transition state study in this section. Because GS1 and GS2 have nearly identical geometries and energies after geometry optimization, only GS2 (0.01 kJ mol^{-1} lower in ΔH_{298}) was considered in the discussion.

All of the optimized conformations of GSH were manually inspected and compared to the input structures. The geometries of GS2, GS3, GS5, GS7 and GS8 match the input structures well, so these five conformations are all local minima at the SMD/M06-2X/6-31+G* level. For GS4, four torsional angles were significantly changed: N-Cα-C-N in the backbone of Cys rotated from -128.9º to 160.5º, N-Cα-C-S in the side chain of Cys rotated from -96.1º to -54.4º, C-N-Cα-C of Gly rotated from 116.9º to 175.5º, and N-Cα-C=O rotated from -85.8º to -6.1º. However, no additional intramolecular interactions were observed after the optimization of GS4. For GS6, two torsional angles were significantly changed: N-Cα-C-N in the backbone bone of Cys rotated from 68.4º to 153.6º, and Cβ-Cγ-C-N in the side chain of Glu rotated from -161.7º to 127.5º. Consequently, the Cα side of Glu gets closer to Gly after geometry optimization to form the interaction between the Cα-NH_3^+ or the Cα-COO^- group of Glu and the Cα-COO^- or the Cα-NH group of Gly, and such interactions closely resemble that in GS8.

From Table 6.11 and Figure 6.15, the conformations of GSH have the order of GS3 > GS8 > GS6 > GS5 > GS2 > GS7 > GS4 in ΔH_{298}, so GS3 is the lowest energy conformation. The geometries of these conformations were examined based on the three amino acid residues: Gly, Cys and Glu. First, the geometry of Gly in GS3 is almost planar in the entire backbone (H-N-Cα-CO_2), and this feature was also observed in GS8, GS7 and GS4. In GS6 and GS5, the CαH_2-COO^- group of Gly rotated slightly but it is still able to interact with other parts of GSH. In GS2, the Cα-COO^- group of Gly rotated away from the Cα-NH group of Gly. Second, the geometry of Cys in GS3 has the

torsional angles C-N-Cα-N, N-Cα-C-N and N-Cα-C-S at -80.3°, 155.9° and 54.8° respectively. This conformation of Cys keeps its Cα-NH group close to the thiolate anion (2.646 Å) and the similar conformations of Cys were observed in GS8, GS6 and GS2. In GS7, the torsion angle N-Cα-C-S changed to -53.1° but the thiolate anion is still close to the Cα-NH group (2.566 Å). In GS4, the torsion angles N-Cα-C-S and C-N-Cα-N changed to -54.4° and 65.6° respectively, so the thiolate anion is further away from the Cα-NH group (3.108 Å). In GS5, the thiolate anion was rotated away from the Cα-NH group. Third, the geometry of Glu can vary a lot in different local minima structures due to its relatively long side chain (-CH_2-CH_2-C=O). In GS3, the rotation of the Glu side chain brings its Cα-NH_3^+ group close to its Cγ-C=O group and the Cα-CO_2^- group of Gly. In GS8, GS6 and GS5, the rotation of the Glu side chain brings its Cα-NH_3^+ group close to the Cα-CO_2^- group of Gly, and its Cα-CO_2^- group close to the Cα-NH group of Gly. In GS2, GS7 and GS4, the rotation of the Glu side chain brings the functional groups of Glu away from that of Gly, thus no intramolecular interactions can be identified between Glu and Gly in these conformations.

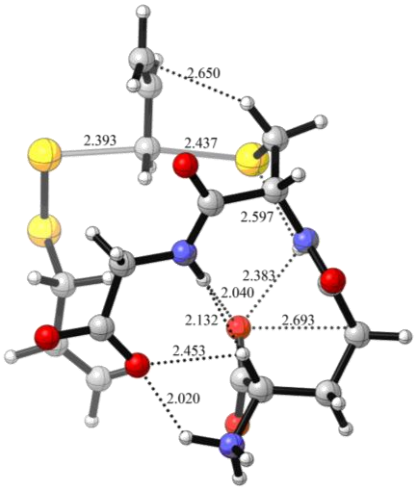
To conclude, the conformations of GSH with multiple intramolecular hydrogen bonds are much lower in energies than those without. The common intramolecular H-bonds are the NH···O interactions (1.789–2.037 Å) in GS3, GS8, GS6 and GS5, the NH···S interactions (2.523–3.108 Å) in all conformations except GS2, and the CH···O interactions (2.413–2.439 Å). In addition, one set of close dihydrogen contact (2.444 Å) was observed in GS3.

Table 6.12 Calculated TS conformations from the reaction of GSH and DADS via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

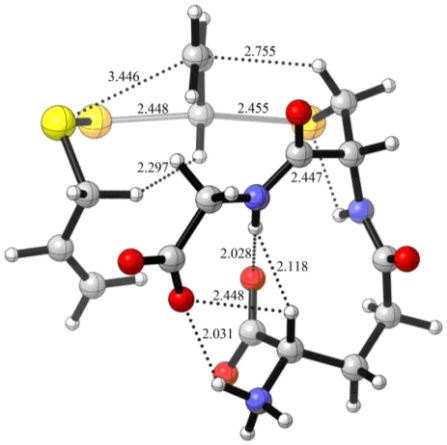
Name	ΔH_{298}
TS1-C-GS-DADS	4.55
TS2-C-GS-DADS	13.13
TS3-C-GS-DADS[a]	ND[b]
TS4-C-GS-DADS[c]	ND
TS5-C-GS-DADS	0.00

[a] TS3-C-GS-DADS changed to the reactant complex after optimization.
[b] ND, not determined.
[c] TS3-C-GS-DADS did not converge after several cycles of optimization.

(a) TS5-C-GS-DADS
S_{lg}-Cα-S_{nuc}: 167.7
ΔH_{298}: 0.00 kJ mol^{-1}



(b) TS1-C-GS-DADS
S_{lg}-Cα-S_{nuc}: 164.8
ΔH_{298}: 4.55 kJ mol^{-1}



(c) TS2-C-GS-DADS
S_{lg}-Cα-S_{nuc}: 173.5
ΔH_{298}: 13.13 kJ mol^{-1}

(d) Superposed GSH in the TS
GSH-TS1: cyan, GSH-TS2: green,
GSH-TS5: orange

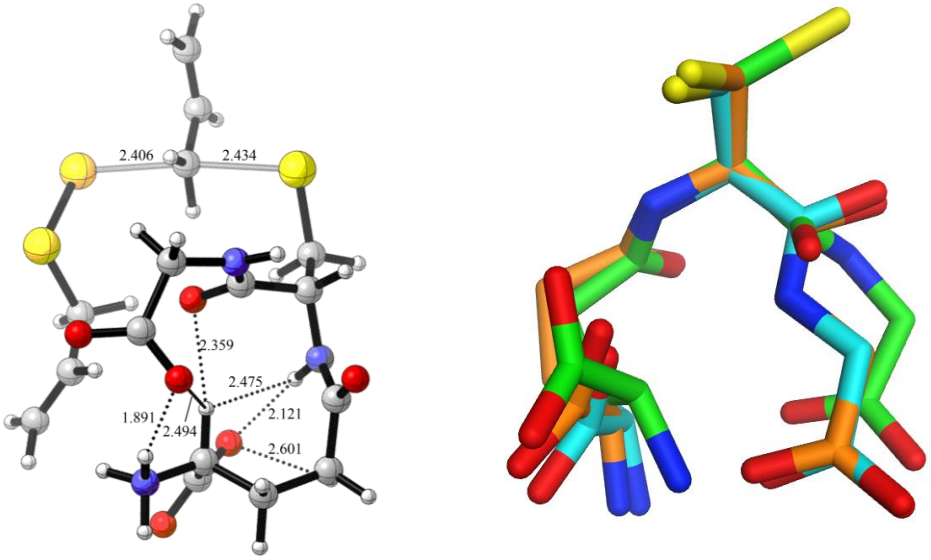


Figure 6.16 SMD/M06-2X/6-31+G* optimized geometries of TS in reaction of GSH and DADS via Cα nucleophilic substitution and the superposition of GSH in these TS with different colorings in carbons. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å, and angles are in °.

As described in section 6.2, a total of five TS conformations were selected and optimized at the SMD/M06-2X/6-31+G* level. Each conformation is the lowest energy conformation from that specific conformational search using the Amber12:EHT[9] force field and the LowModeMD[10] search method. The fixed distances of the breaking Cα-S_{lg} bond (2.45 Å) and the forming Cα-S_{nuc} bond (2.65 Å) were chosen based on the input structures used in the construction of other transition states from the Cα nucleophilic substitution in section 6.3.1. The S-C-S angle was fixed at 180° to resemble the S_N2 transition state geometries.

The TS names and their relative enthalpies (ΔH_{298}) are shown in Table 6.12. The SMD/M06-2X/6-31+G* optimized geometries are shown in Figure 6.16 with the ascending order in ΔH_{298}. The intermolecular interactions

between GSH and DADS not present in the small TS models will be shown and discussed separately later. After geometry optimization, TS3-C-GS-DADS changed to the reactant complex and will not be included in the discussion. TS4-C-GS-DADS did not converge after thousands of SCF cycles and will not be included in the discussion, either.

From Figure 6.16(a)–(c), the geometric features of TS1-C-GS-DADS1, TS2-C-GS-DADS1 and TS5-C-GS-DADS agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking Cα-S_{lg} bond (2.393−2.448 Å), and the other corresponds to the forming Cα-S_{nuc} bond (2.434−2.455 Å). The Cα is pentacoordinate, and almost adopts a trigonal planar geometry considering the three covalent bonds around it. The nucleophile GSH attacks the backside of Cα at an angle of 164.8−173.5^o and causes an inversion at Cα.

To compare these transition states, the GSH conformations and the TS conformations around the reaction center are considered separately.

The GSH conformations, named as GSH-TS1, GSH-TS2 and GSH-TS5, in TS1-C-GS-DADS1, TS2-C-GS-DADS1 and TS5-C-GS-DADS were optimized at the SMD/M06-2X/6-31+G* level with their superposed image shown in Figure 6.16(d). The superposition was carried out by overlaying the N-Cα(-Cβ)-C group of Cys using a Java program written by Dr. Yang Hui. GSH-TS1 and GSH-TS5 are similar to GS8, while GSH-TS2 is different from GS1–GS8. In GSH-TS2, the thiolate anion rotated away from the Cα-NH group in Cys just like GS5. The torsional angle C-Cα-C-N of Gly rotated by 115.1^o from that in GS8, and the torsional angle C-Cγ-Cβ-Cα of Glu rotated

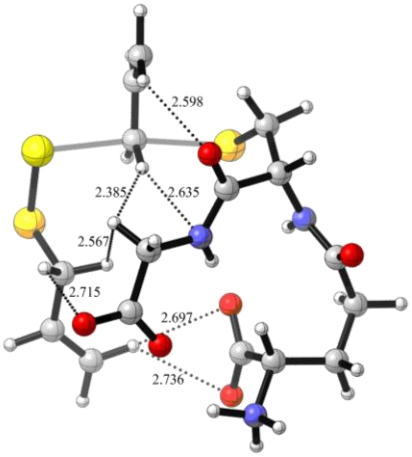
by 7.6º from that in GS8. The Cα-COO^- group of Gly is still close to the Cα-NH_3^+ group of Glu. The three GSH conformers have the order of GSH-TS1/GSH-TS5 > GSH-TS2 in ΔH_{298}, and the difference in energy is only 3.0–3.5 kJ mol^{-1}. From our other studies (results not shown), when MMFFs[27] or OPLS_2005[28] was used in the conformational search, the selected transition states either did not converge after cycles of optimization, or have much higher energies after DFT optimization (> 42 kJ mol^{-1} in ΔH_{298}) when compared with the three transition states presented. MMFFs (Merck Molecular Force Field, static) is an all-atom force field for biopolymers and many organic molecules, and is able to describe the planar amide geometries. OPLS_2005 (Optimized Potential for Liquid Simulations, 2005) is also an all atom force field for biological systems and organic molecules. It is worth noting that in these high energy transition states, the SMD/M06-2X/6-31+G* optimized geometries of their corresponding GSH are also much higher in energy (> 40 kJ mol^{-1} in ΔH_{298}), and the energy differences in the GSH conformations are similar to that in the transition states. Such results suggested that the energies of the GSH conformations in the transition states largely affect the energies of these transition states, and the Amber12:EHT force field is better than MMFFs or OPLS2005 in the conformational search of GSH.

Table 6.13 The comparison between the full TS and the small TS models in terms of their relative enthalpies (ΔH_{298}, kJ mol^{-1})

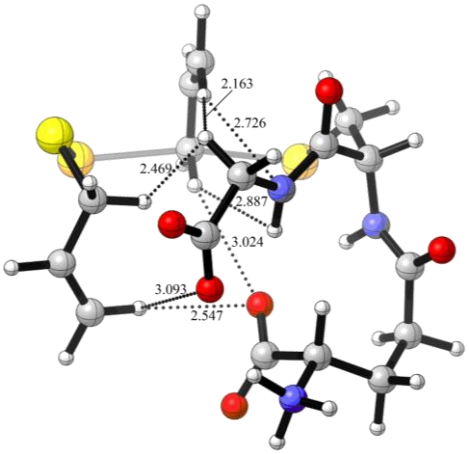
Full TS	ΔH_{298}	Small TS	ΔH_{298}
TS5-C-GS-DADS	0.00	TS-C-MeS-G-DADS12	4.71
TS1-C-GS-DADS	4.55	TS-C-MeS-G-DADS2	0.00
TS2-C-GS-DADS	13.13	TS-C-MeS-A-DADS18	6.60

The TS conformations around the reaction center, i.e. the TS conformation using MeSH as the model compound for GSH, were optimized at the SMD/M06-2X/6-31+G* level. From manual inspection, the optimized geometries of the small TS models are similar to their corresponding full TS models around the reaction center. The energies of the full TS and small TS models are compared and summarized in Table 6.13. The small TS models have the similar order in ΔH_{298} as the full TS models, although the order for the first two transition states from the small model is reversed. Those differences may be due to the additional intermolecular interactions between GSH and DADS. Moreover, TS5-C-GS-DADS and TS1-C-GS-DADS have the unique C-H···π interactions (2.650 Å, 2.755 Å) described in section 6.3.1. This time, the C-H···π interaction is between the GSH-Cys-CβH and the -CH=CH_2 group in DADS. TS1-C-GS-DADS also has the unique LP(S)···π interaction (3.446 Å), which is weaker than the C-H···π interaction as discussed before. One set of close dihydrogen contacts (2.297 Å) was observed in TS1-C-GS-DADS. It is worth noting that TS-C-MeS-G-DADS2 is higher in energy than TS-C-MeS-G-DADS1o presented in section 6.3.1 (≥ +3.62 kJ mol^{-1} in ΔH_{298}).

(a) TS5-C-GS-DADS
ΔH_{298}: 0.00 kJ mol^{-1}



(b) TS1-C-GS-DADS
ΔH_{298}: 4.55 kJ mol^{-1}



(c) TS5-C-GS-DADS CPK model

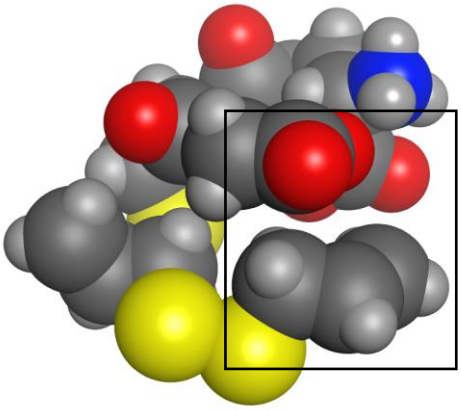


Figure 6.17 Intramolecular interactions in the SMD/M06-2X/6-31+G* optimized geometries of TS in reaction of GSH and DADS via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. The interactions are indicated by the dotted lines. Distances are in Å.

TS5-C-GS-DADS and TS1-C-GS-DADS are all lower in energy than TS2-C-GS-DADS, and have the similar GSH conformations as well as the similar interactions around the reaction center. Hence, the analysis and discussion of the intermolecular interactions between GSH and DADS will be based on these two transition states.

The intermolecular interactions between GSH and DADS in TS5-C-GS-DADS and TS1-C-GS-DADS are shown in Figure 6.17(a)–(b). TS5-C-GS-

DADS has four possible C-H···O interactions (2.598 Å, 2.715 Å, 2.697 Å, 2.736 Å) between the allyl hydrogens and the carboxylate or amide oxygens. TS1-C-GS-DADS has only three possible C-H···O interactions (2.547 Å, 3.024 Å, and 3.093 Å) between the allyl hydrogens and the carboxylate oxygens. In addition, TS5-C-GS-DADS has one possible C-H···N interaction (2.635 Å) between the allyl hydrogen and the amide nitrogen, and TS1-C-GS-DADS has one similar C-H···N interaction with a longer distance (2.726 Å). Some dihydrogen contacts were also observed in these two transition states with close distances (2.163–2.887 Å). Furthermore, the -CH=CH_2 group and the carboxylate group in TS5-C-GS-DADS are in close proximity and almost parallel to each other, so there may be an additional π···π stacking between them as boxed in Figure 6.17(c).

Table 6.14 Interaction analysis for (a) C-H···O interaction, (b) C-H···N interaction, and (c) π···π stacking[a]

(a)

d(C-H···O)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
2.598–2.736	0.0056–0.0074	0.0224–0.0267	π(C=O)→σ*(C-H)	0.67–2.89
			LP(O)→σ*(C-H)	0.67–2.94
			σ(C-H)→π*(C=O)	0.34
			Total	1.35–5.05

(b)

d(C-H···N)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
2.635	0.0083	0.0274	LP(N)→σ*(C-H)	2.70

(c)

d(π···π)[b]	ρ	$\nabla^2\rho$
3.774	0.0045	0.0142

[a] Distances (d) are in Å, ρ and $\nabla^2\rho$ are from the AIM analysis (in a.u.), and E(2) is from the NBO analysis (in kJ mol^{-1}).

[b] d(π···π) is the distance between the midpoint of the C=C bond and the C of the carboxylate.

The C-H···O interactions, the C-H···N interaction and the π···π stacking in the lowest energy transition state TS5-C-GS-DADS were studied by the

AIM and NBO analyses. The results are summarized in Table 6.14. The more detailed results are shown in Table S 8.17.

The C-H···O distances are smaller than or close to the sum of vdW radii for O and H (2.72 Å), and the C-H···N distance is smaller than the sum of vdW radii for O and H (2.75 Å). The π···π distance is measured from the center of the C=C bond to the C of the carboxylate, and its value of 3.774 Å is slightly larger than the sum of vdW radii for 2C's (3.4 Å).

All of the three interaction are non-covalent interactions as evidenced by the small positive ρ and positive $\nabla^2\rho$. For all of the C-H···O and C-H···N interactions, the values of ρ and $\nabla_2\rho$ are similar to that of a weak hydrogen bond. The strengths of the C-H···O interactions can vary from 1.35 kJ mol^{-1} to 5.05 kJ mol^{-1} based on the NBO analysis and they are directly correlated with the corresponding electron density (ρ) as discussed before. The strength of the C-H···N interaction is 2.70 kJ mol^{-1}. Although the donor-acceptor interactions from the NBO analysis are known to overestimate the binding energies, these C-H···O and C-H···N interactions are relatively weak as compared to normal hydrogen bonds (5−30 kJ mol^{-1})[12].

The C-H···O interaction is the result of the donation from π(C=O) or LP(O) to the nearby σ*(C-H) and the possible back donation from σ(C-H) to π*(C=O). The C-H···N interaction is the result of the donation of LP(N) to the nearby σ*(C-H). However, LP(N) is involved in the resonance stabilization of the amide bond, so this interaction is relatively weak.

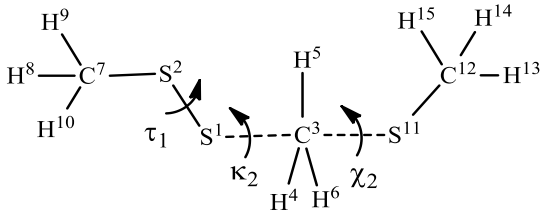
The activation barrier from the full TS model TS5-C-GSH-DADS is 90.91 kJ mol^{-1} in $\Delta H^{\ddagger}_{298}$. As compared to the small TS model, the difference

in $\Delta H^{\ddagger}_{298}$ is small (+3.45 kJ mol^{-1}). Hence, the additional intermolecular interactions between GSH and DADS, and the geometric change in the reaction center did not alter $\Delta H^{\ddagger}_{298}$ much from the small TS model to the full TS model.

In summary, from this preliminary full TS model study, the conformation of DADS in the low-lying transition states can deviate a lot from the lowest energy conformation of the reactant DADS used in the small TS model study, and this is possibly resulted from the intermolecular interactions between GSH and DADS. The unique C-H···π and the LP(S)···π interactions in the small TS models are preserved in some of the low-lying full TS models. The activation barrier from the full TS model is comparable to that from the small TS model based on $\Delta H^{\ddagger}_{298}$.

6.3.5 Transition State Study on the Hypothetical Cα Nucleophilic Substitution of DMDS or DPDS by MeSH

(a)



(b)

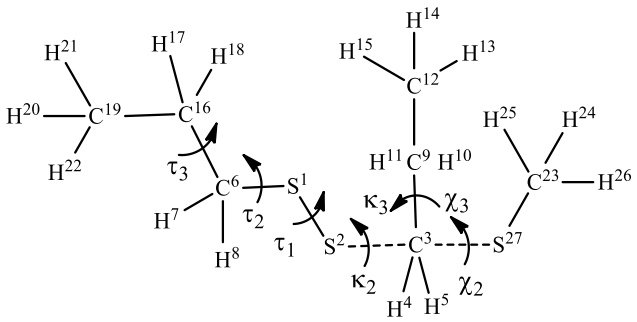


Figure 6.18 Chemical structure of the TS in reaction of MeS^- with (a) DMDS or (b) DPDS via Cα nucleophilic substitution with labeling of atoms and torsional angles.

To construct the transition states in reaction of MeS^- with DMDS or DPDS via Cα nucleophilic substitution, χ_2, χ_3, κ_2 and κ_3 were defined as described in section 6.1. The torsional angles and atom numbering in the TS models are depicted in Figure 6.18(a)−(b). For the reaction of MeS^- and DPDS, the atom numbering in Figure 6.18(b) is based on the nucleophile attack on the Cα of DPDS1 in the -S-propyl group denoted by G(−)G(−), and it will be changed accordingly for the nucleophile attack of DPDS7.

Table 6.15 Calculated TS conformations from the reaction of MeS^- and (a) DMDS or (b) DPDS via Cα nucleophilic substitution and their relative enthalpies (ΔH_{298}, kJ mol^{-1})

(a)

No.	Name	Conformation χ_2; κ_2	ΔH_{298}
1	TS-C-MeS-A-DMDS	anti; G(−)	0.98
2	TS-C-MeS-G-DMDS	G(+); G(−)	0.00
3	TS-C-MeS-G'-DMDS	G(−); G(−)	2.02

(b)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔH_{298}
1	TS-C-MeS-A-DPDS1	G(+)anti; G(−)G(−)	0.00
2	TS-C-MeS-A-DPDS7	G(−)anti; G(+)G(+)	3.26
3	TS-C-MeS-G-DPDS7	G(−)G(+); G(+)G(+)	7.22
4	TS-C-MeS-G'-DPDS7	G(−)G(−); G(+)G(+)	4.19

(a) TS-C-MeS-G-DMDS
(χ_2; κ_2; τ_1):
(62.4; -59.1; 84.1)
S_{lg}-Cα-S_{nuc}: 174.6

(b) TS-C-MeS-A-DPDS1
(χ_3, χ_2; κ_3, κ_2; τ_1, τ_2, τ_3):
(82.2, -169.4; -82.7, -39.0; 87.6, 63.9, -175.9)
S_{lg}-Cα-S_{nuc}: 161.6

Figure 6.19 SMD/M06-2X/6-31+G* optimized geometries of the lowest energy TS in reaction of MeS^- and (a) DMDS, or (b) DPDS via Cα nucleophilic substitution. The breaking/forming bonds are indicted by the semi-transparent lines. Distances are in Å, and angles are in °.

By considering the three possible values of χ_2 (±60° and 180°), three TS conformations were studied from the reaction of MeS^- and DMDS at the SMD/M06-2X/6-31+G* level and the results are reported in Table 6.12(a) with the TS names, conformations, and their relative enthalpies. The transition states are named after the reaction type (C), the nucleophile MeS^- (MeS), the electrophile (DMDS) and the simplified χ_2 notation A, G and G' that correspond to the "anti", "G(+)" and "G(−)" configurations respectively. TS-C-MeS-G-DMDS is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.19(a).

From manual inspection of TS in Table 6.15(a), τ_1 deviates by 2.7°–3.8° from that in DMDS. Nevertheless, the changes in this torsional angle did not vary its torsional configurations as compared to DMDS.

From Table 6.15(a), Similarly, by considering the three possible values of χ_2 (±60° and 180°), three TS conformations were first studied from the reaction of MeS^- and the conformer DPDS7, and the results are reported in

Table 6.15(b) with the TS names, conformations, and their relative energies. The conformation of the forming MeS···Pr, described by (χ_3, χ_2), has the order of G(+)anti|G(–)anti > G(+)G(+)|G(–)G(–) > G(+)G(–)|G(–)G(+) in ΔH_{298}. In addition, the transition states constructed from the nucleophilic substitution on the Cα of DPDS1 in the -S-propyl group denoted by antiG(+)|antiG(−) have high steric hindrance near the reaction center due to the linear propyl chain, thus the linear propyl chain will always be modified after geometry optimization (results are not shown). To simplify the calculation, only one transition state conformation with the most preferred conformation of MeS···Pr denoted by G(+)anti was studied for the reaction of MeS^- and the conformer DPDS1 (Table 6.15(b)). Overall, TS-C-MeS-A-DPDS1 is the lowest energy transition state in ΔH_{298} and its optimized geometry is shown in Figure 6.19(b).

From manual inspection of TS in Table 6.15(b), τ_1 in the first transition state deviates the least (-2.5°) from that in DPDS1, followed by τ_3 (2.8°) and τ_2 (3.6°). Similarly, τ_1 in the last three transition states deviates the least (-2.7−-1.2°) from that in DPDS7, followed by τ_3 (-3.3−-0.5°) and τ_2 (-3.8−-0.1°) (τ_3 = 64.3° and τ_2 = 67.8°). However, the changes in these three torsional angles did not vary their torsional configurations as compared to DPDS1 or DPDS7.

From Figure 6.19, the geometric features of TS-C-MeS-G-DMDS and TS-C-MeS-A-DPDS1 all agree with the S_N2 mechanism. Two half bonds are present in the transition states, one corresponds to the breaking Cα-S_{lg} bond (2.328−2.390 Å) that is longer than the C-S bond in DMDS or DPDS1 (1.819−1.832 Å), and the other corresponds to the forming Cα-S_{nuc} bond

(2.401−2.455 Å) that is longer than the C-S bond in MeSMe or MeSPr1 (1.812−1.822 Å). The Cα is pentacoordinate, and almost adopts a trigonal planar geometry considering the three covalent bonds around it, thus it is approximately sp^2-hybridized instead of sp^3-hybridized in DMDS and DPDS1. The nucleophile MeS^- attacks the backside of Cα at an angle of 161.6−174.6° and causes an inversion at Cα.

By examining the torsional angles, χ_3 and κ_3 are different from τ_3 and have values close to ±85° with the opposite signs in TS-C-MS-A-DPDS1 due to the backside attack. χ_2 is similar to τ_2 and has values close to 60° in TS-C-MeS-G-DMDS or -170° in TS-C-MeS-A-DPDS1. κ_2 is close to -59° in TS-C-MeS-G-DMDS, while κ_2 is increased by 23.5° in TS-C-MeS-A-DPDS1 from the corresponding τ_2 (-62.5°) in DPDS1.

Table 6.16 Activation barriers ($\Delta H^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔH_{298}, kJ mol^{-1}) comparisons for the reaction of MeS^- and (a) diallyl/dialkyl disulfides or (b) S-allyl/alkyl-methyl disulfides.

(a)

No.	Reaction	$\Delta H^{\ddagger}_{298}$	ΔH_{298}
1	DADS + MeS^- → MeSA + ASS^- [a]	87.72	-33.65
2	DMDS + MeS^- → MeSMe + $MeSS^-$	100.44	-34.26
3	DPDS + MeS^- → MeSPr + $PrSS^-$	113.53	-27.40

(b)

No.	Reaction	$\Delta H^{\ddagger}_{298}$	ΔH_{298}
1	MeSSA + MeS^- → MeSA + $MeSS^-$ [a]	89.35	-33.48
2	MeSSPr + MeS^- → MeSPr + $MeSS^-$	112.02	-30.35

(c)

No.	Reaction	$\Delta H^{\ddagger}_{298}$	ΔH_{298}
1	MeSSBn + MeS^- → MeSBn + $MeSS^-$	89.96	-29.97

[a] The energy profiles were reported in section 6.3.1.

The energy profiles, involving the activation barriers and energies of reaction, from the reaction of MeSH and DMDS/DPDS are summarized in

Table 6.16(a). The energy profiles from the reaction of MeSH and DADS are also included in Table 6.16(a) for comparison purpose.

In general, all of the reaction steps in Table 6.16(a) are exothermic as indicated by the negative energies of reaction in ΔH_{298}. Moreover, all of these reactions have the similar ΔH_{298}.

The Cα nucleophilic substitutions on dialkyl disulfides (DMDS or DPDS) are slower than that on DADS by comparing the activation barriers in $\Delta H^{\ddagger}_{298.}$ Such results may be due to the special role of the allyl group in the S_N2 reactions. The -CH=CH_2 group in DADS can help stabilizing the S_N2 transition states by π-conjugation, but the three hydrogens in DMDS or the ethyl group in DPDS cannot not provide such stabilizations. In addition, the Cα nucleophilic substitution on DPDS is slower than that on DADS by comparing the activation barriers in $\Delta H^{\ddagger}_{298.}$ This is probably due to the nature of the Cα reaction center. In the S_N2 reactions, the primary carbon in the methyl group is more reactive than the secondary carbon in the propyl group due to reduced steric hindrance. Therefore, the activation barriers at difference Cα centers should be in the order of propyl Cα > methyl Cα > allylic Cα.

To test on the potential leaving group effects in the reaction of MeSH and DADS or DPDS, the reaction of MeSH and MeSSA or MeSSPr were calculated. The reaction of MeSH and MeSSA has been studied previously in section 6.3.1. The energy profiles from these two reactions are summarized in Table 6.16(b).

By comparing Table 6.16(a) and (b), the energy profiles from the reaction of MeSH and DADS or DPDS closely resemble that from the reaction

of MeSH and MeSSA or MeSSPr, despite the differences in the leaving groups. These results confirm that the order in their activation barriers is not due to the leaving group effects but due to the different characteristics of the reactive Cα centers.

Because a phenyl group next to Cα may help stabilizing the S_N2 transition states by π-conjugation just like the allyl group in MeSSA, it is proposed that MeSSBn (Bn stands for benzyl) can undergo the Cα nucleophilic substitution similar to MeSSA. The energy profiles of this reaction are reported in Table 6.16(c) and the results are comparable to that from the reaction of MeSH and MeSSA (Table 6.16(b)).

6.4 Conclusion

The transitions states from the reaction of MeSH and DADS/DATS were studied computationally and they were constructed based on the most stable conformation(s) of the corresponding reactants to simplify the calculation.

Following the categorization of the reaction steps in Chapter 5, it was observed that:

(1) in the Cα nucleophilic substitutions, the conformation of the forming products usually resembles the most stable form of products, suggesting that some of the resonance stabilizations in the optimal values of τ_3, τ_2 or τ_1 may be still present in the transition states. Two additional and unique interactions, the C-H(MeS^- or AS^-)···π(-$CH=CH_2$) and the LP(S)(DADS or MeSSA)···π(-$CH=CH_2$) interactions, help further stabilize the transition states.

(2) in the S/side-S/mid-S nucleophilic substitutions and the H_2S release step, the conformation of the forming products also tends to resemble the most stable form of products. In addition, the transition state conformations with the lowest ΔH_{298} have the substituent groups at the two sides pointing to the same direction. In the mid-S nucleophilic substitution, the transition state has one additional and unique C-S(DATS)$\cdots\pi$(-CH=CH_2) interaction.

From the analysis of energy profiles, the Cα nucleophilic substitutions are always slower than the S nucleophilic substitutions in reaction of MeS^- and DADS and these results are in agreements with the experimental results from Liang *et al.*[1] This difference may be explained by the higher coordination number at the Cα reaction center than the S reaction center in the transition states, and the additional requirement for re-hybridization of Cα from sp^3 to sp^2. Moreover, the mid-S nucleophilic substitution is slower than the side-S nucleophilic substitution from the reaction of MeS^- and DATS, and this is consistent with a similar reaction study by Myers *et al.*[19] This result is possibly due to the greater stability of the leaving group in the latter case. Consequently, the nucleophilic substitution of DATS by MeS^- is expected to occur mostly on the terminal S atom to form ASS^- directly in one step.

Based on the Cα nucleophilic substitution of DADS by MeS^-, the change of temperature to the reaction temperature of 37°C or the inclusion of more splitting, polarization or diffuse functions in the basis set did not significantly change the activation barrier.

Next, the transition states from the Cα nucleophilic substitution of DADS by the full reactant GSH were studied using an automated

conformational search using the Amber12:EHT force field and the LowModeMD search method. The conformations of GSH from different sources were first studied computationally and several intramolecular hydrogen bonds were identified in the low-lying conformations of GSH. From the analysis of the low-lying transition states, the conformations of GSH in these transition states resemble the low-lying conformations of GSH, and the C-H$\cdots\pi$ and the LP(S)$\cdots\pi$ interactions identified in the small transition states models using MeSH are preserved in some of these transition states. However, the conformation of DADS can deviate a lot from the lowest energy conformation of the reactant DADS used in the small transition state study. This may be caused by the additional intermolecular interactions between GSH and DADS. More importantly, the activation barrier from the full model system is comparable to that from the small model system based on $\Delta H^{\ddagger}_{298}$.

Lastly, the hypothetical Cα nucleophilic substitution of DMDS or DPDS by MeS^- was studied computationally. From the energy profile analysis, the Cα nucleophilic substitutions on DMDS or DPDS are slower than that on DADS. These results are consistent with the unpublished experimental results from Liang *et al.* This can be explained by the increased stabilization of the S_N2 transition states by π-conjugation in the -CH=CH_2 group of DADS. Additionally, the Cα nucleophilic substitution on DPDS is slower than that on DADS because the primary carbon in the methyl group is more reactive than the secondary carbon in the propyl group due to reduced steric hindrance.

6.5 References

1. Liang, D.; Wu, H.; Wong, M. W.; Huang, D., *Org. Lett.* **2015,** *17* (17), 4196.

2. (a) Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008,** *120* (1-3), 215; (b) Zhao, Y.; Truhlar, D. G., *Acc. Chem. Res.* **2008,** *41* (2), 157.

3. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B* **2009,** *113* (18), 6378.

4. Weinhold, F.; Landis., C. R., *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*. Cambridge University Press: Cambridge, UK, 2005.

5. 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. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

6. Bader, R. F. W., *Atoms in Molecules: A Quantum Theory*. Clarendon Press: Oxford, 1994.

7. Popellier, P. L. A.; Bone, R. G. A. *MORPHY98*, UMIST, Manchester, England, EU, 1998.

8. (a) Espinosa, E.; Molins, E.; Lecomte, C., *Chem. Phys. Lett.* **1998,** *285* (3–4), 170; (b) Grabowski, S. J.; Leszczynski, J., *Chem. Phys.* **2009,** *355* (2–3), 169; (c) Koch, U.; Popelier, P. L. A., *J. Phys. Chem.* **1995,** *99* (24), 9747.

9. *Molecular Operating Environment (MOE), 2014.09*, Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal QC, Canada, H3A 2R7, 2014.

10. Labute, P., *J. Chem. Inf. Model.* **2010,** *50* (5), 792.

11. Singh, R.; Whitesides, G. M., In *Sulphur-Containing Functional Groups (1993)*, John Wiley & Sons, Inc.: 2010; pp 633.

12. (a) Haynie, D. T., The First Law of Thermodynamics

Biological Thermodynamics. Cambridge University Press: 2008; (b) Markovitch, O.; Agmon, N., *J. Phys. Chem. A* **2007,** *111* (12), 2253; (c) David, v. d. S.; Paul J., v. M.; Per, L.; Nicuşor, T., *J. Phys. Chem. B* **2006,** *110* (9), 4393.

13. Novoa, J. J.; Mota, F., *Chem. Phys. Lett.* **2000,** *318* (4–5), 345.

14. Iwaoka, M.; Isozumi, N., *Molecules* **2012,** *17* (6), 7266.

15. Nziko, V. d. P. N.; Scheiner, S., *J. Phys. Chem. A* **2015,** *119* (22), 5889.

16. Stewart, R., *The Proton: Applications to Organic Chemistry*. Academic Press: New York, 1985.

17. Stockdreher, Y.; Sturm, M.; Josten, M.; Sahl, H. G.; Dobler, N.; Zigann, R.; Dahl, C., *J. Biol. Chem.* **2014,** *289* (18), 12390.

18. Kreevoy, M. M.; Harper, E. T.; Duvall, R. E.; Wilgus, H. S.; Ditsch, L. T., *J. Am. Chem. Soc.* **1960,** *82* (18), 4899.

19. Myers, A. G.; Cohen, S. B.; Kwon, B. M., *J. Am. Chem. Soc.* **1994,** *116* (4), 1255.

20. Neves, R. P. P.; Fernandes, P. A.; Varandas, A. J. C.; Ramos, M. J., *J. Chem. Theory Comput.* **2014,** *10* (11), 4842.

21. Wright, W., *Acta Crystallogr.* **1958,** *11* (9), 632.

22. Görbitz, C. H., *Acta Chem. Scand.* **1987,** *41b*, 362.

23. Laurence, P. R.; Thomson, C., *Theor. Chim. Acta* **1980,** *57* (1), 25.

24. Lampela, O.; Juffer, A. H.; Rauk, A., *J. Phys. Chem. A* **2003,** *107* (43), 9208.

25. Zhang, R.; Wu, W.; Luo, S., *J. Solution Chem.* **2011,** *40* (10), 1784.

26. Kurian, R. Computational and Experimental Investigations of Metalprotein Thiol/Disulfide Exchange Reactions. Electronic Theses and Dissertations, University of Maine, 2013.

27. (a) Halgren, T. A., *J. Comput. Chem.* **1999,** *20* (7), 720; (b) Halgren, T. A., *J. Comput. Chem.* **1996,** *17* (5-6), 490.

28. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., *J. Comput. Chem.* **2005,** *26* (16), 1752.

Chapter 7 Conclusions and Outlook

7.1 Study on the Structure of Human DNMT1 and DNA Methylation Mechanism

The first half of this thesis (Chapters 3 and 4) presents our molecular modeling study on the structure of hDNMT1 and our mechanistic study on the first two steps in the DNA methylation reaction—the Cys nucleophilic attack and the methyl transfer—based on small model systems.

In Chapter 3, a novel 3D model of the hDNMT1-DNA-SAH complex at the C-terminal domain was generated and validated via homology modeling and MD refinement based on the crystal structure of a productive mDNMT1-DNA-SAH complex (4DA4)[1]. The MD simulations showed that the important H-bonds in this homology model at the active site are similar to that observed in the mDNMT1-DNA complex (4DA4)[1] that remained stable during the MD simulations. The overall structure of this final hDNMT1-DNA-SAH complex resembles that of the mDNMT1-DNA-SAH complex with minor differences in the key active site interactions. Our final model may be used for other molecular modeling studies, such as molecular docking and in silico mutagenesis, to facilitate the study of some existing hDNMT1 inhibitors or the discovery of novel hDNMT1 inhibitors, and to model the effects of mutagenesis on the protein structures and functions. One possible improvement to our homology model is to replace SAH with SAM, so as to

model the pre-reaction hDNMT1-DNA-SAM complex that can be more beneficial to study the DNA reaction mechanism in hDNMT1.

In Chapter 4, the study on the first two steps in the DNA methylation reaction was mainly reported at the (PCM/)M06-2X/6-31G* level based on a truncated small model system. Based on our simple benchmarking study against the DF-MP2 and CCSD methods, the M06-2X functional is better than the B3LYP functional at describing the C-S bond formation and it slightly underestimates the C-S bond stability.

Our results from the reaction step 1 showed that the intermediate 1 (I1) is a stable structure with the surrounding Glu or Arg side chains. The Glu side chain stabilizes the TS1 and I1 through H-bonds with cytosine-N4-H and via direct protonation of cytosine-N3 in the gas phase and solution phase. The Arg side chain stabilizes the TS1 and I1 through the interaction with cytosine-O2 or N3 though the latter interaction is not observed in the protein active site. The roles of Glu and Arg in this reaction step are mostly consistent with the earlier studies[2]. The solvent effects are crucial for calculating the PES and the TS1 structures, but not for the I1 structures.

Our results from the reaction step 2 showed that the I1 is able to react with SAM with or without the surrounding Cys, Glu or Arg side chains. The Cys nucleophilic addition to cytosine-C6 increases the nucleophilicity of Cyt-C5 for the methyl transfer. The Glu side chain may stabilize the TS2 through the H-bond with Cyt-N3 or via direct protonation of Cyt-N3, but it is still uncertain which effect is more important to the TS2. Therefore, energy profiles from this reaction step are required to resolve the uncertainty. The Arg

side chain possibly destabilizes the TS2 through the interaction with Cyt-O2 due to the electron-withdrawing effects of the NH's via H-bonds. The solvent effects are important for calculating the TS2 structures.

The limitations of using this small model system to study the DNA methylation mechanism include the following: (1) cannot locate the preTS1 and preTS2 that resemble the protein active site; (2) fail to locate the TS for the one-step concerted reaction that combines the nucleophilic addition and the methyl transfer. Consequently, the energy profiles from these two reaction steps could not be computed and compared to confirm the roles of Cys, Glu and Arg directly. The unsuccessful attempts on the constrained optimization in our small model systems suggested that the SAM model in our system should be extended to improve such calculations with the careful choices on the restrained parameters. Once the system size is increased, computationally less expensive methods such as semi-empirical methods and force fields may be incorporated and the system can be studied by the QM/QM or QM/MM method.

7.2 Study on H_2S Releasing Reactions from the Reaction of GSH and Organosulfur Compounds

The second half of this thesis (Chapters 5 and 6) presents our computational study on the H_2S releasing reactions from the reaction of GSH and organosulfur compounds at the SMD/M06-2X/6-31+G* level.

Chapter 5 reports on the conformational analysis of the equilibrium structures in the H_2S releasing reactions. To perform our study, the

nucleophile GSH was simplified to MeSH and the S_N2 mechanism was assumed in the nucleophilic substitutions.

The three torsional angles considered in the conformation analysis are τ(C-S-S-X, X=H, C or S) (τ_1), τ(C-S-S-X, X=H, C or S) (τ_2), and τ(C-C-S-X, X=C or S) (τ_3).

In the allyl-containing organosulfur compounds, it was found that the -S-allyl groups all adopt the G(+)G(–)|G(–)G(+) conformation for the (τ_3, τ_2) combination ≈ (±120°, ±300°) combination, while the -SS-allyl groups may adopt the G(+)G(–)+|G(–)G(+)– or the G(+)G(–)–|G(–)G(+)+ conformation for the (τ_3, τ_2, τ_1) combination ≈ (±120°, ±300°, ±90°) or (±120°, ±300°, ±270°). The preferences for these conformations are due to the resonance stabilization by hyperconjugations. Moreover, the trisulfide linkages (-S-S-S-) all adopt the ++|— conformation for the (τ_1, τ_1) combination ≈ (±90°, ±90°) to avoid the steric repulsion in the +–|–+ conformation. Furthermore, to the best of our knowledge, this is probably the first comprehensive computational study on the stable conformers of the allyl-containing organosulfur compounds in an aqueous environment.

In the alkyl-containing organosulfur compounds, it was observed that -S-propyl groups may adopt the antiG(+)|antiG(–), G(+)G(+)|G(–)G(–) or antianti conformation for the (τ_3, τ_2) combination ≈ (180°, ±60°), (±60°, ±60°) or (180°, 180°), while the -SS-propyl groups may adopt antiG(+)+|antiG(–)–, G(+)G(+)–|G(–)G(–)+ or G(+)G(+)+|G(–)G(–)– conformation for the (τ_3, τ_2, τ_1) combination ≈ (180°, ±60°, ±90°), (±60°, ±60°, ±270°) or (±60°, ±60°, ±90°). The conformations of the -S-propyl groups partially agrees with results from

the conformational studies on MeSPr[3]. The central C-S-S-C group prefers to adopt the G(+)+G(+)|G(−)−G(−) conformation for the (τ_3, τ_2, τ_1, τ_2,) combination ≈ (±60°, ±90°, ±60°), which is consistent with experimental studies[4].

Chapter 6 presents a computational study on the transition states and the summary of the energy profiles of the H_2S releasing reactions. The transition states were built based on the most stable conformation(s) of the corresponding reactants to simplify the calculation.

First, the transitions states from the reaction of MeSH and DADS/DATS were studied. In the Cα nucleophilic substitutions, the conformation of the forming products usually resembles the most stable form of products. Moreover, two additional interactions—the C-H(MeS^- or AS^-)···π(-CH=CH_2) and LP(S)(DADS or MeSSA)···π(-CH=CH_2) interactions—were observed that help further stabilize the transition states. In the S/side-S/mid-S nucleophilic substitutions and the H_2S release step, the conformation of the forming products also tends to resemble the most stable form of products. In addition, the transition state conformations with the lowest ΔH_{298} have the substituent groups at the two sides pointing to the same direction. In the mid-S nucleophilic substitution, the transition state has one additional C-S(DATS)···π(-CH=CH_2) interaction.

From the analysis of energy profiles, the Cα nucleophilic substitutions are always slower than the S nucleophilic substitutions from the reaction of MeS^- and DADS due to the higher coordination number at the Cα reaction center than the S reaction center in the transition states, and the additional

requirement for re-hybridization of Cα from sp^3 to sp^2. These results are in agreements with the experimental results from Liang *et al.*[5]. Moreover, the mid-S nucleophilic substitution is slower than the side-S nucleophilic substitution from the reaction of MeS^- and DATS due to the greater stability of the leaving group in the latter reaction. This is consistent with a similar reaction study by Myers *et al.*[6] Hence, the nucleophilic substitution of DATS by MeS^- is expected to occur mostly on the terminal S atom to form ASS^- directly in one step.

Second, the transition states from the Cα nucleophilic substitution of DADS by the full reactant GSH were studied in a preliminary way. The conformations of GSH in these transition states resemble the low-lying conformations of GSH, and the C-H···π and the LP(S)···π interactions identified in the small transition states models are preserved in some of these transition states. However, the conformation of DADS in these transition states can deviate a lot from the lowest energy conformation of DADS used in the small transition state study, possibly due to the additional intermolecular interactions between GSH and DADS. More importantly, the activation barrier from the full model system is comparable to that from the small model system based on $\Delta H^{\ddagger}_{298}$.

Third, the hypothetical Cα nucleophilic substitution from MeSH and DMDS/DPDS was studied computationally. From the energy profile analysis, the Cα nucleophilic substitutions on DMDS or DPDS are slower than that on DADS possibly due to the increased stabilization of the S_N2 transition states by π-conjugation in the $-CH=CH_2$ group of DADS. These results are consistent

with the unpublished experimental results from Liang *et al.* Additionally, the Cα nucleophilic substitution on DPDS is slower than that on DADS because the primary carbon in the methyl group is more reactive than the secondary carbon in the propyl group due to reduced steric hindrance.

Overall, our study on the H_2S releasing reactions from organosulfur compounds and GSH successfully addressed some of the questions raised in the experimental studies, such as why the Cα nucleophilic substitutions are slower than the S nucleophilic substitutions in the reaction of GSH and DADS, and why the allyl groups in the organosulfur compounds resulted in H_2S release when reacted with GSH but the alkyl groups did not. However, our study still has several limitations due to the simplifications and assumptions described earlier that need future work: (1) use of more low-lying conformers of the reactants in the transition state study; (2) study of the S/side-S/mid-S nucleophilic substitutions using the full reactant GSH; (3) possible improvement on the conformational sampling of the full TS model using other search methods.

7.3 References

1. Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J., *Sci.* **2012,** *335* (6069), 709.

2. (a) Wu, J. C.; Santi, D. V., *J. Biol. Chem.* **1987,** *262* (10), 4778; (b) Chen, L.; MacMillan, A. M.; Verdine, G. L., *J. Am. Chem. Soc.* **1993,** *115* (12), 5318; (c) Gabbara, S.; Sheluho, D.; Bhagwat, A. S., *Biochemistry* **1995,** *34*

(27), 8914; (d) Peräkylä, M., *J. Am. Chem. Soc.* **1998,** *120* (49), 12895; (e) Zangi, R.; Arrieta, A.; Cossío, F. P., *J. Mol. Biol.* **2010,** *400* (3), 632.

3. Nogami, N.; Sugeta, H.; Miyazawa, T., *Chem. Lett.* **1975,** (2), 147.

4. (a) Ackermann, K. R.; Koster, J.; Schlücker, S., *Chem. Phys.* **2009,** *355* (1), 81; (b) Sugeta, H.; Go, A.; Miyazawa, T., *Bull. Chem. Soc. Jpn.* **1973,** *46* (11), 3407.

5. Liang, D.; Wu, H.; Wong, M. W.; Huang, D., *Org. Lett.* **2015,** *17* (17), 4196.

6. Myers, A. G.; Cohen, S. B.; Kwon, B. M., *J. Am. Chem. Soc.* **1994,** *116* (4), 1255.

Chapter 8 Appendix

List of Tables

Table S 8.1 Crystal structures of M.HhaI (327), mDNMT1 (1620) and hDNMT1 (1616) including the C-terminal catalytic domain[a]

Table S 8.2 Computational methods in the reported molecular modeling studies of hDNMT1

Table S 8.3 Summary of computational methods in the DNA methylation studies

Table S 8.4 Protonation states of His in the homology model of hDNMT1[a]

Table S 8.5 Calculated conformations of reactants and products (Table 5.2) and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1}) with ranking[a]

Table S 8.6 Geometric information of the lowest energy conformers for reactants and products (Table 5.2)[a]

Table S 8.7 Geometric information of the optimized conformers of MeSSA[a]

Table S 8.8 Donor-Acceptor interactions in the optimized conformers for MeSSA from the NBO analysis[a]

Table S 8.9 Orbital hybridization of LPs on S^9 and S^{10} by the NBO analysis

Table S 8.10 Calculated conformations of reactants and products (Table 5.3) and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1}) with ranking[a]

Table S 8.11 Geometric information of the lowest energy conformers for reactants and products (Table 5.3)[a]

Table S 8.12 Calculated TS conformations from the reaction of MeS^-/AS^- and various allyl-containing organosulfur compounds (except MeSSH) and their relative free energies (ΔG_{298}, kJ mol^{-1})

Table S 8.13 Interaction analysis in transition states from the Cα nucleophilic substitution by AIM and NBO[a]

Table S 8.14 Activation barriers ($\Delta G^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔG_{298}, kJ mol^{-1}) from the reaction of MeSH and (a) DADS or (b) DATS[a]

Table S 8.15 Calculated TS conformations from the reaction of GSH and DADS via Cα nucleophilic substitution and their relative free energies (ΔG_{298}, kJ mol^{-1})

Table S 8.16 The comparison between the full model TS and the small model TS in terms of relative free energies (ΔG_{298}, in kJ mol^{-1})

Table S 8.17 Interaction analysis in TS5-C-GSH-DADS by AIM and NBO[a]

Table S 8.18 Calculated TS conformations from the reaction of MeS^- and (a) DMDS or (b) DPDS via Cα nucleophilic substitution and their relative free energies (ΔG_{298}, kJ mol^{-1})

Table S 8.19 Activation barriers ($\Delta G^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔG_{298}, kJ mol^{-1}) comparisons for the reaction of MeS^- and (a) diallyl/dialkyl disulfides or (b) S-allyl/alkyl-methyl disulfides

List of Figures

Figure S 8.1 Sequence alignment of the C-terminal domain of mDNMT1 and hDNMT1 from the Clustal Omega server[27]. The differences in the residues are highlighted in red.

Figure S 8.2 Atom-atom distances from the 9 ns MD production phase. Time is in ps, and distance is in Å.

Figure S 8.3 PES scan along the C6-S bond for intermediates 1.

Table S 8.1 Crystal structures of M.HhaI (327), mDNMT1 (1620) and hDNMT1 (1616) including the C-terminal catalytic domain[a]

Protein[b]	PDB Information[c]	Ligand Description[d]	References
M.HhaI *(Haemophilus haemolyticus)*	1HMY, 1–327 (2.5 Å)	SAM	Cheng *et al.* (1993)[1]
	2HMY, 1–327 (2.61 Å)	SAM; a non-specific short DNA in solution without co-crystallization	O'Gara *et al.* (1999)[2]
	1MHT, 1–327 (2.6 Å)	SAH & 13-mer unmethylated dsDNA with both chains modified; 5-fluoro-2'-deoxycytidine was methylated and covalently bound to Cys81	Klimasauskas *et al.* (1994)[3]
	3MHT, 1–327 (2.7 Å)	SAH & 12-mer unmethylated dsDNA; 2'-deoxycytidine bound to the active site	O'Gara *et al.* (1996)[4]
	4MHT, 1–327 (2.7 Å)	SAH & 12-mer methylated dsDNA; 5-methyl-2'-deoxycytidine bound to the active site	
	5MHT, 1–327 (2.7 Å)	SAH & 12-mer hemimethylated dsDNA; 2'-deoxycytidine bound to the active site	O'Gara *et al.* (1996)[5]
	6MHT, 1–327 (2.05 Å)	SAH or SAM & 12-mer hemimethylated dsDNA with the unmethylated chain modified; 4'-thio-2'deoxycytidine was methylated and covalently bound to Cys81 in some dsDNA	Kumar *et al.* (1997)[6]
	1M0E, 1–327 (2.39 Å)	SAH & 12-mer unmethylated dsDNA with one chain modified; zebularine covalently bound to Cys81 at C6	Zhou *et al.* (2002)[7]
M.HhaI (*Haemophilus parahaemolyticus*)	2HR1, 1–327 (1.96 Å)	SAH & 12-mer unmethylated dsDNA; 2'-deoxycytidine bound to the active site	Shieh *et al.* (2006)[8]
	2Z6A, 1–327 (2.88 Å)	SAH & 13-mer unmethylated dsDNA; 2'-deoxycytidine bound to the active site	Youngblood *et al.* (2007)[9]
mDNMT1	3AV4, 357–1612 (2.75 Å)	Zn(II)	Takeshita *et al.* (2011)[10]
	3AV5, 357–1608 (3.25 Å)	SAH & Zn(II)	
	3AV6, 357–1608 (3.09 Å)	SAM & Zn(II)	
	3PT6, 651–1600 (3.00 Å, dimer)	SAH, Zn(II) & 19-mer unmethylated dsDNA; DNA bound to the N-terminal domain	Song *et al.* (2011)[11]
	3PT9, 730–1600 (2.50 Å)	SAH & Zn(II)	
	4DA4, 731–1620 (2.60 Å, dimer)	SAH, Zn(II) & 12-mer hemimethylated dsDNA with the unmethylated chain modified; 5-fluoro-2'-deoxycytidine was methylated and covalently bound to Cys1229	Song *et al.* (2012)[12]

hDNMT1	3SWR, 601–1600 (2.49 Å)	Sinefungin & Zn(II); Sinefungin bound to the SAM binding pocket	Hashimoto and Cheng (2011)[e]
	3PTA, 647–1600 (3.6 Å)	SAH, Zn(II) & 19-mer unmethylated dsDNA; DNA bound to the N-terminal domain	Song *et al.* (2011)[11]
	4WXX, 351–1605 (2.62 Å, dimer)	SAH & Zn(II)	Zhang *et al.* (2015)[13]

[a] The numbers in the brackets indicate the total number of residues in the protein.
[b] The names in the brackets indicate the bacterial species.
[c] PDB information includes the PDB ID, the range of amino acid residues in the PDB file, the resolution and the dimer formation. The resolution and the dimer formation are shown in the brackets. It should be noted that some protein structures have missing residues within the range of residues presented.
[d] Number + "-mer": number of nucleotides in one chain of the dsDNA.
[e] There is no published paper on this PDB.

Table S 8.2 Computational methods in the reported molecular modeling studies of hDNMT1

Types	Computational Methods	References
Homology modeling	Model construction using the MODELER module (INSIGHT 2000), cytosine nucleotide and SAM added using the AFFINITY module, model refinement by minimization with a CVFF force field in a water environment using the DISCOVER3 module, model validation using molecular docking of five novel ligands by DOCK5 and AUTODOCK3 & minimization by SYBYL	Siedlecki *et al.* (2003)[14]
	Model construction using the Homology module (Insight II), model validation using WHATIF-Check	Fang *et al.* (2003)[15]
	Model construction using Modeller 8, model refinement by minimization together with a TIP3P water box	Liu *et al.* (2009)[16]
	Model construction using the Prime (Schrödinger), dsDNA from M.HhaI, model validation using PROCHECK, model refinement by minimization with the MMFFs force field in a water environment using the Macromodel	Yoo *et al.* (2011)[17]
	Model construction using Modeller 8, model refinement by minimization together with a TIP3P water box	Weng *et al.* (2014)[18]
	Model construction using Modeller 9.7 (Discovery Studio 3.5), SAH from mDNMT1 was converted to SAM, model validation using Ramachandran plot and Verity3D	Joshi *et al.* (2016)[19]
Molecular dynamics	Solvated in a TIP3P water box using VMD, minimization with the CHARMM27 force field using CHARMM, triplicate 10 ns molecular dynamics, analysis using GROMACS and VMD	Same as above
	dsDNA from M.HhaI, minimization with the OPLS2005 force field using the Maestro (Schrödinger), 100 ps stochastic dynamics and minimization with in a water environment using the Macromodel	Yoo *et al.* (2013)[20]

Table S 8.3 Summary of computational methods in the DNA methylation studies

Study	Methods[a]	References
1	QM (Gaussian 94) *ab initio*: MP2/6-31+G*//HF/3-21+G* or 6-31+G*, DFT: B3LYP/6-31+G* Solvation: IPCM (ε=78.3), default isodensity value Calculation on several bimolecular models	Peräkylä (1998)[21]
2	QM (Gaussian03) DFT: B3LYP/6-31G** Solvation: PCM (ε=20.7) Constrained optimization based on the structure of M.HhaI (2HR1)	Zangi *et al.* (2010)[22]
3	QM/MM (CHARMM31b1) QM: Self-consistent-charge density functional tight-binding (SCC-DFTB)	Zhang *et al.* (2006)[23]

	Adiabatic mapping method in TS calculation Based on the structure of M.HhaI (6MHT)	
4	QM/MM&QM/MM-MD (Modified versions of Q-Chem and TINKER) QM: B3LYP/6-31G*, MM: Amber99SB QM/MM: iterative minimization procedure to map out a minimum energy path QM/MM-MD: umbrella sampling to constrain reaction coordinates, weighted histogram analysis method (WHAM) for full energy profile Based on the structure of M.HhaI (6MHT)	Yang *et al.* (2013)[24]
5	QM & ONIOM (QM/QM) (Gaussian 09); calculation of kinetic isotope effects (KIEs) (ISOEFF) QM: M06-2X/6-31+G** Solvation: PCM (ε=20.7) Constrained optimization on the simple model systems ONIOM (QM/QM): M06-2X/6-31+G**:PM6 SAM based on the lowest-energy conformer from PubChem3D, the hemimethylated dsDNA based on the structure of a 3-mer dsDNA (2JYK), TS based on the structure of mDNMT1 (4DA4) but re-numbered based on hDNMT1 Calculation of KIEs: according to Bigeleisen equations that are based on vibrational frequencies in the equilibrium and TS structures	Du *et al.* (2016)[25]
6	QM/MM-MD (fDynamo and Gaussian 09) QM: M06-2X/6-31G**//AM1 (some testing by B3LYP in single point energy calculation), MM: Amber QM/MM-MD: minimum free energy pathways (MFEP) by the on-the-fly string method, potential of mean force (PMF) for a collective path variable by umbrella sampling to constrain reaction coordinates, WHAM for full energy profile from PMFs Based on the structure of M.HhaI (2HR1) with/without Cys81 protonated and re-modeled using classical MD simulations with Amber99SB force field using NAMD in a TIP3P water box	Aranda *et al.* (2016)[26]

[a] The first line in each cell under Methods gives an overviews of the type of calculations performed (e.g. QM, MM, MD or combination of them) and the software used are included in the brackets after the calculation type.

Table S 8.4 Protonation states of His in the homology model of hDNMT1[a]

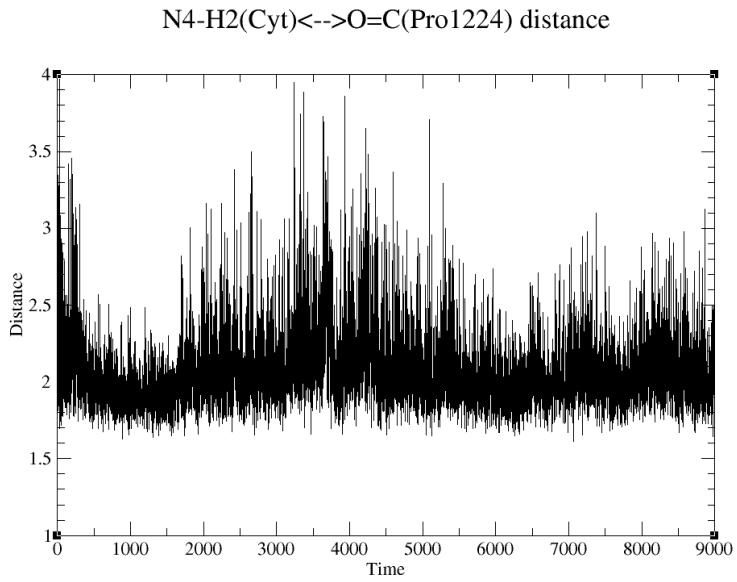
Residues	Tautomer Type
His1156	HID
His1332	HIP
His1412	HIE
His1427	HID
His1459	HIE
His1460	HID
His1507	HIE
Hid1509	HIE
His1541	HIE
His1545	HIE
His1573	HIE

[a] HID is protonated at the δ-N, HIE is protonated at the ε-N and HIP is protonated at both N's.

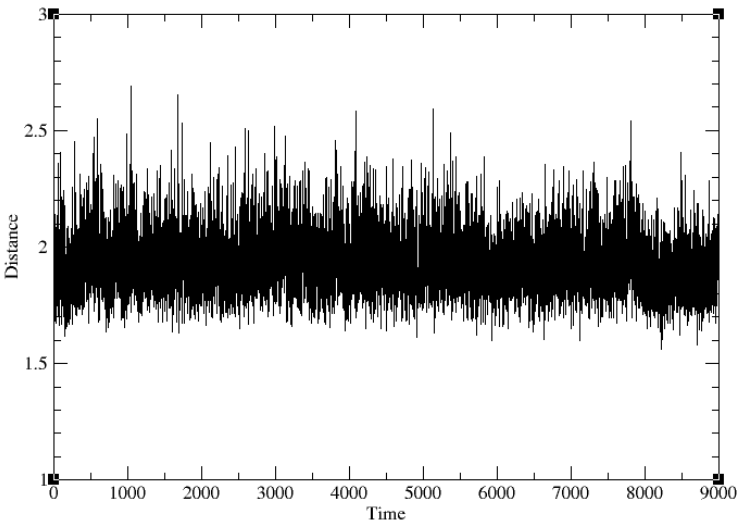
mDNMT1	1137	IKLPKLRTLDVFSGCGGLSEGFHQAGISETLWAIEMWDPAAQAFRLNNPG	1186
hDNMT1	1134	IKLPKLRTLDVFSGCGGLSEGFHQAGISDTLWAIEMWDPAAQAFRLNNPG	1183
mDNMT1	1187	TTVFTEDCNVLLKLVMAGEVTNSLGQRLPQKGDVEMLCGGPPCQGFSGMN	1236
hDNMT1	1184	STVFTEDCNI LLKLVMAGETTNSRGQRLPQKGDVEMLCGGPPCQGFSGMN	1233
mDNMT1	1237	RFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSYRRSMVLKLT	1286
hDNMT1	1234	RFNSRTYSKFKNSLVVSFLSYCDYYRPRFFLLENVRNFVSF KRSMVLKLT	1283
mDNMT1	1287	LRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHV	1336
hDNMT1	1284	LRCLVRMGYQCTFGVLQAGQYGVAQTRRRAIILAAAPGEKLPLFPEPLHV	1333
mDNMT1	1337	FAPRACQLSVVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEI QNGASN	1386
hDNMT1	1334	FAPRACQLSVVVDDKKFVSNITRLSSGPFRTITVRDTMSDLPEVRNGASA	1383
mDNMT1	1387	SEIPYNGEPLSWFQRQLRGS HYQPILRDHICKDMSP LVAARMRHIPLF PG	1436
hDNMT1	1384	LEISYNGEPQSWFQRQLRGAQYQPILRDHICKDMSALVAARMRHIPLAPG	1433
mDNMT1	1437	SDWRDLPNIQVRLGDGVI AHKLQYTF HDVKNGYSSTGALRGVCSCAE- GK	1485
hDNMT1	1434	SDWRDLPNIEVRL SDGTMARKLR YTHHDRKNGRSSSGALRGVCSCVEAGK	1483
mDNMT1	1486	ACDPE SRQFS TLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPM	1535
hDNMT1	1484	ACDPAARQFNTLIPWCLPHTGNRHNHWAGLYGRLEWDGFFSTTVTNPEPM	1533
mDNMT1	1536	GKQGRVLHPEQHRVVSVRECARSQGFPDSYRFFGNILDR HRQVGNAVPPP	1585
hDNMT1	1534	GKQGRVLHPEQHRVVSVRECARSQGFPDTYRLFGNILDKHRQVGNAVPPP	1583
mDNMT1	1586	LAKAIGLEIKLCLL S	1600
hDNMT1	1584	LAKAIGLEIKLCMLA	1598

Figure S 8.1 Sequence alignment of the C-terminal domain of mDNMT1 and hDNMT1 from the Clustal Omega server[27]. The differences in the residues are highlighted in red.

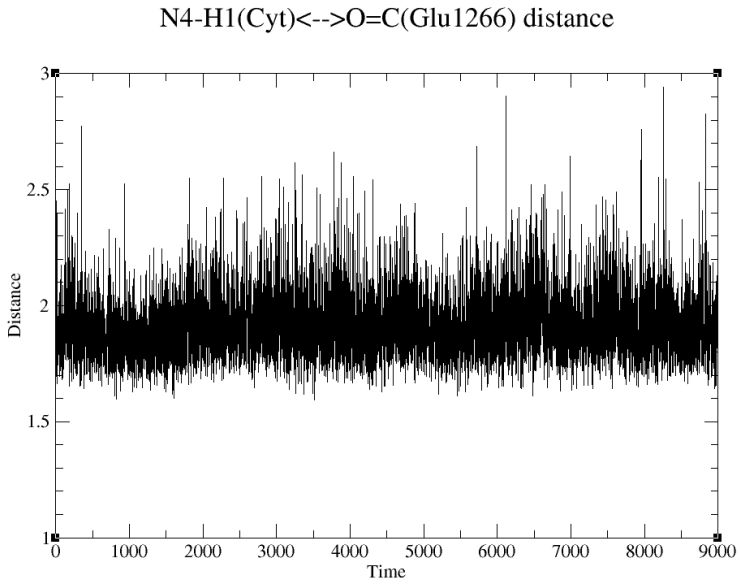
(a)


N4-H2(Cyt)<-->O=C(Pro1224) distance
Distance
Time
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

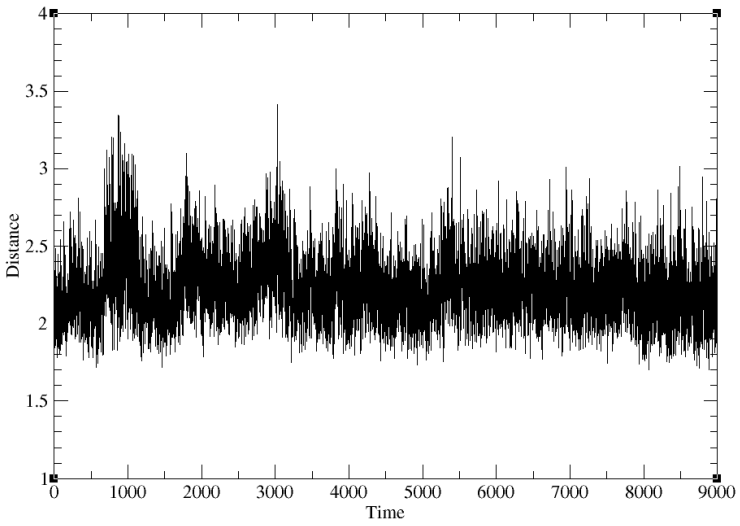
(b)


O-H(Glu1266)<-->N3(Cyt) distance
Distance
Time
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

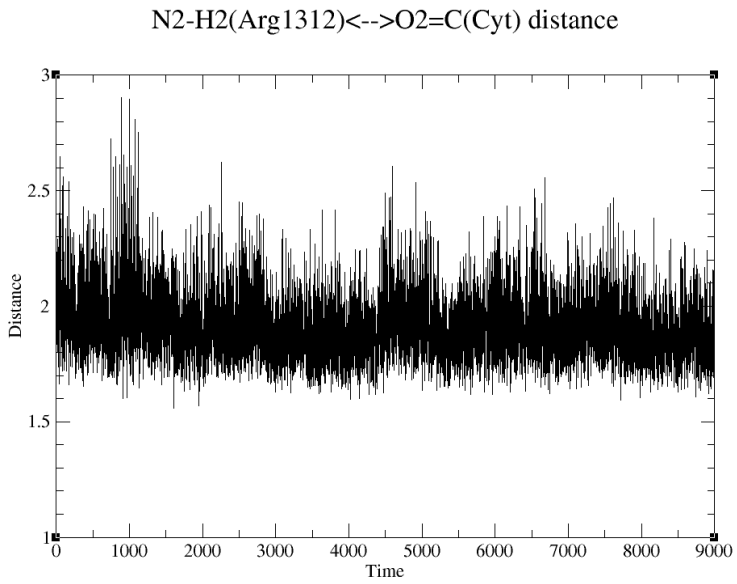
(c)


N4-H1(Cyt)<-->O=C(Glu1266) distance
Distance
Time
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

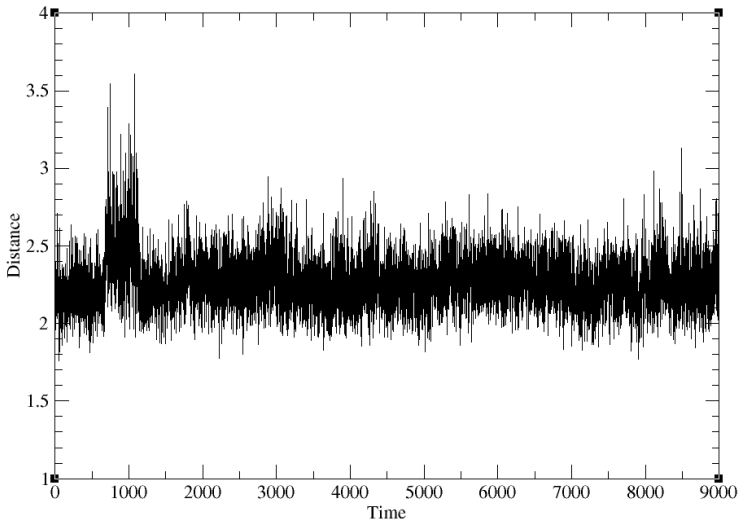
(d)


NE-H(Arg1312)<-->O2=C(Cyt) distance
Distance
Time
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

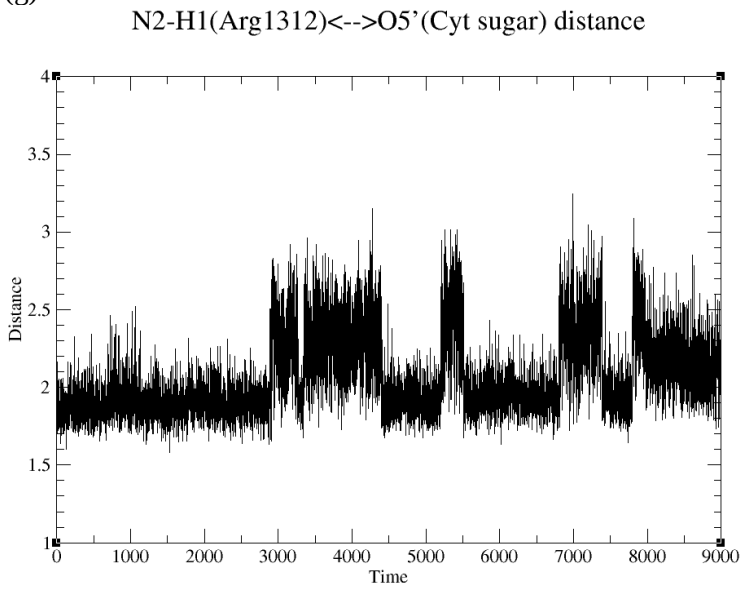
(e)


N2-H2(Arg1312)<-->O2=C(Cyt) distance
Distance
Time
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

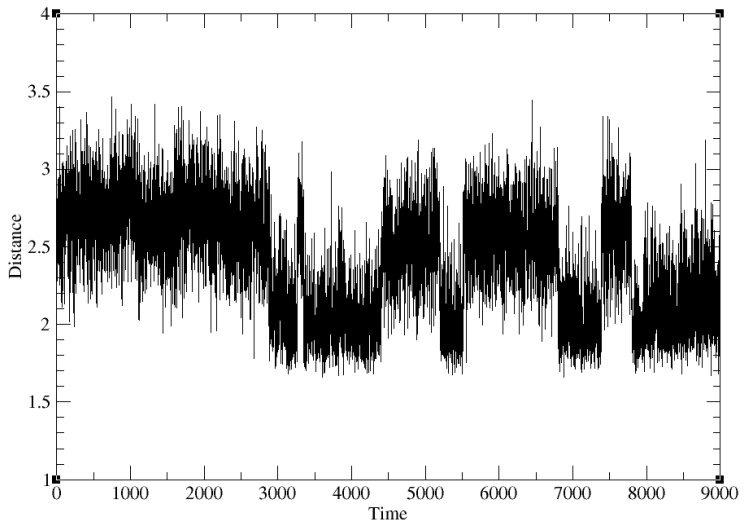
(f)


N2-H2(Arg1312)<-->O4'(Cyt sugar) distance
Distance
Time
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

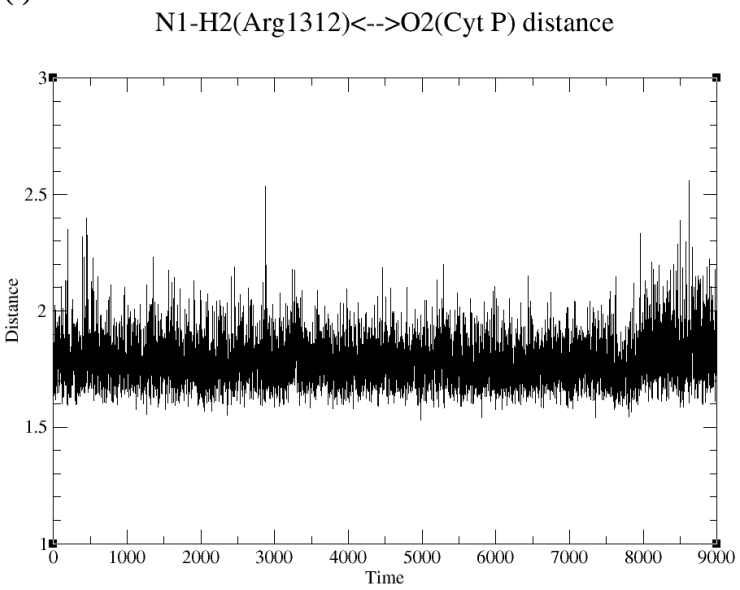
(g)


N2-H1(Arg1312)<-->O5'(Cyt sugar) distance
Distance
Time
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

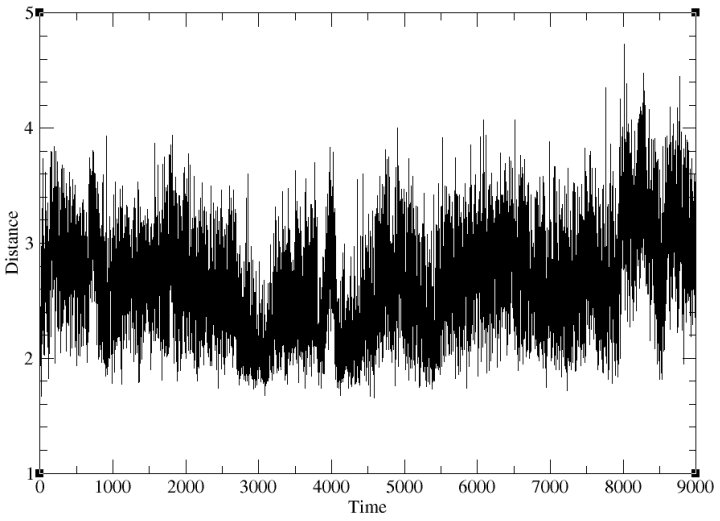
(h)


N2-H1(Arg1312)<-->O2(Cyt P) distance
Distance
Time
4
3.5
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

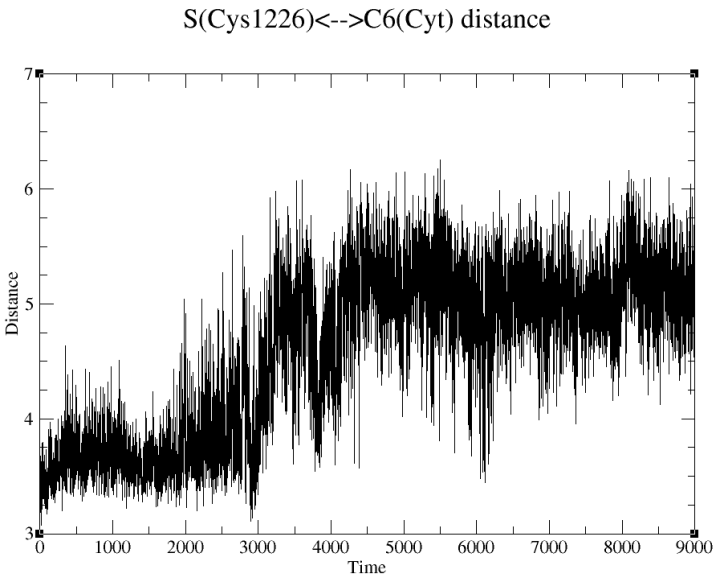
(i)


N1-H2(Arg1312)<-->O2(Cyt P) distance
Distance
Time
3
2.5
2
1.5
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(j)


N2-H2(Arg1310)<-->O2=C(Cyt) distance
Distance
Time
5
4
3
2
1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000

(k)



(l)

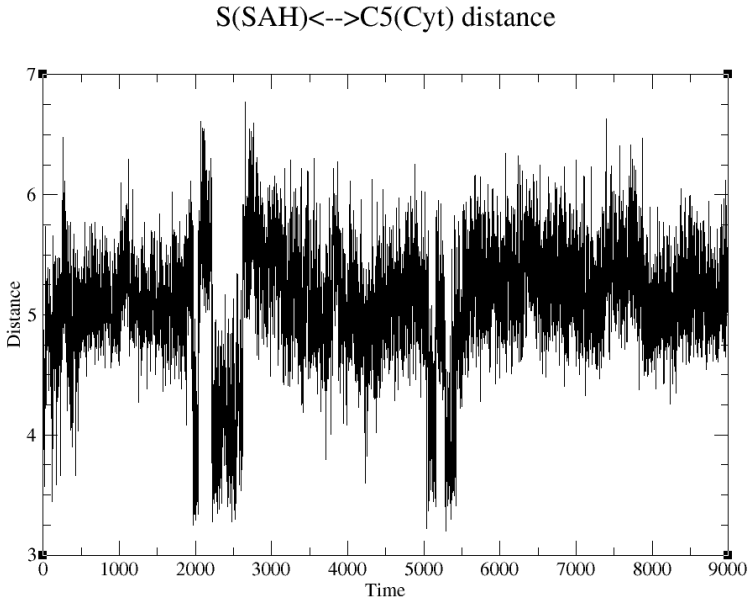
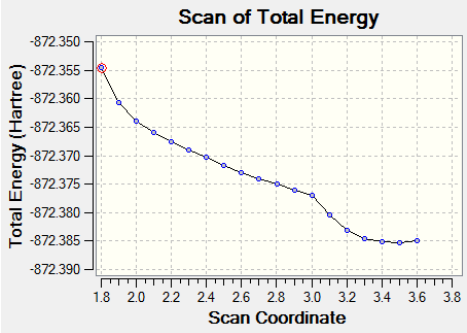
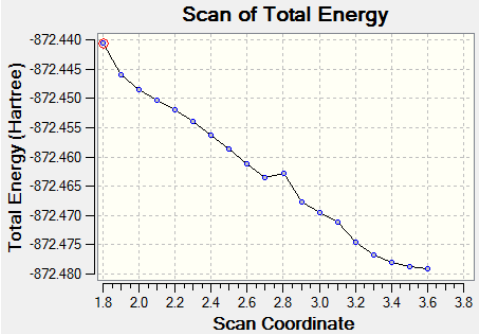


Figure S 8.2 Atom-atom distances from the 9 ns MD production phase. Time is in ps, and distance is in Å.

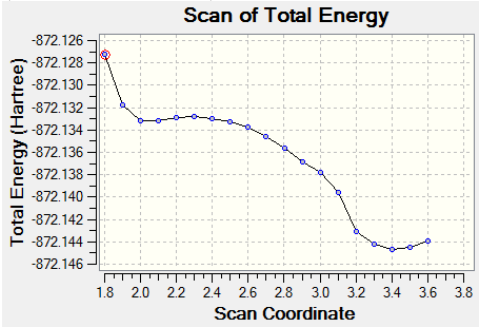
A1-1) CysCytosine
(B3LYP/6-31G*)



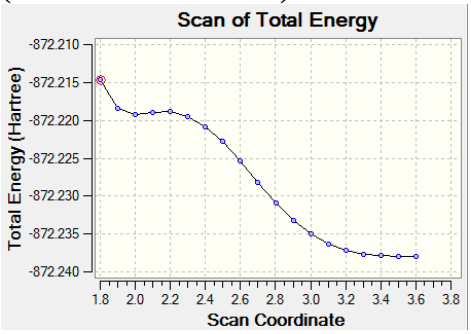
A1-2) CysCytosine
(PCM/B3LYP/6-31G*)



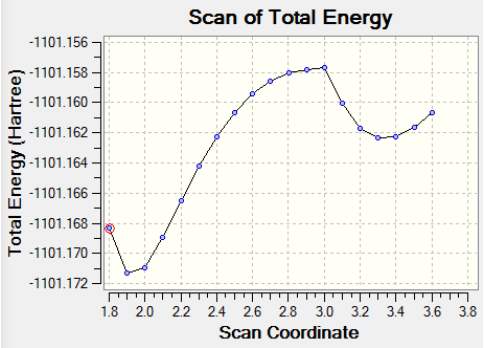
A2-1) CysCytosine
(M06-2X/6-31G*)



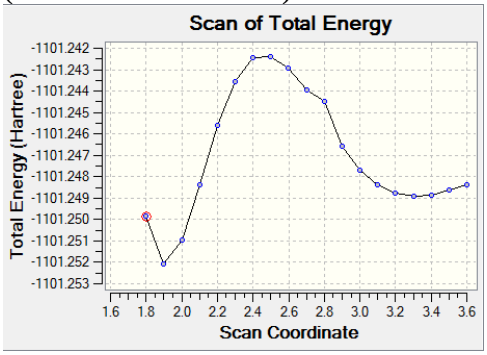
A2-2) CysCytosine
(PCM/M06-2X/6-31G*)



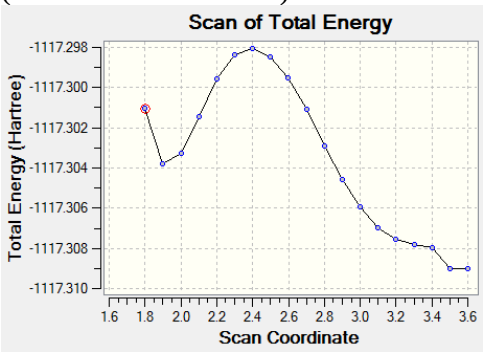
B1) Cys-CytosineH-Glu
(M06-2X/6-31G*)



B2) Cys-CytosineH-Glu
(PCM/M06-2X/6-31G*)



C) CysCytosine-Arg
(PCM/M06-2X/6-31G*)



D) CysCytosineH-Glu-Arg
(PCM/M06-2X/6-31G*)

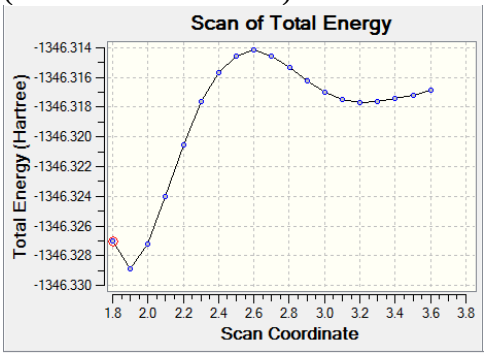


Figure S 8.3 PES scan along the C6-S bond for intermediates 1.

Table S 8.5 Calculated conformations of reactants and products (Table 5.2) and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1}) with ranking[a]

(a) DAS

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)G(–)G(+)	0.00	0.00	1	1
2	G(+)G(+)G(+)G(–)	6.27	6.91	2	2
3	G(+)G(+)G(–)G(+)	7.43	7.16	3	3
4	G(+)antiG(+)G(–)	7.73	7.37	4	4
5	G(+)antiG(–)G(+)	8.04	7.88	5	5
6	G(+)G(–)G(+)G(–)	8.30	9.50	6	6
7	G(+)antiG(+)G(+)	13.15	10.80	7	8
8	G(+)G(+)G(+)G(+)	13.28	12.54	8	10
9	G(+)antiG(–)G(–)	13.66	12.18	9	9
10	G(+)antiantiG(–)	14.16	9.83	10	7
11	G(+)antiantiG(+)	14.36	13.31	11	11
12	G(+)G(+)G(–)G(–)	14.60	14.19	12	12

(b) DADS

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)–G(+)G(–)	0.00	0.00	1	1
2	G(+)G(–)–G(–)G(+)	2.00	2.91	2	2
3	G(+)G(–)+G(–)G(+)	2.83	5.56	3	7
4	G(+)anti-G(+)G(–)	4.06	5.65	4	8
5	G(+)G(+)+G(–)G(+)	4.79	5.42	5	5
6	oppo-G(+)anti+G(–)G(+)[b]	5.67	6.84	6	10
7	G(+)G(+)+G(+)G(–)	6.32	4.86	7	3
8	oppo-G(+)anti–G(–)G(+)	7.01	5.35	8	4
9	G(+)G(–)–antiG(–)	7.09	6.80	9	9
10	G(+)G(–)–G(+)G(+)	7.15	8.77	10	13
11	G(+)G(+)–G(+)G(–)	8.75	12.10	11	20
12	G(+)G(+)+G(+)G(+)	11.06	10.70	12	18
13	oppo-G(+)anti–G(–)G(–)	11.81	11.28	13	19
14	G(+)anti+G(+)G(+)	11.82	9.87	14	15
15	G(+)anti–G(+)G(+)	11.99	8.32	15	11
16	G(+)G(+)–G(–)G(–)	12.00	10.38	16	17
17	oppo-G(+)anti+G(–)G(–)	12.17	9.94	17	16
18	G(+)anti–antiG(+)	12.27	5.42	18	6
19	oppo-G(+)anti–antiG(–)	12.51	9.15	19	14
20	oppo-G(+)anti+antiG(+)	12.66	8.59	20	12
21	G(+)G(+)–G(+)G(+)	13.25	14.52	21	21

(c) DATS

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)++G(–)G(+)	0.00	0.00	1	1
2	G(+)G(–)++G(+)G(–)	1.73	0.51	2	2
3	G(–)G(+)++G(+)G(–)	2.04	2.39	3	4
4	G(+)G(–)–+G(+)G(–)	2.12	4.00	4	6
5	G(–)G(+)–+G(+)G(–)	4.13	6.15	5	11
6	G(+)G(+)+–G(–)G(+)	4.83	8.15	6	18
7	G(+)G(+)+–G(+)G(–)	5.41	8.43	7	19
8	G(–)G(–)++G(–)G(+)	6.55	6.69	8	12
9	G(+)G(+)++G(–)G(+)	6.64	7.42	9	16
10	G(+)G(+)++G(+)G(–)	7.15	5.07	10	8
11	G(+)anti—G(–)G(+)	7.32	1.81	11	3
12	G(+)anti—G(+)G(–)	7.33	5.93	12	10
13	G(+)anti+–G(–)G(+)	7.38	8.87	13	24
14	G(+)anti++G(–)G(+)	7.67	7.23	14	14
15	G(–)G(–)++G(+)G(–)	7.70	3.56	15	5
16	G(+)anti++G(+)G(–)	7.95	4.45	16	7
17	G(+)anti+–G(+)G(–)	8.58	5.16	17	9
18	G(+)anti–+G(+)G(–)	8.92	7.30	18	15
19	G(+)anti–+G(−)G(+)	10.60	8.15	19	17
20	G(+)G(+)++G(+)G(+)	11.99	10.24	20	28
21	G(+)G(+)–+G(+)G(–)	12.13	8.81	21	22
22	G(+)anti—G(–)G(–)	12.65	8.93	22	25
23	G(+)anti++G(+)G(+)	12.78	8.47	23	20
24	G(+)G(+)++G(–)G(–)	13.10	10.65	24	29
25	G(+)anti—antiG(+)	13.44	10.15	25	27
26	G(–)G(–)++G(–)G(–)	13.57	13.70	26	33
27	G(+)anti+–antiG(–)	13.66	6.99	27	13
28	G(+)anti++antiG(+)	13.70	8.67	28	21
29	G(+)G(+)+–G(–)G(–)	13.83	13.06	29	32
30	G(+)anti—G(+)G(+)	13.88	8.87	30	23
31	G(+)anti+–G(–)G(–)	13.92	11.74	31	30
32	G(+)anti++antiG(–)	14.04	10.07	32	26
33	G(+)anti++G(–)G(–)	14.53	16.01	33	37
34	G(+)anti+–antiG(+)	15.16	12.93	34	31
35	G(+)anti–+G(+)G(+)	15.59	13.77	35	34
36	G(+)anti+–G(+)G(+)	15.68	16.51	36	39
37	G(+)G(+)+–G(+)G(+)	15.87	15.55	37	36
38	G(–)anti+–antiG(+)	17.07	14.55	38	35
39	G(+)anti–+G(–)G(–)	18.19	16.45	39	38
40	G(+)G(+)–+G(–)G(+)[c]	ND[d]	ND	ND	ND

41	G(–)G(+)–+G(–)G(+)[c]	ND	ND	ND	ND
42	G(–)G(–)+–G(+)G(+)[c]	ND	ND	ND	ND

(d) ASS^-

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)	0.00	0.00	1	1
2	G(+)G(+)	3.04	3.24	2	3
3	G(+)anti	3.91	2.46	3	2

(e) ASSH

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)+	0.00	0.00	1	1
2	G(–)G(+)+	2.66	4.08	2	2
3	G(–)G(–)+	6.03	5.01	3	4
4	G(+)G(+)+	6.93	4.44	4	3
5	G(+)anti+	7.98	7.86	5	6
6	G(–)anti+	8.18	7.47	6	5

(f) MeSA

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)	0.00	0.00	1	1
2	G(+)G(+)	5.86	4.87	2	3
3	G(+)anti	6.86	3.96	3	2

(g) MeSSA

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)+	0.00	0.52	1	2
2	G(–)G(+)+	0.67	0.00	2	1
3	G(+)G(+)+	5.54	4.54	3	5
4	G(–)anti+	6.30	3.60	4	4
5	G(–)G(–)+	6.56	7.20	5	6
6	G(+)anti+	6.63	3.58	6	3

(h) MeS_3A

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	G(+)G(–)++	0.00	0.53	1	2
2	G(–)G(+)++	0.00	0.00	2	1
3	G(–)G(+)+–	0.56	2.81	3	3
4	G(+)G(–)+–	2.17	8.91	4	11
5	G(+)G(+)++	4.80	3.23	5	4
6	G(–)anti++	5.84	4.67	6	5
7	G(+)anti++	6.35	6.33	7	7
8	G(–)G(–)++	6.43	5.38	8	6
9	G(+)anti+–	6.95	7.50	9	8
10	G(+)G(+)+–	8.13	8.72	10	9
11	G(–)anti+–	8.31	8.87	11	10

12	G(–)G(–)+–	9.04	12.00	12	12

[a] For AS^-, HS^-, MeS^-, $MeSS^-$ and DMDS, only one conformation was generated and optimized, thus they will be included in the main context.
[b] oppo-, the enantiomer of.
[c] Conformations are sterically unfavorable and were not considered in geometry optimization.
[d] ND, not determined.

Table S 8.6 Geometric information of the lowest energy conformers for reactants and products (Table 5.2)[a]

(a)

No.	A1)	A2)	A3)
Name	DAS1	DADS1	DATS1
Conformation	G(+)G(–) G(–)G(+)	G(+)G(–)– G(+)G(–)	G(+)G(–)+ +G(–)G(+)
Bond Lengths			
C=C	1.334	1.334	1.334
C-C	1.496	1.492	1.493
C-S	1.829	1.843	1.841
S-S	NA[b]	2.059	2.067
S-S	NA	NA	2.067
S-C	1.829	1.842	1.841
C-C	1.496	1.492	1.493
C=C	1.334	1.334	1.334
Bond Angles			
C=C-C	123.7	123.2	123.1
C-C-S	112.3	112.6	112.4
C-S-C/S	100.1	103.4	101.7
S-S-S	NA	NA	107.6
S-S-C	NA	102.9	101.7
S-C-C	112.3	112.9	112.4
C-C=C	123.7	123.1	123.1

(b)

No.	B1)	B2)	B3)
Name	AS^-	ASS^-1	ASSH1
Conformation	G(+)	G(+)G(–)	G(+)G(–)+
Bond Lengths			
C=C	1.337	1.335	1.334
C-C	1.496	1.493	1.492
C-S	1.847	1.838	1.840
S-S	NA	2.095	2.069
S-H	NA	NA	1.349

Bond Angles			
C=C-C	125.0	124.2	123.2
C-C-S	112.0	113.2	111.6
C-S-S	NA	104.6	102.0
S-S-H	NA	NA	1.35

(c)

No.	C)	C2-1)	C2-2)	C3-1)	C3-2)
Name	MeSA1	MeSSA1(H)	MeSSA2(G)[c]	$MeS_3A1(H)$	$MeS_3A2(G)$[c]
Conformation	G(+)G(–)	G(+)G(–)+	G(+)G(–)–	G(+)G(–)++	G(+)G(–)—
Bond Lengths					
C=C	1.334	1.334	1.334	1.334	1.334
C-C	1.496	1.492	1.491	1.492	1.492
C-S	1.831	1.841	1.842	1.842	1.840
S-S	NA	2.063	2.062	2.066	2.067
S-S	NA	NA	NA	2.071	2.069
S-C	1.812	1.820	1.818	1.817	1.816
Bond Angles					
C=C-C	123.5	123.1	123.2	122.9	123.0
C-C-S	112.0	112.9	112.7	112.5	112.8
C-S-C/S	99.1	102.9	102.9	101.9	102.4
S-S-S	NA	NA	NA	107.4	106.6
S-S-C	NA	102.6	101.9	101.4	102.0

(d)

No.	D)	E1)	E2)	E3)	F)
Name	HS^-	MeS^-	$MeSS^-$	MeSSH	DMDS
Conformation	NA	NA	NA	G(+)	G(+)
Bond Lengths					
C-S	NA	1.840	1.817	1.817	1.819
S-S	NA	NA	2.100	2.073	2.065
S-H	1.350	NA	NA	1.348	NA
S-C	NA	NA	NA	NA	1.819
Bond Angles					
C-S-S	NA	NA	102.6	102.1	102.1
S-S-C	NA	NA	NA	NA	102.1

[a] Bond lengths are in Å, and bond angles are in °.
[b] NA, not applicable.
[c] The conformation shown corresponds to the enantiomer of the optimized conformer for comparison purpose.

Table S 8.7 Geometric information of the optimized conformers of MeSSA[a]

No.	(a)	(b)	(c)	(d)	(e)	(f)
Name	MeSSA1	MeSSA2	MeSSA3	MeSSA5	MeSSA4	MeSSA6
Conformation	G(+)G(–)+	G(–)G(+)+	G(+)G(+)+	G(–)G(–)+	G(–)anti+	G(+)anti+
Bond Lengths						
C^2=C^3	1.334	1.334	1.334	1.334	1.334	1.334
C^2-H^6	1.090	1.090	1.090	1.090	1.090	1.089
C^1-C^2	1.492	1.491	1.495	1.495	1.497	1.497
C^1-H^n[b]	(n=4) 1.096	(n=5) 1.096	(n=4) 1.095	(n=5) 1.094	(n=5) 1.094	(n=4) 1.093
C^1-H^n	(n=5) 1.093	(n=4) 1.094	(n=5) 1.095	(n=4) 1.096	(n=4) 1.093	(n=5) 1.093
S^{10}-C^1	1.841	1.842	1.839	1.839	1.842	1.842
S^9-S^{10}	2.063	2.062	2.062	2.065	2.065	2.063
S^9-C^{11}	1.820	1.818	1.818	1.820	1.819	1.819
Bond Angles						
C^3=C^2-H^6	120.4	120.4	120.1	120.0	120.2	120.2
C^3=C^2-C^1	123.1	123.2	123.6	123.6	123.4	123.4
H^6-C^2-C^1	116.5	116.4	116.3	116.4	116.5	116.4
C^2-C^1-S^{10}	112.9	112.7	112.9	113.9	108.0	107.7
C^2-C^1-H^n	(n=4) 111.6	(n=5) 111.6	(n=4) 112.0	(n=5) 111.6	(n=5) 111.0	(n=4) 111.3
C^2-C^1-H^n	(n=5) 111.6	(n=4) 111.8	(n=5) 111.1	(n=4) 111.0	(n=4) 111.5	(n=5) 111.3
H^4-C^1-H^5	108.6	108.2	108.2	108.4	109.7	109.8
H^n-C^1-S^{10}	(n=4) 103.2	(n=5) 103.3	(n=5) 104.8	(n=4) 104.1	(n=5) 107.7	(n=4) 107.5
H^n-C^1-S^{10}	(n=5) 108.6	(n=4) 108.77	(n=4) 107.5	(n=5) 107.4	(n=4) 108.97	(n=5) 109.2
C^1-S^{10}-S^9	102.9	102.9	102.6	103.4	102.4	102.7
S^{10}-S^9-C^{11}	102.6	101.9	101.9	103.1	102.0	102.2

[a] Bond lengths are in Å, bond angles and torsional angles are in °.
[b] n=4 or 5, and are in the brackets.

Table S 8.8 Donor-Acceptor interactions in the optimized conformers for MeSSA from the NBO analysis[a]

(a)

Name				MeSSA1 G(+)G(–)+				MeSSA2 G(–)G(+)+			
τ_3				~120°				~-120°			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
6	$\pi(C^2=C^3)$	128	$\sigma^*(C^1-S^{10})$	NA[b]	27.82	0.52	0.053	NA	27.78	0.52	0.053
4	$\sigma(C^1-S^{10})$	130	$\pi^*(C^2=C^3)$	NA	19.46	0.75	0.053	NA	19.25	0.74	0.052
3/2	$\sigma(C^1-H^n)$	131	$\sigma^*(C^2-H^6)$	5	16.15	1.14	0.059	4	15.82	1.14	0.059
2/3	$\sigma(C^1-H^n)$	129	$\sigma^*(C^2=C^3)$	4	11.92	1.29	0.054	5	12.13	1.30	0.055
7	$\sigma(C^2-H^6)$	127/126	$\sigma^*(C^1-H^n)$	5	10.59	1.08	0.047	4	10.63	1.09	0.047
2/3	$\sigma(C^1-H^n)$	130	$\pi^*(C^2=C^3)$	4	9.67	0.68	0.035	5	9.46	0.68	0.035

6	$\pi(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	8.24	0.82	0.036	5	7.95	0.82	0.035
5	$\sigma(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	3.81	1.31	0.031	5	3.72	1.31	0.030
4	$\sigma(C^1-S^{10})$	129	$\sigma^*(C^2=C^3)$	NA	3.51	1.36	0.030	NA	3.39	1.36	0.030
τ_2				~-60º				~60º			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
32	$LP(2)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	17.61	0.80	0.052	NA	16.07	0.80	0.050
32	$LP(2)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	5	10.33	0.81	0.040	4	11.67	0.81	0.043
2/3	$\sigma(C^1-H^n)$	134	$\sigma^*(S^9-S^{10})$	4	10.67	0.69	0.037	5	10.54	0.68	0.037
10	$\sigma(S^9-S^{10})$	126/127	$\sigma^*(C^1-H^n)$	4	4.73	1.13	0.032	4	4.85	1.13	0.032
31	$LP(1)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	2.76	1.20	0.025	NA	3.77	1.21	0.030
31	$LP(1)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	5	4.02	1.20	0.025	4	3.10	1.22	0.027
τ_1				~90º				~90º			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
30	$LP(2)(S^9)$	128	$\sigma^*(S^{10}-C^1)$	NA	28.79	0.49	0.052	NA	29.29	0.49	0.053
32	$LP(2)(S^{10})$	135	$\sigma^*(S^9-C^{11})$	NA	25.02	0.52	0.050	NA	23.85	0.51	0.049

(b)

Name				MeSSA3 G(+)G(+)+				MeSSA5 G(–)G(–)+			
τ_3				~120º				~-120º			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
6	$\pi(C^2=C^3)$	128	$\sigma^*(C^1-S^{10})$	NA	23.30	0.53	0.050	NA	25.82	0.53	0.051
4	$\sigma(C^1-S^{10})$	130	$\pi^*(C^2=C^3)$	NA	17.57	0.75	0.050	NA	17.66	0.75	0.050
3/2	$\sigma(C^1-H^n)$	131	$\sigma^*(C^2-H^6)$	5	16.61	1.14	0.060	4	16.48	1.14	0.060
2/3	$\sigma(C^1-H^n)$	129	$\sigma^*(C^2=C^3)$	4	10.88	1.29	0.052	5	11.84	1.08	0.047
7	$\sigma(C^2-H^6)$	127/126	$\sigma^*(C^1-H^n)$	5	10.54	1.08	0.047	4	10.63	1.29	0.054
2/3	$\sigma(C^1-H^n)$	130	$\pi^*(C^2=C^3)$	4	10.38	0.69	0.037	5	9.92	0.69	0.036
6	$\pi(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	9.33	0.84	0.039	5	8.49	0.83	0.037
5	$\sigma(C^2=C^3)$	126/127	$\sigma^*(C^1-H^n)$	4	3.64	1.31	0.030	5	3.64	1.31	0.030
4	$\sigma(C^1-S^{10})$	129	$\sigma^*(C^2=C^3)$	NA	4.31	1.36	0.034	NA	3.77	1.36	0.031
τ_2				~60º				~-60º			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
32	$LP(2)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	17.49	0.79	0.052	NA	19.20	0.80	0.054
32	$LP(2)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	4	9.62	0.81	0.039	5	8.87	0.81	0.037
2/3	$\sigma(C^1-H^n)$	134	$\sigma^*(S^9-S^{10})$	5	9.54	0.68	0.035	4	10.33	0.68	0.037
10	$\sigma(S^9-S^{10})$	126/127	$\sigma^*(C^1-H^n)$	5	5.23	1.13	0.034	4	4.85	1.13	0.032
31	$LP(1)(S^{10})$	125	$\sigma^*(C^1-C^2)$	NA	3.26	1.20	0.027	NA	2.64	1.20	0.025
31	$LP(1)(S^{10})$	127/126	$\sigma^*(C^1-H^n)$	4	3.26	1.22	0.028	5	4.10	1.21	0.031

τ_1				~90°				~90°			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
30	LP(2)(S^9)	128	σ*(S^{10}-C^1)	NA	27.78	0.50	0.051	NA	28.37	0.50	0.052
32	LP(2)(S^{10})	135	σ*(S^9-C^{11})	NA	23.97	0.51	0.049	NA	25.61	0.52	0.051

(c)

Name				MeSSA4 G(–)anti+				MeSSA6 G(+)anti+			
τ_3				~-120°				~120°			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
6	π(C^2=C^3)	128	σ*(C^1-S^{10})	NA	28.41	0.48	0.051	NA	28.37	0.48	0.051
4	σ(C^1-S^{10})	130	π*(C^2=C^3)	NA	23.64	0.70	0.056	NA	23.60	0.70	0.056
2/3	σ(C^1-H^n)	131	σ*(C^2-H^6)	4	14.39	1.18	0.057	5	14.31	1.18	0.057
3/2	σ(C^1-H^n)	129	σ*(C^2=C^3)	5	12.64	1.12	0.046	4	12.64	1.13	0.046
7	σ(C^2-H^6)	126/127	σ*(C^1-H^n)	4	9.96	1.33	0.057	5	9.92	1.33	0.057
3/2	σ(C1-H^n)	130	π*(C^2=C^3)	5	6.90	0.70	0.300	4	6.95	0.70	0.030
6	π(C^2=C^3)	127/126	σ*(C^1-H^n)	5	5.94	0.88	0.032	4	5.94	0.87	0.032
5	σ(C^2=C^3)	127/126	σ*(C^1-H^n)	5	3.56	1.36	0.031	4	3.51	1.36	0.030
4	σ(C^1-S^{10})	129	σ*(C^2=C^3)	NA	2.80	1.33	0.027	NA	2.80	1.33	0.027
τ_2				~180°				~180°			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
32	LP(2)(S^{10})	126/127	σ*(C^1-H^n)	4	9.16	0.84	0.038	5	10.54	0.85	0.041
32	LP(2)(S^{10})	127/126	σ*(C^1-H^n)	5	7.91	0.84	0.036	4	6.44	0.84	0.032
10	σ(S^9-S^{10})	125	σ*(C^1-C^2)	NA	5.15	1.02	0.032	NA	5.06	1.02	0.031
1	σ(C^1-C^2)	134	σ*(S^9-S^{10})	NA	4.18	0.74	0.024	NA	4.18	0.74	0.024
τ_1				~90°				~90°			
NBO No.	Donor (i)	NBO No.	Acceptor (j)	n	E(2)	E(j)-E(i)	F(i,j)	n	E(2)	E(j)-E(i)	F(i,j)
30	LP(2)(S^9)	128	σ*(S^{10}-C^1)	NA	15.40	0.45	0.036	NA	15.36	0.45	0.036
32	LP(2)(S^{10})	135	σ*(S^9-C^{11})	NA	13.72	0.47	0.035	NA	13.77	0.47	0.035

[a] E(2) is the two-electron stabilizing energy (in kJ mol^{-1}). E(j)-E(i) is the energy difference between the donor and the acceptor (in Hartree), and F(i,j) is the orbital overlap (in a.u.).
[b] NA, not applicable.

Table S 8.9 Orbital hybridization of LPs on S^9 and S^{10} by the NBO analysis

(a)

Name		MeSSA1			MeSSA2		
No.	NBO	Orbital Hybridization			Orbital Hybridization		
		3s	3p	3d	3s	3p	3d
29	LP(1)(S^9)	72.81%	28.17%	0.02%	71.61%	28.37%	0.02%
30	LP(2)(S^9)	0.12%	99.82%	0.06%	0.40%	99.54%	0.06%
31	LP(1)(S^{10})	72.22%	27.77%	0.02%	73.00%	26.99%	0.02%

32	LP(2)(S^{10})	0.72%	99.22%	0.06%	0.00%	99.94%	0.06%

(b)

Name		MeSSA3			MeSSA4		
No.	NBO	Orbital Hybridization			Orbital Hybridization		
		3s	3p	3d	3s	3p	3d
29	LP(1)(S^{9})	71.96%	28.02%	0.02%	71.91%	28.07%	0.02%
30	LP(2)(S^{9})	0.23%	99.71%	0.06%	0.10%	99.84%	0.06%
31	LP(1)(S^{10})	73.05%	26.93%	0.02%	72.02%	27.97%	0.02%
32	LP(2)(S^{10})	0.00%	99.94%	0.06%	0.92%	99.02%	0.06%

(c)

Name		MeSSA5			MeSSA6		
No.	NBO	Orbital Hybridization			Orbital Hybridization		
		3s	3p	3d	3s	3p	3d
29	LP(1)(S^{9})	77.79%	22.20%	0.01%	77.85%	22.14%	0.01%
30	LP(2)(S^{9})	0.30%	99.67%	0.03%	0.25%	99.72%	0.03%
31	LP(1)(S^{10})	78.92%	21.07%	0.01%	78.68%	21.31%	0.01%
32	LP(2)(S^{10})	0.15%	99.83%	0.03%	0.36%	99.61%	0.03%

Table S 8.10 Calculated conformations of reactants and products (Table 5.3) and their relative energies (ΔH_{298} or ΔG_{298}, kJ mol^{-1}) with ranking[a]

(a) MeSPr

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	antiG(+)	0.00	4.34	1	3
2	G(+)G(+)	0.63	4.89	2	4
3	antianti	1.48	0.00	3	1
4	G(+)anti	2.93	3.54	4	2
5	G(+)G(–)	7.88	12.88	5	5

(b) DPDS

No.	Conformation	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	antiG(+)+G(–)G(–)	0.00	0.29	1	2
2	G(+)G(+)+G(–)G(–)	1.14	3.14	2	11
3	antiG(+)–G(+)anti	2.55	0.58	3	3
4	antiG(+)+G(+)anti	2.99	0.64	4	4
5	antiG(+)+G(+)G(+)	3.12	1.37	5	5
6	antiG(+)–G(–)G(–)	3.41	2.27	6	7
7	G(+)G(+)+G(+)G(+)	3.91	0.00	7	1
8	G(+)G(+)–G(–)G(+)	4.07	5.39	8	23
9	G(+)G(+)–G(+)G(+)	4.24	7.02	9	31
10	antianti+G(–)G(–)	4.38	4.87	10	20
11	antiG(+)–G(+)G(+)	5.05	3.79	11	15
12	antianti+G(–)anti	5.20	1.38	12	6
13	antiG(+)–G(–)G(+)	5.30	5.55	13	24
14	G(+)anti+G(–)G(–)	5.32	2.33	14	8
15	antianti+G(+)anti	5.75	2.48	15	9
16	antiG(+)+G(–)anti	6.09	4.73	16	19
17	antianti+G(+)G(+)	6.19	3.71	17	13
18	antiG(+)–antiG(–)	6.48	4.92	18	21
19	G(+)anti–G(+)G(+)	6.51	6.67	19	28

20	antiG(+)+G(+)G(–)	6.55	6.24	20	26
21	G(+)G(+)+G(+)G(–)	6.97	2.69	21	10
22	antiG(+)+antiG(–)	6.98	3.89	22	16
23	antiG(+)–antiG(+)	7.07	3.37	23	12
24	G(+)anti–G(–)G(–)	7.19	4.06	24	17
25	antiG(+)+antiG(+)	7.23	4.96	25	22
26	G(+)anti+G(+)G(+)	7.66	3.76	26	14
27	antianti+antianti	8.64	4.41	27	18
28	antianti+G(+)G(–)	9.23	6.48	28	27
29	antianti+antiG(–)	9.95	8.46	29	34
30	antianti+antiG(+)	10.12	6.76	30	29
31	G(+)G(–)–G(–)G(+)	10.48	9.60	31	36
32	G(+)anti–G(–)G(+)	10.68	9.25	32	35
33	G(+)anti+G(+)G(–)	10.83	8.38	33	33
34	G(+)anti–antiG(+)	11.04	6.81	34	30
35	G(+)anti+antiG(+)	11.40	6.22	35	25
36	G(+)anti+antiG(–)	11.44	8.02	36	32
37	antiG(+)+G(–)G(+)	12.50	12.52	37	39
38	antianti+G(–)G(+)	14.24	10.90	38	37
39	G(+)G(+)+G(–)G(+)	14.32	14.77	39	41
40	G(+)G(-)+G(+)G(–)	14.70	14.46	40	40
41	G(+)anti+G(–)G(+)	14.83	11.84	41	38
42	G(+)anti–G(+)G(–)	15.99	15.14	42	42
43	antiG(+)–G(+)G(–)	17.50	17.20	43	43
44	G(+)G(+)–G(+)G(–)	19.77	21.27	44	44
45	G(+)G(–)+G(–)G(+)	30.36	32.57	45	45

(c) $PrSS^-$

No.	Conformation Name	ΔH_{298}	ΔG_{298}	Rank(ΔH_{298})	Rank(ΔG_{298})
1	antiG(+)	0.00	0.00	1	1
2	G(+)G(+)	0.84	0.19	2	2
3	antianti	1.86	0.20	3	3
4	G(+)anti	3.62	1.64	4	4
5	G(+)G(–)	6.85	7.10	5	5

[a] For MeSMe, there is only one conformation generated and optimized, thus it will be included in the main context.

Table S 8.11 Geometric information of the lowest energy conformers for reactants and products (Table 5.3)[a]

(a)

No.	A1)	A2-1)	A2-2)
Name	MeSMe	MeSPr1(H)	MeSPr3(G)
Conformation	NA[b]	antiG(+)	antianti
Bond Lengths			
C-C	NA	1.528	1.527
C-C	NA	1.524	1.524
C-S	1.812	1.822	1.823
S-C	1.812	1.815	1.814
Bond Angles			
C-C-C	NA	111.1	111.1
C-C-S	NA	114.8	111.0

C-S-C	98.5	99.8	98.6

(b)

No.	A3-1)	A3-2)	B)
Name	DPDS1(H)	DPDS7(G)	PrSS$^-$1
Conformation	antiG(+)+G(−)G(−)	G(+)G(+)+G(+)G(+)	antiG(+)
Bond Lengths			
C-C	1.527	1.525	1.528
C-C	1.521	1.523	1.523
C-S	1.832	1.833	1.825
S-S	2.066	2.067	2.102
S-C	1.832	1.833	NA
C-C	1.524	1.524	NA
C-C	1.525	1.525	NA
Bond Angles			
C-C-C	110.7	113.4	111.7
C-C-S	114.8	114.7	114.6
C-S-S	104.8	102.4	103.7
S-S-C	103.6	102.5	NA
S-C-C	115.4	115.0	NA
C-C-C	114.1	113.7	NA

[a] Bond lengths are in Å, and bond angles are in °.
[b] NA, not applicable.

Table S 8.12 Calculated TS conformations from the reaction of MeS$^-$/AS$^-$ and various allyl-containing organosulfur compounds (except MeSSH) and their relative free energies (ΔG_{298}, kJ mol^{-1})

(a)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔG_{298}
1	TS-C-MeS-A-DADS1	G(+)anti; G(−)G(+)	4.43
2	TS-C-MeS-G-DADS1	G(+)G(+); G(−)G(+)	2.98
3	TS-C-MeS-G'-DADS1[a]	G(+)G(−); G(−)G(+)	6.79
4	TS-C-MeS-A-DADS1o	G(−)anti; G(+)G(−)	2.83
5	TS-C-MeS-G-DADS1o	G(−)G(+); G(+)G(−)	0.00
6	TS-C-MeS-G'-DADS1o	G(−)G(−); G(+)G(−)	8.77

(b)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔG_{298}
1	TS-C-MeS-A-MeSSA1	G(−)anti; G(+)G(−)	6.21
2	TS-C-MeS-G-MeSSA1	G(−)G(+); G(+)G(−)	0.00
3	TS-C-MeS-G'-MeSSA1	G(−)G(−); G(+)G(−)	9.78
4	TS-C-MeS-G'-MeSSA2	G(+)G(−); G(−)G(+)	3.52

(c)

No.	Name	Conformation κ_3, χ_2, χ_1; κ_2, κ_1	ΔG_{298}
1	TS-S-MeS-G-DADS1	G(−)G(−)+; G(+)−	1.15
2	TS-S-MeS-G'-DADS1	G(−)G(−)−; G(+)−	0.00
3	TS-S-MeS-G-DADS1o	G(+)G(+)+; G(−)−	1.86
4	TS-S-MeS-G'-DADS1o	G(+)G(+)−;G(−)−	0.20

(d)

No.	Name	Conformation χ_1; κ_1	ΔG_{298}
1	TS-S-MeS-G-MeSSA1	+; +	4.54
2	TS-S-MeS-G'-MeSSA1	−; +	2.17
3	TS-S-MeS-G-MeSSA2	+; +	0.00
4	TS-S-MeS-G'-MeSSA2	−; +	3.76

(e)

No.	Name	Conformation κ_3, χ_2, χ_1; κ_2, κ_1	ΔG_{298}
1	TS-SS-MeS-G-DATS1	G(+)G(+)+; G(−)+	1.59
2	TS-SS-MeS-G'-DATS1	G(+)G(+)−; G(−)+	0.00

(f)

No.	Name	Conformation χ_1; κ_1	ΔG_{298}
1	TS-SS-MeS-G-MeS$_3$A1	+; +	1.22
2	TS-SS-MeS-G'-MeS$_3$A1	−; +	3.95
3	TS-SS-MeS-G-MeS$_3$A2	+; +	3.99
4	TS-SS-MeS-G'-MeS$_3$A2	−; +	0.00

(g)

No.	Name	Conformation κ_3, κ_2, χ_1', χ_1; κ_1', κ_1	ΔG_{298}
1	TS-MS-MeS-G-DATS1	G(+)G(−)−+; ++	6.86
2	TS-MS-MeS-G'-DATS1	G(+)G(−)−−; ++	0.00

(h)

No.	Name	Conformation κ_3, χ_2, χ_1; κ_2, κ_1	ΔG_{298}
1	TS-S-MeS-G-ASSH1	G(+)G(+)+; G(−)+	0.00
2	TS-S-MeS-G'-ASSH1	G(+)G(+)−; G(−)+	1.76

(i)

No.	Name	Conformation χ_1; κ_1	ΔG_{298}
1	TS-S-MeS-G-MeSSH	+; +	0.94
2	TS-S-MeS-G'-MeSSH	−; +	0.00

(j)

No.	Name	Conformation χ_3, χ_2, χ_2', χ_3'; κ_3, κ_2	ΔG_{298}
1	TS-C-AS-GG'G'-DADS1[b]	G(+)G(−)G(−)G(+); G(−)G(+)	8.46
2	TS-C-AS-G'G'G'-DADS1	G(+)G(−)G(−)G(−);G(−)G(+)	3.91

3	TS-C-AS-G'GG-DADS1	G(+)G(+)G(+)G(−);G(−)G(+)	9.42
4	TS-C-AS-GGG'-DADS1	G(+)G(−)G(+)G(+);G(−)G(+)	5.79
5	TS-C-AS-GG'G-DADS1[b]	G(+)G(+)G(−)G(+);G(−)G(+)	8.47
6	TS-C-AS-G'GA-DADS1	G(+)antiG(+)G(−);G(−)G(+)	2.83
7	TS-C-AS-G'AG'-DADS1	G(+)G(−)antiG(−);G(−)G(+)	5.96
8	TS-C-AS-GG'A-DADS1[d]	G(+)antiG(−)G(+);G(−)G(+)	10.73
9	TS-C-AS-GAG'-DADS1	G(+)G(−)antiG(+);G(−)G(+)	7.55
10	TS-C-AS-G'GG-DADS1o	G(−)G(+)G(+)G(−);G(+)G(−)	0.00
11	TS-C-AS-GGG-DADS1o	G(−)G(+)G(+)G(+);G(+)G(−)	2.19
12	TS-C-AS-GG'G'-DADS1o	G(−)G(−)G(−)G(+);G(+)G(−)	7.75
13	TS-C-AS-G'G'G-DADS1o[e]	G(−)antiG(−)G(−);G(+)G(−)	8.67
14	TS-C-AS-G'GG'-DADS1o	G(−)G(−)G(+)G(−);G(+)G(−)	12.91
15	TS-C-AS-GG'A-DADS1o	G(−)antiG(−)G(+);G(+)G(−)	2.24
16	TS-C-AS-GAG-DADS1o	G(−)G(+)antiG(+);G(+)G(−)	4.48
17	TS-C-AS-G'GA-DADS1o	G(−)antiG(+)G(−);G(+)G(−)	5.62
18	TS-C-AS-G'AG-DADS1o	G(−)G(+)antiG(−);G(+)G(−)	5.64

(k)

No.	Name	Conformation $\chi_3, \chi_2, \chi_2', \chi_3'; \kappa_3, \kappa_2$	ΔG_{298}
1	TS-C-AS-G'GG-MeSSA1	G(−)G(+)G(+)G(−); G(+)G(−)	2.84
2	TS-C-AS-GGG-MeSSA1	G(−)G(+)G(+)G(+); G(+)G(−)	3.88
3	TS-C-AS-GG'G'-MeSSA1	G(−)G(−)G(−)G(+); G(+)G(−)	2.30
4	TS-C-AS-G'G'G-MeSSA1	G(−)G(+)G(−)G(−); G(+)G(−)	3.33
5	TS-C-AS-G'GG'-MeSSA1[f]	G(−)G(−)G(+)G(−); G(+)G(−)	8.50
6	TS-C-AS-GG'A-MeSSA1	G(−)antiG(−)G(+); G(+)G(−)	5.96
7	TS-C-AS-GAG-MeSSA1	G(−)G(+)antiG(+); G(+)G(−)	4.91
8	TS-C-AS-G'AG-MeSSA1	G(−)G(+)antiG(−); G(+)G(−)	0.00
9	TS-C-AS-G'GA-MeSSA1	G(−)antiG(+)G(−); G(+)G(−)	2.67

[a] TS-C-MeS-G'-DADS1 has two imaginary frequencies at 501.91 cm^{-1} and 8.87 cm^{-1}.
[b] TS-C-AS-GG'G'-DADS1 has two imaginary frequencies at 497.72 cm^{-1} and 17.86 cm^{-1}.
[c] χ_2 changed from 60.0° to 108.2° after optimization.
[d] TS-C-AS-GG'A-DADS1 has two imaginary frequencies at 516.99 cm^{-1} and 4.70 cm^{-1}.
[e] χ_2 changed from 60.0° to 173.3° after optimization, so the G'G'G notation should have changed to G'G'A.
[f] χ_2 changed from -60.0° to -112.3° after optimization.

Table S 8.13 Interaction analysis in transition states from the Cα nucleophilic substitution by AIM and NBO[a]

(a)

d(C-H···C)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
TS-C-MeS-G-DADS1o				
2.652	0.0095	0.0305	π(C=C)→σ*(C-H)	3.80
			σ(C-H)→π*(C=C)	1.30

TS-C-AS-G'GG-DADS1o				
2.606	0.0101	0.0306	π(C=C)→σ*(C-H)	6.88
			σ(C-H)→π*(C=C)	1.30
TS-C-MeS-G-MeSSA1				
2.630	0.0099	0.0307	π(C=C)→σ*(C-H)	4.81
			σ(C-H)→π*(C=C)	1.20
TS-C-AS-G'GG-MeSSA1(H)				
2.602	0.0100	0.0306	π(C=C)→σ*(C-H)	6.98
			σ(C-H)→π*(C=C)	1.06
TS-C-AS-G'AG-MeSSA1(G)				
2.545	0.0111	0.0341	π(C=C)→σ*(C-H)	8.14
			σ(C-H)→π*(C=C)	1.30

(b)

d(S⋯C)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
TS-C-MeS-G-DADS1o				
3.497	0.0080	0.0214	LP(p)(S17)→π*(C5=C6)	3.08
TS-C-AS-G'GG-DADS1o				
3.613	0.0076	0.0201	LP(p)(S17)→π*(C5=C6)	2.79
TS-C-MeS-G-MeSSA1				
3.378	0.0099	0.0261	LP(2)(S9)→π*(C2=C3)	3.66
TS-C-AS-G'GG-MeSSA1(H)				
3.327	0.0107	0.0283	LP(2)(S9)→π*(C2=C3)	4.67
TS-C-AS-G'AG-MeSSA1(G)				
3.357	0.0103	0.0270	LP(2)(S9)→π*(C2=C3)	4.09

[a] Distances (d) are in Å, ρ and $\nabla^2\rho$ are from the AIM analysis (in a.u.), and E(2) is from the NBO analysis (in kJ mol^{-1}).

Table S 8.14 Activation barriers ($\Delta G^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔG_{298}, kJ mol^{-1}) from the reaction of MeSH and (a) DADS or (b) DATS[a]

(a)

No.	Reaction	Reaction Type	$\Delta G^{\ddagger}_{298}$	ΔG_{298}
1	MeSH deprotonation	Protonation/deprotonation	58.96	ND
2	DADS + MeS^- → MeSA + ASS^-	Cα nucleophilic substitution	131.64	-37.12
3	DADS + MeS^- → MeSSA + AS^-	S nucleophilic substitution	65.95	-7.95
4	DADS + AS^- → DAS + ASS^-	Cα nucleophilic substitution	135.48	-28.94
5	MeSSA + MeS^- → MeSA + $MeSS^-$	Cα nucleophilic substitution	128.38	-37.55
6	MeSSA + MeS^- → DMDS + AS^-	S nucleophilic substitution	62.57	-9.42
7	MeSSA + AS^- → DAS + $MeSS^-$	Cα nucleophilic substitution	135.10	-29.37
8	ASS^- protonation	Protonation/deprotonation	ND	ND
9	ASSH + MeS^- → MeSSA + HS^-	H_2S release	45.26	-46.93
10	$MeSS^-$ protonation	Protonation/deprotonation	ND	ND
11	MeSSH + MeS^- → DMDS + HS^-	H_2S release	42.82	-50.12

(b)

No.	Reaction	Reaction Type	$\Delta G^{\ddagger}_{298}$	ΔG_{298}
12	DATS + MeS^- → MeSSA + ASS^-	Side-S nucleophilic substitution	49.36	-9.28
13	DATS + MeS^- → MeS_3A + AS^-	Mid-S nucleophilic substitution	63.74	-37.44
14	MeS_3A + MeS^- → DMDS + ASS^-	Side-S nucleophilic substitution	44.16	-37.58

[a] Reaction steps that already appeared from the reaction of MeSH and DADS are omitted.
[b] ND, not determined.

Table S 8.15 Calculated TS conformations from the reaction of GSH and DADS via Cα nucleophilic substitution and their relative free energies (ΔG_{298}, kJ mol^{-1})

Name	ΔG_{298}
TS1-C-GS-DADS	5.91
TS2-C-GS-DADS	7.91
TS3-C-GS-DADS[a]	ND
TS4-C-GS-DADS[c]	ND
TS5-C-GS-DADS	0.00

[a] TS3-C-GS-DADS changed to the reactant complex after optimization.
[b] ND, not determined.
[c] TS3-C-GS-DADS did not converge after several cycles of optimization.

Table S 8.16 The comparison between the full model TS and the small model TS in terms of relative free energies (ΔG_{298}, in kJ mol^{-1})

Full model TS	ΔG_{298}	Small model TS	ΔG_{298}
TS5-C-GS-DADS	0.00	TS-C-MeS-G-DADS12	0.00
TS1-C-GS-DADS	5.91	TS-C-MeS-G-DADS2	2.80
TS2-C-GS-DADS	7.91	TS-C-MeS-A-DADS18	6.21

Table S 8.17 Interaction analysis in TS5-C-GSH-DADS by AIM and NBO[a]

d(C-H···O)	ρ	$\nabla^2\rho$	Donor-Acceptor NBOs	E(2)
2.598	0.0074	0.0267	π(C=O)→σ*(C-H)	2.89
			LP(O)→σ*(C-H)	1.83
			σ(C-H)→π*(C=O)	0.34
2.697	0.0058	0.0226	LP(O)→σ*(C-H)	2.94
2.715	0.0072	0.0265	π(C=O)→σ*(C-H)	2.46
			LP(O)→σ*(C-H)	0.87
2.736	0.0056	0.0224	π(C=O)→σ*(C-H)	0.67
			LP(O)→σ*(C-H)	0.67

[a] Distances (d) are in Å, ρ and $\nabla^2\rho$ are from the AIM analysis (in a.u.), and E(2) is from the NBO analysis (in kJ mol^{-1}).

Table S 8.18 Calculated TS conformations from the reaction of MeS^- and (a) DMDS or (b) DPDS via Cα nucleophilic substitution and their relative free energies (ΔG_{298}, kJ mol^{-1})

(a)

No.	Name	Conformation χ_2; κ_2	ΔG_{298}
1	TS-C-MeS-A-DMDS	anti; G(−)	0.26

2	TS-C-MeS-G-DMDS	G(+); G(−)	1.26
3	TS-C-MeS-G'-DMDS	G(−); G(−)	0.00

(b)

No.	Name	Conformation χ_3, χ_2; κ_3, κ_2	ΔG_{298}
1	TS-C-MeS-A-DPDS1	G(+)anti; G(−)G(−)	0.00
2	TS-C-MeS-A-DPDS7	G(−)anti; G(+)G(+)	0.74
3	TS-C-MeS-G-DPDS7	G(−)G(+); G(+)G(+)	6.80
4	TS-C-MeS-G'-DPDS7	G(−)G(−); G(+)G(+)	5.16

Table S 8.19 Activation barriers ($\Delta G^{\ddagger}_{298}$, kJ mol^{-1}) and energies of reaction (ΔG_{298}, kJ mol^{-1}) comparisons for the reaction of MeS^- and (a) diallyl/dialkyl disulfides or (b) S-allyl/alkyl-methyl disulfides

(a)

No.	Reaction	$\Delta G^{\ddagger}_{298}$	ΔG_{298}
1	DADS + MeS^- → MeSA + ASS^-[a]	131.64	-37.12
2	DMDS + MeS^- → MeSMe + $MeSS^-$	138.17	-36.26
3	DPDS + MeS^- → MeSPr + $PrSS^-$	150.46	-39.00

(b)

No.	Reaction	$\Delta G^{\ddagger}_{298}$	ΔG_{298}
1	MeSSA + MeS^- → MeSA + $MeSS^-$[a]	128.38	-37.55
2	MeSSPr + MeS^- → MeSPr + $MeSS^-$	146.52	-38.90

(c)

No.	Reaction	$\Delta G^{\ddagger}_{298}$	ΔG_{298}
1	MeSSBn + MeS^- → MeSBn + $MeSS^-$	130.58	-37.82

[a] The energy profiles were reported in section 6.3.1.

8.1 References

1. Cheng, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J., *Cell* **1993,** *74* (2), 299.

2. O'Gara, M.; Zhang, X.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1999,** *287* (2), 201.

3. Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X., *Cell* **1994,** *76* (2), 357.

4. O'Gara, M.; Klimasauskas, S.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1996,** *261* (5), 634.

5. O'Gara, M.; Roberts, R. J.; Cheng, X., *J. Mol. Biol.* **1996,** *263* (4), 597.

6. Kumar, S.; Horton, J. R.; Jones, G. D.; Walker, R. T.; Roberts, R. J.; Cheng, X., *Nucleic Acids Res.* **1997,** *25* (14), 2773.

7. Zhou, L.; Cheng, X.; Connolly, B. A.; Dickman, M. J.; Hurd, P. J.; Hornby, D. P., *J. Mol. Biol.* **2002,** *321* (4), 591.

8. Shieh, F. K.; Youngblood, B.; Reich, N. O., *J. Mol. Biol.* **2006,** *362* (3), 516.

9. Youngblood, B.; Shieh, F. K.; Buller, F.; Bullock, T.; Reich, N. O., *Biochemistry* **2007,** *46* (30), 8766.

10. Takeshita, K.; Suetake, I.; Yamashita, E.; Suga, M.; Narita, H.; Nakagawa, A.; Tajima, S., *Proc. Natl. Acad. Sci. U.S.A.* **2011**.

11. Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J., *Sci.* **2011,** *331* (6020), 1036.

12. Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J., *Sci.* **2012,** *335* (6069), 709.

13. Zhang, Z. M.; Liu, S.; Lin, K.; Luo, Y.; Perry, J. J.; Wang, Y.; Song, J., *J. Mol. Biol.* **2015**.

14. Siedlecki, P.; Boy, R. G.; Comagic, S.; Schirrmacher, R.; Wiessler, M.; Zielenkiewicz, P.; Suhai, S.; Lyko, F., *Biochem. Biophys. Res. Commun.* **2003,** *306* (2), 558.

15. Fang, M. Z.; Wang, Y.; Ai, N.; Hou, Z.; Sun, Y.; Lu, H.; Welsh, W.; Yang, C. S., *Cancer Res.* **2003,** *63* (22), 7563.

16. Liu, Z.; Liu, S.; Xie, Z.; Pavlovicz, R. E.; Wu, J.; Chen, P.; Aimiuwu, J.; Pang, J.; Bhasin, D.; Neviani, P.; Fuchs, J. R.; Plass, C.; Li, P. K.; Li, C.;

Huang, T. H.; Wu, L. C.; Rush, L.; Wang, H.; Perrotti, D.; Marcucci, G.; Chan, K. K., *J. Pharmacol. Exp. Ther.* **2009,** *329* (2), 505.

17. Yoo, J.; Medina-Franco, J. L., *J. Comput. Aided Mol. Des.* **2011,** *25* (6), 555.

18. Weng, J.-R.; Lai, I. L.; Yang, H.-C.; Lin, C.-N.; Bai, L.-Y., *Phytother. Res.* **2014,** *28* (1), 49.

19. Joshi, M.; Rajpathak, S. N.; Narwade, S. C.; Deobagkar, D., *Chem. Biol. Drug Des.* **2016**, n/a.

20. Yoo, J.; Kim, J. H.; Robertson, K. D.; Medina-Franco, J. L., *Adv. Protein Chem. Struct. Biol.* **2012,** *87*, 10.1016/B978.

21. Peräkylä, M., *J. Am. Chem. Soc.* **1998,** *120* (49), 12895.

22. Zangi, R.; Arrieta, A.; Cossío, F. P., *J. Mol. Biol.* **2010,** *400* (3), 632.

23. Zhang, X.; Bruice, T. C., *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (16), 6148.

24. Yang, J.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y.; Broyde, S., *Biochemistry* **2013,** *52* (16), 2828.

25. Du, Q.; Wang, Z.; Schramm, V. L., *Proc. Natl. Acad. Sci.* **2016,** *113* (11), 2916.

26. Aranda, J.; Zinovjev, K.; Świderek, K.; Roca, M.; Tuñón, I., *ACS Catal.* **2016,** *6* (5), 3262.

27. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G., *Mol. Syst. Biol.* **2011,** *7* (1).