Understanding the roles of chalcogen and van der Waals dispersion interaction in biomolecules

A Thesis

Submitted in partial fulfillment of the requirements

of the degree of

Doctor of Philosophy

By

Vishal Annasaheb Adhav

20153401

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

(2022)

I would like to dedicate my thesis to my mother "Mai"

CERTIFICATE

Certified that the work incorporated in the thesis entitled **"Understanding the roles of chalcogen and van der Waals dispersion interaction in biomolecules"** submitted by **Vishal Annasaheb Adhav** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Prof. Saikrishnan Kayarat

(Supervisor)

Date: 15/03/2022

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Vishal Annasaheb Adhav

Roll No.20153401

Date: 14/03/2022

Acknowledgments

I want to express my eternal gratitude to my supervisor, Prof. Saikrishnan Kayarat, for his guidance and support during this program. This project would not have been completed without his input in discussing the topic, designing experiments or criticizing the results. His valuable suggestions helped me solve the problems that came while carrying out this project. The way he knows about doing good science very much influenced me and I felt glad while acquiring some of his scientific skills. I will never forget the support and help that he provided me in all the possible ways that he could besides the lab work.

I am sincerely grateful to IISER Pune for allowing me to work and providing adequate infrastructure. I want to thank the Council of Scientific and Industrial Research (CSIR) for providing fellowship during my Ph.D. program. I thank my RAC members Dr. Jeetender Chugh, Prof. Gadre S.R., Prof. Udgaonkar Jayant, for their input in my project in the yearly RAC meetings. They critically monitored the progress of my work, and their comment helped me in moving the project forward. I owe my sincere gratitude to Dr. Balanarayan Pananghat and his students, Prashant, Navin and Mishu, for teaching me computational chemistry and allowing me to work in their laboratory. I am extremely thankful to 5th-year BS-MS students Sanket Shelke, Nevin Zacharia and Gyan Mishra for their efforts and contribution to the various aspect of the project.

I thank the Head of the Biology division, Bio managers and staff from the bio office for ensuring the smooth progress of my work during the tenure. I am thankful for the technical support and help provided by the past lab members, Manasi Kulkarni, Neha Nirwan, Ishtiyaq Ahmed Khan, Mahesh Chand and undergraduate students. I learned many things from them. I am also thankful to the current members of the lab, Om Prakash Chauhan, Sujata Sharma, Vinayak Sadashivam, Singh Akhilesh, Rushik Bhatti, Akash Singh, Ufaq Bhat, Geetanjali Ghosh, Ashwin Uday, Nikki Dutt and current BS-MS students for their help.

I also want to thank Dr. Gayathri Pananghat and her lab members- Shrikant, Jyoti, Mrinmayee, Joyeeta, Vani, Sukanya, Soumyajeet, Apurba and other Ph.D. or BS-MS students for their inputs during lab meetings and their support in the lab. I also thank IISER Pune administrative staff for their help in various administrative and laboratory work. Lastly, I thank all well-wishers for their support and help during the work, whose names are not mentioned here.

Vishal Adhav

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Å	Angstrom
ºC	Degree Celsius
µg	Microgram
µM	Micromolar
3D	Three Dimensional
4mC	4-methylcytosine
5hmC	5-hydroxy-methylcytosine
5-IAF	5-(Iodoacetamido) fluorescein
5mC	5-methylcytosine
5-Methyl-dCTP	2'-deoxy-5-methylcytidine 5'-triphosphate
AAA+	ATPases Associated with diverse cellular Activities
AIM	Atoms In Molecules
au	Atomic Unit
BCP	Bond Critical Point
bp	Base Pair
BSA	Bovine Serum Albumin
BSSE	Basis Set Superposition Error
BTB	Bis-Tris
Ch-bond	Chalcogen bond
CHELPG	Charges from Electrostatic Potentials using a Grid-based method
CP	Critical Point
CSD	Cambridge Structural Database
dCTP	Deoxycytidine triphosphate
DEAE	Diethylaminoethyl
DMU	N−dimethyl urea
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E.coli	Escherichia coli
EDA	Energy Decomposition Analysis
EDTA	Ethylenediaminetetraacetic acid

G09	Gaussian09
GTP	Guanosine triphosphate
h	Hour
H-bond	Hydrogen Bond
IPTG	Isopropyl-D-1-thiogalactopyranoside
IR	Infra-Red
ITC	Isothermal Calorimetry
IUPAC	International Union of Pure and Applied Chemistry
K	Kelvin
kb	Kilobase pair
$kcal.mol^{-1}$	Kilocalorie per mole
KD	Dissociation constant
LBHB	Low Barrier Hydrogen Bond
mC	Modified cytosine
MD	Molecular Dynamics
MetRS	Methionyl-tRNA synthetase
min	Minute
mM	Millimolar
NADH	Nicotinamide adenine dinucleotide
NCI	Non-covalent interaction
nm	Nanometer
NMR	Nuclear Magnetic Resonance
NPT	Constant Number Pressure Temperature
NTD	N-Terminal Domain
NVT	Constant Number Volume Temperature
PCR	Polymerase Chain reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PES	Potential Energy Surface
QTAIM	Quantum theory of Atoms in Molecules
RMSD	Root Mean Square Deviation
RNA	Ribonucleic acid
rpm	Rotation per minute
SASA	Solvent Accessible Surface Area

SHB	Short Hydrogen Bond
SRA	SET and RING-associated
TET	Ten-eleven translocation
UDG	Uracil DNA Glycosylase
V	Volt
vdW	van der Waals

Synopsis

Understanding the roles of chalcogen and van der Waals dispersion interaction in biomolecules

Name: Vishal Annasaheb Adhav

Roll number: 20153401

Thesis supervisor: Prof. Saikrishnan Kayarat

Department: Biology, IISER, Pune

Date of registration: 1st August 2015

Chapter 1. Non-covalent interactions in biomolecular structure, stability, and function

In nature, there are a variety of non-covalent interactions that are critical for the stabilization and functioning of the biomolecules (Dill and MacCallum, 2012; Nick Pace et al., 2014; Xu et al., 2019) . Among the various non-covalent interactions, conventional hydrogen bonds (H-bond) and hydrophobic interactions are considered primary contributors to the stability of biomolecules (Kellis et al., 1988; Pace et al., 2014, 2011). For instance, the favorable enthalpic contribution of H-bond and hydrophobic interactions makes the protein folding process favorable by compensating for the unfavorable conformational entropy (Pace, 2009). It is commonly believed that hydrophobic interactions are considered to drive the folding of proteins, whereas H-bond confers specificities in their structure determination (Dyson et al., 2006; Fersht et al., 1985; Hubbard and Kamran Haider, 2010; Lins and Brasseur, 1995). Apart from this, the secondary interactions such as C-H···O, C-H··· π, π-stacking, and n→π* interactions are also present frequently in biomolecules (Bartlett et al., 2010; Brandl et al., 2001; Burley and Petsko, 1985; Derewenda et al., 1995; Dougherty, 1996; Lucas et al., 2016). These secondary interactions are necessary for sculpting biomolecular structures, their stability, and their function and for molecular recognition. For instance, these interactions together could contribute to the stability of proteins by almost 25% compared to the primary interactions (Newberry and Raines, 2019). Although the calculated strength of these secondary interactions are weaker than H-bonds, their frequent appearances, directional behavior, and cooperative nature make their incorporation in computer-aided molecular modeling to understand the biological process. This chapter provides a general introduction to various non-covalent interactions and discusses their origin, strength,

directional behavior and abundance in biomolecules. The chapter also highlights their significance in proteins structure and stability.

Furthermore, a comprehensive summary of the literature is presented, focusing on the divalent S mediated polar interactions and van der Waals (vdW) dispersion interaction. Based on this, we believe that the polar interactions such as H- and Chalcogen (Ch-) bonds that methionine and cysteines could form via divalent S are often overshadowed by their hydrophobic nature (Iwaoka et al., 2002; Nagano et al., 1999; Nelson, D., and Cox, 2005; Zhou et al., 2009). In particular, though these interactions are noted in protein structures, their interplay and mechanism of stabilizing or functioning proteins remain unaddressed. The presence of vdW dispersion interaction in proteins and its role in stabilizing biomolecules is well established (Holder et al., 2001; Kolář et al., 2011; Li et al., 2018; Ratnaparkhi and Varadarajan, 2000). However, its role in biomolecular recognition is poorly understood.

Thus, together, the significance of the two aspects of weak non-covalent interactions has been studied as part of my Ph.D. project *viz,*

1. understanding the role of sulfur-mediated chalcogen and hydrogen bond in proteins
2. the role of van der Waals dispersion interaction in sequence-specific protein-DNA recognition using McrBC as a model system.

Chapter 2. Characteristics of σ-hole mediated chalcogen bond involving divalent sulfur

Divalent sulfur (S) forms chalcogen bond (Ch-bond) with electron-rich regions of a molecule *via* its σ-hole (Politzer et al., 2013). Its directionality and cooperativity with other non-covalent interactions are important chemically and biologically. Here, dimers made of CH_3–S–X and O/N containing nucleophiles are studied and found to be stabilized by coexisting S···O/N and C–H···O/N interactions. Experimentally accessible electron density and molecular electrostatic potentials (MESP) revealed that reciprocity between S···O/N and C–H···O/N interactions in the stability of cumulative molecular interaction (ΔE) was dependent on the strength of σ-hole on S ($V_{s,max}$). Direct correlation between ΔE of dimers with $V_{s,max}$ of S supports the electrostatic nature of Ch-bond. In case of carbonyl nucleophile having two electron-rich regions, S preferentially approaches the carbonyl π–cloud rather than the lone pair. Direct correlation between the MESP minima (V_{min}) of π-cloud and strength of S–π interaction rationalizes the preferred directionality.

Chapter 3. Rules governing selectivity between sulfur mediated chalcogen versus hydrogen bond

Divalent S can interact with electrophilic centers via lone-pairs or nucleophilic centers via σ-hole (Politzer et al., 2013; Rosenfield et al., 1977). Factors that determine the nature of the bond formed by S with a functional group, which has electrophilic and nucleophilic centers proximal to each other, are not clear. In this chapter, based on the analyses of crystal structures of organic, organometallic, and protein molecules, and their MESP features, it is shown that a stronger lone-pair favors H-bond formation while a stronger σ–hole favors Ch-bond formation. S with lone-pairs and σ–holes of comparable strengths can simultaneously form H- and Ch-bonds with distal electrophiles and nucleophiles, respectively. These observations are presented in the form of the rules for selectivity between S-mediated chalcogen versus hydrogen bond. This could have an implication on assigning the nature of the interaction made by S in proteins where the position of H atoms is not available.

Chapter 4. The role of divalent sulfur mediated chalcogen and hydrogen bond in protein structure, stability, and substrate specificity

As discussed in chapters 2 and 3, the strengths of the lone-pairs and σ–holes on S are influenced by substituent effects. This chapter describes the bioinformatics analysis and energy calculations used to show that S-mediated H-bond and Ch-bond stabilize protein structures by various mechanisms, including capping α-helix termini, protecting free β-sheet edges by negative-design, and augmenting the stability of β-turns. Energy calculations indicated that the contribution of Ch-bond to stability can be as much as a conventional H-bond. Furthermore, we demonstrated that the disruption of a Ch-bond between the enzyme methionyl-tRNA synthetase and its substrate methionine affects substrate binding using mutagenesis based experiment. Thus, the study highlights the importance of S-mediated interactions for understanding protein folding and function, the development of improved strategies for protein structure prediction and design, and structure-based drug discovery.

Chapter 5. The role of van der Waals dispersion interaction in base-specific protein-DNA recognition

van der Waals dispersion interaction is considered one of the important components of protein stability (Holder et al., 2001). However, its role in molecular recognition is not clear. In a recent study, the cohesive solvent-solvent interaction was found to be the primary force behind the association of apolar molecules in solution (Yang et al., 2013). McrBC endonuclease, however, is proposed to recognize its target DNA base (methyl-cytosine) primarily through van der Waals interactions (Sukackaite et al., 2012). In this chapter, the role of this interaction on the base specificity of McrBC, if any, is discussed. The effect of the mutations at the base recognizing

pocket and intercalating residue of McrB on the modulation of target base specificity is presented. The high-resolution crystal structures of these mutant enzymes bound to DNA solved in order to gain insight into the molecular basis on the modulation in DNA base specificity is also presented. Along with this, using dispersion-corrected Density Function Theory (B3LYP-D3), computational studies were carried out to unravel the nature of the interactions with their contribution in stabilizing the flipped DNA base by McrBC, and its mutant enzymes. Together, this study provided direct evidence for the contribution of van der Waals dispersion interaction in biomolecular recognition.

References

Bartlett, G.J., Choudhary, A., Raines, R.T., Woolfson, D.N., 2010. N→π* interactions in proteins. Nat. Chem. Biol. 6, 615–620.

Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J., Hilgenfeld, R., 2001. C-H···π-interactions in proteins. J. Mol. Biol. 307(1), 357–377.

Burley, S.K., Petsko, G.A., 1985. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science. 229(4708), 23-28.

Derewenda, Z.S., Lee, L., Derewenda, U., 1995. The Occurence of C-H· · ·O Hydrogen Bonds in Proteins. J. Mol. Biol. 252(2), 248-262.

Dill, K.A., MacCallum, J.L., 2012. The protein-folding problem, 50 years on. Science. 338(6110), 1042-1046.

Dougherty, D.A., 1996. Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr, and Trp. Science. 271(5246), 163-168.

Dyson, H.J., Wright, P.E., Scheraga, H.A., 2006. The role of hydrophobic interactions in initiation and propagation of protein folding. Proc. Natl. Acad. Sci. 103(35), 13057–13061.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., Winter, G., 1985. Hydrogen bonding and biological specificity analysed by protein engineering. Nature. 314, 235–238.

Holder, J.B., Bennett, A.F., Chen, J., Spencer, D.S., Byrne, M.P., Stites, W.E., 2001. Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. Biochemistry 40(46), 13998–14003.

Hubbard, R.E., Kamran Haider, M., 2010. Hydrogen Bonds in Proteins: Role and Strength.

Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

Iwaoka, M., Takemoto, S., Tomoda, S., 2002. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein engineering. J. Am. Chem. Soc. 124(35), 10613–10620.

Kellis, J.T., Nyberg, K., Šail, D., Fersht, A.R., 1988. Contribution of hydrophobic interactions to protein stability. Nature. 333, 784–786.

Kolář, M., Kubař, T., Hobza, P., 2011. On the role of london dispersion forces in biomolecular structure determination. J. Phys. Chem. B 115(24), 8038–8046.

Li, J., Wang, Y., An, L., Chen, J., Yao, L., 2018. Direct Observation of CH/CH van der Waals Interactions in Proteins by NMR. J. Am. Chem. Soc. 140(9), 3194–3197.

Lins, L., Brasseur, R., 1995. The hydrophobic effect in protein folding. FASEB J. 9(7), 535–540.

Lucas, X., Bauzá, A., Frontera, A., Quiñonero, D., 2016. A thorough anion-π interaction study in biomolecules: On the importance of cooperativity effects. Chem. Sci. 7, 1038-1050.

Nagano, N., Ota, M., Nishikawa, K., 1999. Strong hydrophobic nature of cysteine residues in proteins. FEBS Lett. 458(1), 69-71.

Nelson, D., and Cox, M., 2005. Lehninger principles of biochemistry (4th ed.). W.H. Freeman and Company, New York, 2005.

Newberry, R.W., Raines, R.T., 2019. Secondary Forces in Protein Folding. ACS Chem. Biol. 14(8), 1677–1686.

Nick Pace, C., Martin Scholtz, J., Grimsley, G.R., 2014. Forces stabilizing proteins. FEBS Lett. 588(14), 2177-2184.

Pace, C.N., 2009. Energetics of protein hydrogen bonds. Nat. Struct. Mol. Biol. 16, 681–682.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Schell, D., Thurlkill, R.L., Imura, S., Scholtz, J.M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J.K., Takano, K., Hebert, E.J., Shirley, B.A., Grimsley, G.R., 2014. Contribution of hydrogen bonds to protein stability. Protein Sci. 23(5), 652–661.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Shirley, B.A., Hendricks, M.M., Iimura, S., Gajiwala, K., Scholtz, J.M., Grimsley, G.R., 2011. Contribution of Hydrophobic Interactions to Protein Stability. J. Mol. Biol. 408(3), 514–528.

Politzer, P., Murray, J.S., Clark, T., 2013. Halogen bonding and other σ-hole interactions: A perspective. Phys. Chem. Chem. Phys. 15, 11178–11189.

Ratnaparkhi, G.S., Varadarajan, R., 2000. Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. Biochemistry 39(40), 12365–12374.

Rosenfield, R.E., Parthasarathy, R., Dunitz, J.D., 1977. Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur. 1. Electrophiles and Nucleophiles. J. Am. Chem. Soc. 99(14), 4860–4862.

Sukackaite, R., Grazulis, S., Tamulaitis, G., Siksnys, V., 2012. The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 40(15), 7552–7562.

Xu, Z., Zhang, Q., Shi, J., Zhu, W., 2019. Underestimated Noncovalent Interactions in Protein Data Bank. J. Chem. Inf. Model. 59(8), 3389-3399.

Yang, L., Adam, C., Nichol, G.S., Cockroft, S.L., 2013. How much do van der Waals dispersion forces contribute to molecular recognition in solution? Nat. Chem. 5, 1006–1010.

Zhou, P., Tian, F., Lv, F., Shang, Z., 2009. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Bioinforma. 76(1), 151–163.

Chapter 1

Non-covalent interactions in biomolecular structure, stability, and function

1.1 Non-covalent interactions

J D van der Waals first identified cohesive forces distinct from and weaker than the covalent interactions that hold atoms or molecules together, referred to as non-covalent interactions. (Müller-Dethlefs and Hobza, 2000; Waals, 1873). Kollman defined non-covalent interactions as "*electron stay paired in reactant and product and there is no net change in chemical bonding*" (Kollman, 1977). The overall strength of non-covalent interactions usually ranges from -0.5 to -50 kcal.mol^{-1} and depends on the nature of the interacting atoms or molecules. For example, the calculated strength of the hydrogen bond (H-bond) in the water dimer is ~ -5 kcal.mol^{-1} (Kollman, 1984). Furthermore, for a deeper understanding of forces involved, nature of the interactions, or their characterization, the interaction energy of any two interacting atoms/molecules can be broken down into five components *viz* 1. Electrostatic, 2. Exchange-repulsion 3. Dispersion 4. Polarization and 5. Charge transfer, (Kollman, 1984; Morokuma, 1971; Ziegler and Rauk, 1977). Structural, spectroscopic, thermodynamic, and *ab initio* computational analyses are the primary and most valuable tools employed in literature so far to understand and study non-covalent interactions (Desiraju and Steiner, 1999; Mati and Cockroft, 2010; Müller-Dethlefs and Hobza, 2000).

The key benefit of these interactions in supra/macro-molecular chemistry is that they do not confer much rigidity (easy to form or break) because of their fragile nature, unlike covalent interactions (Mahadevi and Sastry, 2016; Müller-Dethlefs and Hobza, 2000). This unique feature introduces dynamicity within macromolecules such as protein or nucleic acids, which is responsible for their biochemical functions at ambient temperature (Müller-Dethlefs and Hobza, 2000). Additionally, these interactions often work in groups and are critical for determining the structures of macromolecules or their complexes, catalysis, and molecular recognition (Mahadevi and Sastry, 2016). Interactions made by non-polar groups, and polar groups involving Hydrogen (H), Carbon (C), Nitrogen (N), Oxygen (O), Phosphorus (P), and Sulfur (S) atoms, are the primary forces in biomolecular structure determination and their specificities. For instance, the double-helical structure of DNA, in solution, is a combined result of H-bonding, π-π stacking, van der Waals dispersion, and hydrophobic interactions (Kolář et al., 2011; Müller-Dethlefs and Hobza, 2000). This chapter aims to provide general features, the functional relevance of non-covalent interactions, and recent advances made towards understanding their role in biomolecular structure and functions—also a foundation for the questions addressed in this thesis.

Table 1.1 Classification and characteristics of hydrogen bonding interaction.

	Strong H-bond	Moderate H-bond	Weak H-bond
	F-H···F⁻	N-H···O=C	C-H···O
Interaction energy (kcal.mol^{-1})	-15 to -40	-4 to -15	< -4
Δ(X-H), Å	0.005 to 0.2	0.01 to0.05	< 0.01
H···A, Å	1.2 to 1.5	1.5 to 2.2	2.0 to3.0
Shorter than sum of van der Waals	100%	~100%	30-80%

[a]All values are adapted from Desiraju and Steiner, 1999.

1.2 Hydrogen bond

The latest definition provided by IUPAC states that "*The hydrogen bond is an attractive interaction between a hydrogen atom from a molecule or a molecular fragment X-H in which X is more electronegative than H, and an atom or a group of atoms in the same or a different molecule, in which there is evidence of bond formation"*(Arunan et al., 2011). H-bonds are, in general, dominated by electrostatic components and, based on the Energy Decomposition Analysis (EDA), a small contribution from charge transfer and polarization are also reported frequently (Kollman and Allen, 1972; Umeyama and Morokuma, 1977). D-H···A is a general representation for the H-bonds, where D and A is H bond donor and acceptor atom/group, respectively (Figure 1.1 **A**). Three dots between H and A represents the H-bond. The strength of the interaction can be in the range of -1 to -40 kcal.mol^{-1}, depending on the electronegativity of D and A and their environment (Desiraju and Steiner, 1999; Jeffrey and Saenger, 1991). Because of their diverse nature, H-bonds are divided into three groups; strong, moderate, and weak H-bonds (Table 1.1) (Desiraju and Steiner, 1999; Jeffrey and Saenger, 1991). In biomolecules, most H-bonds are either moderate or weak in nature. For example, conventional H-bond (O/N-H···O bond), C-H···O, C-H···π, O/N-H···π, and S-mediated H-bonds are typical in proteins. Below is a brief summary of these H-bonds and their role in biomolecular structure and function.

1.2.1 Conventional hydrogen bond

In 1951, Linus Pauling suggested that the peptide chain can fold into a unique conformation due to H-bonds between backbone amino and carbonyl groups (Mirsky and Pauling, 1936; Pauling et al., 1951; Pauling and Corey, 1951). This revolutionary observation had a significant effect on understanding the architecture of the proteins. In particular, H-bonds are considered as a driving force for forming regular and recurrent conformation of adjacent amino acids in proteins referred

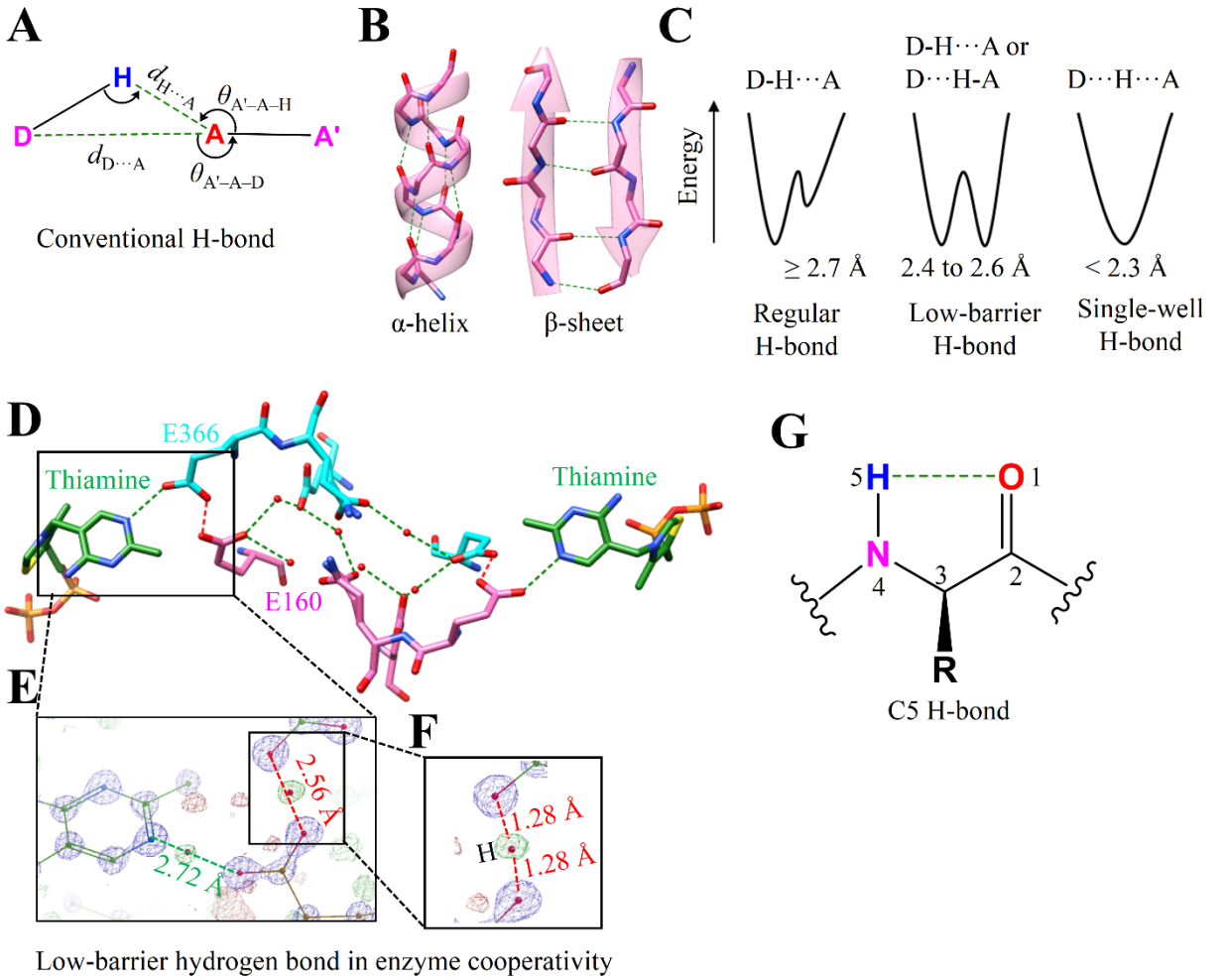


Figure 1.1 Conventional hydrogen bond. (**A**) The geometry and stereochemical parameters used to identify a conventional H-bond in biomolecules. (**B**) Typical structures of α-helix and β-sheet that result from a network of backbone H-bonds. All these H-bonds are shown in green with their distances < 3.5 Å. (**C**) Energy profile of H-bond as a function of the location of H atom between D and A highlighting the difference between regular, low-barrier, and single well H-bond. (**D**) The 0.97 Å resolution crystal structure of the human transketolase with substrate–thiamine intermediates bound (green) zoomed to show a channel and intervening residues communicating two active sites of the homo-dimer (each monomer colored in cyan and pink) (PDB ID: 4KXW). The H-bond network that connects these sites is highlighted in green (regular H-bonds) and red (LBHB). (**E**) Zoomed section showing residues Glu160, Glu366 and the portion of thiamine superposed with the $2F_O - F_C$ electron density map at 6σ contour level and $F_O - F_C$ at 2.7σ contour level. The regular H-bond between thiamine and Glu366 (green) and LBHB between Glu366 and Glu166 (red) are also highlighted. (**F**) A magnified section shows the occurrence of electron density almost precisely halfway between the Glu366 and Glu160, confirming an LBHB. Mutation of Glu160 to Gln disrupts the LBHB, resulting in a five-fold decrease in the enzyme's catalytic constant (k_{cat}). Figure 1.1 **D-F** were generated using the adaptation of data from Dai et al., 2019. (**G**) A definition and geometry of the C5 H-bond.

to as secondary structural elements (Pauling et al., 1951; Pauling and Corey, 1951). The most widespread of such structures in proteins are α-helices and β-sheets (Figure 1.1 **B**). The specificities that determine these structures in the proteins are primarily because of the highly directional behavior of these backbone H-bonds. Ramachandran plot, which relies on the backbone torsion angles of the peptide chain, is a classical biophysical tool to understand the architecture of proteins (Ramachandran et al., 1963). Akin to this, a recently developed rotational descriptor solely based on the pattern of the backbone H-bonds could also be used to identify the secondary and tertiary structures and proposed of great significance in protein designing and engineering (Penner et al., 2014). Besides its role in protein structures, conventional H-bonds are also prevalent in other biomolecules or complexes. They often contribute to enzyme-substrate, receptor-ligand, antigen-antibody, or protein-nucleic acid recognition (Bitencourt-Ferreira et al., 2019; Bulusu and Desiraju, 2020; Chen et al., 2016; Fersht, 1987; Rohs et al., 2010).

A structure-based approach identifies most non-covalent interactions, and H-bond is no exception (Hubbard and Kamran Haider, 2010). In proteins, the identification of H-bond is not trivial because of the loss of H atom's positional information in structures determined using X-ray crystallography. Thus, one often relies on stereochemical criteria such as $d_{D\cdots A} < 3.5$ Å and $\theta_{A'-A-D} > 90.0°$ to identify conventional H-bond, where atom A' is covalently linked to atom A (Hubbard and Kamran Haider, 2010; Torshin et al., 2002). For those cases where H positions are available or predicted, three additional criteria such as $d_{H\cdots A} < 2.5$ Å, $\theta_{D-H-A} > 90.0°$ and $\theta_{A'-A-H} > 90.0°$ are also employed (Hubbard and Kamran Haider, 2010; Torshin et al., 2002) (Figure 1.1 **A**). Note that the H-bond between the carbonyl O and amide N usually has a distance that peak at 2.9 Å in the proteins (Hubbard and Kamran Haider, 2010; Sticke et al., 1992). Although specificities in the determination of protein structures are a direct consequence of conventional H-bond, its contribution to the overall stability of proteins remains controversial. The predicted interaction energy of these H-bonds or calculated using *ab initio* methods ranges within -5 to -6 kcal.mol^{-1} (Fleming and Rose, 2005; Hubbard and Kamran Haider, 2010; Kollman and Allen, 1972). On the contrary, using mutagenesis-based approach, Pace et al. found that H-bond can contribute to protein stability by an average of -1.1 ± 1.0 kcal.mol^{-1} (Pace et al., 2014). This observation is in line with a hypothesis made by Pauling, which states that the effective interaction energy of the H-bond can be up to -2 kcal.mol^{-1} for proteins in an aqueous medium; later on, Fersht et al. confirmed this hypothesis (Fersht et al., 1985; Pauling and Corey, 1951). The disagreement in the calculated and observed strength is mainly because of the H-bond environment. For example, equilibrium hydrogen/deuterium fractionation analysis showed that the strength of the interaction could be up to -6 kcal.mol^{-1} in membrane α-helices because of the low dielectric and low water

environment (Cao et al., 2017). Although the strength of the H-bond interaction is weaker than predicted in proteins and largely depends on the environment, their cumulative effect contributes to the protein stability and folding by compensating for the destabilizing conformational entropy (Bolen and Rose, 2008; Pace, 2009).

Furthermore, proteins structures obtained using the x-ray crystallography, with a resolution better than 1 Å, reveals interesting features of conventional H-bonds (Rajagopal and Vishveshwara, 2005; Zhou and Wang, 2019). For example, analyses of ultra-high resolution protein structures led to the identification of the short H-bonds (SHBs) where $d_{D\cdots A} \leq 2.7$ Å is common (Rajagopal and Vishveshwara, 2005; Zhou and Wang, 2019). In these SHBs, the more sharing of H atom in H-bond is because of the decreased potential barrier and more significant enthalpy contribution than that seen in regular conventional H-bonds. When $d_{D\cdots A}$ is within 2.4 to 2.6 Å, such SHBs could enter the low-barrier H-bond (LBHB) region (Figure 1.1 **C**) (Hosur et al., 2013; Zhou and Wang, 2019). The strength of such H-bond could be within -10 to -20 kcal.mol^{-1} (Kemp et al., 2021; Shan et al., 1996). These LBHBs can easily be identified in the structures determined at ultra-high resolution and are critical for enzyme cooperativity, allostery, and catalysis (Figure 1.1 **D**-**F**) (Cleland and Kreevoy, 1994; Dai et al., 2019; Gerlt et al., 1997).

In addition, the recently characterized intra-residue H-bond between carbonyl O and amino N, referred to as C5 H-bond (Figure 1.1 **G**), is proposed to stabilize flat β-sheets in the amyloid state (Newberry and Raines, 2016). To summarize, some of these findings, in short, point towards the bright future for LBHB in the field of protein designing, and their application in other fields such as synthetic biology might have colossal scope.

1.2.2 C-H···O bond

Compared to conventional H-bonds, C-H···O bonds are weak and often referred to as non-conventional/non-canonical H-bonds (Table 1.1) (Desiraju and Steiner, 1999). The interaction was debated for a long time because of the lower electronegativity of C compared to O/N, otherwise hypothesized to stabilize proteins in 1960 (Ramachandran and Venkatachalam, 1966). Based on the evidence provided by IR spectroscopic data, gas-phase studies, and small molecule crystal structure, C-H···O interaction is now considered bonafide H-bond and has been observed in biomolecules (Derewenda et al., 1995; Desiraju and Steiner, 1999; Horowitz and Trievel, 2012). For example, PDB analysis of protein crystal structures revealed a 0.2 Å shorter contact ($d_{C\cdots O}$

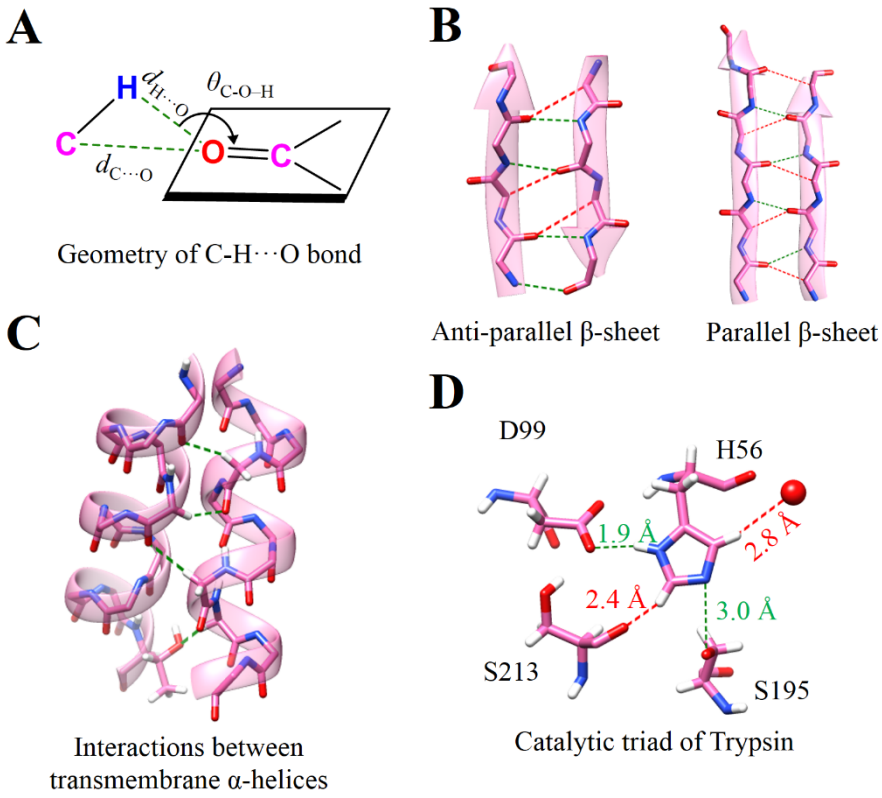


Figure 1.2 C-H···O bond. (**A**) The geometry and stereochemical parameters used to identify the C-H···O bond in biomolecules. The most favorable value for $d_{C\cdots O}$, $d_{H\cdots O}$ and $\theta_{C\text{-}O\text{-}H}$ are ~3.5 Å, ~2.8 Å, and ~135°, respectively (Derewenda et al., 1995). (**B**) A representative example of the network of C-H···O bonds in parallel and antiparallel β-sheets that are coexisting with backbone H-bonds. All these backbone H-bonds (< 3.5 Å) and C-H···O bonds (< 4.0 Å) are shown in green and red, respectively. The mean $d_{H\cdots O}$ is ~2.4 Å (Derewenda et al., 1995) and have similar energetic stabilization for both types of β-sheets (Scheiner, 2006). (**C**) Parallel right-handed helix-helix interaction in glycophorin A stabilized by C^{α}-H···O bond where $d_{H\cdots O}$ < 2.6 Å (PDB ID: 1AFO). (**D**) C^{ε}-H···O bond at the active site of trypsin made up of catalytic triad by Asp-His-Ser residues found to be responsible for its functioning (PDB ID: 1AFO).

~3.5 Å) between C and O than the sum of their van der Waals radii (Figure 1.2 **A**) (Derewenda et al., 1995). A similar observation made by Jiang et al. at the interface of protein-protein complexes, where $d_{C\cdots O}$ was < 3.3 Å (Jiang and Lai, 2002) and correlates with the one made in small molecules, where H positions are more reliable when obtained from x-ray crystallography (Desiraju and Steiner, 1999). C-H bond lengths being a 0.2-0.3 Å longer than their ideal value in the ultra-high resolution protein structures have argued to be because of potential C-H···O interactions which also conclusively provide an evidence for the interaction in proteins (Horowitz and Trievel, 2012). Calculations suggests that the strength of the C-H···O interaction could be within -1 to -2 kcal mol^{-1}. (Nick Pace et al., 2014). Thus, its high abundance in proteins and

reasonable strength could drive the formation and stabilization of secondary structures, particularly β-sheets, similar to that of conventional H-bond (Figure 1.2 **B**) (Horowitz and Trievel, 2012).

Furthermore, since it is well established that the strength of the C-H···O interaction depends on the acidity of the C-H group, H atoms of C^{α} of backbone and C^{ε} of His are prone to participate in the formation of an interaction (Derewenda et al., 1994; Horowitz and Trievel, 2012). In fact, these $C^{\varepsilon}/C^{\alpha}$-H···O interactions found to be of comparable in strength to that of conventional H-bond (Scheiner, 2006; Schmiedekamp and Nanda, 2009). The C^{α}-H···O interactions made by these C^{α}-H groups affect protein structure and function (Horowitz and Trievel, 2012). For example, multiple C^{α}-H···O bonds noted in between α-helices are critical for the determination of the stability and specificities of helix-helix interactions in transmembrane proteins (Figure 1.2 **C**) (Senes et al., 2002). In addition, crystal structure and NMR investigation identified C^{ε}-H···O bond in catalytic triad formed by Asp/Glu-His-Ser residues at the active site of serine hydrolases (Figure 1.2 **D**) (Ash et al., 2000; Derewenda et al., 1994). This enzyme has the potential to carry out acylation and deacylation during hydrolysis. The C^{ε}-H···O bond at the active site controls the conformation of imidazole ring responsible for the enzyme's dual functionality (Ash et al., 2000).

Besides its role in protein structure and function, C-H···O interactions are also crucial in biomolecular recognition. For example, the basis for the high specificity for the methyl group of 5-methylcytosine (5mC) at CpG sites in DNA is because of the C-H···O bond, which highlights its significance in target-specific protein DNA interactions (Nikolova et al., 2018). The high strength of the C-H···O interaction formed by the C-H group adjacent to ammonium cation (N+-C-H), also referred to as N+-C-H···O interaction affects the activity of the inhibiter against G9a-like protein (Itoh et al., 2019). Thus, employing N+-C-H···O interaction in designing drug candidates that carry C-H groups adjacent to ammonium cation could be a useful strategy (Itoh et al., 2019). Together, C-H···O interaction is argued as being equally important as conventional H-bonds in biomolecular recognition. Its ability to stabilize protein-protein complexes by an average of ~17% (sometimes up to ~50%) compared to the sum for all non-covalent forces supports this argument (Jiang and Lai, 2002).

1.2.3 C-H···π interaction

C-H···π is another weak interaction that involves the C-H group as an H-bond donor and π electrons as an H-bond acceptor (Figure 1.3 **A**). It is geometrically similar to the other H-bonds,

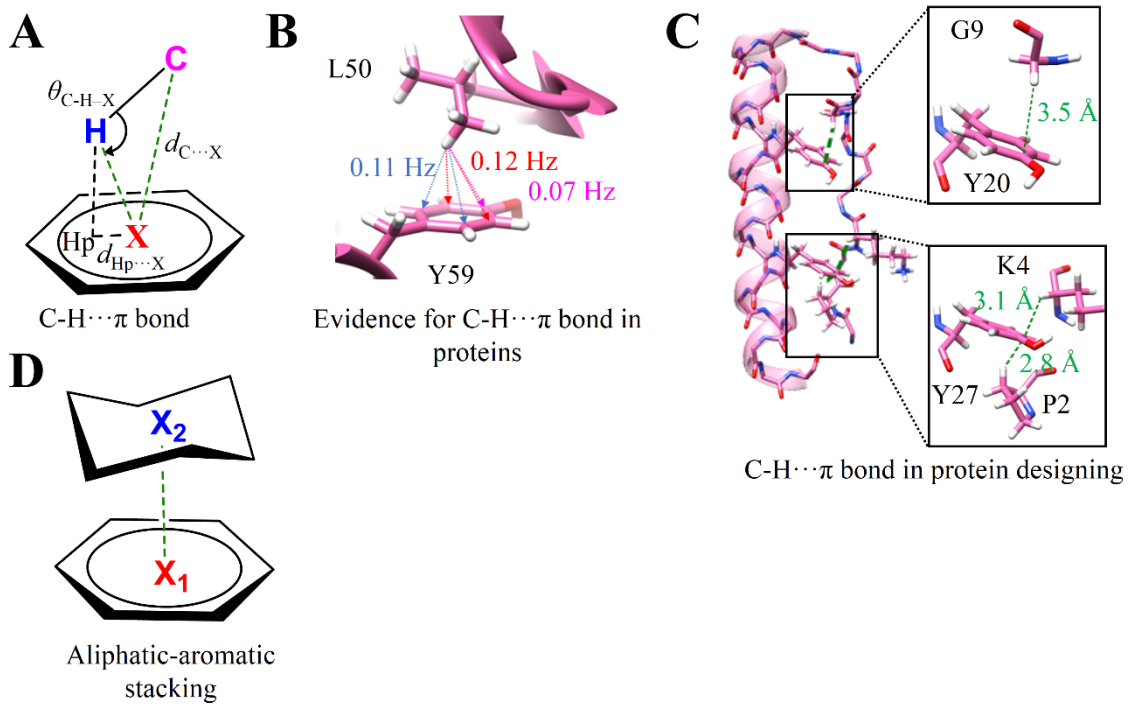


Figure 1.3 C-H··· π bond. (**A**) The geometry and stereochemical parameters used to identify the C-H···π bond in biomolecules (Brandl et al., 2001). (**B**) The observation of *J*-coupling in NMR based experiment between methyl and π groups in ubiquitin (PDB ID: 1UBQ) because of the C-H···π interaction (the values are adapted from Plevin et al., 2010). (**C**) NMR structure of a designed miniproteins, PPα-Tyr, and the magnified region show the C-H···π interactions stabilizing association of α-helix and polyproline II helix in PPα-Tyr (PDB ID: 5LO2). The distances for all C-H···π interactions were < 2.6 Å. (**D**) The geometry of aliphatic and aromatic stacking driven by multiple C-H···π interactions.

and the ability to exist in water makes C-H···π interaction biologically interesting (Newberry and Raines, 2019; Nishio et al., 2014). *ab initio* calculations suggest that the strength of the interaction ranges from -1.5 to -2.5 kcal.mol^{-1}, similar to that of the C-H···O interaction (Nishio et al., 2014). Many recent studies showed that dispersion and electrostatic components drive the formation of interaction (Nishio et al., 2014; Tsuzuki, 2012). Surprisingly, the possible role of the hydrophobicity of CH and π containing groups behind their association is negligible, otherwise expected because of their non-polar nature. Osmometry-based measurements support this, as the strength of C-H···π interaction between aliphatic-aromatic motifs has been found to be threefold higher than the hydrophobically driven aliphatic-aliphatic counterparts (Cheng et al., 2020).

To identify the C-H···π interaction in proteins, Brandl–Weiss provided the following cut-offs for three structure-based parameters *viz*. **1.** Distance between C and center of mass (X) of the π ring ≤ 4.5 Å, **2.** The angle between C-H···X ≥ 120, and **3.** Distance between Hp and X are within

1.0 or 1.2 Å; Hp is H's vertical projection on the π system (Figure 1.3 **A**) (Brandl et al., 2001). Given the uncertainties of the H atoms position in most of the protein structures, the CHPI program uses the computationally optimized H position to identify and obtain reliable geometry of the C-H···π interaction (Umezawa and Nishio, 2000). Many similar studies characterized this interaction using structural database analysis (Nishio et al., 2014); however, An observation of *J*-coupling between methyl and π group forming C-H···π interaction provided direct NMR-based evidence for the interaction in proteins (Figure 1.3 **B**) (Plevin et al., 2010). Also, an upfield shift for H of C^{α} of the tripeptide interacting with either Trp, Phe, Tyr, or His, is another evidence for the C-H···π interactions in peptides (Ganguly et al., 2012). The experimentally measured stability provided to this tripeptide by the C-H···π interaction was up to -1.0 kcal.mol^{-1}, half of the calculated strength of the interaction (Ganguly et al., 2012).

In proteins, three-quarters of Trp, half of Phe/Tyr, and a quarter of His residues forms C-H···π interaction that often water shielded and proposed to contribute to the favorable folding enthalpy (Brandl et al., 2001). In addition, a network of C-H···π interactions which are frequent in intra-, and inter-secondary structures could also stabilize the proteins similar to conventional H-bonds (Brandl et al., 2001; Kumar and Balaji, 2014). For example, an altered thermostability of the designed miniproteins by single C-H···π interaction might support this argument (Figure 1.3 **C**) (Baker et al., 2017). Apart from this, the interaction also plays a crucial role in biomolecular recognition; in particular, its roles in carbohydrate-protein interaction are well documented (Asensio et al., 2013; Hudson et al., 2015). The stacking of the aliphatic sugar ring of carbohydrates with the aromatic protein residues is mainly because of multiple C-H···π interactions. It is considered a driving force behind the carbohydrate-protein association (Figure 1.3 **D**) (Asensio et al., 2013; Hudson et al., 2015). Interestingly, a recent study showed that 39% of all carbohydrate and protein complexes in PDB had C-H···π interaction and could contribute favorably up to -8 kcal.mol^{-1} to the overall binding (Houser et al., 2020). Similarly, analysis of 130 structures of protein-DNA complexes from PDB revealed that 40% of all contacts had sugar-π stacking comprising C-H···π and lone pair···π interactions, suggesting its critical role in protein-DNA recognition (Wilson et al., 2014).

1.2.4 S-mediated H-bond

Divalent S widely occur in proteins in the form of cysteine and methionine. Early studies showed that the S could also act as a donor (S-H group of cysteine) as well as acceptor (S of methionine), similar to that of O or N (Figure 1.4 **A**) (Copley et al., 1939; Heafield et al., 1942). However, there are certain differences between S-mediated and conventional H-bonds. For example, though their

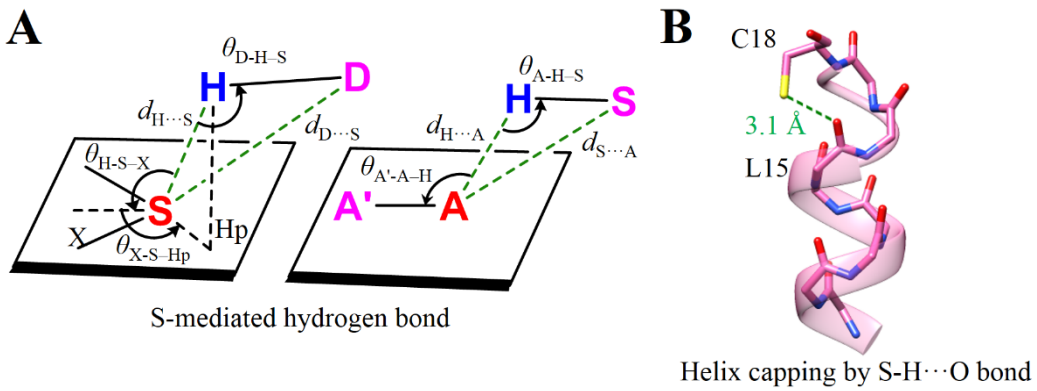


Figure 1.4 Divalent sulfur mediated H-bond. (**A**) The geometry and stereochemical parameters used to identify the S-mediated H-bond in biomolecules (Zhou et al., 2009). The favorable values of these parameters are as follow, $d_{H\cdots S}$ = 2.74 Å, $d_{D\cdots S}$ = 3.52 Å, $d_{H\cdots A}$ = 2.51 Å, $d_{S\cdots A}$ = 3.50 Å, $\theta_{D\text{-}H\text{-}S}$ = 141.1°, $\theta_{H\text{-}S\text{-}X}$ = 119° $\theta_{X\text{-}S\text{-}Hp}$ = 137°, $\theta_{A\text{-}H\text{-}S}$ = 136.5° and $\theta_{A'\text{-}A\text{-}H}$ = 117.4°, (**B**) A representative example of S-mediated H-bond (S-H···O interaction) in stabilizing the C-terminus of α-helix by helix capping (PDB ID: 1CPV) (Aurora and Rose, 1998; Gregoret et al., 1991).

strengths are comparable (in the ranges of -4.5 to -5.5 kcal.mol^{-1}), it is dominated by dispersion effect than electrostatic for S-mediated H-bonds (Biswal et al., 2012; Biswas and Wategaonkar, 2009; Howard and Kjaergaard, 2008; Rao Mundlapati et al., 2015). In addition, S-mediated H-bond length in proteins usually appears longer (~ 3.4 Å between D and A) than the conventional H-bond (~ 2.9 to 3.0 Å between D and A), presumably because of the larger size of the S atom and its diffused electron cloud (Figure 1.4 **A**) (Gregoret et al., 1991; Sticke et al., 1992). This was validated by employing a large dataset of protein structures and optimized H positions where the average $d_{H\cdots A}$ was ~2.8 Å ($d_{S\cdots A}$ was ~3.5 Å) and $\theta_{S\text{-}H\cdots A}$ was ~140° (instead of being close to 180°) (Figure 1.4 **A**) (Zhou et al., 2009).

Divalent S in proteins, in general, has a higher preference (5:1 ratio) to donate H than accept, which makes the sulfhydryl group of reduced cysteine interesting (Zhou et al., 2009). For instance, H-bond between sulfhydryl group of cysteine (i+4th residue) and carbonyl O (i^{th} residue) is proposed to stabilize the C-terminal of the α-helix by a mechanism commonly referred to as helix capping (Figure 1.4 **B**) (Aurora and Rose, 1998; Gregoret et al., 1991). In a surprising scenario, buried cysteines prefer to interact with the charged residues. On the contrary, deprotonated cysteines (negatively charged) prefer to interact with neutral residues, presumably because of the local dielectric constant (Mazmanian et al., 2016). However, removing one of such H-bond between S of iron (Fe) chelated cysteine with the side chain of Thr/Ser decreases the midpoint potential of the Fe-S assembly, in turn compromising the enzyme activity (Denke et al.,

1998). These observations are a few examples of the potential benefits of S-mediated H-bonds in understanding the protein structure, stability, folding, and enzyme functioning.

1.3 Hydrophobic interaction

In a ground-breaking review published in 1959, Walter Kauzmann suggested that the unique configuration of the proteins could be a result of the hydrophobic effect and the force that stabilizes folded state of the proteins (Kauzmann, 1959). Later on, for globular proteins, Tanford noted that hydrophobic interaction alone could explain the instability of the unfolded state (Tanford, 1962). Together, these observations broke the paradigm of conventional H-bonds being the primary requirement for the architecture of proteins (Nick Pace et al., 2014). The structure of myoglobin solved by Kendrew where he noticed that non-polar residues mainly packed the interior provided conclusive evidence for the hydrophobic interaction driving protein folding (Kendrew, 1963; Kendrew et al., 1960; Lee and Richards, 1971). In fact, closely packed solid spheres can occupy up to 71% space, whereas the protein atoms can occupy 75 % of the space in the interior (Klapper, 1971; Lee and Richards, 1971; Nick Pace et al., 2014); thus, protein interior can be considered solid rather than liquid, highlighting the packing potential of the hydrophobic interaction (Klapper, 1971; Lee and Richards, 1971; Nick Pace et al., 2014). Interestingly, Dill proposed that any polymer carrying residues incompatible in a given solvent when compacted to maximum might have a protein-like interior (Chan and Dill, 1990). Together, this highlights the universal non-specific nature of the hydrophobic interaction and its possible contribution to protein folding and structures.

IUPAC defined hydrophobicity as "*The association of non-polar groups or molecules in an aqueous environment which arises from the tendency of water to exclude non-polar molecules*" (van de Waterbeemd et al., 1997). Additionally, IUPAC suggested the following definition for hydrophobic interaction "*The tendency of hydrocarbons (or of lipophilic hydrocarbon-like groups in solutes) to form intermolecular aggregates in an aqueous medium and analogous intramolecular interactions. The name arises from the attribution of the phenomenon to the apparent repulsion between water and hydrocarbons. However, the phenomenon ought to be attributed to the effect of the hydrocarbon-like groups on the water-water interaction. The misleading alternative term "hydrophobic bond' is discouraged"* (Muller, 1994). Thus, hydrophobic interaction originates from unfavorable interactions between hydro-carbons/non-polar and water molecules (Figure 1.5) (Chandler, 2005). Also, because of the "ordered-water" at their interface, the entropy change for the transfer of non-polar molecules from oil to water is large and negative (Chandler, 2005). This effect is often accomplished by a compromised H-bond

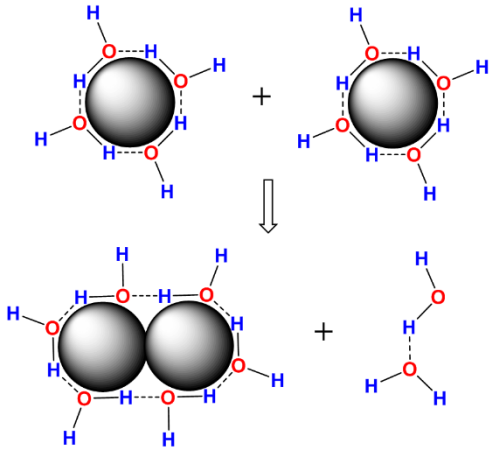


Figure 1.5 Principle of the hydrophobic effect. The compromised H-bonds and ordering water molecules around the non-polar molecules (gray spheres), shown in the top panel, drive their association (bottom panel) in an aqueous solution.

network around the non-polar molecules, thus providing a basis for the interaction (Figure 1.5) (Chandler, 2005).

In the model of protein folding, hydrophobic interactions have received much attention (Dyson et al., 2006). Hydrophobic interactions are proposed to initiate the folding at the early state of unfolded protein, referred to as a statistical coil, which subsequently folds into the native structure (Anfinsen, 1973; Dyson et al., 2006). This presumably occurs between nearby non-polar residues to avoid a large unfavorable entropy change. Studies have shown that the removal of single $–CH_2–$ group destabilizes proteins by 1.1 ± 0.5 kcal.mol^{-1} (Kellis et al., 1988; Pace et al., 2011). Also, an average of 60% contribution to the stability of proteins, mostly by aromatic residues, support the primary role of hydrophobic interaction in deciding the architecture of proteins, as predicted by Walter Kauzmann (Pace et al., 2011). Since most non-polar residues are present at the core of proteins, the contribution of hydrophobic interaction behind protein-protein association is negligible and dominated mainly by polar interactions (Yan et al., 2008).

It is well established that double-helical assembly of DNA results from H-bonds, π-π stacking, hydrophobic effect, and dispersion interaction. Recent studies showed that the formation and stabilization of DNA assembly in water result from the hydrophobic interaction. The contribution of H-bonds is negligible (Feng et al., 2019; Lindman et al., 2021). In fact, the dry environment created by the hydrophobic effect at the interior of the DNA enhances the ability of H-bonds to introduce base specificity, highlighting their cooperative nature in the architecture of DNA (Feng et al., 2019). Furthermore, addition of semi hydrophobic agents in aqueous solution

are found to decrease the stacking energy in DNA, which result in the formation of holes that can act as intercalation sites; this mechanism is called hydrophobic catalysis (Feng et al., 2019). Thus, high sensitivity towards the external environment, high tunability, low functional effect, and introduction of amphiphilic nature to the nucleic acid are particulars of hydrophobic interactions. These properties can make hydrophobic interaction useful for creating higher-order self-assembly of DNA-based biomaterials, drug/gene-delivery or stimuli-responsive systems, and controlling inter-cellular interactions (Xiao et al., 2020).

1.4 Other directional non-covalent interactions

1.4.1 π···π stacking interaction

As discussed in section 1.3, non-polar residues come close to each other by entropy-driven, hydrophobic effect and packing of these residues in the interior of proteins, often non-specific and non-directional in nature. The aromatic side chains of Phe, Tyr, Trp, or His residues, are also considered non-polar, however, can stack on each other to form specific, directional, enthalpy-driven interaction referred to as π···π stacking interaction (Figure 1.6 **A**-**B**) (Burley and Petsko, 1985; Hunter et al., 1991; Meyer et al., 2003). The short-range charge transfer and long-range dispersion effect contributes to its attractive nature and complete quantum-mechanical origin (Carter-Fenk and Herbert, 2020; Tsuzuki et al., 2002). Two geometries of these stacking interactions, *viz*., parallel-displaced and T-stacking, are standard (Figure 1.6 **A**-**B**); the former occurs more often in proteins and is energetically more favorable than the later (Chelli et al., 2002). Note that section 1.2.2 discusses the C-H···π interaction, primarily drives the T-stacking interactions (Figure 1.6 **A**).

Database analyses indicates that ~60% of aromatic side-chains can participate in π-π stacking interaction. The equilibrium distance between the centroids of the two aromatic rings is ~4.5 Å, mainly for parallel-displaced geometry (Figure 1.6 **B**) (Burley and Petsko, 1985). Also, particular preferences for their appearance in proteins could make them helpful in protein designing. For example, Tyr-Tyr stacking often occurs at the proteins interface, and Tyr-Phe or Phe-Phe is preferentially found at the interior of proteins, highlighting the possible contribution of an aqueous environment behind their occurrences (Chelli et al., 2002). A double-mutant cycle indicates that each pair of interactions can stabilize the protein up to -1.3 kcal.mol^{-1}, in line with the calculated strength, which ranges between -1 to -2 kcal.mol^{-1} (Serrano et al., 1991). Furthermore, 80% and 20% of these interactions are proposed to stabilize tertiary and quaternary structures, respectively, thus, highlighting their dominant role in protein structures (Burley and Petsko, 1985). Interestingly, π···π stacking interaction by sp^2 hybridized groups such as backbone

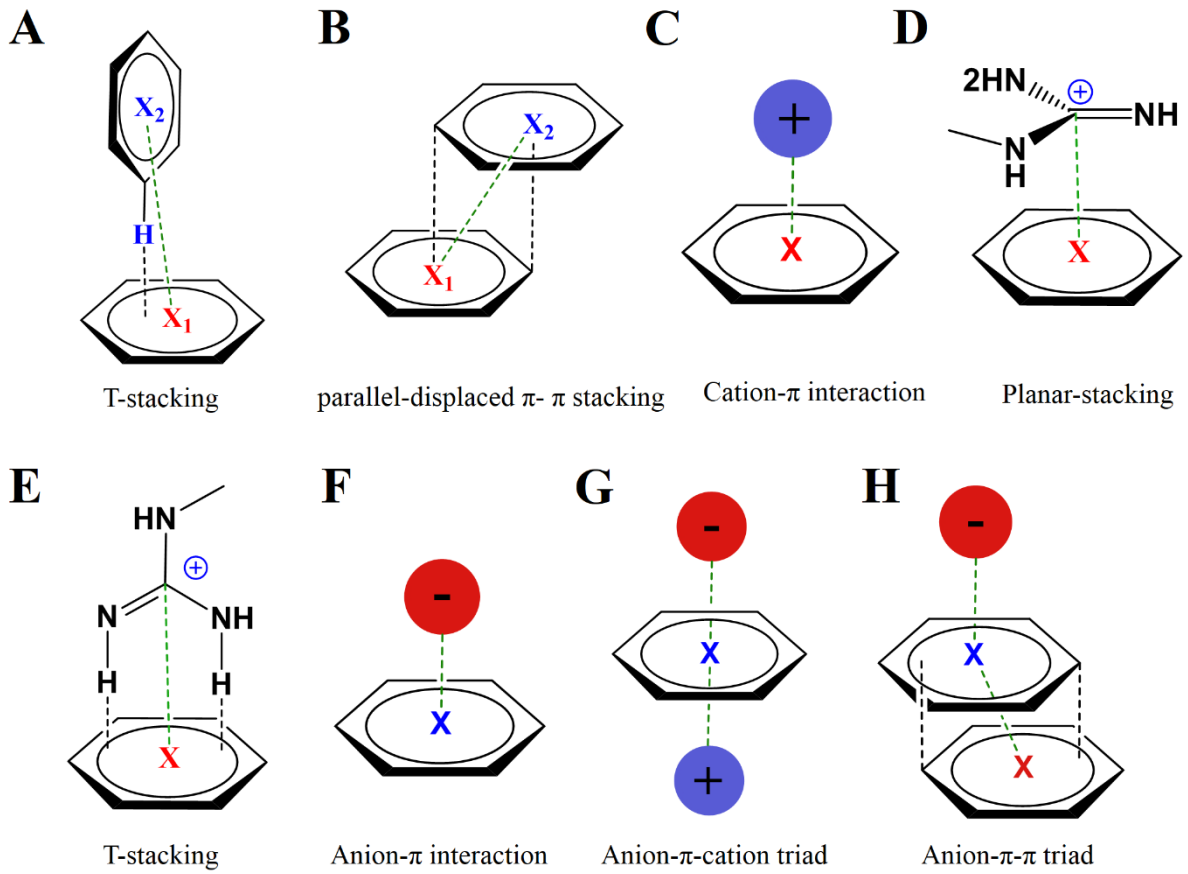


Figure 1.6 π-stacking interactions. A representation of the two geometries of π-π stacking (**A**) T-stacking and (**B**) parallel-displaced, common in biomolecules. (**C**) A general representation of cation···π interaction. Two commonly occurring geometries *viz* (**D**) planar-staking and (**E**) T-stacking of Arg···π stacking interaction in proteins. (**F**) A general representation of anion···π interaction. (**G**) Anion- π-cation and (**H**) Anion-π- π triads representing the cooperative nature of anion···π interaction.

carbonyl is also expected and drives protein phase separation, an essential feature of forming membrane-less organelles and the nuclear pore (Vernon et al., 2018). Also, stacking interactions between nucleotide bases and aromatic residues are abundant in protein-nucleic acid and protein-nucleotide complexes (Wilson et al., 2014). The following preferences, T > C > A ~ G or Phe > Tyr >Trp ~ His, were found to participate in the formation of interaction in protein-nucleic acid complexes, which might have crucial implications in base-specific nucleic acid recognition (Wilson et al., 2014).

1.4.2 Cation···π interaction

Electrostatic interaction between positively charged ions or molecules with an electron-rich or π-basic ring of the aromatic side-chain carrying negative quadrupole moment makes cation···π interaction highly directional (Figure 1.6 **C**) (Mahadevi and Sastry, 2016; Mecozzi et al., 1996).

In addition to its dominant electrostatic nature, a more significant polarization effect makes this interaction stronger than π-π stacking interaction (Figure 1.6 **A**-**B**) (Wheeler and Houk, 2009). Interestingly, the contribution from dispersion, which drives π···π association, in cation-π interaction instead remains debated (Mecozzi et al., 1996; Wheeler and Houk, 2009). Perutz and Levitt were the first to observe short amino-aromatic contacts and propose that aromatic rings act as a hydrogen bond acceptor and suggested its nucleophilic nature (Levitt and Perutz, 1988).

Upon suggestions from Perutz, a systematic analysis by Burley and Petsko revealed that positively charged residues such as Lys, Arg, and His were within 6 Å of radius from the centroid of the aromatic ring of Trp, Tyr, and Phe residues (Burley and Petsko, 1986). Using this distance cut-off criteria, Dougherty's group later found that an average of 1 out of 77 these residues in proteins could participate in favorable cation···π interaction, thus contributing to protein stability (Gallivan and Dougherty, 1999). In this case, Trp has a high preference compared to others, and one-fourth of all Trp in PDB participate in cation···π interaction; on the other hand, Arg has a higher preference than Lys (Gallivan and Dougherty, 1999). Compared to T-stacked, planar-stacking geometry of cation···π interactions made by Arg is energetically more favorable and more often seen in proteins (Flocco and Mowbray, 1994; Wheeler and Houk, 2009), possibly because of additional contribution from π···π stacking (Figure 1.6 **D-E**). Additionally, structural motifs involving cation···π interactions are highly conserved, suggesting their possible roles not only in stabilizing a particular fold but also in function (Pinheiro et al., 2017). A stacking of Arg with the backbone of α-helix (stacking with the backbone amide group) stabilizes *i*→*i*-4 and *i*-3→*i*-7 backbone H-bonds cooperatively, as evidenced by their shortened H-bond distances (Wang et al., 2018). Moreover, Arginine stacking drives tau protein aggregation to form a fibril responsible for various neurodegenerative diseases (Ferrari et al., 2020). Besides, this interaction also appears crucial in protein-nucleic acid recognition. For example, ~71% of protein-DNA and ~65% protein-RNA complexes have at least one cation···π interaction at their interface. Almost half of them involve Arg residue in protein-DNA complexes (Wintjens et al., 2000; Zhang et al., 2014). For both cases, a stronger preference of Arg towards Guanine base suggests its possible role in target base specificity (Wintjens et al., 2000; Zhang et al., 2014).

In the case of protein complexes, at least one cation-π interaction is present in half of the protein-protein and one-third of homodimer complexes and suggested to contribute up to -3 kcal.mol^{-1} for their overall stability (Crowley and Golovin, 2005). Similarly, studies on drug-receptor and protein-protein interaction showed that cation-π interaction could contribute to binding by -2 to -5 kcal.mol^{-1} (Dougherty, 2013). Also, there are many instances for cation···π

interactions in membrane proteins such as ion channels and G protein-coupled receptors, which are crucial for their biochemical functions. For example, cation···π interaction drives the binding of nicotine to the acetylcholine receptor (Dougherty, 2013; Infield et al., 2021).

Re-engineering natural enzymes is one of the aims of applied biocatalysis and synthetic biology (Li et al., 2018; Narayan and Sherman, 2013). The introduction of cation···π interaction in endo-polygalacturonases increases the thermotolerance and catalytic efficiency, suggesting its scope in these fields (Tu et al., 2016).

1.4.3 Anion···π interaction

On the contrary to cation···π, anion···π interaction forms between electron-rich ions/molecules (anions) and electron-deficient or π-acidic rings of an aromatic group carrying positive quadrupole moment (Figure 1.6 **F**) (Schottel et al., 2008). In addition to its electrostatic nature, some contribution from the anion-induced polarization adds to its attractive nature (Schottel et al., 2008). Although, this interaction is widely found in small molecules and extensively used in crystal engineering, designing, supramolecular chemistry, and catalysis (Ahmad Rather and Ali, 2021). Its functional role in biomolecules was identified in 2011 (Estarellas et al., 2011a). The partial anionic nature (because of lone pairs of O) of catalytic water that interacts with the π-basic ring of the substrate of the urate oxidase was thought crucial for its functioning. The crystal structure of cyanide or chloride bound to urate oxidase and their ability to inhibit the activity of this enzyme confirmed the nature of the interaction (Estarellas et al., 2011a). Also, using directed evolution, streptavidin variant was identified, which was able to carry out enantioselective addition by stabilizing anionic transition state *via* anion···π interaction (Cotelle et al., 2016). The ability of nitrate ion to inhibit its activity confirmed the nature of the interaction, suggesting the potential use of interaction in asymmetric synthesis and synthetic biology (Cotelle et al., 2016; Estarellas et al., 2011b; Zhao et al., 2018).

In PDB, contacts of chloride and phosphate ions with aromatic residues are abundant, mostly within 3.5 to 4.5 Å (Robertazzi et al., 2011). Also, ~61% of PDB structures are found to have at least one anion-π interaction, primarily by Asp and Glu residues (Lucas et al., 2016). Interestingly, His has a high preference compared to other aromatic residues. Although His residue, in proteins, is often neutral, these contacts were presumed to be with protonated His, which is electron-deficient but retaining its aromatic nature (Robertazzi et al., 2011).

Unlike His, how does the electron-rich π-ring of other aromatic residues, which are always neutral, interact with anions? This could be possible by cooperatively incorporating another

stacking interaction forming anion-π-π or anion-π-cation triads (Figure 1.6 **G-H**) (Lucas et al., 2016). Such a triad formation introduces a positive quadrupole moment for an aromatic ring interacting with anions, resulting in favorable interaction in proteins or their complexes. Such interplay among the stacking interactions is well established in the small molecules (Mahadevi and Sastry, 2016). These triads appear essential for the folding or specificities in biomolecular structures because of the following preference for interaction **1.** In proteins: Glu-Tyr-Phe, **2.** At the protein-protein interface: Asp-Trp-Phe, **3.** Inter-chain: Asp/Glu-Phe-Arg, **4.** In RNA: Glu-Adenine-Arg, and **5.** In DNA: Asp-His-Thymine or Asp-Cytosine-Adenine (Lucas et al., 2016). A significant experimentally observed strength (-1.3 kcal.mol^{-1} for Asp and Phe interaction), an ability to stabilize secondary structures (Smith et al., 2017), and the cooperative nature of an anion···π interaction could help to develop the strategy in protein designing.

1.4.4 Salt bridge

Salt bridges are electrostatic interactions that form between oppositely charged residues such as Asp or Glu and Lys, Arg, or protonated His in proteins (Figure 1.7 **A-C**). The strength of these interactions is strongly dependent on the dielectric constant of their environment (Bosshard et al., 2004). It usually ranges from -2 to -5 kcal.mol^{-1} for buried salt bridges and ~ -0.5 kcal.mol^{-1} for those which are exposed to water, presumably because of the charge dispersing in an aqueous medium (Anderson et al., 1990; Fersht, 1972; Serrano et al., 1990). PDB survey indicates that the favorable distance for the salt bridges is ~ 4 Å. Most of the proteins have, on average, four salt bridges, and ~80% of the proteins have at least one (Barlow and Thornton, 1983; Sarakatsannis and Duan, 2005). Salt bridges that form between the residues that are separated by three intervening residues are abundant in α-helices, and possibly because of the 3.6 residues per turn of α-helix might facilitate the formation of the interaction (Figure 1.7 **D**) (Barlow and Thornton, 1983; Sarakatsannis and Duan, 2005). In the peptide model, such a salt bridge is shown to increase the stability of α-helix (Marqusee and Sauer, 1994).

Although salt bridges are diverse in proteins, their role in the overall stability and functioning of the proteins remains debatable (Anderson et al., 1990; Hendsch and Tidor, 1994; Marqusee and Baldwin, 1987; Strop and Mayo, 2000; Tissot et al., 1996). For instance, replacing a buried salt bridge triad of two Arg and one Glu with hydrophobic residues creates a more stable variant of Arc repressor, suggesting its poor role in stabilizing Arc repressor (Waldburger et al., 1995). On the contrary, a group of four conserved salt bridges identified in Cytochrome P450cam could provide stability up to -6 kcal.mol^{-1} and was proposed essential to keep the heme group intact and binding of the substrate (Lounnas and Wade, 1997).

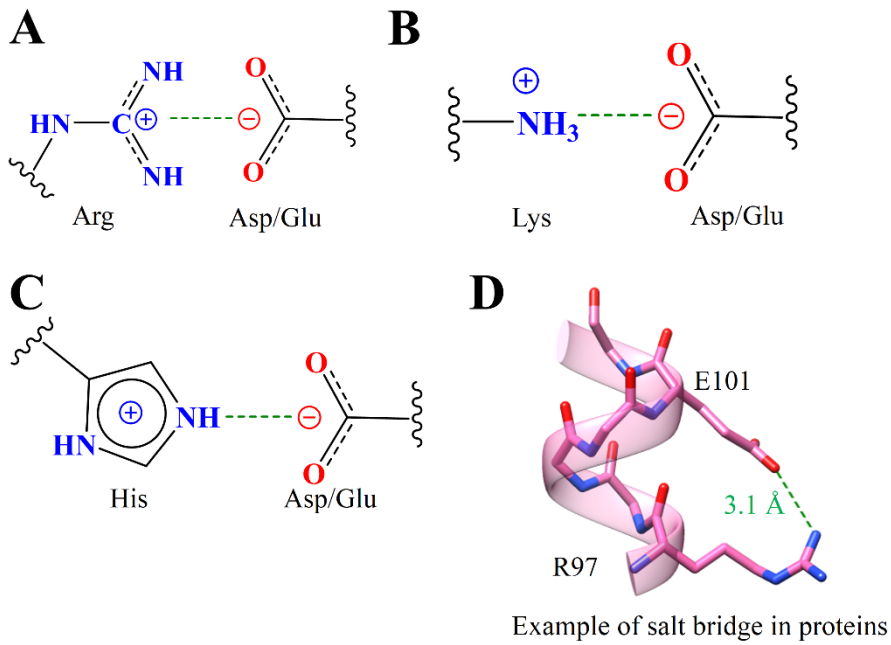


Figure 1.7 Salt bridge. A general representation of salt bridge that occurs between the oppositely charged residues such (**A**) Arg-Glu/Asp, (**B**) Lys-Glu/Asp, and (**C**) protonated His-Glu/Asp. (**D**) A representative example of a salt bridge often occurs in α-helices and forms between the residues separated by three intervening residues.

These discrepancies might result from the perturbation of delicate balance between favorable charge-charge/charged-dipole interactions and unfavorable desolvation of residues that forms a salt bridge, which is either buried or exposed to water (Bosshard et al., 2004). In PDB, almost all salt bridges have Solvent Accessible Surface Area (SASA) > 20% (peaked at ~32%). This suggests their partially exposed nature, which might not need to overcome the entropic cost, otherwise needed at the buried site (Sarakatsannis and Duan, 2005). This could be the reason for their unique location of appearance in protein's architecture (Wimley et al., 1996). The accuracy of force fields to model salt bridges increases by incorporating solvent polarization, which otherwise overestimates their strength (Debiec et al., 2014).

Salt bridges increase the thermostability of proteins (Kumar and Nussinov, 2001). An example is the enhanced thermostability by interchain salt bridges in the triple helical structure of the collagen (Gurry et al., 2010). More salt bridges are noted in thermophilic or halophilic proteins than in mesophilics that helps to introduce a rigidity that is critical for maintaining their structures, thus, highlighting another benefit (Kumar et al., 2000; Nayek et al., 2014). Moreover, salt bridge interactions are often evolutionarily conserved (Ban et al., 2019), suggesting their possible functional relevance as well (Mhaindarkar et al., 2018). In summary, it could be a general strategy

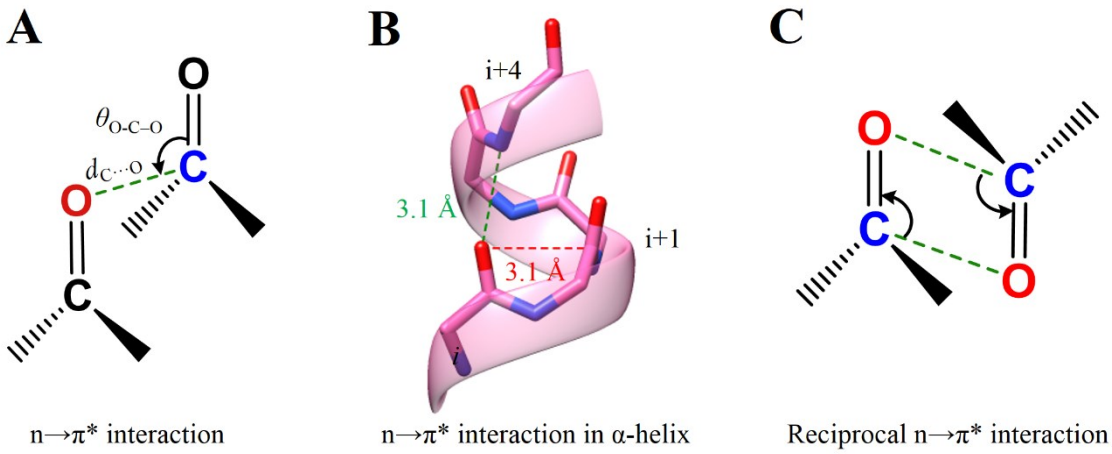


Figure 1.8 n→π* interaction. (**A**) A general representation of n→π* interaction formed between the two carbonyl groups with geometrical parameters that characterize the interaction. (**B**) n→π* interaction formed between the i^{th} and $i+1^{th}$ residues in α-helix (red). (**C**) A representation of the reciprocal n→π* interaction.

to design enzymes with higher thermostability for medicinal or industrial applications by employing geometric preferences of salt bridges (Donald et al., 2011; Lee et al., 2014).

1.4.5 n→π* interaction

A carbonyl group (C=O) forms a permanent dipole due to the polarization of electron density towards O because it being more electronegative than C. As a result, O is partially negative, and C is partially positively charged. Thus, it can participate in C=O···C=O interactions, which are abundant in proteins (Figure 1.8 **A**) (Bartlett et al., 2010). Earlier, this interaction was considered a dipolar interaction of electrostatic origin (Fischer et al., 2008; Worley et al., 2012). However, Raines and others confirmed its charge transfer nature that arises from the delocalization of electron density of the lone pair of O to antibonding orbital (π*) of the carbonyl group. The "n→π* interaction" is considered a general representation of this interaction (Choudhary et al., 2009; Kamer et al., 2013; Sahariah and Sarma, 2019), and the estimated interaction energy ranges from -0.3 to -0.7 kcal.mol^{-1} (Newberry and Raines, 2019). One-third of all residues in proteins could participate in n→π* interaction because of the abundance of the carbonyl group. Most of them are from α-helices (> 70% of all residue in α-helices) and are argued to stabilize this secondary structure in proteins (Figure 1.8 **B**) (Bartlett et al., 2010). The favorable distance for the interaction is ~3.0 Å with θ_{O-C-O} ~ 102.0° (Figure 1.8 **B**) (Bartlett et al., 2010). Also, the two n→π* interactions formed back and forth within the two carbonyl groups (Figure 1.8 **C**), referred to as reciprocal n→π* interactions, are often part of polyproline π-helices and can affect protein folding (Rahim et al., 2017).

1.5 Chalcogen bonding interaction

Group VI elements, mainly divalent S, Se, or Te, can interact with various nucleophiles to form another set of non-covalent interactions called chalcogen bond (Ch-bond) (Figure 1.9 **A**) (Bauzá et al., 2015; Murray et al., 2007a, 2007b; Politzer et al., 2017; Politzer and Murray, 2013). An extensive work by Politzer and his group on understanding the origin of the interaction revealed its electrostatic nature with a small contribution from dispersion and charge transfer components (Murray et al., 2007a, 2007b; Politzer et al., 2017; Politzer and Murray, 2013). Recently IUPAC defined Ch-bond as "*net attractive interaction between an electrophilic region associated with a chalcogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity,"* also highlighting its electrostatic nature (Aakeroy et al., 2019). Here, an electrophilic region associated with chalcogen atoms is called σ-hole that can interact with the nucleophiles to form a Ch-bond (Politzer et al., 2017) (Figure 1.9 **B**). On the contrary, an NMR-based experiment suggests that charge transfer is the primary stabilizing force for the chalcogen bonds (Pascoe et al., 2017). This charge transfer arises from electron donation from nucleophiles to the antibonding orbital (σ*) of the chalcogen atom (n→ σ*) (Pascoe et al., 2017). Although this observation, in addition to some theoretical analysis, initiated a considerable debate on the origin of the Ch-bond, its electrostatic model is still widely considered (Bunchuay et al., 2020; Docker et al., 2021; Rozhkov et al., 2021; Scilabra et al., 2019; Tarannam et al., 2021).

The divalent S is abundantly present in proteins in the form of methionine and cysteine, and Ch-bonds formed by them are important because of their unique role in biomolecular structure and function. CSD and PDB data analyses provided many instances for this interaction in small and protein molecules. For example, Rosenfield et al. observed the presence of short contacts of divalent S with various nucleophiles whose approach was distinct from the approach of electrophiles (Rosenfield et al., 1977) (Figure 1.9 **C**). The contribution of frontier orbital interaction was proposed to be responsible for such a directional behavior of the interactions. Later, a similar analysis by Pal and Chakrabarti for proteins revealed that 22% of S from Met residues had short contact with carbonyl O where $d_{S\cdots O}$ was < 5 Å (Pal and Chakrabarti, 2001) (Figure 1.9 **D**). Most of these contacts had θ being greater than 40˚, and Φ between 30˚ to 60˚ (Figure 1.9 **D**). A more comprehensive analysis by Iwoaka et al. revealed that the divalent S prefer to approach the π over the lone pair region of carbonyl O, confirming the directional behavior of the interaction in proteins (Iwaoka et al., 2002) (Figure 1.9 **E**). Also, 1.3 and 8.9% of Met and Cys in proteins could participate in Ch-bond formation, respectively, suggesting their possible role in stabilizing proteins (Iwaoka et al., 2002). In addition to this, PDB data also suggest that Ch-bonds

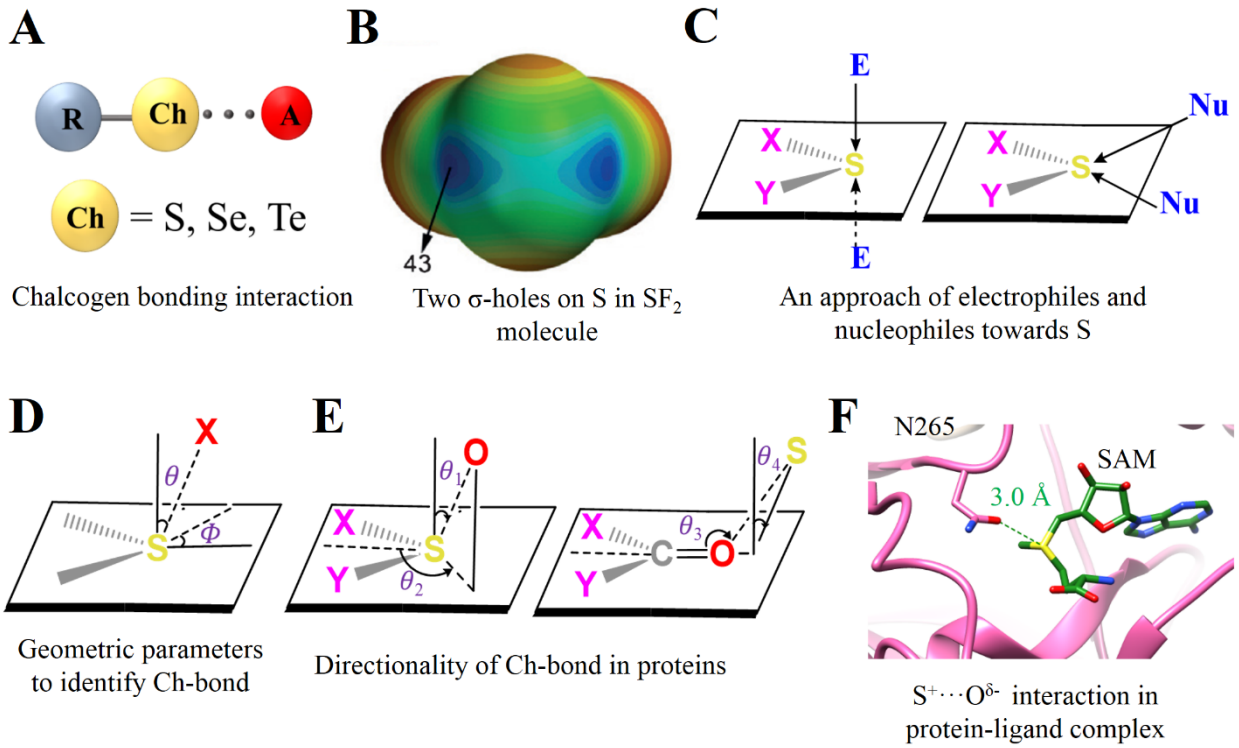


Figure 1.9 Chalcogen bonding interaction. (**A**) A general representation of the Ch-bond formed between the chalcogen atom such as S, Se, or Te with electron-rich molecules. Here, the chalcogen atom represents the Ch-bond donor, and A represents the Ch-bond acceptor (Aakeroy et al., 2019). (**B**) Molecular electrostatic potential surface for SF_2 molecule computed at B3LYP/6-31+G* level. The two blue regions (the positive electrostatic region that can attract nucleophiles) on S represent the σ-holes, and its strength ($V_{S,max}$) is shown in kcal.mol^{-1} (adapted from Bauzá et al., 2015). (**C**) A representation shows a distinct approach of electrophiles and nucleophiles towards divalent S. (**D**) The geometric parameters used to investigate Ch-bond in proteins. The most common value of $d_{S\cdots O}$ is 3.6 Å, and θ and φ ranges within > 50° and 30-60°, respectively (adapted from Pal and Chakrabarti, 2001). (**E**) Another set of stereochemical parameters used to identify S···O bond in proteins. The favorable values of θ_1, θ_2, θ_3, and θ_4 are ~ 90°, ~130°, ~90° and ~0° or ~180°, respectively (adapted from Iwaoka et al., 2002). (**F**) An electrostatic $S^{+}\cdots O^{\delta-}$ interaction at the active site of histone-lysine N-methyltransferase SET7/9 (PDB ID: 4J83).

are common in protein-ligand complexes that mainly involve aromatic S and could be used as a part of a strategy for therapeutics design (Kristian et al., 2018).

Furthermore, Fick et al. demonstrated that the sulfonium ion of AdoMet forms a stronger S···O bond because of the positively charged trivalent S that allows SET7/9 protein to discriminate between the product and reactant (Fick et al., 2016) (Figure 1.9 **F**). Note that, though such a trivalent S could form electrostatic interaction because of its cationic nature, no evidence is available yet for a similar role of σ-hole mediated Ch-bond in the functioning of an enzyme.

Furthermore, Ch-bonds among the phospholipase A2 domains, ribonuclease A, lysosome, and insulin families are conserved (Iwaoka and Isozumi, 2012, 2006). These Ch-bonds could contribute to their enzymatic activities, conformational rearrangement at the active site, thermodynamic stability, or folding similar to other interactions, which primarily remains scarce. However, the favorable interaction energy (up to -3 kcal.mol^{-1}), highly directional properties and independence on the solvent polarity are unique to Ch-bonds that might be helpful to introduce specificity similar to other polar interactions (Pascoe et al., 2017; Iwaoka et al., 2002).

1.6 van der Waals dispersion interaction

The attractive or negative term in the Lennard-Jones equation (equation 1) that arises from the quantum mechanical electron correlation effect represents non–directional van der Waals (vdW) dispersion interaction (Wagner and Schreiner, 2015). This interaction is sometimes also referred to as London dispersion or van der Waals interactions due to the pioneering work of J. D. van der Waals and F. London.

$$E_{LJ} = \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \qquad (1)$$

The original definition provided by London states that *"an interaction characterized by a short-period mutual perturbation of the inner electron motion of molecules whose magnitude is the major attractive contribution in the simplest non-polar and also weakly polar molecules"* (Wagner and Schreiner, 2015). The strength of interacting instantaneous fluctuating dipoles between neighboring atoms depends on their polarizability. The distance separating these dipoles by a relation $1/r^6$ is fundamental of vdW dispersion interaction (equation 1) (Hwang et al., 2016; Reilly and Tkatchenko, 2015; Wagner and Schreiner, 2015). The interactions among non-polar or weakly polar molecules primarily account for vdW dispersion interactions since interactions between polar molecules are dominated by electrostatic (Reilly and Tkatchenko, 2015). The gas-phase calculated strength of the methane dimer ($CH_4 \cdots CH_4$) is around -0.4 kcal.mol^{-1} and is often considered as a benchmark for the analysis involving vdW dispersion interactions (Figure 1.10 **A**) (Echeverría et al., 2011). The direct correlation between the number of C atoms n-alkanes and interaction energy indicates that vdW dispersion interaction is additive and non-saturating in nature. Thus, it is responsible for a linear correlation between the sizes of the hydrocarbons with their increasing melting point (Echeverría et al., 2011; Parks, 1954; Reilly and Tkatchenko, 2015). The short C-H$\cdots$H-C contacts stabilize these n-alkane dimers. The ability of these contacts to stabilize dodecahedrane dimer by ~ -3 kcal.mol^{-1} is clear evidence for the more potent effect in branch chain alkanes and polyhedranes (Echeverría et al., 2011). The underlying mechanism for

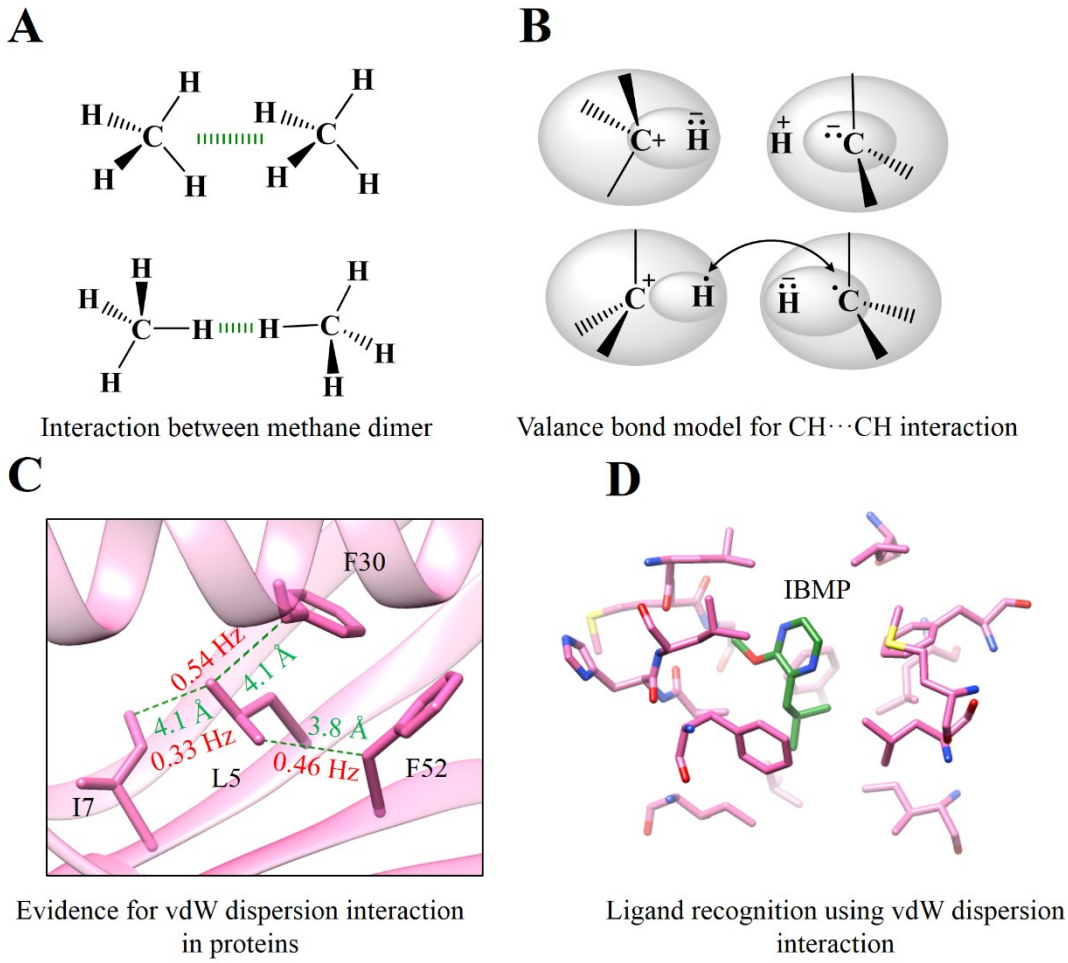


Figure 1.10 van der Waals dispersion interaction. (**A**) The two standard geometries of methane dimer forming the vdW dispersion interaction. (**B**) A valance bond model showing the domination of CH···CH interaction by charge alteration (top panel) for small alkanes and recoupling of bonding electrons to form H···H, C···C, and C···H bonds for large alkanes (bottom panel) (adapted from Danovich et al., 2013). (**C**) The observation of $^{vdW}J_{CC}$ coupling in NMR based experiment between non-polar residues in protein GB3 (PDB ID: 1IGD) because of the CH···CH vdW dispersion interaction (the values are adapted from Li et al., 2018). (**D**) The enthalpy-driven bonding of 2-methoxy- 3-isobutylpyrazine (IBMP) to the variant of MUP (PDB ID: 1YP6). The IBMP does not form any polar interaction with surrounding protein residues.

C-H···H-C contacts depends on the size of the alkanes revealed by the valance bond model (Figure 1.10 **B**) (Danovich et al., 2013). For smaller alkanes, charge alteration C^{+}-H^{-}···H^{+}-C^{-} dominates the interaction (Figure 1.10 **B**, top panel). Whereas, for large alkanes, the bonding electrons of CH groups reorganize themselves in a way that recouples these electrons to form H···H, C···C, and C···H bonds (Figure 1.10 **B**, bottom panel) (Danovich et al., 2013).

A saturated C in the form of aliphatic non-polar residues such as Ala, Val, Leu, Ile, and Met are common in proteins. An ultra-high-resolution PDB database analysis revealed ~4.5 pairs

of contacts per residue with a favorable distance of ~4.1 and ~5.4 Å for the two most preferred C···C interactions (Figure 1.10 **A**) (Li et al., 2018). An observation of through space van der Waals coupling for different C···C interaction ($^{vdW}J_{C\cdots C}$) up to a value between ~0.2 to ~0.5 Hz confirms the presence of vdW dispersion interaction in proteins (Figure 1.10 **C**) (Li et al., 2018). The stabilization of proteins by $–CH_2–$ up to -1.1 kcal.mol^{-1} could be a result of favorable desolvation entropy of non-polar groups (hydrophobic interaction) followed by an enthalpic contribution from their packing/optimization at the core of proteins by vdW dispersion interaction (Nick Pace et al., 2014). It is challenging to segregate the vdW dispersion interaction from hydrophobic interaction (Nick Pace et al., 2014). However, a similar energetic contribution for packing non-polar residues in membrane protein compared to the soluble proteins hints at the possible contribution of vdW dispersion interaction in the protein's stability because of the lack of hydrophobic interaction in membrane proteins (Joh et al., 2009). This observation was evidenced by a few other studies (Ratnaparkhi and Varadarajan, 2000, Holder et al., 2001), particularly by Holder et al., who found that the average stability of the buried $–CH_3$ group is greater than the hydrophobic transfer energy from octanol to water (Holder et al., 2001). In addition, the interaction between electron-deficient carbonyl C and electron-rich side-chain sp^3 C (C···C) plays its stabilizing role in the self-assembly of antiparallel β-sheets of a short peptide (Misra et al., 2021). This short peptide has a therapeutic application suggesting the role of vdW dispersion in designing such molecules.

Besides this, vdW dispersion interaction is also crucial for protein-ligand interaction. The enthalpy-dominated binding of Tween 40 to human serum albumin or 2-methoxy-3-isobutylpyrazine (IBMP) to the mouse major urinary protein(MUP) suggests a contribution from vdW dispersion interaction instead of the expected hydrophobic effect because of the non-polar nature of ligands (Figure 1.10 **D**) (Barratt et al., 2005; Ghosh et al., 2015). The solvent isotopic substitution ITC measurements and all-atom molecular dynamics simulations with the inclusion of water confirmed that this enthalpic contribution does not come from ligand or protein pocket desolvation (Barratt et al., 2005). Thus, various force fields have added Lennard-Jone's potential to account for vdW dispersion interaction in biomolecular simulations or docking studies. Removal of attractive term or vdW dispersion (equation 1) component from a force field results in the double-helical structure of DNA oligomer collapsing when simulated for 100 ns. This observation provides strong computational evidence for the role of this interaction in the stability and specificity of the biomolecular structures (Kolář et al., 2011).

1.7 Summary and Objectives

One of the unaccomplished aims of protein science is to accurately predict protein structures or their biochemical properties from sequence information (Gligorijević et al., 2021; Jumper et al., 2021; Kulmanov and Hoehndorf, 2020; Lee et al., 2017; Marks et al., 2012; Ovchinnikov et al., 2017; Whisstock and Lesk, 2003). One of the ways to achieve this goal is to scrutinize the forces that stabilize or destabilize the proteins or their complexes (Newberry and Raines, 2019). The precise knowledge of these forces and their incorporation in the force fields could help accurately model proteins and understand processes involving protein folding or protein-ligand recognition. One of the most significant destabilizing forces that make these processes unfavorable is the loss of conformational entropy that could contribute to 200 kcal.mol^{-1} per 100 residues for folding the peptide chain (Pace, 2009).

On the other hand, the conventional H-bond and hydrophobic interactions are the most acknowledged among the different stabilizing forces. The favorable contribution of H-bond and hydrophobic interaction is about -100 kcal.mol^{-1} and -110 kcal.mol^{-1} per 100 residues, respectively (Pace, 2009). This compensates for the unfavorable conformational entropy towards folding the peptide chain and making the folding process favorable as overall free energy change (ΔG) becomes up to -10 kcal.mol^{-1} per 100 residues (Pace, 2009). In addition, these interactions are also apt for other processes that involve biomolecular recognition.

Although the contribution of these two primary interactions in understanding protein folding and its stability is recognizable, however, their insufficiency in explaining the discrepancies in computational and experimental results makes the inclusion of the various secondary weak interactions necessary. Although these weak individual interactions appear diminutive in terms of their interaction energy, their cumulative effect could strongly alter the stability and biochemical properties of proteins because of their frequent occurrence. For instance, the enthalpic contribution of n→π*, π-stacking, and C-H···O interaction is around -8.5, -5, and -5 kcal.mol^{-1} per 100 residues, respectively (Newberry and Raines, 2019). The review summarizing various secondary interactions by Newberry et al. concludes that the cumulative enthalpic contribution of these interactions is around -27 kcal.mol^{-1} per 100 residues, almost 25% in comparison to conventional H-bonds or hydrophobic interactions, which strongly suggests their incorporation in biomolecular force fields is necessary to enhance computer-aided prediction and modelling (Newberry and Raines, 2019). Understanding these weak/secondary interactions is in progress. Estimating their critical properties precisely, such as experimental stabilization, environment dependency, and interplay among them, needs much attention.

As part of my thesis, I have investigated two such non-covalent interactions in the context of biomolecules *viz* S-mediated H- and Ch-bond and vdW dispersion interaction with the following two objectives.

Objective 1: Understanding the roles of and interplay between sulfur-mediated chalcogen and hydrogen bond in proteins.

Objective 2: The role of van der Waals dispersion interaction in sequence-specific protein-DNA recognition.

1.8 References

Aakeroy, C.B., Bryce, D.L., Desiraju, G.R., Frontera, A., Legon, A.C., Nicotra, F., Rissanen, K., Scheiner, S., Terraneo, G., Metrangolo, P., Resnati, G., 2019. Definition of the chalcogen bond (IUPAC Recommendations 2019). Pure Appl. Chem. 91(11), 1889-1892.

Ahmad Rather, I., Ali, R., 2021. Anion-π Catalysis: A Novel Supramolecular Approach for Chemical and Biological Transformations. Current Topics in Chirality - From Chemistry to Biology, IntechOpen, London.

Anderson, D.E., Becktel, W.J., Dahlquist, F.W., 1990. pH-Induced Denaturation of Proteins: A Single Salt Bridge Contributes 3-5 kcal/mol to the Free Energy of Folding of T4 Lysozyme. Biochemistry. 29(9), 2403–2408.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science. 181(4096), 223-230.

Arunan, E., Desiraju, G.R., Klein, R.A., Sadlej, J., Scheiner, S., Alkorta, I., Clary, D.C., Crabtree, R.H., Dannenberg, J.J., Hobza, P., Kjaergaard, H.G., Legon, A.C., Mennucci, B., Nesbitt, D.J., 2011. Definition of the hydrogen bond (IUPAC Recommendations 2011). Pure Appl. Chem. 83(8), 1637-1641.

Asensio, J.L., Ardá, A., Cañada, F.J., Jiménez-Barbero, J., 2013. Carbohydrate-aromatic interactions. Acc. Chem. Res. 46(4), 946–954.

Ash, E.L., Sudmeier, J.L., Day, R.M., Vincent, M., Torchilin, E. V., Haddad, K.C., Bradshaw, E.M., Sanford, D.G., Bachovchin, W.W., 2000. Unusual ^{1}H NMR chemical shifts support (His) $C^{\varepsilon 1}$-H···O=C H-bond: Proposal for reaction-driven ring flip mechanism in serine protease catalysis. Proc. Natl. Acad. Sci. U. S. A. 97(19), 10371-10376.

Aurora, R., Rose, G.D., 1998. Helix capping. Protein Sci. 7(1), 21-38.

Baker, E.G., Williams, C., Hudson, K.L., Bartlett, G.J., Heal, J.W., Goff, K.L.P., Sessions, R.B., Crump, M.P., Woolfson, D.N., 2017. Engineering protein stability with atomic precision in a monomeric miniprotein. Nat. Chem. Biol. 13, 764–770.

Ban, X., Lahiri, P., Dhoble, A.S., Li, D., Gu, Z., Li, C., Cheng, L., Hong, Y., Li, Z., Kaustubh, B., 2019. Evolutionary Stability of Salt Bridges Hints Its Contribution to Stability of Proteins. Comput. Struct. Biotechnol. J. 17, 895-903.

Barlow, D.J., Thornton, J.M., 1983. Ion-pairs in proteins. J. Mol. Biol. 168(4), 867-885.

Barratt, E., Bingham, R.J., Warner, D.J., Laughton, C.A., Phillips, S.E.V., Homans, S.W., 2005. Van der Waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. J. Am. Chem. Soc. 127(33), 11827–11834.

Bartlett, G.J., Choudhary, A., Raines, R.T., Woolfson, D.N., 2010. N→π* interactions in proteins. Nat. Chem. Biol. 6, 615–620.

Bauzá, A., Mooibroek, T.J., Frontera, A., 2015. The Bright Future of Unconventional σ/π-Hole Interactions. ChemPhysChem. 16(12), 2496-2517.

Biswal, H.S., Gloaguen, E., Loquais, Y., Tardivel, B., Mons, M., 2012. Strength of NH···S hydrogen bonds in methionine residues revealed by gas-phase IR/UV spectroscopy. J. Phys. Chem. Lett. 3(6), 755–759.

Biswas, H.S., Wategaonkar, S., 2009. Nature of the N-H···S hydrogen bond. J. Phys. Chem. A. 113(46), 12763–12773.

Bitencourt-Ferreira, G., Veit-Acosta, M., de Azevedo, W.F., 2019. Hydrogen bonds in protein-ligand complexes. Methods in Molecular Biology. 2053, Humana, New York, NY.

Bolen, D.W., Rose, G.D., 2008. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu. Rev. Biochem. 77(1), 339-362.

Bosshard, H.R., Marti, D.N., Jelesarov, I., 2004. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. J. Mol. Recognit. 17(1) 1-16.

Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J., Hilgenfeld, R., 2001. C-H···π-interactions in proteins. J. Mol. Biol. 307, 357–377.

Bulusu, G., Desiraju, G.R., 2020. Strong and Weak Hydrogen Bonds in Protein–Ligand Recognition. J. Indian Inst. Sci. 100, 31–41.

Bunchuay, T., Docker, A., Eiamprasert, U., Surawatanawong, P., Brown, A., Beer, P.D., 2020. Chalcogen Bond Mediated Enhancement of Cooperative Ion-Pair Recognition. Angew. Chemie - Int. Ed. 59(29), 12007-12012.

Burley, S.K., Petsko, G.A., 1986. Amino-aromatic interactions in proteins. FEBS Lett. 203(2), 139-143.

Burley, S.K., Petsko, G.A., 1985. Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science. 229(4708), 23-28.

Cao, Z., Hutchison, J.M., Sanders, C.R., Bowie, J.U., 2017. Backbone Hydrogen Bond Strengths Can Vary Widely in Transmembrane Helices. J. Am. Chem. Soc. 139(31), 10742–10749.

Carter-Fenk, K., Herbert, J.M., 2020. Reinterpreting π-stacking. Phys. Chem. Chem. Phys. 22, 24870-24886

Chan, H.S., Dill, K.A., 1990. Origins of structure in globular proteins. Proc. Natl. Acad. Sci. U. S. A. 87(16) 6388-6392.

Chandler, D., 2005. Interfaces and the driving force of hydrophobic assembly. Nature. 437, 640–647.

Chelli, R., Gervasio, F.L., Procacci, P., Schettino, V., 2002. Stacking and T-shape competition in aromatic-aromatic amino acid interactions. J. Am. Chem. Soc. 124(21), 6133–6143.

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S.M., Savidge, T.C., 2016. Regulation of protein-ligand binding affinity by hydrogen bond pairing. Sci. Adv. 2, e1501240.

Cheng, X., Shkel, I.A., O'Connor, K., Thomas Record, M., 2020. Experimentally determined strengths of favorable and unfavorable interactions of amide atoms involved in protein self-assembly in water. Proc. Natl. Acad. Sci. U. S. A. 117, 27339–27345.

Choudhary, A., Gandla, D., Krow, G.R., Raines, R.T., 2009. Nature of amide carbonyl-carbonyl interactions in proteins. J. Am. Chem. Soc. 131(21), 7244–7246.

Cleland, W.W., Kreevoy, M.M., 1994. Low-barrier hydrogen bonds and enzymic catalysis. Science. 264(5167), 1887-1890.

Copley, M.J., Marvel, C.S., Ginsberg, E., 1939. Hydrogen Bonding by S-H. VII. Aryl Mercaptans. J. Am. Chem. 61(11), 3161–3162.

Cotelle, Y., Lebrun, V., Sakai, N., Ward, T.R., Matile, S., 2016. Anion-π enzymes. ACS Cent. Sci. 2(6), 388–393.

Crowley, P.B., Golovin, A., 2005. Cation-π interactions in protein-protein interfaces. Proteins Struct. Funct. Genet. 59(2), 231-239.

Dai, S., Funk, L.M., von Pappenheim, F.R., Sautner, V., Paulikat, M., Schröder, B., Uranga, J., Mata, R.A., Tittmann, K., 2019. Low-barrier hydrogen bonds in enzyme cooperativity. Nature. 573, 609–613.

Danovich, D., Shaik, S., Neese, F., Echeverría, J., Aullón, G., Alvarez, S., 2013. Understanding the nature of the CH···HC interactions in alkanes. J. Chem. Theory Comput. 9(4), 1977–1991.

Debiec, K.T., Gronenborn, A.M., Chong, L.T., 2014. Evaluating the strength of salt bridges: A comparison of current biomolecular force fields. J. Phys. Chem. B. 118(24), 6561–6569.

Denke, E., Merbitz-Zahradnik, T., Hatzfeld, O.M., Snyder, C.H., Link, T.A., Trumpower, B.L., 1998. Alteration of the midpoint potential and catalytic activity of the Rieske iron-sulfur protein by changes of amino acids forming hydrogen bonds to the iron-sulfur cluster. J. Biol. Chem. 273(150), 9085-9093.

Derewenda, Z.S., Derewenda, U., Kobos, P.M., 1994. (His)C^{ε}-H···O=C < hydrogen bond in the active sites of serine hydrolases. J. Mol. Biol. 241(1), 83-93.

Derewenda, Z.S., Lee, L., Derewenda, U., 1995. The occurrence of C - H ··· O hydrogen bonds in proteins. J. Mol. Biol. 252(2), 248–262.

Desiraju, G.R., Steiner, T., 1999. The Weak Hydrogen Bond In Structural Chemistry and Biology

(International Union of Crystallography, Monographs on Crystallography, 9), Oxford University Press, Oxford.

Docker, A., Guthrie, C.H., Kuhn, H., Beer, P.D., 2021. Modulating Chalcogen Bonding and Halogen Bonding Sigma-Hole Donor Atom Potency and Selectivity for Halide Anion Recognition. Angew. Chemie - Int. Ed. 60(40), 21973-21978.

Donald, J.E., Kulp, D.W., DeGrado, W.F., 2011. Salt bridges: Geometrically specific, designable interactions. Proteins Struct. Funct. Bioinforma. 79(3), 898-915.

Dougherty, D.A., 2013. The cation-π interaction. Acc. Chem. Res. 46(4), 885–893.

Dyson, H.J., Wright, P.E., Scheraga, H.A., 2006. The role of hydrophobic interactions in initiation and propagation of protein folding. Proc. Natl. Acad. Sci. 103(35), 13057-13061.

Echeverría, J., Aullón, G., Danovich, D., Shaik, S., Alvarez, S., 2011. Dihydrogen contacts in alkanes are subtle but not faint. Nat. Chem. 3, 323–330.

Estarellas, C., Frontera, A., Quiñonero, D., Deyà, P.M., 2011a. Relevant Anion-π Interactions in Biological Systems: The Case of Urate Oxidase. Angew. Chemie-Int. Ed. 123(2), 435-438.

Estarellas, C., Frontera, A., Quiñonero, D., Deyà, P.M., 2011b. Anion-π interactions in flavoproteins. Chem. - An Asian J. 6(9), 2316-2318.

Feng, B., Sosa, R.P., Mårtensson, A.K.F., Jiang, K., Tong, A., Dorfman, K.D., Takahashi, M., Lincoln, P., Bustamante, C.J., Westerlund, F., Nordén, B., 2019. Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. Proc. Natl. Acad. Sci. U. S. A. 116(35), 17169-17174.

Ferrari, L., Stucchi, R., Konstantoulea, K., van de Kamp, G., Kos, R., Geerts, W.J.C., van Bezouwen, L.S., Förster, F.G., Altelaar, M., Hoogenraad, C.C., Rüdiger, S.G.D., 2020. Arginine π-stacking drives binding to fibrils of the Alzheimer protein Tau. Nat. Commun. 11, 571.

Fersht, A.R., 1987. The hydrogen bond in molecular recognition. Trends Biochem. Sci. 12, 301-304.

Fersht, A.R., 1972. Conformational equilibria in α- and δ-chymotrypsin: The energetics and importance of the salt bridge. J. Mol. Biol. 64(2), 497-509.

Fersht, A.R., Shi, J.P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., Winter, G., 1985. Hydrogen bonding and biological specificity analysed by protein engineering. Nature. 314, 235–238.

Fick, R.J., Kroner, G.M., Nepal, B., Magnani, R., Horowitz, S., Houtz, R.L., Scheiner, S., Trievel, R.C., 2016. Sulfur-Oxygen Chalcogen Bonding Mediates AdoMet Recognition in the Lysine Methyltransferase SET7/9. ACS Chem. Biol. 11, 748–754.

Fischer, F.R., Wood, P.A., Allen, F.H., Diederich, F., 2008. Orthogonal dipolar interactions between amide carbonyl groups. Proc. Natl. Acad. Sci. U. S. A. 105(45), 17290-17294.

Fleming, P.J., Rose, G.D., 2005. Do all backbone polar groups in proteins form hydrogen bonds? Protein Sci. 14, 1911–1917.

Flocco, M.M., Mowbray, S.L., 1994. Planar stacking interactions of arginine and aromatic side-chains in proteins. J. Mol. Biol. 235(2), 709-717.

Gallivan, J.P., Dougherty, D.A., 1999. Cation-π interactions in structural biology. Proc. Natl. Acad. Sci. U. S. A. 96(17), 9459-9464.

Ganguly, H.K., Majumder, B., Chattopadhyay, S., Chakrabarti, P., Basu, G., 2012. Direct evidence for CH···π interaction mediated stabilization of Pro-cisPro bond in peptides with Pro-Pro-aromatic motifs. J. Am. Chem. Soc. 134(10), 4661–4669.

Gerlt, J.A., Kreevoy, M.M., Cleland, W.W., Frey, P.A., 1997. Understanding enzymic catalysis: The importance of short, strong hydrogen bonds. Chem. Biol. 4, 259–267.

Ghosh, N., Mondal, R., Deshmukh, A., Dutta, S., Mukherjee, S., 2015. Weak interactive forces govern the interaction between a non-ionic surfactant with human serum albumin. Chem. Phys. Lett. 634, 77-82.

Gligorijević, V., Renfrew, P.D., Kosciolek, T., Leman, J.K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B.C., Fisk, I.M., Vlamakis, H., Xavier, R.J., Knight, R., Cho, K., Bonneau, R., 2021. Structure-based protein function prediction using graph convolutional networks. Nat. Commun. 12, 3168.

Gregoret, L.M., Rader, S.D., Fletterick, R.J., Cohen, F.E., 1991. Hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Bioinforma. 9(2), 99-107.

Gurry, T., Nerenberg, P.S., Stultz, C.M., 2010. The contribution of interchain salt bridges to triple-helical stability in collagen. Biophys. J. 98(11), 2634-2643.

Heafield, T.G., Hopkins, G., Hunter, L., 1942. Hydrogen bonds involving the sulphur atom. Nature. 149, 218.

Hendsch, Z.S., Tidor, B., 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci. 3(2), 211-226.

Holder, J.B., Bennett, A.F., Chen, J., Spencer, D.S., Byrne, M.P., Stites, W.E., 2001. Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. Biochemistry. 40(46), 13998–14003.

Horowitz, S., Trievel, R.C., 2012. Carbon-Oxygen Hydrogen Bonding in Biological Structure. J. Biol.

Chem. 287(50), 41576–41582.

Hosur, M. V., Chitra, R., Hegde, S., Choudhury, R.R., Das, A., Hosur, R. V., 2013. Low-barrier hydrogen bonds in proteins. Crystallogr. Rev. 19(1), 3–50.

Houser, J., Kozmon, S., Mishra, D., Hammerová, Z., Wimmerová, M., Koča, J., 2020. The CH–π Interaction in Protein–Carbohydrate Binding: Bioinformatics and In Vitro Quantification. Chem. - A Eur. J. 26(47) 10769-10780.

Howard, D.L., Kjaergaard, H.G., 2008. Hydrogen bonding to divalent sulfur. Phys. Chem. Chem. Phys. 10, 4113-4118.

Hubbard, R.E., Kamran Haider, M., 2010. Hydrogen Bonds in Proteins: Role and Strength. in: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

Hudson, K.L., Bartlett, G.J., Diehl, R.C., Agirre, J., Gallagher, T., Kiessling, L.L., Woolfson, D.N., 2015. Carbohydrate-Aromatic Interactions in Proteins. J. Am. Chem. Soc. 137(48), 15152–15160.

Hunter, C.A., Singh, J., Thornton, J.M., 1991. π-π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. J. Mol. Biol. 218(4), 837-846.

Hwang, J., Li, P., Smith, M.D., Shimizu, K.D., 2016. Distance-Dependent Attractive and Repulsive Interactions of Bulky Alkyl Groups. Angew. Chemie - Int. Ed. 55(28), 8086-8089.

Infield, D.T., Rasouli, A., Galles, G.D., Chipot, C., Tajkhorshid, E., Ahern, C.A., 2021. Cation-π Interactions and their Functional Roles in Membrane Proteins: Cation-π interactions in membrane proteins. J. Mol. Biol. 433(17), 167035.

Itoh, Y., Nakashima, Y., Tsukamoto, S., Kurohara, T., Suzuki, M., Sakae, Y., Oda, M., Okamoto, Y., Suzuki, T., 2019. N+-C-H···O Hydrogen bonds in protein-ligand complexes. Sci. Rep. 9, 767.

Iwaoka, M., Isozumi, N., 2012. Hypervalent nonbonded interactions of a divalent sulfur atom. Implications in protein architecture and the functions. Molecules. 17(6), 7266-7283.

Iwaoka, M., Isozumi, N., 2006. Possible roles of S···O and S···N interactions in the functions and evolution of phospholipase A2. Biophysics (Oxf). 2, 23-34.

Iwaoka, M., Takemoto, S., Tomoda, S., 2002. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein engineering. J. Am. Chem. Soc. 124(35), 10613–10620.

Jeffrey, G.A., Saenger, W., 1991. Hydrogen Bonding in Biological Structures, Hydrogen Bonding in Biological Structures. Springer-Verlag, Berlin, Germany.

Jiang, L., Lai, L., 2002. CH···O hydrogen bonds at protein-protein interfaces. J. Biol. Chem. 277(40), 37732–37740.

Joh, N.H., Oberai, A., Yang, D., Whitelegge, J.P., Bowie, J.U., 2009. Similar energetic contributions of packing in the core of membrane and water-soluble proteins. J. Am. Chem. Soc. 131(31), 10846–10847.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature. 596, 583–589

Kamer, K.J., Choudhary, A., Raines, R.T., 2013. Intimate interactions with carbonyl groups: Dipole-dipole or n →π*? J. Org. Chem. 78(5), 2099–2103.

Kauzmann, W., 1959. Some Factors in the Interpretation of Protein Denaturation. Adv. Protein Chem. 14, 1–63

Kellis, J.T., Nyberg, K., Šail, D., Fersht, A.R., 1988. Contribution of hydrophobic interactions to protein stability. Nature. 333, 784–786.

Kemp, M.T., Lewandowski, E.M., Chen, Y., 2021. Low barrier hydrogen bonds in protein structure and function. Biochim. Biophys. Acta - Proteins Proteomics. 1869(1), 140557.

Kendrew, J.C., 1963. Myoglobin and the structure of proteins. Science. 139(3561), 1259-1266.

Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., Shore, V.C., 1960. Structure of myoglobin: A three-dimensional fourier synthesis at 2Å. resolution. Nature. 185, 422–427.

Klapper, M.H., 1971. On the nature of the protein interior. BBA - Protein Struct. 229(3), 557-566.

Kolář, M., Kubař, T., Hobza, P., 2011. On the role of London dispersion forces in biomolecular structure determination. J. Phys. Chem. B 115(24), 8038–8046.

Kollman, P., 1984. Chapter 2 Non-covalent forces of importance in biochemistry. New Compr. Biochem. 6, 55-71.

Kollman, P.A., 1977. Non-covalent Interactions. Acc. Chem. Res. 10(10), 365–371.

Kollman, P.A., Allen, L.C., 1972. The theory of the hydrogen bond. Chem. Rev. 72(3), 283–303.

Kristian, K., Fanfrlík, J., Lepšík, M., 2018. Chalcogen Bonding in Protein−Ligand Complexes: PDB Survey and Quantum Mechanical Calculations. ChemPhysChem. 19(19), 2540-2548.

Kulmanov, M., Hoehndorf, R., 2020. DeepGOPlus: Improved protein function prediction from sequence. Bioinformatics. 36(2), 422–429.

Kumar, M., Balaji, P. V., 2014. C-H...pi interactions in proteins: Prevalence, pattern of occurrence, residue propensities, location, and contribution to protein stability. J. Mol. Model. 20, 2136.

Kumar, S., Nussinov, R., 2001. How do thermophilic proteins deal with heat? Cell. Mol. Life Sci. 58, 1216–1233

Kumar, S., Tsai, C.J., Ma, B., Nussinov, R., 2000. Contribution of salt bridges toward protein thermostability. J. Biomol. Struct. Dyn. 17, 79–85.

Lee, B., Richards, F.M., 1971. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55(3), 379-400.

Lee, C.W., Wang, H.J., Hwang, J.K., Tseng, C.P., 2014. Protein thermal stability enhancement by designing salt bridges: A combined computational and experimental study. PLoS One. 9(11), e112751.

Lee, J., Freddolino, P.L., Zhang, Y., 2017. Ab initio protein structure prediction, in: From Protein Structure to Function with Bioinformatics: Second Edition, Springer, Dordrecht.

Levitt, M., Perutz, M.F., 1988. Aromatic rings act as hydrogen bond acceptors. J. Mol. Biol. 201(4), 751-754.

Li, J., Wang, Y., An, L., Chen, J., Yao, L., 2018. Direct Observation of CH/CH van der Waals Interactions in Proteins by NMR. J. Am. Chem. Soc. 140(9), 3194–3197.

Li, R., Wijma, H.J., Song, L., Cui, Y., Otzen, M., Tian, Y., Du, J., Li, T., Niu, D., Chen, Y., Feng, J., Han, J., Chen, H., Tao, Y., Janssen, D.B., Wu, B., 2018. Computational redesign of enzymes for regio- and enantioselective hydroamination. Nat. Chem. Biol. 14, 664–670.

Lindman, B., Medronho, B., Alves, L., Norgren, M., Nordenskiöld, L., 2021. Hydrophobic interactions control the self-assembly of DNA and cellulose. Q. Rev. Biophys. 54, E3.

Lounnas, V., Wade, R.C., 1997. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. Biochemistry. 36(18), 5402–5417.

Lucas, X., Bauzá, A., Frontera, A., Quiñonero, D., 2016. A thorough anion-π interaction study in biomolecules: On the importance of cooperativity effects. Chem. Sci. 7, 1038-1050.

Mahadevi, A.S., Sastry, G.N., 2016. Cooperativity in Non-covalent Interactions. Chem. Rev. 116(5), 2775–2825.

Marks, D.S., Hopf, T.A., Sander, C., 2012. Protein structure prediction from sequence variation. Nat. Biotechnol. 30, 1072–1080.

Marqusee, S., Baldwin, R.L., 1987. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. Proc. Natl. Acad. Sci. U. S. A. 84(24), 8898-8902.

Marqusee, S., Sauer, R.T., 1994. Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in λ repressor. Protein Sci. 3(12), 2217-2225.

Mati, I.K., Cockroft, S.L., 2010. Molecular balances for quantifying non-covalent interactions Chem. Soc. Rev. 39, 4195-4205.

Mazmanian, K., Sargsyan, K., Grauffel, C., Dudev, T., Lim, C., 2016. Preferred Hydrogen-Bonding Partners of Cysteine: Implications for Regulating Cys Functions. J. Phys. Chem. B 120(39), 10288–10296.

Mecozzi, S., West, A.P., Dougherty, D.A., 1996. Cation-π interactions in aromatics of biological and medicinal interest: Electrostatic potential surfaces as a useful qualitative guide. Proc. Natl. Acad. Sci. U. S. A. 93(20), 10566-10571.

Meyer, E.A., Castellano, R.K., Diederich, F., 2003. Interactions with Arenes Interactions with Aromatic Rings in Chemical and Biological Recognition. Angew. Chemie - Int. Ed . 42(11),1210-50.

Mhaindarkar, D., Gasper, R., Lupilov, N., Hofmann, E., Leichert, L.I., 2018. Loss of a conserved salt bridge in bacterial glycosyl hydrolase BgIM-G1 improves substrate binding in temperate environments. Commun. Biol. 1, 171.

Mirsky, A.E., Pauling, L., 1936. On the Structure of Native, Denatured, and Coagulated Proteins. Proc. Natl. Acad. Sci. 22(7), 439-447.

Misra, S., Singh, P., Mahata, R.N., Brandaõ, P., Roy, S., Mahapatra, A.K., Nanda, J., 2021. Supramolecular antiparallel β-sheet formation by tetrapeptides based on amyloid Sequence. J. Phys. Chem. B. 125(17), 4274-4285.

Morokuma, K., 1971. Molecular Orbital Studies of Hydrogen Bonds. III. C=O···O Hydrogen Bond in H2CO···H2O and H2CO···2H2O. J. Chem. Phys. 55, 1236.

Müller-Dethlefs, K., Hobza, P., 2000. Non-covalent Interactions: A Challenge for Experiment and Theory. Chem. Rev. 100 (1), 143-168.

Muller, P., 1994. Glossary of terms used in physical organic chemistry: (IUPAC Recommendations 1994). Pure Appl. Chem. 66(5), 1077-1184.

Murray, J.S., Lane, P., Clark, T., Politzer, P., 2007a. σ-hole bonding: Molecules containing group VI atoms. J. Mol. Model. 13, 1033–1038.

Murray, J.S., Lane, P., Politzer, P., 2007b. A predicted new type of directional non-covalent interaction. International Journal of Quantum Chemistry. 107(12), 2286-2292.

Narayan, A.R.H., Sherman, D.H., 2013. Re-engineering nature's catalysts. Science. 339(6117), 283-284.

Nayek, A., Gupta, P.S. Sen, Banerjee, S., Mondal, B., Bandyopadhyay, A.K., 2014. Salt-bridge energetics

in halophilic proteins. PLoS One. 9(4), e93862.

Newberry, R.W., Raines, R.T., 2019. Secondary Forces in Protein Folding. ACS Chem. Biol. 14(8), 1677–1686.

Newberry, R.W., Raines, R.T., 2016. A prevalent intraresidue hydrogen bond stabilizes proteins. Nat. Chem. Biol. 12, 1084–1088.

Nick Pace, C., Martin Scholtz, J., Grimsley, G.R., 2014. Forces stabilizing proteins. FEBS Lett. 588(14), 2177-2184.

Nikolova, E.N., Stanfield, R.L., Dyson, H.J., Wright, P.E., 2018. CH···O Hydrogen Bonds Mediate Highly Specific Recognition of Methylated CpG Sites by the Zinc Finger Protein Kaiso. Biochemistry. 57(14), 2109–2120.

Nishio, M., Umezawa, Y., Fantini, J., Weiss, M.S., Chakrabarti, P., 2014. CH-π hydrogen bonds in biological macromolecules. Phys. Chem. Chem. Phys. 16, 12648-12683.

Ovchinnikov, S., Park, H., Varghese, N., Huang, P.S., Pavlopoulos, G.A., Kim, D.E., Kamisetty, H., Kyrpides, N.C., Baker, D., 2017. Protein structure determination using metagenome sequence data. Science. 355(6322), 294-298.

Pace, C.N., 2009. Energetics of protein hydrogen bonds. Nat. Struct. Mol. Biol. 16, 681–682.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Schell, D., Thurlkill, R.L., Imura, S., Scholtz, J.M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J.K., Takano, K., Hebert, E.J., Shirley, B.A., Grimsley, G.R., 2014. Contribution of hydrogen bonds to protein stability. Protein Sci. 23(5), 652–661.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Shirley, B.A., Hendricks, M.M., Iimura, S., Gajiwala, K., Scholtz, J.M., Grimsley, G.R., 2011. Contribution of Hydrophobic Interactions to Protein Stability. J. Mol. Biol. 408(3), 514–528.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19(1), 115–128.

Parks, G.S., 1954. Selected Values of Physical and Thermodynamic Properties of Hydrocarbons and Related Compounds. J. Am. Chem. Soc. 76(7), 2032

Pascoe, D.J., Ling, K.B., Cockroft, S.L., 2017. The Origin of Chalcogen-Bonding Interactions. J. Am. Chem. Soc. 139(42), 15160–15167.

Pauling, L., Corey, R.B., 1951. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. Proc. Natl. Acad. Sci. 37(11), 729-740.

Pauling, L., Corey, R.B., Branson, H.R., 1951. The structure of proteins; two hydrogen-bonded helical

configurations of the polypeptide chain. Proc. Natl. Acad. Sci. U. S. A. 37(4), 205-211.

Penner, R.C., Andersen, E.S., Jensen, J.L., Kantcheva, A.K., Bublitz, M., Nissen, P., Rasmussen, A.M.H., Svane, K.L., Hammer, B., Rezazadegan, R., Nielsen, N.C., Nielsen, J.T., Andersen, J.E., 2014. Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture. Nat. Commun. 5, 5803.

Pinheiro, S., Soteras, I., Gelpí, J.L., Dehez, F., Chipot, C., Luque, F.J., Curutchet, C., 2017. Structural and energetic study of cation-π-cation interactions in proteins. Phys. Chem. Chem. Phys. 19, 9849-9861.

Plevin, M.J., Bryce, D.L., Boisbouvier, J., 2010. Direct detection of CH/π interactions in proteins. Nat. Chem. 2, 466–471.

Politzer, P., Murray, J.S., 2013. Halogen bonding: An interim discussion. ChemPhysChem. 14(2),278-294.

Politzer, P., Murray, J.S., Clark, T., Resnati, G., 2017. The σ-hole revisited. Phys. Chem. Chem. Phys. 19, 32166-32178.

Rahim, A., Saha, P., Jha, K.K., Sukumar, N., Sarma, B.K., 2017. Reciprocal carbonyl-carbonyl interactions in small molecules and proteins. Nat. Commun. 8, 78.

Rajagopal, S., Vishveshwara, S., 2005. Short hydrogen bonds in proteins. FEBS J. 272(8), 1819–1832.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7(1), 95-99.

Ramachandran, G.N., Venkatachalam, C.M., 1966. The stability of the two-bonded collagen triple helix. BBA - Biophys. Incl. Photosynth. 120(3), 457-458.

Rao Mundlapati, V., Ghosh, S., Bhattacherjee, A., Tiwari, P., Biswal, H.S., 2015. Critical assessment of the strength of hydrogen bonds between the sulfur atom of methionine/cysteine and backbone amides in proteins. J. Phys. Chem. Lett. 6(8), 1385-1389.

Ratnaparkhi, G.S., Varadarajan, R., 2000. Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. Biochemistry. 39(40), 12365–12374.

Reilly, A.M., Tkatchenko, A., 2015. Van der Waals dispersion interactions in molecular materials: Beyond pairwise additivity. Chem. Sci. 6, 3289-3301.

Robertazzi, A., Krull, F., Knapp, E.W., Gamez, P., 2011. Recent advances in anion-π interactions. CrystEngComm. 13, 3293-3300.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S., 2010. Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. 79, 233-269.

Rosenfield, R.E., Parthasarathy, R., Dunitz, J.D., 1977. Directional Preferences of Nonbonded Atomic

Contacts with Divalent Sulfur. 1. Electrophiles and Nucleophiles. J. Am. Chem. Soc. 99(14), 4860–4862.

Rozhkov, A. V., Katlenok, E.A., Zhmykhova, M. V., Ivanov, A.Y., Kuznetsov, M.L., Bokach, N.A., Kukushkin, V.Y., 2021. Metal-Involving Chalcogen Bond. The Case of Platinum(II) Interaction with Se/Te-Based σ-Hole Donors. J. Am. Chem. Soc. 143(38), 15701-15710.

Sahariah, B., Sarma, B.K., 2019. Relative orientation of the carbonyl groups determines the nature of orbital interactions in carbonyl-carbonyl short contacts. Chem. Sci. 10, 909-917.

Sarakatsannis, J.N., Duan, Y., 2005. Statistical characterization of salt bridges in proteins. Proteins Struct. Funct. Genet. 60(4), 732-739.

Scheiner, S., 2006. Contributions of NH···O and CH···O hydrogen bonds to the stability of β-sheets in proteins. J. Phys. Chem. B. 110(37), 18670-18679.

Schmiedekamp, A., Nanda, V., 2009. Metal-activated histidine carbon donor hydrogen bonds contribute to metalloprotein folding and function. J. Inorg. Biochem. 103(7), 1054-1060.

Schottel, B.L., Chifotides, H.T., Dunbar, K.R., 2008. Anion-π interactions. Chem. Soc. Rev. 37, 68-83.

Scilabra, P., Terraneo, G., Resnati, G., 2019. The Chalcogen Bond in Crystalline Solids: A World Parallel to Halogen Bond. Acc. Chem. Res. 52(5), 1313-1324.

Senes, A., Ubarretxena-Belandia, I., Engelman, D.M., 2002. The C --H···O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. Proc. Natl. Acad. Sci. 98(16) 9056-9061.

Serrano, L., Bycroft, M., Fersht, A.R., 1991. Aromatic-aromatic interactions and protein stability. Investigation by double-mutant cycles. J. Mol. Biol. 218(2), 465-475.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M., Fersht, A.R., 1990. Estimating the Contribution of Engineered Surface Electrostatic Interactions to Protein Stability by Using Double-Mutant Cycles. Biochemistry. 29(40), 9343-9352.

Shan, S.O., Loh, S., Herschlag, D., 1996. The energetics of hydrogen bonds in model systems: Implications for enzymatic catalysis. Science. 272(5258), 97-101.

Smith, M.S., Lawrence, E.E.K., Billings, W.M., Larsen, K.S., Bécar, N.A., Price, J.L., 2017. An Anion-π Interaction Strongly Stabilizes the β-Sheet Protein WW. ACS Chem. Biol. 12(10), 2535-2537.

Sticke, D.F., Presta, L.G., Dill, K.A., Rose, G.D., 1992. Hydrogen bonding in globular proteins. J. Mol. Biol. 226(4), 1143–1159.

Strop, P., Mayo, S.L., 2000. Contribution of surface salt bridges to protein stability. Biochemistry. 39(6), 1251-1255.

Tanford, C., 1962. Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. J. Am. Chem. Soc. 84 (22), 4240–4247.

Tarannam, N., Shukla, R., Kozuch, S., 2021. Yet another perspective on hole interactions. Phys. Chem. Chem. Phys. 23, 19948-19963.

Tissot, A.C., Vuilleumier, S., Fersht, A.R., 1996. Importance of two buried salt bridges in the stability and folding pathway of barnase. Biochemistry. 35(21), 6786-6794.

Torshin, I.Y., Weber, I.T., Harrison, R.W., 2002. Geometric criteria of hydrogen bonds in proteins and identification of "bifurcated" hydrogen bonds. Protein Eng. 15(5), 359–363.

Tsuzuki, S., 2012. CH/π interactions. Annu. Reports Prog. Chem. - Sect. C. 108, 69–95.

Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M., Tanabe, K., 2002. Origin of attraction and directionality of the π/π interaction: Model chemistry calculations of benzene dimer interaction. J. Am. Chem. Soc. 124(1), 104-112.

Tu, T., Li, Y., Su, X., Meng, K., Ma, R., Wang, Y., Yao, B., Lin, Z., Luo, H., 2016. Probing the role of cation-π interaction in the thermotolerance and catalytic performance of endo-polygalacturonases. Sci. Rep. 6, 38413,

Umeyama, H., Morokuma, K., 1977. The Origin of Hydrogen Bonding. An Energy Decomposition Study. J. Am. Chem. Soc. 99(5), 1316-1332.

Umezawa, Y., Nishio, M., 2000. CH/π interactions in the crystal structure of TATA-box binding protein/DNA complexes. Bioorganic Med. Chem. 8(11), 2643-2650.

van de Waterbeemd, H., Carter, R.E., Grassy, G., Kubinyi, H., Martin, Y.C., Tute, M.S., Willett, P., 1997. Glossary of terms in computational drug design (IUPAC Recommendations 1997). Pure Appl. Chem. 69(5), 1137-1152.

Vernon, R.M.C., Chong, P.A., Tsang, B., Kim, T.H., Bah, A., Farber, P., Lin, H., Forman-Kay, J.D., 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. Elife. 7, e31486.

Waals, van der, D, J., 1873. On the Continuity of the Gaseous and Liquid States. Elsevier Sci. Publ. B.V.

Wagner, J.P., Schreiner, P.R., 2015. London Dispersion in Molecular Chemistry - Reconsidering Steric Effects. Angew. Chemie - Int. Ed. 54(42), 12274-12296.

Waldburger, C.D., Schildbach, J.F., Sauer, R.T., 1995. Are buried salt bridges important for protein stability and conformational specificity? Nat. Struct. Biol. 2, 122–128.

Wang, J., Chen, J., Li, J., An, L., Wang, Y., Huang, Q., Yao, L., 2018. Arginine side chain stacking with peptide plane stabilizes the protein helix conformation in a cooperative way. Proteins Struct. Funct. Bioinforma. 86(6), 684-692.

Wheeler, S.E., Houk, K.N., 2009. Substituent effects in cation/π interactions and electrostatic potentials above the centers of substituted benzenes are due primarily to through-space effects of the substituents. J. Am. Chem. Soc. 131(9), 3126-3127.

Whisstock, J.C., Lesk, A.M., 2003. Prediction of protein function from protein sequence and structure. Q. Rev. Biophys. 36(3), 307-340.

Wilson, K.A., Kellie, J.L., Wetmore, S.D., 2014. DNA-protein π-interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. 42(10), 6726–6741.

Wimley, W.C., Gawrisch, K., Creamer, T.P., White, S.H., 1996. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. Proc. Natl. Acad. Sci. U. S. A. 93(7), 2985-2990.

Wintjens, R., Liévin, J., Rooman, M., Buisine, E., 2000. Contribution of cation-π interactions to the stability of protein-DNA complexes. J. Mol. Biol.302(2), 393-408.

Worley, B., Richard, G., Harbison, G.S., Powers, R., 2012. ^{13}C NMR reveals no evidence of n-π* interactions in proteins. PLoS One. 7(8), e42075.

Xiao, F., Chen, Z., Wei, Z., Tian, L., 2020. Hydrophobic Interaction: A Promising Driving Force for the Biomedical Applications of Nucleic Acids. Adv. Sci. 7(16),2001048.

Yan, C., Wu, F., Jernigan, R.L., Dobbs, D., Honavar, V., 2008. Characterization of protein-protein interfaces. Protein J. 27, 59–70,

Zhang, H., Li, C., Yang, F., Su, J., Tan, J., Zhang, X., Wang, C., 2014. Cation-pi interactions at non-redundant protein-RNA interfaces. Biochem. 79, 643–652.

Zhao, Y., Cotelle, Y., Liu, L., López-Andarias, J., Bornhof, A.B., Akamatsu, M., Sakai, N., Matile, S., 2018. The Emergence of Anion-π Catalysis. Acc. Chem. Res. 51(9), 2255-2263.

Zhou, P., Tian, F., Lv, F., Shang, Z., 2009. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Bioinforma. 76(1), 151–163.

Zhou, S., Wang, L., 2019. Unraveling the structural and chemical features of biological short hydrogen bonds. Chem. Sci. 10, 7734-7745.

Ziegler, T., Rauk, A., 1977. On the calculation of bonding energies by the Hartree Fock Slater method. Theor. Chim. Acta. 46, 1–10.

Chapter 2

Characteristics of σ-hole mediated chalcogen bond involving divalent sulfur

2.1 Introduction

Non–covalent interactions formed by divalent elements of group 16 with nucleophiles, called chalcogen bond (Ch-bond), have drawn considerable attention in recent years (Dube et al., 2012; Garrett et al., 2015; Metrangolo and Resnati, 2012; Wang et al., 2009; Zhang and Wang, 2009). Identified in many crystal structures of small, supra, and biomolecules, these interactions are suggested to play essential roles in self–assembly (Gleiter et al., 2003; Gopalakrishna et al., 2013; Werz et al., 2002), catalysis (Biot and Bonifazi, 2018; Garrett et al., 2016; Mahmudov et al., 2017; Robinson et al., 2016) and biological processes (Fick et al., 2016; Iwaoka et al., 2002a; Iwaoka and Isozumi, 2012; Reid et al., 2014; Rosenfield et al., 1977; Zhang et al., 2015). For example, unlike hydrogen bonds, the chalcogen bond can facilitate catalysis in an aprotic solvent (Benz et al., 2017). Also, certain macrocycles employ divalent selenium in anion recognition through a chalcogen interaction (Lim and Beer, 2018). In proteins, divalent S in the methionine or cysteine could form Ch-bond with the carbonyl O of the peptide backbone. Such Ch-bonds are suggested to be important for protein stability because of their comparable strength to that of the conventional hydrogen bond (Iwaoka et al., 2002b; Pal and Chakrabarti, 2001).

Ch-bond belongs to a broader class of interactions called σ–hole bonding (Dyduch et al., 2013; Hennemann et al., 2012; Mohajeri et al., 2009; Politzer et al., 2008; Politzer and Murray, 2010). According to Murray et al., "*the term 'sigma–hole' originally referred to the electron-deficient outer lobe of a half-filled p (or nearly p) orbital involved in forming a covalent bond. If the electron deficiency is sufficient, there can result in a region of positive electrostatic potential which can interact attractively (non–covalently) with negative sites on other molecules (sigma–hole bonding)*" (Murray et al., 2009). In general, group 14, 15, 16, and 17 elements form such σ–holes and the corresponding interactions are called tetral, pnicogen, chalcogen, and halogen interaction, respectively (Bauzá et al., 2015; Brezgunova et al., 2013; Cavallo et al., 2016; Desiraju et al., 2013; Legon, 2010; Metrangolo and Resnati, 2008; Politzer et al., 2013; Wang et al., 2016).

The molecular electrostatic potentials (MESP) is a powerful tool to examine these σ–hole interactions (Hennemann et al., 2012; Murray et al., 2009). Besides, it is often used to understand various chemical aspects such as bonding, reactivity, and delocalization (Murray et al., 1994; Murray and Politzer, 2011; Politzer et al., 2001, 1985; Politzer and Murray, 2002; Sjoberg and Politzer, 1990; Suresh et al., 2000). MESP for a molecule is defined by (in atomic unit, au):

$$V(\boldsymbol{r}) = \sum_{\mathrm{A}}^{\mathrm{N}} \frac{\mathrm{Z_A}}{|\mathbf{r} - \mathbf{R_A}|} - \int \frac{\rho(\mathbf{r}')\mathrm{d}^3\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} \qquad (1)$$

Here, $\{Z_A\}$ represents the nuclear charges located at position vectors $\{\mathbf{R_A}\}$. $\rho(r)$ is the molecular electron density measurable from X-ray diffraction experiments. This charge density is defined as:

$$\rho(\mathbf{r}) = N \sum_{\sigma} \int |\psi(\mathbf{x}, \mathbf{x}_2, .., \mathbf{x}_N)|^2 \, d^3r_2 \ldots . d^3r_N \quad (2)$$

Here, molecular electron density distribution, $\rho(r)$, can be obtained from the many-electron wave function $\psi(\mathbf{x_1}, \mathbf{x_2}, .., \mathbf{x_N})$, where $\{\mathbf{x_i}\}$ is the spin–space position vector of the i^{th} electron and N is the total number of electrons.

MESP depends on the balance between the electronic and nuclear components. The sign of $V(r)$ is negative when the second term in equation 1 is dominant over the first term and positive when the first term is dominant over the second term. The negative regions of MESP signify positions of favorable electrophilic attack. These negative regions usually appear near chemically interesting facets of lone pairs, π-bonds, aromatic π–clouds, facile–strained bonds, etc. (Gadre and Pundlik, 1995; Kumar et al., 2014; Mohan et al., 2013; Solà, 2013; Suresh and Gadre, 1999). A simple way of analyzing MESP is to compute $V(r)$ on a molecular surface (van der Waals surface or 0.01 au density isosurface), which is usually represented in the literature as $V_S(r)$. The most negative $V_S(r)$ value represents an electron-rich region of the molecule. It is termed as $V_{S,min}$, while the most positive value of $V_S(r)$ represents an electron-deficient region of the molecule labeled as $V_{S,max}$. In the case of the σ–hole bonding interactions, the $V_{S,max}$ values depend on the electronegativity and polarizability of the corresponding atoms in the molecule (Politzer et al., 2013). Using computational methods for halogenated complexes, a linear relationship between the magnitude of σ–hole, $V_{S,max}$, and complexation energy (ΔE) has been observed (Deepa et al., 2014; Politzer et al., 2013).

In general, the emerging importance of Ch-bond makes it essential to understand their characteristic features. In this study, we aim to employ MESP to rationalize the strength and directional qualities of the Ch-bond formed between divalent S and O/N in molecules of biological relevance. We also study if the interaction can modulate the strength of cumulative molecular interactions giving rise to cooperativity, which, to the best of our knowledge, has not been investigated before (Adhikari and Scheiner, 2014, 2011; Lundemba et al., 2020; Scheiner, 2011;

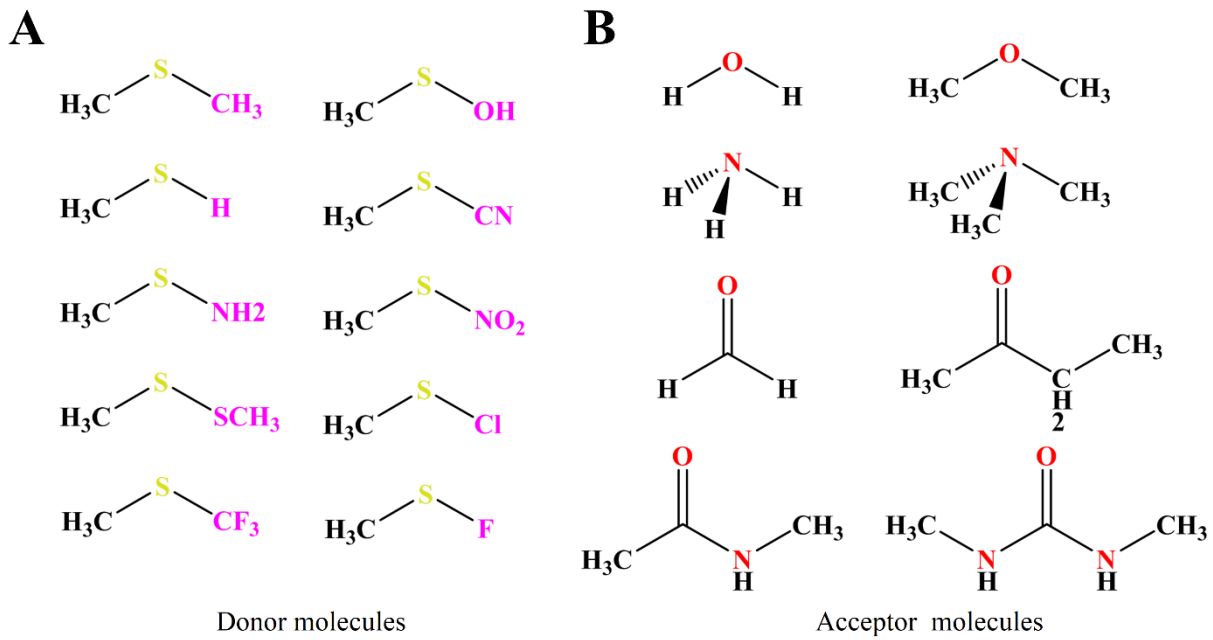


Figure 2.1 Monomers used for studying divalent S mediated Ch-bond. (**A**) Divalent S molecular scaffolds employed in this study. (**B**) N and O containing nucleophiles used for complexation with divalent S monomers.

Shukla and Chopra, 2016). Additionally, the recent spectroscopic investigations showed that the formation of S···O interaction is driven by orbital interaction (Pascoe et al., 2017). Also, based on the analysis of protein crystal structures, the direction of approach of divalent S towards the π–electron instead of the lone–pair (n_o) region of the carbonyl group in the peptide backbone was attributed to orbital delocalization (Iwaoka et al., 2002b). Both these observations contradict the nature of S···O interaction that has a predominantly electrostatic component similar to other σ–hole interactions (Metrangolo and Resnati, 2001; Murray et al., 2009; Politzer et al., 2013). This led us to ask if the origin of such directional selectivity of S···O interaction can be explained based on MESP, a quantity that can be experimentally measured (Murray and Politzer, 2011; Suresh et al., 2000).

To address this question, we chose a variety of divalent S–containing molecules having substituents with different electron-withdrawing capacities (CH_3SX, where X= –F, –Cl, $–NO_2$, –CN, –OH, $–CF_3$, $–SCH_3$, $–NH_2$, –H, and $–CH_3$; Figure 2.1 **A**). Strength and directionality of S···O/N interaction were investigated by complexing these S monomers with simple nucleophiles such H_2O, $(CH_3)_2O$, H_2CO, NH_3, and $N(CH_3)_3$ (Figure 2.1 **B**). Furthermore, we provided a rationale based on MESP for the preferential approach of divalent S towards π–region instead of

the n_o–region of backbone carbonyl group that was noted in protein structures (Iwaoka et al., 2002b). For this, we used ethyl-methyl ketone (referred to as ketone), N–methyl acetamide (referred to as amide), and N–dimethyl urea (referred to as DMU) (Figure 2.1 **B**). Additionally, we showed that this directional selectivity for Ch-bond was an effect of the cooperative nature of other coexisting weak interactions. Understanding such coordination among weak interactions in proteins could help to develop strategies in protein designing and engineering.

2.2 Computational Section

All the monomers and dimeric complexes used for analysis had been optimized using the M06 density functional (Goerigk, 2015; Mardirossian and Head-Gordon, 2016; Zhao and Truhlar, 2008) using the basis set 6–31G++(2D, 2P) implemented within the Gaussian 09 (G09) program (Frisch et al., 2009). Appendix **A** provides the coordinates for the all-optimized structures. The M06 functional is an apt choice, recommended for main group thermochemistry, organometallics, and non–covalent interactions (Zhao and Truhlar, 2008). In order to confirm that all these monomeric and dimeric optimized complexes are real minima on the potential energy surface, we performed Hessian evaluations. It was confirmed that all the structures presented are minima on the potential energy minima (PES) from the positive eigenvalues of the Hessian. We have studied the test complex, $H_2O{:}S(CH_3)F$, using various quantum chemical methods (Table 2.1) and found that the stabilization energy was comparable for the dispersion added density functional theory (B3LYP–D) (DiLabio et al., 2013) and M06 functional. The intermolecular interaction energy of dimeric complexes (ΔE^*_{AB}) can be calculated using Equation 3, where the sum of the energy of monomer A and B (E_A and E_B) is subtracted from the overall energy of complexation (E_{AB}). However, equation 4 can account for basis set superposition error (BSSE) where E_A^{AB} and E_B^{AB} are energies of monomers A and B calculated using a basis set used earlier for dimer and provide more accurate complexation energy (ΔE_{AB}) as suggested by Boys and Bernadi (Boys and Bernardi, 1970; Řezáč and Hobza, 2016).

$$\Delta E^*_{AB} = E_{AB} - E_A - E_B \qquad (3)$$

$$\Delta E_{AB} = E_{AB} - E_A^{AB} - E_B^{AB} \qquad (4)$$

BSSE correction was estimated using the counterpoise correction method employed in G09 programs. Molecular electrostatic potential (MESP) maps were calculated and visualized with an appropriate isovalue using Gauss view 05 (Dennington et al., 2009).

Table 2.1 Structural parameters and binding energies for H_2O:$S(CH_3)F$ complex using different quantum methods.

Method	Basis set	d (Å)[a]	θ (deg.)[b]	ΔE (kcal.mol^{-1})[c]
MP2	6-31G++(2D, 2P)	2.82	179.7	-3.77
MP4	6-31G++(2D, 2P)	2.89	179.8	-3.39
B3LYP	6-31G++(2D, 2P)	2.89	177.7	-3.07
B3LYP-D	6-31G++(2D, 2P)	2.85	178.6	-4.52
M06	6-31G++(2D, 2P)	2.76	178.4	-4.60

[a]Distance between S and O, [b]angle between F, S, and O involve in S···O interaction. [c]BSSE corrected complexation energies using the counterpoise correction method.

Topological properties associated with the electron density at the bond critical point were obtained from the quantum theory of Atoms in Molecules (AIM) implemented in the AIM2000 program (Bader, 2002, 1991, 1990; Biegler-König and Schönbohm, 2002; König et al., 2001). The Quantum theory of Atoms in Molecules (QTAIM) is a powerful approach to characterize weak non–covalent interactions (Bone and Bader, 1996). QTAIM explores the electron density, ρ(r), of the molecule. The density saddle points along the bond path between the atoms are defined as bond critical point (BCP). The topological features associated with the BCP, such as electron density, ρ(r), and Laplacian of electron density, $\nabla^2\rho$(r), can be used to investigate the strength and nature of the bonding. The line of maximum ρ(r) connecting the neighboring atoms is termed as a bond path. $\nabla^2\rho(r) < 0$ implies charge concentration at the BCP, indicating the covalent nature of the chemical bond. However, when $\nabla^2\rho(r) > 0$, the concentration of the electron density is towards the nuclei. This depletion is a measure of the ionic nature of the interaction (Bader, 2002). The wavefunction file was generated using G09 program at M06/6–31G++ (2D, 2P) for the AIM analysis. Molecular electrostatic potential critical points (MESP CPs) were found and characterized using the code developed and implemented in the DAMQT program (Fernández Rico et al., 2004a, 2004b; Kumar et al., 2015; López et al., 2015, 2009). These MESP CPs are generally represented as (R, S), where rank (R) refers to the number of non–zero eigenvalues while signature (S) refers to the algebraic sum of the signs of these eigenvalues. When R=3, there are four types of the non–degenerated CPs; (3,–3), that represents the local minimum, (3,–1) and (3,+1) that represent the saddle points, and (3,+3) that represents the local maximum for MESP.

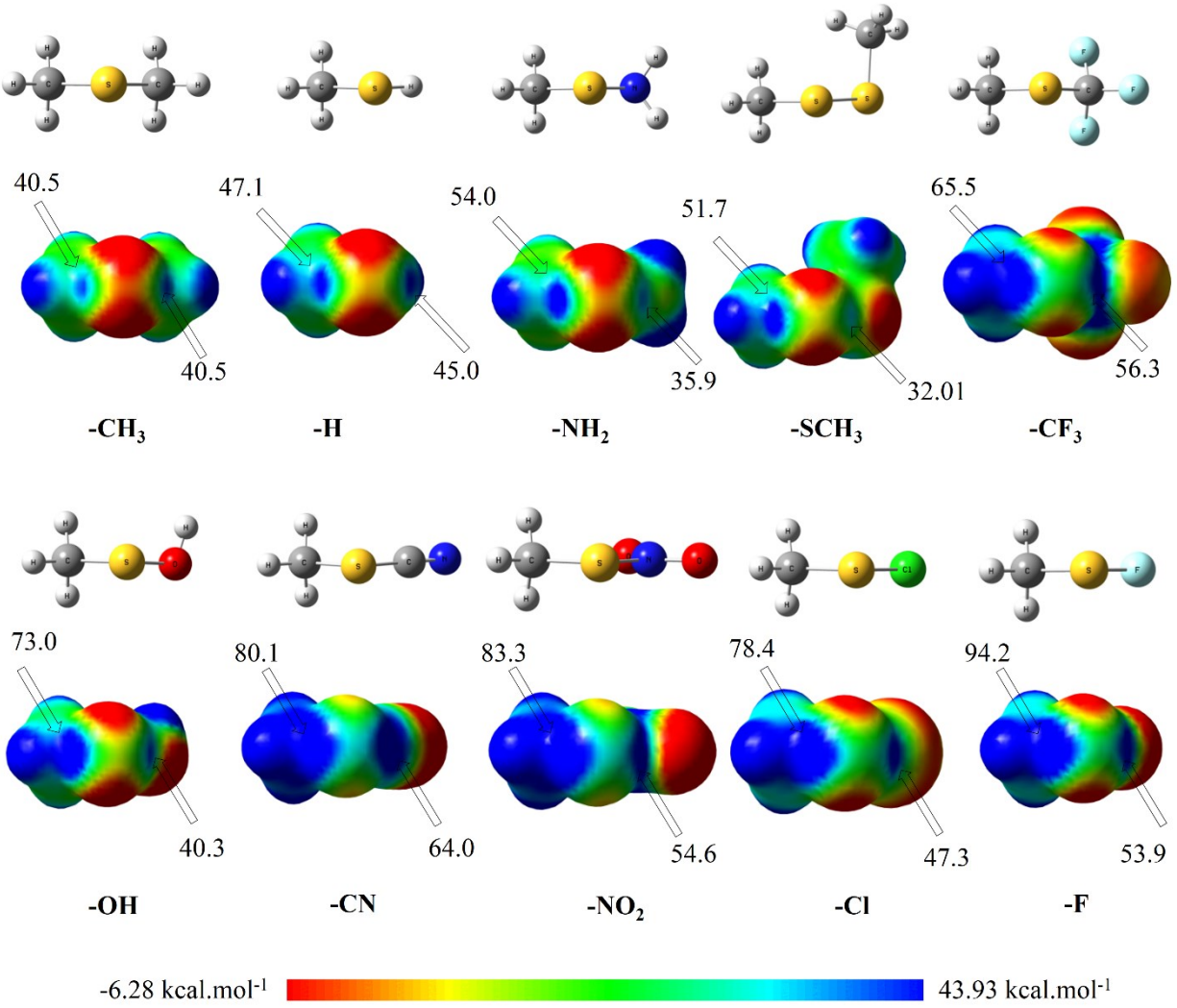


Figure 2.2 Structure and MESP analysis of divalent S monomers. Energy minimized structures of divalent S scaffold (CH_3-S-X) where, -X= –CH_3, –H, –NH_2, –SCH_3, –CF_3, –OH, –CN, –NO_2, –Cl, –F and their MESP maps with a color-coding range from –6.28 (red) to 43.93 kcal.mol^{-1} (blue) textured on 0.01 au density isosurface. The two σ–holes formed on the extension S–X/CH_3 bonds in these fragments are marked by arrows, and their magnitude ($V_{S,max}$) are shown in kcal.mol^{-1}.

2.3 Results

2.3.1 Surface electrostatic potential maps of divalent S-containing monomers

Figure 2.2 shows the surface electrostatic potential maps textured on 0.01 au density isosurface for X–S–CH_3 molecular motifs, where -X represents the different substituents (see Figure 2.1 **A**). These MESP maps revealed two σ–holes (positive regions colored blue) along the extension of the S–X and S–C covalent bonds of S (Figure 2.2). The values of surface MESP maximum, $V_{S,max}$, are given in kcal.mol^{-1}, represents the strength of these σ–holes. These values of σ–hole opposite

to C–X bond increased as the electron-withdrawing nature of the substituents increased (Figure 2.2) and ranged from 40.5 kcal.mol^{-1} for –CH_3 to 94.2 kcal.mol^{-1} for –F. Interestingly, despite –Cl and –OH being more electronegative, values of $V_{S,max}$ for the σ–hole with them as substituents were lower than–NO_2 and –CN (Figure 2.2). This is possible because –NO_2 and –CN have the potential to withdraw electron density through inductive as well as through resonance effects. In terms of MESP, this would amount to through bond and through-space interactions. Arranging the substituents based on the values of $V_{S,max}$ resulted in the following series for potential σ–hole interactions.

$$–F > –NO_2 > –CN > –Cl > –OH > –CF_3 >–NH_2 > –SCH_3 > –H > –CH_3$$

a similar trend of $V_{S,max}$ for divalent selenium (Shukla and Chopra, 2015) and sulfur was reported previously (Lundemba et al., 2020; Shukla and Chopra, 2016).

In contrast, variation in the magnitude of σ–hole opposite to S–CH_3 bond did not seem to be strongly correlated with the electron-withdrawing nature of the substituents. This could be because of the proximity between the corresponding σ–hole and substituents with negative electrostatic potential regions because of their lone pairs or π–bonds. In addition to the two σ–holes, S also had two electronegative regions (red regions in Figure 2.2), corresponding to the lone–pair regions of divalent S, which can interact with electrophiles.

The MESP derived charges (CHELPG charges from G09) on the S attached to different substituents were 0.028, –0.032, –0.045, –0.078, –0.045, –0.111, –0.101, –0.104, –0.308, –0.205 au for –F, –Cl, –NO_2, –CN, –OH, –CF_3, –SCH_3, –NH_2, –H, –CH_3, respectively. This result implied that as the electron-withdrawing power of the substituents increased negative charge on S decreased. –F group substitution at S made it entirely positive (0.028 au). As $V_{S,max}$ of only the σ–hole opposite to C–X bond correlated with the electron-withdrawing power of the substituents, we limited ourselves to studying this σ–hole's interaction with the nucleophiles in Figure 2.1 **B**.

2.3.2 Structures of H_2O:$S(CH_3)X$, H_2CO:$S(CH_3)X$, and NH_3:$S(CH_3)X$ complexes

Next, H_2O, H_2CO, and NH_3 were allowed to interact so that the lone pair region of O and N could point towards the σ–hole on the S opposite to the C–X bond. Figures 2.3 **A-C** show geometries of some of the representative H_2O/H_2CO/NH_3:$S(CH_3)X$ energy–minimized dimers, along with their corresponding MESP maps (for coordinates of all optimized dimers, refer to appendix **A**). Their structural parameters are provided in Table 2.2. Of the 30 structures evaluated, the geometric

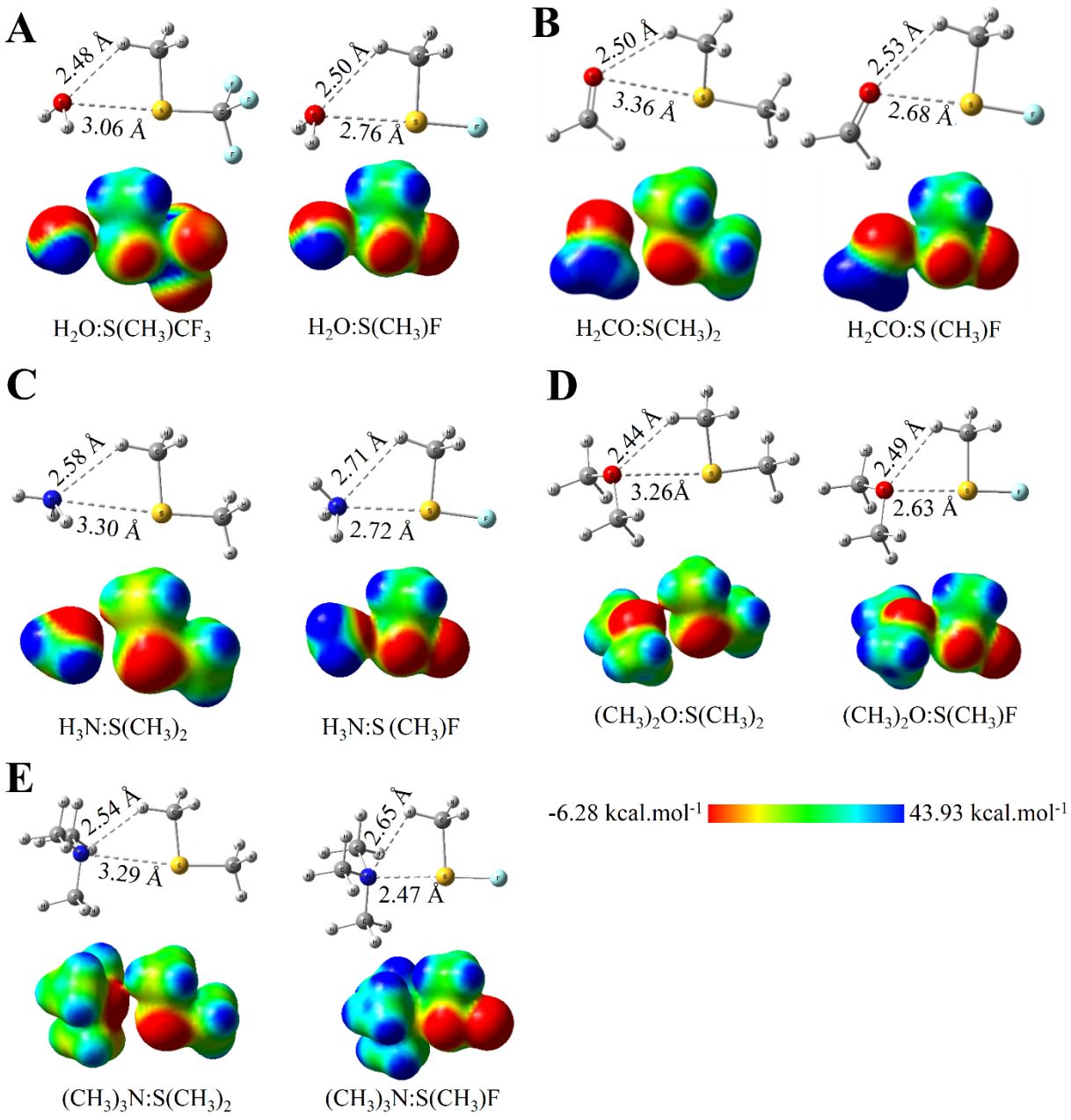


Figure 2.3 Structure and MESP analysis of H_2O:$S(CH_3)X$, H_2CO:$S(CH_3)X$, and NH_3:$S(CH_3)X$ complexes. (A) Energy minimized structure and MESP maps for H_2O:$S(CH_3)CF_3$/$S(CH_3)F$ complexes. **(B)** Energy minimized structures and MESP maps for H_2CO:$S(CH_3)_2$/$S(CH_3)F$ complexes. **(C)** Energy minimized structures and MESP maps for H_3N:$S(CH_3)_2$/$S(CH_3)F$ complexes. **(D)** Energy minimized structures and MESP maps for $(CH_3)_2O$:$S(CH_3)_2$/$S(CH_3)F$ complexes. **(E)** Energy minimized structures and MESP maps for $(CH_3)_3N$:$S(CH_3)_2$/$S(CH_3)F$ complexes.

Table 2.2 Structural parameters, complexation energies, and topological parameters for different H_2O:$S(CH_3)X$, NH_3:$S(CH_3)X$, and H_2CO:$S(CH_3)X$ complexes.

	-X	d^a	θ^b	d^c	θ^d	ΔE^e	ρ^f	$\nabla^2\rho^g$	ρ^h	$\nabla^2\rho^i$
H_2O	CF_3	3.06	174.0	2.48	127.0	-2.98	0.01055	0.0344	0.00944	0.0351
	-CN	3.01	176.8	2.44	126.5	-3.64	0.01109	0.0376	0.00992	0.0376
	-OH	2.99	173.7	2.54	123.6	-3.10	0.01195	0.0382	0.00907	0.0329
	$-NO_2$	2.97	179.6	2.47	125.2	-4.01	0.01226	0.0389	0.00956	0.0363
	-Cl	2.92	177.3	2.51	122.5	-3.80	0.01369	0.0427	0.00927	0.0350
	-F	2.76	178.5	2.50	117.6	-4.60	0.01799	0.0566	-	-
NH_3	$-CH_3$	3.30	176.2	2.58	131.2	-2.63	0.00650	0.0322	9.53	0.0286
	-H	3.32	173.2	2.56	133.4	-2.86	0.00643	0.0224	9.97	0.0294
	$-NH_2$	3.21	178.7	2.66	127.5	-2.48	0.00770	0.0262	8.53	0.0254
	$-SCH_3$	3.27	174.0	2.65	129.4	-2.87	0.00731	0.0237	8.61	0.0256
	$-CF_3$	3.11	176.3	2.67	125.3	-3.62	0.01011	0.0314	8.71	0.0264
	-CN	3.07	178.3	2.64	125.2	-4.40	0.01109	0.0341	9.22	0.0277
	-OH	3.02	172.2	2.77	120.2	-3.45	0.01149	0.0365	7.70	0.0237
	$-NO_2$	3.01	174.6	2.70	122.2	-4.80	0.01304	0.0370	8.56	0.0268
	-F	2.72	174.3	2.71	111.8	-5.87	0.01540	0.0703	-	-
H_2CO	$-CH_3$	3.36	164.7	2.50	131.9	-2.53	0.00594	0.02296	0.00882	0.03186
	-H	3.31	164.4	2.50	132.6	-2.70	0.00635	0.02408	0.00881	0.03216
	$-NH_2$	3.31	166.7	2.53	130.9	-2.42	0.00622	0.02360	0.00838	0.03023
	$-SCH_3$	3.29	170.8	2.53	131.2	-2.62	0.00677	0.02432	0.00818	0.03032
	$-CF_3$	2.99	176.3	2.54	124.1	-3.31	0.01182	0.03910	0.00820	0.03245
	-OH	2.94	170.0	2.57	120.7	-3.40	0.01278	0.04199	0.00798	0.03103
	-Cl	2.85	174.6	2.54	119.7	-4.00	0.01550	0.04902	-	-
	-F	2.68	176.4	2.53	114.6	-4.86	0.02080	0.06628	-	-

[a]Distance between S and N/O in Å and [b]angle between X, S, and N/O in degrees involved in S···N/O interaction. [c]Distance between H and N/O in Å and [d]angle between C, H, and N in degrees involved in C-H···N/O interaction. [e]BSSE corrected complexation energies in kcal.mol^{-1}using the counterpoise correction method. [f, g]Topological properties ρ(e Å^{-3}) and $\nabla^2\rho$ (e Å^{-5}) for S···N/O interaction. [h, i]Topological properties ρ (e Å^{-3}) and $\nabla^2\rho$ (e Å^{-5}) for C-H···N/O interaction.

minima of 23 complexes had real frequencies (Table 2.2). In contrast, the remaining had either real minimum involving O–H···S hydrogen bond instead of S···O interaction or had an imaginary frequency.

The distance (S···O/N) and angle (X–S···O/N) measurements of the optimized complexes indicated a strong substitution effect on S···O/N interaction (Figure 2.3 **A-C** and Table 2.2). The sum of van der Waals radii of S and N, and S and O are 3.35 Å and 3.32 Å, respectively (Bondi, 1964). The distance between S and O/N in the optimized structures were less than their van der Waals radii sum irrespective of the substituents. As the electron-withdrawing power of the substituents increased, the distance between S···O/N decreased. For $–CF_3$ and –CN substitutions, the distance between S and O was 3.06 Å and 3.01 Å, respectively, which decreased to 2.76 Å on

substitution with –F when interacting with H_2O (Table 2.2). We also observed a similar trend for H_2CO:$S(CH_3)X$ or H_3N:$S(CH_3)X$ complexes (Table 2.2). For example, in the H_3N:$S(CH_3)X$ complexes, the distances between S and O were 3.30 Å and 2.72 Å for $–CH_3$ and –F substitutions, respectively. The angle measured between X, S, and O/N ranged from 172˚ to 180˚ (Table 2.2), matching the angular range for σ–hole mediated interactions in literature (Brezgunova et al., 2013; Murray et al., 2009).

Along with S···O/N interaction, C–H···O/N interaction was noted in all the optimized complexes (Figure 2.3 **A-C**). In the H_2O:$S(CH_3)X$ complexes, when -X = $–CF_3$ and –F, the distance between H and O was 2.48 Å and 2.50 Å, and C–H···O angles were 127.0˚ and 117.6˚ respectively (Figure 2.3 **A-C** and Table 2.2). In the case of $–CH_3$, –H, $–SCH_3$ and $–NH_2$ substituted divalent S molecules, MESP revealed that the negative region corresponding to the lone–pair of N in NH_3 was pointing towards –H of CH_3 group rather than σ–hole of S (A representative example shown in Figure 2.3 **C**). Hence, for these substituents, the C–H···N interaction appeared more favorable than S···N interaction. However, in the case of $–CF_3$, –CN, –OH, and –F substituted groups, the lone–pair pointed towards the σ–hole of S (A representative example shown in Figure 2.3 **C**). Hence, we concluded that NH_3:$S(CH_3)X$ complexes were stabilized by C–H···N interaction in case of X= $–CH_3$, –H, $–SCH_3$ or $–NH_2$ and S···N interaction in case of -X = $–CF_3$, –CN, –OH or –F. We think that similar trend would be observed for other nucleophiles too.

2.3.3 Structures of $(CH_3)_2O$:$S(CH_3)X$ and $(CH_3)_3N$:$S(CH_3)X$ complexes

Next, we investigated the dimers involving nucleophiles such as $(CH_3)_2O$ and $N(CH_3)_3$, which occur widely in chemical and biological systems. The complex nature of these nucleophiles could represent the realistic scenario for studying the S···N/O interaction found in crystal structures. In addition, $-CH_3$ is a weak electron-donating group that could increase the nucleophilicity of N and O and, thus, could provide more stability to dimers than the nucleophiles discussed in the previous section. The optimized dimeric complexes were real minima on PES as no imaginary frequency was found. Figure 2.3 **D-E** shows some representative complexes with their respective MESP maps (for the coordinates of all optimized complexes, refer to appendix **A**). Table 2.3 provides structural parameters associated with these complexes.

Analysis of these energy minimized structures revealed that the distances between the S and O/N were shorter than those observed in $H_2CO/H_2O/NH_3$:$S(CH_3)X$ complexes (Table 2.2 and 2.3). For $–CH_3$ and –F substitution on divalent S, the distances between S and O were 3.26 Å and

Table 2.3 Structural parameters, complexation energies, and topological parameters for different $(CH_3)_2O$:S(CH_3)X and $(CH_3)_3N$:S(CH_3)X complexes.

	-X	d^a	θ^b	d^c	θ^d	ΔE^e	ρ^f	$\nabla^2\rho^g$	ρ^h	$\nabla^2\rho^i$
$(CH_3)_2O$	$-CH_3$	3.26	172.9	2.44	132.8	-3.07	0.00730	0.0259	10.40	0.0358
	-H	3.05	171.4	2.50	124.4	-3.35	0.01123	0.0362	9.39	0.0342
	$-NH_2$	3.02	173.8	2.51	123.5	-3.21	0.01164	0.0370	9.45	0.0343
	$-SCH_3$	3.03	169.8	2.49	123.6	-3.53	0.01180	0.0368	9.73	0.0354
	$-CF_3$	2.98	174.5	2.49	124.0	-4.23	0.01277	0.0401	9.64	0.0354
	-CN	2.95	175.2	2.48	123.6	-4.81	0.01333	0.0422	9.92	0.0363
	-OH	2.88	172.7	2.49	120.4	-4.37	0.01481	0.0463	9.87	0.0367
	$-NO_2$	2.84	174.8	2.50	120.3	-5.22	0.01628	0.0491	-	-
	-Cl	2.77	175.6	2.50	117.7	-5.26	0.01878	0.0567	-	-
	-F	2.63	176.9	2.49	113.1	-6.41	0.02372	0.0728	-	-
$(CH_3)_3N$	$-CH_3$	3.29	171.1	2.54	131.0	-3.26	0.00829	0.0240	10.68	0.0315
	-H	3.20	172.1	2.58	128.1	-3.62	0.00988	0.0281	10.09	0.0299
	$-NH_2$	3.05	174.8	2.61	121.8	-3.61	0.01290	0.0361	9.82	0.0303
	$-SCH_3$	3.09	168.8	2.61	122.2	-3.94	0.01232	0.0336	9.72	0.0298
	$-CF_3$	3.02	173.2	2.59	122.8	-4.85	0.01367	0.0378	10.10	0.0315
	-CN	2.99	174.7	2.58	122.3	-5.62	0.01406	0.0408	10.36	0.0324
	-OH	2.78	175.1	2.65	113.2	-5.71	0.02098	0.0595	-	-
	$-NO_2$	2.75	173.6	2.65	114.2	-6.97	0.02340	0.0624	-	-
	-Cl	2.64	174.7	2.65	110.3	-8.26	0.02822	0.0760	-	-
	-F	2.47	177.3	2.65	100.3	-11.26	0.03911	0.0996	-	-

[a]Distance between S and N/O in Å and [b]angle between X, S, and N/O in degrees involved in S···N/O interaction. [c]Distance between H and N/O in Å and [d]angle between C, H, and N in degrees involved in C-H···N/O interaction. [e]BSSE corrected complexation energies in kcal.mol^{-1}using the counterpoise correction method. [f, g]Topological properties ρ (e Å^{-3}) and $\nabla^2\rho$ (e Å^{-5}) for S···N/O interaction. [h, i]Topological properties ρ (e Å^{-3}) and $\nabla^2\rho$ (e Å^{-5}) for C-H···N/O interaction.

2.63 Å, respectively. While for S and N, it was found to be 3.29 Å and 2.47 Å, respectively (Figure 2.3 **D-E)**. The distance between S and O/N was within this range for other substituents, and X–S···O/N angle was within 170˚ to 180˚ (Table 2.3). These observations were similar to that of complexes from the previous section. MESP maps clearly showed that the lone–pair region of N and O was not completely pointing towards –H as was noted in $(CH3)_2S$:NH_3, H(CH_3)S:H_2CO, and $NH_2(CH_3)$S:H_2O complexes. Instead, the lone pair region of N was pointing between –H and the σ–hole of S (Figure 2.3 **E**). Consequently, in the case of NH_3:S(CH_3)X, it was inferred that for weak electron-withdrawing substituents, the predominant interaction was C–H···N (see the previous section). We also concluded that increasing the nucleophilicity of NH_3 may result in S···N interaction contributing towards stabilizing the complexes. This could also be true for other nucleophiles.

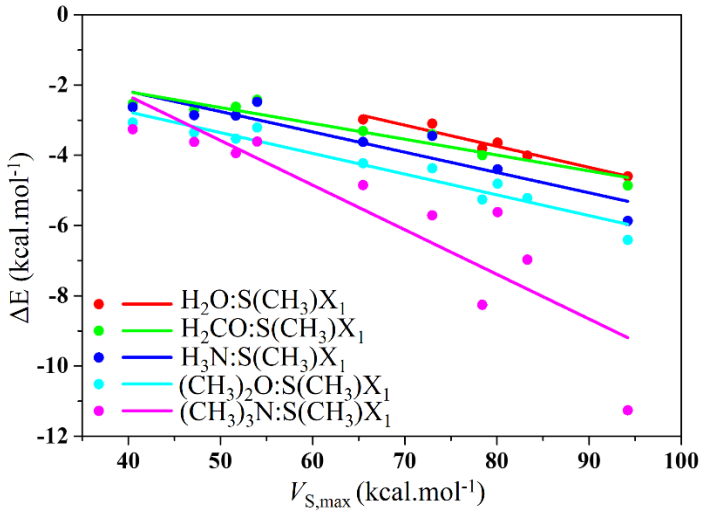


Figure 2.4 A plot of Δ*E* versus $V_{S,max}$. Correlation between BSSE corrected complexation energies for H_2O:$S(CH_3)X$ (r^2=0.93), H_2CO:$S(CH_3)_2$/$S(CH_3)X$ (r^2=0.90), H_3N:$S(CH_3)X$ (r^2=0.84), $(CH_3)_2O$:$S(CH_3)X$ (r^2=0.92) and $(CH_3)_3N$:$S(CH_3)X$ (r^2=0.77) complexes and magnitude of σ–hole ($V_{S,max}$) in kcal.mol^{-1}. - X=CH_3, –H, $–NH_2$, $–SCH_3$, $–CF_3$, –OH, –CN, $–NO_2$, –Cl, and –F.

2.3.4 Complexation energies (Δ*E*) for H_2O:$S(CH_3)X$, NH_3:$S(CH_3)X$, H_2CO:$S(CH_3)X$, $(CH_3)_2O$:$S(CH_3)X$ and $(CH_3)_3N$:$S(CH_3)X$ complexes

To further characterize the interaction between the fragments in the dimer, we calculated interaction energies that stabilize them. To measure the cumulative/overall strengths of these dimers (Figure 2.3 and Table 2.2 and 2.3), we calculated their BSSE corrected Δ*E*. These Δ*E* values range from –2.48 to –11.26 kcal.mol^{-1} (Tables 2.2 and 2.3). This implies strong stabilization by coexisting S···O/N and C–H···N/O interactions. Δ*E* involving S···N interaction was higher than S···O interaction, presumably because of the higher electron-donating ability of N over O (Table 2.2 and 2.3). Also, Δ*E* involving the S···O interaction formed by $O(CH_3)_2$ was higher than H_2O. This could be because of the weak electron-donating effect of the methyl group increasing the nucleophilicity of O.

Earlier, Adhikari *et al*. performed a similar analysis with NH_3 at MP2/aug–cc–pVDZ level of theory, where the nucleophile was allowed to interact with various $S(CH_3)X$ monomers (X = – CH_3, –H, $–NH_2$, $–CF_3$, –OH, $–NO_2$, –Cl, and –F) (Adhikari and Scheiner, 2014). They reported Δ*E* values for these complexes ranging from –2.32 to –4.42 kcal.mol^{-1}. Our analysis employed M06 density functional (see Computational Section), and the calculated Δ*E* values ranged from –2.63 to –5.87 kcal.mol^{-1}. The values reported by Adhikari et al.(2014) were slightly lower than

those obtained by us possibly because of the underestimation of the interaction energies by MP2/aug–cc–pVDZ theoretical level. For instant, the MP2 method underestimates the Δ*E* of the Ne_2 or methane/2–buyne:HF/LiF dimers, which are primarily driven by dispersion and induction interactions, respectively (Roy et al., 2012).

Next, we plotted $V_{S,max}$ of the σ–hole opposite to C–X bond as a function of Δ*E* for all the complexes studied (Figure 2.4). From the plot, it is clear that as the electron-withdrawing power of substituents increased, the magnitude of σ–hole also increased, which increased the complexation energy. The trend for complexation energy with varying substituents was in the following order,

$$–F > –Cl > –NO_2 > –CN \approx –OH > –CF_3 > –SCH_3 \approx –NH_2 > –H \approx –CH_3$$

This order showed that $–NO_2$ and –CN substitution on S provided higher stabilization than –OH and was consistent with $V_{S,max}$ values on S for these substituents (see above). We also noted that the overall strength of Δ*E* was strongly dependent on the nature of the nucleophiles. Following is a pattern (in increasing order) for the ability of the nucleophiles to provide higher stabilization (Figure 2.4),

$$N(CH_3)_3 > O(CH_3)_3 > NH_3 > H_2CO > H_2O$$

The correlation between Δ*E* and $V_{S,max}$ is clear evidence for the S···O/N interaction, which is predominantly of an electrostatic nature (Figure 2.4).

2.3.5 Atoms in molecules (AIM) analysis for H_2O:$S(CH_3)X$, NH_3:$S(CH_3)X$, H_2CO:$S(CH_3)X$, $(CH_3)_2O$:$S(CH_3)X$ and $(CH_3)_3N$:$S(CH_3)X$ complexes

We next performed AIM analysis to find density values at BCP for C–H···O/N and S···O/N interactions in these complexes to seek their contribution in overall Δ*E* (Table 2.2 and 2.3 with some representative examples are shown in Figure 2.5). Density values at BCP, $\rho(r)$ characterizing S···O/N interaction, were in the range of 0.00594–0.03911 ($e/Å^3$) (Table 2.2 and 2.3). We noted that as the electron-withdrawing power of the substituent increased, the density value at the BCP increased, implying that the strength of the S···O/N increased (Figure 2.5 and 2.6). Also, the strength of the S···O/N interaction increased with change in the nucleophile in the following order (Table 2.2-2.3 and Figure 2.6).

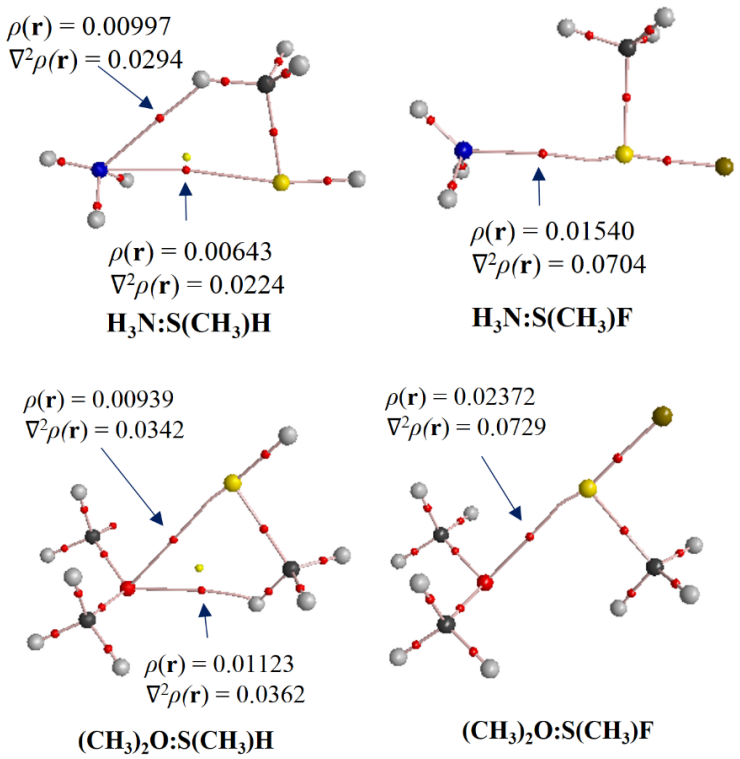


Figure 2.5 Representative examples of AIM analysis. Some of the representative examples of AIM analysis with the topological properties ρ (e Å^{-3}) and $\nabla^2\rho$ (e Å^{-5}) at BCP characterizing for C-H···N/O interaction.

$$N(CH_3)_3 > O(CH_3)_3 > NH_3 > H_2CO > H_2O$$

Based on these observations, we concluded that the strength of the S···O/N interaction depends on the nature of the substituent and the nucleophile, and such trends were absent for C–H···O/N interaction (Figure 2.4 and 2.6).

Interestingly, for some of the complexes involving –F, –Cl, and –NO_2 substitution, we did not find a BCP for C–H···O interaction, possibly because of a strong S···O/N interaction (Figure 2.5 and Table 2.2-2.3). The ρ values for S···O/N interaction, particularly for –CH_3, –H, –SCH_3, and –NH_2 substitution, were lower than C–H···N interaction, suggesting their weaker nature in these cases (Figure 2.6). This is consistent with the inference drawn from the MESP maps (Figure 2.3). $\nabla^2\rho$(r) had positive values for all the complexes, which increased as the electron-withdrawing power of the substituent increased (Table 2.2-2.3). This trend is consistent with the ΔE values discussed in the previous section.

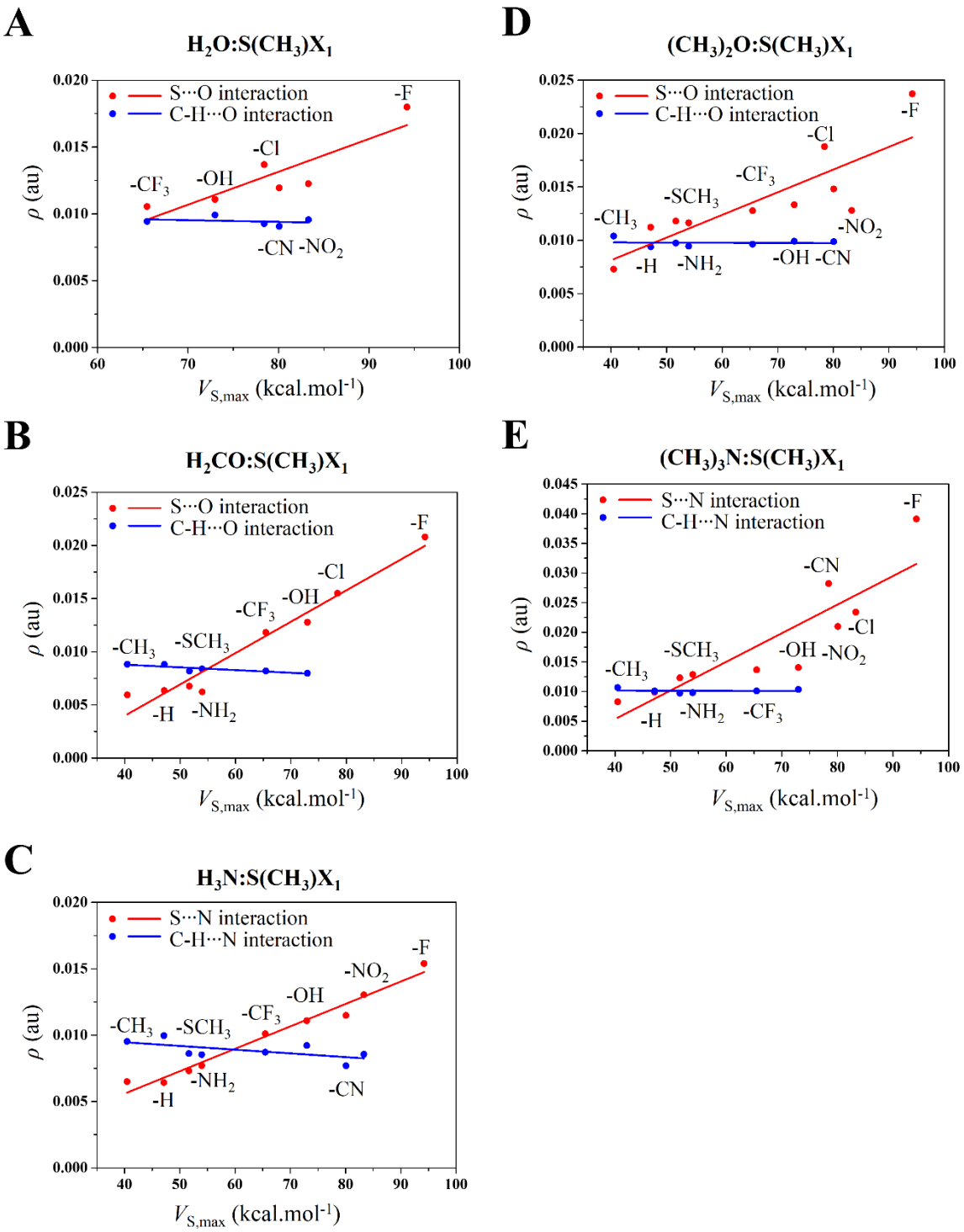

A
H2O:S(CH3)X1
S···O interaction
C-H···O interaction
-F
-Cl
-CF3
-OH
-CN
-NO2
ρ (au)
VS,max (kcal.mol-1)
B
H2CO:S(CH3)X1
S···O interaction
C-H···O interaction
-F
-Cl
-OH
-CF3
-CH3
-SCH3
-H
-NH2
ρ (au)
VS,max (kcal.mol-1)
C
H3N:S(CH3)X1
S···N interaction
C-H···N interaction
-F
-NO2
-OH
-CF3
-CH3
-SCH3
-NH2
-H
-CN
ρ (au)
VS,max (kcal.mol-1)
D
(CH3)2O:S(CH3)X1
S···O interaction
C-H···O interaction
-F
-Cl
-CF3
-SCH3
-CH3
-NO2
-H
-NH2
-OH
-CN
ρ (au)
VS,max (kcal.mol-1)
E
(CH3)3N:S(CH3)X1
S···N interaction
C-H···N interaction
-F
-CN
-Cl
-SCH3
-OH
-NO2
-CH3
-H
-NH2
-CF3
ρ (au)
VS,max (kcal.mol-1)

Figure 2.6 A plot of $\rho(\mathbf{r})$ versus $V_{S,max}$. (A) Correlation between molecular electron density $\rho(\mathbf{r})$ at BCPs characterizing C–H···O and S···O interactions with $V_{S,max}$ in kcal.mol^{-1} for H_2O:S(CH_3)X complexes. **(B)** Correlation between molecular electron density $\rho(\mathbf{r})$ at BCPs characterizing C–H···O and S···O interactions with $V_{S,max}$ in kcal.mol^{-1} for $(CH_3)_2O$:S(CH_3)X complexes. **(C)** Correlation between molecular electron density $\rho(\mathbf{r})$ at BCPs characterizing C–H···N and S···N interactions with $V_{S,max}$ in kcal.mol^{-1} for H_3N:S(CH_3)X complexes. **(D)** Correlation between molecular electron density $\rho(\mathbf{r})$ at BCPs characterizing C–H···O and S···O interactions with the $V_{S,max}$ in kcal.mol^{-1} for $(CH_3)_2O$:S(CH_3)X complexes. **(E)** Correlation between molecular electron density $\rho(\mathbf{r})$ at BCPs characterizing C–H···N and S···N interaction with $V_{S,max}$ in kcal.mol^{-1} for $(CH_3)_3N$:S(CH_3)X complexes. -X = CH_3, –H, –NH_2, –SCH_3, –CF_3, –OH, –CN, –NO_2, –Cl, and –F.

Among the two interactions, the strength of the C–H···O/N interaction was independent of the nature of the substituents attached to S and the type of nucleophile (Figure 2.6). Hence, we concluded that the correlation between ΔE and $V_{S,max}$ for all the complexes discussed above is dominated by S···O/N interaction (Figure 2.4 and 2.6). Also, since the change in the strength of the S···O/N interaction modulated the cumulative molecular interaction, we concluded cooperation between the two coexisting interactions (Figure 2.4 and 2.6). In summary, the AIM analysis revealed that S···O/N and C–H···O/N interactions were coexisting and contributed to the overall stability of the systems in a reciprocal way. Hence, we propose that these coexisting interactions can serve as a signature pair of interactions involved in molecular recognition.

2.3.6 Directional behavior of the S···O interaction involving carbonyl functional groups

Apart from the nucleophiles studied above, carbonyl functional groups, which are widely found in organic molecules and biomolecules, can also act as nucleophiles. This class of nucleophiles has two electron-rich regions, the π–region and the lone pair region, with which the σ-hole of S can interact. Crystal structure analyses indicate that a divalent S can approach the carbonyl O in two ways (Figure 2.7 **A**). 1) Perpendicular to the carbonyl group, referred to as a π–direction approach; 2) along the lone–pair direction referred to as n_o–direction. Previously, orbital interaction was used to explain the preferential approach of S in π–direction over n_o–direction resulting in S···O bond (Iwaoka et al., 2002b). However, as the origin of S···O bond is electrostatic in nature, we aim to find the electrostatic basis for the directional selectivity of S···O interaction and its cooperativity with other weak interactions.

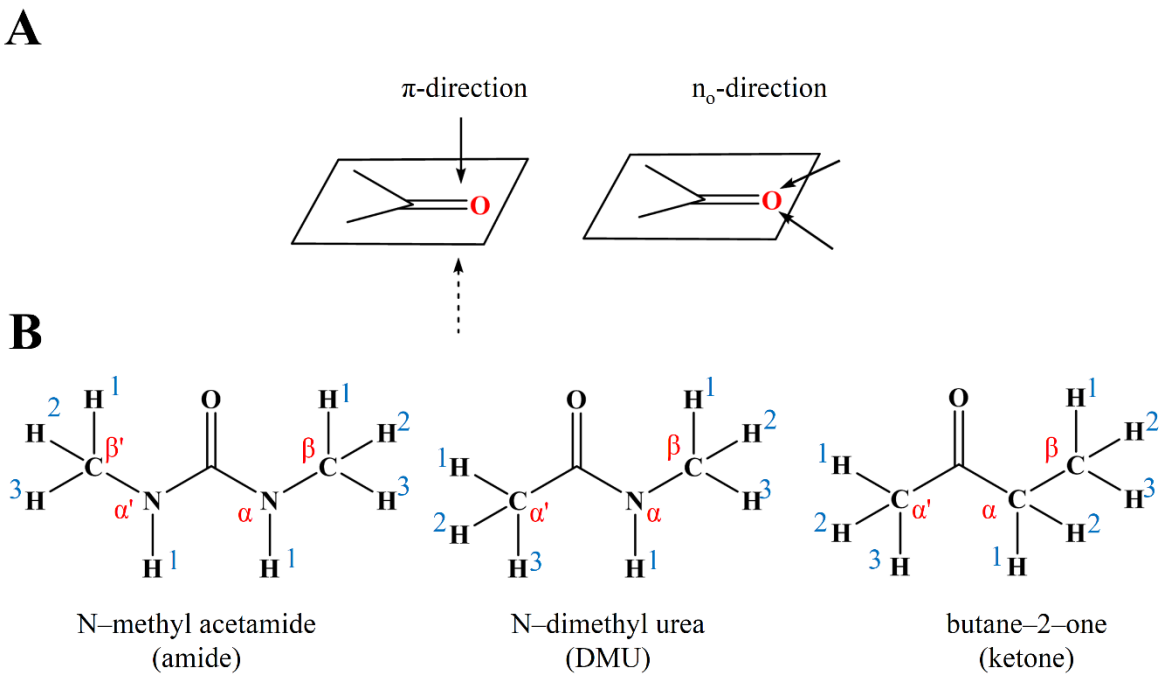


Figure 2.7 Directional selectivity of S···O interaction. (A) The two directional preferences of a divalent S to approach a carbonyl functional group, namely π–direction and n_o–direction. **(B)** Structures of DMU, Amide, and Ketone with their atoms labeled appropriately.

For this, we selected carbonyl-containing molecules such as N–dimethyl urea (DMU), which has two –NH_2 group*s* in conjugation, N–methyl acetamide (amide) which has one –NH2 group in conjugation, and butane–2–one (ketone) that does not have any conjugation with the carbonyl group. This selection of nucleophiles provided a way to systematically change the π–electronic environment of the carbonyl group. We allowed the various carbonyl motifs (Figure 2.7 **B**) to interact with S(CH_3)X fragments in the π–direction and n_o–direction. Some of the representative energy minimized complexes in π–direction and n_o–direction are shown in Figure 2.8 (for all complexes and their structural parameters, see Appendix **A**). Using structural analysis (refer to Figure 2.8 and Table 2.4 and 2.5), as discussed in the previous sections, we noticed the presence and substituent influenced geometry of the S···O=C interaction in all complexes. The distance between S and O decreased as the electron-withdrawing power of the substituents increased. However, unlike nucleophiles discussed in the previous sections where X–S···O/N angle ranged within 170˚–180˚, here, the linearity of X–S···O=C angle increased as the electron-withdrawing power of substituents increased (Table 2.4 and 2.5). For example, in DMU:S(CH_3)X complexes, in which S approached O=C along the π–direction, X–S···O=C angles were 178.2˚, 174.7˚ and 165.9˚ for –F, –CN and –CH_3 substituents, respectively (Table 2.4).

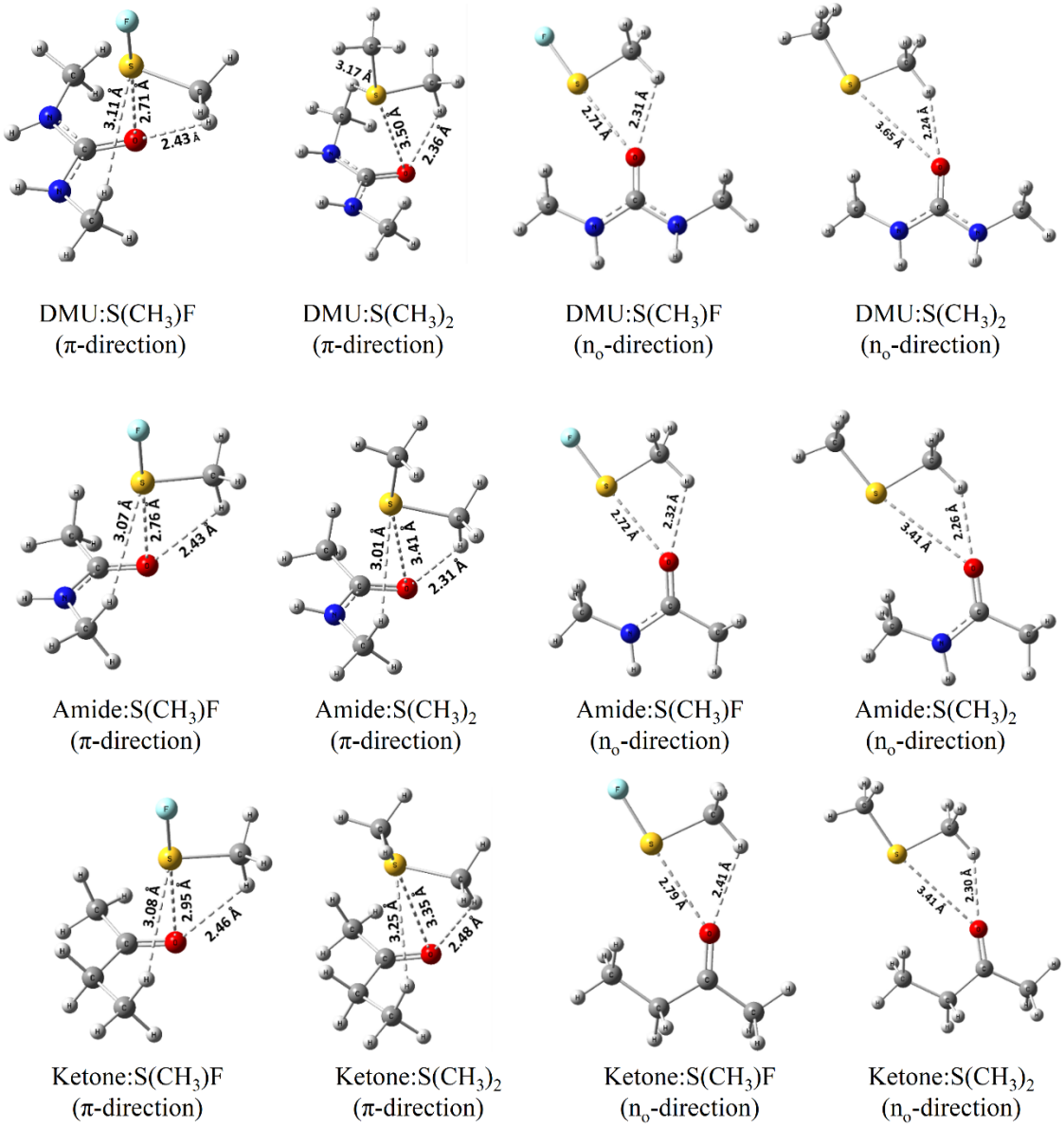


Figure 2.8 Energy minimized structures of representative dimers in π–direction and n_o-direction. Some representative energy minimized complexes in the π–direction and n_o-direction. Interaction distances are given in Å.

Furthermore, these structures revealed that the four interactions *viz*. S···O=C, C–H···O=C, and two C–H···S interactions could stabilize these complexes (Table 2.4 and 2.5). Amongst the coexisting interactions, C–H···O=C was the most prominent interaction in all the above–optimized complexes (Figure 2.8). As noted in the previous section, the strength of the C–H···O bond was neither affected significantly by the substituent on the divalent S nor by the nature of the nucleophile. In general, the distance between H and O in the C–H···O=C interaction in all these complexes was in the range of 2.31 to 2.48 Å. However, the angle for the C–H···O=C interaction

Table 2.4 Structural parameters and complexation energies for DMU:S(CH_3)X, Amide:S(CH_3)X and ketone:S(CH_3)X complexes in π-direction.

	-X	d^a	θ^b	d^c	θ^d	d^e	θ^f	d^g	θ^h	ΔE^i
DMU	-CH_3	3.50	165.9	2.36	142.2	3.27	118.5	3.17	117.7	-5.26
	-H	3.31	166.8	2.33	137.8	3.63	117.5	3.12	124.5	-5.04
	-NH_2	3.33	174.3	2.36	137.2	3.10	122.5	3.28	119.6	-5.60
	-SCH_3	3.30	169.4	2.42	134.5	3.29	122.5	3.22	121.9	-5.81
	-CF_3	3.00	173.6	2.34	128.7	3.31	124.1	3.31	122.3	-6.11
	-CN	2.97	174.7	2.34	128.0	3.32	123.8	3.32	122.2	-5.90
	-OH	2.95	173.9	2.44	124.4	3.47	120.4	3.03	123.5	-7.01
	-NO_2	2.89	176.1	2.40	124.3	3.29	120.4	3.33	124.5	-7.45
	-Cl	2.90	179.8	2.41	124.4	3.28	119.4	3.28	119.4	-7.02
	-F	2.71	178.2	2.43	117.6	3.11	123.0	3.31	124.1	-7.87
Amide	-CH_3	3.41	167.2	2.31	141.1	3.01	123.7	3.29	113.2	-4.69
	-H	3.32	167.4	2.29	139.7	3.04	124.0	3.34	112.3	-4.91
	-SCH_3	3.32	159.4	2.41	131.8	3.26	115.0	3.24	122.1	-4.96
	-NH_2	3.26	175.7	2.34	135.9	3.01	125.9	3.34	111.8	-4.89
	-CF_3	3.02	173.2	2.36	128.6	3.09	127.1	3.48	109.1	-5.63
	-CN	3.00	174.7	2.32	129.4	3.09	127.7	3.53	108.6	-5.47
	-OH	2.99	175.4	2.42	126.2	3.08	125.4	3.29	111.4	-6.40
	-NO_2	2.91	177.9	2.41	125.0	3.28	124.5	3.48	109.2	-6.88
	-Cl	2.93	176.6	2.44	125.1	3.07	127.3	3.50	107.9	-6.34
	-F	2.76	177.1	2.43	119.3	3.07	127.9	3.36	108.1	-7.13
Ketone	-CH_3	3.35	165.0	2.48	132.7	3.25	126.2	3.20	111.9	-4.46
	-H	3.32	162.3	2.32	137.1	3.08	130.2	3.26	112.3	-4.66
	-NH_2	3.32	170.6	2.50	133.2	3.13	128.6	3.25	112.0	-4.22
	-SCH_3	3.31	165.8	2.45	133.3	3.14	129.3	3.27	112.8	-4.72
	-CF_3	3.29	161.6	2.34	135.4	3.04	129.5	3.37	111.5	-4.78
	-CN	3.04	171.0	2.38	128.6	3.24	128.4	3.47	103.5	-5.35
	-OH	3.08	167.5	2.42	128.2	3.27	125.1	3.23	105.2	-5.00
	-NO_2	3.00	176.3	2.43	126.9	3.28	128.0	3.51	103.1	-5.43
	-Cl	3.00	175.7	2.42	127.1	3.07	130.1	3.51	104.1	-5.09
	-F	2.95	177.1	2.46	124.8	3.08	129.2	3.47	102.6	-5.49

[a]Distance between S and O in Å, [b]angle measured between X, S, and O in degrees involved in the formation of S···O=C interaction. [c]Distance between interacting H and O in Å, [d]angle measured between C, H and O in degrees involved in C-H···O=C interaction between –CH_3 group in S(CH_3)X and the carbonyl O atom. [e]Distance between $H_{\beta1}$ and S in Å, [f]angle measured between C_β, $H_{\beta1}$ and S in degrees involved in C_β-$H_{\beta1}$···S interaction. [g]Distance between $H_{\beta'1}/H_{\alpha'1}$ and S in Å, [h]angle measured between $C_{\beta'}/C_{\alpha'}$, $H_{\beta'1}/H_{\alpha'1}$ and S in degrees involved in $C_{\beta'}/C_{\alpha'}$- $H_{\beta'1}/H_{\alpha'1}$···S interaction. [i]BSSE corrected complexation energies in kcal.mol^{-1} using the counterpoise correction method. For nomenclature refer figure 2.7 **B**.

Table 2.5 Structural parameters and complexation energies for DMU:S(CH_3)X, Amide:S(CH_3)X and ketone:S(CH_3)X complexes in n_o-direction.

	-X	d^a	θ^b	d^c	θ^d	d^e	θ^f	d^g	θ^h	ΔE^i
DMU	-CH_3	3.65	160.5	2.24	151.7	3.02	124.3	3.46	96.8	-2.78
	-H	3.33	165.2	2.25	140.8	3.00	123.4	3.43	96.0	-3.09
	-NH_2	3.31	169.6	2.29	138.9	3.00	124.5	3.47	95.0	-2.76
	-SCH_3	3.32	170.5	2.28	139.3	2.99	123.3	3.42	95.8	-3.51
	-CF_3	3.26	167.9	2.25	139.2	3.20	115.5	3.41	102.3	-4.88
	-CN	2.99	173.3	2.27	130.0	3.24	113.0	3.43	101.7	-4.21
	-OH	2.98	175.5	2.31	127.9	3.01	120.9	3.43	94.9	-6.17
	-NO_2	2.92	177.0	2.30	127.8	3.02	118.6	3.39	95.7	-6.7
	-Cl	2.92	176.2	2.29	127.5	3.00	118.3	3.36	96.0	-5.67
	-F	2.71	177.0	2.31	120.1	2.95	118.5	3.36	93.4	-6.52
Amide	-CH_3	3.41	165.8	2.26	142.1	2.97	124.4	3.43	95.4	-2.85
	-H	3.38	165.8	2.29	141.4	3.04	120.3	3.39	98.6	-3.24
	-SCH_3	3.33	170.4	2.29	139.4	3.00	124.2	3.47	95.1	-4.33
	-NH_2	3.32	169.6	2.30	139.1	2.99	124.2	3.46	95.1	-3.65
	-CF_3	3.25	168.8	2.29	137.8	3.04	121.2	3.43	96.6	-4.93
	-CN	2.99	173.7	2.28	129.7	3.22	108.5	3.27	105.6	-4.39
	-OH	2.99	174.8	2.30	128.4	3.00	120.9	3.43	94.7	-6.19
	-NO_2	2.93	177.3	2.31	127.8	3.20	108.7	3.27	104.6	-6.83
	-Cl	2.92	176.3	2.31	127.2	2.99	115.1	3.27	98.5	-5.81
	-F	2.72	117.2	2.32	119.8	2.98	112.9	3.23	97.7	-6.69
Ketone	-CH_3	3.41	166.5	2.30	141.0	3.06	123.7	3.46	98.1	-2.04
	-H	3.39	168.1	2.43	138.3	3.02	123.1	3.42	97.9	-2.45
	-NH_2	3.34	171.1	2.38	137.2	3.13	121.5	3.48	99.6	-2.18
	-SCH_3	3.33	173.4	2.40	136.6	3.07	121.1	3.41	99.5	-2.73
	-CF_3	3.07	173.2	3.36	130.2	3.19	112.1	3.30	105.6	-3.88
	-CN	3.01	174.1	2.32	129.4	3.19	114.5	3.38	102.7	-4.94
	-OH	3.00	175.3	2.38	126.5	3.04	120.8	3.42	96.9	-3.36
	-NO_2	2.97	177.2	2.34	128.1	3.18	108.3	3.18	108.0	-5.39
	-Cl	2.95	176.6	2.34	127.2	3.16	108.3	3.16	107.8	-4.67
	-F	2.79	178.1	2.41	120.1	2.97	115.4	3.24	99.2	-5.35

[a]Distance between S and O in Å, [b]angle measured between X, S, and O in degrees involved in the formation of S···O=C interaction. [c]Distance between interacting H and O in Å, [d]angle measured between C, H and O in degrees involved in C-H···O=C interaction between the –CH_3 group in S(CH_3)X and the carbonyl O atom. [e]Distance between $H_{\beta'1}$ and S in Å, [f]angle measured between $C_{\beta'}$ $H_{\beta'1}$ and S in degrees involved in $C_{\beta'}$-$H_{\beta'1}$···S interaction. [g]Distance between $H_{\beta'2}$ and S in Å, [h]angle measured between $C_{\beta'}$, $H_{\beta'2}$ and S in degrees involved in $C_{\beta'}$-$H_{\beta'2}$···S interaction. [i]BSSE corrected complexation energies in kcal.mol^{-1} using counterpoise correction method. For nomenclature refer figure 2.7 **B**.

varied systematically. The C–H···O angle in DMU:S(CH_3)X complexes in the π-direction was 117.6˚, 128.0˚ and 142.2˚ for –F, –CN and –CH_3 substitutions (Table 2.4). Thus, the linearity of the C–H···O angle increased as the electron-withdrawing power of the substituent decreased (Table 2.4 and 2.5). Hence, we concluded that a C–H···O=C interaction could significantly contribute to the overall stability of the complexes, particularly for X= –CH_3, –H, –NH_2, and –SCH_3. In the case of the two C–H···S bonds formed between S and the –CH_3 groups of the carbonyl (Table 2.4 and 2.5), their distances ranged from 3.1 to 3.5 Å. At the same time, C–H···S angles were found to be between 100˚ to 130˚ for all the substituents (Table 2.4 and 2.5). In order to investigate the stability of and directional preferences in these complexes and to study the contribution of these four interactions towards this, we computed Δ*E* as a function of different substitutions on S and different carbonyl functional groups.

2.3.7 Δ*E* for the approach of S in π–and n_o–direction of carbonyl functional groups

Δ*E* values for all the complexes are given in Tables 2.4 and 2.5. Figure 2.9 **A**-**E** represents a plot of $V_{S,max}$ of S vs. Δ*E* for these complexes, with S(CH_3)X approaching along the π– or n_o–direction. In general, the approach of S monomers along the π–direction towards the carbonyl motifs was preferred over the n_o–direction (Figure 2.9 **A**–**C**). For example, the Δ*E* for DMU:S$(CH_3)_2$ complexes in the π–direction and n_o–direction were –5.26 and –2.78 kcal.mol^{-1}, respectively (Table 2.4 and 2.5). A similar trend was noted for other nucleophiles. However, the difference in Δ*E* of π–and n_o–direction was found to decrease as the electron-withdrawing power of substituents increased (Figure 2.9 **A**–**C**). For example, the Δ*E* for DMU:S(CH_3)F dimers in π–direction and n_o–direction were –7.87 and –6.52 kcal.mol^{-1}, respectively (Table 2.4 and 2.5). This decrease in the difference in Δ*E* along the two directions was possible because of the cooperation between S···O=C and C–H···S interactions. The strength of S···O=C interaction increased for these substituents because of the increase in the magnitude of the $V_{S,max}$, contrastingly the strength of the C–H···S interaction decreased because of the decrease in the values for V_{min} of S (Figure 2.10).

In order to investigate the directional preferences for the approach of the S (Figure 2.7 **B**), plots of $V_{S,max}$ of S vs. Δ*E* of all complexes in π–direction (Figure 2.9 **D**) and n_o–direction (Figure 2.9 **E**) were examined. The correlation between Δ*E* and $V_{S,max}$ was independent of the direction of approach of S. Interestingly, Δ*E* for all complexes in π–direction increased in the order of DMU:S(CH_3)X > amide:S(CH_3)X > ketone:S(CH_3)X, while the values for complexes in n_o–direction increased in the order of DMU:S(CH_3)X ≈ amide:S(CH_3)X > ketone:S(CH_3)X (Figure 2.9 **D**-**E**).

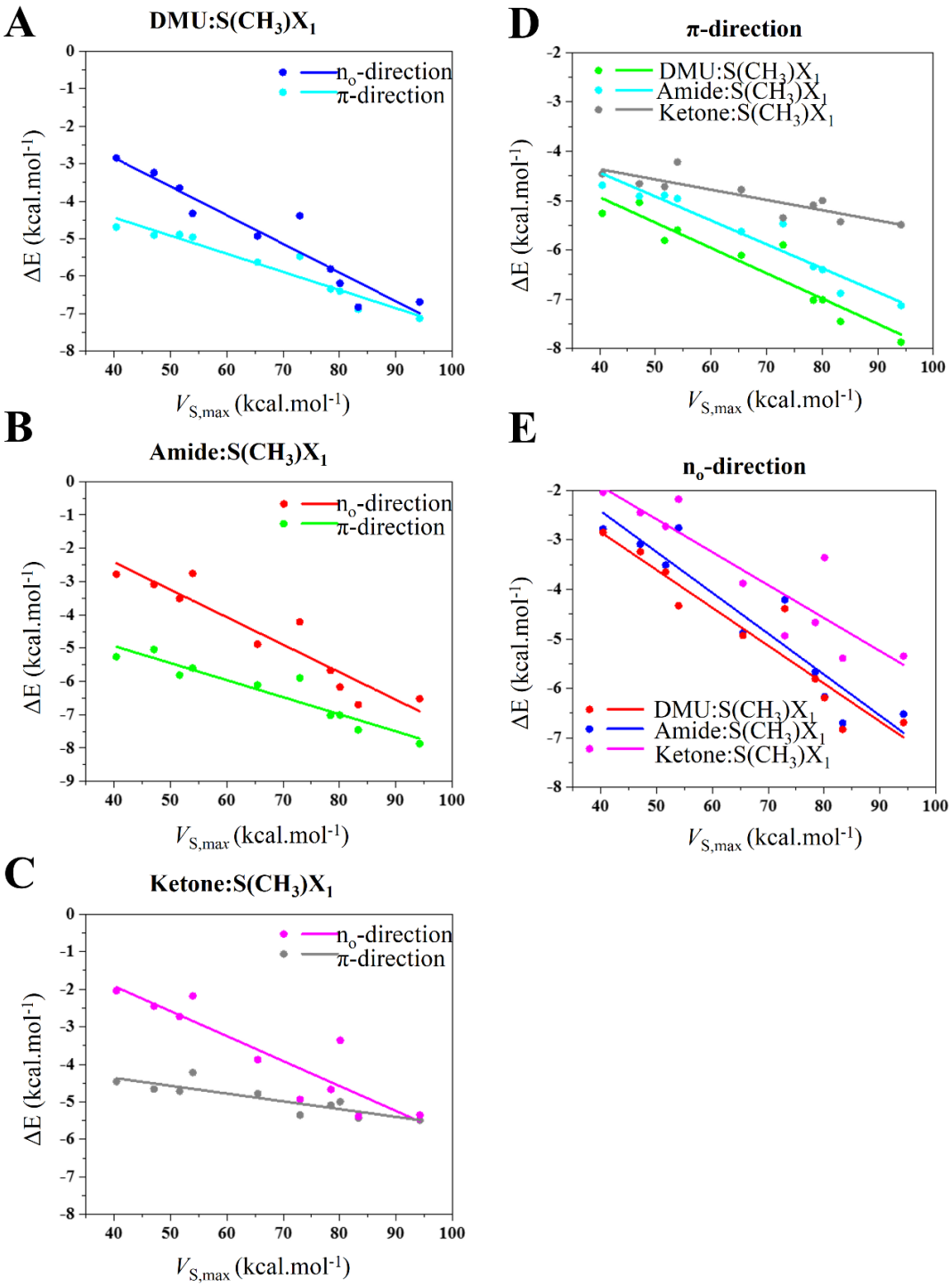

A
DMU:S(CH3)X1
no-direction
π-direction
ΔE (kcal.mol-1)
VS,max (kcal.mol-1)
B
Amide:S(CH3)X1
no-direction
π-direction
ΔE (kcal.mol-1)
VS,max (kcal.mol-1)
C
Ketone:S(CH3)X1
no-direction
π-direction
ΔE (kcal.mol-1)
VS,max (kcal.mol-1)
D
π-direction
DMU:S(CH3)X1
Amide:S(CH3)X1
Ketone:S(CH3)X1
ΔE (kcal.mol-1)
VS,max (kcal.mol-1)
E
no-direction
DMU:S(CH3)X1
Amide:S(CH3)X1
Ketone:S(CH3)X1
ΔE (kcal.mol-1)
VS,max (kcal.mol-1)

Figure 2.9 Plots of Δ*E* versus $V_{S,max}$ for dimers made of S monomers and carbonyl functional groups. Correlation of BSSE corrected complexation energies (Δ*E* in kcal mol^{-1}) with magnitude of σ–hole ($V_{S,max}$ in kcal.mol^{-1}) for (**A**) DMU:S(CH$_3$)X complexes in π–direction (r^2=0.87) and n_o–direction (r^2=0.86). (**B**) Amide:S(CH$_3$)X complexes in π–direction (r^2=0.91) and n_o–direction (r^2=0.89). (**C**) Ketone:S(CH$_3$)X complexes in π–direction (r^2=0.72) and n_o–direction (r^2=0.77). (**D**) DMU:S(CH$_3$)X (r^2=0.87), amide:S(CH$_3$)X (r^2=0.91) and ketone:S(CH$_3$)X complexes (r^2=0.72) in π–direction. (**E**) DMU: S(CH$_3$)X (r^2=0.86), amide: S(CH$_3$)X (r^2=0.89) and ketone:S(CH$_3$)X (r^2=0.77) complexes in n_o–direction. -X=CH$_3$, –H, –NH$_2$, –SCH$_3$, –CF$_3$, –OH, –CN, –NO$_2$, –Cl, and –F.

Strong substitution effects were observed for DMU:S(CH$_3$)X and amide:S(CH$_3$)X complexes in the π–direction (Figure 2.9 **D**). Δ*E* for DMU:S(CH$_3$)$_2$ and DMU:S(CH$_3$)F complexes were –5.26 and –7.86 kcal.mol^{-1}, while Δ*E* for amide:S(CH$_3$)$_2$ and amide:S(CH$_3$)F complexes were –4.69 and –7.13 kcal mol^{-1} (Table 2.4). However, the substitution on S did not show such a strong effect in the case of ketone:S(CH$_3$)X complexes. For ketone:S(CH$_3$)$_2$ and ketone:S(CH$_3$)F, Δ*E* was –4.46 and –5.49 kcal.mol^{-1}, respectively (Table 2.4). Consistent with our previous observations, the contribution from C–H···O=C interaction to the overall complexation stability did not change significantly with change in nucleophiles and substitutions. Consequently, the change in Δ*E* observed on varying the nucleophile was primarily due to the change in S···O=C interaction strength. In contrast, Δ*Es* for complexation of the S(CH$_3$)X monomers to DMU and amide were found to be similar in the n_o–direction and slightly lower for ketone:S(CH$_3$)X complexes (Figure 2.9 **E**).

In summary, the directional preference of approach of the divalent S towards a carbonyl group was

π–direction > n_o–direction (Figure 2.9 **A**-**C**).

The preference for the approach of the divalent S in π–direction varied as

DMU > Amide > Ketone (Figure 2.9 **D**),

While in n_o–direction, it varied as

DMU ≈ Amide > Ketone (Figure 2.9 **E**).

Thus, these results provide a molecular rationale for observing crystal structure analysis of organic and biomolecules that divalent S preferentially approach as a carbonyl group along the π–direction rather than the n_o–direction (Iwaoka et al., 2002b).

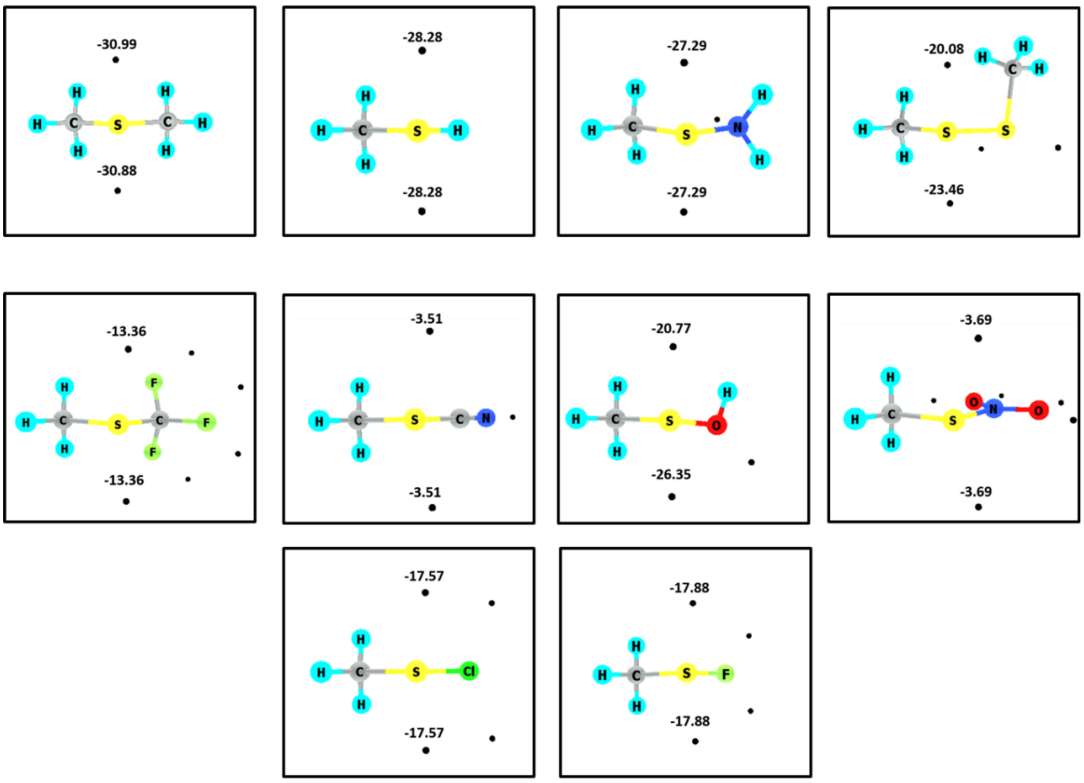


Figure 2.10 MESP minimum (V_{min}) values represent the lone-pair region of divalent sulfur monomers with different substituents. The V_{min} values are given in kcal.mol^{-1}.

2.3.8 The MESP basis for directional selectivity of S towards the carbonyl functional groups

To understand the basis of the selectivity discussed above, we first analyzed the contribution of the C–H···S interaction in the overall stabilization of all the complexes. For this, an MESP analysis of all the divalent S containing motifs was carried out, which revealed the presence of two V_{min} on the S, representing the lone pair regions (Figure 2.10). In the case of the S approaching along the π–direction, the two lone pairs of S pointed towards the H of the two –CH_3 moieties of the carbonyl group (Figures 2.7 **B** and 2.8). Contrastingly, for approach in n_o–direction, the lone pairs pointed away from the H (Figure 2.8 and 2.10). Hence, the strength of the C–H···S interaction formed along the π–direction was expected to be higher than the interaction along the n_o–direction.

In addition, the substituents attached to the divalent S primarily affected the V_{min} values of S (Figure 2.10). As the electron-withdrawing power of the substituents decreased, the V_{min} values of the two lone pairs of S became more negative (Figure 2.10). Thus, the strength of the C–H···S

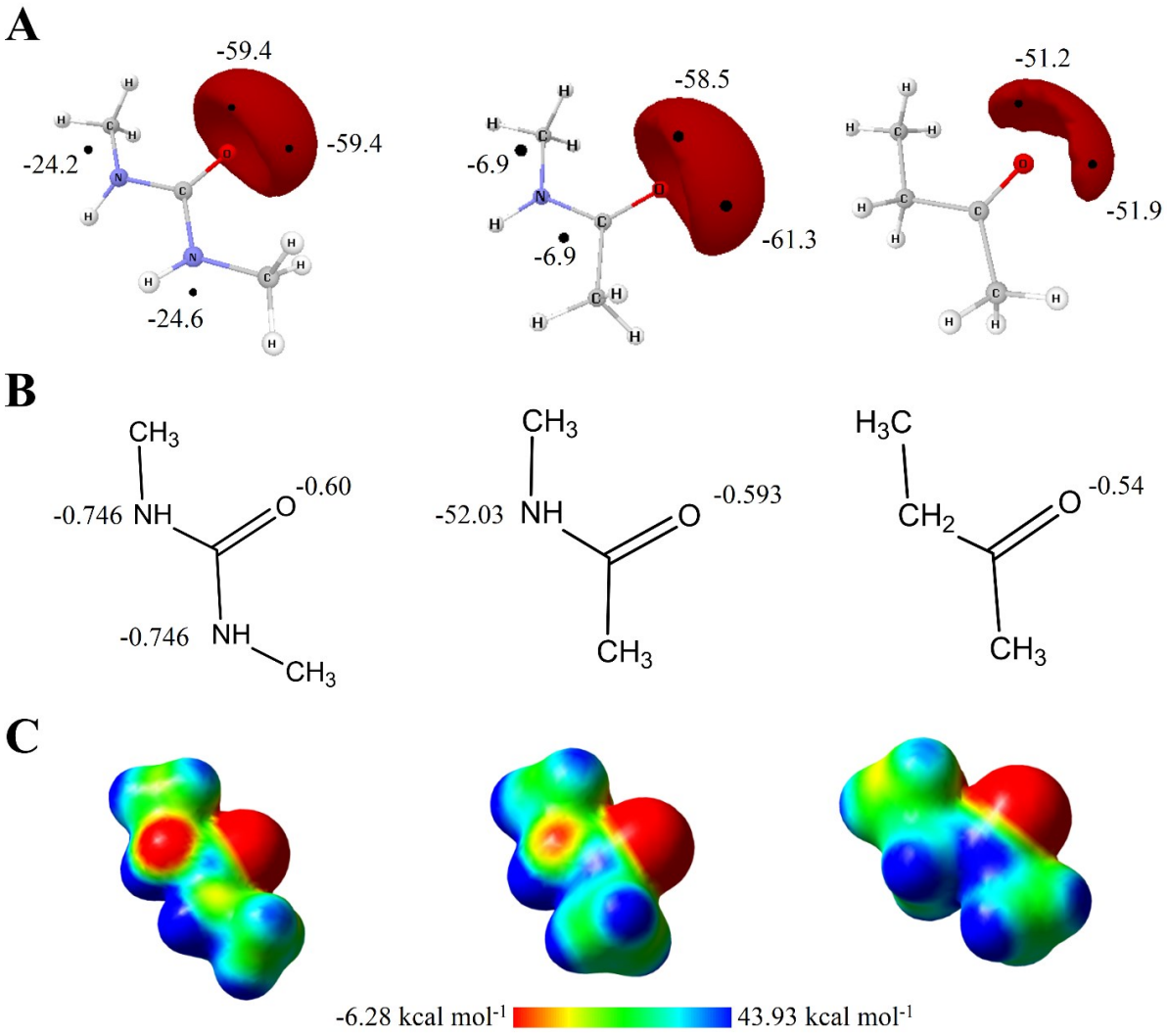


Figure 2.11 MESP analysis of carbonyl functional groups. (**A**) MESP isosurfaces (contour value = –0.065) and MESP minimum (V_{min} in kcal.mol^{-1}) for DMU, amide, and ketone represent the electron-dense region within the molecules. (**B**) MESP derived charges (in au) on N and O of DMU, amide, and ketone. (**C**) MESP maps shown for DMU, amide, and ketone with the color-coding range from –6.28 (red) to 43.93 kcal.mol^{-1} (Blue) and textured on 0.01 au density isosurface.

interaction was weaker in complexes having strong electron-withdrawing substituents (–F, –Cl, –NO_2, and –OH) compared to substituents such as –CH_3, –H, –NH_2, and –SCH_3.

Next, to investigate the contribution from S···O=C interaction, we studied the MESP topography features of DMU, ketone, and amide. MESP V_{min} values for these systems are shown in Figure 2.11 **A**. In addition, we calculated MESP derived charges and surface MESP maps for DMU, amide, and ketone motifs (Figure 2.11 **B** and 2.11 **C**). For DMU, values of V_{min} were –24.20 and –24.15 kcal.mol^{-1} in the π–region, and –59.36 kcal.mol^{-1} for each lone pair of O (in n_o–

region). Similarly, amide could be characterized by three V_{min}, one in π–region (–6.90 kcal.mol^{-1}) and two representing lone pair of O (–58.54, and –61.30 kcal.mol^{-1}). Two V_{min} characterized the ketone (–51.20 and –51.89 kcal mol^{-1}) in the lone pair region of O (Figure 2.11 **A**). In the case of all complexes along the π–direction, σ–hole of S interacted preferentially with the V_{min} located in the π–region (Figure 2.8 and 2.11 **A**). The value of V_{min} in the π–region of DMU was more negative than amide. Consequently, the overall ΔE for DMU:S(CH_3)X complexes was higher than that for amide:S(CH_3)X complexes (Figure 2.9 **D** and 2.11 **A**). As ketone does not have any MESP minima in the π–region, it provides lower stability than DMU and amide (Figure 2.9 **D**).

In case of n_o–direction, the strength of DMU:S(CH_3)X and amide:S(CH_3)X complexes were similar but higher than ketone: S(CH_3)X complexes (Figure 2.9 **E**). This is because the values of the two V_{min} on O in n_o–direction for DMU and amide were comparable but deeper than ketone. In complexes with –H, –CH_3, –SCH_3, and –NH_2 substituents, the distance between S and O ranged between 3.31 Å to 3.65 Å (Table 2.4), which is around or greater than the sum of van der Waals radii of S and O (3.32 Å). We suggest that the stability of these complexes was primarily due to the C–H···O interaction (Figure 2.8). For other substituents along n_o–direction, the distance between S and O ranged between 3.26 Å to 2.71 Å, indicating the presence of S···O=C interaction mediated by the σ–hole of S and one of the two V_{min} on O (Figure 2.8, Table 2.5).

In summary, the various interactions can be arranged based on their contribution to the stability of all complexes in the following two directions.

1) For approach in π–direction,

when, X= –CH_3, –H, –NH_2 and –SCH_3,

C–H···S interaction > S···O=C interaction.

When X= –F, –Cl, –NO_2, –OH, –CN and –CF_3,

S···O=C interaction > C–H···S interaction

The strength of the C–H···O=C interaction was independent of the substituents.

2) For approach along n_o –direction,

when X= –H, –CH_3, –SCH_3, and –NH_2,

C–H···O=C interaction > S···O=C interaction.

When X= –F, –Cl, –NO_2, –OH, –CN and –CF_3,

S···O=C interaction > C–H···O=C interaction

The strength of the C–H···S interaction would be poor because of the lone pair of S pointing away independent of the substituents.

2.4 Discussion

Different weak interactions can exist together in characteristic pairs and cooperate to strengthen each other with functional implications. For example, the presence of a cooperating cation–π interaction can strengthen a π–π stacking interaction by as much as 5 fold (Mahadevi and Sastry, 2016). Similarly, there are reports of the possible co-existence of chalcogen and hydrogen bonds (Adhikari and Scheiner, 2014, 2011; Burling and Goldstein, 1992; Esrafili and Mohammadian-Sabet, 2015; Guru Row and Parthasarathy, 1981; Iwaoka and Babe, 2015; Iwaoka and Isozumi, 2012; Rosenfield et al., 1977; Shukla and Chopra, 2015; Zhang et al., 2015). However, to the best of our knowledge, cooperativity between chalcogen bonds with other interactions has not been analyzed before. In addition, it is essential to ask if one can delineate the contributions of the individual coexisting interactions. Knowledge of the contributions of coexisting interactions under different conditions can significantly help to tune their molecular strength, which is of considerable importance in molecular engineering and design. Here in this study, in addition to characterizing the strength and directionality, we also delineated the contributions from coexisting S···O/N interaction and various hydrogen bonds.

In the initial part of the study, we showed that the strength of the S···O/N interaction depends on the nature of the substituent attached to S and interacting nucleophiles, similar to other σ-hole interactions (Politzer et al., 2013). Also, the strength could be comparable to the hydrogen bond formed by the water dimer (–5.0 kcal.mol^{-1}) (Feyereisen et al., 1996). We noted that weak C–H···O hydrogen bonds could also stabilize the complex with S···O/N interaction. Interestingly the strength of the C–H···O interaction was not affected by the nature of the substituent attached to divalent S and the type of the nucleophile. Thus, a direct correlation between interaction energy and strength of the σ-hole that S could form is evidence for the electrostatic nature of the Ch-bond, similar to the halogen bond (Politzer et al., 2013). The cooperating C–H···O/N and S···O/N should be considered as a signature pair of weak interactions. This pair could occur frequently and may have implications in molecular recognition processes in biology, supramolecular chemistry, and catalysis.

Similarly, we also investigated the cooperative facets among weak interactions, including S···O=C interaction, using various carbonyl groups as a model system and MESP. This analysis

revealed the contribution of the electrostatic component in deciding the preferential direction of the interaction observed earlier in crystal structure analysis (Iwaoka et al., 2002b). We showed that the directional preference for the approach of the divalent S towards the various carbonyl groups depends on the nature of the carbonyl groups and the other weak interactions. A robust directional selectivity for the approach of the divalent S towards the π–region of the carbonyl groups was observed. This was because as the electron density in the π–region increased, the strength of the S···O=C interaction also increased. Moreover, poor directional selectivity was seen when S approached the lone–pair region of the carbonyl group. Additionally, C–H···O, C–H···S, and S···O interaction cooperatively stabilized the dimers involving the carbonyl groups, and their contribution, except for the C–H···O interaction, was dependent on the nature of the substituents attached to divalent S. MESP topology of the carbonyl groups provided a rationale for this directional selectivity.

2.5 References

Adhikari, U., Scheiner, S., 2014. Effects of charge and substituent on the S···N chalcogen bond. J. Phys. Chem. A 118(17), 3183–3192.

Adhikari, U., Scheiner, S., 2011. The S···N noncovalent interaction: Comparison with hydrogen and halogen bonds. Chem. Phys. Lett. 118(17), 3183–3192.

Bader, R.F.W., 2002. Atoms in Molecules, Encyclopedia of Computational Chemistry. John Wiley and Sons, Ltd.

Bader, R.F.W., 1991. A Quantum Theory of Molecular Structure and Its Applications. Chem. Rev. 91(5), 893-928.

Bader, R.F.W., (1985). Atoms in Molecules. Acc. Chem. Res., 18, 9–15

Bauzá, A., Mooibroek, T.J., Frontera, A., 2015. The Bright Future of Unconventional σ/π-Hole Interactions. ChemPhysChem 16(12), 2496–2517.

Benz, S., López-Andarias, J., Mareda, J., Sakai, N., Matile, S., 2017. Catalysis with Chalcogen Bonds. Angew. Chemie - Int. Ed. 56(3), 812–815.

Biegler-König, F., Schönbohm, J., 2002. Update of the AIM2000-program for atoms in molecules. J. Comput. Chem. 23(15), 1489–1494.

Biot, N., Bonifazi, D., 2018. Programming Recognition Arrays through Double Chalcogen-Bonding Interactions. Chem. - A Eur. J. 24(21), 5439-5443.

Bondi, A., 1964. van der Waals Volumes and Radii. J. Phys. Chem. 68(3), 441–451.

Bone, R.G.A., Bader, R.F.W., 1996. Identifying and Analyzing Intermolecular Bonding Interactions in van der Waals Molecules. J. Phys. Chem. 100(26), 10892–10911.

Boys, S.F., Bernardi, F., 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys. 19(4), 553–566.

Brezgunova, M.E., Lieffrig, J., Aubert, E., Dahaoui, S., Fertey, P., Lebègue, S., Ángyán, J.G., Fourmigué, M., Espinosa, E., 2013. Chalcogen bonding: Experimental and theoretical determinations from electron density analysis. geometrical preferences driven by electrophilic-nucleophilic interactions. Cryst. Growth Des. 13(8), 3283–3289.

Burling, F.T., Goldstein, B.M., 1992. Computational Studies of Nonbonded Sulfur-Oxygen and Selenium-Oxygen Interactions in the Thiazole and Selenazole Nucleosides. J. Am. Chem. Soc. 114(7), 2313–2320.

Cavallo, G., Metrangolo, P., Milani, R., Pilati, T., Priimagi, A., Resnati, G., Terraneo, G., 2016. The halogen bond. Chem. Rev. 116(4), 2478–2601.

Deepa, P., Sedlak, R., Hobza, P., 2014. On the origin of the substantial stabilisation of the electron-donor 1,3-dithiole-2-thione-4-carboxyclic acid···I_2 and DABCO···I_2 complexes. Phys. Chem. Chem. Phys. 16(14), 6679–6686.

Dennington, R., Keith, T., Millam, J., 2009. GaussView, Version 5. Semichem Inc., Shawnee Mission. KS.

Desiraju, G.R., Ho, P.S., Kloo, L., Legon, A.C., Marquardt, R., Metrangolo, P., Politzer, P., Resnati, G., Rissanen, K., 2013. Definition of the halogen bond (IUPAC Recommendations 2013). Pure Appl. Chem. 85(8), 1711-1713.

DiLabio, G.A., Johnson, E.R., Otero-de-la-Roza, A., 2013. Performance of conventional and dispersion-corrected density-functional theory methods for hydrogen bonding interaction energies. Phys. Chem. Chem. Phys. 15(31), 12821.

Dube, J.W., Hänninen, M.M., Dutton, J.L., Tuononen, H.M., Ragogna, P.J., 2012. Homoleptic pnictogen-chalcogen coordination complexes. Inorg. Chem. 51(16), 8897–8903.

Dyduch, K., Mitoraj, M.P., Michalak, A., 2013. ETS-NOCV description of σ-hole bonding, Journal of Molecular Modeling. 19(7), 2747–2758.

Esrafili, M.D., Mohammadian-Sabet, F., 2015. Bifurcated chalcogen bonds: A theoretical study on the structure, strength and bonding properties. Chem. Phys. Lett. 634(12), 210–215.

Fernández Rico, J., López, R., Ema, I., Ramírez, G., 2004a. Electrostatic potentials and fields from density expansions of deformed atoms in molecules. J. Comput. Chem. 25(11), 1347–1354.

Fernández Rico, J., López, R., Ramírez, G., Ema, I., Ludena, E. V., 2004b. Analytical method for the representation of atoms-in-molecules densities. J. Comput. Chem. 25(11), 1355–1363.

Feyereisen, M.W., Feller, D., Dixon, D.A., 1996. Hydrogen Bond Energy of the Water Dimer. J. Phys. Chem. 100(8), 2993–2997.

Fick, R.J., Kroner, G.M., Nepal, B., Magnani, R., Horowitz, S., Houtz, R.L., Scheiner, S., Trievel, R.C., 2016. Sulfur-Oxygen Chalcogen Bonding Mediates AdoMet Recognition in the Lysine Methyltransferase SET7/9. ACS Chem. Biol. 11(3), 748–754.

Gadre, S.R., Pundlik, S.S., 1995. Topographical Analysis of Electron Density and Molecular Electrostatic Potential for Cyclopropa- and Cyclobutabenzenes. J. Am. Chem. Soc. 117(37), 9559–9563.

Garrett, G.E., Carrera, E.I., Seferos, D.S., Taylor, M.S., 2016. Anion recognition by a bidentate chalcogen bond donor. Chem. Commun. 52(64), 9881–9884.

Garrett, G.E., Gibson, G.L., Straus, R.N., Seferos, D.S., Taylor, M.S., 2015. Chalcogen bonding in solution: Interactions of benzotelluradiazoles with anionic and uncharged Lewis bases. J. Am. Chem. Soc. 137(12), 4126–4133.

Gleiter, R., Werz, D.B., Rausch, B.J., 2003. A world beyond hydrogen bonds? - Chalcogen-chalcogen interactions yielding tubular structures. Chem. - A Eur. J. 9(12), 2676–2683.

Goerigk, L., 2015. Treating London-Dispersion Effects with the Latest Minnesota Density Functionals: Problems and Possible Solutions. J. Phys. Chem. Lett. 6(19), 3891–3896.

Gopalakrishna, T.Y., Reddy, J.S., Anand, V.G., 2013. Antiaromatic supramolecules: F···S, F···Se, and F···π intermolecular interactions in 32 π expanded isophlorins. Angew. Chemie - Int. Ed. 52(6), 1763–1767.

Guru Row, T.N., Parthasarathy, R., 1981. Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur in Terms of Its Orbital Orientations. 2. S…S Interactions and Nonspherical Shape of Sulfur in Crystals. J. Am. Chem. Soc. 103(2), 477–479.

Hennemann, M., Murray, J.S., Politzer, P., Riley, K.E., Clark, T., 2012. Polarization-induced σ-holes and hydrogen bonding. J. Mol. Model. 18(6), 2461–2469.

Iwaoka, M., Babe, N., 2015. Mining and structural characterization of S···X chalcogen bonds in protein database. Phosphorus, Sulfur Silicon Relat. Elem. 190(8), 1257–1264.

Iwaoka, M., Isozumi, N., 2012. Hypervalent nonbonded interactions of a divalent sulfur atom. Implications in protein architecture and the functions. Molecules. 17(6), 7266–7283.

Iwaoka, M., Takemoto, S., Okada, M., Tomoda, S., 2002a. Weak Nonbonded S≥X (X = O, N, and S) Interactions in Proteins. Bull. Chem. Soc. Jpn. 75, 1611–1625.

Iwaoka, M., Takemoto, S., Tomoda, S., 2002b. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein engineering. J. Am. Chem. Soc. 124(36), 10613–10620.

König, F.B., Schönbohm , J., Bayles, D., 2001. AIM2000-a program to analyze and visualize atoms in molecules. J. Comput. Chem. 22(5), 545–559.

Kumar, A., Gadre, S.R., Mohan, N., Suresh, C.H., 2014. Lone pairs: An electrostatic viewpoint. J. Phys. Chem. A 118(2), 526–532.

Kumar, A., Yeole, S.D., Gadre, S.R., Lõpez, R., Rico, J.F., Ramírez, G., Ema, I., Zorrilla, D., 2015. DAMQT 2.1.0: A new version of the DAMQT package enabled with the topographical analysis of electron density and electrostatic potential in molecules. J. Comput. Chem. 36(31), 2350–2359.

Legon, A.C., 2010. The halogen bond: an interim perspective. Phys. Chem. Chem. Phys. 12(28), 7736.

Lim, J.Y.C., Beer, P.D., 2018. Sigma-Hole Interactions in Anion Recognition. Chem. 4(4), 731-783.

López, R., Rico, J.F., Ramírez, G., Ema, I., Zorrilla, D., 2015. DAMQT 2.0: A new version of the DAMQT package for the analysis of electron density in molecules. Comput. Phys. Commun. 192, 289–294.

López, R., Rico, J.F., Ramírez, G., Ema, I., Zorrilla, D., 2009. DAMQT: A package for the analysis of electron density in molecules. Comput. Phys. Commun. 180(9), 1654–1660.

Lundemba, A.S., Bibelayi, D.D., Wood, P.A., Pradon, J., Yav, Z.G., 2020. σ-Hole interactions in small-molecule compounds containing divalent sulfur groups R1-S-R2. Acta Crystallogr. B. Struct. Sci. Cryst. Eng. Mater. B76, 707-718.

M. J. Frisch, G., Trucks, 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.., 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, J. A., Jr.; Peralta, J.E.., Ogliaro, F.., Bearpark, M.., Heyd, J.J.., Brothers, E.., Kudin, K.N.., Staroverov, V. N.; Kobayashi, R.., Normand, J.; Raghavachari, K.., Rendell, A.., Burant, J.C.., Iyengar, S.S.., Tomasi, J.., Cossi, M.., Rega, N.., Millam, J.M.., Klene, M.., Knox, J.E.., Cross, J.B.., Bakken, V.., Adamo, C.., Jaramillo, J.., Gomperts, R.., Stratmann, R.E.., Yazyev, O.., Austin, A.J.., Cammi, R.., Pomelli, C.., Ochterski, J.W.., Martin, R.L.., Morokuma, K.., Zakrzewski, V.G.., Voth, G.A.., Salvador, P.., Dannenberg, J. J.; Dapprich, S.., Daniels, A.D.., Farkas, Ö.., Foresman, J.B.., Ortiz, J.V.., Cioslowski, J.., Fox, D.J., 2009. Gaussian 09, Revision E. 01; Gaussian. Gaussian, Inc. Wallingford, CT.

Mahadevi, A.S., Sastry, G.N., 2016. Cooperativity in Noncovalent Interactions. Chem. Rev. 116(5), 2775–2825.

Mahmudov, K.T., Kopylovich, M.N., Guedes Da Silva, M.F.C., Pombeiro, A.J.L., 2017. Chalcogen bonding in synthesis, catalysis and design of materials. Dalt. Trans. 46(31), 10121–10138.

Mardirossian, N., Head-Gordon, M., 2016. How Accurate Are the Minnesota Density Functionals for Noncovalent Interactions, Isomerization Energies, Thermochemistry, and Barrier Heights Involving Molecules Composed of Main-Group Elements? J. Chem. Theory Comput. 12(9), 4303–4325.

Metrangolo, P., Resnati, G., 2012. Enzyme mimics: Halogen and chalcogen team up. Nat. Chem. 4(6), 437–438.

Metrangolo, P., Resnati, G., 2008. Chemistry: Halogen versus hydrogen. Science. 321 (5891), 918–919.

Metrangolo, P., Resnati, G., 2001. Halogen bonding: A paradigm in supramolecular chemistry. Chem. - A Eur. J. 7(12), 2511–2519.

Mohajeri, A., Pakiari, A.H., Bagheri, N., 2009. Theoretical studies on the nature of bonding in σ-hole complexes. Chem. Phys. Lett. 467(4-6), 393–397.

Mohan, N., Suresh, C.H., Kumar, A., Gadre, S.R., 2013. Molecular electrostatics for probing lone pair–π interactions. Phys. Chem. Chem. Phys. 15(42), 18401.

Murray, J.S., Lane, P., Politzer, P., 2009. Expansion of the σ-hole concept, Journal of Molecular Modeling. 15(6), 723–729.

Murray, J.S., Paulsen, K., Politzer, P., 1994. Molecular surface electrostatic potentials in the analysis of non-hydrogen-bonding noncovalent interactions. Proc. Indian Acad. Sci. - Chem. Sci. 106(2), 267–275.

Murray, J.S., Politzer, P., 2011. The electrostatic potential: An overview. Wiley Interdiscip. Rev. Comput. Mol. Sci. 1(2), 153–163.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19(1), 115–128.

Pascoe, D.J., Ling, K.B., Cockroft, S.L., 2017. The Origin of Chalcogen-Bonding Interactions. J. Am. Chem. Soc. 139(42), 15160–15167.

Politzer, P., Laurence, P.R., Jayasuriya, K., 1985. Molecular electrostatic potentials: An effective tool for the elucidation of biochemical phenomena. Environ. Health Perspect. 61, 191–202.

Politzer, P., Murray, J.S., 2010. An overview of σ-hole bonding, an important and widely-occurring noncovalent interaction. Practical Aspects of Computational Chemistry. Springer, Dordrecht.

Politzer, P., Murray, J.S., 2002. The fundamental nature and role of the electrostatic potential in atoms and molecules. Theor. Chem. Acc. 108 (3), 134–142.

Politzer, P., Murray, J.S., Clark, T., 2013. Halogen bonding and other σ-hole interactions: A perspective. Phys. Chem. Chem. Phys. 15(27), 11178–11189.

Politzer, P., Murray, J.S., Concha, M.C., 2008. σ-hole bonding between like atoms; a fallacy of atomic charges. Journal of Molecular Modeling. 14, 659–665.

Politzer, P., Murray, J.S., Peralta-Inga, Z., 2001. Molecular surface electrostatic potentials in relation to noncovalent interactions in biological systems. International Journal of Quantum Chemistry. 85(6), 676-684

Reid, R.C., Yau, M.K., Singh, R., Lim, J., Fairlie, D.P., 2014. Stereoelectronic effects dictate molecular conformation and biological function of heterocyclic amides. J. Am. Chem. Soc. 136(34), 11914–11917.

Řezáč, J., Hobza, P., 2016. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. Chem. Rev. 116(9), 5038–5071.

Robinson, E.R.T., Walden, D.M., Fallan, C., Greenhalgh, M.D., Cheong, P.H.-Y., Smith, A.D., 2016. Non-bonding 1,5-S···O interactions govern chemo- and enantioselectivity in isothiourea-catalyzed annulations of benzazoles. Chem. Sci. 7(12), 6919–6927.

Rosenfield, R.E., Parthasarathy, R., Dunitz, J.D., 1977. Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur. 1. Electrophiles and Nucleophiles. J. Am. Chem. Soc. 99(14), 4860–4862.

Roy, D., Marianski, M., Maitra, N.T., Dannenberg, J.J., 2012. Comparison of some dispersion-corrected and traditional functionals with CCSD(T) and MP2 ab initio methods: Dispersion, induction, and basis set superposition error. J. Chem. Phys. 137(13), 134109.

Scheiner, S., 2011. On the properties of X ··· N noncovalent interactions for first-, second-, and third-row X atoms On the properties of X · · · N noncovalent interactions for first-, second-, and third-row X atoms. J. Chem. Phys. 134(16), 164313.

Shukla, R., Chopra, D., 2016. Understanding the effect of substitution on the formation of S..F chalcogen bond. J. Chem. Sci. 128, 1589–1596.

Shukla, R., Chopra, D., 2015. Exploring the Role of Substitution on the Formation of Se···O/N Noncovalent Bonds. J. Phys. Chem. B 119(47), 14857–14870.

Sjoberg, P., Politzer, P., 1990. Use of the electrostatic potential at the molecular surface to interpret and predict nucleophilic processes. J. Phys. Chem. 94(10), 3959–3961.

Solà, M., 2013. Forty years of Clar's aromatic π-sextet rule. Front. Chem. 1, 22.

Suresh, C.H., Gadre, S.R., 1999. Clar's aromatic sextet theory revisited via molecular electrostatic potential topography. J. Org. Chem. 64(7), 2505–2512.

Suresh, C.H., Koga, N., Gadre, S.R., 2000. Molecular electrostatic potential and electron density topography: structure and reactivity of (substituted arene)Cr(CO)3 complexes. Organometallics 19(16), 3008–3015.

Wang, H., Wang, W., Jin, W.J., 2016. σ-Hole Bond vs π-Hole Bond: A Comparison Based on Halogen Bond. Chem. Rev. 116(9), 5072–5104.

Wang, W., Ji, B., Zhang, Y., 2009. Chalcogen bond: A sister noncovalent bond to halogen bond. J. Phys. Chem. A 113(28), 8132–8135.

Werz, D.B., Gleiter, R., Rominger, F., 2002. Nanotube formation favored by chalcogen-chalcogen interactions. J. Am. Chem. Soc. 124(36), 10638–10639.

Zhang, X., Gong, Z., Li, J., Lu, T., 2015. Intermolecular Sulfur···Oxygen Interactions: Theoretical and Statistical Investigations. J. Chem. Inf. Model. 55(10), 2138–2153.

Zhang, Y., Wang, W., 2009. The bifurcate chalcogen bond: Some theoretical observations. J. Mol. Struct. THEOCHEM 916(1-3), 135–138.

Zhao, Y., Truhlar, D.G., 2008. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. Theor. Chem. Acc. 120(1-3), 215–241.

Chapter 3

Rules governing selectivity between sulfur mediated chalcogen versus hydrogen bond

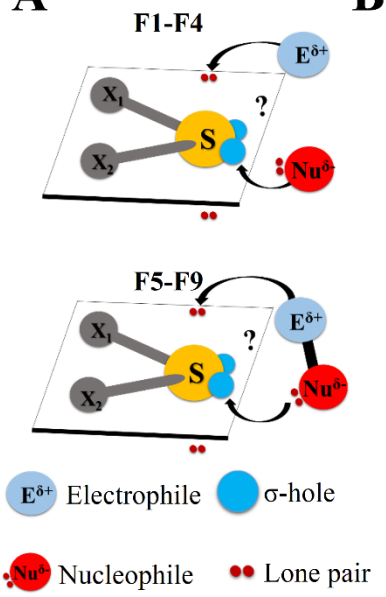
3.1 Introduction

Divalent sulfur (S) has the unique feature of participating in non-covalent interactions with nucleophiles and electrophiles (Rosenfield et al., 1977). Though these polar bonding properties of S have been known for long (Guru Row and Parthasarathy, 1981; Rosenfield et al., 1977), it has gained prominence in the field of organic chemistry over the last decade (Andersen et al., 2014; Beno et al., 2015; Motherwell et al., 2018; Pascoe et al., 2017; Scilabra et al., 2019; Wang et al., 2020). In contrast, these bonding properties of S in biomolecules are often overlooked. For example, many standard biochemistry textbooks categorized methionine and cysteine as non-polar hydrophobic amino acids, though they carry the divalent S (Berg et al., 2013; Boyle, 2005). At the same time, it is true that the side chain of methionine and cysteine are considerably hydrophobic (Kyte and Doolittle, 1982; Nagano et al., 1999). However, the lone pairs of S can participate in H-bond formation. Analyses of protein structures have revealed the prevalence of H-bond between acceptor S and donor O or N (Rao Mundlapati et al., 2015; Zhou et al., 2009). Similarly, S of a disulfide bond can interact with backbone C=O through n→π* interaction (Kilgore and Raines, 2018), highlighting the possible structural or functional roles of lone pair mediated interactions that S could form in proteins.

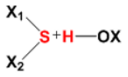
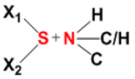
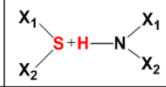
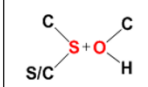
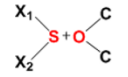
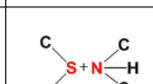
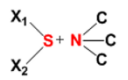
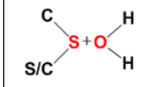
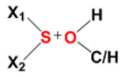
Additionally, the divalent S has two electropositive regions on the extension of its two covalent bonds, referred to as σ–holes, which can interact with various nucleophiles (Murray et al., 2012; Politzer et al., 2017, 2014). The interaction made by a σ–hole of S with a nucleophile is categorized as a chalcogen bond (Ch-bond) and found useful in self–assembly and catalysis of organic molecules (Benz et al., 2017; Chen et al., 2018; Garrett et al., 2015; Lim et al., 2017; Lim and Beer, 2018; Mahmudov et al., 2017). Similar to H-bonds, Ch-bonds are highly directional in nature. A recent spectroscopic study on thiophenes has shown that Ch-bond can be as strong as conventional H-bond (Pascoe et al., 2017). However, unlike the latter, the strength of Ch-bond is independent of solvent polarity (Pascoe et al., 2017). The occurrence of Ch-bond in proteins has been documented previously (Iwaoka et al., 2002; Iwaoka and Isozumi, 2012; Pal and Chakrabarti, 2001) and hypothesized to be functionally significant (Iwaoka et al., 2002; Iwaoka and Isozumi, 2012, 2006).

Many functional groups in proteins contain both electrophilic and nucleophilic centers with which S can interact, primarily through its lone pairs and σ–holes, respectively. Interestingly, the previous analysis using the crystal structures of small molecules showed a distinct approach of electrophiles and nucleophiles towards S. For example, the electrophiles such as metal cations, H, C, etc., approaches the S roughly perpendicular to the X_1-S-X_2 plane (Rosenfield et al., 1977),

A



B

	Fragment	N_T	N_c (%)		Fragment	N_T	N_c (%)
F1		7,854	360 (**4.7**)	F6		6,770	90 (**1.3**)
F2		7,803	494 (**6.3**)	F7*		N.D.	609 (Methionine) 154 (Cystine)
F3		11,552	367 (**3.2**)	F8*		N.D.	420 (Methionine) 79 (Cystine)
F4		7,537	21 (**0.3**)	F9*		N.D.	4,130 (Methionine) 624 (Cystine)
F5		7,950	516 (**6.5**)				

*Fragment is from protein structures obtained from PDB

Figure 3.1 Nature of divalent S-mediated interactions. (**A**) The approach of an electrophile or a nucleophile (upper panel) and a covalently linked electrophile-nucleophile fragment (lower panel) towards S. (**B**) Fragments analyzed using the CSD and the PDB with a summary of the analysis. N_T is the total number of independent pairs of fragments found in the CSD structures. N_C is the total number of S···O or S···H-O and S···N or S···H-N contacts found in the CSD and the PDB, having distance between S and the atom in red less than the sum of their van der Waals radii. X_1 and X_2 are any elements. Values in the parentheses are equal to N_C/N_T X 100 and represent the frequency of occurrence of the contacts mentioned above in the CSD.

where X_1 and X_2 are substituents covalently bonded to S, presumably to interact with the lone pairs (Figure 3.1 **A**). On the other hand, anions, N, O, etc., approaches the S on the extension of its covalent bonds to interact with the σ–hole (Rosenfield et al., 1977) (Figure 3.1 **A**). Hence, the direction of the approach of functional groups towards S and the nature of the bond formed are interlinked and could influence protein conformation. This prompted us to ask whether S forms a lone pair or σ–hole mediated interaction with a functional group having both electrophilic and nucleophilic centers (Figure 3.1 **A**) and what determines the choice between them.

Here, we have addressed these questions through extensive computational, cheminformatics, and bioinformatics analyses. Analysis of crystal structures in Cambridge Structural Database (CSD) (Groom and Allen, 2014) and the Protein Data Bank (PDB) (Rose et al., 2017) led us to deduce rules that determine the formation of either H-bond (lone pair mediated interaction) or Ch-bond (σ–hole mediated interaction) by S. Subsequently, Molecular Electrostatic

Potential (MESP) (Politzer et al., 2001, 1985; Politzer and Murray, 2002) analysis was used to obtain a molecular rationale for the choice of the non-covalent bond that S could form.

3.2 Methods

3.2.1 Computational details

All dimeric complexes *viz*. $(CH_3)_2S{:}OH_2$, $(CH_3)_2S{:}NH_3$, $Cl(CH_3)S{:}O(CH_3)_2$, and $Cl(CH_3)S{:}N(CH_3)_3$ were optimized employing the M06 functional (Zhao and Truhlar, 2008) and the 6-31++G(2D,2P) basis set using the Gaussian09 program (Frisch et al., 2009). In order to confirm that all these monomeric and dimeric optimized complexes are minima on the potential energy surface (PES), we performed Hessian evaluations with Int=ultrafine. It was confirmed that all the structures presented are minima on the PES from the positive eigenvalues of the Hessian. The atomic coordinates of the optimized complexes are provided in Appendix **B**. $(CH_3)_2S{:}OH_2$ and $(CH_3)_2S{:}NH_3$ complexes were used for PES scan to investigate energetically favorable region for S···H-O and S···H-N interactions, respectively. Similarly, PES scans were performed for $Cl(CH_3)S{:}O(CH_3)_2$ for S···O and $Cl(CH_3)S{:}N(CH_3)_3$ for S···N interactions. To carry out spherical energy scan using these complexes, d (distance between S and H/O/N) was kept constant, while θ (the angle between the centroid (c) of a triangle defined by C-S-X, S, and H/O/N), and δ (the torsion angle between C, c, S and H/O/N in degrees) were varied (Figure 3.2). Complexation energies (ΔEs) for different θ and δ values were calculated for all these complexes using the following equation (Řezáč and Hobza, 2016).

$$\Delta E^*_{AB} = E_{AB} - E_A - E_B$$

Here, E_A = Ground state energy of monomer A, E_B = Ground state energy of monomer B, E_{AB} = Total energy of complex AB and ΔE^*_{AB} = Complexation energy of complex AB (*without correction for basis set superposition error).

All the monomers mentioned in Figure 3.7 **A** and **B** were optimized using the M06 method (Zhao and Truhlar, 2008) at 6-311++G(3DF,3PD) basis set level. The atomic coordinates of the optimized monomers are provided in Appendix **B**. MESP topographical analyses characterizing the strength of lone pairs (V_{min}) of divalent S were carried out using the rapid topography mapping FORTRAN code developed by Gadre and co-workers and implemented in DAMQT (Kumar et al., 2015; Yeole et al., 2012). Texturing of MESP of molecules at defined density surface ($V_{S,max}$) was carried out using Gaussview 5.0 program (Dennington et al., 2009).

3.2.2 CSD analysis

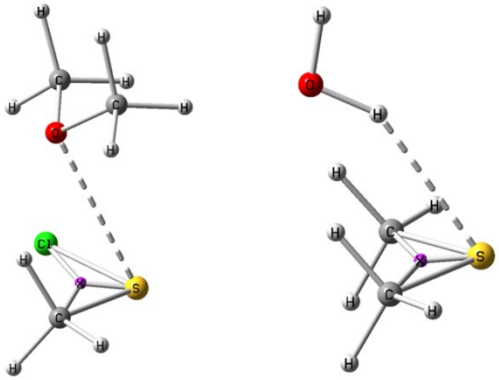


Figure 3.2 Definition of the PES scan parameters. The definition of θ and δ is as follow; *d* (Distance between S and H/O/N) = constant, θ (Angle between c (centroid of triangle made by C, S, and C/Cl atoms), S, and H/O/N) = 60˚ to 178˚ and δ (Dihedral angle between c, S, and H/O/N) = -90˚ to 90˚. *d* values for all the complexes (F1-F4) were obtained from energy minimized structures and kept constant during the PES scan. θ and δ were varied in steps of 2˚. For $Cl(CH_3)S{:}N(CH_3)_3$ complex, θ was varied from 90˚ to 178˚ instead 60˚ to 178˚ when $-40^\circ \leq \delta \leq 40^\circ$ to avoid steric clashes between atoms.

Fragments provided in Figure 3.1 **B** were queried in structural data retrieved from CSD (Groom and Allen, 2014) (version 5.39, Feb 2018) using ConQuest 1.21 (Bruno et al., 2002a). The following criteria were used for the search: **1**. 3D coordinates determined for all the atoms; **2.** structures with crystallographic *R factor* ≤ 10%; **3.** no disorder in crystallographic data; **4.** no error in 3D atomic co-ordinates; and **5.** no polymeric structures. Obtained data were further processed and analyzed using Mercury 3.10.1 (Bruno et al., 2002b). Using this, we searched for intermolecular S···O/N and S···H-O/N contacts, where $d_{S\cdots O} \leq 3.32$ Å, $d_{S\cdots N} \leq 3.35$ Å and $d_{S\cdots H} \leq$ 2.8 Å (Zhou et al., 2009). The approach of H or O/N towards S in space was investigated using the 3D parameters facility provided in ConQuest 1.21. To segregate H- form Ch-bonds based on their θ and δ values, we calculated the mean of these values in F1-F4 (refer to results section). The range of θ and δ for H- and Ch-bond was obtained by taking their mean ± standard deviation (1 sigma) of the calculated values were used as limits for distinguishing between H- and Ch-bonds. The values for the limits were rounded off to the closest value, which was a multiple of 5. The mean of δ values was calculated from their modulus. This angular range of θ and δ for H- and Ch-bond (see Results section for their values) was used throughout the study. All plots were generated using the OriginPro 9.0 program (Seifert, 2014).

3.2.3 PDB analyses

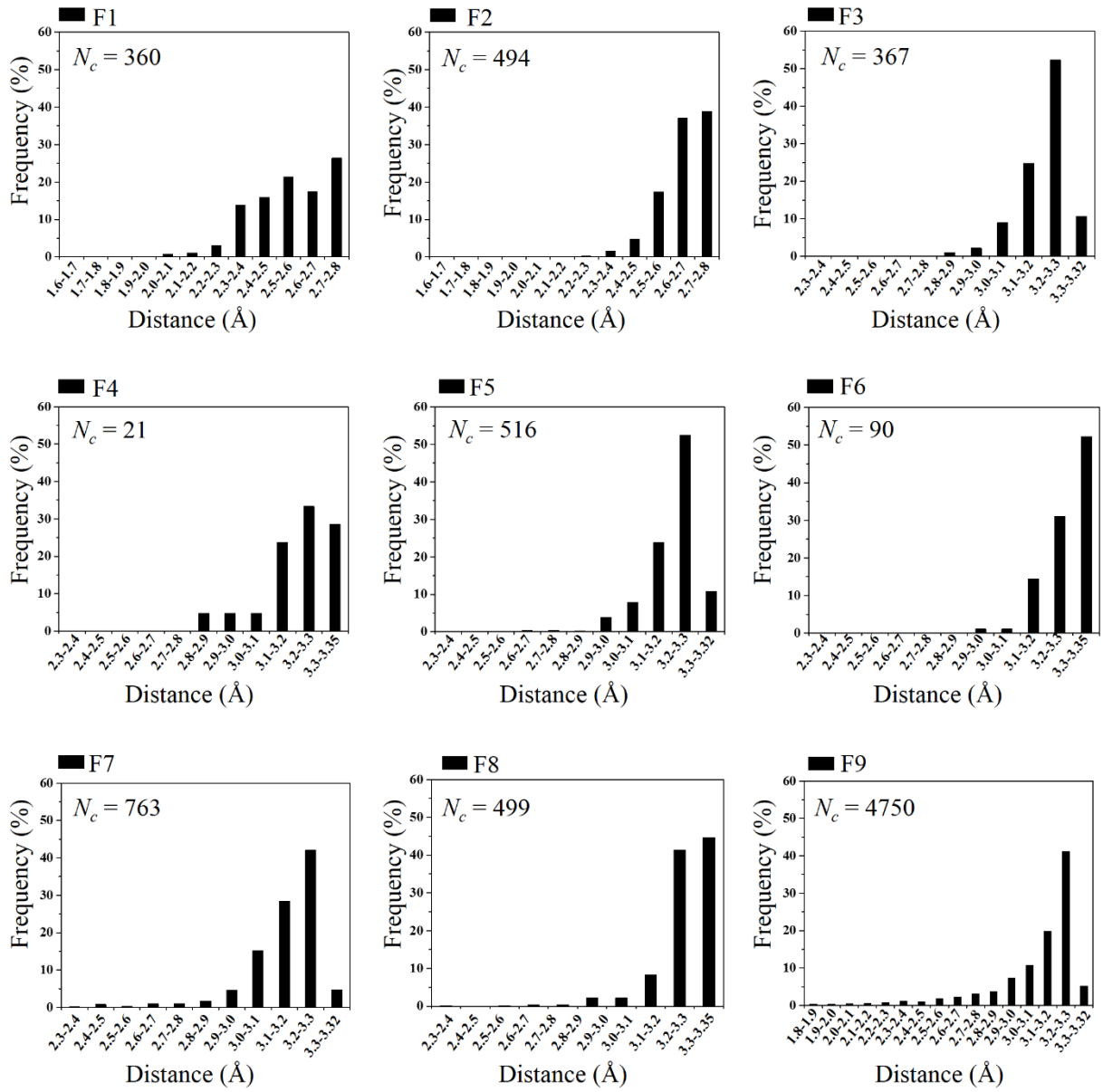


Figure 3.3 A distribution of *d* values. Histograms show the frequency (%) of the contacts mentioned in Figure 3.1 **B** with respect to their *d* values. For S···H-O/N contact, the distance cut of 2.8 Å was used**.** In the case of S···O and S···N contacts, the distance cut of 3.32 and 3.35 Å were used, respectively, which is equal to the sum of van der Waals radii of S and O or N.

Protein structures determined using X-ray crystallography in the PDB (Rose et al., 2017) were downloaded using PISCES (Wang and Dunbrack, 2005) in January 2018. Two sets of protein structures were generated using the following criteria. **Set 1** contained protein structures having pairwise sequence identity ≤ 90%, *resolution* ≤ 2.0 Å, and R_{free} ≤ 25%. **Set 2** contained protein structures having pairwise sequence identity ≤ 90%, *resolution* ≤ 2.5 Å, and R_{free} ≤ 30%. This resulted in **Set 1** having 16851 structures and **Set 2** having 25423 structures. These structures were

analyzed using in-house scripts written in Python 3.7.1 (Shelke, 2020). Search for H-bonds and Ch-bonds were made using the criterion of d (Å) ≤ less than equal to the sum of van der Waals radii of S and O/N. The criteria for direction obtained using the CSD analysis were used to distinguish between H- and Ch-bond. Figures of protein structures were made using the Chimera 1.13.1rc program (Pettersen et al., 2004). To search for metal-chelating cysteine, we used the distance between S and metal that could range within 1.9 to 2.8 Å. The distance range ensured that metals with different ionic radii were identified. All the analyses in this chapter were performed using the PDB from the Set 1 database.

3.3 Results

3.3.1 The geometry of S-mediated H-bond and Ch-bond

We selected the S···H-O/N (H-bond) and S···O/N (Ch-bond) interactions that S could form using its lone pair and σ–hole, respectively, as a representative to address the aforementioned questions. But, most structures in the PDB solved using X-ray crystallography do not have information on the position of H atoms, so it is not trivial to identify if the non-covalent bond between S and O/N is H- or Ch-bond. Hence, to overcome the ambiguity, we used the direction of approach of O or N towards S to distinguish between the two interactions. This is because an electrophile (H in this case) expects to approach S differently from a nucleophile (partially negatively charged O/N in this case) as the former interacts with the lone pair, in contrast to the latter which interacts with the σ-hole (Figure 3.1 **A**). To identify the preferred direction of approach of the two groups, we analyzed high-resolution crystal structures of organic molecules in the CSD (with experimentally determined hydrogen positions available) for the insight of the geometry of interactions made by electrophiles (F1 and F2) or nucleophiles (F3 and F4) with S (Figure 3.1 **B**). Our choice of fragments throughout this study was motivated by their relevance to proteins and their ligands.

Fragments referred to as F1 and F2 were investigated to study the preferred geometry of H-bonds, and F3 and F4 were studied to understand the geometry of Ch-bonds in the CSD. We used distance cut off such as $d_{S\cdots H} \leq 2.8$ Å (Zhou et al., 2009), $d_{S\cdots O} \leq 3.32$ Å or $d_{S\cdots N} \leq 3.35$ Å, to ensure the potential H- and Ch-bond. Note that these distances cut off are equal to the sum of the van der Waals radii of the interacting atoms. The analysis summary is presented in Figure 3.1 **B.** The distribution of the d values indicates the significant number of S···H or S···O/N contacts in the analysis that are shorter than their sum of van der Waals radii (Figure 3.3). We next employed θ and δ to obtain the preferred angular distribution for H- and Ch-bonds formed by S (Figure 3.4 **A**). The use of these geometric parameters also facilitated a direct comparison of the angular distribution with the complexation energy (ΔE, see Methods for the definition) obtained

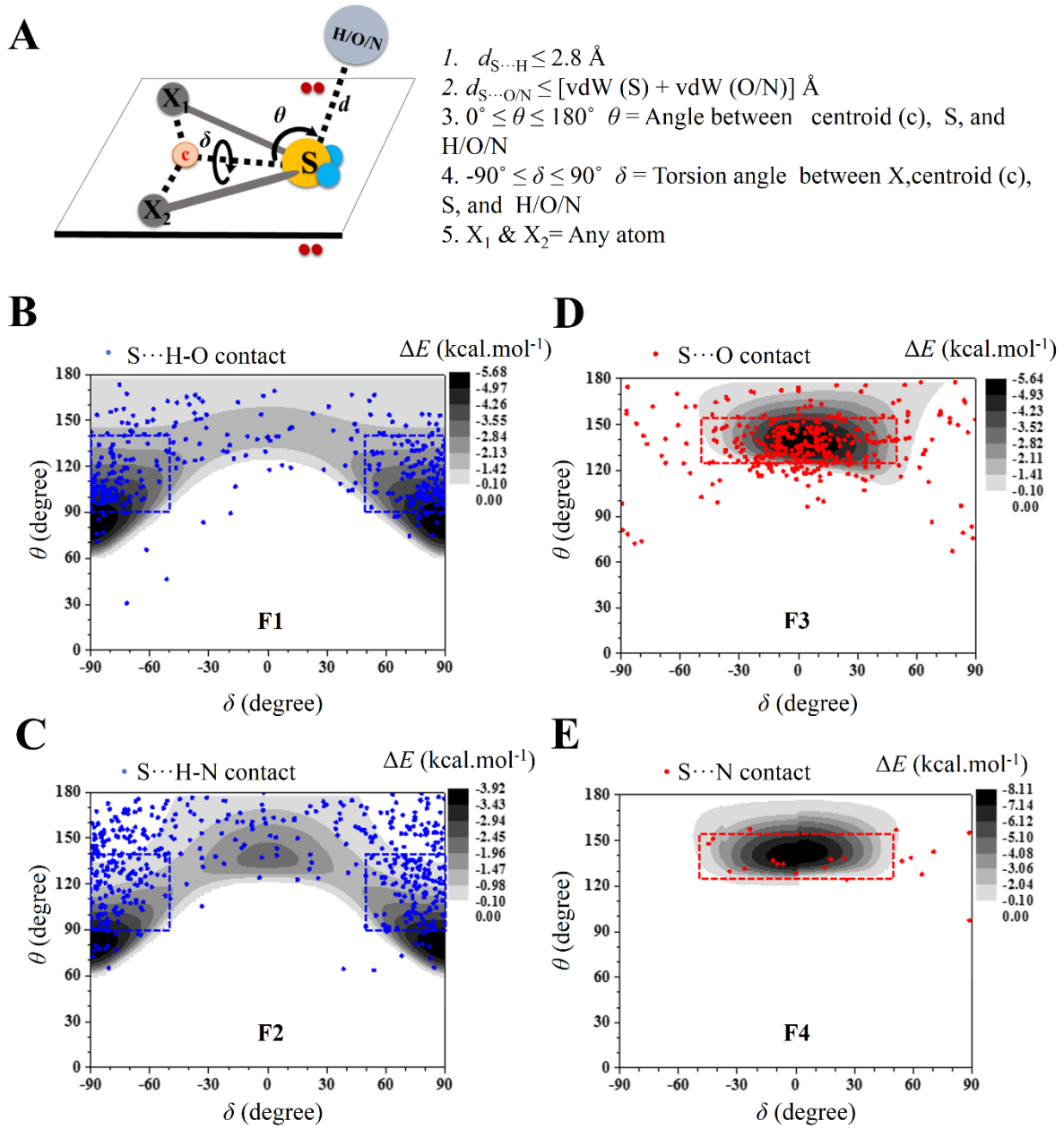


Figure 3.4 Definition of angular parameters and their mapping to distinguish H- from Ch-bond. (A) Definition of geometrical parameters *d*, θ and δ. The angular parameters θ and δ were used to investigate the approach of H and N/O towards divalent S. The mapping of θ and δ values for (**B**) S···H-O and (**C**) S···H-N contacts (blue dots) in F1 and F2, respectively, with computationally calculated Δ*E*s in the background (grayscale). $(CH_3)_2S{:}OH_2$ and $(CH_3)_2S{:}NH_3$ complexes were used as a model system to calculate Δ*E*s for F1 and F2, respectively. (**D**) S···O contacts (red dots) in F3 and (**E**). S···N contacts (red dots) in F4, $Cl(CH_3)S{:}O(CH_3)_2$ and $Cl(CH_3)S{:}N(CH_3)_3$ complexes were used as the model system to calculate Δ*E*s in the background (grayscale), respectively.

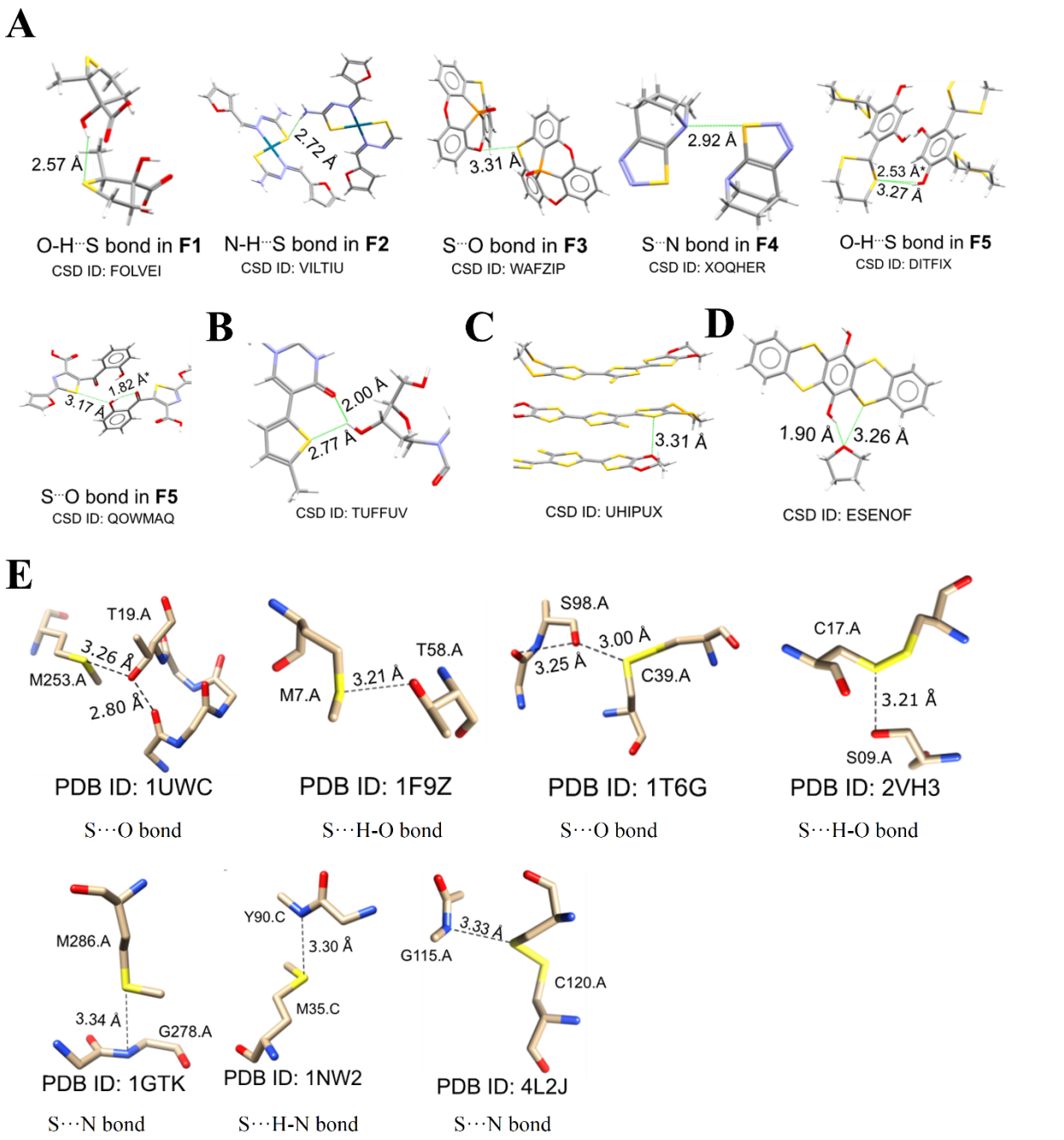


Figure 3.5 A representative examples from CSD and PDB analysis. (**A**) Representative examples of H- and Ch-bond in F1-F5 with their CSD ID. In the case of F5, $d_{S\cdots H}$ is marked by an asterisk. Note that H-O/N groups that formed Ch-bond with S in F5 could form H-bond with a neighboring acceptor atom (**B**) Representative example of outliers for S···H-O contacts in F1 (**C**) Representative example of outliers of S···O contacts in F3 from those clustered around at $\theta = 75^{\circ}$ and $\delta = 90^{\circ}$ or -90° in Figure 3.4 **D** and (**D**) Representative examples of outliers of S···O contacts in F3. All examples are shown with their CSD ID. (**E**) Representative examples of methionine- or cystine-mediated Ch-bond and H-bond in F7-F8. All structures are shown with their PDB IDs.

from Potential Energy Surface (PES) scans for different values of θ and δ. The calculations were carried out using density functional at M06 (Zhao and Truhlar, 2008) and the 6-31G++G(2D,2P) basis set using the Gaussian09 program (Frisch et al., 2009). Values of interaction energy (ΔE) from the PES scan for the model systems (refer to Methods) were calculated and mapped onto the plot (Figure 3.4 **B-E**).

Two distinct clusters were observed in the θ-δ plot for S···H-O contacts from (**i**) $90° \leq \theta \leq 140°$ and $-90° \leq \delta < -50°$and (**ii**) $90° \leq \theta \leq 140°$ and $50° < \delta \leq 90°$; which matched well with the location of the PES scan minima (Figure 3.4 **B-C** and Figure 3.5 **A**). The clusters represented the preferred direction, which was indicative of the direction of approach of the electrophile towards the lone pairs of S. In the case of S···O contacts, a single cluster was observed at a different region of the θ-δ plot ($115° \leq \theta \leq 155°$ and $-50° \leq \delta \leq 50°$), which correlated well with the PES scan minimum (Figure 3.4 **C** and Figure 3.5 **A**). The direction corresponds to the approach of the nucleophile towards the σ-hole of S (Aakeroy et al., 2019; Politzer et al., 2013). Outliers in the plots were due to the presence of other strong interactions, such as other H-bonds and stacking interactions, within the molecules (Figure 3.5 **B**). The number of S···N interactions (F4 in Figure 3.1 **B**) was much less than S···O in the CSD (F3 in Figure 3.1 **B**), possibly because of N being conjugated in most of the structures resulting in the lack of lone pair electrons for the formation of Ch-bond.

3.3.2 Geometrical features that distinguish S-mediated H-bond from Ch-bond in small molecules

Next, we sought to find the nature of bonding between S and a functional group having both electrophilic and nucleophilic centers (Figure 3.1 **A**, bottom panel) using the directional criteria for H- and Ch-bonds obtained above. We studied fragments F5 and F6 in which the electrophile and the nucleophile were separated by a single covalent bond (Figure 3.1 **B**) and could form either S···H-O/N or S···O/N interaction. Fragments with electrophiles and the nucleophiles separated by a single covalent bond were studied based on the assumption that either one of them, but not both, would interact with S due to structural constraints. A distance less than the sum of their van der Waals radii between S and O/N ensured that the interacting groups could form either a H- or Ch-bond. From this set of interactions, contacts satisfying $d_{S\cdots H} \leq 2.8$ Å were assigned as H-bond and the rest as Ch-bond. Note that the use of this filtering strategy excluded those H-bonds ($d_{S\cdots H} \leq 2.8$ Å) having the distance between S and O/N greater than the sum of their van der Waals radii.

The CSD analysis revealed three clusters in the θ-δ plot (Figure 3.6 **A**). Two of these clusters matched with those seen for the H-bond in F1 and F2 (Figure 3.4 **B** and **C**). Most

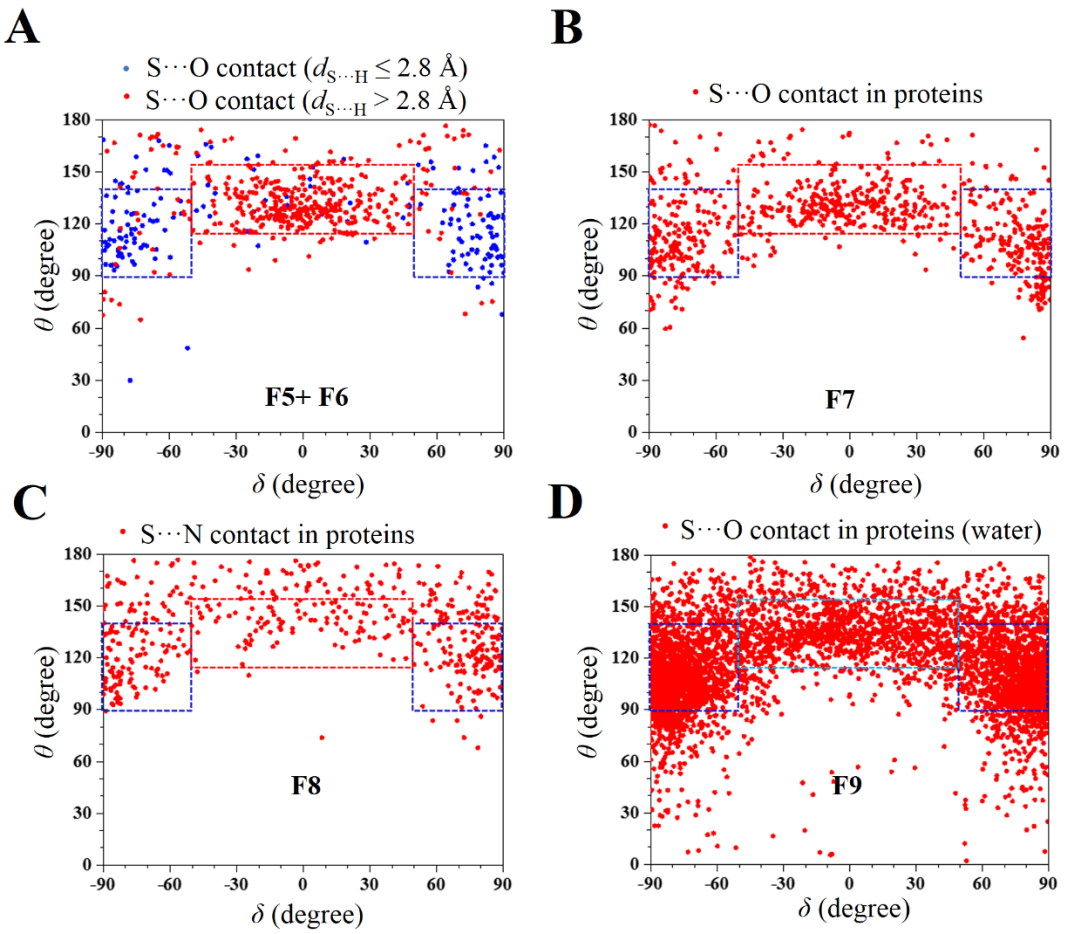


Figure 3.6 Mapping of θ and δ values to distinguish H- from Ch-bond in proteins. (A) The mapping of θ and δ values for contacts in F5 and F6. The contacts, where $d_{S\cdots H} \leq 2.8$ Å, are labeled as blue dots and rest a red dots ($d_{S\cdots H} > 2.8$ Å). (**B**) S···O contacts formed by methionine and cystine in F7 (**C**) S···N contacts formed by methionine and cystine in F8. (**D**) S···O contacts formed by methionine and cystine with water molecules in F9.

interactions in the two clusters satisfied the H-bond criterion of $d_{S\cdots H} \leq 2.8$ Å (blue dots in Figure 3.6 **A**). S···O/N interactions with $d_{S\cdots H} > 2.8$ Å (red dots in Figure 3.6 **A**) primarily clustered at a region that matched with the cluster formed by F3 and F4 forming Ch-bond (Figure 3.4 **D** and **E**). A few interactions in this cluster had $d_{S\cdots H} \leq 2.8$ Å. Note that H-O/N groups that formed Ch-bond with S could form H-bond with a neighboring acceptor atom (Figure 3.5 **A**).

3.3.3 Geometrical features that distinguish S-mediated H-bond from Ch-bond in proteins

In proteins, interactions of methionine or cystine (Disulfide Bridge) with side chains of serine, threonine, and tyrosine or the backbone amide or water (fragments F7-F9 in Figure 3.1 **B**) are equivalent to fragments F5 and F6 and could form either H- or Ch-bond. Due to the absence of coordinates of H in most crystal structures in the PDB, we used the angular limits for θ and δ

obtained by analyzing F1-F4 to identify S-mediated H- and Ch-bonds in proteins. For this reason and because there were very few examples of X_1-S-H in the CSD analysis discussed above, the thiol group of cysteines was not included in this or any other analyses that follow. All the protein structures analyzed here and later (unless mentioned otherwise) had a *resolution* ≤ 2 Å, R_{free} ≤ 25%, and pairwise sequence identity ≤ 90% amongst them. Contacts were identified using the distance criteria of $d_{S\cdots O} \leq 3.32$ Å and $d_{S\cdots N} \leq 3.35$ Å. To minimize the effect of structural constraints on the direction parameters, we excluded those fragments where the S and O/N were separated by less than 7 covalent bonds. As in the case of F5 and F6 in the CSD, the θ-δ plot for F7-F9 showed segregation of angular values into three clusters that corresponded to either H-bond (blue boxes in Figure 3.6 **B-D**) or Ch-bond (red box in Figure 3.6 **B-D**). The representative examples for these interactions are presented in Figure 3.5 **E**.

3.3.4 The electronic environment of S determines the formation of H- and Ch-bond

Our analysis in the previous section provided a geometric parameter to identify the nature of the interaction between S and a group containing a nucleophile and an electrophile separated by a covalent bond. We next sought to find what dictated the choice of the bond formed. In general, the formation of H- or Ch-bond is observed to depend on the strength of lone pairs and σ–holes on S, respectively (Adhikari and Scheiner, 2014; Kumar et al., 2014). The magnitude of the lone pairs and σ–holes are affected by the chemical properties of the substituent covalently bonded to S (Adhikari and Scheiner, 2014; Kumar et al., 2014). Consequently, we studied the strength of lone pairs and σ–holes of S in certain model systems relevant to biomolecules using MESP (Figure 3.7 **A** and **B**). As expected, the MESP analysis revealed the presence of two V_{min} (MESP minimum) on S in all the model systems, which correspond to the lone pairs, and whose values were substituent dependent (Figure 3.7 **A**). The electrostatic potential maps also showed the presence of two positive σ–holes along the extension of the S–X bonds except in the case of $[Fe(SCH_3)_4]$ (Figure 3.7 **B**). The strength of the σ–hole ($V_{S,max}$) increased with the electron-withdrawing power of the substituent. This confirmed that substituents on S modulated the electronic effects of the lone pairs and σ–holes, which, consequently, was expected to affect the nature of the bond formed.

Next, we categorized all contacts listed in Figure 3.1 **B** in the CSD based on the substituents linked to S, i.e., S(Ar) = S in aromatic ring; M-S-M, M = any metal; M-S-Y, Y = any element except M; E-S-Y, E = any electron-withdrawing group; R-S-R, R = saturated C, H or S, and checked if S formed H-bond or Ch-bond (Figure 3.7 **C** and Table 3.1). The classification into H-bond or Ch-bond was based on the distance and angular (θ and δ) criteria defined above. 87%

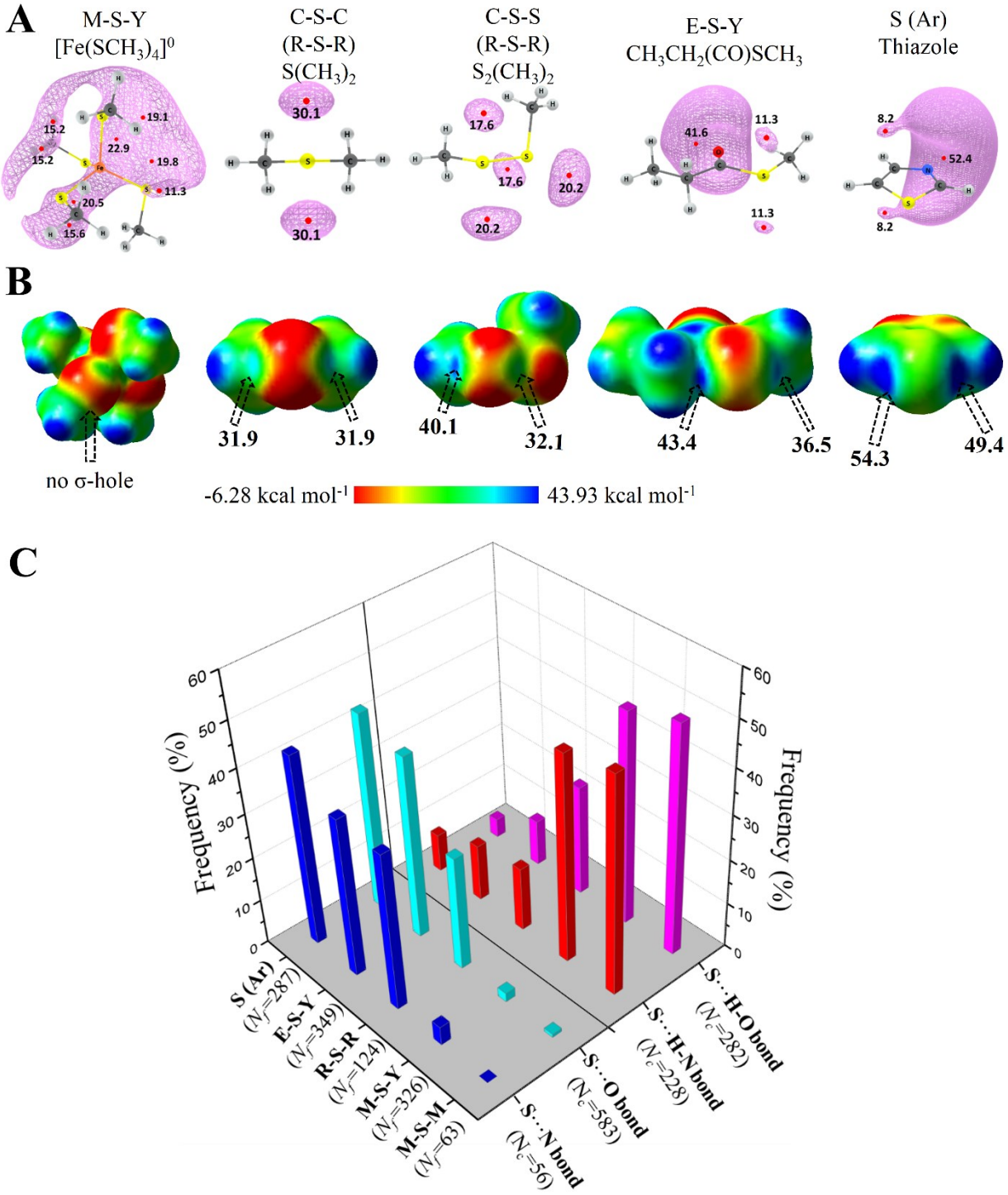


Figure 3.7 Rules for forming H- and Ch-bonds by S from CSD analysis. MESP analysis showing **(A)** MESP minimum values, V_{min} in kcal.mol^{-1} represents the lone pair of the S-containing monomers used in this study and (**B**) the two σ–holes on S in these monomers marked by arrows with their magnitude, $V_{S,max}$ in kcal.mol^{-1}. (**C**) The histogram shows Ch-bond and H-bond frequency formation in the CSD with the different electronic environments of S. M = any metal; Y = any element except M; R = saturated C, H, and S; E = any electron-withdrawing group; and S(Ar)= Aromatic S.

Table 3.1 Classification of the CSD data based on the nature of S. S-mediated H-bonds and Ch-bonds were identified using the distance (d) and angular (θ and δ) criteria defined in the text.

Fragment	S···H-O contacts (N_c)	S···H-N contacts (N_c)	S···O contacts (N_c)	S···N contacts (N_c)	Total Fragments (N_f)
M-S-M	35	27	1	0	63
M-S-Y	172	136	15	3	326
R-S-R	32	15	68	9	124
E-S-Y	32	32	263	22	349
S (Ar)	11	18	236	22	287
Total	282	228	583	56	1149

of S(Ar) and 77% of E-S-Y formed Ch-bond (Figure 3.7 **C** and Table 3.1). In sharp contrast, more than 95% of M-S-M formed H-bond (Figure 3.7 **C** and Table 3.1). In comparison, saturated C/S/H substituents (R-S-R) appeared to have a lesser influence on the choice of the bond formed. The number of Ch-bond (62%) was only slightly higher than H-bond (38%) (Figure 3.7 **C** and Table 3.1). The observations matched the expectations from the MESP analysis performed on representative molecules (Figure 3.7 **A** and **B**). For example, the high occurrence of Ch-bond in S(Ar) and E-S-Y is correlated with their high $V_{S,max,}$ and low V_{min} values from the MESP calculations. Absence of a σ–hole in the MESP map of M-S-Y matched with the higher number of H-bond by metal-chelated S. The above analysis thus provided a set of rules to predict the nature of the bond formed by S based on its electronic environment. We conclude that if S is part of an aromatic ring or is bonded to an electron-withdrawing group, it is most likely to form a Ch-bond, while H-bond is preferentially formed when S is coordinated with a metal.

3.3.5 S in disulfide linkages and aromatic ligands preferentially form Ch-bond

We analyzed protein structures in the PDB to find if these rules hold true for biomolecules. We focused on interactions made by S of constitutive methionine and cystine residues with the hydroxyl, amino or carbonyl/carboxyl group of backbone amide, side chains of serine, threonine, tyrosine, arginine, histidine, lysine, asparagine, glutamine, tryptophan, aspartate, and glutamate, and bound water (Figure 3.8-3.9 and Table 3.2). Our analysis revealed that the disulfide-linked Cys-S was more frequently involved in Ch-bond (87%) than H-bond (13%) (Figure 3.10 and Table 3.3). In comparison, Met-S appeared to form H-bond (59%) only marginally more than Ch-bond (41%) (Figure 3.10 and Table 3.3). The MESP analysis showed that S bonded to two methyl groups (C-S-C), as in methionine, had comparable values of V_{min} and $V_{S,max}$ (Figure 3.7 **A**-**B**), while $V_{S,max}$ on a disulfide-linked S (C-S-S) was larger than V_{min} (Figure 3.7 **A**-**B**), thus providing a rationale for cystine to form Ch-bond preferentially. Additionally, we also analyzed the interaction of aromatic S, often part of ligands or drug molecules, with the above functional groups in proteins.

Table 3.2 A summary of the number of interactions seen in PDB. H-bond and Ch-bond were identified using the distance (d) and angular (θ and δ) criteria defined in the text.

Interaction	Interacting residues	N_c	
		Resolution ≤ 2.0 Å and R_{free} ≤ 25%	*Resolution* ≤ 2.5 Å and R_{free} ≤ 30%
S···O contact	S(Ar) and Peptide backbone	51	92
	S(Ar) and Glu/Gln/Asn/Asp	20	31
	S(Ar) and Ser/Thr/Tyr	16	30
	S(Ar) and H_2O	63	114
	Cystine and Peptide backbone	1179	2041
	Cystine and Glu/Gln/Asn/Asp	140	265
	Cystine and Ser/Thr/Tyr	163	310
	Cystine and H_2O	671	1047
	Methionine and Peptide backbone	1059	2247
	Methionine and Glu/Gln/Asn/Asp	407	731
	Methionine and Ser/Thr/Tyr	652	1170
	Methionine and H_2O	4338	6221
	Metal chelated cysteine and Peptide backbone	38	67
	Metal chelated cysteine and Glu/Gln/Asn/Asp	82	138
	Metal chelated cysteine and Ser/Thr/Tyr	222	341
	Metal chelated cysteine and H_2O	1036	1377
S···N contact	S(Ar) and Peptide backbone	8	18
	S(Ar) and Arg/His/Lys	10	18
	S(Ar) and Trp/Asn/Gln	6	9
	cystine and Peptide backbone	95	229
	cystine and Arg/His/Lys	90	190
	cystine and Trp/Asn/Gln	55	107
	Methionine and Peptide backbone	476	965
	Methionine and Arg/His/Lys	565	1079
	Methionine and Trp/Asn/Gln	389	757
	Metal chelated cysteine and Peptide backbone	1153	2098
	Metal chelated cysteine and Arg/His/Lys	433	725
	Metal chelated cysteine and Trp/Asn/Gln	63	94

*Pairwise sequence identity ≤ 90%. N_c = number of contact.

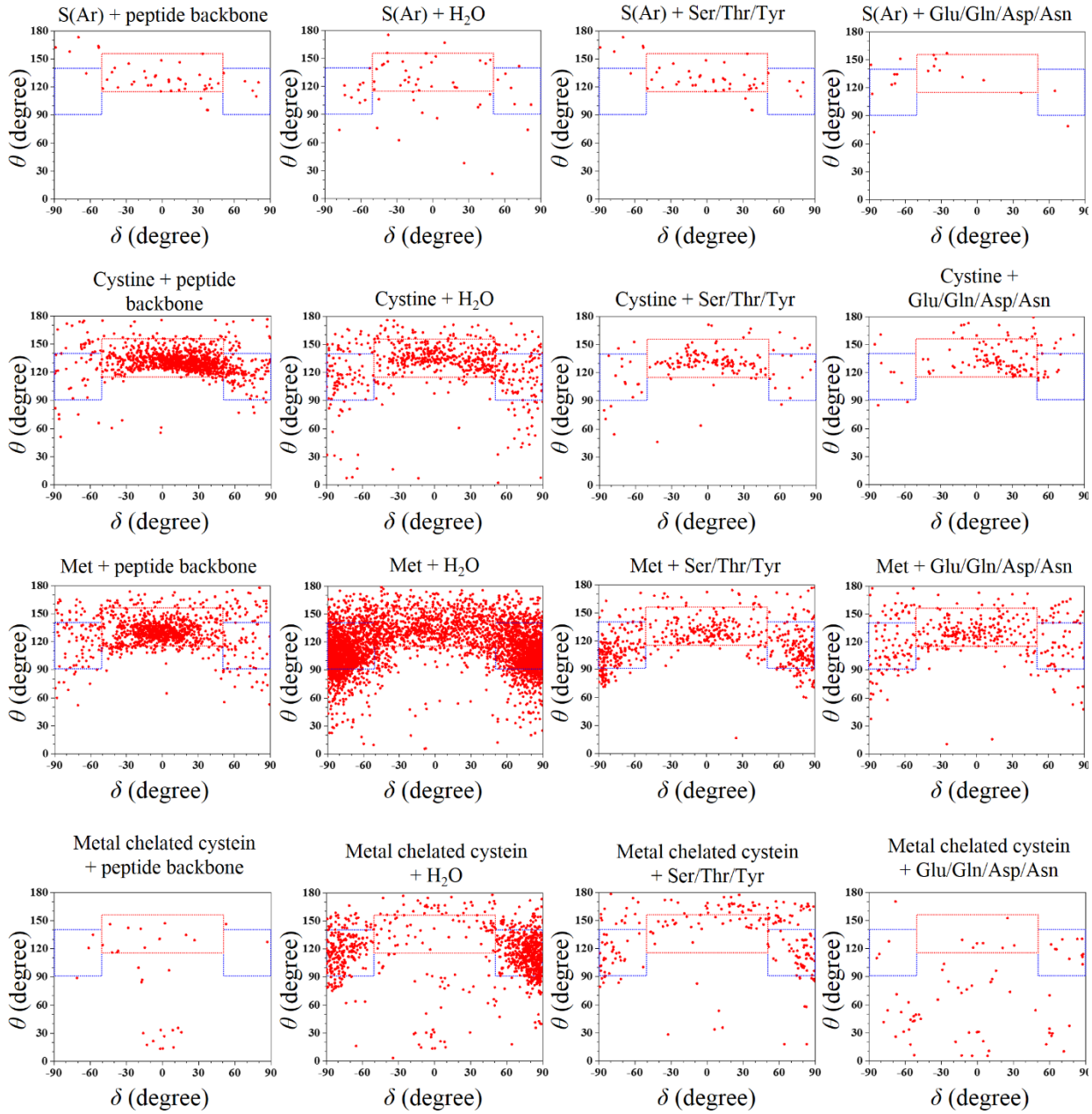


Figure 3.8 *θ-δ* plots for all S···O contacts in PDB. The plot of θ versus δ for S···O contacts from PDB analysis. The distance cutoff 3.32 Å (sum of van der Waals radii of S and O) was used to identify these contacts (see Table 3.2). *Resolution* ≤ 2.0 Å, Pairwise sequence identity ≤ 90%, and R_{free} ≤ 25%.

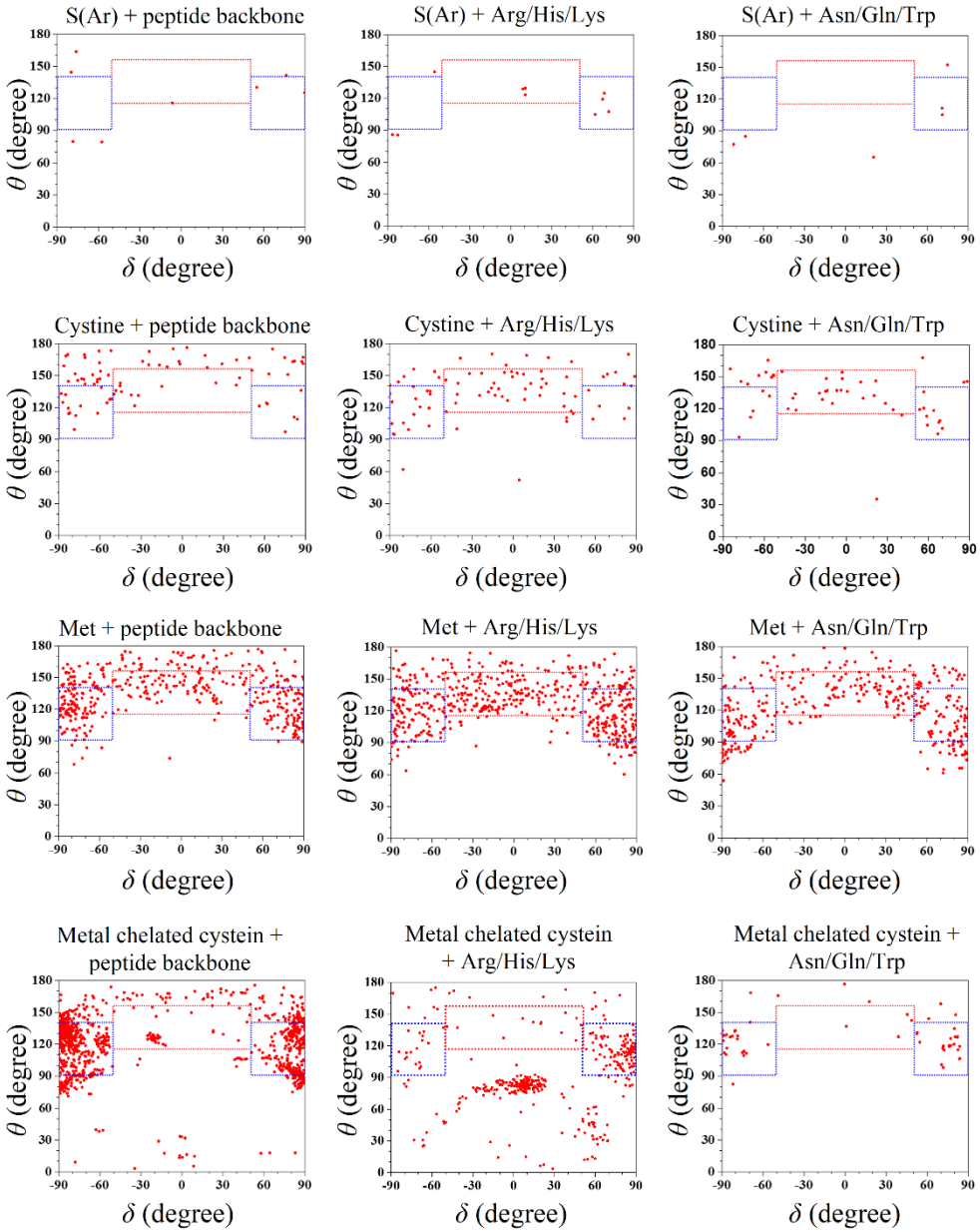


Figure 3.9 *θ-δ* plots for all S···N contacts in PDB. The plot of θ versus δ for S···N contacts from PDB analysis. The distance cutoff of 3.35 Å (sum of van der Waals radii of S and N) was used to identify these contacts (see Table 3.2). *Resolution* ≤ 2.0 Å, Pairwise sequence identity ≤ 90% and R_{free} ≤ 25%.

Table 3.3 Classification of the PDB data based on the nature of S. H-bonds and Ch-bonds were identified using the angular range of θ and δ defined in the text.

Fragment	S···H-O contacts (N_c)	S···H-N contacts (N_c)	S···O contacts (N_c)	S···N contacts (N_c)	Total Fragments (N_f)
M-S-C	775	941	109	59	1884
C-S-C	2726	600	1958	390	5674
C-S-S	175	65	1462	77	1779
S (Ar)	17	8	82	4	111
Total	3690	1614	3611	530	9448

Structures satisfying the criteria of *resolution* ≤ 2.0 Å, pairwise sequence identity ≤ 90%, and R_{free} ≤ 25% were used for the analysis.

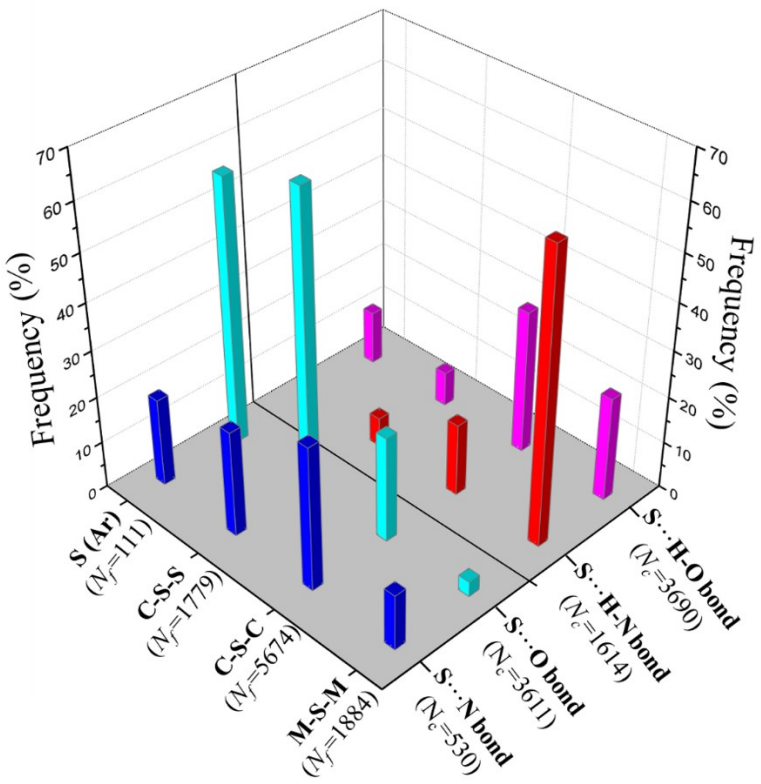


Figure 3.10 Rules for forming H- and Ch-bonds by S in proteins. Histogram showing frequency formation of Ch-bond and H-bond in the PDB with the different electronic environment of S. M = any metal; C-S-S = S from disulfide bridge; C-S-C = S from methionine and S(Ar)= Aromatic S from ligand molecules.

Aromatic S preferentially formed Ch-bond with groups containing O (83%) consistent with MESP analysis (Figure 3.10 and Table 3.3). S-mediated interaction with functional groups containing N did not show these features, possibly because of a smaller number of examples (Table 3.2). In summary, using database analysis, we found the following preferences for the formation of H- or Ch-bond by divalent S,

From CSD

M-S-Y/M fragment could form only H-bond

R-S-R fragment could form both Ch- and H-bond at almost equal preference

E-S-Y/Ar(S) fragment could preferentially form Ch-bond

From PDB

Metal Chelated Cysteines could form only H-bond

Methionine could form Ch- or H-bond at almost equal preference

Cystine could preferentially form Ch-bond

3.4 Discussion

In this study, we showed that the nature of the interaction formed by S was determined by its electronic environment, which was modulated by the covalently bonded substituents. For instance, Met-S showed an equal preference for forming H-bond and Ch-bond. At the same time, Cys-S had a higher preference for forming Ch-bond due to the lower strength of its lone pairs compared to σ-holes. The strength of σ-hole on S, when in an aromatic ring or covalently bonded to electron-withdrawing groups, is higher due to depletion of the electron density of the lone pair regions, thus favoring Ch-bond formation. Accordingly, we noted a higher frequency of Ch-bond formed with aromatic ligands. This observation is consistent with previous reports that S in the aromatic rings of drugs containing thiophene, thiazole, and thiadiazole groups interact with O in target proteins *via* Ch-bond (Koebel et al., 2016; Kristian et al., 2018; Thomas et al., 2015; Zhang et al., 2015).

Interestingly, the Ch-bonds were mostly with O rather than N. This presumably is because the strongly delocalized lone pairs of N in backbone amide or the side chain of arginine, lysine, histidine, asparagine, glutamine, or tryptophan precluded the formation of Ch-bond. Furthermore, cooperativity among S-mediated and other weak interactions is likely to modulate their strengths with implication to protein function, which remains to be studied. For example, the propensity and

strength of Ch-bond may increase upon delocalization of lone pair electron density of Cys-S to form an n→π* interaction with a vicinal carbonyl group (Kilgore and Raines, 2018).

3.5 References

Aakeroy, C.B., Bryce, D.L., Desiraju, G.R., Frontera, A., Legon, A.C., Nicotra, F., Rissanen, K., Scheiner, S., Terraneo, G., Metrangolo, P., Resnati, G., 2019. Definition of the chalcogen bond (IUPAC Recommendations 2019). Pure Appl. Chem. 91(11), 1889-1892.

Adhikari, U., Scheiner, S., 2014. Effects of charge and substituent on the S···N chalcogen bond. J. Phys. Chem. A. 118(17), 3183–3192.

Andersen, C.L., Jensen, C.S., Mackeprang, K., Du, L., Jørgensen, S., Kjaergaard, H.G., 2014. Similar strength of the NH···O and NH···S hydrogen bonds in binary complexes. J. Phys. Chem. A. 118(46), 11074-11082.

Beno, B.R., Yeung, K., Bartberger, M.D., Pennington, L.D., Meanwell, N.A., 2015. A Survey of the Role of Noncovalent Sulfur Interactions in Drug Design. 58(11), 4383-4438.

Benz, S., López-Andarias, J., Mareda, J., Sakai, N., Matile, S., 2017. Catalysis with Chalcogen Bonds. Angew. Chemie - Int. Ed. 56(3), 812–815.

Berg, J.M., Tymockzko, J.L., Stryer, L., 2013. Biochemistry, W H Freeman.

Boyle, J., 2005. Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. Biochem. Mol. Biol. Educ.

Bruno, I.J., Cole, J.C., Edgington, P.R., Kessler, M., Macrae, C.F., McCabe, P., Pearson, J., Taylor, R., 2002a. ConQuest- New software for searching the CSD and visualizing crystal structures. Acta Crystallogr. Sect. B. B58, 389-397.

Bruno, I.J., Cole, J.C., Edgington, P.R., Kessler, M., Macrae, C.F., McCabe, P., Pearson, J., Taylor, R., 2002b. Mercury 3.8 software. Acta Crystallogr. Sect. B Struct. Sci. B58, 389-397.

Chen, L., Xiang, J., Zhao, Y., Yan, Q., 2018. Reversible Self-Assembly of Supramolecular Vesicles and Nanofibers Driven by Chalcogen-Bonding Interactions. J. Am. Chem. Soc. 140(23), 7079-7082.

Dennington, R., Keith, T., Millam, J., 2009. GaussView, Version 5. Semichem Inc. , Shawnee Mission. KS.

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.; Sonnenber, D.J., 2009. Gaussian 09. Gaussian, Inc. Wallingford CT 2–3.

Garrett, G.E., Gibson, G.L., Straus, R.N., Seferos, D.S., Taylor, M.S., 2015. Chalcogen bonding in solution: Interactions of benzotelluradiazoles with anionic and uncharged Lewis bases. J. Am. Chem. Soc. 137(12), 4126–4133.

Groom, C.R., Allen, F.H., 2014. The Cambridge Structural Database in retrospect and prospect. Angew. Chemie - Int. Ed. 53(3), 662-671.

Guru Row, T.N., Parthasarathy, R., 1981. Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur in Terms of Its Orbital Orientations. 2. S…S Interactions and Nonspherical Shape of Sulfur in Crystals. J. Am. Chem. Soc. 103(2), 477–479.

Iwaoka, M., Isozumi, N., 2012. Hypervalent nonbonded interactions of a divalent sulfur atom. Implications in protein architecture and the functions. Molecules. 17(6), 7266–7283.

Iwaoka, M., Isozumi, N., 2006. Possible roles of S···O and S···N interactions in the functions and evolution of phospholipase A2. Biophysics (Oxf). 2, 23–34.

Iwaoka, M., Takemoto, S., Tomoda, S., 2002. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein engineering. J. Am. Chem. Soc. 124(35), 10613–10620.

Kilgore, H.R., Raines, R.T., 2018. N→π∗ Interactions Modulate the Properties of Cysteine Residues and Disulfide Bonds in Proteins. J. Am. Chem. Soc. 140(50), 17606-17611.

Koebel, M.R., Cooper, A., Schmadeke, G., Jeon, S., Narayan, M., Sirimulla, S., 2016. S···O and S···N Sulfur Bonding Interactions in Protein-Ligand Complexes: Empirical Considerations and Scoring Function. J. Chem. Inf. Model. 56(12), 2298-2309.

Kristian, K., Fanfrlík, J., Lepšík, M., 2018. Chalcogen Bonding in Protein−Ligand Complexes: PDB Survey and Quantum Mechanical Calculations. ChemPhysChem. 19(19), 2540-2548.

Kumar, A., Gadre, S.R., Mohan, N., Suresh, C.H., 2014. Lone pairs: An electrostatic viewpoint. J. Phys. Chem. A 118(2), 526–532.

Kumar, A., Yeole, S.D., Gadre, S.R., Lõpez, R., Rico, J.F., Ramírez, G., Ema, I., Zorrilla, D., 2015. DAMQT 2.1.0: A new version of the DAMQT package enabled with the topographical analysis of electron density and electrostatic potential in molecules. J. Comput. Chem. 36(31), 2350–2359.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157(1), 105-132.

Lim, J.Y.C., Beer, P.D., 2018. Sigma-Hole Interactions in Anion Recognition. Chem. 4(4), 731–783.

Lim, J.Y.C., Marques, I., Thompson, A.L., Christensen, K.E., Félix, V., Beer, P.D., 2017. Chalcogen Bonding Macrocycles and [2]Rotaxanes for Anion Recognition. J. Am. Chem. Soc. 139(8), 3122–3133.

Mahmudov, K.T., Kopylovich, M.N., Guedes Da Silva, M.F.C., Pombeiro, A.J.L., 2017. Chalcogen bonding in synthesis, catalysis and design of materials. Dalt. Trans. 46(31), 10121–10138.

Motherwell, W.B., Moreno, R.B., Pavlakos, I., Arendorf, J.R.T., Arif, T., Tizzard, G.J., Coles, S.J., Aliev, A.E., 2018. Non-covalent Interactions of π Systems with Sulfur: The Atomic Chameleon of Molecular Recognition. Angew. Chemie - Int. Ed. 57(5), 1193-1198.

Murray, J.S., Lane, P., Clark, T., Riley, K.E., Politzer, P., 2012. σ-Holes, π-holes and electrostatically-driven interactions. J. Mol. Model. 18, 541–548.

Nagano, N., Ota, M., Nishikawa, K., 1999. Strong hydrophobic nature of cysteine residues in proteins. FEBS Lett. 558(1), 69-71.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19(1), 115–128.

Pascoe, D.J., Ling, K.B., Cockroft, S.L., 2017. The Origin of Chalcogen-Bonding Interactions. J. Am.

Chem. Soc. 139(42), 15160–15167.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25(13), 1605-1612.

Politzer, P., Laurence, P.R., Jayasuriya, K., 1985. Molecular electrostatic potentials: An effective tool for the elucidation of biochemical phenomena. Environ. Health Perspect. 61,191-202.

Politzer, P., Murray, J.S., 2002. The fundamental nature and role of the electrostatic potential in atoms and molecules. Theor. Chem. Acc. 108, 134–142.

Politzer, P., Murray, J.S., Clark, T., 2014. σ-Hole bonding: A physical interpretation. Top. Curr. Chem. 358, 19–42.

Politzer, P., Murray, J.S., Clark, T., 2013. Halogen bonding and other σ-hole interactions: A perspective. Phys. Chem. Chem. Phys. 15, 11178–11189.

Politzer, P., Murray, J.S., Clark, T., Resnati, G., 2017. The σ-hole revisited. Phys. Chem. Chem. Phys. 19, 32166-32178.

Politzer, P., Murray, J.S., Peralta-Inga, Z., 2001. Molecular surface electrostatic potentials in relation to non-covalent interactions in biological systems. International Journal of Quantum Chemistry. 85(6), 676–684.

Rao Mundlapati, V., Ghosh, S., Bhattacherjee, A., Tiwari, P., Biswal, H.S., 2015. Critical assessment of the strength of hydrogen bonds between the sulfur atom of methionine/cysteine and backbone amides in proteins. J. Phys. Chem. Lett. 6(8), 1385-1389.

Řezáč, J., Hobza, P., 2016. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. Chem. Rev. 116(9), 5038-5071.

Rose, P.W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Di Costanzo, L., Duarte, J.M., Dutta, S., Feng, Z., Green, R.K., Goodsell, D.S., Hudson, B., Kalro, T., Lowe, R., Peisach, E., Randle, C., Rose, A.S., Shao, C., Tao, Y.P., Valasatava, Y., Voigt, M., Westbrook, J.D., Woo, J., Yang, H., Young, J.Y., Zardecki, C., Berman, H.M., Burley, S.K., 2017. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. Nucleic Acids Res. 45, D271–D281

Rosenfield, R.E., Parthasarathy, R., Dunitz, J.D., 1977. Directional Preferences of Nonbonded Atomic Contacts with Divalent Sulfur. 1. Electrophiles and Nucleophiles. J. Am. Chem. Soc. 99(14), 4860–4862.

Scilabra, P., Terraneo, G., Resnati, G., 2019. The Chalcogen Bond in Crystalline Solids: A World Parallel to Halogen Bond. Acc. Chem. Res. 52(5), 1313-1324.

Seifert, E., 2014. OriginPro 9.1: Scientific Data Analysis and Graphing Software—Software Review. J. Chem. Inf. Model. 54(5), 1552-1552.

Shelke, S.S., 2020. Divalent Sulfur mediated interactions in proteins architecture, stability and molecular recognition. Master dissertation, IISER, Pune.

Thomas, S.P., Jayatilaka, D., Guru Row, T.N., 2015. S···O chalcogen bonding in sulfa drugs: Insights from

multipole charge density and X-ray wavefunction of acetazolamide. Phys. Chem. Chem. Phys. 17, 25411–25420.

Wang, G., Dunbrack, R.L., 2005. PISCES: Recent improvements to a PDB sequence culling server. Nucleic Acids Res. 33, W94–W98.

Wang, W., Zhu, H., Feng, L., Yu, Q., Hao, J., Zhu, R., Wang, Y., 2020. Dual Chalcogen-Chalcogen Bonding Catalysis. J. Am. Chem. Soc. 142(6), 3117-3124.

Yeole, S.D., López, R., Gadre, S.R., 2012. Rapid topography mapping of scalar fields: Large molecular clusters. J. Chem. Phys. 137(7), 074116.

Zhang, X., Gong, Z., Li, J., Lu, T., 2015. Intermolecular Sulfur···Oxygen Interactions: Theoretical and Statistical Investigations. J. Chem. Inf. Model. 55(10), 2138–2153.

Zhao, Y., Truhlar, D.G., 2008. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. Theor. Chem. Acc. 120, 215–241.

Zhou, P., Tian, F., Lv, F., Shang, Z., 2009. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Bioinforma. 76(1), 151–163.

Chapter 4

The role of sulfur mediated chalcogen and hydrogen bonds in protein structure, stability, and substrate specificity

4.1 Introduction

Non-covalent interactions are fundamental for protein folding, structural stability, and function. Traditionally, hydrogen bonds (H-bonds), hydrophobic effects, electrostatic, and van der Waals interactions are assumed to be the major drivers of protein folding and stability (Dill and MacCallum, 2012; Nick Pace et al., 2014). However, the essentiality of other weak interactions in sculpting protein structures is also being discovered. For example, the importance of weak H-bonds, such as C-H···O interaction, cation/anion-π, and n→π^* interactions for the stability of protein structures is well known (Bartlett et al., 2010; Derewenda et al., 1995; Gallivan and Dougherty, 1999; Lucas et al., 2016; Manikandan and Ramakumar, 2004; Newberry and Raines, 2019). Apart from C, O, N, and H that form these non-covalent interactions, divalent sulfur (S) is present in methionines (Met-S^{δ}) and/or cystines (Cys-S^{γ}) of proteins. Sulfur has unique bonding properties that allow it to participate in polar interactions such as hydrogen (H-bond) or chalcogen bond (Ch-bond) (Iwaoka et al., 2002; Pal and Chakrabarti, 2001; Zhou et al., 2009). Consequently, methionine and cystine are expected to contribute to distinct polar interactions in proteins, making them imperative to study to better understand the structures and folding of proteins.

In Chapter 3, we showed that the electronic environment of divalent sulfur (S) determines the nature of polar bonds it can form in proteins. For instance, S^{γ} of the metal-chelated cysteines primarily participate in the H-bond formation, while S^{γ} of disulfide-bridge cysteines prefers to form Ch-bond. Interestingly, S^{δ} of the methionine appears to form both interactions at almost equal preference (Figure 4.1). Additionally, this preference of S to participate in the formation of H- or Ch-bonds could directly be correlated to the strength of its lone pairs and σ-holes (Figure 4.1). Consequently, these observations led us to examine the roles of H- and Ch-bond made by S and their interplay in the structure, stability, and folding of the proteins, if any. Additionally, the precise role of Ch-bond in protein-substrate interaction is also unaddressed.

Here, we have addressed these questions through extensive computational and bioinformatics analyses. We identified many H- and Ch-bonds formed by S in different protein secondary structures. Using the database analysis, we showed that the nature of the H-bond formed by metal-chelated cysteines could alter the conformation of the peptide backbone. Computational calculations showed that the interactions augment conformational stability by various mechanisms, including capping α-helix termini, protecting free β-sheet edges by negative-design, and augmenting the stability of β-turns. Furthermore, using a mutagenesis-based approach, we demonstrate a critical role of Ch-bond in the substrate specificity of the enzyme methionyl-tRNA

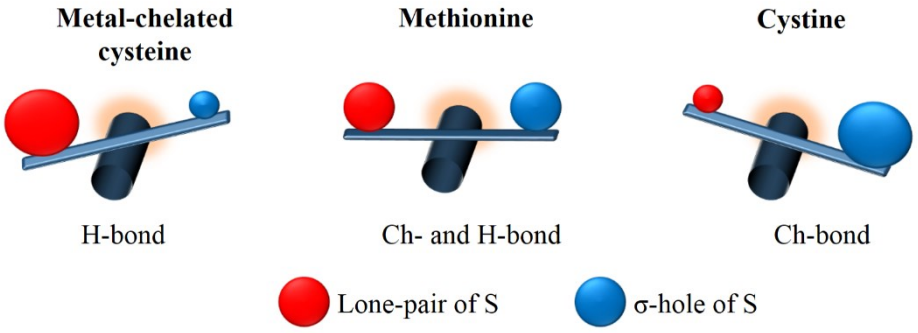


Figure 4.1 Seesaw changes in the strength of the lone pairs and σ–holes on S. A graphical representation showing the substituent-dependent change in the strength of the lone pairs and σ–holes affecting the nature of bond formation by S in proteins.

synthetase. Together, the study reported here unravels the underappreciated roles of S-mediated interactions, particularly of Ch-bonds, in protein structure, stability, and molecular recognition.

4.2 Methods

4.2.1 Computational details

Model systems used to investigate the role of S···O interactions in α-helix is from the coordinates of 1PVH (PDB ID 1PVH), and 4KT1 (PDB ID 4KT1) for β-strand (Figure 4.7 **A**). All the side chains atoms in these fragments were deleted, and the C_β atom was replaced by H. The hydrogen atoms were added manually using Gaussview 5.0 program (Dennington et al., 2009), following which partial energy minimization was carried out (all atoms other than hydrogen were frozen). Coordinates of the altered structures are provided in Appendix **C**. The torsion angles (χ) between N, C_α, C_β, and S (Figure 4.7 **A**) were varied for the PES scan. A similar strategy was used to optimize the representative structures for studying Cases 1 to 3 of the CXXXXC motif shown in Figure 4.8 (PDB IDs 2FD6 for Case 1, 3CEL for Case 2, and 1HTR for Case 3). Coordinates of the optimized structures are provided in Appendix **C**. AIM analysis was carried out using AIM2000 (Biegler-König and Schönbohm, 2002; König et al., 2001) for the three structures.

4.2.2 PDB analyses

The two PDB datasets, defined below (also discussed in the previous Chapter 3), was used for the analysis carried out in this chapter (Shelke, 2020)

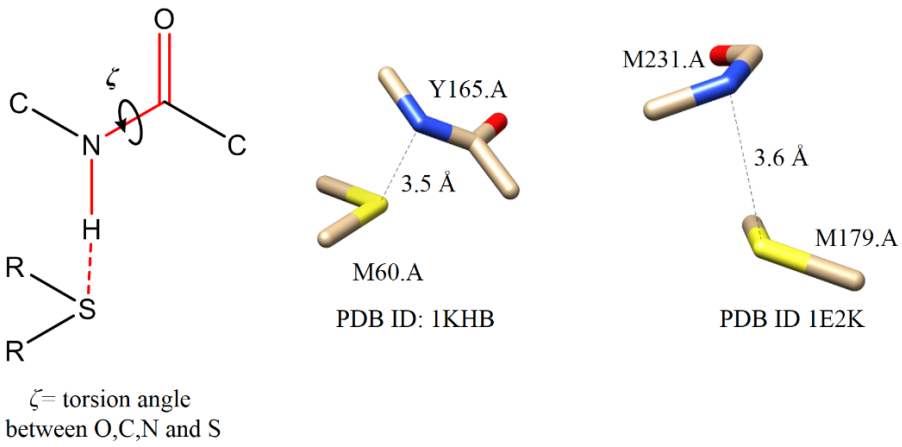


Figure 4.2 Definition of ζ. The definition of ζ, along with some of the representative examples, where ζ value was greater than 240° or less than 120°, indicating the respective H was pointing away from the lone pair region of S.

Dataset 1: pair-wise sequence identity ≤ 90%, *resolution* ≤ 2.0, and R_{free} ≤ 25%,

Dataset 2: pair-wise sequence identity ≤ 90%, *resolution* ≤ 2.5 Å, and R_{free} ≤ 30%.

Search for metal-chelating cysteine rendered using a distance between S and metal between 1.9 to 2.8 Å. The distance range ensured that metals with different ionic radii were identified. The criteria for H-bond used was $d \leq 3.6$ Å (Zhou et al., 2009), $90° \leq \theta \leq 140°$, $-90° \leq \delta < -50°$ or $50° < \delta \leq 90°$, and $120^{\circ} \leq \zeta \leq 240^{\circ}$ (the criteria to identify H-or Ch-bond is discussed in Chapter 3). An additional criterion of an imaginary torsion angle ζ was applied to exclude structures where N-H was not pointing towards the lone pair regions of S (Figure 4.2 and Figure 4.3 **A**).

To understand the role of H-bond and Ch-bond in capping of α-helix, we searched for S···H-N ($d_{S\cdots N} \leq 3.6$ Å, $90° \leq \theta \leq 140°$, $-90° \leq \delta < -50°$ or $50° < \delta \leq 90°$, and $120^{\circ} \leq \zeta \leq 240^{\circ}$) and S···O ($d_{S\cdots O} \leq 3.32$ Å, $115° \leq \theta \leq 155°$ and $-50° \leq \delta \leq 50°$) interactions made by the peptide backbone with S. α-helices were identified using the header information in the PDB coordinate files (Shelke, 2020). For N-terminal capping, we searched for S···H-N interactions where the amino group was of N_1, N_2, or N_3 residue at helix terminus as defined previously (Aurora and Rose, 1998). While, for C-terminal capping, we searched for S···O interactions where the carbonyl O was of C_1, C_2, or C_3 residue at helix terminus as defined previously (Table 4.2) (Aurora and Rose, 1998).

To investigate S···H-N ($d_{S\cdots N} \leq 3.6$ Å, $90° \leq \theta \leq 140°$, $-90° \leq \delta < -50°$ or $50° < \delta \leq 90°$ and $120° \leq \zeta \leq 240°$), S···O ($d_{S\cdots O} \leq 3.32$ Å, $115° \leq \theta \leq 155°$ and $-50° \leq \delta \leq 50°$) and S···N ($d_{S\cdots N} \leq 3.35$ Å, $115° \leq \theta \leq 155°$ and $-50° \leq \delta \leq 50°$) interactions in α-helices and β-strands, we obtained the information of secondary structures from the header in the PDB coordinate files. Only internal-helical residues were considered, and the last four residues at the N- and the C-termini of the helix were excluded. In the case of β-strands, all the residues belonging to the strand were considered (Shelke, 2020) (Table 4.3).

To understand the role of S···O interaction in the CXXXXC motif, we searched for disulfide-linked cysteines separated by four intervening residues. The cyclic peptide was classified into four different groups, namely; Case 1: Co-existing S···O bond between the S of *i* or *i*+5 cysteine and the carbonyl O of *i*+2 residue ($d_{S\cdots O} \leq 3.32$ Å, $115° \leq \theta \leq 155°$ and $-50° \leq \delta \leq 50°$), and H-bond between *i*+2 and *i*+5 residues ($d_{O\cdots N} \leq 3.5$ Å); Case 2: Co-existing S···O bond between the S of the *i* or *i*+5 residue and the carbonyl O of the *i*+2 residue, and H-bonds between *i*+2 and *i*+5 residues and *i*+1 and *i*+4 residues; Case 3: H-bond between *i*+2 and *i*+5 residues and absence of S···O bond between *i* or *i*+5 residue and *i*+2 residue; and Case 4: absence of any of the above mentioned non-covalent bonds (Table 4.4).

4.2.3 Cloning and purification of MetRS and MetRS^{Y260F}

The gene encoding 1 to 551 amino acids of MetRS was amplified from genomic DNA of *E. coli* K12 strain using forward primer (5'-AACTTTAATAAGGAGATATACCATGGGCACTCAAG TCGCGAAGAAAATTCTGGTGA-3') and reverse primer (5'-CAGCAGCGGTTTCTTTACCA GACTCGAGTTATTTTACTTCTTCTTTAGAGGCTT-3') (Shelke, 2020). The amplified gene was cloned into a pRSF vector using the restriction sites of NcoI and XhoI. The integrity of the positive clone was confirmed by complete gene sequencing. MetRS-Tyr260 was mutated to phenylalanine using the site-directed mutagenesis method using the primer 5'-TGGCTGGACGCACCGATTGGCTTTATGGGTTCTTTCAAGAATCTG-3'. MetRS and MetRS^{Y260F} were purified using a previously mentioned protocol (Mellot et al., 1989) using streptomycin sulfate and ammonium sulfate precipitation, followed by purification using Hi-Trap DEAE sepharose FF and Superdex 75 columns (GE Healthcare). The purified proteins were stored in a binding buffer (20 mM Tris-HCl pH 7.6, 0.1 mM EDTA, 10 mM DTT).

4.2.4 Isothermal Titration Calorimetry (ITC) binding experiments

For ITC based measurements, purified MetRS was dialyzed in buffer containing 20 mM Tris-Cl, pH 7.6, and 0.1 mM EDTA at 4 ºC overnight. Ligands (methionine or norleucine) were also dissolved in the same dialysis buffer. The concentration of MetRS was determined using Nanodrop2000 by measuring absorbance at 280 nm. ITC measurements were carried out by titrating 1 mM of ligands solution into 25 uM of MetRS at 25 ºC using the MicroCal PEAQ-ITC. The titration speed of the stirrer was 1000 rpm, and the injection interval time was 3 mins. Data for each experiment was modeled using a one-site binding model provided in the MicroCal PEAQ-ITC Analysis software with fitted offset control. Stoichiometry of the binding was fixed to 1 (N=1) to obtain reliable values of K_d, $\Delta G°$, $\Delta H°$, and $T\Delta S°$ as suggested previously for systems having low c-values (Kantonen et al., 2017). Error in K_d and thermodynamic quantities were calculated as the standard error of the mean using three independent experiments.

4.2.5 Fluorescence-based binding experiments

Change in intrinsic fluorescence of MetRS/MetRS^{Y260F} upon titration with methionine or norleucine was measured using the Horiba fluroMax 4 spectrophotometer. All the readings were acquired using an excitation wavelength of 295 nm and an emission wavelength of 330 nm at 25 °C. These measurements were carried out using an enzyme concentration of 0.8 μM and a reaction volume of 200 ul. Measurements were performed using 0 μM, 10 μM, 50 μM, 100 μM, 200 μM, 500 μM, 700 μM, 1000 μM, 1500 μM and 2000 μM of methionine or norleucine. Data were fitted to a simple equilibrium model, and binding parameters were obtained using GraphPad PRISM 8.4.1.

4.3 Results

4.3.1 Metal-coordinated S participate in H-bond formation

We first analyzed the PDB for non-covalent interactions formed by metal-chelated cysteines, which occur in many metalloproteins. Consistent with the rules defined in chapter 3 and independent of the identity of the metal, the thiolate of cysteine preferentially formed H-bond with a neighboring donor, in particular the backbone N-H (Figure 4.3 **A**-**B** and Table 4.1). As the resolution of the PDB structures analyzed were in general lower (resolution ≤ 2 Å) than the CSD structures, we relaxed the distance criterion for H-bond formation to $d_{S\cdots N} \leq 3.6$ Å (Table 4.1).

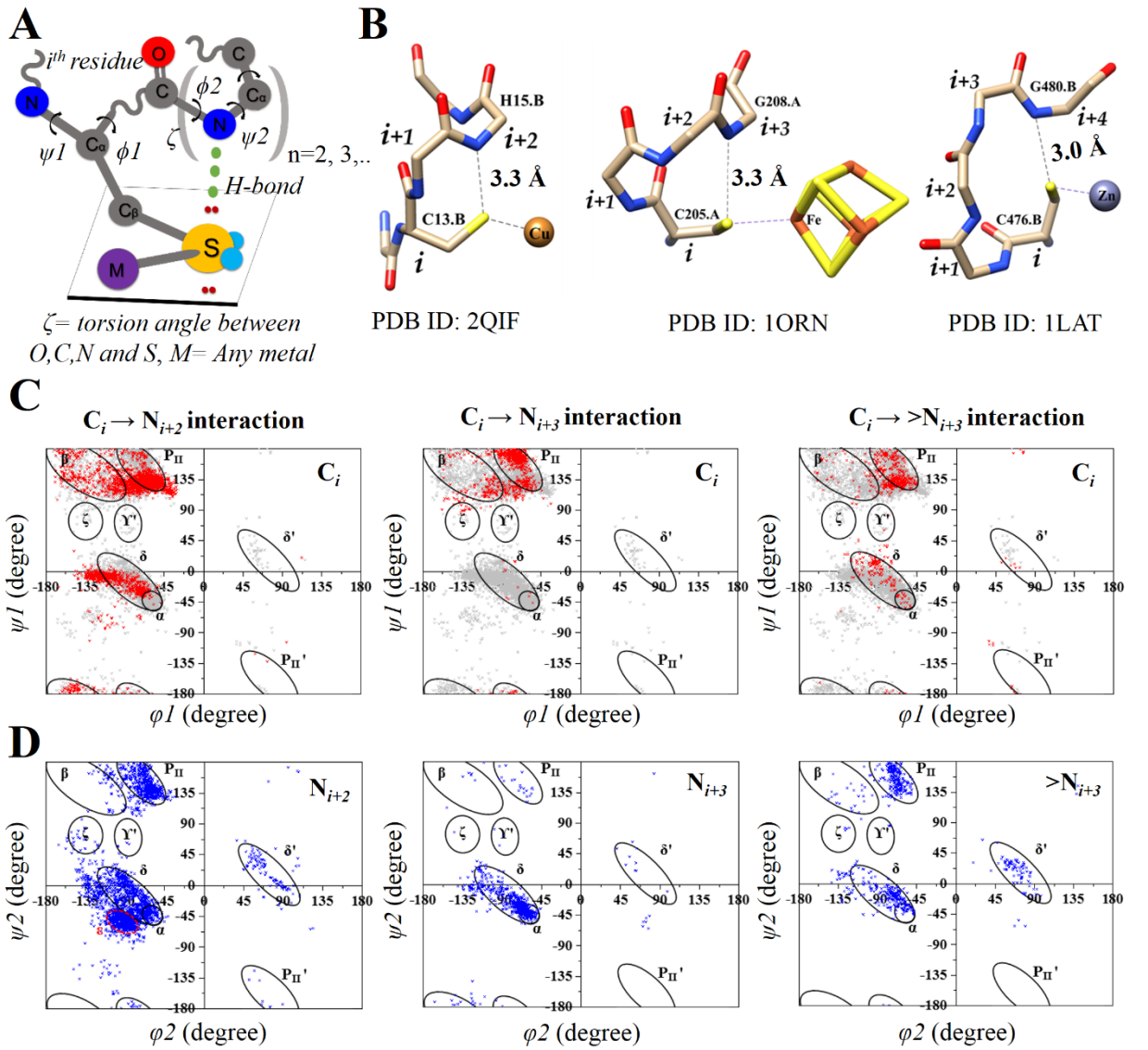


Figure 4.3 H-bond by metal-chelating cysteines. (**A**) A schematic representation of S···H-N interaction between C_i^{th} and N_{i+n}^{th} residues along with the definition of Ramachandran angles and geometry of the interaction (C represent metal-chelating cysteine and N represents any H-bond donor residue). (**B**) Some representative examples of H-bond between C_i^{th} and N_{i+n}^{th} residues. (**C**) *Φ* and *Ψ* values of all metal-chelating cysteine in the Ramachandran plot (refer to Table 4.1 for contacts details). Points that satisfy the distance and direction criteria of the H-bond are shown as red points. (**D**) *Φ* and *Ψ* values of the N_{i+n}^{th} residue act as H-donor in the H-bond with metal-chelating cysteine shown as the blue point. Highlighted in the Ramachandran plots are the commonly populated regions. A red circle marks the ε region defined in this study.

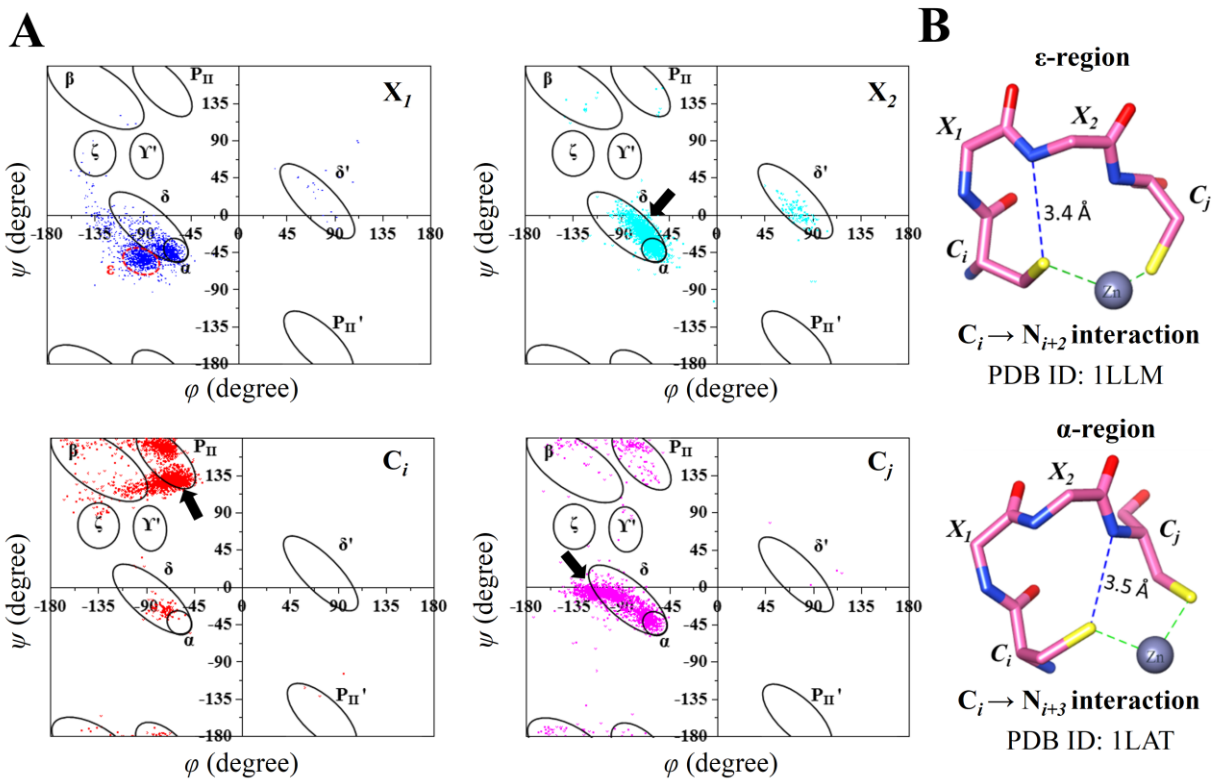


Figure 4.4 Identification ε-region in Ramachandran plot resulting from S-mediated H-bond. (**A**) Ramachandran plot of residues C_i, X_1, X_2, and C_j of all metal-bound CXXC motifs in the PDB (Dataset II). The ε-region that populates by X_2 residue is shown in the red circle, and back arrow indicates the corresponding conformation of C_i, X_2, and C_j residues. (**B**) A representative example CXXC motif (C_i-X_1-X_2-C_j) in PDB showing H-bond between C_i and X_2 residues, where X_2 residue is populating in the ε-region (top panel). A representative example CXXC motif (C_i-X_1-X_2-C_j) in PDB forming H-bond between C_i and C_j residues, where X_2 residue is populating in the α-region (bottom panel).

The conformations of the two interacting residues were dependent on their relative spacing, as observed from the Ramachandran plot of the backbone torsion angles Φ and Ψ of the cysteine (C_i^{th}) and the residue (N_{i+n}^{th}) forming the H-bond (Figure 4.3 **C**-**D**). For example, when the cysteine at position *i* was separated from the H-bond donor by one amino acid (N_{i+2}), the H-bond donor residue populated P_{II}, α, δ and a unique region adjacent to α and δ (marked by red dotted circle) (Figure 4.3 **C**-**D**, left panel).

Inspection of structures populating this unique region revealed that in most of them, the cysteine C_i was part of a metal-bound CXXC motif, such as in zinc-binding proteins. These metal-bound CXXC motifs had a unique backbone conformation. Each of the four residues had preferred Φ and Ψ values (Figure 4.4). The S-mediated H-bond was integral to the unique conformation of this set of metal-bound CXXC motifs (Figure 4.4). We suggest that the region populated by N_{i+2}

Table 4.1 A summary of the results of the PDB analysis performed to investigate the H-bond in M-S-C fragment.

Criteria	Dataset I	Dataset II
Metal-S-C fragments	5301	8887
$C_i \rightarrow N_{i+2}$ interaction	1918	2998
$C_i \rightarrow N_{i+3}$ interaction	498	863
$C_i \rightarrow > N_{i+3}$ interaction	343	641
C_i-X_1-X_2-C_j motif	1121	1978

Analysis performed using Dataset I (*Resolution* ≤ 2.5 Å, pair-wise sequence identity ≤ 90% and R_{free} ≤ 30%) and Dataset II (*Resolution* ≤ 2.0 Å, pair-wise sequence identity ≤ 90% and R_{free} ≤ 25%).

of this motif in the Ramachandran plot (-75° ≤ ϕ ≤ -30° and -120° ≤ ψ ≤ -70°) be referred to as ε (Figure 4.3 **C**-**D**, left panel and Figure 4.4), following the nomenclature used by Karplus for other commonly populated regions of the plot (Hollingsworth and Karplus, 2010).

When the donor residue was at i+3 position (N_{i+3}), then the Ramachandran angles of the residue populated α and δ regions (Figure 4.3 **C**-**D**, central panel). N_{i+3} in α region were often at the N-terminus of α-helices. By forming H-bond with the backbone N-H of N_{i+3}, the metal-bound thiolate capped the N-terminus of the helix. Capping satisfies the H-bond forming abilities of the free backbone N-H or C=O of the terminal residues of an α-helix (Aurora and Rose, 1998). The above observation is consistent with a previous study on synthetic peptides, which found that the negatively charged thiolate of cysteine had a high preference to cap the N-terminus of α-helix (Doig and Baldwin, 1995).

4.3.2 Role of S in helix capping

Capping is essential for the stability of α-helices in proteins and peptides (Aurora and Rose, 1998). The role of polar side chains of serine, threonine, and asparagine, the acidic side chain of aspartate, the backbone amide of a neighboring residue, and metal-chelated S of cysteine in helix capping are well documented (Aurora and Rose, 1998; Doig and Baldwin, 1995), but not those of methionine and cystine. Since the N-terminus and the C-terminus of α-helices have free backbone N-H (electrophile) and free backbone C=O (nucleophile), respectively, we asked if Met-S^{δ} or Cys-S^{γ} would interact and cap them.

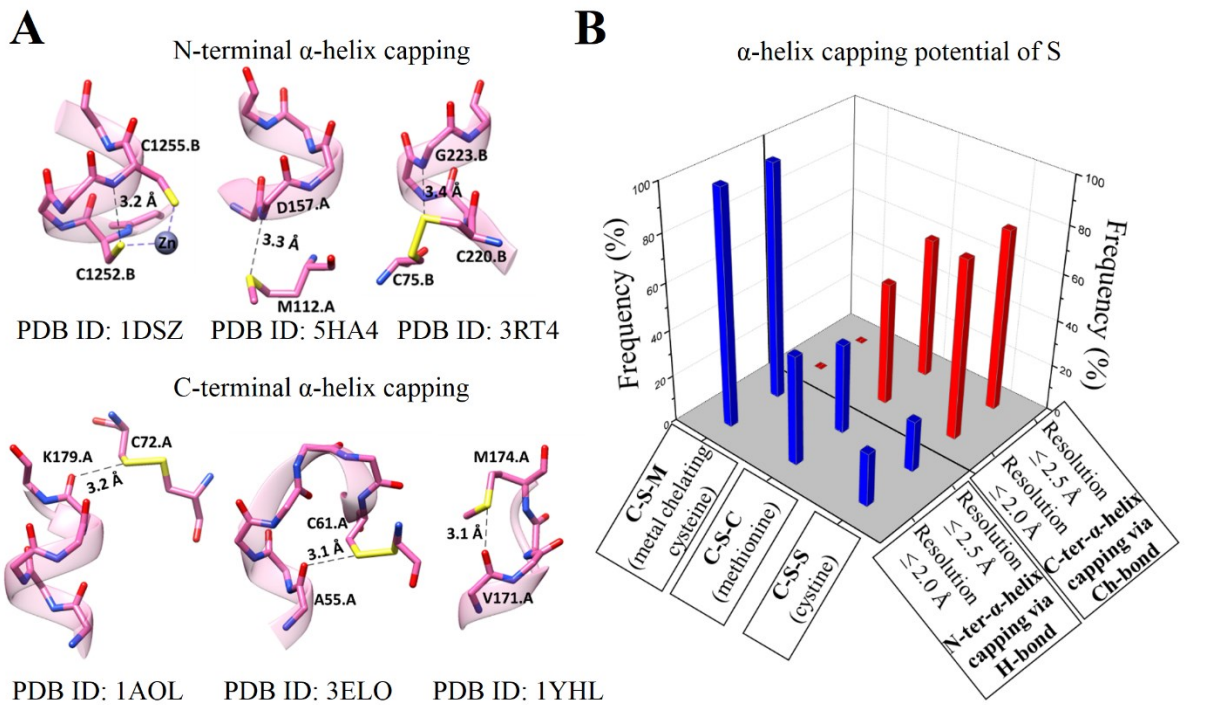


Figure 4.5 α-helices capping by the S-mediated interactions. (A) Representative examples of H-bond capping the N-terminus of α-helices and Ch-bond capping the C-terminus of α-helices. (**B**) Histogram showing the frequency of H-bond and Ch-bond interactions capping the N- and C-termini of α-helices by metal-chelating cysteine, methionine, and cystine.

Table 4.2 A summary of the results of the PDB analysis performed to identify H-bond and Ch-bond formed by Cys-S^{γ} or Met-S^{δ} that cap α-helices in proteins.

Fragment (residue)	Total Number of α-helix capping contacts (N_T) [a]		N-terminal α-helix capping contacts (N_N) [b]		C-terminal α-helix capping contacts (N_C) [c]	
	Dataset I	Dataset II	Dataset I	Dataset II	Dataset I	Dataset II
C-S-S (Cystine)	107	161	24	36	83	125
C-S-C (Methionine)	168	249	77	95	91	154
C-S-M (Metal-chelated cysteine)	526	898	526	898	0	0

[a]Total number of S···O/ S···H-N contacts found capping α-helices. [b]Total number of S···H-N contacts found capping the N-termini of α-helices. [c]Number of S···O contacts found capping the C-termini of α-helices. Refer to Methods for the definition dataset I and II.

We analyzed protein structures in the PDB and found several examples of Met-S^{δ} or Cys-S^{γ} interacting with backbone amide at either the N-terminus or the C-terminus of α-helix (Figure 4.5 **A** and Table 4.2). Amongst the examples involving Met-S^{δ}, 46% of the interactions were H-bonds with backbone N-H of N-terminal residues and 54% Ch-bonds with backbone C=O of C-terminal residues (Figure 4.5 **B** and Table 4.2). In contrast, amongst the examples involving Cys-S^{γ}, 78% capped the C-terminus by Ch-bond, while the other 22% capped the N-terminus (Figure 4.5 **B** and Table 4.2). Our analysis revealed that Cys-S^{γ} or Met-S^{δ} can also cap the N- or the C-terminus of α-helices by H-bond or Ch-bond, respectively. This is the first study reporting the role of Ch-bond in helix capping (this work was carried out in collaboration with Shelke, 2020).

4.3.3 Augmentation of the stability of regular secondary structures by S

In addition to H-bond between i and i+4 residues and the capping interactions, other non-covalent interactions such as C-H···O and n→π* are essential for structural stability of α-helices (Bartlett et al., 2010; Manikandan and Ramakumar, 2004). Similarly, the structural stability of β-sheets is dependent not only on inter-strand H-bonds between backbone O and N-H but also on inter-strand C-H···O interactions (Derewenda et al., 1995). An earlier study reported instances of methionine forming intra-helical and inter-strand Ch-bonds with backbone O (Pal and Chakrabarti, 2001). This prompted us to find if S could also contribute to the stability of the regular secondary structures, namely α-helices and β-strands through H-bond and Ch-bond with backbone O or N-H.

We analyzed the PDB for interactions made by Met-S^{δ} or Cys-S^{γ} with backbone O or N-H of residues constituting α-helices and β-strands. In the case of α -helices, since the interactions with residues at the termini were analyzed in the previous sections, only internal residues were considered. Met-S^{δ} and Cys-S^{γ} formed Ch-bond with backbone O of α-helices and β-strands (Figure 4.6 **A** and Table 4.3). Examples of Ch-bonds with backbone N of α-helical and β-strand residues, though present, were considerably less in number (Table 4.3). However, as would be expected, N-H···S bond was not observed in intra-helical regions, as the backbone N-H of i^{th} residue was H-bonded to the backbone C=Os of the $i-4^{th}$ residue.

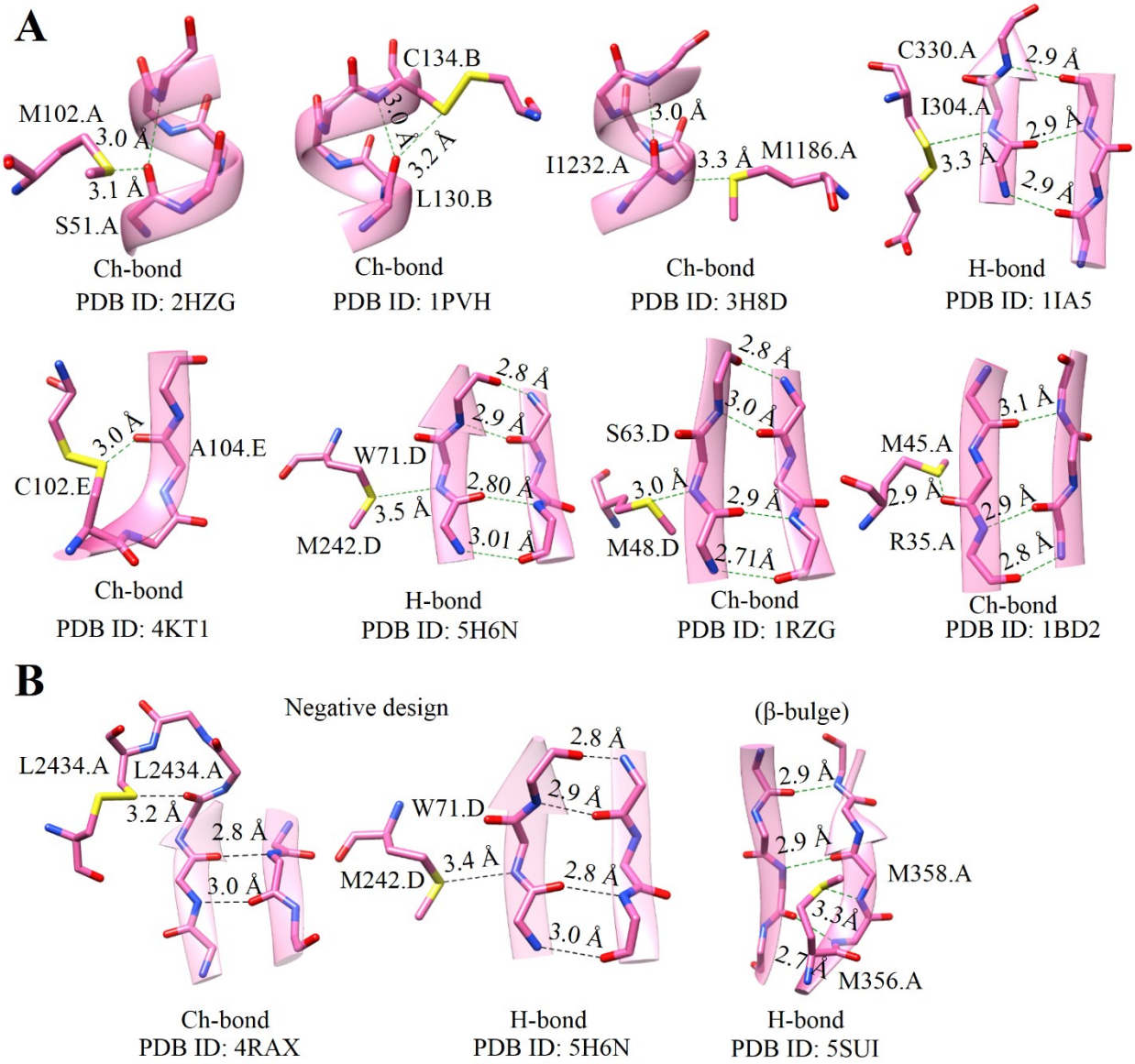


Figure 4.6 Stabilizing role of S-mediated interactions in α-helix and β-sheet. (**A**) Representative examples of H-bonds and Ch-bonds formed by Met-S^{δ} and Cys-S^{γ} with residues in α-helix and β-sheet. (**B**) An example of H-bonds formed by Met-S^{δ} introducing β-bulge and negative-design involving Ch-bond and H-bond.

Table 4.3 A summary of the results of PDB analysis performed to identify H-bond and Ch-bond formed by Cys-S^{γ} or Met-S^{δ} with residues in α-helix (only internal residues) and β-sheets. (*Resolution* ≤ 2.5 Å, pair-wise sequence identity ≤ 90% and R_{free} ≤ 30%).

Fragment (residue)	S···H-N interaction		S···O interaction		S···N interaction	
	α-helix	β- strand	α-helix	β- strand	α-helix	β- strand
C-S-S (Cystine)	0	13	267	93	0	10
C-S-C (Methionine)	0	243	104	175	39	40

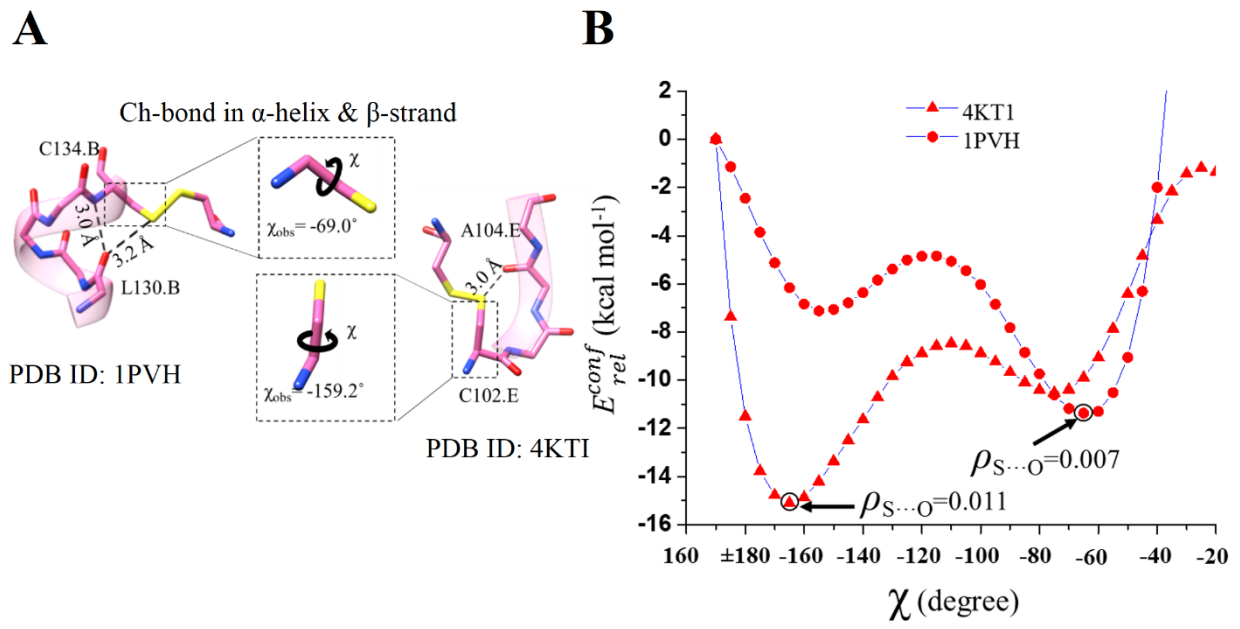


Figure 4.7 Stabilization of α-helix and β-sheet by Ch-bond. (**A**) Representative examples of Ch-bond found in α-helix and β-sheet. (**B**) The plot for conformational energies as a function of χ (E at χ = 170˚ was assigned as 0.0 kcal.mol^{-1}). Values of ρ at BCP for S···O interaction for the most energetically favorable structures are given in au.

Interestingly, we found N-H···S bond in β-strand regions (Figure 4.6 **A** and Table 4.3), many of which involved backbone N-H of β-strands at the edge of β-sheets (Figure 4.6 **B**). The H-bond appeared to stabilize the free N-H. We also found a Ch-bond formed by S with free backbone C=O of edge strands (Figure 4.6 **B**). Many elements of negative-design that stabilize the edge strand of β-sheets have been documented previously, including the interaction of other regions of the protein with the edge β-strand, disruption of backbone H-bond formation by proline, or a β-bulge, or use of inward-pointing charged residues to prevent strand-mediated dimerization (Koga et al., 2012; Richardson and Richardson, 2002). Our analysis revealed that H-bond or Ch-bond formed by backbone N-H or C=O of edge β-strand with a neighboring Met-S$^{\delta}$ or Cys-S$^{\gamma}$ is another element of negative-design that can stabilize β-sheets (Figure 4.6 **B**). Additionally, we found that in some proteins, the free backbone N-H of the insertion residue of a classical β-bulge formed H-bond with Met-S$^{\delta}$ located two residues ahead (Figure 4.6 **B**).

To gauge the potential of Ch-bond to lock protein conformation and stabilize regular secondary structures, we performed a potential energy surface (PES) scan by varying the torsion angle χ about the covalent bond between C$^{\alpha}$ and C$^{\beta}$ of the cysteine with its S forming the Ch-bond (Figure 4.7 **A**). One fragment each from α-helix and a β-strand were chosen for the calculations

(Figure 4.7 **A**). For ease of calculation, the side chain groups were excluded from the calculations (see Methods). A plot of relative conformational energy vs. χ showed that the minimum conformational energy corresponded to the χ value of the respective crystal structures (Figure 4.7 **B**). The structures corresponding to the energy minimum were subjected to Atoms in Molecules (AIM) analysis which showed the presence of Bond Critical Point (BCP) between S and O. ρ values of these BCPs were in the range suggested previously, i.e., 0.002-0.035 au (Figure 4.7 **B**) (Bader, 1991). The analysis, thus, strongly suggested that Ch-bond could provide extra stability to a particular conformation in protein molecules.

4.3.4 Ch-bond stabilizes β-turn containing motifs

Once we established that S-mediated interactions could contribute to the stability of regular secondary structures, we proceeded to find if they also augment the stability of β-turns. We found such examples in the motif CXXXXC, in which a disulfide bond links cysteine at i and i+5 positions. The motif, to be referred to as Case 1, adopts a unique conformation in which residues i+2 to i+5 form a type II β-turn, with the backbone C=O of i+2 residue additionally forming a Ch-bond with the disulfide-bonded S of the cysteine at i+5 position (Figure 4.8 and Table 4.4). The motif was found in diverse proteins, notably among the members of the three-finger protein fold (PDB ID 6GBI) bacterial pore-forming toxin proaerolysin (PDB ID 1PRE), carboxylesterase Notum (PDB ID 4UZ1), lipoprotein lipase (PDB ID 6E7K), Type IV Pilins (PDB ID 5G24) and FAB (PDB ID 3QYC). Each of the six residues of the motif occupies a unique position in the Ramachandran plot (Figure 4.9 **A**). The Ch-bond and the H-bond between backbone C=O of i+2 and backbone N-H of i+5 residues, the latter characteristic of a type II β-turn (Venkatachalam, 1968), appeared to stabilize the unique conformation of the motif.

The CXXXXC motif in the protein Cel7A had a type I β-turn instead of type II turn, which was immediately preceded by a type II' β-turn (to be referred to as Case 2; Figure 4.8; Table 4.4). The CXXXXC motif of ionotropic glutamate receptor (PDB ID 4YKI), α-conotoxin (PDB ID 1HJE), Xpd4 helicase (PDB ID 2VL7), and archaeal prim-pol domain (PDB ID 1RO2) also has a conformation similar to Case 2 but without a preceding type II' β-turn. A striking feature of the CXXXXC structural motifs in Cases 1 and 2 was the conservation of the Ramachandran angles of the constitutive residues (Figure 4.9 **A** and **B**). In comparison, the Ramachandran angles of the six residues in the CXXXXC structural motifs lacking the Ch-bond were clearly less conserved. These structures could be divided into two groups. Case 3 had a type I β-turn and the corresponding H-bond between backbone C=O of i+2 and backbone N-H of i+5 residues (Figure 4.8, Figure 4.9 **C**

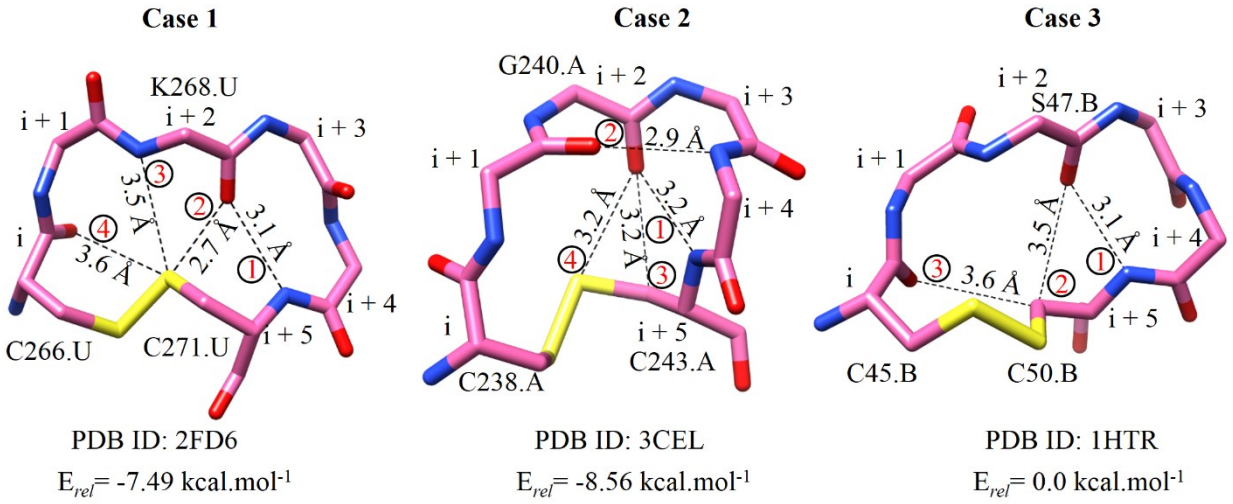


Figure 4.8 Identification of Ch-bond in β-turns. Representative examples of disulfide-linked CXXXXC motif in protein structures. The left panel is Case 1, having a type II β-turn. The central panel is Case 2, having a type I and a type II' β-turn. The right panel is Case 3, having a type I β-turn. Co-existing S···O, S···H-N, C···H-O and O···H-N interactions if present are marked by dotted lines and numbered in red in each panel. E_{rel} calculated for the three structures are given below the respective panel (refer to Methods section for details).

Table 4.4 A summary of the results of PDB analysis performed to study different CXXXXC motifs.

Criteria	Number of contacts	
	Dataset II	Dataset I
Case 1 ($d_{S\cdots O} \leq 3.32$ Å, $115^\circ \leq \theta \leq 155^\circ$, $-50^\circ \leq \delta \leq 50^\circ$, $d_{N5\cdots O2} \leq 3.5$ Å, $d_{N4\cdots O1} > 3.5$ Å)	72	46
Case 2 ($d_{S\cdots O} \leq 3.32$ Å, $115^\circ \leq \theta \leq 155^\circ$, $-50^\circ \leq \delta \leq 50^\circ$, $d_{N5\cdots O2} \leq 3.5$ Å, $d_{N4\cdots O1} \leq 3.5$ Å)	07	05
Case 3 ($d_{S\cdots O} > 3.32$ Å, , $d_{N5\cdots O2} \leq 3.5$ Å, $d_{N4\cdots O1} > 3.5$ Å)	43	19
Case 4 ($d_{S\cdots O} > 3.32$ Å, $115^\circ > \theta > 155^\circ$, $-50^\circ > \delta > 50^\circ$, $d_{N5\cdots O2} > 3.5$ Å) ($d_{N4\cdots O1} > 3.5$ Å)	176	97

N4 and N5 are the backbone amino group of i+4 and i+5 residues. O1 and O2 are backbone O of i+1 and i+2 residues. Refer to "methods" for the definition dataset I and II.

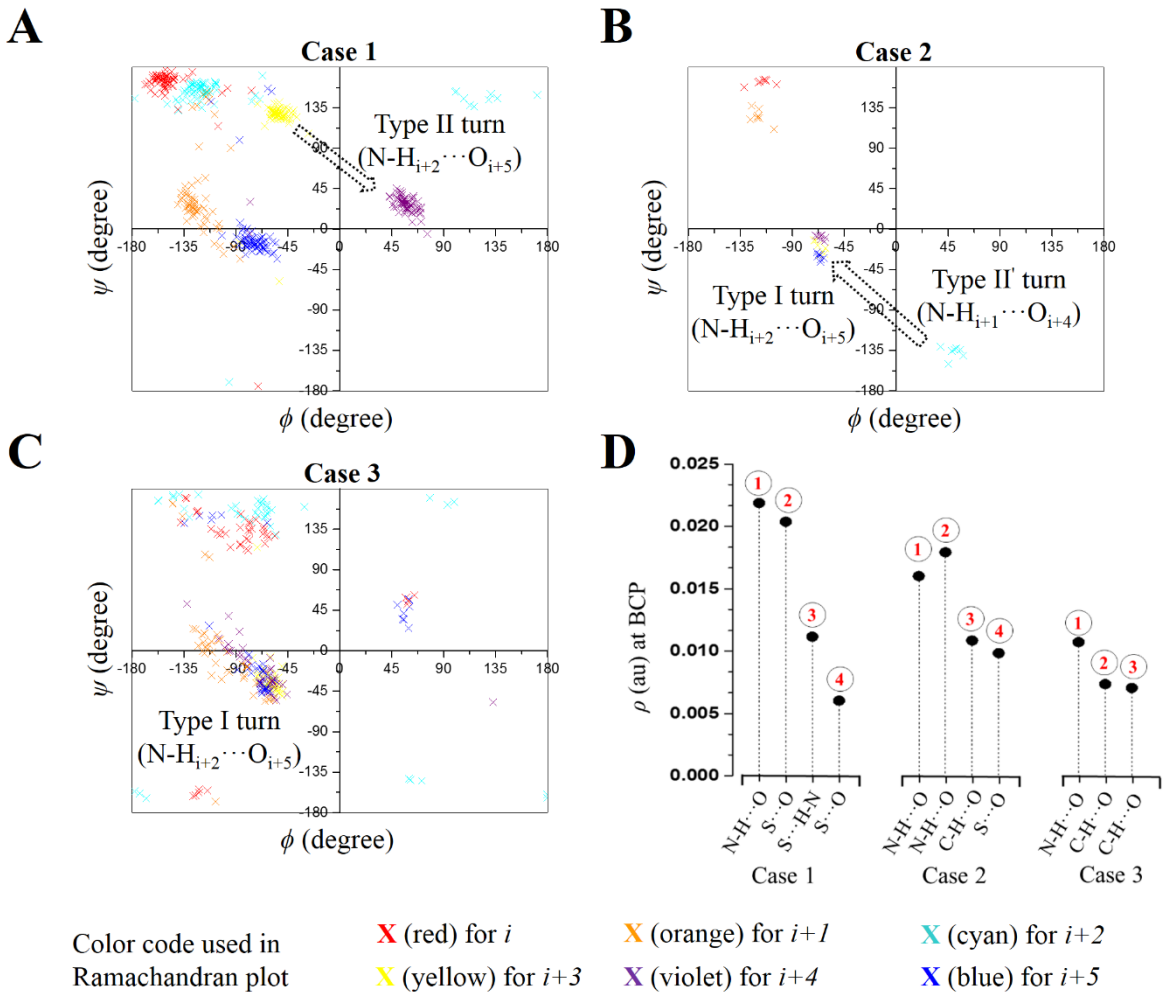


Figure 4.9 Contribution of Ch-bond to the stability β-turns. (**A-C**) *Φ* and *Ψ* values of the residues from i to i + 5 in CXXXXC motifs belonging to Cases 1-3 shown in the Ramachandran plot (refer to Figure 4.8 for cases 1-3). Structures satisfied the criteria of *resolution* ≤ 2.5 Å, R_{free} ≤ 30% and pair-wise sequence identity ≤ 90% amongst them. (**D**) A plot showing ρ (kcal.mol^{-1}) at BCP for interactions obtained using AIM analysis for Cases 1-3.

and Table 4.4) and Case 4, which did not have any β-turn (Table 4.4). This indicated that the Ch-bond played a role in the unique conformation adopted by the motifs.

To confirm this, we performed quantum chemical and AIM analyses of Cases 1 to 3 (for models, refer to Methods). Conformations of Cases 1 and 2 were found to be energetically favorable compared to the conformation in Case 3 (see E_{rel} values in Figure 4.8 and Table 4.5). Interestingly, ρ values of BCP characterizing S···O bond and N-H···O bond were comparable in Case 1 (Figure 4.9 **D** and Table 4.5). This indicated similar strengths of H- and Ch-bonds and demonstrated that the Ch-bond contributed significantly to the stability of the structural motif. In

Table 4.5 A summary of the results of AIM analysis for Case 1-3.

	Interaction No.	Interaction	ρ (kcal.mol-1) at BCP	E(SCF) in au	E_{rel} (kcal.mol^{-1})
Case1 (PDB ID: 2FD6)	1	N-H···O	0.02188	-2009.27372	-7.49
	2	S···O	0.02038		
	3	N-H···S	0.01116		
	4	S···O	0.00603		
Case2 (PDB ID: 3CEL)	1	N-H···O	0.01602	-2009.27541	-8.56
	2	N-H···O	0.01793		
	3	C-H···O	0.01085		
	4	S···O	0.00984		
Case3 (PDB ID: 1HTR)	1	N-H···O	0.01073	-2009.26177	0.00
	2	C-H···O	0.00736		
	3	C-H···O	0.00705		

Case 2, in addition to the H-bonds resulting from the backbone conformations of type I and type II' turns, the Ch-bond also contributed to the stability of the structure (Figure 4.9 **D** and Table 4.5).

4.3.5 Enzyme specificity mediated by Ch-bond

While analyzing S-mediated interactions between proteins and their ligands, we noted that the backbone amide of Leu13 (Leu13-N) and phenolic oxygen of Tyr260 (Tyr260-O$^{\eta}$) of the enzyme methionyl-tRNA synthetase (MetRS) from *Escherichia coli* (*E*.coli) were within 3.5 Å of S$^{\delta}$ its substrate methionine (Figure 4.10 **A**). The two interactions were previously described as hydrogen bonds and proposed to be essential for the substrate specificity of MetRS (Crepin et al., 2003; Schmitt et al., 2009; Serre et al., 2001; Tanrikulu et al., 2009). The interaction between Met-S$^{\delta}$ and Leu13-N satisfied the directional criteria of an H-bond (θ = 111.5° and δ = 75.8°). However, the geometry of Met-S$^{\delta}$ and Tyr260-O$^{\eta}$ (θ = 129.6° and δ = 8.3°) revealed the interaction to be a Ch-bond rather than an H-bond. Also, Tyr260-O$^{\eta}$ was H-bonded to the acceptor His301-N$^{\varepsilon}$ (His301-N$^{\delta}$ was H-bonded to the acceptor Ile297-O) and to a water molecule (Figure 4.10 **A**), implying that Tyr260-O$^{\eta}$ may not form H-bond with the acceptor Met-S$^{\delta}$. The pair of Tyr260 and His301 is highly conserved in MetRS from different species (Serre et al., 2001) but absent in other

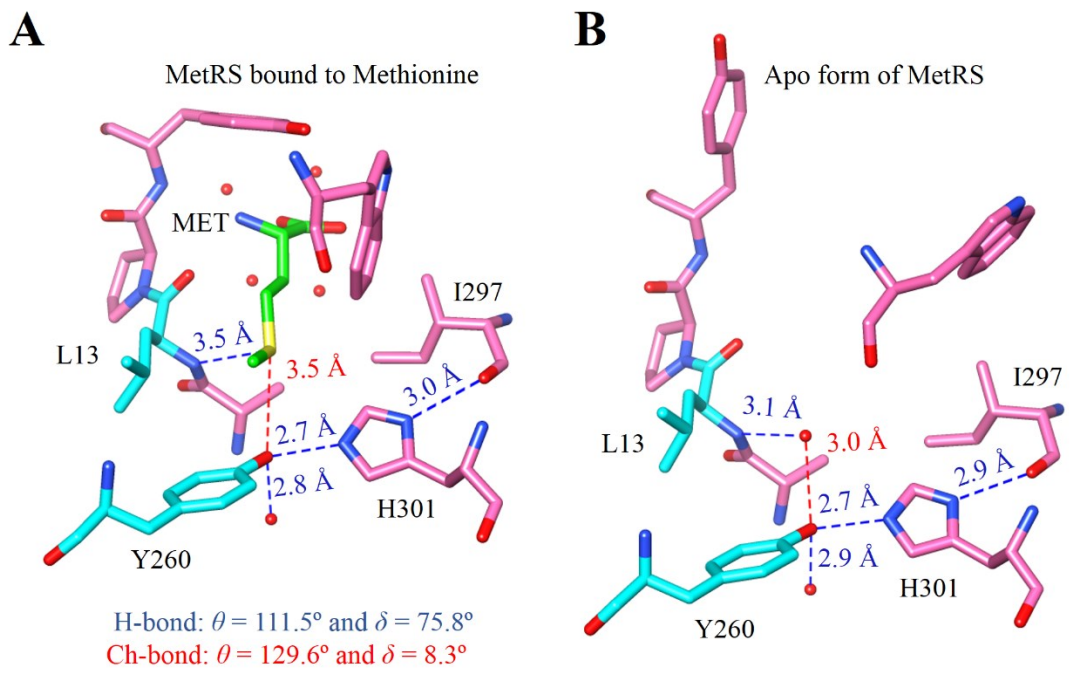


Figure 4.10 Methionine recognition by MetRS. (**A**) A zoomed view of the interactions made by MetRS with Met-S$^{\delta}$ (PDB ID: 6SPO). The H-bond and Ch-bond made by Met-S$^{\delta}$ are shown using blue and red broken lines, respectively. θ and δ values of the two interactions are also provided. (**B**) A zoomed view of the interactions made by bound water with Met-S$^{\delta}$ (PDB ID: 1QQT). Red sphere - the water molecule forming an H-bond with Tyr260.

aminoacyl-tRNA synthetases (Crepin et al., 2003), indicating the importance of the interaction network for methionine specificity.

Previously, structural comparison between apo and methionine bound MetRS showed that the methionine introduces a significant conformational change in the binding pocket of MetRS, indicative of an induced-fit mechanism for the recognition (Figure 4.10 **A-B**) (Schmitt et al., 2009; Serre et al., 2001). During this process, water molecules bound to Leu13-N and Tyr260-O$^{\eta}$ in the binding pocket of MetRS are replaced by Met-S$^{\delta}$ mediated H and Ch-bonds (Figure 4.10 **B**) (Serre et al., 2001). It appears that the exchange between these polar interactions is necessary for the specificity (Serre et al., 2001). Note that the water molecule acting simultaneously as an H-bond donor and an H-bond acceptor is energetically more favorable than being an acceptor of two H-bonds simultaneously (Ohno et al., 2005; Vennelakanti et al., 2021). In addition to this, the methionine side chain appeared to participate in hydrophobic and van der Waals interactions with other non-polar amino acids in binding pockets (Serre et al., 2001).

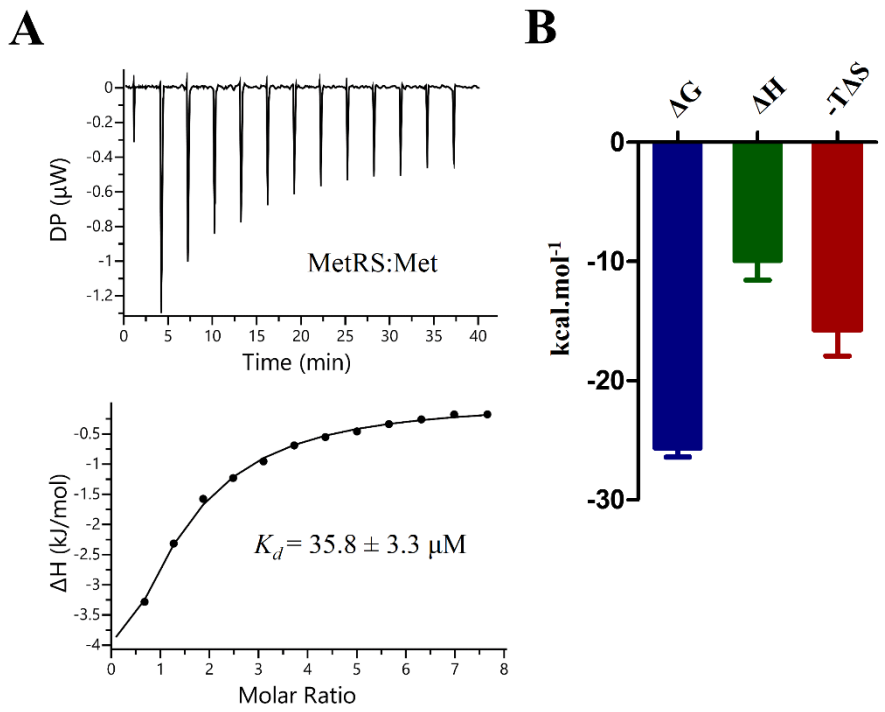


Figure 4.11 Binding of methionine to MetRS. (**A**) ITC thermograms for the titration of methionine with MetRS with the upper panel showing raw data (negative heat pulses indicates exothermic binding). The lower panel shows integrated heats, giving a differential binding curve. Data was fitted using a single-site binding model with n=1. (**B**) A plot of Δ*G*, Δ*H*, and −TΔ*S* derived from the ITC data obtained from methionine binding to MetRS.

These observations prompted us to find the energetic contribution of the hydrophobic, side-chain van der Waals packing and S-mediated interactions in methionine recognition. For this purpose, we carried out isothermal calorimetry (ITC) based titration of methionine with MetRS. We found that the binding is prominently entropy-driven and contributed ~60% to Δ*G*° (Figure 4.11 **A** and **B**). This entropic contribution could be because of the hydrophobic interaction between methionine and MetRS. Moreover, we also noted an enthalpy contribution (~ 40% to Δ*G*°), possibly because of the van der Waals and/or S-mediated polar interactions described above (Figure 4.11 **A** and **B**). Altogether suggesting the hydrophobic nature of the Met, van der Waals interaction because of side-chain packing and/or S-mediated interactions drives the association of methionine and MetRS.

Next, we sought to address if S-mediated H- and Ch-bonds are important for methionine specificity. For this purpose, we employed methionine isostere; norleucine (Figure 4.12 **A**). Although norleucine has been successfully incorporated in recombinant proteins (Cirino et al., 2003), to our best knowledge, the binding of norleucine to MetRS has not been investigated.

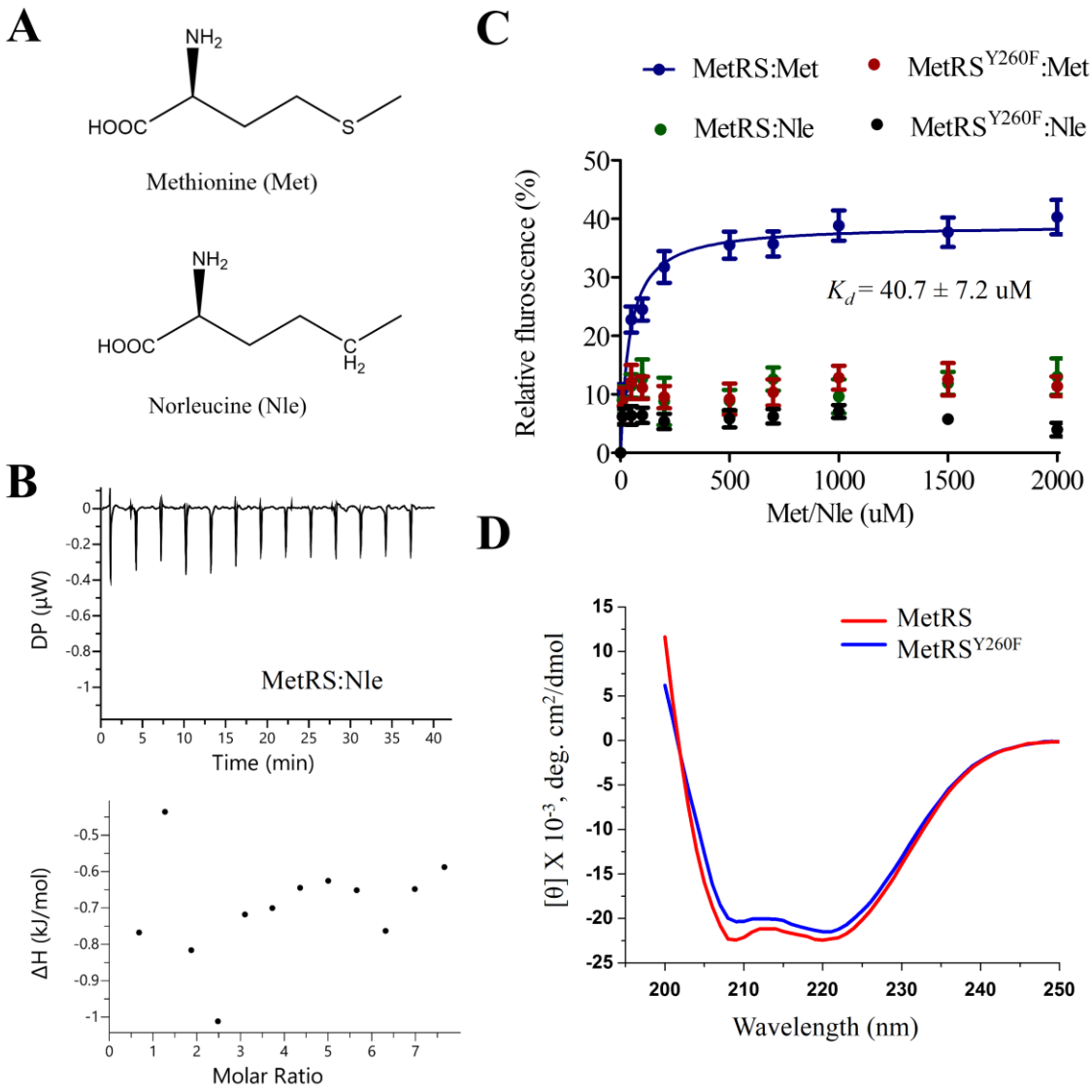


Figure 4.12 Contribution of Ch-bond for methionine specificity in MetRS. (**A**) The structure of methionine and norleucine. (**B**) ITC thermograms for the titration of norleucine with MetRS with the upper panel showing raw data. The lower panel shows integrated heats and could not fit a single-site binding model. (**C**) Plot showing the change in intrinsic fluorescence of MetRS and MetRS^{Y260F} on titration with methionine or norleucine. The values obtained for MetRS were fitted to a simple equilibrium model using the standard least-square procedure (R^2=0.98). The values obtained for others could not fit a simple equilibrium model. (**D**) Superposition of the circular dichroism (CD) spectra of MetRS and MetRS^{Y260F} indicates that the mutation did not affect the overall secondary structure of the proteins.

Norleucine being structurally similar to methionine, is used as a probe for differentiating the contribution of the hydrophobic interaction and side-chain van der Waals packing interactions from S-mediated interactions (Figure 4.12 **A**). Unlike Met, we do not observe any binding of norleucine to MetRS by ITC and intrinsic tryptophan fluorescence measurements (Figure 4.12 **B**

and **C**). This could be because of the lacking potential of norleucine-CH_2^{δ} group to replace H-bonds made by bound water with Leu13-N and Tyr260-O^{η} in the binding pocket of apo MetRS with S-mediated H- and Ch-bond (Figure 4.10 **B**). Therefore, this is clear evidence that the side chain van der Waals packing or hydrophobic interactions together are not sufficient for recognizing methionine in the absence of S-mediated interactions (Figure 4.10 and Figure 4.12 **A-C**).

To investigate the importance of the Ch-bond between Met-S^{δ} and Tyr260-O^{η}, we mutated Tyr260 to phenylalanine ($MetRS^{Y260F}$) so that the phenolic hydroxyl group was absent. The mutation was expected to disrupt the Ch-bond without affecting the H-bond between Leu13-N and Met-S^{δ} or the overall structure of the protein (Figure 4.12 **D**). Change in intrinsic fluorescence of the enzyme upon increasing concentrations of methionine was used to monitor ligand binding (Mellot et al., 1989). However, $MetRS^{Y260F}$ showed no significant change in fluorescence intensity even at high methionine concentrations (Figure 4.12 **C**), implying poor ligand binding. This demonstrated that the Ch-bond between Tyr260-O^{η} and Met-S^{δ} was critical for the substrate specificity of MetRS and the H-bond between Leu13-N and Met-S^{δ} was not sufficient for the same. In conclusion, these experiments clearly highlighted a critical role played by the Ch-bond in enzyme-substrate recognition.

4.4 Discussion

In this study, we have tried to understand the role of H- and Ch-bonds formed by divalent S in proteins. Bioinformatics analysis and computational methods showed that the S-mediated interactions contributed to the stability of protein conformation and secondary structures. Therefore, we concluded that S-mediated H- and Ch-bonds, like other weak interactions, are an essential aspect of the energy landscape in protein folding that compensates for unfavorable conformational entropy change through favorable interactions (Dobson, 2003; Grantcharova et al., 2001; Newberry and Raines, 2019).

We noted that the absence of σ-holes on S coordinated to metal favored H-bond formation, particularly with the backbone amide. As seen in the case of certain metal-chelated CXXC motifs, the H-bond appears integral to the local conformation of the polypeptide backbone. The metal coordination is structurally and functionally significant in various biologically essential proteins such as zinc-finger proteins (Laity et al., 2001), the multifunctional iron-sulfur cluster containing NADH dehydrogenase, hydrogenases, cytochrome C reductase, and those involved in DNA

replication and repair (Fuss et al., 2015). Disruption of the metal coordination, and consequently the H-bond, can affect the local backbone conformation, thus the functional activity of the protein (Saikrishnan et al., 2012). Furthermore, Met-S^{δ} and Cys-S^{γ} were found to provide stability to secondary structure elements by (**i**) capping helices through H- and Ch-bond, (**ii**) H-bond with intra-helical or inter-strand residues, and (**iii**) by H- and Ch-bonds with backbone amides of edge β-strands that could serve as an element of negative-design stabilizing β-sheets and preventing aggregation of proteins (Richardson and Richardson, 2002).

Though the occurrence of Ch-bond in biomolecules has been documented before, their functional significance had not been experimentally demonstrated. Here, through mutagenesis, we showed that an σ-hole mediated Ch-bond is integral to the specificity of the enzyme MetRS for its substrate Met. The substrate specificity of an aminoacyl-tRNA synthetase is necessary for error-free protein translation, a fundamental cellular process. Interestingly, the Ch-bond was previously described as H-bond (Serre et al., 2001). Based on the geometry of the interaction, we could identify it as a Ch-bond. We believe that interactions similar to this have not been recognized as Ch-bonds that could be essential for biomolecular functions. The geometrical parameters for S-mediated H-bond and Ch-bond reported here and previously (Iwaoka et al., 2002; Pal and Chakrabarti, 2001; Zhou et al., 2009) can also be used to model methionine and cystine in protein structures, particularly those determined using low-resolution X-ray crystallography or electron cryo-microscopy data.

An outstanding challenge of biological chemistry has been to predict the tertiary and quaternary structure of proteins *ab* initio from their primary amino acid sequence and accurately compute the thermodynamic properties of proteins, their stability, and ligand binding (Dill and MacCallum, 2012). Efforts to overcome the challenge have been hindered by our limited understanding of the forces of interaction that drive protein folding and the bonding properties of amino acids. The wide variety of interactions made by S in proteins requires that these non-covalent interactions too are considered in the energy functions used for determining protein structures, folding pathways, and binding properties. Also, the design and engineering of proteins and peptides would benefit from understanding the bonding properties of methionine and cysteine/cystine.

In summary, the study reveals that S-mediated H- and Ch-bonds contribute to structural stability and substrate specificity of proteins, very much like the interactions formed by polar amino acids. However, S-mediated interactions can have properties different from other polar non-

covalent interactions. For instance, the resistance of Ch-bond strength to solvent polarity (Pascoe et al., 2017) thus brings additional diversity to the repertoire of weak interactions essential for biomolecular functions. This could be why, despite their high biosynthetic cost (Doig, 2017), nature selected S-containing amino acids as part of the twenty building blocks of proteins.

4.5 References

Aurora, R., Rose, G.D., 1998. Helix capping. Protein Sci. 7(1), 21-38.

Bader, R.F.W., 1991. A Quantum Theory of Molecular Structure and Its Applications. Chem. Rev. 91(5), 893-928.

Bartlett, G.J., Choudhary, A., Raines, R.T., Woolfson, D.N., 2010. N→π* interactions in proteins. Nat. Chem. Biol. 6, 615–620.

Biegler-König, F., Schönbohm, J., 2002. Update of the AIM2000-program for atoms in molecules. J. Comput. Chem. 23(15), 1489–1494.

Cirino, P.C., Tang, Y., Takahashi, K., Tirrell, D.A., Arnold, F.H., 2003. Global incorporation of norleucine in place of methionine in cytochrome P450 BM-3 heme domain increases peroxygenase activity. Biotechnol. Bioeng. 83(6), 729-734.

Crepin, T., Schmitt, E., Mechulam, Y., Sampson, P.B., Vaughan, M.D., Honek, J.F., Blanquet, S., 2003. Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in Escherichia coli methionyl-tRNA synthetase. J. Mol. Biol. 332(1), 59-72.

Dennington, R., Keith, T., Millam, J., 2009. GaussView, Version 5. Semichem Inc. , Shawnee Mission. KS.

Derewenda, Z.S., Lee, L., Derewenda, U., 1995. The occurrence of C - H ⋯ O hydrogen bonds in proteins. J. Mol. Biol. 252(2), 248–262.

Dill, K.A., MacCallum, J.L., 2012. The protein-folding problem, 50 years on. Science. 338(6110), 1042–1046.

Dobson, C.M., 2003. Protein folding and misfolding. Nature. 426, 884–890.

Doig, A.J., 2017. Frozen, but no accident – why the 20 standard amino acids were selected. FEBS J. 284(9), 1296-1305.

Doig, A.J., Baldwin, R.L., 1995. N- and C- capping preferences for all 20 amino acids in α- helical peptides. Protein Sci. 4(7), 1296-1305.

Fuss, J.O., Tsai, C.L., Ishida, J.P., Tainer, J.A., 2015. Emerging critical roles of Fe-S clusters in DNA replication and repair. Biochim. Biophys. Acta - Mol. Cell Res. 1853(6), 1253-1271.

Gallivan, J.P., Dougherty, D.A., 1999. Cation-π interactions in structural biology. Proc. Natl. Acad. Sci. U. S. A. 96(17), 9459-9464

Grantcharova, V., Alm, E.J., Baker, D., Horwich, A.L., 2001. Mechanisms of protein folding. Curr. Opin.

Struct. Biol. 11(1), 70-82.

Hollingsworth, S.A., Karplus, P.A., 2010. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. Biomol. Concepts. 1(3-4), 271-283.

Iwaoka, M., Takemoto, S., Tomoda, S., 2002. Statistical and theoretical investigations on the directionality of nonbonded S···O interactions. Implications for molecular design and protein engineering. J. Am. Chem. Soc. 124(35), 10613–10620.

König, F.B., Schönbohm, J., Bayles, D., 2001. AIM2000-a program to analyze and visualize atoms in molecules. J. Comput. Chem. 22(5), 545–559.

Kantonen, S.A., Henriksen, N.M., Gilson, M.K., 2017. Evaluation and Minimization of Uncertainty in ITC Binding Measurements: Heat Error, Concentration Error, Saturation, and Stoichiometry. Biochim. Biophys. Acta - Gen. Subj. 1861(2), 485-498.

Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., Baker, D., 2012. Principles for designing ideal protein structures. Nature. 491, 222–227.

Laity, J.H., Lee, B.M., Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. 11(1), 39-46.

Lucas, X., Bauzá, A., Frontera, A., Quiñonero, D., 2016. A thorough anion-π interaction study in biomolecules: On the importance of cooperativity effects. Chem. Sci. 7, 1038-1050.

Manikandan, K., Ramakumar, S., 2004. The occurrence of C-H...O hydrogen bonds in α-helices and helix termini in globular proteins. Proteins Struct. Funct. Genet. 56(4), 768-781.

Mellot, P., Mechulam, Y., Le Corre, D., Blanquet, S., Fayat, G., 1989. Identification of an amino acid region supporting specific methionyl-tRNA synthetase: tRNA recognition. J. Mol. Biol. 208(3), 429-443.

Newberry, R.W., Raines, R.T., 2019. Secondary Forces in Protein Folding. ACS Chem. Biol. 14(8), 1677–1686.

Nick Pace, C., Martin Scholtz, J., Grimsley, G.R., 2014. Forces stabilizing proteins. FEBS Lett. 588(14), 2177-2184.

Ohno, K., Okimura, M., Akai, N., Katsumoto, Y., 2005. The effect of cooperative hydrogen bonding on the OH stretching-band shift for water clusters studied by matrix-isolation infrared spectroscopy and density functional theory. Phys. Chem. Chem. Phys. 7, 3005-3014.

Pal, D., Chakrabarti, P., 2001. Non-hydrogen bond interactions involving the methionine sulfur atom. J. Biomol. Struct. Dyn. 19(1), 115–128.

Pascoe, D.J., Ling, K.B., Cockroft, S.L., 2017. The Origin of Chalcogen-Bonding Interactions. J. Am. Chem. Soc. 139(42), 15160-15167.

Richardson, J.S., Richardson, D.C., 2002. Natural -sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. 99(5), 2754-2759.

Saikrishnan, K., Yeeles, J.T., Gilhooly, N.S., Krajewski, W.W., Dillingham, M.S., Wigley, D.B., 2012. Insights into Chi recognition from the structure of an AddAB-type helicase-nuclease complex. EMBO J. 31(6), 1568-1578.

Schmitt, E., Tanrikulu, I.C., Yoo, T.H., Panvert, M., Tirrell, D.A., Mechulam, Y., 2009. Switching from an Induced-Fit to a Lock-and-Key Mechanism in an Aminoacyl-tRNA Synthetase with Modified Specificity. J. Mol. Biol. 394(5), 843-851.

Serre, L., Verdon, G., Choinowski, T., Hervouet, N., Risler, J.L., Zelwer, C., 2001. How methionyl-tRNA synthetase creates its amino acid recognition pocket upon L-methionine binding. J. Mol. Biol. 306(4), 863-876.

Shelke, S.S., 2020. Divalent Sulfur mediated interactions in proteins architecture, stability, and molecular recognition. Master dissertation, IISER, Pune.

Tanrikulu, I.C., Schmitt, E., Mechulam, Y., Goddard, W.A., Tirrell, D.A., 2009. Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo. Proc. Natl. Acad. Sci. U. S. A. 106(36), 15285-15290.

Venkatachalam, C.M., 1968. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers. 6(10), 1425-1436.

Vennelakanti, V., Qi, H.W., Mehmood, R., Kulik, H.J., 2021. When are two hydrogen bonds better than one? Accurate first-principles models explain the balance of hydrogen bond donors and acceptors found in proteins. Chem. Sci. 12, 1147-1162.

Zhou, P., Tian, F., Lv, F., Shang, Z., 2009. Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. Proteins Struct. Funct. Bioinforma. 76(1), 151–163.

Chapter 5

The role of van der Waals dispersion interaction in base-specific protein-DNA recognition: McrBC as a case study

5.1 Introduction

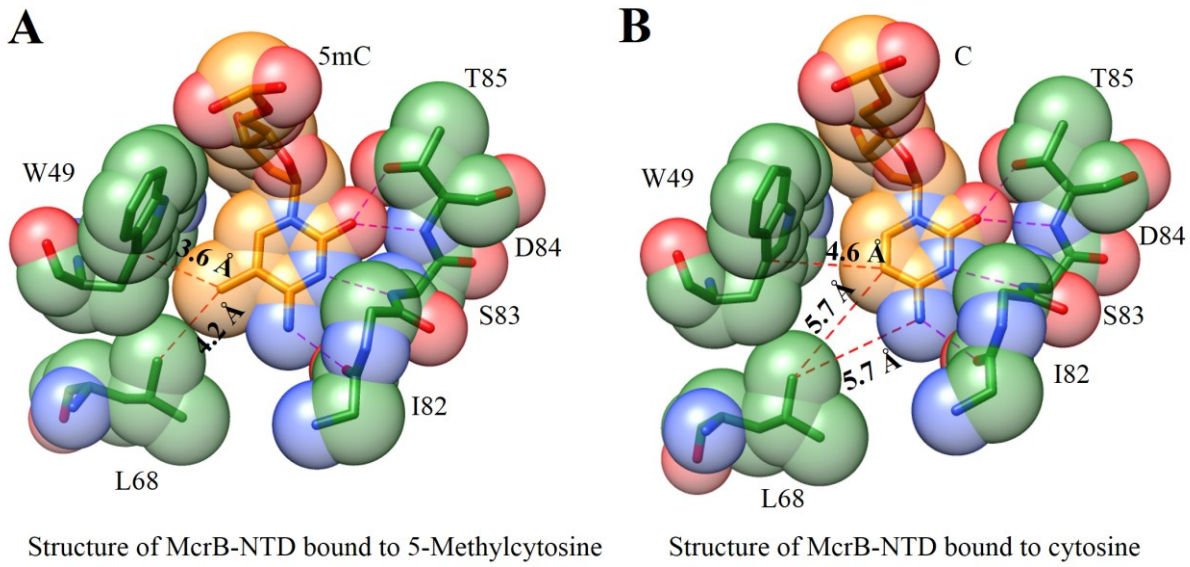
The human genome comprises ~2.9 billion base pairs, of which > 1.5% code for proteins and the rest are considered "junk" mainly because of insufficient knowledge about their functional purpose (Abdellah et al., 2004; Craig Venter et al., 2001). The ability of cells to efficiently distinguish between and extract the information from these protein-coding and non-coding DNA sequences is a puzzle for the biologist (Rohs et al., 2010). The underlying mechanism for this has broad implications in molecular biology. In 1953, Watson and Crick determined the 3D structure of the DNA (Watson and Crick, 1953), and afterward, many crystal structures were determined for the protein bound to DNA (Luscombe et al., 2000; Sagendorf et al., 2020). These structures provided an initial glimpse of the mechanism for target specificity in protein-DNA recognition. In particular, the topology of the DNA and proteins appeared to contribute dominantly (Garvie and Wolberger, 2001; Rohs et al., 2009). For instance, the enhancement in the target specificity could result from the local and global structural deformation of DNA. This results in its A-, B- or Z-form, bending, kinking, and narrowing minor groove conformations (Rohs et al., 2010). A mechanism by which proteins recognize such deformations in the DNA is commonly referred to as shape readout (Rohs et al., 2010). On the other hand, the 70 superfamilies of the DNA binding proteins share common DNA binding motifs: Helix-turn-helix, helix-loop-helix, β-barrel, zinc finger, or leucine zipper are some representative examples (Garvie and Wolberger, 2001; Rohs et al., 2010). These motifs often have electrostatic complementary to a target sequence of the DNA. Additionally, these motifs form direct/indirect conventional hydrogen bonds (H-bonds) or hydrophobic interactions with DNA bases complementary to their amino acids, by the mechanism referred to as base readout (Rohs et al., 2010; Seeman et al., 1976). Thus, shape and base readouts are the most common conserved features noticed in almost all protein-DNA complexes. Together, considered a key in explaining DNA target specificity (Rohs et al., 2010).

Furthermore, some enzymes often use a common mechanism referred to as the DNA base flipping that enables direct access to the base to carry out or check the status of chemical modifications on it (Cheng and Blumenthal, 1996; Hong and Cheng, 2016; Hopkins and Reich, 2004; Roberts and Cheng, 1998; Várnai and Lavery, 2002). In this case, the target base swings out from the intra-helical to the extra-helical region (Roberts and Cheng, 1998). The mismatch or damaged bases in duplex DNA often have altered mechanical and chemical properties than the regular Watson-Crick base pairs (Moe and Russu, 1992). These properties affect DNA breathing or base-pair stability and introduce DNA deformation that initiates the binding of DNA repair enzyme, in turn, base flipping (Chen and Prohofsky, 1995; Cheng and Blumenthal, 1996). On the

other hand, some sequence-dependent proteins induce DNA distortion by kinking or bending the target sequence, thus facilitating the base flipping (Huang et al., 2003). The extrusion of a target base can occur either from the minor or major groove of DNA (Cheng and Blumenthal, 1996). Often, recognition pockets of these enzymes complement the target base in terms of its electronic or steric properties and aid with key H-bonds or stacking interactions between the flipped base and protein residues. Thus, they together are proposed to establish target base specificity (Roberts and Cheng, 1998).

Numerous crystal structure-based reports have identified the contribution of base flipping in DNA target specificity. For instance, the crystal structures of methyltransferases such as HhaI, HaeIII, TaqI, or PvuII that identify potential target bases to carry out methylation provided initial evidence for this mechanism (Hong and Cheng, 2016). Also, Dioxygenases ten-eleven translocation (TET) enzymes that oxidize 5mC or T4 phage glucosyltransferases that glucosylate 5hmC also use this mechanism for DNA base modification (Hardwick et al., 2018). The SRA domain from UHRF1, which is crucial for maintaining the methylation pattern of DNA, recognizes the methylated base by flipping mechanism (Arita et al., 2008). The DNA repair enzymes, such as T4 endonuclease V and UDG, remove damaged bases to maintain genome integrity (Dodson et al., 1994; Schormann et al., 2014). Furthermore, target specificity of type I-III restriction enzymes such as EcoKI, EcoRI, and EcoP15I rest on the base flipping mechanism (Allan et al., 1999; Chand et al., 2015; Gupta et al., 2015; Su et al., 2005).

The *Escherichia coli* McrBC is a modification-dependent restriction enzyme (Panne et al., 2001; Sutherland et al., 1992). McrB has an N-terminal DNA binding domain (McrB-NTD) that recognizes its target sequence, RmC) (R=A or G and mC=modified cytosine). The modified cytosine can either be 5-methylcytosine (5mC), 4-methylcytosine (4mC), or 5-hydroxy-methylcytosine (5hmC) (Sukackaite et al., 2012; Zagorskaitė et al., 2018). The C-terminal domain of McrB has a AAA+ fold that binds to GTP (Nirwan et al., 2019b; Panne et al., 2001). McrB oligomerizes as a hexameric ring in the presence of GTP (Nirwan et al., 2019b). McrC carries out nucleolytic cleavage by complexing with McrB in the presence of GTP (Nirwan et al., 2019a; Niu et al., 2020). The active McrBC complex requires at least two G/A(5mC) target sites separated by more than 30 bp for its nuclease activity (Nirwan et al., 2019b). The previously determined structure of the *E. coli* McrB-NTD bound to DNA revealed that the protein uses a base flipping mechanism to identify its target base, 5-methylcytosine, which is preceded by either guanine or adenine (Sukackaite et al., 2012). The structure reveals that the architecture of the base recognizing pocket of McrB complements its target 5mC in terms of its electronic and steric



Structure of McrB-NTD bound to 5-Methylcytosine Structure of McrB-NTD bound to cytosine

Figure 5.1 A base recognizing pocket of McrBC. Crystal structures showing the flipped (**A**) 5-Methylcytosine and (**B**) cytosine into base recognizing pocket of McrB-NTD (PDB ID 3SSE and 3SSC, respectively). The polar interactions between 5mC or cytosine and pocket residues are highlighted in magenta, whereas non-polar interactions are shown in red dotted lines. The H-bond distances are below 3.5 Å. van der Waals (vdW) models for these structures are also presented using the vdW radius for each atom defined previously (Tsai et al., 1999) and implemented in the Chimera program.

properties (Sukackaite et al., 2012). For instance, this pocket is incompatible with accommodating purines, presumably because of the barrier created by Trp49 residue (Figure 5.1 **A**). The side chain and backbone H-bond network between the flipped base and pocket residues ensure the specificity for 5mC over thymine (Figure 5.1 **A**). Interestingly, non-polar residues such as Leu68 and Trp49 of the binding pocket of McrB have close contact with the methyl group of flipped 5mC (Figure 5.1 **A**). We believe that this close contact results from the C-H···π and van der Waals dispersion (vdW) interaction ($CH_3 \cdots CH_3$ contact) of the methyl group of 5mC with C^{γ} of Trp49 and $C^{\delta 1}$ of Leu68, respectively (Figure 5.1 **A**) (Brandl et al., 2001; Li et al., 2018).

Using ultra-high resolution structures of proteins, Li et al. noted that the favorable distance for vdW C···C contact usually peaked at 4.05 or 5.35 Å (Li et al., 2018). Thus, the distance of 4.2 Å for the contact between the methyl group of 5mC and $C^{\delta 1}$ of Leu68 is a clear indication for vdW dispersion interaction between McrB and its target base 5mC (Figure 5.1 **A**). The lack of contact between the cytosine (C5) and Leu68 (the distance is 5.7 Å between them) in the structure of McrB-NTD bound to non-methylated DNA could support this argument (Figure 5.1 **B**). However, this structure is also accomplished by the loss of C-H···π interaction as the distance increase from

3.6 to 4.6 Å when 5mC changed to cytosine (Figure 5.1 **A** and **B**). Hence, we speculated that both these weak interactions, together, could discriminate 5mC from cytosine. Their absence could generate a void (Figure 5.1 **B**) that might be unstable (Xue et al., 2019), together thus must be responsible for the target base specificity in McrBC endonuclease.

Although it is well established that the vdW dispersion interaction could stabilize the proteins (Holder et al., 2001; Joh et al., 2009; Misra et al., 2021; Ratnaparkhi and Varadarajan, 2000). However, unlike C-H···π or other weak interactions (Chakrabarti and Samanta, 1995; Mandel-Gutfreund et al., 1998; Wilson et al., 2014), its role in the biomolecular recognition is unclear. In particular, a recent NMR-based study showed that the contribution of the vdW dispersion interactions behind the association of non-polar molecules are negligible in the solvent-accessible system and might not have a significant role in molecular recognition (Yang et al., 2013). Additionally, there is a lack of evidence for this interaction in base-specific protein-DNA interaction. Since the target base specificity in the McrBC enzyme results from the interactions between non-polar residues Leu68, Trp49, and 5mC/C without the participation of water solvent. Hence, we ask if this vdW dispersion interaction could contribute to the target base specificity in the McrBC enzyme. If so, can we change the target base specificity of this endonuclease by modulating the vdW dispersion interaction?

We addressed the questions mentioned above using computational, mutagenesis, biochemical, and crystallographic investigations. We showed that the introduction of dispersion interaction in the McrB pocket resulted in specificity for the cytosine, even in lack of coexisting C-H···π interaction. Also, from the high-resolution crystal structures of various mutants of McrB, we noted that this pocket of McrB is not rigid, unlike other enzymes (Kimber et al., 2014), and connected to the intercalating residue by water-mediated H-bonds and thus, could stabilize the flipped C. Interestingly, this water-mediated H-bond network and its cooperation with dispersion interactions is appeared to be critical for the cytosine specificity. Together, these observations highlighted the importance of weak vdW dispersion interactions that are often overshadowed in the light of the other non-covalent interaction in base-specific protein-DNA interaction. Thus, it could be of broad implications in designing and engineering an active pocket of an enzyme that uses a base flipping mechanism to create their variants/library with diverse biochemical functions.

5.2 Materials and methods

5.2.1 Computational analysis

The geometries of various dimers mentioned in this study were obtained from the X-ray structures of the McrB-NTD bound to DNA to understand the nature of bonding and estimate their interaction energies. The reliable position of the H atoms in these dimers was obtained using the Gaussian09 program (Frisch et al., 2009) by employing the B3LYP-D3 functional (Becke, 1993; Grimme, 2006; Lee et al., 1988) and 6-311G++(3D, 3P) basis set. During this geometry minimization process, the position/coordinate of the other heteroatoms was kept frozen; otherwise, it could lead to a different structure resulting from other strong forces such as H-bond and π-π stacking. This also mimics the geometry of these dimers similar to that of crystal structure, thus providing a suitable way to understand the strength or nature of the interaction.

The intermolecular interaction energy of these dimers (ΔE_{AB}) was calculated using Equation 1, which accounts for basis set superposition error (BSSE), as suggested by Boys and Bernadi (Boys and Bernardi, 1970; Řezáč and Hobza, 2016).

$$\Delta E_{AB} = E_{AB} - E_A^{AB} - E_B^{AB} \qquad (4)$$

The BSSE correction was estimated using the counterpoise correction method employed in G09 programs. The contribution of the vdW dispersion interaction in BSSE corrected total ΔE for the dimers was estimated using B3LYP-D3 (Becke, 1993; Grimme, 2006; Lee et al., 1988) and wb97xd (Becke, 1997; Chai and Head-Gordon, 2008; Grimme, 2006; Wu and Yang, 2002) methods by comparing with the B3LYP.

Next, to visualize and understand the nature of the interactions in various dimers studied here, we performed non-covalent interaction (NCI) indexing using the NCIPLOT4 program (Contreras-García et al., 2011; Johnson et al., 2010). This analysis utilizes the properties of the electron density, such as the reduced gradient of density, *s* defined as (Contreras-García et al., 2011; Johnson et al., 2010),

$$S = \frac{1}{2\sqrt[3]{3\pi^2}} \frac{|\nabla\rho|}{\sqrt[3]{\rho^2}}$$

Here, ρ is the electron density. To know the non-covalent interaction if attractive or repulsive, the sign of an eigenvalue λ_2 of the electron density Hessian matrix, $\nabla^2\rho=\lambda_1+\lambda_2+\lambda_3$ is used, where $\lambda_1<\lambda_2<\lambda_3$. The sign of λ_2 is negative for attractive and positive for repulsive interaction. In addition,

Table 5.1 List of the DNA primers used in this chapter.

Primer name	Sequence (5' – 3')
T7 Promoter	TAATACGACTCACTATAGGG
T7 Terminator	GCTAGTTATTGCTCAGCGG
mbnrev	ATGATGATGATGATGATGACCTTAGAGGTGGAATAACACTC
MB60MSPI-1F	GCCGGGTAACCGGGTAAGTCCGGGTAAGA(5mC)CGGTAGTTCGG ATCGAGGGGTAGGCCGC
MB60MSPINM-2R	AGTCAAATTGCATATGCTGGTCTTTCAGCGCCGGTAATCGTCTTG TGAAGGATCCGCGGC
MB-N-13bp-F	TGAGACCGGTAGC
MB-N-13bp-R	AGCTACCGGTCTC
MB-N-13bp5mC-F	TGAGA(5mC)CGGTAGC

the quantity sign(λ_2)ρ defines the strength of the interaction (Contreras-García et al., 2011; Johnson et al., 2010).

5.2.2 Molecular dynamics (MD) simulations

The mutants of McrB-NTD bound to DNA after removing the bound BTB ligand were subjected to MD simulations using Gromacs 5.0 software package (Abraham et al., 2015). The AMBER94 force field, as suggested previously for protein-nucleic acid complexes, was employed for this purpose. (Cornell et al., 1996). The systems were solvated in a cubic water box using the TIP3w water model. To neutralize the system, 100 mM KCl was added by replacing water molecules. In order to remove the steric clash, energy minimization of the system was carried out using the steepest descent algorithm and 50,000 cycles. This energy minimized system was further equilibrated into NVT and NPT phases for 500 ps. The temperature (300 K) and pressure (1 bar) of each system were maintained using Vrescale, a modified Berendsen thermostat temperature coupling method (Berendsen et al., 1984), and the Parrinello-Rahman pressure coupling method (Parrinello and Rahman, 1981), respectively. In the end, the productive MD simulation for each system was carried out using these well-equilibrated systems for 20 ns at 300 K and 1 bar. The RMSD and H-bond analysis was carried out VMD program (Humphrey et al., 1996).

5.2.3 Cloning of Full length and N-terminus McrB (McrB-NTD) constructs

For full-length McrB mutants, forward primer carrying the desired mutation and T7 terminator primer were used to amplify the corresponding gene from wild type McrB, previously cloned in pHIS17 vector (Nirwan et al., 2019b). These amplified products were subsequently cloned in the pHIS17 vector using the restriction-free cloning method (Mishra, 2017; Nirwan et al., 2019b; Zacharia, 2020). A similar approach was used to clone McrB-NTD mutants, where 1-160 amino acids carrying genes were amplified from full-length McrB mutants (Zacharia, 2020). For this

purpose, we used T7 promoter and MBN-reverse primers (Table 5.1). All positive clones were fully sequenced to confirm desired mutations. All primers used in this study are presented in Table 5.1.

5.2.4 Purification of McrC, full length, and N- McrB constructs

All Full-length McrB or McrB-NTD constructs had C-terminus 6X His tag while McrC was without any tag. All purifications steps were performed as described previously (Mishra, 2017; Nirwan et al., 2019b; Zacharia, 2020). A plasmid of each construct was transformed into *E. coli* T7 Express LysY/Iq (C3013I, NEB) or *E. coli* BL21(AI) grown in 2 L of LB containing 100 ug/ml ampicillin at 37°C. After OD reached ~0.6 at 600 nm, the temperature of the shaker was reduced to 18°C, and cultures were induced with 0.5 mM IPTG or 0.06% w/v L-arabinose. These induced cultures were grown at 18°C for overnight (12-16 h), and cells were pelleted (5000 rpm, 20 minutes, 4°C).

The pellets were suspended in 50 ml lysis buffer (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 25 mM imidazole, 5 mM $MgCl_2$, 10% (v/v) glycerol) and was lysed by sonication at 4°C. The cell lysate was ultra-centrifuged at 4°C at 37000 rpm for 45 minutes. This supernatant obtained was loaded onto a 5 ml Ni-NTA column pre-equilibrated with Buffer A (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 25 mM imidazole). Protein was eluted using a step gradient of Buffer B (500 mM NaCl, 50 mM Tris-Cl pH 8.0, 500 mM imidazole) by 5 % to 100%. Purest fractions, inferred from 12% SDS PAGE gel, were pooled together. For full-length McrB constructs and McrC, these fractions were dialyzed against 2 L of B50 buffer (50 mM NaCl, 50 mM Tris-Cl pH 8.0, 1 mM DTT). This dialyzed protein was then centrifuged at 18000 rpm at 4°C for 20 mins.

Dialyzed full-length McrB mutants were loaded onto an 8 ml MonoQ 10/100GL column (GE Life Sciences) equilibrated with Buffer B50 (50 mM NaCl, 50 mM Tris-Cl pH 8.0, and 1 mM DTT). Fractions of 1.5 ml were collected in 20 column volumes over a linear gradient of 0 to 50% of Buffer B1000 (1000 mM NaCl, 50 mM Tris-Cl pH 8.0, and 1 mM DTT). The pure fractions were then pooled and concentrated using 10 kDa vivaspin2 concentrator (GE Life Sciences). The protein solution was washed with B100 (50 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM DTT) and stored at -80°C. In case of McrC, Dialyzed Ni-NTA fractions were loaded onto an 8 ml MonoS 10/100 GL column (GE Life Sciences) equilibrated with Buffer B50. Fractions of 1.5 ml were collected in 20 column volumes over a linear gradient of 0 to 50% of Buffer B1000. The pure fractions were then pooled and concentrated using 10 kDa vivaspin2 concentrator (GE Life Sciences). The protein solution was washed with B100 and stored at -80°C.

In case of McrB-NTD constructs, the purest Ni-NTA fractions were dialyzed against 2 L of Buffer C (10 mM Tris-Cl pH 8.0, 0.2 M KCl, 0.1 mM EDTA, and 1 mM DTT). Dialyzed protein was then centrifuged at 4°C for 20 mins. The supernatant was loaded onto a 24 ml Superdex75 column (GE Life sciences) equilibrated with Buffer C. Fractions containing protein were concentrated and stored at -80°C in Buffer C (Sukackaite et al., 2012).

5.2.5 DNA cleavage assay

Two DNA substrates *viz*. 1.2 kb or 114 bp linear DNA was used for cleavage assay throughout this study. 1.2 kb DNA was amplified by PCR using T7 Promoter and T7 Terminator primers from the pHIS17 vector containing the MBΔN gene (Mishra, 2017; Nirwan et al., 2019b; Zacharia, 2020). Deoxyribonucleotide triphosphates (dNTP) mix containing either 2'-deoxy-5-methylcytidine 5'-triphosphate (5-methyl-dCTP) or dCTP was used to generate methylated or non-methylated 1.2 kb DNA, respectively. While 114 bp DNA was generated as mentioned previously (Nirwan et al., 2019b) using MB60MSPI-1F and MB60MSPI-2R primers either containing two (2X) or single (1X) R(5mC) sites.

The nucleolytic cleavage of DNA was carried out in a 10 µl reaction mixture containing cleavage buffer (10 mM Tris-Cl pH 8, 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT) as described previously (Mishra, 2017; Nirwan et al., 2019b; Zacharia, 2020). A 75 nM of 114 bp DNA with 75 nM of McrBC complex (4:1 molar ratio of McrB to McrC) or 75 ng of the 1.2 kb DNA with 50 nM of the McrBC complex was incubated in the presence or absence of 1 mM GTP (Jena Bioscience) at 37°C for 1 h. The reaction was stopped by adding 2 µl of 6X STES buffer (40% (w/v) sucrose, 0.2 M Tris–Cl pH 7.5, 40 mM EDTA, 1% (w/v) SDS) followed by denaturation by heating at 65°C for 15 minutes. The cleavage reactions with 114 bp DNA were loaded on a pre-electrophoresed 10% native PAGE gel. The gel was run at 150 V in 1X TBE buffer and stained with a solution containing 2 µg/ml ethidium bromide for 5 min. The cleavage reaction with the 1.2 kb DNA was loaded on a 0.8 % agarose gel and was run at 110 V in 1X TAE buffer. All gels were imaged on an E-Gel imager (Thermo Fisher Scientific).

Quantification of nuclease activity was done using the ImageJ program. The fraction of uncut DNA (1X R(5mC) 114 bp DNA) left in each reaction was used to calculate the nuclease efficiency (also referred to as the percentage of DNA substrate cleaved) by McrB mutants. Four separate and independent reactions were carried out to calculate their nuclease efficiencies (Zacharia, 2020).

5.2.6 Fluorescence anisotropy measurements

Fluorescence anisotropy measurements were carried out in a HORIBA FluoroMax 4 spectrophotometer. As mentioned previously, 13 bp DNA was used for measurements (Zagorskaitė et al., 2018). One of the 5' ends of this duplex oligo was labeled with 5-(Iodoacetamido) fluorescein (5-IAF) (Sigma-Aldrich) using the protocol as described earlier (Zearfoss and Ryder, 2012). All measurements were carried out at 25°C with excitation wavelength 492 nm and emission wavelength 515 nm (5 nm slit width for each). 50 nM of labeled 13 bp was used to monitor change in anisotropy, and binding reactions were carried out in 150 ul of anisotropy buffer (20 mM MES-KOH pH 6.0, 100 mM KCl, 1 mM DTT). The concentration of McrB-NTD mutants was increased from 62.5 nM to 15 μM. Every anisotropy value was taken as an average of 10 readings, each with an integration time of 1 second. The binding curve was plotted using the software GraphPad Prism 5.0 using the equation Y = Bmax*X/(KD + X), where Y is the anisotropy change calculated by subtracting from unbound labeled 13 bp DNA. X is the McrB-NTD mutant's concentration, and Bmax is the estimated saturation value of anisotropy increase during each experiment. The curve was plotted by averaging data points from three or more independent experiments (Zacharia, 2020).

5.2.7 Crystallization of McrB-NTD with 13 bp DNA and data collection

Crystallization of all McrB-NTD mutants bound to 13 bp DNA was carried out as described previously (Sukackaite et al., 2012). McrB-NTD mutant stored in Buffer C (10 mM Tris-Cl pH 8.0, 0.2 M KCl, 0.1 mM EDTA and 1 mM DTT) was mixed in 1.6:1 molar ratio with DNA (160 uM McrB-NTD and 100 uM DNA). Crystallization was carried out using the hanging drop vapor diffusion method, and 1 ul protein:DNA complex solution was mixed with 1 ul of crystallization buffer (0.1-0.2 M Bis-Tris pH 5.5 and 12-24 % PEG 4000). Crystallization was carried out at 25°C, and all crystals reached their maximum size in 21 days. Prior to diffraction experiments, crystals were transferred to crystallization buffer, additionally containing 25% PEG 400 and flash cryo-cooled at 100 K. Diffraction data for each McrB-NTD mutant bound to DNA was collected at IO4 beamline at Diamond Light Source, UK. All collected data were processed using MOSFLM (Battye et al., 2011) and AIMLESS (Evans and Murshudov, 2013) programs except for the $McrB^{L68F}$-NTD -DNA complex. This data was processed using XDS (Kabsch et al., 2010), SCALA (Kabsch et al., 2010), TRUNCATE (French and Wilson, 1978) because of its highly mosaic nature.

Table 5.2 Crystallographic data-collection and refinement statistics. Values in parentheses are for the highest resolution shell.

Complex	McrB^{L68Y}-NTD and 13 non-methylated DNA	McrB^{L68F}-NTD and 13 bp non-methylated DNA	McrB^{L68F}-NTD and 13 bp hemi-methylated DNA	McrB$^{Y41F-L68Y}$-NTD and 13 bp non-methylated DNA	McrB$^{Y41F-L68F}$-NTD and 13 bp non-methylated DNA
Wavelength	0.9795	0.97950	0.9795	0.9762	0.9795
Resolution range (Å)	62.2 -1.55 (1.60 - 1.55)	37.1-1.8 (1.86 - 1.80)	39.7- 2.1 (2.18 - 2.10)	39.5 - 2.1 (2.18 - 2.10)	49.67 - 1.65 (1.71 - 1.65)
Space group	$P2_12_12_1$	$C121$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	36.1, 68.9, 143.8, 90, 90, 90	74.7, 36.1, 142.8, 90, 98.6, 90	36.1, 70.3, 144.2, 90, 90, 90	35.9, 68.4, 144.8, 90, 90, 90	36.1, 68.9, 143.3, 90, 90, 90
Multiplicity	12.1 (12.3)	3.9 (3.4)	4.4 (4.2)	4.1 (4.3)	9.1 (8.5)
Completeness (%)	99.9 (99.6)	99.4 (98.9)	97.9 (96.90)	94.4 (93.0)	100 (100)
Mean I/sigma(I)	15.7 (2.8)	8.0 (1.3)	7.7 (2.0)	6.4 (1.2)	12.4 (1.8)
R_{merge}	0.076 (0.869)	0.086 (0.770)	0.117 (0.678)	0.12 (0.853)	0.085 (0.736)
CC1/2	0.999 (0.916)	0.995 (0.710)	0.993 (0.536)	0.98 (0.603)	0.998 (0.824)
Reflections used in refinement	53047 (5216)	35145 (3400)	21633 (2091)	20233 (1931)	44059 (4302)
Reflections used for *R*-free	5293 (492)	1733 (197)	1125 (101)	1978 (180)	4301 (401)
R_{work}	0.1636 (0.2589)	0.1844 (0.2939)	0.1857 (0.2710)	0.1789 (0.2918)	0.1632 (0.2326)
R_{free}	0.1983 (0.2910)	0.2198 (0.3355)	0.2452 (0.3285)	0.2362 (0.3648)	0.1973 (0.2738)
RMS(bonds)	0.007	0.008	0.005	0.008	0.007
RMS(angles)	0.92	1.04	0.76	0.96	0.89
Ramachandran favored (%)	96.42	96.03	97.37	96.71	96.73
Ramachandran allowed (%)	3.58	3.97	2.63	3.29	3.27
Ramachandran outliers (%)	0	0	0	0	0
Clash score	4.71	5.19	4.71	4.69	6.86
Average *B*-factor (Å^2)	34.83	42.38	38.34	39.59	33.73
macromolecules	34.18	41.55	38.12	39.59	32.65
ligands	34.13	53.3	39.06	41.55	37.52
solvent	42.29	48	41.16	39.4	43.49

Table 5.3 A Chain wise RMSD for all mutant McrB-NTD structures determined in this study. A PDB ID: 3SSE was used for WT McrB-NTD

RMSD		WT		L68F		L68F*		L68Y		Y41F-L68F		Y41F-L68Y	
		A	B	A	B	A	B	A	B	A	B	A	B
WT	A	0											
	B		0										
L68F	A	0.39		0									
	B		0.39		0								
L68F*	A	0.57		0.60		0							
	B		0.43		0.27		0						
L68Y	A	0.26		0.32		0.54		0					
	B		0.29		0.42		0.50		0				
Y41F-L68F	A	0.28		0.32		0.58		0.14		0			
	B		0.25		0.29		0.52		0.12		0		
Y41F-L68Y	A	0.3		0.38		0.59		0.18		0.22		0	
	B		0.28		0.45		0.51		0.17		0.19		0

*Structure of McrB^{L68F}-NTD bound to hemi-methylated DNA

5.2.8 Structure solution of McrB-NTD bound to 13 bp DNA

All protein-DNA complexes were crystallized in the $P2_12_12_1$ space group except the McrB^{L68F}-NTD-DNA complex, which crystallized in the C121 space group (Table 5.2). Structures were solved by molecular replacement using the PHASER molecular replacement program (McCoy, 2006), and previously reported structure (PDB ID 3SSE) was used as a search model for all cases. Multiple cycles of model building and refinement were carried out using COOT (Emsley et al., 2010) and *phenix.refine* programs (Afonine et al., 2012), respectively. Except in the McrB^{L68F}-NTD and DNA complex, densities for the terminal residue of DNA in all other structures were poor, and no unique orientation was determined; hence two alternate orientations of DNA were built. All structural images were made using Chimera program version 1.15 (Pettersen et al., 2004). van der Waals (vdW) model for the structural snapshot shown in Figures 5.1 and 5.5 prepared using the vdW radius for each atom defined previously (Tsai et al., 1999) and implemented in the Chimera program (Pettersen et al., 2004). A summary of crystallographic data-collection and refinement statistics is provided in Table 5.2. The RMSD values for structural comparison of all the crystal structures determined in this study are presented in Table 5.3.

5.3 Results

5.3.1 van der Waals dispersion interaction distinguishes 5-methylcytosine from cytosine in McrB

We first sought to understand the nature and strength of the $CH_3 \cdots CH_3$ interaction between the methyl group of 5mC and $C^{\delta 1}$ of Leu68. For this purpose, the coordinates of the heteroatoms of the dimer made of Leu68 and 5mC were obtained from the PDB ID 3SSE, H atoms modelled and the structure energy minimized. We next calculated the basis set superposition error (BSSE) corrected interaction energy (ΔE) with and without dispersion components. Figure 5.2 **A** summarizes the results obtained. Use of computational method with the dispersion component (B3LYP-D3/6-311G++(3D, 3P) resulted in ΔE = -0.87 kcal.mol^{-1} and computational method without the dispersion component (B3LYP/6-311G++(3D, 3P) resulted in ΔE = 0.24 kcal.mol^{-1} (Figure 5.2 **A**). We used a similar strategy to calculate the interaction strength between the Leu68 and cytosine. The analysis suggested that this dimer was only marginally stable, as ΔE was equal to -0.23 kcal.mol^{-1} in comparison to the value for Leu:5mC dimer (Figure 5.2 **A**).

In addition to this, we also carried out the non-covalent interaction (NCI) analysis, which relies on the electron density and its second derivative for the interacting dimers. The sign of the second eigenvalue (λ_2) of the Hessian matrix and magnitude of density indicates the nature ($\lambda_2 <$ 0 corresponding to attractive and $\lambda 2 > 0$ corresponding to repulsive) and strength of the interaction (Johnson et al., 2010). The plot of the reduced density gradient, $s(\rho)$, and density isosurface for Leu:5mC dimer clearly showed that $CH_3 \cdots CH_3$ interaction stabilized this dimer and the magnitude of sign(λ_2)ρ for this interaction was ~ -0.01 au, which correlates to its interaction energy of ~ -0.9 kcal.mol^{-1} (Figure 5.2 **B**). Whereas, in the case of Leu:cytosine dimer, the sign(λ_2)ρ values were ~ 0.00 au and could be correlated to its relatively poor interaction energy of -0.23 kcal.mol^{-1} (Figure 5.2 **C**). Together, these observations supported the hypothesis that the vdW dispersion interaction could help the McrBC enzyme to distinguish 5-methylcytosine from cytosine.

A

Method and basis set	BSSE corrected interaction energies (ΔE) in kcal.mol^{-1}	
	Leucine:5-methylcytosine dimer	Leucine:cytosine dimer
B3LYP/6-311++G(3D, 3P)	0.24	0.03
B3LYP-D3/6-311++G(3D, 3P)	-0.87	-0.23
wb97xd/6-311++G(3D, 3P)/auto	-0.88	-0.24

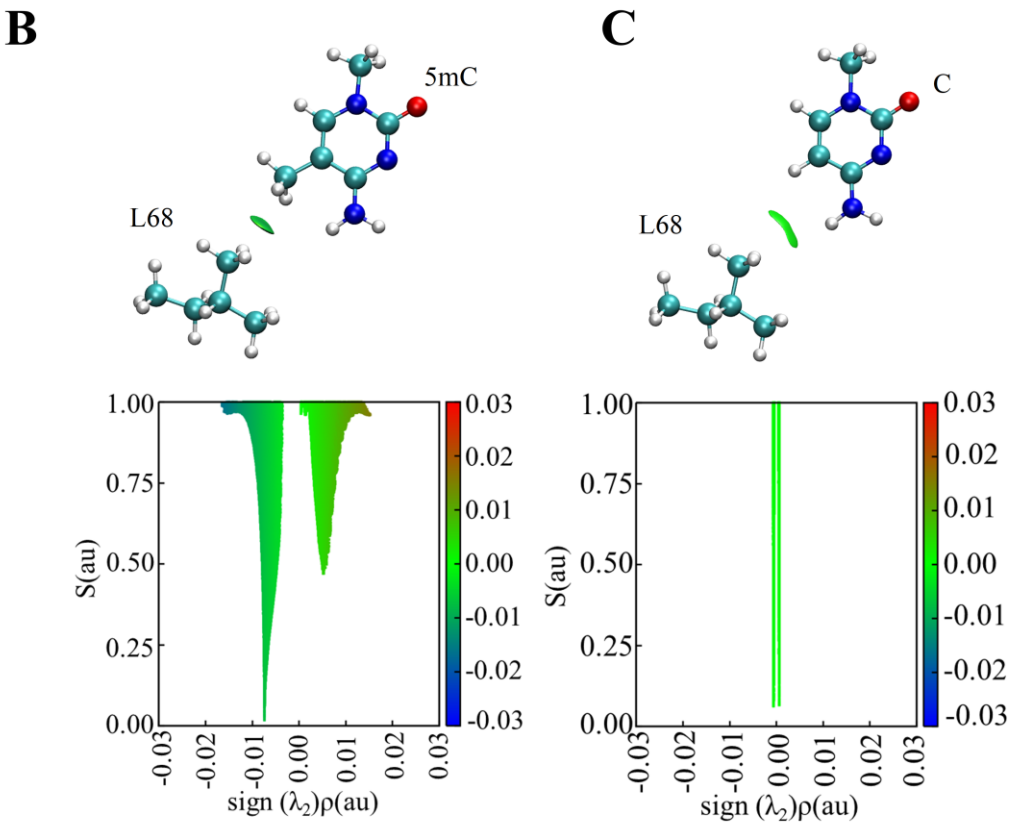


Figure 5.2 Computational analysis for the interaction of leucine with 5-methylcytosine or cytosine. (**A**) BSSE corrected interaction energies (ΔE) in kcal.mol^{-1} for leucine:5-methylcytosine and leucine:cytosine dimers using B3LYP, B3LYP-D3, and wb97xd functional. A non-covalent interaction (NCI) analysis of (**B**) leucine:5-methylcytosine and (**C**) leucine:cytosine dimers showing the plot of the NCI isosurface (s= 0.5) colored according to the RGB scheme over the range of $-0.03 < sign(\lambda_2)\rho < 0.03$. Blue indicates attractive, green indicates vdW interaction, and red indicates repulsion interaction.

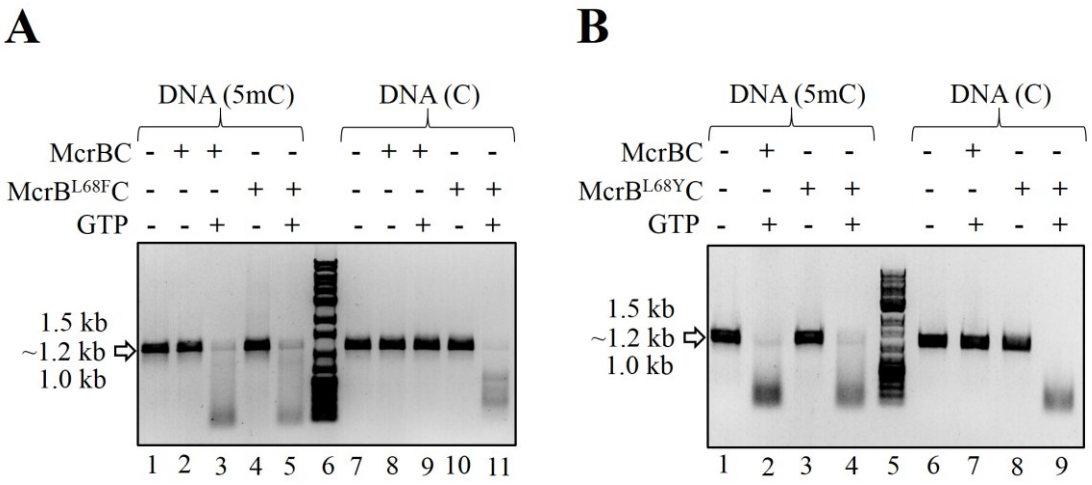


Figure 5.3 DNA cleavage by McrB^{L68F}C and McrB^{L68Y}C complexes. A representative 0.8 % agarose gel of DNA cleavage assay of (**A**) McrB^{L68F}C and (**B**) McrB^{L68Y}C complexes with methylated (5mC) and non-methylated ~1.2 kb DNA substrate. Refer to Materials and Methods for details.

5.3.2 Broadening target base specificity by mutagenesis

We next sought a role for Leu68 in McrB in target base specificity by mutating it to amino acids having different charge and/or size. As the nuclease activity depends on the recognition of the target DNA sequence, wild type McrB and the McrB mutants in complex with McrC were tested for their ability to cleave a substrate DNA. Among the mutants of McrB, the polar mutant proteins were insoluble. This probably was because of the introduction of a polar residue in the non-polar base recognizing pocket of McrB. Interestingly, the L68F mutation (referred to as McrB^{L68F}) was toxic to *E.coli* BL21 (AI) expression cells. We suspected that the nuclease activity of this mutant degraded the genomic DNA. The suspicion arose because the *E.coli* BL21 (AI) expression cells had endogenous McrC. The complexation of McrBL68F with McrC might have degraded the host genomic DNA with non-methylated target sites. We could overcome this issue by using *E. coli* T7 Express LysY/Iq (C3013I, New England Biolabs Inc.), a Δ*mcrC* expression strain. The strain was used for purifying the mutant (see methods).

We next tested the nucleolytic activity of the McrB^{L68F}C mutant and compared it to the wild type McrBC enzyme. For this purpose, a 1.2 kb methylated DNA substrate carrying multiple R(5mC) sites was subjected for cleavage with McrB^{L68F}C. This complex retained the GTP-dependent nuclease activity for methylated DNA similar to the McrBC complex (Figure 5.3 **A**). This complex cleaved the DNA even when it was non-methylated (Figure 5.3 **A**). Note that the

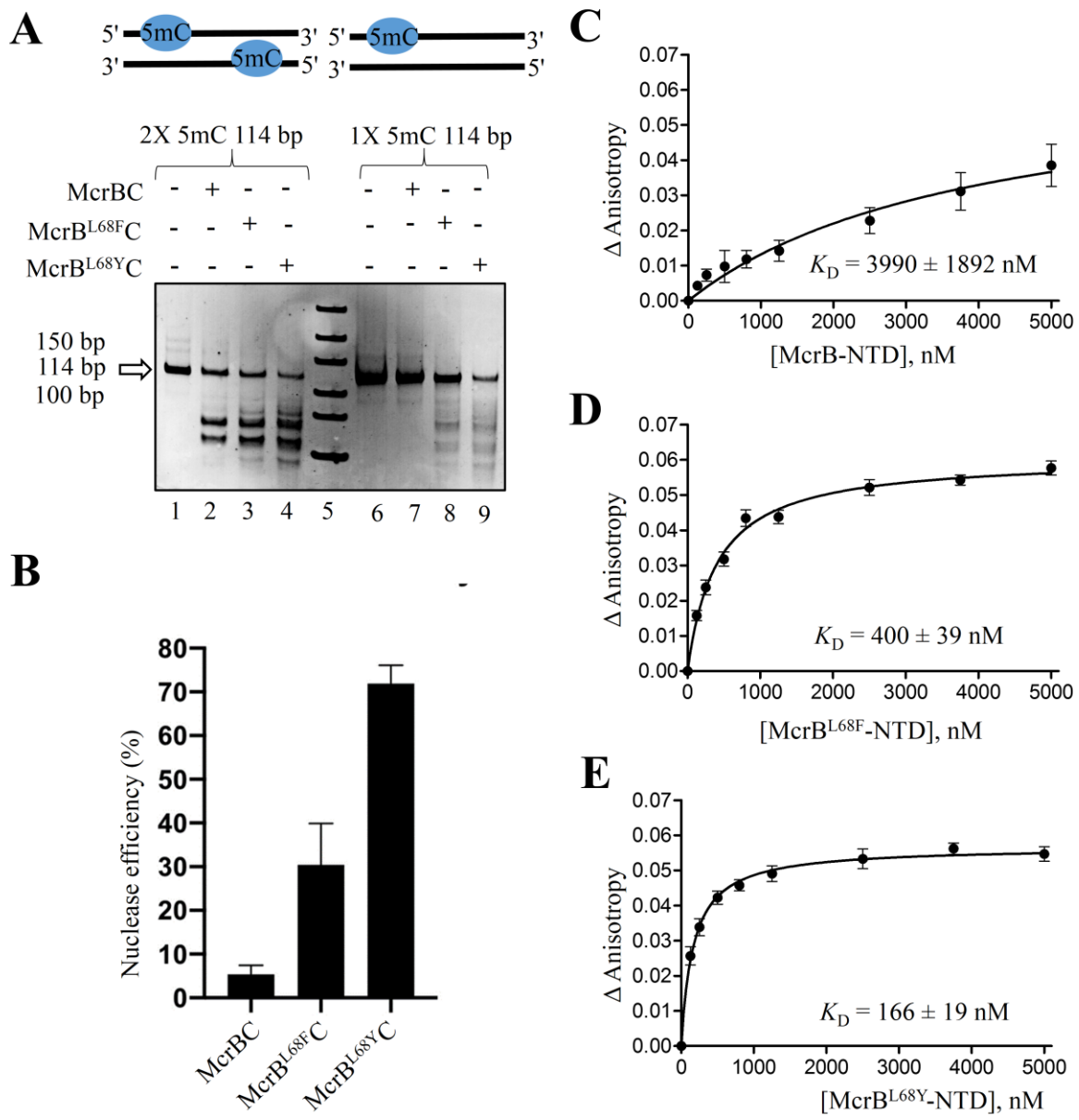


Figure 5.4 Nuclease and binding efficiencies of McrBC, McrB^{L68F}C, and McrB^{L68Y}C complexes. (**A**) The top panel shows a schematic representation of 114 bp used for DNA cleavage assay containing a single 5mC site. The bottom panel shows a representative 10 % native PAGE gel for DNA cleavage assay of McrBC, McrB^{L68F}C McrB^{L68Y}C with the 114 bp DNA. (**B**) A histogram showing nuclease efficiency of McrBC, McrB^{L68F}C, and McrB^{L68Y}C towards the 114 bp DNA containing a single 5mC. The nuclease efficiency of each complex was calculated using four separate and independent DNA cleavage reactions. A plot of anisotropy change as a function of varying concentrations of (**C**) McrB-NTD, (**D**) McrB^{L68F}-NTD, and (**E**) McrB^{L68Y} proteins. These anisotropy measurements were carried out using a covalently linked 5-IAF fluorescence label to 13 bp DNA, and each data point represents an average of at least three independent experiments. The binding constant (K_D) for each is also presented and calculated from non-linear curve fitting using Graphpad Prism 5.0 program.

Wild type McrBC does not cut non-methylated DNA. The cleavage pattern of the non-methylated DNA was, however, different from the methylated DNA substrate (Figure 5.3 **A**). The partial cleavage of this DNA substrate by the McrB^{L68F}C complex could have been because of its weaker nuclease activity towards the non-methylated DNA (Figure 5.3 **A**).

Phenylalanine has a larger size compared to leucine. Thus, we concluded that the McrB^{L68F} mutation stabilized the flipped cytosine as it snuggly fitted into the base recognition pocket of McrB, which now had a reduced size due to the mutation. To see if this is true, we next introduced Tyr at position 68 because of its similar structure to Phe, except for the additional hydroxyl (-OH$^{\eta}$) group. McrB^{L68Y}C, as expected, retained its nuclease activity for non-methylated DNA (Figure 5.3 **B**). Interestingly, the cleavage pattern of methylated and non-methylated DNA was similar (Figure 5.3 **B**). The McrB^{L68Y}C enzyme appeared nucleolytically more efficient when compared to the McrB^{L68F}C enzyme (Figures 5.3 **A** and **B**).

To quantify the nuclease efficiency of these mutant enzymes, we employed a 114 bp DNA substrate containing two R(5mC) sites (referred to as 2X 5mC 114 bp) or a single R(5mC) site (referred to as 1X 5mC 114 bp) for cleavage assays (Figure 5.4 **A**, top panel) (Nirwan et al., 2019b). Both McrB^{L68F}C and McrB^{L68Y}C appeared as efficient as the wild type McrBC enzyme in cleaving 114 bp DNA, having two R(5mC) 114 bp DNA (Figure 5.4 **A)**. Interestingly, McrB^{L68F}C and McrB^{L68Y}C enzymes were also active against the 1X 5mC 114 bp DNA substrate, unlike the wild type enzyme, which does not cleave non-methylated DNA (Figure 5.4 **A**, the bottom panel). We quantified their cleavage activity towards 1X 5mC 114 bp DNA substrate and compared it with McrBC (Figure 5.4 **B**). While McrB^{L68F}C could cleave 30 % of the non-methylated DNA, McrB^{L68Y}C could cleave as much as 70 % (Figure 5.4 **B**). In summary, the ability of McrB^{L68F}C and McrB^{L68Y}C to cleave methylated as well as non-methylated DNA suggests their broadened target DNA specificity in comparison to the wild type enzyme.

5.3.3 The higher binding affinity of mutant enzymes for non-methylated DNA

We next sought to quantify the affinities of McrB^{L68F}C and McrB^{L68Y}C enzymes for non-methylated DNA. For this purpose, we used the N-terminal domain of McrB^{L68F} and McrB^{L68Y} (referred to as McrB^{L68F}-NTD and McrB^{L68Y}-NTD) because of its primary role in DNA recognition and binding (Sukackaite et al., 2012). A previous study showed that the affinity of the wild type McrB-NTD for the R(5mC) site containing 13 bp DNA was 19-fold higher than the equivalent non-methylated DNA (Zagorskaitė et al., 2018). We carried out binding studies using the same 13 bp DNA with covalently linked fluorescent molecule (5-IAF) at one of 5' sites.

Anisotropy measurements showed that the K_D for binding of McrB-NTD to this DNA was 3990 ± 1892 nM (Figure 5.4 **C**). In the case of McrB^{L68F}-NTD, we found that the affinity towards non-methylated DNA increased by 8 folds compared to McrB-NTD (Figure 5.4 **C** and **D**). Also, McrB^{L68Y}-NTD further enhanced binding by 16 folds (Figure 5.4 **C** and **E**). Interestingly, we noted that the K_D value for binding McrB-NTD to methylated DNA (K_D = 209 ± 90 nM) reported previously (Zagorskaitė et al., 2018) was comparable to the K_D value for binding of McrB^{L68Y}-NTD to non-methylated DNA, which was 166 ± 19 nM (Figure 5.4 **E**).

5.3.4 Structural basis for the broadened specificity of McrB mutants

To identify the molecular basis for the specificity towards the non-methylated DNA by McrB^{L68F}and McrB^{L68Y}, we crystallized their NTD bound to the 13 bp non-methylated DNA (same as in PDB ID: 3SSE). All the crystals were grown in conditions similar to those of wild-type McrB-NTD bound to non-methylated DNA (PDB ID: 3SSE). The structure of McrB^{L68F}-NTD bound to the 13 bp non-methylated DNA was determined at a resolution of 1.8 Å (Table 5.2). The structure revealed a shorter spacing between Phe68 and the flipped cytosine (Figure 5.5 **A**). The distance between Phe-C$^{\zeta}$ and Cyt-N4 was 3.9 Å, while the distance between Phe-C$^{\varepsilon 1}$ and Cyt-C5 was 4.5 Å (Figure 5.5 **A**). The distances for the corresponding contacts in the structure of wild type McrB-NTD bound to non-methylated DNA were 5.7 Å each (Figure 5.1 **B**). Comparing these two structures clearly demonstrated that the replacement of leucine by the larger side chain of phenylalanine at position 68 of McrB filled the void when cytosine occupied the base recognition pocket (Figure 5.5 **A** and Figure 5.1 **B**). This led us to conclude that the shorter distance between McrB-Phe68 and cytosine helped stabilize the flipped cytosine when bound to McrB^{L68F}-NTD, resulting in a ~8-fold increase in the affinity for non-methylated DNA (Figure 5.4 **C** and **D**). This observation, in turn, also provided a rationale for the nuclease activity of McrB^{L68F}C towards non-methylated DNA (Figure 5.3 **A**).

As the McrB^{L68F}C mutant enzyme was also nuclease active against methylated DNA, we wanted to know how Phe68 residue interacts with 5mC. To address this, we determined the crystal structure of the N-terminal variant of this mutant bound to a 13 bp hemimethylated DNA at a resolution of 2.1 Å (Table 5.2). We did not notice any significant change in the conformation of either Phe68 or 5mC (Figure 5.5 **A-B** and Figure 5.1 **B**). All the interactions made by Phe68 with 5mC and C were conserved. The sole exception was a much shorter distance of 3.6 Å between the C$^{\varepsilon 1}$ of Phe68 and the methyl group of 5mC (Figure 5.5 **B**).

We next tried to rationalize the higher affinity of McrB^{L68Y}-NTD for the non-methylated DNA in comparison to both McrB-NTD and McrB^{L68F}-NTD. For this, we determined the structure of McrB^{L68Y}-NTD bound to the 13 bp non-methylated DNA at a resolution of 1.55 Å (Table 5.2). The structure of McrB^{L68Y}-NTD bound to DNA revealed a conventional H-bond between O$^{\eta}$ of

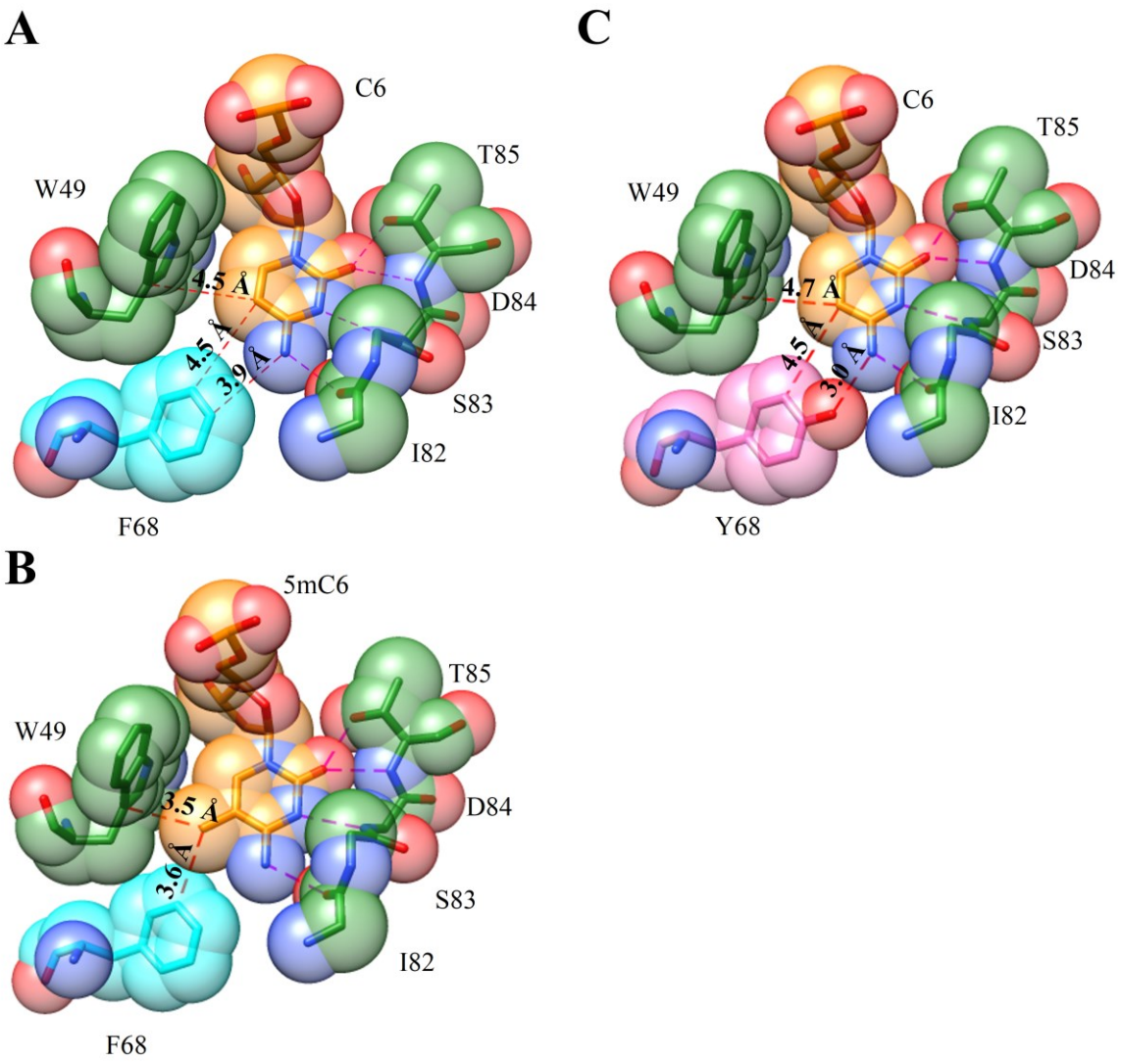


Figure 5.5 Crystal structures revealing the binding pocket of McrB^{L68F} and McrB^{L68Y} mutants. Crystal structures showing the flipped (**A**) cytosine and (**B**) 5mC into the base recognizing pocket of McrB^{L68F}-NTD. (**C**) Crystal structures showing the flipped 5mC into the base recognizing pocket of McrB^{L68Y}-NTD. The polar interactions between cytosine or 5mC and pocket residues (all below 3.5 Å) are highlighted in magenta, whereas non-polar interactions are shown in red. van der Waals (vdW) models for these structures are also presented, and vdW radius for each atom is defined previously **(**Tsai et al., 1999**)** and implemented in the Chimera program.

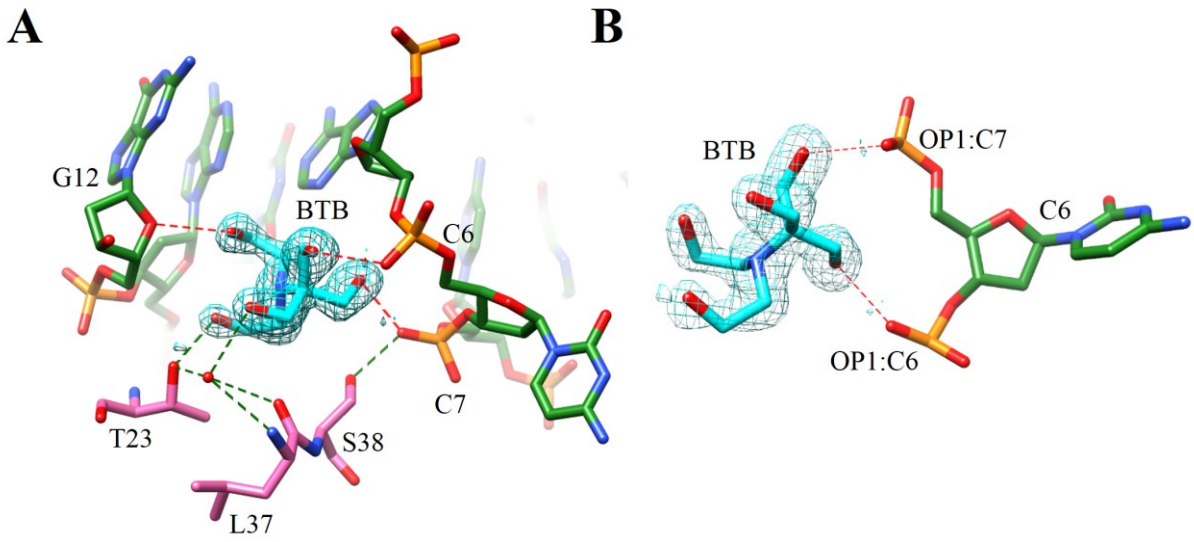


Figure 5.6 The structure of BTB bound at the interface of McrB-NTD and DNA. (**A**) A snapshot showing a bound Bis-Tris (PDB ID: BTB) in the crystal structure of McrB^{L68Y}-NTD and 13 bp DNA at resolution 1.55 Å. A segmented Fo – Fc electron density map at 1.2σ contour level for BTB is also shown. The direct or water-mediated H-bonds (d < 3.5 Å) of BTB with protein residues are highlighted as green, whereas those with DNA are as red dotted lines. (**B**) A zoomed section showing the direct H-bonds between BTB and OP1 of the phosphate group linking Cyt6 and Cyt7.

Tyr68 and N4 of cytosine (H-bond distance was 1.55 Å) in addition to the vdW contact between their C$^{\varepsilon 1}$ and C5 atoms having a distance of 4.5 Å (Figure 5.5 **C**). The additional H-bond formed by the phenolic hydroxyl group provided higher stability to the flipped cytosine. Thus, the higher affinity or nuclease activity of the McrB^{L68Y}C enzyme towards non-methylated DNA could be explained in light of this unique H-bond (Figures 5.4 **D-E** and Figure 5.4 **B**). While determining the crystal structures of DNA-bound McrB-NTD and its mutants, we noticed a strong electron density in the F_O-F_C map at the protein and DNA interface. We interpreted this electron density to be that of Bis-Tris from the crystallization condition (Figure 5.6 **A**). The direct or water-mediated H-bond network formed by Bis-Tris with the protein and DNA could contribute to the stability of the protein-DNA complex (Figure 5.6 **B**).

5.3.5 vdW dispersion interaction contributes to cytosine specificity

Next, to know the contribution of Phe68 or Tyr68 contacting cytosine in stabilizing the flipped state, we performed a computational analysis similar to that discussed in the previous section. Figure 5.7 **A** provides a summary of the analysis. The interaction energy for Phe:cytosine dimer was -0.97 kcal.mol^{-1}, which was comparable to the Leu:5mC dimer. The interaction energy was dominated by the dispersion component. This suggested that the Phe residue could stabilize

A

Method and basis set	BSSE corrected interaction energies (ΔE) in kcal.mol^{-1}	
	Phenylalanine:cytosine dimer	Tyrosine:cytosine dimer
B3LYP/6-311++G(3D, 3P)	0.81	-2.43
B3LYP-D3/6-311++G(3D, 3P)	-0.97	-4.69
wb97xd/6-311++G(3D, 3P)/auto	-0.71	-4.17

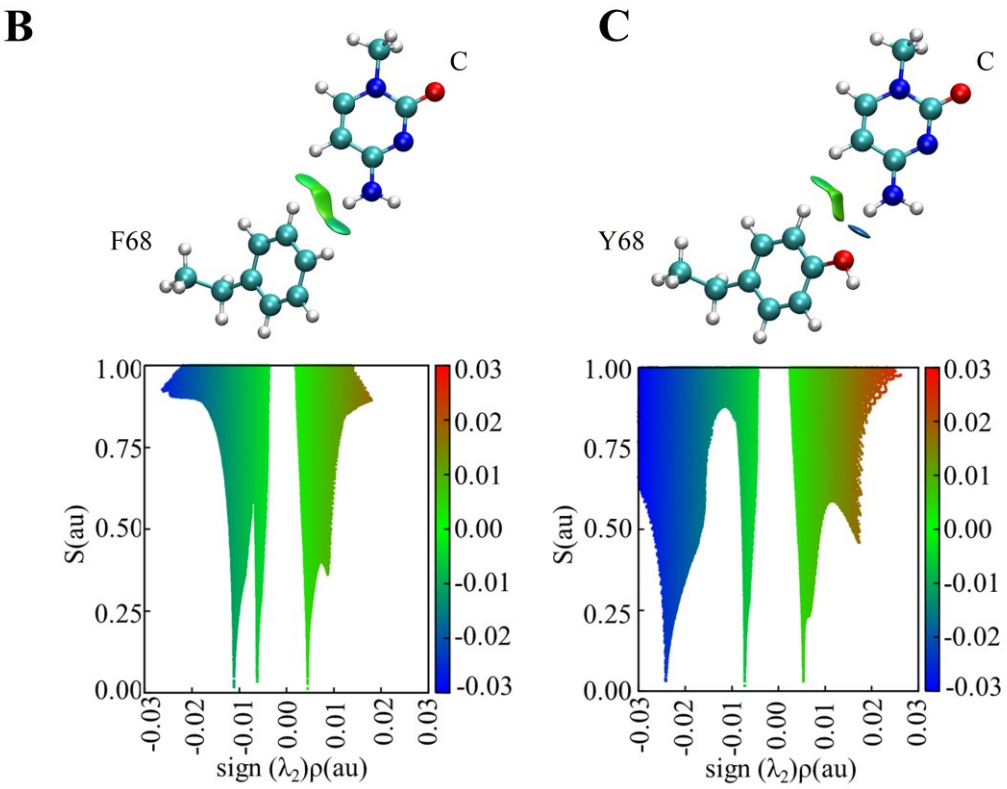


Figure 5.7 Computational analysis for the interaction of phenylalanine or tyrosine with cytosine. (**A**) BSSE corrected interaction energies (ΔE) in kcal.mol^{-1} for Phe:cytosine and Tyr:cytosine dimers using B3LYP, B3LYP-D3, and wb97xd functional. A non-covalent interaction (NCI) analysis of (**B**) Phe:cytosine and (**C**) Try:cytosine dimers showing the plot of the NCI isosurface (s= 0.5) colored according to the RGB scheme over the range $-0.03 < sign(\lambda_2)\rho < 0.03$. Blue indicates attractive, green indicates vdW interaction, and red indicates repulsion interaction.

cytosine similar to Leu stabilizing 5mC in the WT scenario and facilitating McrB^{L68F}C to cleave a non-methylated DNA. The structure of the DNA bound McrB-NTD also revealed the presence of C-H···π interaction between the aromatic ring of Tyr49 and the methyl group of 5mC that could contribute to the specificity for 5mC (Figure 5.1 **A**). In contrast, the structure of the non-methylated DNA-bound McrB^{L68F}-NTD did not have the C-H···π interaction (Figure 5.5 **A**). This led us to conclude that the Phe:cytosine interaction alone was sufficient for cytosine specificity.

The NCI analysis indicated that the H-$C^{\zeta}\cdots$H-N4 (sign(λ_2)ρ = ~ -0.011 au) and H-$C^{\varepsilon 1}$ and H-C5 (sign(λ_2)ρ = ~ -0.007 au) vdW dispersion interactions stabilized the dimer of Phe68:cytosine (Figure 5.7 **B)**. Similar NCI analysis for the Tyr:cytosine dimer, revealed an additional H-bond between H-O^{η} $\cdots$H-N4 with sign(λ_2)ρ = ~ -0.024 au that correlates with its increased interaction energy of -4.69 kcal.mol^{-1} (Figure 5.7 **A** and **C**).

5.3.6 A critical role of intercalating residue towards cytosine specificity

The above analysis revealed that the replacement of McrB-Leu68 with a larger hydrophobic amino acid, i.e., phenylalanine or tyrosine, resulted in broadening of the specificity of McrB. Upon mutation, McrB could bind to both methylated and non-methylated DNA with comparable affinities. In an effort to make McrB specific to only non-methylated DNA, we mutated McrB-Leu68 to tryptophan, which has an even larger hydrophobic side chain. However, the expression of McrB^{L68W} resulted in the protein being insoluble upon lysis. As an alternate strategy, we looked for a residue whose mutation would potentially alter the specificity of McrB to bind only non-methylated DNA. This led us to focus on the residue McrB-Tyr41 that inserts into the cavity in the DNA formed upon base-flipping of the cytosine (Sukackaite et al., 2012). Such intercalating residues are known to stabilize the flipped state of the base (Sukackaite et al., 2012, Hendershot and Brien, 2014).

We wondered if mutation of McrB-Tyr41 would preferentially stabilize a flipped cytosine over 5-methylcytosine. A previous study had shown that mutation of McrB-Tyr41 to the much smaller amino acid alanine reduced the affinity for methylated DNA. This is probably because of the ability of Tyr41 residue to introduce deformation in the DNA, which in turn is responsible for the recognition of the target site on DNA by a mechanism commonly referred to as indirect readout (Martin et al., 1999; Sukackaite et al., 2012). However, we found that the double mutant McrB$^{L68F\text{-}Y41A}$ could recognize 5mC but not C (Figure 5.8 **A**). Based on this observation, our preliminary hypothesis was that mutation of McrB^{L68F}-Tyr41 to the much larger tryptophan would push the flipped cytosine further into its base-specificity pocket, which would stabilize the flipped cytosine while sterically blocking the larger 5-methylcytosine from entering the pocket. The mutation, however, resulted in the formation of inactive protein since it was not able to recognize 5mC as well as cytosine, possibly in the misfolded state (Figure 5.8 **B**).

In parallel, we mutated McrB^{L68F}-Tyr41 to phenylalanine, an amino acid that is structurally similar to tyrosine except for the absence of the side chain hydroxyl group. Interestingly, the double mutant McrB$^{Y41F\text{-}L68F}$ did not cleave non-methylated DNA while still

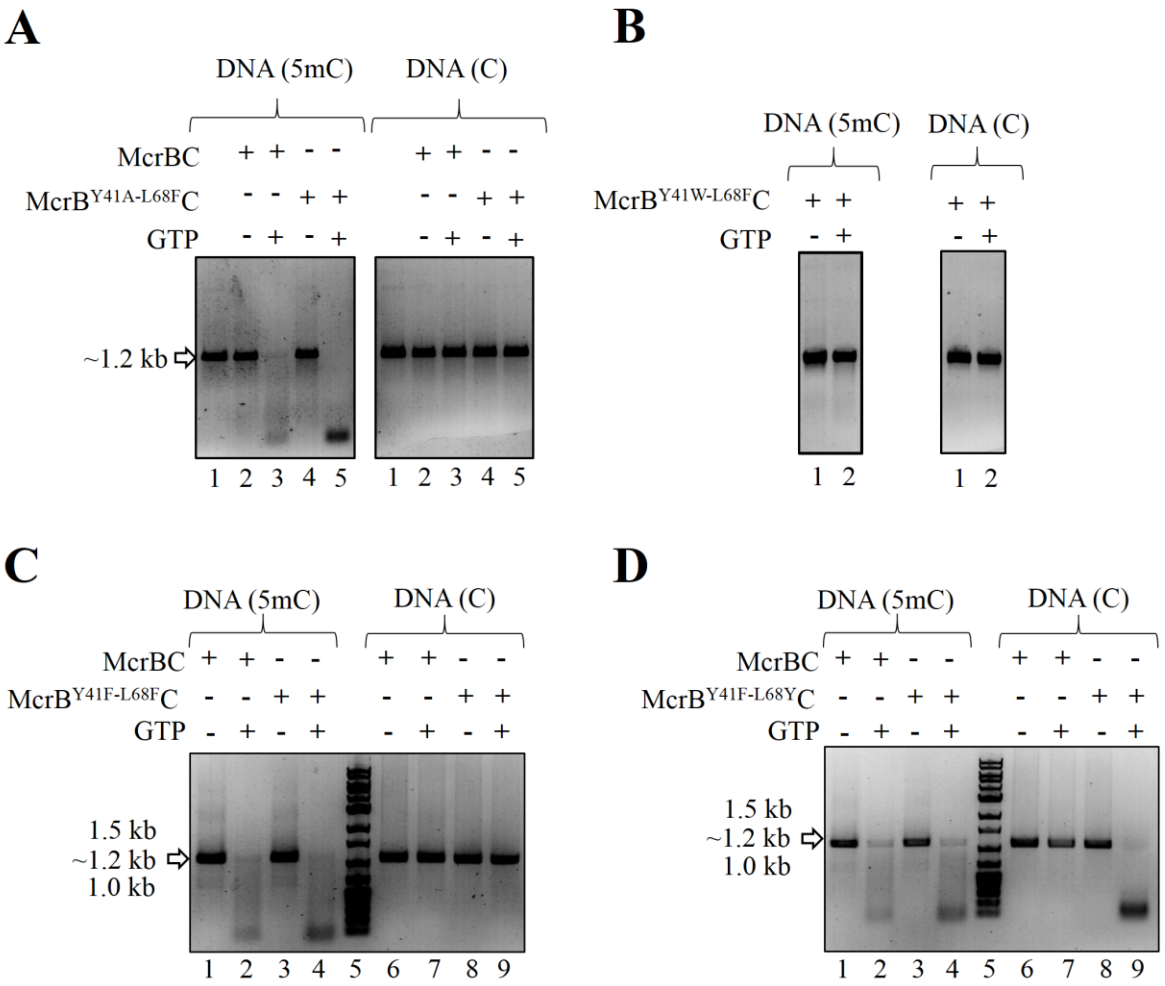


Figure 5.8 DNA cleavage experiments showing the role of the intercalating residue in cytosine specificity. A representative 0.8 % agarose gel of DNA cleavage assay of (**A**) McrB$^{Y41A-L68F}$C, (**B**) McrB$^{Y41W-L68Y}$C, (**C**) McrB$^{Y41F-L68F}$C and (**D**) McrB$^{Y41F-L68Y}$C complexes with methylated (5mC) and non-methylated ~1.2 kb DNA. Refer Materials and Methods for details.

cleaving methylated DNA (Figure 5.8 **C**). This we suspected was because of the loss in the cytosine specificity that was gained by the single mutant McrB^{L68F}. Interestingly, the nuclease activity against both methylated and non-methylated DNA was retained in the double mutant McrB$^{Y41F-L68Y}$C (Figure 5.8 **D**). Quantification of the nuclease activity of McrB$^{Y41F-L68Y}$C against non-methylated DNA showed that its nuclease activity was comparable to that of McrB^{L68F}C (Figure 5.9 **A** and **B**).

We suspected poor binding of McrB$^{Y41F-L68F}$ or McrB$^{Y41F-L68Y}$ to non-methylated DNA that resulted in the corresponding mutant McrBC to have no or reduced nuclease activity against the non-methylated DNA, respectively. Anisotropic DNA binding studies confirmed our suspicion. The K_D value for binding of McrB$^{Y41F-L68F}$-NTD to non-methylated DNA was 1997 ±

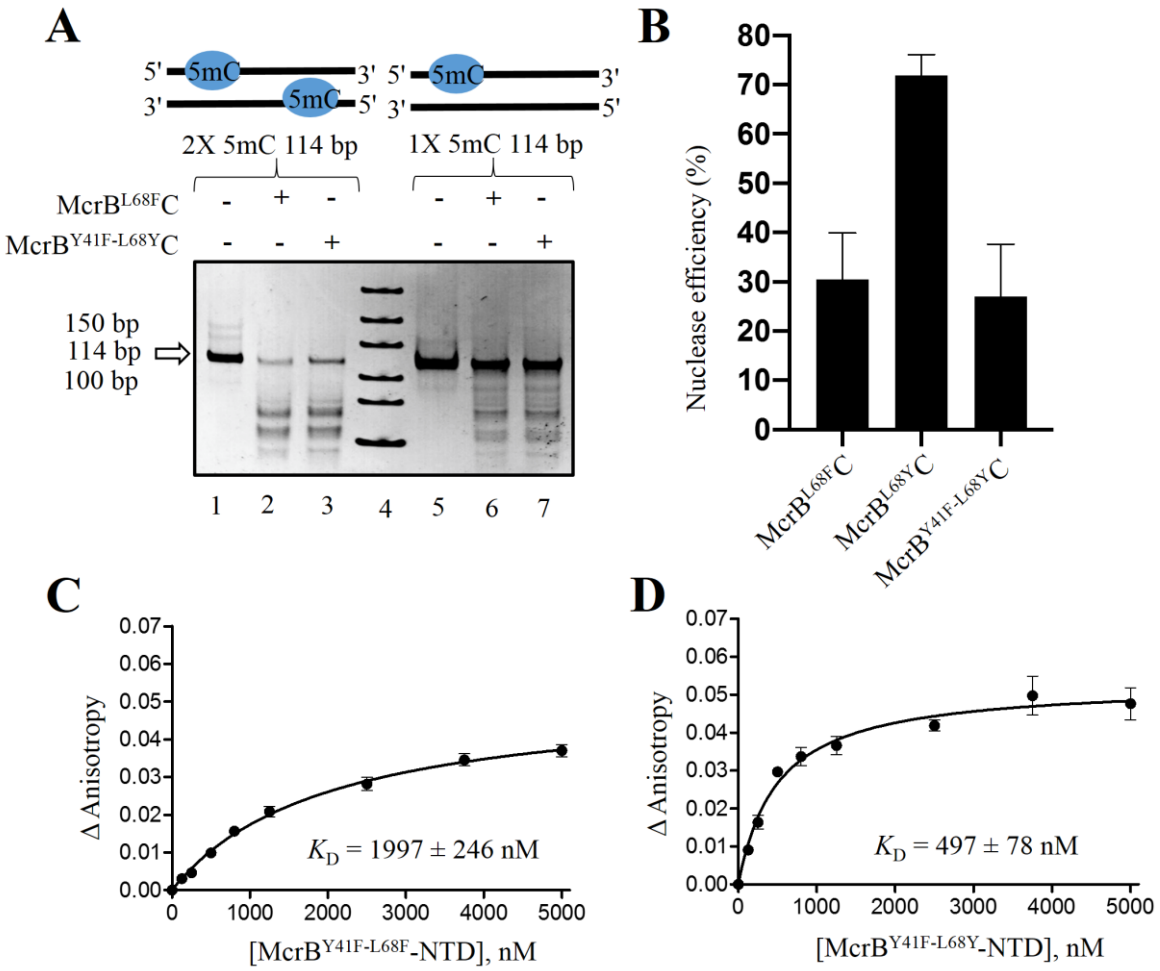


Figure 5.9 Nuclease and binding efficiencies of McrB$^{Y41F-L68F}$C and McrB$^{Y41F-L68Y}$C complexes. (**A**) The top panel shows a schematic representation of 114 bp used for DNA cleavage assay containing a single 5mC site. The bottom panel shows a representative 10 % native PAGE gel for DNA cleavage assay of McrB^{L68F}C and McrB$^{Y41F-L68Y}$C with 114 bp substrate. (**B**) A histogram shows a nuclease efficiency (%) of McrB$^{Y41F-L68Y}$C compared to McrBC^{L68F}, McrB^{L68Y}C complexes towards 114 bp DNA containing a single 5mC. The nuclease efficiency of each complex was calculated using four separate and independent DNA cleavage reactions. A plot of anisotropy change as a function of varying concentrations of (**C**) McrB^{L68F}-NTD and (**D**) McrB^{L68Y}-NTD. These anisotropy measurements were carried out using a covalently linked 5-IAF fluorescence label to 13 bp DNA, and each data point represents an average of at least three independent experiments. The binding constant (K_D) for each is also presented and calculated from non-linear curve fitting using Graphpad Prism 5.0 program.

246 nM, an almost 5-fold increase in comparison to the single mutant McrB^{L68F}-NTD that had a K_D value of 400 ± 39 nM (Figure 5.4 **D** and 5.9 **C**). The reduction in affinity of McrB$^{Y41F-L68F}$-NTD for non-methylated DNA explains why McrB$^{Y41F-L68F}$C does not cleave non-methylated DNA (Figure 5.8 **C**). Similarly, in the case of McrB$^{Y41F-L68Y}$-NTD, the K_D value was 497 ± 78 nM, which is almost comparable to McrB^{L68F}-NTD (Figure 5.9 **D** and Figure 5.4 **D**), thus, explaining the comparable nuclease efficiencies of McrB$^{Y41F-L68Y}$C and McrB^{L68F}C (Figure 5.9 **B**). Note that the binding of McrB$^{Y41F-L68Y}$-NTD (K_D = 497 ± 78 nM) was weaker than McrB^{L68Y}-NTD (K_D = 166 ± 19 nM) (Figure 5.9 **D** and Figure 5.4 **E**). The difference in affinities was consistent with the relatively poor nuclease activity of McrB$^{Y41F-L68Y}$C in comparison to the McrB^{L68Y}C towards non-methylated DNA (Figure 5.4 **E**, Figure 5.9 **D**, and Figure 5.9 **B**).

In summary, first, we replaced the intercalating Tyr41 with phenylalanine to generate a double mutant of McrB such as McrB$^{Y41F-L68F}$ and McrB$^{Y41F-L68Y}$. Interestingly, we noted that, unlike McrB^{L68Y}C, McrB^{L68F}C, or McrB$^{Y41F-L68Y}$C, McrB$^{Y41F-L68F}$C did not cleave non-methylated DNA. We correlated this to the reduced affinity of McrB$^{Y41F-L68F}$ for non-methylated sites (RC) compared to the other three mutants. Together, these observations hinted towards possible coordination between the base recognizing McrB pocket and the intercalating Tyr41 in the target recognition, which might be critical for cytosine specificity.

5.3.7 Contribution of structural water towards cytosine specificity

To obtain the molecular basis for the different specificities displayed by McrB$^{Y41F-L68F}$-NTD and McrB$^{Y41F-L68Y}$-NTD, we determined the crystal structure of the DNA binding domain of these mutants bound to the 13 bp non-methylated DNA to a resolution of 1.65 and 2.10, respectively (Table 5.2). We compared these structures with the structures of McrB^{L68Y}-NTD and McrB^{L68F}-NTD. Comparison of the structures revealed a network of water-mediated interaction connecting O2 of the flipped cytosine to the hydroxyl group of the intercalating Tyr41 *via* the hydroxyl group McrB-Thr85 (Figure 5.10 **A**). This network was conserved in both McrB^{L68F}-NTD and McrB^{L68Y}-NTD structures. In addition, the network of water molecules connecting Lys116, Gua8, and Tyr41 appeared to increase the overall hydration of the minor groove of DNA and solvate the –OH group of Tyr41 (Figure 5.10 **A**). Together, these water-mediated interactions made the intercalation favorable and increased the overall stability of the protein-DNA complex. In contrast, in the structures of McrB$^{Y41F-L68F}$-NTD and McrB$^{Y41F-L68Y}$-NTD bound to DNA, the water-mediated interaction connecting flipped cytosine to the intercalating residue was absent because of the absence of the hydroxyl group due to the replacement of tyrosine by phenylalanine (Figure 5.10

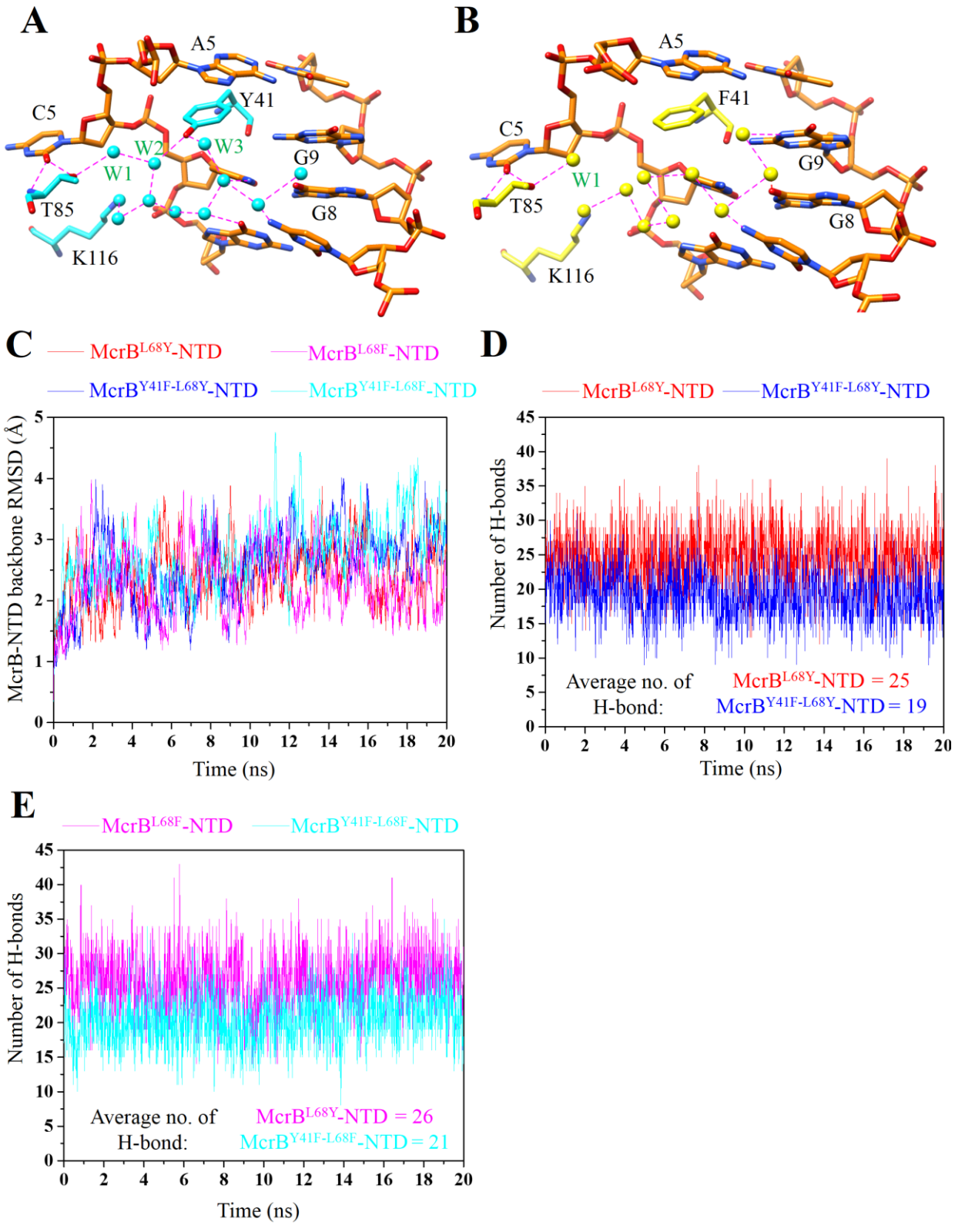

A
A5
Y41
C5
W2
W3
W1
G9
T85
G8
K116
B
A5
F41
C5
W1
G9
T85
G8
K116
C
McrB^L68Y-NTD
McrB^L68F-NTD
McrB^Y41F-L68Y-NTD
McrB^Y41F-L68F-NTD
McrB-NTD backbone RMSD (Å)
Time (ns)
D
McrB^L68Y-NTD
McrB^Y41F-L68Y-NTD
Number of H-bonds
Average no. of H-bond:
McrB^L68Y-NTD = 25
McrB^Y41F-L68Y-NTD = 19
Time (ns)
E
McrB^L68F-NTD
McrB^Y41F-L68F-NTD
Number of H-bonds
Average no. of H-bond:
McrB^L68Y-NTD = 26
McrB^Y41F-L68F-NTD = 21
Time (ns)

Figure 5.10 Solvation of the intercalating residue. (**A**) A structural snapshot showing the intercalating Tyr41 and surrounded water molecules in cyan as in the crystal structure of McrB^{L68Y}-NTD bound to non-methylated DNA. A structure of McrB^{L68Y}-NTD instead of other mutant enzymes bound to DNA was used because of its relatively better resolution (Table 5.2). (**B**) A structural snapshot showing the intercalating Phe41 and water molecules in yellow. A structure of McrB$^{Y41F-L68F}$-NTD instead of McrB$^{Y41F-L68Y}$-NTD bound to DNA was used because of its relatively better resolution (Table 5.2). The H-bonds (d < 3.5 Å) within water molecules or with protein or DNA residues are highlighted in magenta. (**C**) A plot of RMSD, calculated through least square fitting protein backbone versus simulation time (20 ns), to investigate the stability of the McrB-NTD mutants. A plot of the number of H-bond within 5 Å of intercalating residue versus simulation time (20 ns) for (**D**) McrB^{L68Y}-NTD or McrB$^{Y41F-L68Y}$-NTD bound to DNA and (**E**) McrB^{L68F}-NTD or McrB$^{Y41F-L68F}$-NTD bound to DNA. The average no. of H-bond for each case is also presented.

B). In general, the water-mediated hydrogen bond network in McrB$^{Y41F-L68F}$-NTD and McrB$^{Y41F-L68Y}$-NTD was different from that in McrB^{L68F}-NTD or McrB^{L68Y}-NTD. Additionally, the exposure of the hydrophobic edge of Phe41, made of $C^{\varepsilon 1}$, $C^{\varepsilon 2,}$ and C^{ζ} atoms, to the solvent region (Figure 5.10 **B**) would result in the intercalation being entropically unfavorable.

To further study the water arrangement around the intercalating residue, we carried out molecular dynamics (MD) simulation of the four variants of McrB discussed so far, such as McrB^{L68Y}-NTD, McrB^{L68F}-NTD, McrB$^{Y41F-L68Y}$-NTD, or McrB$^{Y41F-L68Y}$-NTD bound to non-methylated DNA. The simulation was carried out for 20 ns (Figure 5.10 **C**-**E**). The backbone RMSD values of the simulated structures ranged from 2 to 4 Å, indicating that all protein-DNA complexes were stable during the simulation run (Figure 5.10 **C**). We next counted the number of H-bonds within a 5 Å radius of residue 41 of the McrB-NTD mutants (Figure 5.10 **D** and **E**). These H-bonds were mainly in the region of the minor groove of the DNA. On average, in the case of McrB^{L68Y}-NTD and McrB^{L68F}-NTD, the number of H-bond within a 5 Å radius of Tyr41 was 25 and 26, respectively. Whereas, in McrB$^{Y41F-L68Y}$-NTD and McrB$^{Y41F-L68Y}$-NTD having Phe41, it was only 19 and 21, respectively. This could result in the intercalation of Phe41 unfavorable because of the loss in H-bond and exposure of the hydrophobic region to solvent. On the other hand, solvation of the water exposed –OH group could make intercalation of Tyr41 favorable. Together, these effects could affect the stability of the base-flipped state, which would affect cytosine specificity.

5.4 Discussion

C5 methylation of cytosine is the most common epigenetic mark of all forms of life (Kumar et al., 2018; Medvedeva et al., 2014; Moore et al., 2013; Ooi et al., 2009; Walsh and Xu, 2006; Yin et al., 2017). A common mode of discrimination of methylated from a non-methylated base is by base flipping (Cheng and Blumenthal, 1996; Han et al., 2015; Roberts and Cheng, 1998). Hence, understanding the mechanism of base flipping and recognition of 5mC by proteins is essential. In this study, we used McrBC as a model to investigate the same, and the results of the study are summarized in Figure 5.11. The conserved residues Leu68 and Trp49 of McrB recognize 5mC using vdW dispersion interaction and C-H···π interactions, respectively (Figure 5.11) (Brandl et al., 2001; Li et al., 2018; Sukackaite et al., 2012). Upon replacing Leu68 with phenylalanine, the mutant McrBC was able to recognize cytosine in addition to 5mC and cleaved methylated and non-methylated DNA (Figure 5.11). The cytosine specificity was primarily achieved by stabilizing vdW dispersion interaction between Phe68 and cytosine (Figure 5.11). This conclusion was based on computational analysis and high-resolution crystal structure. Interestingly, this structure lacked the C-H···π interaction involving Trp49 leading us also to conclude that cytosine specificity was the sole result of dispersion interaction (Figure 5.11). This C-H···π interaction, however, appears crucial for 5mC specificity in the case of McrB^{L68F} mutant because it probably compensated for the unfavorable steric interaction of McrB-Phe68 with 5mC ($\Delta E = \sim 0.5$ kcal.mol^{-1}). Thus, we find that by altering the cavity size of the base-specificity pocket, such as by L68F or L68Y McrB mutants, the McrBC complex could accommodate cytosine or 5mC stably bind to either methylated or non-methylated DNA and cleave both (Figure 5.11). Note that all mutants of McrB discussed in this study were able to cleave methylated DNA, which could be because of the conservation of the C-H···π interaction, similar to the noted in the structure of the McrB^{L68F}-NTD bound to hemimethylated DNA (Figure 5.5 **B**).

Interestingly, unlike for 5mC specificity, the cytosine specificity of McrB mutants was strongly dependent on the nature of intercalating residue (Figure 5.11). While McrB^{L68F}C could cut non-methylated DNA, the double mutant McrB$^{L68F\ -Y41F}$C could not. Removal of just the phenolic hydroxyl group of the side chain in the intercalating residue 41 of McrB was sufficient to prevent the recognition of non-methylated DNA (Figure 5.11). Additionally, the ability of the McrB$^{Y41F-L68Y}$C to cut non-methylated DNA, while McrB$^{Y41F-L68F}$C failed to cut the same, was surprising. Both these McrB mutants had phenylalanine as the intercalating residue 41, which disrupted the water network observed in the wild type McrB having tyrosine at position 41 (Figure 5.11). Based on our analysis, we propose that introducing vdW dispersion interaction between the

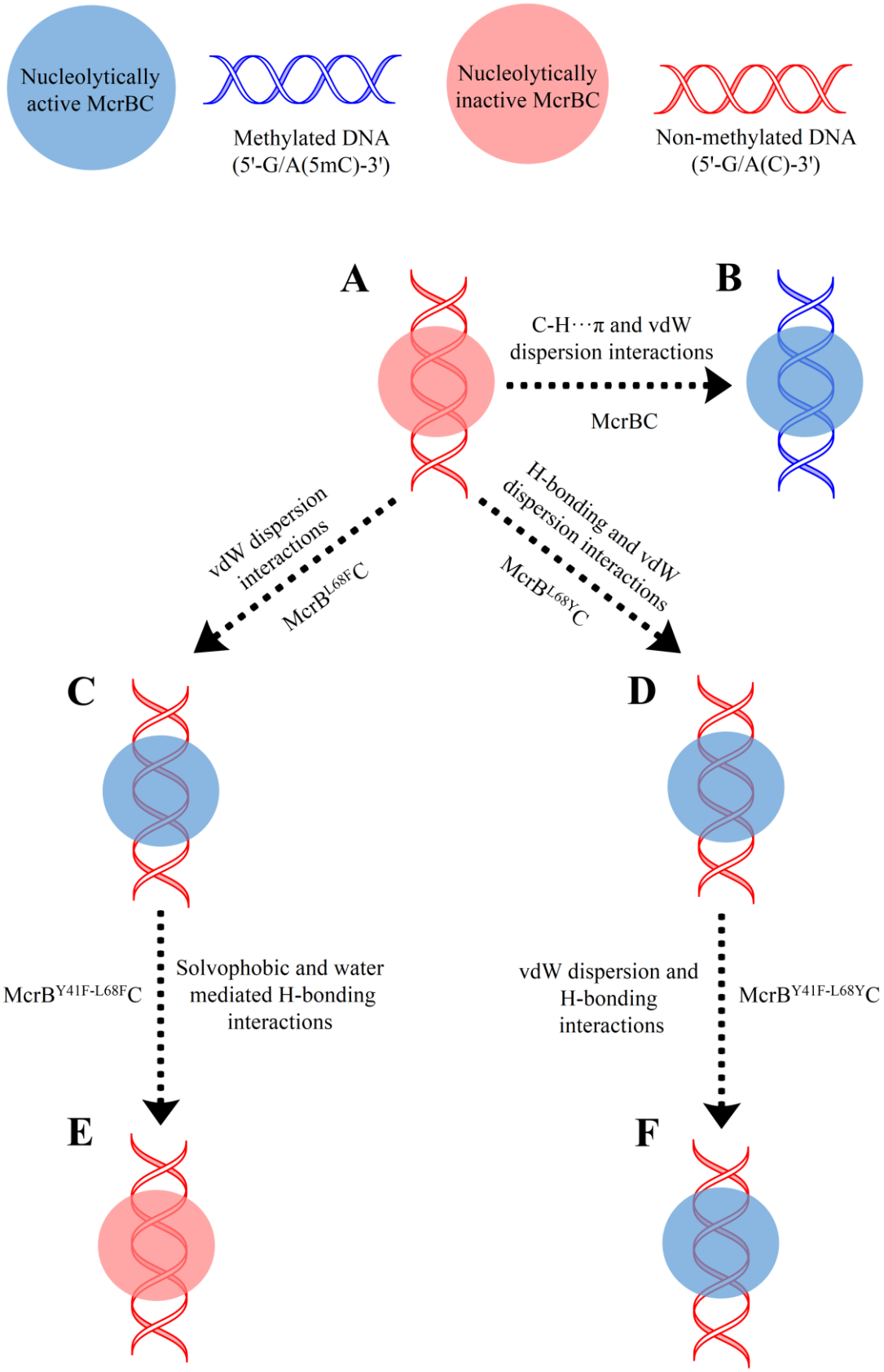

Nucleolytically active McrBC
Methylated DNA (5'-G/A(5mC)-3')
Nucleolytically inactive McrBC
Non-methylated DNA (5'-G/A(C)-3')
A
B
C-H···π and vdW dispersion interactions
McrBC
vdW dispersion interactions
McrB^L68FC
H-bonding and vdW dispersion interactions
McrB^L68YC
C
D
McrB^Y41F-L68FC
Solvophobic and water mediated H-bonding interactions
vdW dispersion and H-bonding interactions
McrB^Y41F-L68YC
E
F

Figure 5.11 Modulation of non-covalent interactions alters base specificity of McrB. A summary of this study highlights the delicate role of the van der Waals dispersion interaction and its coordination with the other non-bonded interactions made by the base recognizing pocket and the intercalating residue of McrB with DNA and solvent molecules. (**A**) The wild type McrBC is nuclease inactive (depicted as a red sphere) towards a non-methylated DNA (depicted as red double helix). (**B**) Wild type McrBC is, nuclease active (depicted as a blue sphere) against a modified DNA (blue double helix) having methylated cytosine that satisfies the target site requirement. The recognition of the methylated cytosine is established by vdW dispersion and C-H$\cdots\pi$ interactions. McrBC can cleave methylated or non-methylated DNA upon mutation of McrB-Leu68 to (**C**) phenylalanine or (**D**) tyrosine. Recognition of a flipped cytosine by McrB^{L68F} is mediated by vdW dispersion interaction. Recognition by McrB^{L68Y} is primarily mediated by vdW dispersion interaction and H-bond. (**E**) However, McrB^{L68F} fails to recognize the flipped cytosine on mutation of intercalating McrB-Tyr41 to phenylalanine, and McrB$^{Y41F-L68F}$ fails to cleave non-methylated DNA. This possibly is due to solvophobic effects resulting from rearrangement of water molecules and the consequent disruption of a network of H-bonds linking the intercalating residue to McrB-Thr85 in the base-specificity pocket. (**F**) However, the loss in stability of the flipped cytosine upon mutation of McrB-Tyr41 to phenylalanine can be compensated by replacing the phenylalanine at position 68 of McrB by tyrosine, i.e. the double mutant McrB$^{Y41F-L68Y}$, which on complexation with McrC can cleave both methylated and non-methylated DNA.

McrB residue at position 68 and the flipped cytosine by mutations such as McrB-L68F or McrB-L68Y stabilized the flipped base and allowed recognition of cytosine (Figure 5.3 **A**). Rearrangement in the water network around the intercalating residue McrB-Tyr41 upon mutation to phenylalanine destabilized the flipped state of cytosine and affected its recognition (Figure 5.8 **A**, Figure 5.11). Interestingly, the introduction of tyrosine at position 68 of McrB results in an H-bond between the hydroxyl O$^{\eta}$ of Tyr68 and N8 of the flipped, which compensates for the destabilizing effect of McrB-Y41F and, thus, stabilizes the flipped base and facilitates recognition of cytosine by McrB$^{Y41F-L68Y}$C. Our observations and conclusions also highlight the fine-tuning of the coordination between the base recognizing pocket and intercalation residue that dictates the specificity of McrB for methylated and non-methylated cytosine (Figure 5.11). This analysis is also an excellent example of coordination among different weak interactions (H-bond, vdW dispersion, and hydrophobic interactions in the case discussed here) and also the compensatory roles that they can play towards molecular recognition (Figure 5.11). The study also demonstrates that the packing of a flipped base can be modified by modulating the weak non-covalent interactions, such as H-bond and vdW dispersion interaction, to stabilize the base and facilitate its recognition.

5.5 References

Abdellah et al., 2004. Finishing the euchromatic sequence of the human genome. Nature. 431, 931–945.

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindah, E., 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1–2, 19-25.

Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. Sect. D Biol. Crystallogr. D68, 352-367.

Allan, B.W., Garcia, R., Maegley, K., Mort, J., Wong, D., Lindstrom, W., Beechem, J.M., Reich, N.O., 1999. DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. J. Biol. Chem. 274(27), 19269-19275.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., Shirakawa, M., 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 455, 818–821.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. Sect. D Biol. Crystallogr. D67, 271-281.

Becke, A.D., 1997. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. J. Chem. Phys. 107(20), 8554.

Becke, A.D., 1993. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98(7), 5648.

Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Dinola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81(8), 3684.

Boys, S.F., Bernardi, F., 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. Mol. Phys. 19(4), 553–566.

Brandl, M., Weiss, M.S., Jabs, A., Sühnel, J., Hilgenfeld, R., 2001. C-H···π-interactions in proteins. J. Mol. Biol. 307(1), 357–377.

Chai, J. Da, Head-Gordon, M., 2008. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. Phys. Chem. Chem. Phys. 10, 6615-6620.

Chakrabarti, P., Samanta, U., 1995. CH/π Interaction in the Packing of the Adenine Ring in Protein Structures. J. Mol. Biol. 251(1), 9-14.

Chand, M.K., Nirwan, N., Diffin, F.M., Aelst, K. Van, Kulkarni, M., Pernstich, C., Szczelkun, M.D.,

Saikrishnan, K., 2015. Translocation-coupled DNA cleavage by the Type. Nat. Chem. Biol. 11, 870–877.

Chen, Y.Z., Prohofsky, E.W., 1995. Sequence and temperature dependence of the interbase hydrogen-bond breathing modes in B-DNA polymers: Comparison with low-frequency Raman peaks and their role in helix melting. Biopolymers. 35(6), 573-582.

Cheng, X., Blumenthal, R.M., 1996. Finding a basis for flipping bases. Structure. 4(6), 639–645.

Contreras-García, J., Johnson, E.R., Keinan, S., Chaudret, R., Piquemal, J.P., Beratan, D.N., Yang, W., 2011. NCIPLOT: A program for plotting noncovalent interaction regions. J. Chem. Theory Comput. 7(3), 625-632.

Cornell, W., Cieplak, P., Bayly, C., Gould, I., Merz, K., Ferguson, D., Spellmeyer, D., Fox, T., Caldwell, J., Kollman, P., 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117(19), 5179-5197.

Craig et al., 2001. The sequence of the human genome. Science. 291(5507), 1304-1351.

Dodson, M.L., Michaels, M.L., Lloyd, R.S., 1994. Unified catalytic mechanism for DNA glycosylases. J. Biol. Chem. 269(52), 32709-32712.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. Sect. D Biol. Crystallogr. D66, 486-501.

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. Sect. D Biol. Crystallogr. D69, 1204-1214.

French, S., Wilson, K., 1978. On the treatment of negative intensity observations. Acta Crystallogr. Sect. A. A34, 517-525.

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.; Sonnenber, D.J., 2009. Gaussian 09. Gaussian, Inc. Wallingford CT 2–3.

Garvie, C.W., Wolberger, C., 2001. Recognition of specific DNA sequences. Mol. Cell. 8(5), 937-946.

Grimme, S., 2006. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J. Comput. Chem. 27(15), 1787-1799.

Gupta, Y.K., Chan, S.H., Xu, S.Y., Aggarwal, A.K., 2015. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nat. Commun. 6, 7363.

Han, T., Yamada-Mabuchi, M., Zhao, G., Li, L., Liu, G., Ou, H.Y., Deng, Z., Zheng, Y., He, X., 2015. Recognition and cleavage of 5-methylcytosine DNA by bacterial SRA-HNH proteins. Nucleic Acids

Res. 43(2), 1147–1159.

Hardwick, J.S., Lane, A.N., Brown, T., 2018. Epigenetic Modifications of Cytosine : Biophysical Properties , Regulation , and Function in Mammalian DNA. BioEssays. 40(3), 1700199.

Hendershot, J.M., Brien, P.J.O., 2014. Critical role of DNA intercalation in enzyme-catalyzed nucleotide flipping. Nucleic Acids Res. 42(20), 12681–12690.

Holder, J.B., Bennett, A.F., Chen, J., Spencer, D.S., Byrne, M.P., Stites, W.E., 2001. Energetics of side chain packing in staphylococcal nuclease assessed by exchange of valines, isoleucines, and leucines. Biochemistry 40(46), 13998–14003.

Hong, S., Cheng, X., 2016. DNA base flipping: A general mechanism for writing, reading, and erasing DNA modifications. Adv. Exp. Med. Biol. 945, 321-341.

Hopkins, B.B., Reich, N.O., 2004. Simultaneous DNA Binding, Bending, and Base Flipping. J. Biol. Chem. 279(35), 37049–37060.

Huang, N., Banavali, N.K., Mackerell, A.D., 2003. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. Proc. Natl. Acad. Sci. U. S. A. 100(1), 68-73.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics. J. Mol. Graph. 14(1), 33-38.

Joh, N.H., Oberai, A., Yang, D., Whitelegge, J.P., Bowie, J.U., 2009. Similar energetic contributions of packing in the core of membrane and water-soluble proteins. J. Am. Chem. Soc. 131(31), 10846–10847.

Johnson, E.R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A.J., Yang, W., 2010. Revealing noncovalent interactions. J. Am. Chem. Soc. 132(18), 6498–6506.

Kabsch, W., 2010. *XDS*. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132.

Kimber, S.T., Brown, T., Fox, K.R., 2014. A Mutant of Uracil DNA Glycosylase That Distinguishes between Cytosine and 5-Methylcytosine, PLOS ONE. 9(4), e95394.

Kumar, S., Chinnusamy, V., Mohapatra, T., 2018. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. Front. Genet. 9, 640.

Lee, C., Yang, W., Parr, R.G., 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B. 37(2), 785.

Li, J., Wang, Y., An, L., Chen, J., Yao, L., 2018. Direct Observation of CH/CH van der Waals Interactions in Proteins by NMR. J. Am. Chem. Soc. 140(9), 3194–3197.

Luscombe, N.M., Austin, S.E., Berman, H.M., Thornton, J.M., 2000. An overview of the structures of protein-DNA complexes. Genome Biol. 1, reviews001.1.

Mandel-Gutfreund, Y., Margalit, H., Jernigan, R.L., Zhurkin, V.B., 1998. A role for CH···O interactions in Protein-DNA recognition. J. Mol. Biol. 277(5), 1129-1140.

Martin, A.M., Sam, M.D., Reich, N.O., Perona, J.J., 1999. Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. Nat. Struct. Biol. 6, 269–277.

McCoy, A.J., 2006. Solving structures of protein complexes by molecular replacement with Phaser. Acta Crystallographica Section D: Biological Crystallography. D63, 32-41.

Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I. V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R.R., Bajic, V.B., 2014. Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 15, 119.

Mishra, G.G., 2017. Understanding the molecular basis of base-specific interaction of Type IV restriction system McrBC. Master dissertation, IISER, Pune.

Misra, S., Singh, P., Mahata, R.N., Brandaõ, P., Roy, S., Mahapatra, A.K., Nanda, J., 2021. Supramolecular antiparallel β-sheet formation by tetrapeptides based on amyloid Sequence. J. Phys. Chem. B. 125(17), 4274-4285.

Moe, J.G., Russu, I.M., 1992. Kinetics and Energetics of Base-Pair Opening in 5′;-d(CGCGAATTCGCG)-3′ and a Substituted Dodecamer Containing G·T Mismatches. Biochemistry. 31(36), 8421-8428.

Moore, L.D., Le, T., Fan, G., 2013. DNA methylation and its basic function. Neuropsychopharmacology. 38, 23–38.

Nirwan, N., Itoh, Y., Singh, P., Bandyopadhyay, S., Vinothkumar, K.R., Amunts, A., Saikrishnan, K., 2019a. Structure-based mechanism for activation of the AAA+ GTPase McrB by the endonuclease McrC. Nat. Commun. 10, 3058.

Nirwan, N., Singh, P., Mishra, G.G., Johnson, C.M., Szczelkun, M.D., Inoue, K., Vinothkumar, K.R., Saikrishnan, K., 2019b. Hexameric assembly of the AAA+ protein McrB is necessary for GTPase activity. Nucleic Acids Res. 47(2), 868–882.

Niu, Y., Suzuki, H., Hosford, C.J., Walz, T., Chappie, J.S., 2020. Structural asymmetry governs the assembly and GTPase activity of McrBC restriction complexes. Nat. Commun. 11, 5907.

Ooi, S.K.T., O'Donnell, A.H., Bestor, T.H., 2009. Mammalian cytosine methylation at a glance. J. Cell Sci. 122(16), 2787–2791.

Panne, D., Müller, S.A., Wirtz, S., Engel, A., Bickle, T.A., 2001. The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. EMBO J. 20, 3210-3217.

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52(12), 7182.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25(13), 1605-1612.

Ratnaparkhi, G.S., Varadarajan, R., 2000. Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. Biochemistry 39(40), 12365–12374.

Řezáč, J., Hobza, P., 2016. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. Chem. Rev. 116(9), 5038-5071.

Roberts, R.J., Cheng, X., 1998. Base flipping. Annu. Rev. Biochem. 67, 181–198.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., Mann, R.S., 2010. Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. 79, 233-269.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., Honig, B., 2009. The role of DNA shape in protein-DNA recognition. Nature. 461, 1248–1253.

Sagendorf, J.M., Markarian, N., Berman, H.M., Rohs, R., 2020. DNAproDB: An expanded database and web-based tool for structural analysis of DNA-protein complexes. Nucleic Acids Res. 48(D1), D277–D287.

Schormann, N., Ricciardi, R., Chattopadhyay, D., 2014. Uracil-DNA glycosylases - Structural and functional perspectives on an essential family of DNA repair enzymes. Protein Sci. 23(12), 1667-1685.

Seeman, N.C., Rosenberg, J.M., Rich, A., 1976. Sequence specific recognition of double helical nucleic acids by proteins. Proc. Natl. Acad. Sci. U. S. A. 73(3), 804-808.

Su, T., Tock, M.R., Egelhaaf, S.U., Poon, W.C.K., Dryden, D.T.F., 2005. DNA bending by M . EcoKI methyltransferase is coupled to nucleotide flipping. Nucleic Acids Res. 33(10), 3235–3244.

Sukackaite, R., Grazulis, S., Tamulaitis, G., Siksnys, V., 2012. The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 40(15), 7552–7562.

Sutherland, E., Coe, L., Raleigh, E.A., 1992. McrBC: a multisubunit GTP-dependent restriction endonuclease. J. Mol. Biol. 225(2), 327-348.

Tsai, J., Taylor, R., Chothia, C., Gerstein, M., 1999. The packing density in proteins: Standard radii and volumes. J. Mol. Biol. 290(1), 253-266.

Várnai, P., Lavery, R., 2002. Base flipping in DNA: Pathways and energetics studied with molecular dynamic simulations. J. Am. Chem. Soc. 124(25), 7272-7273.

Walsh, C.P., Xu, G.L., 2006. Cytosine methylation and DNA repair. Current Topics in Microbiology and

Immunology. 301. Springer, Berlin, Heidelberg.

Watson, J.D., Crick, F.H.C., 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 171, 737–738.

Wilson, K.A., Kellie, J.L., Wetmore, S.D., 2014. DNA-protein π-interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. 42(10), 6726–6741.

Wu, Q., Yang, W., 2002. Empirical correction to density functional theory for van der Waals interactions. J. Chem. Phys. 116(2), 515.

Xue, M., Wakamoto, T., Kejlberg, C., Yoshimura, Y., Nielsen, T.A., Risør, M.W., Sanggaard, K.W., Kitahara, R., Mulder, F.A.A., 2019. How internal cavities destabilize a protein. Proc. Natl. Acad. Sci. U. S. A. 16(42), 21031–21036.

Yang, L., Adam, C., Nichol, G.S., Cockroft, S.L., 2013. How much do van der Waals dispersion forces contribute to molecular recognition in solution? Nat. Chem. 5, 1006–1010.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., Nitta, K.R., Taipale, M., Popov, A., Ginno, P.A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., Taipale, J., 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 356, eaaj2239.

Zacharia, N.K., 2020. Molecular basis and modulation of target DNA recognition in the Type IV restriction endonuclease McrBC. Master dissertation, IISER, Pune.

Zagorskaitė, E., Manakova, E., Sasnauskas, G., 2018. Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC. FEBS Lett. 592(19), 3335-3345.

Zearfoss, N.R., Ryder, S.P., 2012. End-labeling oligonucleotides with chemical tags after synthesis. Methods Mol. Biol. 941, 181–193.

Summary and future directions

This thesis first investigated the directional and the cooperative facets of the σ-hole mediated Ch-bond formed by divalent S in Chapter 2. Our computational analysis revealed that the two interactions, S···O/N and C–H···O/N could coexist, together strengthening the overall molecular interaction. Interestingly, we noted that C–H···O/N interaction dominates over S···O/N interactions when the strength of σ-hole is poor. Whereas, the S···O/N interaction dominates over C–H···O/N interaction when the σ-hole is stronger. We also studied the directional selectivity of Ch-bond in the context of a complex nucleophile having multiple nucleophilic centers. We provided the MESP basis for the directional preference that was observed in the crystal structures and the correlation between the strength of σ-hole and interaction energy supported the electrostatic basis of the interaction. Also, the cumulative molecular interaction appeared to be stabilized by the cooperative nature of weak S···O, C–H···O and C-H···S interactions. Thus, we proposed that Ch-bond collaborating with other weak interactions could form a signature pair of the interaction, whose cumulative strength and directional features could alter or stabilize the local conformations of proteins, which may have functional relevance. Thus, identifying such pairs of interactions in biomolecules and investigating their interplay, similar to that reported for H-bond and n→π* interaction (Bartlett et al., 2013), could be of broad implications in the field such as protein modeling or designing.

Hydrophobicity of methionine and redox properties of cysteine have been given greater precedence in their roles in protein structures and functions over their ability to form weak polar interactions. In chapters 3 and 4, we investigated the interplay between divalent S mediated σ–hole and lone pair interactions. The reciprocity observed between σ–hole and lone pair strength was correlated with their participation in interactions in CSD and PDB. Such reciprocity was also found to be crucial for the capping of α-helices in proteins. In addition, various mechanisms have been identified through which Ch-bond could stabilize the regular and non-regular secondary structures in proteins. We also showed that the σ–hole interaction made by methionyl-tRNA-synthase with its methionine substrate was crucial for specificity. This study underlines the significance of weak polar interactions made by divalent S in protein folding, structure, and function. However, the experimental quantification of the contribution of Ch-bond to the folding or stabilization of the biomolecules remained to be investigated. Also, it would be interesting to find if Ch-bond could play any role in biomolecular recognition, similar to C–H···O or π-stacking

interaction, that are abundant at protein-protein or protein-nucleic acid interfaces (Jiang and Lai, 2002; Wilson et al., 2021).

Chapter 5 provided evidence for vdW dispersion interaction in the base-specific protein-DNA recognition. We showed that DNA base specificity could be altered by introducing the dispersion interaction with the cytosine, an intended target base, which wild-type McrBC does not recognize. The high-resolution crystal structures and dispersion-corrected density function theory supported this and unraveled the mechanism of stabilizing the flipped base of DNA. The calculated strength of the dispersion interactions, in this case, was ~ -0.9 kcal.mol^{-1}. However, our study does not reveal the contribution of the hydrophobic interactions in the stabilization of the flipped state of DNA-bound McrBC. As suggested previously, it is not trivial to distinguish the contribution of the hydrophobic form dispersion interaction in solution for biomolecular recognition (Pace et al., 2011). However, due to the lack of direct interaction of solvent in the base flipping process, we believe the dispersion interaction dominates over hydrophobic. However, ITC based binding experiment would be needed to confirm this. Such an experiment would allow us to find entropic contribution (corresponding to hydrophobic interaction) and enthalpic (corresponding to dispersion interaction) in the stabilization of flipped cytosine for the mutants of McrB, which are discussed in this thesis. This could also reveal the experimental quantification of the dispersion interaction, which is not determined for protein-nucleic acid complexes to the best of our knowledge.

In summary, this thesis studied two aspects of weak interactions. Firstly, polar interaction made by divalent S and its role in proteins were investigated. Additionally, evidence for σ–hole mediated Ch-bond protein-substrate recognition was obtained. Secondly, this thesis identified and showed the delicate coordination of vdW dispersion interaction with other non-covalent interactions in modulating base-specific McrB-DNA recognition. Together, the research work reported in this thesis attempted to understand the significance of weak interactions in the structure, stability, and function of biomolecules.

References

Bartlett, G.J., Newberry, R.W., Vanveller, B., Raines, R.T., Woolfson, D.N., 2013. Interplay of hydrogen bonds and n →π* interactions in proteins. J. Am. Chem. Soc. 135(49), 18682-18688.

Jiang, L., Lai, L., 2002. CH···O hydrogen bonds at protein-protein interfaces. J. Biol. Chem. 277(40), 37732–37740.

Pace, C.N., Fu, H., Fryar, K.L., Landua, J., Trevino, S.R., Shirley, B.A., Hendricks, M.M., Iimura, S., Gajiwala, K., Scholtz, J.M., Grimsley, G.R., 2011. Contribution of Hydrophobic Interactions to Protein Stability. J. Mol. Biol. 408(3), 514–528.

Wilson, K.A., Kung, R.W., D'Souza, S., Wetmore, S.D., 2021. Anatomy of non-covalent interactions between the nucleobases or ribose and π-containing amino acids in RNA-protein complexes. Nucleic Acids Res. 49(4), 2213–2225.

Appendix A

Table A1. Cartesian coordinates of all optimized $H_3N{:}S(CH_3)X$ complexes.

$H_3N{:}S(CH_3)_2$

S -0.43127300 -0.40952400 0.00104900
C -0.13292700 1.37643500 0.00012900
H -0.56108100 1.83751500 0.89525800
H 0.95381700 1.50233400 0.00154600
H -0.55842600 1.83607800 -0.89701100
N 2.86932000 -0.23228400 -0.00049900
C -2.24220500 -0.38745100 -0.00109200
H -2.62197500 0.11302900 -0.89695700
H -2.58408800 -1.42469800 -0.00111900
H -2.62414000 0.11371500 0.89346300
H 3.88184700 -0.19255600 -0.00078400
H 2.59027800 -0.77131000 0.81228800
H 2.58969900 -0.76963300 -0.81419800

$H_3N{:}S(CH_3)H$

S -0.90190000 -0.64111900 -0.00594000
C -0.85706500 1.17914900 -0.00379100
H -1.28575200 1.58079500 0.91579000
H 0.20640500 1.42977100 -0.05129600
H -1.37044500 1.58595600 -0.87650100
N 2.35435400 0.02407300 -0.00250800
H -2.23706800 -0.78162800 0.08383000
H 2.30837500 -0.57330800 -0.82072200
H 2.15670300 -0.56474600 0.79953200
H 3.31409900 0.33765400 0.08470200

$H_3N{:}S(CH_3)NH_2$

S -0.42754000 -0.34495400 0.00024800
C -0.19346900 1.44660000 -0.00002200
H -0.63519000 1.88746600 0.89637400
H 0.89156000 1.59352200 0.00040000
H -0.63444600 1.88710800 -0.89695900
N 2.78679200 -0.28342900 -0.00007800
N -2.14988500 -0.41634900 -0.00027100
H -2.48856000 -0.89487400 -0.82603100
H -2.48908200 -0.89445000 0.82552100
H 3.79241200 -0.15770800 -0.00004300
H 2.55322200 -0.84185300 0.81366900
H 2.55318100 -0.84111000 -0.81432200

$H_3N{:}S(CH_3)SCH_3$

S -0.34856100 -0.28046400 -0.14587400
C -0.72577600 1.49913200 -0.11125500
H -0.50571400 1.94911700 -1.08107700
H -1.79809700 1.55569100 0.10148000
H -0.16041400 1.99699600 0.67994600
N -3.58927500 -0.39931600 0.28241000
S 1.66015800 -0.29445300 -0.60695800
C 2.41851500 -0.05037800 1.03123100
H 3.50169700 -0.06731100 0.87849100
H 2.12969600 0.91711100 1.44915000
H 2.12939500 -0.85881500 1.70542500
H -3.48107700 -0.86437100 -0.61259800
H -4.58375000 -0.28320600 0.44008800
H -3.24878900 -1.04385100 0.98769500

$H_3N{:}S(CH_3)CF_3$

S 0.55393700 -0.25567700 -0.06020700
C 0.86757500 1.52922700 -0.00504300
H 0.48454300 1.95543200 0.92365500
H 1.95541800 1.61120600 -0.03583300
H 0.42739100 2.02119700 -0.87382000
N 3.66489400 -0.44434300 0.01961600
C -1.24561400 -0.16950500 0.00617100
F -1.69825300 0.41423100 1.11975400
F -1.77413800 0.51120000 -1.01454100
F -1.73207900 -1.40752400 -0.03221300
H 3.57669800 -1.09613900 -0.75263400
H 4.62940500 -0.13171000 0.03119200
H 3.51776300 -0.97824400 0.86968100

$H_3N{:}S(CH_3)CN$

S -0.08115700 -0.32085600 -0.00039100
C 0.16847000 1.48688300 0.00000700
H -0.26464100 1.92340700 0.89991700
H 1.25405200 1.59931100 -0.00137900
H -0.26701500 1.92411500 -0.89841200
N 2.99678300 -0.39376700 0.00022100
C -1.77874100 -0.32900700 0.00008000
N -2.94207600 -0.36673900 0.00039100
H 3.90251200 0.06263900 0.00027900
H 2.97586000 -0.99933100 0.81401100
H 2.97641600 -1.00016100 -0.81296600

$H_3N{:}S(CH_3)OH$

S 0.36510600 -0.34000400 -0.00305700
C 0.33974100 1.45965400 0.00496400
H 0.84330800 1.84347200 -0.88535100
H -0.71985200 1.72909500 -0.01855300
H 0.80177100 1.86008800 0.91112300
N -2.66258100 -0.24850100 0.01099900
O 2.03031300 -0.59794800 -0.10416900
H 2.36765500 -0.67303200 0.79420800
H -2.57201500 -0.80598300 -0.83149700
H -2.60722000 -0.89038700 0.79415700
H -3.59822600 0.14197300 0.01140400

$H_3N{:}S(CH_3)NO_2$

S -0.28513600 -0.22247800 0.00900900
C -0.53805700 1.55837000 0.00574100
H -0.14769300 2.00323900 -0.91020000
H -1.62458900 1.66554300 0.04754400
H -0.07466100 2.01106800 0.88284800
N -3.29086900 -0.47035800 -0.00899600
H -3.41413900 -0.94807900 0.87977400
H -4.14035200 0.06082200 -0.17905200
H -3.24283000 -1.18834100 -0.72351500
N 1.51444100 -0.24205600 -0.00450400
O 2.11892600 0.80430000 -0.01246800
O 1.98979500 -1.35529000 0.00228200

$H_3N{:}S(CH_3)F$

S -0.28030800 -0.35973300 -0.00313600
C -0.38520100 1.43380400 -0.00015900
H -0.90182200 1.77675600 0.89893400
H 0.64279800 1.80001700 0.00009700
H -0.90257000 1.78006600 -0.89753600
F -1.94666200 -0.60051100 0.00296600
N 2.44410900 -0.23761300 0.00028700
H 2.66272100 -0.64281200 0.90368500
H 2.56471500 -0.96709900 -0.69352200
H 3.14148600 0.47386900 -0.18922200

Table A2. Cartesian coordinates of all optimized $H_2O{:}S(CH_3)X$ complexes.

$H_2O{:}S(CH_3)F$ (MP2)

S -0.32250500 -0.38615400 -0.00004600
C -0.26410900 1.40999900 0.00000000
H -0.73897700 1.80282300 0.89539700
H 0.79529200 1.66202400 0.00020200
H -0.73885000 1.80300200 -0.89536300
O 2.49710100 -0.18716900 0.00001700
F -2.00698900 -0.51330400 0.00001500
H 2.75731100 -0.71597300 -0.75936700
H 2.75604700 -0.71631400 0.75959700

$H_2O{:}S(CH_3)F$ (MP4)

S -0.33982700 -0.39230600 -0.00003500
C -0.27366300 1.41201100 -0.00001700
H -0.74953900 1.80523000 0.89574100
H 0.78650100 1.66325900 -0.00005600
H -0.74960200 1.80524800 -0.89573500
O 2.54505300 -0.18390900 -0.00004200
F -2.02274200 -0.51024000 0.00004500
H 2.81827600 -0.70314800 -0.75865700
H 2.81782400 -0.70232900 0.75929600

$H_2O{:}S(CH_3)F$ (B3LYP)

S -0.33165600 -0.37392100 -0.00000400
C -0.29249600 1.43093000 0.00000200
H -0.76889500 1.82327300 0.89919900
H 0.76948200 1.68641800 -0.00004100
H -0.76893200 1.82322000 -0.89920100
O 2.55202600 -0.18809400 -0.00000100
F -2.00839900 -0.54991900 0.00000600
H 2.74463300 -0.74089600 -0.76516500
H 2.74457300 -0.74083200 0.76522200

$H_2O{:}S(CH_3)F$ (B3LYP-D)

S -0.32524500 -0.37886200 0.00037800
C -0.26240300 1.42727800 0.00007100
H -0.73675000 1.82413600 0.89876900
H 0.80200900 1.67160700 0.00027700
H -0.73632100 1.82375500 -0.89902100
O 2.51973900 -0.19253000 0.00006000
F -2.00464400 -0.52929200 -0.00043100
H 2.66577500 -0.75681100 -0.76703600
H 2.66750100 -0.76070200 0.76394200

$H_2O{:}S(CH_3)CF_3$

S 0.55523800 -0.30657300 0.07746800
C 0.93732800 1.46592700 0.00308300
H 0.50957000 1.98319900 0.86338700
H 2.02675300 1.51841400 0.03752000
H 0.57303100 1.89446400 -0.93209200
O 3.62076200 -0.38651000 -0.03862900
H 3.59930000 -0.98470800 -0.78891800
H 3.87477400 -0.92903100 0.71089800
C -1.23913500 -0.15151100 -0.00700300
F -1.75006700 0.55719900 1.00313700
F -1.65911500 0.43859400 -1.12946700
F -1.77109400 -1.37041500 0.03769300

$H_2O{:}S(CH_3)CN$

S -0.04731200 0.27872800 -0.48380600
C -0.73809300 1.97021500 -0.52378300
H -0.56909100 2.41753000 -1.50331800
H -1.80587900 1.83564900 -0.34303300
H -0.28818400 2.57311800 0.26488400
O -2.97798600 -0.29395800 -0.05957000
H -3.12277300 -0.77688200 0.75695400
H -3.28626700 -0.87563600 -0.75780700
C 1.57775500 0.68452100 -0.75925100
N 2.69949000 0.93231800 -0.94647900

$H_2O{:}S(CH_3)OH$

S 0.36639500 -0.36561600 0.02506400
C 0.20090000 1.42704900 0.00298600
H 0.63272000 1.83090600 -0.91560500
H -0.87679400 1.61477300 0.02350900
H 0.67033700 1.87935400 0.88025000
O -2.62799900 -0.22777700 -0.00361800
H -2.69508400 -0.86952900 0.70681200
H -2.48951000 -0.74793100 -0.79901300
O 2.03801800 -0.50656700 -0.13032100
H 2.41045700 -0.54525900 0.75661700

$H_2O{:}S(CH_3)NO_2$

S -0.27985900 -0.27513400 0.00024400
C -0.63009400 1.48876000 0.00015100
H -0.22503500 1.95840100 -0.89686900
H -1.72196100 1.54385100 0.00006200
H -0.22519200 1.95848900 0.89719600
O -3.25222800 -0.40459400 0.00000400
H -3.51629300 -0.92478900 0.76224600
H -3.51550400 -0.92167400 -0.76462700
N 1.51497400 -0.20834700 -0.00004900
O 2.06713300 0.86648900 -0.00024400
O 2.04227900 -1.29767800 -0.00006900

$H_2O{:}S(CH_3)Cl$

S -0.05681200 -0.29057100 -0.02944900
C -0.32870200 1.49248000 0.00262100
H 0.11760600 1.96359000 -0.87537500
H -1.41681700 1.60232200 -0.02272100
H 0.07119200 1.92502700 0.92190000
O -2.97817600 -0.36801200 0.01865900
H -3.02796600 -0.94512200 0.78433900
H -3.16529800 -0.93007600 -0.73667900
Cl 2.00752300 -0.29278800 0.01380700

$H_2O{:}S(CH_3)F$

S -0.30553200 -0.37682000 0.00009800
C -0.27110300 1.41481500 -0.00005000
H -0.75331700 1.80524400 0.89866000
H 0.78968200 1.67962700 0.00264400
H -0.74904100 1.80566100 -0.90085100
O 2.45121700 -0.18900100 -0.00005800
F -1.97018800 -0.53194600 -0.00010100
H 2.67437700 -0.72545400 -0.76454400
H 2.67538600 -0.72534200 0.76420400

Table A3. Cartesian coordinates of all optimized $(CH_3)_3N{:}S(CH_3)X$ complexes.

$(CH_3)_3N{:}S(CH_3)_2$

S 1.59806300 -0.36953500 0.10809200
C 1.34848500 1.42340600 0.13689000
H 1.67907900 1.87621900 -0.80331800
H 0.27302200 1.57580400 0.26379300
H 1.88678400 1.87530100 0.97571800
N -1.68372200 -0.03830700 0.03302100
C -1.83138300 0.24838300 -1.37651700
H -1.66905600 1.31708700 -1.56122900
H -1.08610100 -0.31389000 -1.94987600
H -2.83985800 -0.01514600 -1.75623000
C -2.59527800 0.75373600 0.82703600
H -2.43788000 0.55068900 1.89169100
H -2.41687000 1.82151200 0.65224400
H -3.65791100 0.54118800 0.58863900
C -1.85106300 -1.45036600 0.29790600
H -1.13321900 -2.02718000 -0.29477100
H -1.66093200 -1.65715700 1.35657700
H -2.87318100 -1.80353100 0.05006000
C 3.40290100 -0.37302600 -0.06446000
H 3.87743000 0.10724600 0.79701400
H 3.72607400 -1.41533200 -0.11135800
H 3.70770000 0.13510000 -0.98470600

$(CH_3)_3N{:}S(CH_3)H$

S -1.96205600 0.60860100 0.05138500
C -1.97673200 -1.21035800 0.11917000
H -2.32591400 -1.64014900 -0.82162400
H -0.93400200 -1.49193000 0.28503500
H -2.58497300 -1.56879300 0.95132600
N 1.17657700 -0.01809400 0.03740500
C 1.31962100 -0.26121900 -1.38142700
H 1.04074200 -1.29586900 -1.61467200
H 0.65581300 0.40558100 -1.94245300
H 2.35923800 -0.09893500 -1.73227600
C 1.98675200 -0.93469400 0.80807200
H 1.82792200 -0.76629400 1.87859200
H 1.70773200 -1.96994900 0.57819700
H 3.06987700 -0.81894300 0.59771800
C 1.49138400 1.35555100 0.36487300
H 0.85023100 2.03114600 -0.21116400
H 1.30752300 1.53711100 1.42924900
H 2.54945400 1.60551700 0.14485700
H -3.29291700 0.76486700 -0.09091000

$(CH_3)_3N{:}S(CH_3)NH_2$

S 1.48014600 -0.30017000 0.08571300
C 1.37184700 1.50719000 0.07183000
H 1.71968000 1.90590400 -0.88427600
H 0.31616000 1.75023100 0.22059500
H 1.96787300 1.92005500 0.88873700
N -1.55995300 -0.06246300 0.02386200
C -1.79248500 0.21614000 -1.37582900
H -1.57660700 1.26956600 -1.59070200
H -1.12818400 -0.40007400 -1.99152400
H -2.84047800 0.01216400 -1.67645700
C -2.37191700 0.78207400 0.87036000
H -2.14325900 0.58611800 1.92348800
H -2.16035700 1.83811000 0.66414500
H -3.45767100 0.61613200 0.71492200
C -1.77600300 -1.46112100 0.32037900
H -1.12219500 -2.07855800 -0.30506800
H -1.53201400 -1.66185500 1.36928300
H -2.82580600 -1.76923000 0.13891200
N 3.20024100 -0.44680700 -0.03353200
H 3.57263600 -0.93104100 0.77445500
H 3.45721300 -0.95560600 -0.87065800

$(CH_3)_3N{:}S(CH_3)SCH_3$

S 0.73466700 -0.17863900 -0.08800700
C 0.49825400 1.62619200 -0.01137300
H 0.64516800 2.07107800 -0.99782900
H -0.53328000 1.77297600 0.32230100
H 1.18614000 2.07065000 0.71179100
N -2.35101900 -0.10124400 0.10710300
C -2.63146600 0.23753500 -1.27146200
H -2.49819800 1.31454800 -1.42830800
H -1.93768000 -0.29284700 -1.93263400
H -3.66792100 -0.02514300 -1.56513000

C -3.20924500 0.63540700 1.00867400
H -2.95983100 0.39265900 2.04709200
H -3.07106200 1.71367300 0.86492900
H -4.28255200 0.40630500 0.84867200
C -2.47085700 -1.52715800 0.32548100
H -1.79107500 -2.06532300 -0.34371200
H -2.19712600 -1.77112100 1.35760700
H -3.50174600 -1.89147400 0.14158700
S 2.73180800 -0.29926000 -0.60743800
C 3.54920100 -0.13976800 1.01249900
H 4.62259300 -0.25309000 0.83367200
H 3.36061900 0.84477200 1.44748400
H 3.20417100 -0.92581800 1.68696900

$(CH_3)_3N{:}S(CH_3)CF_3$
S -0.48331500 0.16610100 0.14140200
C -0.28646700 -1.63701300 0.09241200
H -0.60191400 -2.03278700 -0.87460700
H 0.78272100 -1.80229100 0.23535200
H -0.84346700 -2.10589600 0.90526500
N 2.54191800 0.09773100 0.02610400
C 2.75291900 -0.16136100 -1.38284600
H 2.59327900 -1.22448700 -1.59865100
H 2.03979400 0.41906100 -1.97846600
H 3.77847100 0.10338000 -1.70882700
C 3.42847600 -0.70333600 0.84301300
H 3.22527900 -0.52242300 1.90376700
H 3.27018700 -1.76905900 0.63950400
H 4.49627600 -0.47608100 0.65134500
C 2.69754300 1.50694100 0.32180300
H 1.99739700 2.09452600 -0.28186300
H 2.47755300 1.69385900 1.37838600
H 3.72426800 1.86544300 0.10958900
C -2.28326300 0.17769900 -0.00048300
F -2.72326800 -0.40481600 -1.12081600
F -2.89555100 -0.44642600 1.01058900
F -2.69845500 1.44206800 -0.00858900

$(CH_3)_3N{:}S(CH_3)CN$
S 1.18310500 -0.24786400 0.10903700
C 1.05118000 1.57348700 0.07290700
H 1.39411800 1.95678000 -0.88850600
H -0.01370500 1.77054500 0.20914600
H 1.62601700 2.00502300 0.89237100
N -1.80184200 -0.08939100 0.01769400
C -2.04956700 0.18258500 -1.38376300
H -1.86141400 1.24051400 -1.60197300
H -1.37809100 -0.42038800 -2.00496300
H -3.09328000 -0.04623700 -1.67621300
C -2.64202300 0.73295400 0.86389700
H -2.41184400 0.54318700 1.91753300
H -2.46310600 1.79483900 0.65781000
H -3.71999100 0.53356100 0.70364600
C -1.99433600 -1.49535600 0.31196600
H -1.33081600 -2.10262600 -0.31335900
H -1.75209100 -1.69419000 1.36147600
H -3.03702100 -1.81938100 0.12645400
C 2.87924600 -0.34193300 0.00435700
N 4.03677700 -0.44007000 -0.06686400

$(CH_3)_3N{:}S(CH_3)OH$
S 1.33505100 -0.31517200 -0.02131100
C 1.41544200 1.48998300 0.00273100
H 2.02555700 1.82924900 -0.83788200
H 0.39458600 1.86145500 -0.11124000
H 1.83465500 1.85523500 0.94431300
N -1.44089600 -0.05823500 0.00390500
C -1.96398200 0.51061900 -1.21874700
H -1.75136400 1.58530700 -1.25780800
H -1.48270800 0.03451700 -2.08034600
H -3.06004200 0.37676200 -1.30971100
C -1.96326000 0.60436300 1.17818500
H -1.48382000 0.19551600 2.07490000
H -1.74787600 1.67827100 1.13397900
H -3.05958000 0.48019100 1.27974100
C -1.67637100 -1.48524900 0.05887500
H -1.21002200 -1.97059800 -0.80545800
H -1.22436600 -1.89925500 0.96681300
H -2.75643100 -1.72923300 0.06015800
O 3.00721000 -0.60635500 -0.08602400
H 3.31822300 -0.71448400 0.81810300

$(CH_3)_3N{:}S(CH_3)NO_2$
S -0.69000700 0.09983900 0.03580500
C -0.63979800 -1.70325100 0.02794700
H -1.04516500 -2.09516400 -0.90584500
H 0.41646100 -1.96312100 0.12061600
H -1.19816300 -2.09771000 0.87761200
N 2.06446600 0.11407500 0.00126700
C 2.56421300 -0.41985000 -1.24887900
H 2.43292300 -1.50767600 -1.27551900
H 2.00485200 0.01449300 -2.08491600
H 3.63918500 -0.20187900 -1.39681000
C 2.71738800 -0.48997000 1.14446700
H 2.26442000 -0.11679800 2.06970200
H 2.60033400 -1.57939800 1.11773200
H 3.80102300 -0.26633500 1.17196900
C 2.18393200 1.55846200 0.03763600
H 1.64066800 1.99849900 -0.80557100
H 1.74521600 1.94381400 0.96443400
H 3.23879000 1.88678400 -0.01528400
N -2.49222300 0.28332900 0.00003400
O -3.20114000 -0.69674000 -0.02642700
O -2.85642800 1.43835100 0.00303400

$(CH_3)_3N{:}S(CH_3)Cl$
S 0.87956900 -0.16777500 0.00016300
C 0.91145700 1.64282200 0.00011900
H 1.41678200 2.00790400 -0.89651600
H -0.12839200 1.97686800 0.00022500
H 1.41697400 2.00800600 0.89660200
N -1.76933100 -0.11033200 0.00001300
C -2.32346100 0.47906900 -1.20211900
H -2.16415500 1.56306300 -1.20187000
H -1.82434900 0.05765200 -2.08130800
H -3.41068400 0.29409800 -1.29120500
C -2.32465500 0.48082000 1.20075400
H -1.82532900 0.06186400 2.08099900
H -2.16690400 1.56502700 1.19824600
H -3.41167000 0.29460700 1.28977700
C -1.94295900 -1.55032800 0.00102300
H -1.46576800 -1.98064100 -0.88576900
H -1.46700500 -1.97919300 0.88917700
H -3.01133100 -1.83480300 0.00049600
Cl 2.96657400 -0.40541300 -0.00001300

$(CH_3)_3N{:}S(CH_3)F$
S -1.16732900 -0.35113400 -0.01664900
C -1.36197500 1.44427700 -0.00572700
H -1.56682400 1.82009100 0.99977400
H -0.45504300 1.90399800 -0.40248800
H -2.19619200 1.67932500 -0.66986800
F -2.85441200 -0.62094200 0.00915400
N 1.29066700 -0.05914900 -0.00067900
C 1.74454300 0.70425900 1.14755100
H 1.43788800 1.75086600 1.04915100
H 1.29385700 0.29613200 2.05827100
H 2.84422700 0.67471100 1.25061200
C 1.79271500 0.48930700 -1.24713300
H 1.37716100 -0.07486600 -2.08825000
H 1.48909700 1.53684800 -1.35015900
H 2.89559100 0.44646000 -1.29880400
C 1.65524600 -1.45947600 0.13005100
H 1.23070900 -1.86493000 1.05398700
H 1.24781600 -2.02479800 -0.71455300
H 2.75084300 -1.59337600 0.15263300

Table A4. Cartesian coordinates of all optimized $(CH_3)_2O{:}S(CH_3)X$ complexes.

$(CH_3)_2O{:}S(CH_3)_2$
S -1.18514100 -0.32069500 0.17359600
C -1.14088300 1.21123300 -0.79116500
H -1.59885500 2.03233500 -0.23124400
H -0.08590700 1.43017100 -0.97548400
H -1.65829900 1.07886200 -1.74622500
C 2.29945100 -1.02070100 -0.64133400
H 1.70604500 -1.72666900 -0.03908800
H 2.06216500 -1.17199500 -1.69669700
H 3.36921100 -1.22849300 -0.47911400
C 2.23123000 0.58866800 1.03361500
H 1.65414200 -0.07980500 1.69195900
H 3.30133600 0.48585400 1.27443500
H 1.92070000 1.61976400 1.22102300
O 1.98612900 0.30656600 -0.31731100
C -2.98284200 -0.47766000 0.33471200
H -3.45150200 -0.57577900 -0.64937100
H -3.18487500 -1.38033100 0.91537000
H -3.40266500 0.38544000 0.86041900

$(CH_3)_2O{:}S(CH_3)H$
S 1.47353800 -0.60469800 -0.36908600
C 1.85413800 0.86888500 0.63011500
H 2.15412300 1.70866300 0.00039900
H 0.92293800 1.11353500 1.14480100
H 2.62634100 0.65254700 1.37002100
C -1.89968100 -1.02709000 0.37351400
H -1.53761600 -1.54954900 -0.52647900
H -1.60104700 -1.59641100 1.25644400
H -2.99894600 -0.97376600 0.33127000
C -1.63211300 1.04354300 -0.64204600
H -1.23315900 0.59300300 -1.56493200
H -2.71903100 1.18192200 -0.75459900
H -1.16404500 2.02079500 -0.49611400
O -1.33518200 0.25207700 0.47619400
H 2.72122900 -0.80422300 -0.83447500

$(CH_3)_2O{:}S(CH_3)NH_2$
S -1.11824400 -0.31716000 0.02571500
C -1.19084800 1.35017400 -0.67221200
H -1.51635900 2.06802300 0.08450700
H -0.17359100 1.58102300 -1.00200400
H -1.87162200 1.36857200 -1.52594600
C 2.21408400 -1.12970000 -0.46507500
H 1.70495800 -1.74313400 0.29524400
H 1.92348300 -1.48408700 -1.45635900
H 3.30236400 -1.24584400 -0.34199700
C 2.12219200 0.74574500 0.90172900
H 1.58466500 0.19837100 1.69220900
H 3.20255900 0.70771600 1.11266700
H 1.79612500 1.78912600 0.91099500
O 1.83373700 0.21536300 -0.36319500
N -2.78789300 -0.52760500 0.40575600
H -3.17381100 -1.31468100 -0.10134900
H -2.91408800 -0.67751300 1.39920200

$(CH_3)_2O{:}S(CH_3)SCH_3$
S -0.35903600 0.06409200 -0.10487700
C -0.22891400 -1.36384200 1.01778200
H -0.37868900 -2.29394800 0.46585100
H 0.78415700 -1.32217200 1.42872000
H -0.95845800 -1.28042500 1.82673500
C 2.85475300 1.38436100 0.22464800
H 2.27725800 1.76066800 -0.63488700
H 2.52954700 1.90849100 1.12599500
H 3.92212700 1.59213200 0.05058300
C 2.99705400 -0.74004900 -0.70401700
H 2.42294000 -0.43191600 -1.59213200
H 4.07163200 -0.63505700 -0.92032800
H 2.77901900 -1.78972200 -0.48968200
O 2.63127300 0.01388700 0.42061800
S -2.30843800 -0.09210200 -0.76018500
C -3.21845400 0.74047000 0.58045500
H -4.27374300 0.73487400 0.29171200
H -3.10126000 0.19852900 1.52198300
H -2.87176600 1.76997100 0.68827300

$(CH_3)_2O{:}S(CH_3)CF_3$
S -0.13515700 -0.12579000 -0.11845700
C -0.06386300 1.59993000 -0.67534900
H -0.41998000 2.27132300 0.10777700
H 0.99306600 1.78353500 -0.87754200
H -0.64107300 1.72892700 -1.59225500
C 3.13169000 -1.12971700 -0.60213600
H 2.53778300 -1.83313600 0.00314700
H 2.88919400 -1.27713200 -1.65665000
H 4.19990400 -1.34220500 -0.44304600
C 3.09103900 0.47847200 1.07656900
H 2.50047600 -0.17227900 1.74074500
H 4.15899100 0.34448900 1.30715500
H 2.81490000 1.51938400 1.26479500
O 2.82581800 0.19978100 -0.27294200
C -1.92131500 -0.21830800 0.10955800
F -2.60492100 -0.00178900 -1.01714100
F -2.22639100 -1.43803200 0.54480800
F -2.38008800 0.66306700 1.00265200

$(CH_3)_2O{:}S(CH_3)CN$
S -0.81999500 -0.22982000 -0.03602100
C -0.83972200 1.48273500 -0.67169500
H -1.21895000 2.16109900 0.09266800
H 0.20523200 1.70099000 -0.89979600
H -1.44523300 1.53472700 -1.57643600
C 2.48690900 -1.07626200 -0.61314000
H 1.95728200 -1.83005500 -0.00893000
H 2.23478200 -1.22347100 -1.66542600
H 3.57015900 -1.21465400 -0.47857100
C 2.36958000 0.50194300 1.09333900
H 1.85218000 -0.20611600 1.75989000
H 3.45000200 0.45372700 1.29577200
H 2.01252900 1.51324900 1.30554300
O 2.09347100 0.22287600 -0.25470300
C -2.48985600 -0.41290300 0.20724500
N -3.62961400 -0.56692500 0.38496400

$(CH_3)_2O{:}S(CH_3)OH$
S -1.06882800 -0.32049400 0.06834900
C -1.25090900 1.30872400 -0.68234200
H -1.64170500 2.01938500 0.05014200
H -0.24490900 1.60314400 -0.99578300
H -1.90458200 1.26473500 -1.55728800
C 2.17927900 -1.11416700 -0.47151700
H 1.76287400 -1.73455500 0.33796600
H 1.83412200 -1.50599500 -1.43065700
H 3.27779600 -1.16850000 -0.42864300
C 2.08593700 0.78549100 0.86495000
H 1.64780800 0.22070200 1.70237500
H 3.17924700 0.82006100 0.98816000
H 1.69332300 1.80547300 0.88475600
O 1.73389200 0.21223600 -0.36705900
O -2.66934300 -0.59005800 0.52748600
H -3.10496100 -1.05426300 -0.19457000

$(CH_3)_2O{:}S(CH_3)NO_2$
S -0.42084500 -0.08209700 -0.11269700
C -0.43452800 1.68979800 -0.42484200
H -0.78347800 2.23586600 0.45260000
H 0.61077300 1.93101900 -0.63675100
H -1.05467300 1.91745800 -1.29247100
C 2.77278600 -1.01374800 -0.75720800
H 2.24070300 -1.82864500 -0.24124400
H 2.49796600 -1.02208700 -1.81389600
H 3.85554400 -1.18216700 -0.66204100
C 2.71654000 0.33668100 1.13835800
H 2.21209800 -0.45364900 1.71596100
H 3.80108300 0.26223700 1.30596400
H 2.36674700 1.31102300 1.48990500
O 2.40868700 0.23676300 -0.22873800
N -2.18397600 -0.35154700 0.10638200
O -2.47146400 -1.50466900 0.33539600
O -2.94399800 0.58377300 0.01866900

$(CH_3)_2O{:}S(CH_3)Cl$
S -0.63947100 -0.15279800 -0.04578000
C -0.73685400 1.54582600 -0.65188100
H -1.13480400 2.20830200 0.11965200
H 0.29512600 1.81718800 -0.89229300
H -1.35062100 1.58993500 -1.55378700
Cl -2.66040100 -0.47056200 0.26055800
C 2.49993700 -1.09914400 -0.62442400
H 2.03942100 -1.84937400 0.03754700
H 2.16464200 -1.27590100 -1.64836000
H 3.59361500 -1.20582100 -0.57393700
C 2.46410000 0.51160800 1.05829500
H 2.00796900 -0.19435300 1.76932300
H 3.55685300 0.49124500 1.18406000
H 2.09691600 1.51795700 1.27555900
O 2.10326800 0.19942500 -0.26334000

$(CH_3)_2O{:}S(CH_3)F$
S -0.97242400 -0.31479500 0.09227700
C -1.30413400 1.26515000 -0.69360800
H -1.69844500 1.98366400 0.02861300
H -0.34377400 1.60955800 -1.08590400
H -2.00473200 1.12412400 -1.51947800
C 2.11026900 -1.07996400 -0.50763500
H 1.81112700 -1.73936800 0.32220100
H 1.71618400 -1.48232200 -1.44279800
H 3.20806300 -1.04461700 -0.55623700
C 2.02410500 0.80868500 0.85369000
H 1.74249400 0.20260800 1.72826800
H 3.11633600 0.93572800 0.84020700
H 1.54926900 1.78967100 0.93573000
O 1.57553800 0.21026500 -0.33823900
F -2.56927500 -0.66529800 0.45157800

Table A5. Cartesian coordinates of all optimized $H_2C{=}O{:}S(CH_3)X$ complexes.

$H_2C{=}O{:}S(CH_3)_2$
S 0.77206000 -0.32763800 -0.25916500
C 0.65851800 1.43022400 0.16224200
H -0.37193100 1.73315300 -0.03693000
H 0.88580100 1.58794400 1.22098500
H 1.34279600 2.01743700 -0.45738000
O -2.50280000 0.42125300 -0.19268200
C -2.70268600 -0.68190800 0.23195100
H -1.86999100 -1.34498400 0.54870100
H -3.72981100 -1.09351800 0.32931500
C 2.51327300 -0.59022800 0.16422000
H 2.69063100 -0.38961700 1.22512800
H 2.74964800 -1.63569400 -0.04540000
H 3.15766700 0.04893600 -0.44680900

$H_2C{=}O{:}S(CH_3)H$

S 1.17249400 -0.68524300 -0.09541600
C 1.42397000 1.10428100 0.12914100
H 0.46088400 1.56039100 -0.10911800
H 1.68403900 1.33728400 1.16286300
H 2.18642000 1.48195500 -0.55385200
O -1.86537100 0.64795400 -0.21065100
C -2.33225100 -0.39176900 0.16120600
H -1.69137200 -1.25307100 0.44577400
H -3.42939100 -0.54386000 0.24062000
H 2.40217200 -1.07752400 0.28349800

$H_2C{=}O{:}S(CH_3)SCH_3$
S -0.00968300 0.09142100 -0.45067800
C -0.22103600 1.64498200 0.47326200
H -1.30242100 1.78433400 0.56492100
H 0.22378400 1.56177800 1.46768500
H 0.22555300 2.47379300 -0.07894700
O -3.22918600 0.16836800 0.23472000
C -3.33322300 -0.99881300 -0.01818500
H -2.45471500 -1.61643100 -0.30120800
H -4.31271200 -1.51999100 0.01918200
S 2.03972200 -0.07435500 -0.57345700
C 2.44412900 -0.85780900 1.02005600
H 3.52993800 -0.99139800 1.03436500
H 2.15084500 -0.21245000 1.85152500
H 1.95337000 -1.82979200 1.10008200

$H_2C{=}O{:}S(CH_3)CF_3$
S -0.28558200 -0.05266600 -0.09115300
C -0.33796700 1.75468700 0.04843300
H -1.39952300 2.00358300 0.01737800
H 0.09009200 2.07788200 0.99854900
H 0.17716000 2.21913600 -0.79365900
O -3.26632300 0.28185500 -0.05904900
C -3.56380000 -0.87271800 0.06810700
H -2.79449700 -1.66549600 0.18478600
H -4.62129700 -1.20677400 0.07615200
C 1.50397700 -0.24926700 0.00280800
F 2.15170700 0.38111000 -0.97975800
F 2.02155300 0.20580600 1.14687300
F 1.78615100 -1.54655300 -0.08583400

$H_2C{=}O{:}S(CH_3)NH_2$
S 0.78816400 -0.28439100 -0.23348700
C 0.69901900 1.46855100 0.19727500
H -0.33161300 1.77181000 -0.01027200
H 0.92278700 1.61216600 1.25667600
H 1.39216800 2.04518600 -0.41895100
O -2.45549300 0.40134700 -0.22608400
C -2.68147000 -0.68702500 0.22268400
H -1.86471400 -1.36172400 0.55697000
H -3.71740300 -1.07361400 0.32623600
N 2.43762800 -0.59288800 0.15682000
H 2.51332200 -1.29836900 0.87905700
H 2.95009000 -0.89491000 -0.66273600

$H_2C{=}O{:}S(CH_3)OH$
S 0.65972700 -0.28750400 -0.13872500
C 0.89364500 1.47369700 0.15158200
H -0.09882300 1.92175200 0.05046400
H 1.27438500 1.63898500 1.16215300
H 1.56869400 1.90870900 -0.58981100
O -2.19284300 0.44092900 -0.22370800
C -2.64811200 -0.59155200 0.18222100
H -1.99518100 -1.43676700 0.48598000
H -3.74226200 -0.75027100 0.27444000
O 2.22615900 -0.83286100 0.16249800
H 2.69783000 -0.83975000 -0.67675800

$H_2C{=}O{:}S(CH_3)Cl$
S 0.23564100 -0.11074600 -0.10293300
C 0.33187300 1.67994600 0.08856400
H -0.70695700 2.01771900 0.03701500
H 0.76212700 1.93900100 1.05801000
H 0.90934400 2.12155200 -0.72599100
O -2.58655900 0.30718300 -0.10281000
C -2.98414100 -0.80483800 0.10837700
H -2.28529500 -1.64922000 0.28451400
H -4.06663100 -1.04103600 0.13791900
Cl 2.24842600 -0.54848200 0.02919400

$H_2C{=}O{:}S(CH_3)F$
S 0.57238300 -0.29916900 0.04386500
C 1.00202600 1.43758200 -0.05331300
H 1.50712100 1.65390200 -0.99707500
H 0.05150500 1.97522800 -0.00966800
H 1.62455800 1.72020700 0.79821000
F 2.13868300 -0.87921700 -0.01612400
O -1.99786400 0.48796000 0.07736200
C -2.62648100 -0.52459000 -0.06378000
H -3.73378500 -0.52537700 -0.06507100
H -2.12602700 -1.50593300 -0.19945800

Table A6. Cartesian coordinates of all optimized $DMU{:}S(CH_3)X$ complexes in π-direction.

$DMU{:}S(CH_3)_2$
S -1.60944800 -0.20818000 -0.29195200
C -1.76111900 0.09083400 1.48875200
H -2.36334700 -0.69424700 1.95520800
H -0.74653600 0.07131800 1.89529100
H -2.21338100 1.07010200 1.67347200
C 1.60876000 0.09292100 0.03104400
O 1.52631300 0.11225200 1.24917100
N 1.80462400 -1.07020200 -0.67054500
H 1.61205900 -1.05906200 -1.65932300
C 1.65852100 -2.33712100 0.01230400
H 2.00755300 -3.13643600 -0.64503500
H 2.27226300 -2.33638400 0.91500900
H 0.61844300 -2.53815900 0.30244600
C 1.11872800 2.46958600 -0.12438100
H 1.41921300 3.31174900 -0.75220500
H 0.02868100 2.49589600 0.01641500
H 1.59482400 2.57450700 0.85175900
N 1.56498300 1.23539100 -0.73207500
H 1.35938100 1.12177800 -1.71255500
C -3.36941600 -0.16864700 -0.72095200
H -3.45379600 -0.33265400 -1.79656700
H -3.91010900 -0.96097900 -0.19544300
H -3.80467800 0.80367000 -0.47285900

$DMU{:}S(CH_3)H$
S 2.86481600 0.16892900 -0.04773000
C 2.53281300 -1.61807700 0.06323000
H 2.99964000 -2.15304800 -0.76563500
H 2.87487300 -2.01944500 1.01871500
H 1.44647300 -1.70888600 -0.00293100
C -1.43862200 0.22071900 0.00112100
O -0.41530300 -0.44148500 -0.00992200
N -1.43850600 1.59008700 0.09838600
H -2.25831600 2.06882700 -0.24064600
C -0.17899200 2.30195900 0.00846700
H -0.35042400 3.35298900 0.25002400
H 0.52997400 1.88781100 0.72949300
H 0.27556700 2.22681800 -0.98711200
N -2.68942100 -0.34783900 -0.08199300
H -3.45922500 0.19165600 0.28196000
C -2.80853400 -1.78850600 -0.00777900
H -2.53846700 -2.18594200 0.97848500
H -3.83704600 -2.07226500 -0.24058300
H -2.14704200 -2.24517400 -0.74645700
H 4.20485400 0.09337300 0.06276300

$DMU{:}S(CH_3)NH_2$
S 2.39532000 0.07940400 -0.05573300
C 2.06126300 -1.69052900 0.08191400
H 2.52325900 -2.22588000 -0.75082500
H 2.43764700 -2.07282900 1.03356900
H 0.97261200 -1.78663900 0.03885200
C -1.89527200 0.23703200 -0.00799300
O -0.88199800 -0.43683600 -0.07557300
N -1.87320200 1.60436400 0.11705500
H -2.69682000 2.10109300 -0.18458300
C -0.60653600 2.29814700 -0.00231300
H -0.75822600 3.35210800 0.23970300
H 0.11142600 1.87554800 0.70480000
H -0.17412700 2.21314200 -1.00692600
N -3.15655800 -0.31266600 -0.05255200
H -3.90166600 0.22879500 0.35711300
C -3.29377900 -1.75280100 -0.00402500
H -3.00749300 -2.17362100 0.96789800
H -4.33079800 -2.01873100 -0.21979500
H -2.65416600 -2.20208900 -0.76595200
N 4.12112600 0.04707700 0.02977200
H 4.52676300 0.43618800 -0.81273700
H 4.44882700 0.57461900 0.82976500

$DMU{:}S(CH_3)SCH_3$
S 1.00673700 -0.00035400 0.01843100
C 1.07576600 -0.82637700 1.63842000
H 1.82859300 -0.35634300 2.27520100
H 0.07945100 -0.69121800 2.07101900
H 1.29226000 -1.88876000 1.50826700
C -2.30409200 0.09100100 0.00019400
O -2.07798000 0.00126600 1.19747500
N -2.44052100 1.30251500 -0.62654400
H -2.40702200 1.32885100 -1.63237300
C -2.10195700 2.51353600 0.08925000
H -2.39711400 3.37496300 -0.51340100
H -2.64855500 2.54571800 1.03395600
H -1.02920000 2.57979900 0.31260500
C -2.13544100 -2.31882600 -0.29699700
H -2.56926700 -3.08118700 -0.94777700
H -1.04844500 -2.47100000 -0.24122300
H -2.55232300 -2.43811300 0.70460600
N -2.48474900 -1.00885700 -0.80182000
H -2.42164800 -0.87472000 -1.79839900
S 2.88369300 -0.37585400 -0.74851900
C 3.87589200 0.93550500 0.03449900
H 3.87758600 0.81991500 1.12090800
H 4.89677600 0.81808600 -0.34123500
H 3.49173800 1.91855800 -0.24418800

$DMU{:}S(CH_3)CF_3$
S -0.78637700 -0.14284500 0.16263900
C -0.98204300 0.70169700 1.75700300
H -1.58546500 0.10027100 2.43867800
H 0.03692600 0.79541800 2.13758900
H -1.42190200 1.68964700 1.60997200
C 2.44626600 0.01350500 -0.01684800
O 2.04432800 0.12912600 1.13170300
N 2.68372300 -1.20638000 -0.59068600
H 3.28033000 -1.23997300 -1.40148900
C 2.55378600 -2.40591000 0.20838900
H 2.60036000 -3.27623900 -0.44899800
H 3.33439900 -2.48849300 0.97396600
H 1.58317800 -2.40194200 0.71074800
C 2.25846700 2.40523500 -0.37440300
H 2.62547900 3.15833200 -1.07456900
H 1.16677100 2.49137200 -0.29021800
H 2.69317100 2.60212900 0.60745600
N 2.67787300 1.09569300 -0.83085700
H 2.62908300 0.92879200 -1.82412200
C -2.52251600 -0.13595600 -0.32192000
F -3.03839700 1.09451000 -0.40435000
F -3.30623400 -0.81460600 0.52034800
F -2.62979600 -0.70250500 -1.52237300

$DMU{:}S(CH_3)CN$
S -1.44020700 -0.22561100 -0.01332900
C -1.77129300 0.61909000 1.57209600
H -2.42023600 0.00168900 2.19295500
H -0.78259400 0.72273800 2.02343100
H -2.21773100 1.59626700 1.38611400
C 1.78146600 0.04101000 -0.00318500
O 1.28809100 0.12097300 1.11333400
N 2.10297600 -1.15949100 -0.57603200
H 2.75559600 -1.15768700 -1.34330900
C 1.97299900 -2.37581000 0.19859400
H 2.13925200 -3.23138600 -0.45841400
H 2.68120100 -2.41809500 1.03426300
H 0.96148300 -2.44346900 0.60686800
C 1.54892600 2.43435900 -0.32741300
H 1.93302000 3.20889300 -0.99382200
H 0.45211000 2.48520700 -0.31010900
H 1.91572000 2.63047800 0.68203100
N 2.03588100 1.14591100 -0.77572700
H 2.09345600 1.00264700 -1.77191200
C -3.02991500 -0.27865600 -0.60548600
N -4.10825600 -0.32859800 -1.04097500

$DMU{:}S(CH_3)OH$
S -1.57965900 -0.43316200 -0.19221800
C -2.21142700 0.39290300 1.27725900
H -2.86677500 -0.26567100 1.85291300
H -1.32322600 0.63963200 1.86702600
H -2.73916200 1.30712700 0.99267000
C 1.51336400 0.20052800 0.08874700
O 0.97624600 0.14142800 1.18465700
N 2.09099200 -0.89215800 -0.50113000
H 2.76024400 -0.73412300 -1.23701700
C 2.16691200 -2.13753400 0.23109200
H 2.48710200 -2.92941900 -0.44908800
H 2.86126300 -2.08667400 1.07842700
H 1.17667100 -2.38950100 0.61827600
C 0.78101100 2.49250800 -0.22392400
H 1.07829500 3.37303800 -0.79748200
H -0.29507200 2.31494900 -0.36009700
H 0.96821900 2.68706300 0.83377000
N 1.58011900 1.36005600 -0.64835500
H 1.66831100 1.23759000 -1.64595100
O -3.01459900 -0.60724100 -1.06421800
H -3.39144300 -1.46263200 -0.83411400

$DMU{:}S(CH_3)NO_2$
S 1.09186500 -0.14526500 0.18545300
C 1.41449500 0.95922100 1.56711000
H 1.79030600 1.91956600 1.21107800
H 0.43320400 1.08045300 2.03412900
H 2.11972400 0.50195000 2.26173900
C -2.11012300 0.04509800 -0.00435300
O -1.62226400 0.27072000 1.09600700
N -2.35577100 1.03926400 -0.91688300
H -2.42360900 0.76449500 -1.88453000
C -1.86032100 2.37378100 -0.65118000
H -2.28747900 3.06374900 -1.38151700
H -2.17559200 2.68579300 0.34624600
H -0.76483600 2.43334300 -0.70036900
C -2.30979200 -2.32569000 0.51070000
H -2.42979400 -3.26076400 -0.03939100
H -1.31530500 -2.31493900 0.96290300
H -3.05234000 -2.28189100 1.31594500
N -2.43077200 -1.21865400 -0.41426300
H -3.06003400 -1.32092400 -1.19405400
N 2.75996900 -0.30919100 -0.45486500
O 2.83954600 -1.02263300 -1.43074900
O 3.66726500 0.27178900 0.09336500

$DMU{:}S(CH_3)Cl$
S -1.28711300 -0.00116900 -0.00048900
C -1.72228500 0.00528200 1.74924600
H -2.28597200 -0.89358300 2.00655900
H -0.75520300 0.00766800 2.26085400
H -2.28665000 0.90565500 1.99972100
C 1.90298300 0.00044300 0.03176400
O 1.38107900 0.00100600 1.14061700
N 2.22489300 -1.15878500 -0.61710300
H 2.49555500 -1.11928700 -1.58567000
C 1.80726700 -2.43370600 -0.07438100
H 2.32585700 -3.23256900 -0.60821700
H 2.07889300 -2.48799800 0.98179000
H 0.72433000 -2.58801300 -0.15932200
C 1.80327300 2.43419200 -0.07941900
H 2.32071700 3.23290300 -0.61461400
H 0.72009600 2.58653100 -0.16503000
H 2.07449000 2.49110700 0.97670800
N 2.22349400 1.15892400 -0.61934800
H 2.49082000 1.11770100 -1.58879800
Cl -3.19081700 -0.00280600 -0.80759900

$DMU{:}S(CH_3)F$
S 1.55648300 -0.37461100 -0.21081400
C 2.29832100 0.46344100 1.18906700
H 2.79328100 1.38259400 0.86611300
H 1.45981600 0.69984900 1.84960500
H 3.00052700 -0.19881000 1.69980700
C -1.49272000 0.15070000 0.10067600
O -0.77693700 0.09737000 1.09443600
N -1.73871300 1.32520800 -0.56131000
H -2.01947600 1.25288100 -1.52698600
C -0.97247600 2.49950500 -0.19409600
H -1.35396700 3.35812900 -0.74978400
H -1.09407900 2.69428900 0.87339900
H 0.09796400 2.37441400 -0.40372100
C -1.98614500 -2.22857000 0.26460700
H -2.32233300 -3.02333800 -0.40406300
H -0.94323200 -2.41514000 0.53020500
H -2.58376400 -2.25280900 1.18351000
N -2.08627400 -0.96036700 -0.42744700
H -2.85493100 -0.83060200 -1.06541800
F 2.98055400 -0.63233100 -1.05281700

Table A7. Cartesian coordinates of all optimized $DMU{:}S(CH_3)X$ complexes in n_o-direction.

$DMU{:}S(CH_3)_2$
S 2.55989800 0.15706000 -0.16542400
C 2.03639700 -1.53775700 0.19565000
H 2.47301200 -2.23375200 -0.52720800
H 2.33277900 -1.81859000 1.21100200
H 0.94711700 -1.55000100 0.11264100
C -1.99559300 0.22007100 -0.00829500
O -1.02810900 -0.51365700 -0.11421500
N -1.88709700 1.57849800 0.15797100
H -2.68447500 2.13348500 -0.11067000
C -0.58322100 2.20001600 0.03722800
H -0.66805200 3.25300000 0.31334000
H 0.12122200 1.71722900 0.71896500
H -0.17269500 2.12239200 -0.97702700
N -3.28928300 -0.24977500 -0.04703000
H -3.99278800 0.32330300 0.39210200
C -3.51261000 -1.67979800 -0.03597500
H -3.22769700 -2.14569600 0.91560600

H -4.56908100 -1.87636900 -0.23074900
H -2.92081500 -2.14376400 -0.82725000
C 4.34408400 -0.06752100 0.05932800
H 4.73476000 -0.79981700 -0.65425300
H 4.56722100 -0.39214400 1.08057800
H 4.82631600 0.89590000 -0.12077400

DMU:S(CH_3)H
S 2.86481600 0.16892900 -0.04773000
C 2.53281300 -1.61807700 0.06323000
H 2.99964000 -2.15304800 -0.76563500
H 2.87487300 -2.01944500 1.01871500
H 1.44647300 -1.70888600 -0.00293100
C -1.43862200 0.22071900 0.00112100
O -0.41530300 -0.44148500 -0.00992200
N -1.43850600 1.59008700 0.09838600
H -2.25831600 2.06882700 -0.24064600
C -0.17899200 2.30195900 0.00846700
H -0.35042400 3.35298900 0.25002400
H 0.52997400 1.88781100 0.72949300
H 0.27556700 2.22681800 -0.98711200
N -2.68942100 -0.34783900 -0.08199300
H -3.45922500 0.19165600 0.28196000
C -2.80853400 -1.78850600 -0.00777900
H -2.53846700 -2.18594200 0.97848500
H -3.83704600 -2.07226500 -0.24058300
H -2.14704200 -2.24517400 -0.74645700
H 4.20485400 0.09337300 0.06276300

DMU:S(CH_3)NH_2
S 2.39532000 0.07940400 -0.05573300
C 2.06126300 -1.69052900 0.08191400
H 2.52325900 -2.22588000 -0.75082500
H 2.43764700 -2.07282900 1.03356900
H 0.97261200 -1.78663900 0.03885200
C -1.89527200 0.23703200 -0.00799300
O -0.88199800 -0.43683600 -0.07557300
N -1.87320200 1.60436400 0.11705500
H -2.69682000 2.10109300 -0.18458300
C -0.60653600 2.29814700 -0.00231300
H -0.75822600 3.35210800 0.23970300
H 0.11142600 1.87554800 0.70480000
H -0.17412700 2.21314200 -1.00692600
N -3.15655800 -0.31266600 -0.05255200
H -3.90166600 0.22879500 0.35711300
C -3.29377900 -1.75280100 -0.00402500
H -3.00749300 -2.17362100 0.96789800
H -4.33079800 -2.01873100 -0.21979500
H -2.65416600 -2.20208900 -0.76595200
N 4.12112600 0.04707700 0.02977200
H 4.52676300 0.43618800 -0.81273700
H 4.44882700 0.57461900 0.82976500

DMU:S(CH_3)SCH_3
S -1.58277500 0.13795400 -0.08499500
C -1.24446800 -1.57037800 -0.61105100
H -1.79360900 -2.27793200 0.01487700
H -1.51453600 -1.69894500 -1.66087200
H -0.16724500 -1.70250800 -0.47452700
C 2.70184600 0.21106300 0.06475800
O 1.68323600 -0.45907200 0.03807000
N 2.69698100 1.58013700 -0.03420900
H 3.50223800 2.06382800 0.33149000
C 1.43309400 2.28818700 0.00809500
H 1.61208700 3.34181800 -0.21552900
H 0.75549400 1.88186200 -0.74657200
H 0.93704800 2.20287900 0.98294200
N 3.95150800 -0.34893100 0.19710100
H 4.73355900 0.19644700 -0.12982300
C 4.08518800 -1.78881200 0.13626700
H 3.85030400 -2.19422100 -0.85565200
H 5.10855700 -2.06134800 0.40256100
H 3.40532000 -2.24738800 0.85695200
S -3.61014500 0.29967600 -0.42873400
C -4.30396800 -0.43012500 1.08954300
H -3.97079900 0.13096500 1.96490100
H -5.39237800 -0.36279500 1.00022400
H -4.01478600 -1.48022000 1.17821700

DMU:S(CH_3)CF_3
S 1.30891200 0.11661000 -0.06124500
C 0.98972900 -1.66013300 0.11252100
H 1.39654500 -2.20511900 -0.74085300
H 1.40662100 -2.02970000 1.05074600
H -0.09903200 -1.73905600 0.12800700
C -2.95627500 0.22767700 0.00567300
O -1.90875800 -0.39730400 0.05425900
N -3.00571400 1.59661600 0.07293000
H -3.82549100 2.04556800 -0.30399300
C -1.76820100 2.34942200 0.04774700
H -1.99373600 3.40419900 0.21597900
H -1.11292500 2.00369000 0.85039300
H -1.22997800 2.24063800 -0.90176200
N -4.17815800 -0.38728900 -0.12034100
H -4.98686600 0.13365300 0.18024300
C -4.25466500 -1.83006300 -0.03255200
H -4.00389200 -2.20635400 0.96675300
H -5.26669700 -2.14715100 -0.29198500
H -3.55882900 -2.27591800 -0.74600100
C 3.10688900 0.03783500 -0.01856800
F 3.62517400 -0.69721900 -1.00739600
F 3.58352800 -0.47949600 1.11856300
F 3.58951900 1.27388200 -0.13546900

DMU:S(CH_3)CN
S 1.96032100 0.05734400 -0.13826000
C 1.78397200 -1.61082200 0.58436800
H 2.20515600 -2.35314700 -0.09354700
H 2.27172500 -1.64649400 1.55846700
H 0.70564600 -1.74232200 0.68351200
C -2.06895900 0.22543800 0.00651900
O -0.99402000 -0.34565400 0.11700500
N -2.20004400 1.58535800 0.11520900
H -3.02171300 2.00204100 -0.29352000
C -1.01352800 2.41170900 0.19841300
H -1.30824600 3.43071000 0.45629400
H -0.35843500 2.03103400 0.98486100
H -0.44493600 2.42792800 -0.73968000
N -3.23792000 -0.45051600 -0.23515100
H -4.10032700 0.01584000 -0.00233200
C -3.23950600 -1.89745500 -0.21850800
H -3.07706900 -2.30905300 0.78515600
H -4.19695500 -2.25601600 -0.60158800
H -2.44561900 -2.26791500 -0.86994000
C 3.65575300 0.10457700 -0.21099100
N 4.81673700 0.16722200 -0.27296700

DMU:S(CH_3)OH
S 2.21990300 0.05809500 -0.15168900
C 2.07878700 -1.62927700 0.45784700
H 2.49788000 -2.34445400 -0.25518800
H 2.57603100 -1.71686400 1.42691500
H 1.00441600 -1.80146500 0.56781600
C -1.78823800 0.22731900 -0.00401100
O -0.73176300 -0.38341300 -0.01949200
N -1.86216600 1.58738400 0.15500900
H -2.70279500 2.03886700 -0.16987300
C -0.64535800 2.37427100 0.14412400
H -0.87689600 3.38832200 0.47644300
H 0.07881200 1.93661900 0.83476000
H -0.17996200 2.41612100 -0.84847400
N -3.00381900 -0.40043100 -0.14477900
H -3.81130200 0.08284600 0.21648000
C -3.05257500 -1.84696800 -0.14381500
H -2.79669700 -2.27863200 0.83179600
H -4.05740200 -2.16830900 -0.42573700
H -2.34525600 -2.23413700 -0.87973300
O 3.90287400 0.22493400 -0.12494500
H 4.23203200 -0.04133800 -0.98917200

DMU:S(CH_3)NO_2
S 1.51333200 -0.02760900 -0.03889200
C 1.35079000 -1.81216200 0.09354300
H 1.82296800 -2.30155900 -0.75904100
H 1.78197100 -2.16660700 1.03047500
H 0.26905300 -1.97071300 0.08222000
C -2.44092900 0.25358500 0.00469400
O -1.39215600 -0.37613400 0.01128000
N -2.48658800 1.61824300 0.10237600
H -3.32665500 2.07606900 -0.21403300
C -1.25742500 2.38374400 0.05840500
H -1.48222400 3.42557400 0.29394600
H -0.56026500 2.00309800 0.80821100
H -0.76646800 2.33350700 -0.92103700
N -3.66239400 -0.36124600 -0.10084600
H -4.46941900 0.15529900 0.21105500
C -3.74032900 -1.80558900 -0.04854400
H -3.46836600 -2.20717500 0.93510200
H -4.75910200 -2.11319800 -0.29192800
H -3.06342600 -2.23533800 -0.78961100
N 3.29954700 0.10517400 -0.02402100
O 3.70389400 1.24546900 -0.10365400
O 3.97351300 -0.89557200 0.06059300

DMU:S(CH_3)Cl
S 1.76629700 0.00310900 -0.07661200
C 1.63139500 -1.74308900 0.34815300
H 2.07521000 -2.36227700 -0.43410600
H 2.09930800 -1.93944600 1.31498000
H 0.55256400 -1.91581500 0.40515400
C -2.17229700 0.24839200 0.00568300
O -1.13012300 -0.38633900 0.07071200
N -2.22266900 1.61390700 0.10781000
H -3.03830000 2.06992700 -0.26999800
C -0.99319100 2.37975500 0.13379400
H -1.22136800 3.40582100 0.42923400
H -0.31128200 1.94934700 0.87050400
H -0.48235400 2.38955300 -0.83706100
N -3.38973500 -0.36084200 -0.17345600
H -4.20812800 0.15353900 0.11154900
C -3.47095100 -1.80543700 -0.13237600
H -3.24378700 -2.21273400 0.86033200
H -4.47786200 -2.11089700 -0.42360300
H -2.76093100 -2.23150800 -0.84405100
Cl 3.83025000 0.15272200 -0.12910800

DMU:S(CH_3)F
S 2.06778900 0.07298500 -0.09855300
C 2.06944900 -1.68136300 0.26433400
H 2.56658500 -2.23253100 -0.53708600
H 2.54985700 -1.87153900 1.22671700
H 1.01212100 -1.95569500 0.31026900
C -1.69943500 0.22750800 0.00135200
O -0.61764500 -0.34310100 0.01605500
N -1.82970700 1.58542700 0.10813000
H -2.69013100 1.99108100 -0.22470500
C -0.65060100 2.42766400 0.09604000
H -0.93584700 3.43927200 0.39126600
H 0.07568000 2.04692100 0.81761300
H -0.16841000 2.46181000 -0.88855800
N -2.88110500 -0.45943400 -0.12261400
H -3.71686100 0.00097800 0.20181700
C -2.86716300 -1.90629700 -0.07901400
H -2.57669600 -2.29614400 0.90424100
H -3.86243100 -2.27709600 -0.33163700
H -2.15870400 -2.28707800 -0.81740000
F 3.73595100 0.25222800 -0.13322300

Table A8. Cartesian coordinates of all optimized Amide:S(CH_3)X complexes in π-direction.

Amide:$S(CH_3)_2$
S 1.50933300 -0.17834700 -0.29713800
C 1.69982900 0.24203600 1.45494200
H 2.12224700 1.24566500 1.56443300
H 0.69776500 0.20847900 1.89226200
H 2.34267600 -0.49055300 1.95168500
C -1.99459400 -1.79950500 -0.41150200
H -2.46928400 -1.79968600 -1.39678700
H -2.60154500 -2.37765700 0.28806900
H -1.01761900 -2.28856700 -0.48770200
C -1.78030600 -0.42247800 0.15374900
O -1.49879600 -0.24384900 1.32745700
N -1.90477900 0.60906600 -0.72262800
H -1.98111300 0.39116400 -1.70261800
C -1.56135100 1.95953000 -0.33602100
H -1.97950200 2.66233000 -1.05927000
H -1.98495100 2.16816200 0.64867400
H -0.47310200 2.09810900 -0.28257700
C 3.25411400 -0.13000000 -0.78331900
H 3.83051200 -0.87128900 -0.22174600
H 3.30761600 -0.36975500 -1.84747800
H 3.67464500 0.86698700 -0.62108200

Amide:S(CH_3)H
S -1.81176600 0.35600200 -0.72513800
C -2.36179000 -0.22435800 0.91233000
H -2.75685200 -1.24012400 0.85211700
H -1.46094800 -0.21562600 1.53228200
H -3.10492800 0.45299000 1.33557800
C 1.80582900 1.71585300 -0.21441400
H 2.48822400 1.65607700 -1.06685200
H 2.28720200 2.25686200 0.60279100
H 0.91679400 2.28336000 -0.50874400
C 1.35308900 0.37468000 0.29290800
O 0.77646300 0.25213100 1.36157500
N 1.61102800 -0.68842700 -0.51272800
H 1.96067000 -0.50994900 -1.43981500
C 1.08614300 -2.00050700 -0.20641000
H 1.61651000 -2.75079100 -0.79578400
H 1.23885900 -2.20932900 0.85469800
H 0.01104900 -2.06353100 -0.42103800
H -2.99684500 0.25195400 -1.35301500

Amide:S(CH_3)NH_2
S 1.51169200 -0.05223800 -0.29843800
C 1.86681100 -0.07964100 1.47985200
H 2.36985400 0.84118600 1.78954600
H 0.89014200 -0.15342700 1.97071400
H 2.47370500 -0.95180200 1.73959000
C -1.96808400 -1.76301500 -0.58739300
H -2.49921100 -1.65859200 -1.53753800
H -2.52595400 -2.43068200 0.07246400
H -0.99076800 -2.22002700 -0.77440400
C -1.73609600 -0.45847600 0.12296600
O -1.32838000 -0.41396200 1.27301700
N -1.98615200 0.66810100 -0.59170900
H -2.19535500 0.57850400 -1.57215600
C -1.65460600 1.97278300 -0.06190400
H -2.07706600 2.74144700 -0.71132800
H -2.07708900 2.08160700 0.93986100
H -0.56745000 2.11072900 0.00451100
N 3.04680400 0.05089900 -1.03919100
H 3.53916900 0.90182300 -0.79297100
H 3.62727600 -0.75616700 -0.84220500

Amide:S(CH_3)SCH_3
S -0.85672900 0.11434800 0.15325100
C -0.92291400 -1.35231900 1.23006100
H -1.74348400 -1.26455700 1.94577200
H -1.03647400 -2.25480600 0.62566100
H 0.03726200 -1.37175200 1.75525400
C 2.42417200 -1.49358200 -1.09963900
H 2.81153400 -1.18689100 -2.07517200
H 1.38751000 -1.82596200 -1.22082100
H 3.00413000 -2.33989800 -0.72566600
C 2.44336700 -0.39801400 -0.06926300
O 2.23720900 -0.62140200 1.11285900
N 2.69379900 0.85428600 -0.53177300
H 2.71956900 0.99943700 -1.52749600
C 2.60013300 2.00445800 0.34048300
H 3.06022300 2.86553400 -0.14779300
H 1.55875900 2.24392000 0.58955100
H 3.13204300 1.79823500 1.27175600
S -2.64085400 -0.01196600 -0.87167100
C -3.80625700 0.74379800 0.30665600
H -4.78892400 0.72454100 -0.17399500
H -3.84733900 0.16339500 1.23136000
H -3.51875200 1.77585700 0.51605600

Amide:S(CH_3)CF_3
S -0.60499800 -0.04357800 0.13940800
C -0.86578900 -1.21251300 1.50165400
H -1.34023100 -2.12496600 1.13719100
H 0.13978800 -1.43357200 1.86364100
H -1.45854600 -0.74647900 2.28997000
C -2.33651200 0.21336800 -0.28870100
F -3.06111000 0.68212100 0.73058500
F -2.94710300 -0.90046400 -0.70229000
F -2.39918300 1.10267300 -1.27822900
C 2.85407000 -1.35399700 -1.31886900
H 3.37879400 -0.89659200 -2.16206500
H 1.87239100 -1.70095700 -1.65752100
H 3.41032700 -2.22825200 -0.97386200
C 2.64481600 -0.42305600 -0.15762900
O 2.21640900 -0.81762200 0.91687900
N 2.92931200 0.88408900 -0.37273800
H 3.23002300 1.17522800 -1.28775000
C 2.69658600 1.88040000 0.65172200
H 3.05945100 2.84629500 0.29730700
H 1.62975900 1.96124800 0.88851100
H 3.22929800 1.61347400 1.56853200

Amide:S(CH_3)CN
S 1.30127700 -0.27138200 -0.02096800
C 1.57712300 0.49388700 1.61425400
H 1.97514100 1.50093000 1.48859800
H 0.58142100 0.52008100 2.06134600
H 2.25152500 -0.12621600 2.20439100
C -2.41244700 -1.79852000 -0.29265600
H -3.04980000 -1.82551000 -1.18044800
H -2.92710100 -2.27766400 0.54272300
H -1.50386900 -2.37646100 -0.49086000
C -2.00502100 -0.41237700 0.11894000
O -1.44897800 -0.19331800 1.18578600
N -2.25629000 0.58180100 -0.76603300
H -2.66766000 0.35512400 -1.65604600
C -1.82896000 1.93925300 -0.50286900
H -2.20914900 2.59321200 -1.28869100
H -2.21706000 2.27763600 0.46166700
H -0.73490400 2.01104400 -0.47605400
C 2.88802200 -0.18986500 -0.61564900
N 3.96494000 -0.14720100 -1.05532900

Amide:S(CH_3)OH

S 1.48485300 -0.30161900 -0.17450400
C 1.96306000 0.73241300 1.21926000
H 2.32210400 1.69948500 0.85818800
H 1.04581100 0.86345100 1.80137000
H 2.72344600 0.24316400 1.83305900
C -2.02097600 -1.83869100 -0.05010100
H -2.67422500 -1.96418200 -0.91795500
H -2.48659100 -2.29156100 0.82778300
H -1.07865600 -2.36666300 -0.23121000
C -1.69103100 -0.40549000 0.25910700
O -1.11925100 -0.08532300 1.29045400
N -2.03935700 0.51105500 -0.67708300
H -2.39757600 0.18756300 -1.56028400
C -1.62776300 1.89222400 -0.54638000
H -2.09062000 2.48057300 -1.34015300
H -1.94895600 2.28585500 0.42123900
H -0.53638000 1.98633000 -0.61300600
O 2.92672900 -0.30158800 -1.05017200
H 3.43993300 -1.06295500 -0.76094800

Amide:S(CH₃)NO₂
S -0.95514000 -0.10277200 -0.03902600
C -1.33954000 -1.58238400 -0.98443700
H -1.88594600 -1.32140200 -1.89125900
H -0.35759100 -1.99632400 -1.22895400
H -1.90978600 -2.28502000 -0.37618500
C 2.54656700 -0.74615300 1.74976800
H 3.17674400 -0.04407800 2.30198700
H 3.00111800 -1.73902900 1.77074700
H 1.57590300 -0.82053700 2.25037700
C 2.31028900 -0.35327600 0.31930600
O 1.74856500 -1.10010500 -0.47157200
N 2.72427700 0.88127900 -0.04725700
H 3.15012500 1.48053300 0.63982900
C 2.51243800 1.37539100 -1.39158200
H 3.00393200 2.34339800 -1.49738000
H 2.93190400 0.67817700 -2.12193300
H 1.44518600 1.49157200 -1.60695600
N -2.61898600 0.48439400 0.28471800
O -3.56210000 -0.14278100 -0.13655600
O -2.65707900 1.50987100 0.92857700

Amide:S(CH₃)Cl
S 1.15322800 -0.02653800 -0.00602500
C 1.60005500 -0.66341000 1.61944000
H 2.14882100 0.09019100 2.18739300
H 0.63691900 -0.87198000 2.09460500
H 2.17999600 -1.58327200 1.52409400
C -2.36753400 -1.58019100 -1.01303200
H -2.98025000 -1.28030700 -1.86749500
H -2.86653200 -2.38650000 -0.47105100
H -1.41340900 -1.97298800 -1.37883100
C -2.07392900 -0.46136500 -0.05445200
O -1.52070800 -0.65965500 1.01863000
N -2.42797300 0.78431600 -0.44891700
H -2.83478100 0.91264000 -1.36023000
C -2.12755500 1.94058100 0.36941900
H -2.55618000 2.82858800 -0.09727600
H -2.55609600 1.81887900 1.36805000
H -1.04522400 2.07456600 0.47616600
Cl 3.04766700 0.30400700 -0.75964100

Amide:S(CH₃)F
S 1.44152500 0.14495400 -0.24402700
C 2.22885300 -0.50301800 1.22727600
H 2.83579900 0.26725700 1.70747400
H 1.39893600 -0.78804800 1.87952600
H 2.82911100 -1.38168200 0.98194900
C -1.86375000 -1.77775600 -0.71089900
H -2.60436800 -1.65307400 -1.50537900
H -2.18435400 -2.57203400 -0.03315100
H -0.91120200 -2.08712300 -1.15357900
C -1.61069500 -0.53180400 0.08703900
O -0.89132800 -0.53749700 1.07886500
N -2.17928300 0.60877000 -0.36272400
H -2.75695500 0.57856700 -1.18599600
C -1.95314500 1.87237200 0.30899900
H -2.40069500 2.67379600 -0.28018300
H -2.39851400 1.86991600 1.30863300
H -0.87945400 2.05699900 0.41372900
F 2.84281300 0.48844300 -1.08832700

Table A9. Cartesian coordinates of all optimized Amide:S(CH₃)X complexes in n_o -direction.

Amide:S(CH₃)₂
S 2.05852800 0.24266200 -0.06937800
C 1.94566900 -1.55767000 0.08078400
H 2.38884600 -1.88918100 1.02497700
H 0.88131800 -1.80173100 0.06566500
H 2.45172600 -2.04316900 -0.75935900
C -3.54340900 -1.10111500 -0.02967300
H -4.44827400 -0.48695900 -0.02399900
H -3.55499200 -1.76295700 0.84002900
H -3.54436900 -1.73516700 -0.91976600
C -2.26455400 -0.30602200 -0.00749900
O -1.17243900 -0.84786300 -0.01886800
N -2.39958300 1.04349400 0.03756800
H -3.32224700 1.44162300 -0.00968700
C -1.24547300 1.91784600 0.04033100
H -1.55382300 2.91796500 0.35015500
H -0.49576400 1.53903700 0.74059900
H -0.77721400 1.97502900 -0.94876300
C 3.86243400 0.39793800 0.00992800
H 4.33541400 -0.13286800 -0.82237200
H 4.10760900 1.46003100 -0.06174100
H 4.24391200 0.00832400 0.95905200

Amide:S(CH₃)H
S -2.50017800 0.43956200 0.23850500
C -2.56022800 -1.29350500 -0.31691000
H -3.14185400 -1.90490000 0.37498300
H -2.96731200 -1.36664700 -1.32681800
H -1.52114400 -1.62757100 -0.31729700
C 2.96107500 -1.25051000 0.15693500
H 3.90674900 -0.70141700 0.14650600
H 2.95980600 -1.97879400 -0.65773900
H 2.88126400 -1.81013300 1.09241900
C 1.74437700 -0.37449100 0.01717400
O 0.61858000 -0.84205800 -0.01490700
N 1.97565200 0.95919900 -0.07994800
H 2.91941400 1.29625300 0.00971100
C 0.88869700 1.90929400 -0.18545700
H 1.27011900 2.85245300 -0.58107300
H 0.12547700 1.51619600 -0.86204000
H 0.41198900 2.09122900 0.78449400
H -3.82339500 0.67768000 0.16922000

Amide:S(CH₃)NH₂
S 2.04540300 0.18815700 -0.07124500
C 1.98573000 -1.60574300 0.13334100
H 2.50092900 -2.09642800 -0.69563000
H 2.44089600 -1.89043500 1.08470900
H 0.92345100 -1.86679500 0.12784500
C -3.48990600 -1.11723700 -0.06225000
H -4.40421400 -0.51725300 -0.05943800
H -3.50104600 -1.79063600 0.79847500
H -3.47065700 -1.73951400 -0.96043400
C -2.22441800 -0.30268400 -0.01441000
O -1.12374500 -0.82697100 -0.01384400
N -2.38109100 1.04441200 0.04092800
H -3.30906000 1.42869200 -0.01657700
C -1.24140600 1.93671000 0.06847400
H -1.56726300 2.92499900 0.39794400
H -0.48996700 1.55448500 0.76507800
H -0.76859500 2.02262500 -0.91653500
N 3.75647300 0.42078100 -0.02446800
H 4.08566100 0.83168200 -0.88958200
H 4.01570000 1.02122400 0.74866800

Amide:S(CH₃)SCH₃
S -1.23353200 0.10754000 -0.05688400
C -1.06680500 -1.60974800 -0.63373700
H -1.70092800 -2.27343800 -0.04112800
H -1.32644200 -1.67567800 -1.69186600
H -0.01337700 -1.86189100 -0.48117800
C 4.32697200 -1.06973300 0.24164900
H 5.22906100 -0.45165900 0.23992300
H 4.38036700 -1.78459600 -0.58318900
H 4.28930600 -1.64721600 1.16885800
C 3.04851400 -0.28507000 0.11278500
O 1.95981900 -0.83447300 0.07978500
N 3.17900500 1.06238000 0.02769200
H 4.09595300 1.46765900 0.11217500
C 2.02425900 1.93019400 -0.06891200
H 2.34543000 2.92019200 -0.39711900
H 1.31422200 1.52408200 -0.79452100
H 1.50364800 2.02077300 0.89115900
S -3.22856700 0.48868900 -0.41904800
C -4.01478800 -0.21462300 1.06601700
H -5.09000700 -0.03944400 0.96415800
H -3.83097500 -1.29010900 1.12565600
H -3.64319000 0.28467300 1.96305200

Amide:S(CH₃)CF₃
S 0.95013800 0.02223500 -0.12385200
C 0.84887200 -1.74372100 0.27675300
H 1.36731600 -2.33249500 -0.48158500
H 1.25915500 -1.93186500 1.27025000
H -0.21927400 -1.96527300 0.26253000
C -4.56341700 -1.03271900 -0.14512600
H -5.45862100 -0.40508200 -0.13152600
H -4.61347800 -1.74635100 0.68099600
H -4.54430100 -1.61102100 -1.07228700
C -3.27599500 -0.26140800 -0.03437200
O -2.19153000 -0.82189200 -0.02995000
N -3.39202000 1.08546600 0.06931700
H -4.30730100 1.49899300 0.00923900
C -2.23034100 1.94520500 0.15104000
H -2.53733400 2.93095400 0.50440300
H -1.50598800 1.52423600 0.85351300
H -1.73576200 2.05127900 -0.82113700
C 2.74282200 0.17130600 -0.04046000
F 3.23981100 -0.11850400 1.16641500
F 3.07918300 1.42800800 -0.32567600
F 3.37413100 -0.62775200 -0.90578400

Amide:S(CH₃)CN
S 1.63491700 0.07844200 -0.08647300
C 1.68332100 -1.71041400 0.27995900
H 2.21651000 -2.23490500 -0.51284900
H 2.15462400 -1.87644200 1.24862400
H 0.63336500 -2.00487800 0.30230200
C -3.58395200 -1.16612000 -0.17973500
H -4.53424700 -0.62572700 -0.18813900
H -3.58763100 -1.88701400 0.64137500
H -3.48614900 -1.73288600 -1.10919500
C -2.38037300 -0.27676000 -0.03176400
O -1.24796800 -0.73288900 0.02684200
N -2.61854800 1.05415800 0.03586300
H -3.56546600 1.38599300 -0.03776800
C -1.53726500 2.00790500 0.16811800
H -1.95630600 3.00143100 0.33305600
H -0.89974700 1.74292300 1.01661600
H -0.91354500 2.02842000 -0.73171600
C 3.30854100 0.36245400 -0.10308400
N 4.45026700 0.58994600 -0.12078200

Amide:S(CH₃)OH
S 1.90689100 0.18237500 -0.11695800
C 2.02526100 -1.56515700 0.29707700
H 2.59459100 -2.11010000 -0.46066600
H 2.48326600 -1.68502500 1.28191800
H 0.99206600 -1.92403400 0.31309400
C -3.26842600 -1.25454400 -0.09371500
H -4.23124600 -0.74974400 0.02375500
H -3.15666700 -2.00244100 0.69514600
H -3.25638500 -1.78639400 -1.04857400
C -2.08337800 -0.32809000 -0.04647000
O -0.93976400 -0.74722600 -0.12971400
N -2.35703200 0.99119000 0.09746600
H -3.31743500 1.29011300 0.11609600
C -1.30126300 1.98208400 0.13049500
H -1.71817000 2.93756500 0.45270600
H -0.52078700 1.67513400 0.83228300
H -0.83686900 2.10525400 -0.85409000
O 3.53895600 0.60880200 0.00252800
H 3.92991300 0.49497700 -0.86943800

Amide:S(CH₃)NO₂
S 1.20695400 -0.04601800 -0.03670200
C 1.26591000 -1.82943300 0.17158600
H 1.79418300 -2.29305100 -0.66214000
H 1.73551800 -2.08874100 1.12103900
H 0.21297300 -2.12252300 0.17050300
C -3.98851800 -1.13304500 -0.16056600
H -4.92699600 -0.57237200 -0.16467300
H -3.99981200 -1.84580900 0.66766500
H -3.91174800 -1.71070000 -1.08516000
C -2.76517700 -0.26819100 -0.03176300
O -1.64135900 -0.74980000 0.00021100
N -2.97513900 1.06549900 0.05058900
H -3.91741900 1.41452300 0.00087800
C -1.87504900 1.99937000 0.17333700
H -2.27559200 3.00181700 0.32951200
H -1.24128300 1.73214600 1.02396700
H -1.25347700 2.00069100 -0.72778400
N 2.96435300 0.30375100 -0.05300800
O 3.75690800 -0.60274700 0.06054500
O 3.22506300 1.48046500 -0.18390700

Amide:S(CH₃)Cl
S 1.44440400 0.01508300 -0.05480100
C 1.54753200 -1.77074900 0.16623600
H 2.07935600 -2.22974100 -0.66970000
H 2.03013000 -2.01085100 1.11577100
H 0.50349300 -2.09655300 0.17673800
C -3.73368500 -1.11095600 -0.09496600
H -4.66172100 -0.53292300 -0.09687400
H -3.73967400 -1.79558800 0.75678200
H -3.68796200 -1.72025400 -1.00105200
C -2.49152800 -0.26577300 -0.02050700
O -1.37742200 -0.76793800 -0.00544800
N -2.67795500 1.07371800 0.03733000
H -3.61458600 1.43690700 -0.01708400
C -1.56026500 1.99261900 0.09801500
H -1.93230100 2.99266400 0.32526000
H -0.86237200 1.68328300 0.88128300
H -1.01326800 2.02014900 -0.85048500
Cl 3.46948000 0.44518900 -0.05024800

Amide:S(CH₃)F
S 1.75271200 0.18094900 -0.07740700
C 2.09482500 -1.55052000 0.22757900
H 2.65654000 -1.97779500 -0.60603300
H 2.63759300 -1.67050400 1.16772800
H 1.11134300 -2.02218600 0.29807700
C -3.10665600 -1.32266800 -0.14750300
H -4.09448800 -0.85757300 -0.09311300
H -3.01746600 -2.07481300 0.63965100
H -3.00792200 -1.84066100 -1.10524500
C -1.96827700 -0.34935300 -0.01228600
O -0.80891900 -0.73481200 0.04635400
N -2.29771200 0.96106400 0.03843800
H -3.26581600 1.22046800 -0.05017500
C -1.29133300 1.99851400 0.13876600
H -1.77733800 2.94387000 0.38430100
H -0.57489400 1.74876000 0.92585300
H -0.73835300 2.11207300 -0.79963200
F 3.35682200 0.66870900 -0.15579300

Table A9. Cartesian coordinates of all optimized Ketone:S(CH₃)X complexes in π-direction.

Ketone:S(CH₃)₂
S 1.54991800 -0.41462800 -0.13029800
C 1.78135600 0.60046100 1.35040600
H 1.95097400 1.64607600 1.07455500
H 0.86007100 0.52432900 1.93260900
H 2.62339400 0.22834000 1.94099800
C -1.89672100 -1.88975200 -0.23341800
H -2.79584300 -2.02999800 -0.84300100
H -1.91542500 -2.55136300 0.63384600
H -1.03667200 -2.15128600 -0.86202900
C -1.75185500 -0.45894900 0.20569600
O -1.45683800 -0.16908800 1.34449400
C -1.62876300 1.99264900 -0.42459100
H -1.86043700 2.71625100 -1.21128800
H -2.17231300 2.27659900 0.48084800
H -0.55894200 2.06076800 -0.19997400
C -1.99085800 0.58719500 -0.85501300
H -3.05278800 0.51384500 -1.13597400
H -1.43673100 0.28284100 -1.75466000
C 3.13540900 -0.07718700 -0.93983200
H 3.96684400 -0.40755000 -0.31006300
H 3.15425200 -0.63829500 -1.87649300
H 3.23823200 0.98969500 -1.16005300

Ketone:S(CH₃)H
S -1.84137200 0.42657200 -0.67951200
C -2.34415600 -0.41347300 0.85730000
H -2.52777000 -1.47429600 0.67732900
H -1.49485200 -0.30667400 1.53762900
H -3.22611200 0.06102400 1.28985500
C 1.67022200 1.81935700 -0.14542800
H 2.66363600 1.86059600 -0.60420300
H 1.59830900 2.52770200 0.68125200
H 0.94482900 2.10441300 -0.91787600
C 1.33292000 0.43162300 0.32359500
O 0.76277200 0.23879100 1.37562800
C 1.12508000 -2.03408000 -0.23456100
H 1.50304200 -2.82508100 -0.88830400
H 1.34921700 -2.30190000 0.80165900
H 0.03534800 -1.99373900 -0.34071200
C 1.73498000 -0.69579100 -0.59422900
H 2.83513900 -0.73493100 -0.57119300
H 1.48271600 -0.39820800 -1.62217400
H -2.91799800 0.09978900 -1.41615200

Ketone:S(CH₃)NH₂
S 1.55124400 -0.35150000 -0.12532300
C 1.83202800 0.68555000 1.32708300
H 2.07301000 1.70626500 1.02055400
H 0.89165000 0.67657800 1.88667200

H 2.63637100 0.27132000 1.93836600
C -1.92667900 -1.89467900 -0.15838600
H -2.80815900 -2.03424100 -0.79289700
H -1.98997100 -2.52209100 0.73190500
H -1.05229100 -2.20330300 -0.74507600
C -1.74599200 -0.45356200 0.23050400
O -1.41439300 -0.13507800 1.35150900
C -1.57407100 1.97136300 -0.48527500
H -1.79679600 2.67287900 -1.29418800
H -2.09342000 2.30508300 0.41730400
H -0.49873000 2.01196700 -0.28063000
C -1.98515900 0.56306000 -0.85817600
H -3.05518000 0.50813800 -1.11110700
H -1.46246900 0.21261500 -1.76022200
N 3.07743200 -0.18326200 -0.90515000
H 3.53711100 -1.08118500 -0.99398200
H 2.97133000 0.23303100 -1.82205400

Ketone:S(CH$_3$)SCH$_3$
S -0.85917700 0.18593800 0.08520000
C -0.86171700 -0.81749000 1.60430600
H -0.95903200 -1.87548400 1.35229700
H 0.10547800 -0.63267200 2.08259400
H -1.67135700 -0.50422500 2.26726200
C 2.49345600 2.00402500 -0.35136300
H 3.28555000 2.16837300 -1.08903900
H 2.60421100 2.68680000 0.49238100
H 1.53844900 2.21427100 -0.84978400
C 2.46230900 0.57922700 0.12724700
O 2.22989200 0.30307300 1.28424400
C 2.43306200 -1.88901300 -0.44985700
H 2.68142800 -2.61726800 -1.22683500
H 3.01205600 -2.12528800 0.44724700
H 1.37318700 -2.00584200 -0.19937900
S -2.63635700 -0.35324500 -0.80647100
C -3.82546300 0.72354200 0.05551800
H -4.80420600 0.52505000 -0.39146100
H -3.86094100 0.48259800 1.12049900
H -3.56059600 1.77232900 -0.09074100
C 2.71132900 -0.47723700 -0.91953100
H 3.75675000 -0.35761800 -1.24318500
H 2.11056400 -0.21704200 -1.80339600

Ketone:S(CH$_3$)CF$_3$
S 0.61875000 0.19134500 0.19531700
C 0.73997500 -0.83581200 1.68606400
H 1.33989000 -0.33615700 2.44791800
H -0.29090700 -0.94853300 2.02979200
H 1.15174100 -1.81485400 1.43673200
C -2.56811300 -1.63061400 -1.15891500
H -3.32663000 -1.50137600 -1.93735900
H -2.66685500 -2.60781800 -0.68354800
H -1.58675900 -1.57638700 -1.64771700
C -2.62641200 -0.53879800 -0.12829200
O -2.41466400 -0.76200400 1.04486400
C -2.78495200 1.94047500 0.39650100
H -3.06445000 2.91192300 -0.02034100
H -3.41965200 1.73951300 1.26382000
H -1.75255200 2.01008100 0.75637600
C -2.92563600 0.84690500 -0.64172700
H -3.94390700 0.81532400 -1.05803400
H -2.27540700 1.02255200 -1.51211200
C 2.35507800 0.14389600 -0.29022000
F 3.16638000 0.67078800 0.62742200
F 2.49550200 0.84265200 -1.41306900
F 2.79513800 -1.09522500 -0.51879600

Ketone:S(CH$_3$)CN
S 1.32558700 -0.37333400 0.13902200
C 1.49985800 0.95743000 1.37841600
H 1.74931400 1.89495200 0.88100200
H 0.51526800 1.02100500 1.84669300
H 2.26026300 0.68348200 2.10925900
C -2.22470700 -1.91993800 -0.09955300
H -3.17125900 -2.07134700 -0.62803100
H -2.20574300 -2.49399200 0.82788700
H -1.42846700 -2.28935600 -0.75859400
C -1.96881000 -0.46640500 0.17974700
O -1.49816200 -0.09844500 1.23559000
C -1.87093100 1.92456000 -0.66284800
H -2.15015300 2.57806000 -1.49354700
H -2.34296100 2.30226600 0.24848400
H -0.78697000 1.99701500 -0.52468600
C -2.28601400 0.49453800 -0.93681600
H -3.36646500 0.41779900 -1.12943900
H -1.81269300 0.10041900 -1.84883500
C 2.86730300 -0.27776600 -0.56003800
N 3.91651200 -0.23627300 -1.06181500

Ketone:S(CH$_3$)OH
S 1.53654400 -0.38277700 -0.06626900
C 1.82710100 0.97593400 1.07739900
H 1.98005500 1.90404700 0.52075700
H 0.91221100 1.04772100 1.67487900
H 2.67947200 0.77024200 1.72939200
C -1.79777300 -1.94331900 -0.03365100
H -2.77060700 -2.15404300 -0.49045500
H -1.66187900 -2.53892200 0.87024400
H -1.03152200 -2.22594400 -0.76605800
C -1.64402500 -0.48102600 0.27742700
O -1.19296000 -0.10262300 1.33764000
C -1.70756700 1.92048500 -0.53839500
H -2.09513000 2.57416500 -1.32468200
H -2.11094300 2.25544500 0.42138700
H -0.62025700 2.04371300 -0.50037000
C -2.07119800 0.47561600 -0.80662700
H -3.15817200 0.34973900 -0.92649900
H -1.63867700 0.11709500 -1.75200900
O 2.90628200 -0.25717000 -1.03685600
H 3.58494400 -0.82660900 -0.65948100

Ketone:S(CH$_3$)NO$_2$
S -0.95796300 -0.08530500 -0.05953000
C -1.32056300 -1.52705200 -1.06862400
H -1.85972000 -1.23358100 -1.96976700
H -0.33364400 -1.92467600 -1.32279700
H -1.88863000 -2.25987400 -0.49498900
C 2.40313100 -0.82954300 1.79168400
H 3.33877300 -0.49158500 2.24784000
H 2.29928200 -1.91173800 1.88256000
H 1.58473900 -0.35164300 2.34557400
C 2.29936800 -0.41235200 0.35270700
O 1.81913900 -1.14589100 -0.48687200
C 2.56297800 1.38311800 -1.42086900
H 2.93257100 2.39667300 -1.59806900
H 3.08780800 0.70157200 -2.09608000
H 1.50249900 1.35459700 -1.68905700
C 2.76917400 0.98036300 0.02392800
H 3.82772900 1.04406100 0.31585400
H 2.25127600 1.66319500 0.71542300
N -2.62674600 0.47661600 0.28492700
O -3.56022500 -0.13925500 -0.17014400
O -2.67298600 1.46944200 0.97558300

Ketone:S(CH$_3$)Cl
S 1.15882500 -0.01475800 -0.00715600
C 1.57403400 -0.54691100 1.66361300
H 2.11363400 0.24113800 2.19197500
H 0.60258000 -0.72173600 2.13658000
H 2.15129500 -1.47298700 1.63928600
C -2.26380500 -1.70359200 -0.93401700
H -3.19137300 -1.63015700 -1.51018700
H -2.23831800 -2.63068700 -0.35941800
H -1.43491200 -1.71834500 -1.65335500
C -2.06835600 -0.52223800 -0.02678500
O -1.57513300 -0.64546400 1.07496200
C -2.13859900 1.99191000 0.31968500
H -2.47228700 2.93233200 -0.12751600
H -2.61933700 1.88454800 1.29609300
H -1.05922900 2.05776500 0.49127100
Cl 3.05603600 0.26679800 -0.75747900
C -2.47050600 0.82024900 -0.58019500
H -3.54764600 0.76835300 -0.79848800
H -1.99375700 0.91754400 -1.56809800

Ketone:S(CH$_3$)F
S 1.47992200 0.07927500 -0.27345100
C 2.18792000 -0.32201900 1.31684100
H 2.73396500 0.53038000 1.72575700
H 1.31955500 -0.54383200 1.94583500
H 2.82498300 -1.20578700 1.24409700
C -1.85932000 -1.84383400 -0.62951600
H -2.86263200 -1.90030100 -1.06329300
H -1.66937500 -2.70174300 0.01708400
H -1.14455400 -1.87083400 -1.46195600
C -1.64599000 -0.56373600 0.12630700
O -1.02237200 -0.53932800 1.16760000
C -1.86574400 1.96097700 0.24443000
H -2.29935300 2.83114600 -0.25594500
H -2.24570800 1.92793700 1.26936600
H -0.78139100 2.10062400 0.30155200
C -2.20447800 0.68688400 -0.50064900
H -3.29181200 0.54501700 -0.59020900
H -1.84441000 0.71413000 -1.54091800
F 2.89851500 0.34553800 -1.10015800

Table A10. Cartesian coordinates of all optimized Ketone:S(CH$_3$)X complexes in n_o -direction.

Ketone:S(CH$_3$)$_2$
S 2.08664500 0.24794800 -0.04331200
C 1.98458400 -1.55773800 0.04327600
H 2.49963800 -2.01155800 -0.80912500
H 2.42243900 -1.92105900 0.97823500
H 0.92310200 -1.81217200 0.01042600
C -3.49436200 -1.23678800 -0.00202200
H -4.21785400 -0.91846900 0.75580800
H -3.21409900 -2.27961200 0.15353700
H -3.99186700 -1.14401900 -0.97466200
C -2.27165200 -0.35883700 0.00489500
O -1.15643100 -0.82107100 -0.08068400
C -1.28024400 1.97293000 -0.05810400
H -1.51232400 3.03511800 0.05938400
H -0.51533400 1.69838800 0.67453000
H -0.83750600 1.82280000 -1.04753100
C -2.52226900 1.12448400 0.11360000
H -2.99047500 1.29130900 1.09580900
H -3.30554400 1.38611000 -0.61313300
C 3.88998500 0.40514600 0.04383200
H 4.13172800 1.46959700 0.00331700
H 4.27139200 -0.01040600 0.98192500
H 4.36557500 -0.09981100 -0.80293400

Ketone:S(CH$_3$)H
S 2.51410000 0.48112900 0.04424100
C 2.66975000 -1.33079800 -0.04391200
H 3.25160700 -1.63151400 -0.91680500
H 3.11405500 -1.73160000 0.86876800
H 1.64962800 -1.70458200 -0.14187100
C -2.92978100 -1.37891900 0.02190500
H -3.66680200 -1.09079000 0.77875100
H -2.59023500 -2.40166600 0.19162000
H -3.43606500 -1.32819800 -0.94927600
C -1.75818300 -0.43334200 0.01222000
O -0.61846300 -0.83486400 -0.05463800
C -0.91111900 1.95968900 -0.10275800
H -1.22120300 3.00648400 -0.03763000
H -0.14616000 1.77717000 0.65806400
H -0.43571200 1.80221700 -1.07581600
C -2.09646600 1.03537400 0.07997500
H -2.58607400 1.19902000 1.05217100
H -2.88592700 1.22882500 -0.66136500
H 3.82980100 0.74347500 0.15805300

Ketone:S(CH$_3$)NH$_2$
S 2.07699000 0.19265400 -0.01874500
C 2.05424700 -1.61390900 0.02468200
H 2.57116300 -2.01853300 -0.84851900
H 2.52330400 -1.97511500 0.94276400
H 0.99903800 -1.90140600 0.00335900
C -3.44251400 -1.25912200 -0.00522600
H -4.17877200 -0.94967900 0.74390700
H -3.14453900 -2.29556800 0.15951800
H -3.93254100 -1.18117900 -0.98294400
C -2.23634500 -0.35875900 0.00843800
O -1.11186800 -0.80130100 -0.05749000
C -1.29078300 1.99422900 -0.06235200
H -1.55186500 3.05276800 0.02382500
H -0.53961700 1.75776800 0.69721000
H -0.81931700 1.83285400 -1.03657200
C -2.51719300 1.12077600 0.09665300
H -3.00761000 1.28965000 1.06749000
H -3.29121400 1.35828000 -0.64804600
N 3.78488000 0.44772200 0.02982500
H 4.09386200 0.94525400 -0.79645000
H 4.04257400 0.97949900 0.85235500

Ketone:S(CH$_3$)SCH$_3$
S -1.23957900 0.03033700 -0.20100300
C -1.14533500 -1.78585200 -0.26838000
H -1.72917700 -2.22931700 0.54189100
H -1.50241800 -2.14354500 -1.23592000
H -0.08728100 -2.03012700 -0.13940000
C 4.32408400 -1.11454100 0.17991700
H 5.06236400 -0.79069300 -0.56114000
H 4.09755700 -2.17394400 0.05182300
H 4.77848300 -0.96396200 1.16630800
C 3.06442700 -0.29573000 0.08828200
O 1.96936600 -0.81118100 0.11757600
C 1.96265500 1.98903200 0.00850900
H 2.15931200 3.05934900 -0.09873400
H 1.28770200 1.67807900 -0.79443000
H 1.43059000 1.82875400 0.95127200
S -3.24899200 0.35809400 -0.53096700
C -3.94062500 0.14927100 1.14159500
H -5.01791000 0.32120900 1.05813900
H -3.76684200 -0.86606200 1.50656300
H -3.50335100 0.87898800 1.82590900
C 3.25165300 1.19591200 -0.03152200
H 3.80013500 1.37020600 -0.96957900
H 3.95187600 1.50707500 0.75780500

Ketone:S(CH$_3$)CF$_3$
S 0.90594800 -0.03202500 -0.04974400
C 0.90995200 -1.83687900 0.13376800
H 1.40990000 -2.30582800 -0.71522700
H 1.38401700 -2.12196900 1.07445500
H -0.14457100 -2.11595000 0.14655500
C -4.41576000 -1.19256200 -0.09450800
H -5.18109300 -0.88836300 0.62699400
H -4.14089800 -2.23661200 0.06134500
H -4.86050500 -1.08570200 -1.09090900
C -3.20107800 -0.31001000 -0.00289500
O -2.08032600 -0.77052300 -0.00183900
C -2.22207100 2.03398300 0.03422600
H -2.48173100 3.09442800 0.09415200
H -1.55225600 1.79655400 0.86580900
H -1.66128100 1.87000200 -0.89084100
C -3.46618900 1.17204900 0.07547200
H -4.04050600 1.34113100 0.99879800
H -4.16725500 1.42538000 -0.73339000
C 2.69218400 0.19834300 -0.02094500
F 2.95128500 1.49838300 -0.14686800
F 3.25863200 -0.22187800 1.11513600
F 3.32357000 -0.44161900 -1.01024900

Ketone:S(CH$_3$)CN
S 1.66393700 0.08852500 -0.02898900
C 1.72035600 -1.73448300 0.07886900
H 2.23695100 -2.14077800 -0.79064600
H 2.21145100 -2.03609900 1.00398500
H 0.67216900 -2.03699900 0.08085300
C -3.55181900 -1.28335400 -0.05973700
H -4.32219500 -1.02055700 0.67245700
H -3.21302900 -2.30840800 0.09609600
H -4.01775800 -1.20829100 -1.04931300
C -2.39135300 -0.32990100 0.00579000
O -1.24470400 -0.72307200 -0.01588900
C -1.55302600 2.06758800 -0.01420600
H -1.87339100 3.11138200 0.04147400
H -0.83873200 1.88544500 0.79415700
H -1.02039600 1.92306600 -0.95919300
C -2.74054700 1.13371000 0.08645500
H -3.28303700 1.27533100 1.03341000
H -3.49012000 1.33863100 -0.69186400
C 3.33681000 0.37495800 -0.01987700
N 4.47860100 0.60205000 -0.01489200

Ketone:S(CH$_3$)OH
S 1.91347000 0.19168800 0.01032700
C 2.10415700 -1.59779000 -0.05807400
H 2.66206900 -1.90084000 -0.94819300
H 2.60393000 -1.95229000 0.84665000
H 1.08817600 -1.99987900 -0.10595200
C -3.21131800 -1.37583500 0.08707100
H -3.88303400 -1.13433000 0.91735000
H -2.81854400 -2.38728300 0.19947700
H -3.80985400 -1.32427800 -0.82997200
C -2.08265000 -0.38466400 -0.00289100
O -0.93535000 -0.74353300 -0.15115400
C -1.31875600 2.03375100 -0.11031300
H -1.66276700 3.06927800 -0.03819000
H -0.53969500 1.87509700 0.64126400
H -0.85253500 1.89218100 -1.09007400
C -2.46878700 1.06952200 0.08988400
H -2.94386900 1.20991300 1.07247900
H -3.27920800 1.24104300 -0.63415500
O 3.53336200 0.63036500 0.20628200
H 3.89983100 0.77981500 -0.67100000

Ketone:S(CH$_3$)NO$_2$
S 1.21971900 -0.04955400 0.01403900
C 1.28841500 -1.84455000 0.06895600
H 1.78205600 -2.23462100 -0.82173800
H 1.79743700 -2.18105800 0.97270900
H 0.23771700 -2.14604600 0.08627000
C -3.99385500 -1.20042400 -0.10247100
H -4.77127900 -0.89357900 0.60472300
H -3.71762700 -2.24197200 0.06660200
H -4.42407000 -1.10251400 -1.10606300
C -2.78581200 -0.31205800 0.00296600
O -1.66426700 -0.77189000 0.04606200

C	-1.81095600	2.03434700	0.03983600
H	-2.07550100	3.09496600	0.05226700
H	-1.18941400	1.82729300	0.91590200
H	-1.19748400	1.84480200	-0.84620000
C	-3.05399500	1.16995800	0.03786500
H	-3.67589900	1.35887600	0.92561700
H	-3.71273200	1.40332800	-0.81164300
N	2.97753200	0.30408300	-0.02983600
O	3.77024500	-0.60871000	-0.01398700
O	3.23474600	1.48699600	-0.07421700

Ketone:$S(CH_3)Cl$

S	1.47401200	0.02083800	0.00359700
C	1.58269400	-1.77797000	0.06475500
H	2.07555900	-2.16288000	-0.83049200
H	2.10629300	-2.09952400	0.96739500
H	0.54022900	-2.10918500	0.09339300
C	-3.70362900	-1.22712900	-0.08263100
H	-4.47872600	-0.92994600	0.63130800
H	-3.40851000	-2.26317400	0.08855400
H	-4.14515500	-1.14072100	-1.08232900
C	-2.50841200	-0.31882300	0.00783900
O	-1.38039500	-0.76054600	0.04918800
C	-1.57196400	2.04325600	0.02032300
H	-1.85352100	3.09974300	0.02573100
H	-0.94365200	1.85145400	0.89507000
H	-0.95973600	1.85525800	-0.86692900
Cl	3.49588600	0.45044500	-0.04236000
C	-2.80115200	1.15935800	0.03075500
H	-3.42243400	1.34589600	0.91951200
H	-3.46665900	1.37431900	-0.81838400

Ketone:$S(CH_3)F$

S	1.79764200	0.17785000	-0.05882000
C	2.15432800	-1.56048500	0.18478600
H	2.74339900	-1.94580200	-0.65039500
H	2.67219500	-1.71526400	1.13391300
H	1.17882400	-2.05310300	0.20808000
C	-3.12153000	-1.39443600	-0.11532200
H	-3.90721600	-1.19127800	0.61971700
H	-2.74013500	-2.40915000	0.00564700
H	-3.58290500	-1.30202300	-1.10556600
C	-2.00242500	-0.39679100	-0.00513900
O	-0.84213800	-0.74860100	-0.02916900
C	-1.26805100	2.03425300	0.03221300
H	-1.62945900	3.05903900	0.15264600
H	-0.51569800	1.83986700	0.80239800
H	-0.76519400	1.95914700	-0.93672800
F	3.39334000	0.68377700	-0.09518100
C	-2.41144000	1.04637900	0.12653000
H	-2.93646700	1.13195600	1.09028100
H	-3.18786800	1.24229000	-0.62730200

Appendix B

Table B1. Cartesian coordinates of all the complexes used in this study at M06/6-31G++(2D,2P) level.

(CH3)2S:OH2
S 0.74528500 0.00015400 -0.66108300
C 0.67450500 1.37034200 0.52612900
H 1.54708000 1.34679200 1.18481200
H 0.68372000 2.29924800 -0.04768600
H -0.24857200 1.31653700 1.11108800
O -2.40371600 -0.00012100 0.09302300
H -1.61503900 -0.00074500 -0.47234500
H -3.15158400 -0.00355100 -0.50583900
C 0.67585600 -1.36998700 0.52624200
H -0.24653100 -1.31669800 1.11238100
H 0.68481900 -2.29893400 -0.04753100
H 1.54910800 -1.34627400 1.18404700

Cl(CH3)S:O(CH3)2
O -2.10323500 0.19995400 -0.26241100
C 0.73603800 1.54512700 -0.65239400
H -0.29649900 1.81539700 -0.89175000
H 1.34886600 1.58938500 -1.55495200
H 1.13414100 2.20844400 0.11838000
S 0.63988200 -0.15306000 -0.04532600
Cl 2.66101700 -0.47013900 0.26024100
C -2.46384500 0.50992000 1.05981200
H -2.09391300 1.51471700 1.27961300
H -3.55667900 0.49216600 1.18510400
H -2.00978000 -0.19887300 1.76935500
C -2.50110400 -1.09744800 -0.62624300
H -3.59483700 -1.20349800 -0.57535000
H -2.16656700 -1.27210300 -1.65077700
H -2.04078500 -1.84954800 0.03374100

(CH3)2S:NH3
S -0.87353800 -0.00016400 -0.64123700
C -0.67199900 -1.36532100 0.53492500
H -1.46904200 -1.34455300 1.28389200
H -0.73753000 -2.29828500 -0.02897500
H 0.30623100 -1.29840200 1.02224900
C -0.67299400 1.36514400 0.53490000
H 0.30451100 1.29815800 1.02369000
H -0.73742400 2.29799700 -0.02931800
H -1.47105000 1.34488500 1.28281000
N 2.55034700 0.00044400 0.03686200
H 3.11565500 -0.81290500 -0.17848300
H 3.11685100 0.81284600 -0.17890400
H 1.76594300 0.00082900 -0.61415400

Cl(CH3)S:N(CH3)3
C 0.91139900 1.64279200 -0.00000100
H -0.12849000 1.97673200 0.00026100
H 1.41656400 2.00796100 -0.89669000
H 1.41700100 2.00799600 0.89642700
S 0.87956400 -0.16778600 0.00008000
Cl 2.96654600 -0.40539500 -0.00002800
N -1.76929000 -0.11031100 0.00000100
C -1.94295800 -1.55031600 -0.00016600
H -1.46641300 -1.97979600 -0.88770600
H -1.46636600 -1.98001300 0.88724500
H -3.01133400 -1.83475800 -0.00017300
C -2.32397500 0.48006700 -1.20141600
H -2.16533100 1.56415500 -1.20007400
H -1.82477600 0.05983600 -2.08113000
H -3.41110200 0.29455800 -1.29046300
C -2.32405800 0.47982100 1.20150000
H -1.82512400 0.05918800 2.08117200
H -2.16510000 1.56386400 1.20052700
H -3.41124700 0.29456000 1.29029100

Table B2. Cartesian coordinates of all the monomers mentioned in figure 3.8 at M06/6-311++G(3DF,3PD) level.

[Fe(SCH3)4]0
C 2.35803800 -2.04518300 -0.49278500
S 1.48892000 -0.72120200 -1.38459200
C -2.96600800 -0.91035900 0.90900100
S -1.57894000 -1.57011700 -0.05269100
Fe -0.07179200 -0.08901800 -0.13919600
S -1.15694300 1.47451600 -1.00668900
S 0.54266100 0.19859800 1.84069600
H 2.80476300 -1.66933700 0.42557900
H -3.52860400 -0.18789400 0.32111000
H 3.13475800 -2.43510900 -1.14859700
H 1.65915100 -2.84400900 -0.25289700
H -2.60799500 -0.43904800 1.82233000
H -3.61329500 -1.74731900 1.16822800
C -0.06093100 2.90727700 -0.82209000
H 0.93812300 2.68003800 -1.19171700
H -0.47792600 3.72235200 -1.41234300
H -0.01116100 3.21543200 0.22190300
C 1.99668000 1.28213100 1.75182100
H 2.81854500 0.77748800 1.24713800
H 1.76368200 2.20488700 1.22253700
H 2.28870100 1.52506400 2.77258100

S(CH3)2
S -0.00001200 0.66660300 0.00000000
C -1.35765100 -0.51548100 -0.00001300
H -2.28798600 0.04842300 0.00015400
H -1.32680900 -1.14416300 -0.89044500
H -1.32665600 -1.14425800 0.89036100
C 1.35768800 -0.51547200 0.00001100
H 1.32658700 -1.14426900 -0.89033600
H 2.28801400 0.04842900 -0.00022200
H 1.32682300 -1.14408400 0.89049100

S2(CH3)2
S -0.89790500 -0.50437100 0.48497800
C -1.79236300 0.81026300 -0.38229700
H -2.78105200 0.87020600 0.07247200
H -1.28793100 1.76742000 -0.26111400
H -1.89168600 0.57073100 -1.43860000
S 0.89790500 -0.50437100 -0.48497800
C 1.79236300 0.81026300 0.38229800
H 2.78106100 0.87018800 -0.07245500
H 1.28794500 1.76742400 0.26109300
H 1.89166400 0.57074300 1.43860500

CH3CH2(CO)SCH3
C -1.51541600 -0.68075400 -0.00007200
H -1.45273900 -1.34064900 -0.87026100
H -1.45268700 -1.34087800 0.86993600
C -0.28350300 0.18826100 -0.00002100
S 1.21862200 -0.78866200 0.00005200
O -0.29846100 1.38563700 -0.00005300
C 2.45183300 0.52205500 -0.00001500
H 3.07121900 0.44510400 -0.88964000
H 1.91682400 1.47108500 -0.00018900
H 3.07106400 0.44535500 0.88974000
C -2.79944400 0.11719400 0.00006100
H -3.66279300 -0.54670000 -0.00000300
H -2.86097000 0.75971100 0.87729500
H -2.86101000 0.75993600 -0.87700300

Thiazole
C -0.01973800 1.19511500 0.00021000
C -1.28258500 -0.58357200 -0.00009400
C -0.07838500 -1.20926200 0.00019300
S 1.17634200 -0.04691800 -0.00009900
H 0.27105400 2.23666100 0.00020500
N -1.24298400 0.77912800 -0.00008500
H -2.23876500 -1.08787800 -0.00018700
H 0.13138200 -2.26568400 0.00031000

Appendix C

Table C1. Atomic coordinates of the partial energy minimized structure (based on PDB ID 1PVH) used to study S···O interaction in α-helix at M06/6-311G+(D) level (See Figure 4.6 **A**).

C -5.76416500 0.82099500 -0.73635600
S -4.65589300 -0.12684100 0.36192000
C 0.95680200 3.34817900 2.11379500
C 0.98420800 2.59571800 0.82509500
O 0.23354400 1.67953800 0.56315300
N 1.96566700 2.98107700 -0.04172000
C 1.99944000 2.41868300 -1.36920700
C 2.47223700 0.97671700 -1.42536100
O 1.93753500 0.16200900 -2.14161900
N 3.55925000 0.67875400 -0.65426900
C 3.99842400 -0.69375000 -0.56314100
C 3.00385700 -1.60099900 0.14381300
O 2.82886600 -2.75281200 -0.17345900
N 2.36425000 -0.99320300 1.18009000
C 1.38124100 -1.69272500 1.96537500
C 0.11774600 -2.07722200 1.20681100
O -0.47719200 -3.10258100 1.47024200
N -0.28278300 -1.16645600 0.29735000
C -1.37277700 -1.44208100 -0.60566400
C -0.94032600 -2.30087000 -1.77422900
C -2.01214100 -0.17328700 -1.10371300
S -2.82583800 0.85620900 0.16872000
H -0.05245300 3.61771000 2.39793300
H 1.36187300 2.72536700 2.90411600
H 1.54279100 4.25450200 2.07838400
H 1.01821100 2.46564600 -1.84567200
H 2.65370800 3.02973300 -1.98000300
H 3.92205300 1.35661800 -0.00504000
H 4.17226500 -1.07202000 -1.56414100
H 4.94704700 -0.75313600 -0.02405100
H 2.66095500 -0.07211600 1.45706600
H 1.08839600 -1.05977900 2.80007200
H 1.81716400 -2.58801200 2.39261800
H 0.28511600 -0.34817400 0.11252700
H -1.20183300 -1.82514800 -2.70500800
H 0.12535900 -2.45392600 -1.75835700
H -1.42073300 -3.26775200 -1.74380000
H -2.09860900 -2.01253200 -0.05848700
H -2.72407900 -0.42215300 -1.89429800
H -1.24300600 0.43228200 -1.57961700
H -6.38165200 1.52221900 -0.18946800
H -5.20862000 1.37990000 -1.47647600
H -6.42611900 0.14491000 -1.26361800
H 2.49587100 3.80939900 0.17507300

Table C2. Atomic coordinates of the partial energy minimized structure (based on PDB ID 4KT1) used to study S···O interaction in β-sheet at M06/6-311G+(d) level (See Figure 4.6 **A**).

N -3.16733400 -2.50521600 0.35365200
C -2.08128600 -1.62051200 0.00550700
C -0.75909200 -2.37578400 0.02211800
O -0.58349700 -3.33220500 0.74999400
C -1.99946300 -0.45564700 0.97896500
S -1.03445200 0.98910400 0.43466800
N 0.19472900 -1.88886100 -0.81040000
C 1.53543900 -2.40960500 -0.82993000
C 2.60261400 -1.47496000 -0.28517300
O 3.61815900 -1.88324300 0.23660900
N 2.31892700 -0.16377000 -0.46481800
C 3.15742900 0.89456000 0.01283900
C 2.35877100 2.17334600 0.15274200
O 1.47739500 2.45297600 -0.63144000
N 2.74790800 2.98891500 1.16164700
C -3.38299300 2.83387900 0.07104100
S -2.30616800 1.82214700 -0.98839300
H -3.94836100 2.21531600 0.77282800
H -2.79810900 3.57831400 0.61513500
H -4.08687400 3.34490100 -0.59250000
H 2.29169000 3.88419200 1.26271300
H 3.43661200 2.71846200 1.84569600
H 3.66469200 0.57356800 0.92695800
H 3.94295400 1.12108800 -0.71088600
H 1.49056000 0.10102200 -0.97628100
H 1.56141800 -3.32731600 -0.25073700
H 1.77924300 -2.67836300 -1.85457100
H -0.02137200 -1.08735100 -1.38775600
H -2.26301100 -1.20169100 -0.98021200
H -3.61917500 -2.17602800 1.19677500
H -2.77995100 -3.42086500 0.55431600
H -1.53850000 -0.79949700 1.90034900
H -3.00647000 -0.13512200 1.24724200

Table C3. Atomic coordinates of the partial energy minimized structure (based on PDB ID: 2FD6) used as Case 1 in Figures 7 and 8 **D**, at M06/6-311G+(D) level.

N 4.77034500 -2.18314600 1.09447800
C 4.30253500 -1.34063200 0.02277100
C 3.72912700 -0.07470300 0.64912100
O 3.19271300 -0.09137900 1.73490300
C 3.23949900 -2.01391200 -0.84396500
S 1.81922200 -2.77764800 0.02315200
N 3.80437200 1.05600000 -0.11336400
C 3.42158500 2.34695800 0.40697900
C 2.28204600 3.01727000 -0.33304400
O 2.34546100 4.14852400 -0.76847000
N 1.18823800 2.22631300 -0.41571800
C -0.02409400 2.62085500 -1.07254000
C -1.17548400 2.09375800 -0.25921100
O -1.20867700 0.92755600 0.09950700
N -2.13887900 2.95452400 0.09853400
C -3.19147500 2.50305500 0.97019900
C -3.87028700 1.25992900 0.42848000
O -4.29950000 1.20574200 -0.70316300
N -3.93572200 0.23320900 1.31028800
C -4.58758900 -1.00383600 0.96697100
C -3.84269900 -2.01399100 0.10812100
O -4.34353500 -3.09942800 -0.11106900
N -2.66003900 -1.60094100 -0.35481200
C -1.81785300 -2.40640300 -1.18552100
C -0.56256500 -1.66551600 -1.52599800
S 0.43901700 -1.22544400 -0.02285400
H 3.12833700 2.20898400 1.44970800
H -0.03771900 3.70584100 -1.17662800
H -2.79929200 2.31994000 1.96980400
H -5.47398700 -0.73290600 0.38546700
H -4.77551900 -1.52571600 1.90806300
H 5.13265300 -3.04080600 0.69209700
H 3.93100900 -2.43827600 1.61928500
H 5.13721900 -1.11143100 -0.68585700
H 3.72117800 -2.88052600 -1.31854400
H 2.85659400 -1.25574700 -1.53600900
H -2.34823300 -2.68944100 -2.10153100
H 0.06118300 -2.32764300 -2.14365600
H -0.88029300 -0.73581500 -2.02362800
H 4.26042100 1.02007700 -1.01515300
H 1.21957300 1.29527500 -0.01756200
H -2.10254000 3.90389700 -0.22467800
H -3.63195900 0.35583800 2.26596300
H -2.35057100 -0.67459900 -0.11257400
H -3.93012800 3.28820800 1.06567900
H -0.07415000 2.19655400 -2.08034700
H 4.26486400 3.02117300 0.41409900
H -1.54685000 -3.32024800 -0.68248000

Table C4. Atomic coordinates of the partial energy minimized structure (based on PDB ID 3CEL) used as Case 2 in Figures 7 and 8 **D**, at M06/6-311G+(D) level.

N -5.25217800 -0.26789100 0.35645600
C -3.94984600 -0.29633300 0.96200100
C -3.26081800 1.05309500 0.72057400
O -3.63915700 1.85528000 -0.10091900
C -3.13533900 -1.46362500 0.42960100
S -2.98197200 -1.51929100 -1.39623200
N -2.18532800 1.29669300 1.53353200
C -1.41614800 2.50392100 1.42602500
C -0.04777300 2.31690300 0.81364600
O 0.82625900 1.64938000 1.34139700
N 0.14200800 2.98637600 -0.34841300
C 1.43416300 3.03649200 -0.98213000
C 2.14986300 1.70589800 -1.11405200
O 1.63110000 0.70777900 -1.56713800
N 3.45722200 1.74969100 -0.73165500
C 4.28430800 0.57300400 -0.80839000
C 3.97336100 -0.52131700 0.18531700
O 4.49034100 -1.61356800 0.07789900
N 3.10459400 -0.16694400 1.15035800
C 2.55724800 -1.14747200 2.03959400
C 1.57498000 -2.11959700 1.41475400
O 1.13656600 -3.06019100 2.04938800
N 1.23645900 -1.80783700 0.15780200
C 0.40146800 -2.63022600 -0.65287700
C -0.02341500 -1.91291200 -1.88953900
S -1.18502600 -0.50663300 -1.68544200
H 2.10784100 3.70141800 -0.41390400
H 1.22984600 3.36515300 -2.01052200
H 4.23007900 0.15326700 -1.80510800
H 1.98083900 -0.61875200 2.80692600
H 3.40549400 -1.73003100 2.39567500
H -5.71622200 -1.15397700 0.51128500
H -5.78938100 0.44709400 0.82376900
H -4.02620000 -0.49121700 2.05210900
H -3.65386900 -2.39929300 0.74597000
H -2.11273900 -1.37445000 0.82689400
H -1.97302200 3.23364300 0.84649100
H -0.46861800 -2.90122000 -0.09520100
H 0.87480300 -1.43516200 -2.28374900
H -0.49107800 -2.65987700 -2.54470000
H -1.27997900 2.93936500 2.41206500
H 5.31811300 0.82574100 -0.62821900
H 0.91548100 -3.56011500 -0.92856200
H -1.91057700 0.62248200 2.23131900
H 1.66409800 -1.00082400 -0.26791100
H 2.71336100 0.75996800 1.16677000
H 3.86975900 2.59749600 -0.37050800
H -0.54869300 3.66704100 -0.66281500

Table C5. Atomic coordinates of the partial energy minimized structure (based on PDB ID 1HTR) used as Case 3 in Figures 7 and 8 **D**, at M06/6-311G+(D) level.

N -5.69747500 -0.35880900 -0.75416300
C -4.49873900 -0.33900300 0.05640300
C -3.47153000 0.48024400 -0.67934000
O -2.98679000 0.11280700 -1.72586500
C -3.94135700 -1.69736000 0.35932900
S -2.49707000 -1.53702700 1.49106000
N -3.19487300 1.67625800 -0.09588400
C -2.40999600 2.64739100 -0.79259300
C -1.08227800 2.95817700 -0.15656900
O -0.37268800 3.85758800 -0.55130300
N -0.74215300 2.16949100 0.90504800
C 0.57082900 2.34004100 1.44769200
C 1.67243900 1.92996100 0.47656600
O 1.45371300 1.26830500 -0.51208800
N 2.94373000 2.29344500 0.83155000
C 4.02589000 2.02580600 -0.09609500
C 4.28905300 0.54842700 -0.31876000
O 4.67954800 0.11286600 -1.37460700
N 4.10207900 -0.22577800 0.78783800
C 4.19565500 -1.65582900 0.69106600
C 3.17157700 -2.25236800 -0.24738500
O 3.43394500 -3.18221800 -0.97817600
N 1.93402400 -1.70038600 -0.15469700
C 0.94899000 -2.08833400 -1.12822300
C -0.42009700 -1.65250200 -0.72575200
S -1.03521900 -2.73066000 0.63141700
H -2.25953000 2.25991600 -1.80620700
H 0.69337400 3.38085300 1.74375900
H 3.83701200 2.49840700 -1.06194800
H 5.19215400 -1.92959600 0.36528000
H -5.92426500 -1.29471800 -1.03838300
H -6.49493600 0.03354700 -0.27896700
H -4.72725900 0.08889000 1.01841400
H -4.72239100 -2.25723700 0.88350600
H -3.60786700 -2.12176000 -0.59100500
H 0.97302900 -3.16187200 -1.18750700
H -0.32373200 -0.64527600 -0.31323200
H -1.08078100 -1.74171400 -1.60140900
H 1.71734900 -0.91942900 0.46972800
H 3.80231600 0.20281600 1.64176200
H 4.01998900 -2.11897600 1.66127400
H 4.95460900 2.46005100 0.27458300
H 3.12794600 2.82640200 1.66999200
H -1.37544100 1.44359700 1.23411800
H -2.98429700 3.55872300 -0.90384800
H 0.68205600 1.74670800 2.35039800
H -3.54763500 1.88470600 0.83041300
H 1.18317400 -1.71327600 -2.13992800

Appendix D

Table D1. Coordinates of leucine and 5-methylcytosine dimer

C	6.597738	-1.591613	0.035649
C	6.054611	-0.347594	-0.665523
C	4.969674	0.422797	0.115975
C	4.682363	1.763922	-0.526879
C	3.675544	-0.359892	0.220153
N	-3.901695	-0.927084	-0.157104
C	-4.238813	0.431815	-0.292085
N	-3.34903	1.419677	-0.05494
C	-2.095677	1.118816	0.327379
C	-1.717323	-0.315208	0.473768
C	-0.324657	-0.709649	0.894398
C	-2.673881	-1.289909	0.212021
O	-5.396224	0.713206	-0.630068
N	-1.20769	2.106631	0.566064
C	-4.924514	-1.921654	-0.464684
H	6.885772	0.335015	-0.862524
H	5.650143	-0.620677	-1.645702
H	5.353234	0.597861	1.127604
H	5.576915	2.386608	-0.588286
H	3.929953	2.322411	0.033569
H	4.304379	1.633447	-1.544406
H	2.928512	0.18812	0.797765
H	3.808808	-1.328492	0.703138
H	3.256041	-0.542036	-0.773055
H	5.824226	-2.344796	0.189816
H	-0.061527	-0.294427	1.871269
H	-0.227402	-1.792132	0.96623
H	0.433189	-0.368324	0.182864
H	-2.449854	-2.342439	0.298929
H	-0.2304	1.905685	0.662012
H	-5.788943	-1.775546	0.178887
H	-5.255597	-1.807033	-1.494568
H	-4.507708	-2.914104	-0.313368
H	-1.48545	3.04375	0.328444
H	7.39151	-2.055316	-0.550791
H	7.012519	-1.338791	1.013537

Table D2. Coordinates of leucine and cytosine dimer

C	6.89797	1.597854	-0.239753
C	6.289174	0.410519	0.504832
C	5.132061	-0.346856	-0.148102
C	4.988957	-1.710333	0.48323
C	3.829917	0.423954	-0.025535
C	-5.088716	2.01218	0.327118
N	-4.074843	0.952232	0.048137
C	-4.407129	-0.384821	0.280638
O	-5.537961	-0.65577	0.698589
N	-3.482585	-1.347342	0.039324
C	-2.272898	-1.014172	-0.414282
N	-1.398547	-2.00517	-0.638327
C	-1.910694	0.347318	-0.655952
C	-2.832219	1.286634	-0.410545
H	7.097542	-0.303516	0.691643
H	5.963369	0.737653	1.498509
H	5.362935	-0.476547	-1.211871
H	4.804254	-1.622659	1.557301
H	5.889465	-2.313106	0.354
H	4.154009	-2.268244	0.053865
H	3.010821	-0.091453	-0.531724
H	3.903287	1.423654	-0.456913
H	3.549926	0.537387	1.024852
H	6.168004	2.389602	-0.419279
H	-0.438466	-1.817547	-0.855888
H	-0.931602	0.618035	-1.0209
H	-2.642344	2.338971	-0.562007
H	-5.97342	1.830337	-0.276976
H	-5.374327	1.969736	1.37481
H	-4.653566	2.977941	0.087662
H	7.717268	2.030872	0.33407
H	7.294697	1.290985	-1.20961
H	-1.664872	-2.937653	-0.374072

Table D3. Coordinates of tyrosine and cytosine dimer

C	-5.543128	2.120504	0.093485
N	-4.47308	1.099471	-0.027059
C	-4.766316	-0.216148	0.300453
O	-5.911806	-0.507153	0.65052
N	-3.800999	-1.145619	0.200254
C	-2.573793	-0.802707	-0.194402
N	-1.647821	-1.761892	-0.259805
C	-2.245803	0.543514	-0.538072
C	-3.217195	1.459037	-0.439193
C	6.085751	2.313853	-0.408209
C	5.696683	1.29034	0.643324
C	4.531978	0.377155	0.32746
C	3.243423	0.889439	0.119468
C	4.700612	-1.008214	0.303751
C	2.173132	0.02976	-0.129024
C	3.641641	-1.853078	0.047171
C	2.375375	-1.333568	-0.153965
O	1.313954	-2.182016	-0.405872
H	6.566058	0.669502	0.870319
H	5.469649	1.819772	1.574412
H	3.071384	1.956671	0.14822
H	5.684121	-1.426398	0.47369
H	1.17839	0.42374	-0.285512
H	3.795401	-2.925638	0.022899
H	1.621669	-3.092208	-0.411502
H	-0.671189	-1.587927	-0.423864
H	-1.257447	0.81813	-0.872936
H	-3.06758	2.501859	-0.675831
H	5.260082	2.989082	-0.638434
H	6.922441	2.920893	-0.061569
H	6.381413	1.831195	-1.341081
H	-1.914916	-2.675873	0.06296
H	-6.37617	1.850845	-0.550669
H	-5.903303	2.150045	1.118757
H	-5.138047	3.086622	-0.194259

Table D4. Coordinates of phenylalanine and -5-methylcytosine dimer

C	6.423536	-1.950483	0.314229
C	6.045702	-0.897321	-0.727729
C	4.822461	-0.094876	-0.397983
C	3.582526	-0.700836	-0.269014
C	4.902805	1.28841	-0.287849
C	2.449194	0.060508	0.005177
C	3.778908	2.051317	-0.010262
C	2.551352	1.43769	0.132515
C	-5.467704	-1.980533	-0.444078
N	-4.448358	-0.965702	-0.153764
C	-4.713593	0.399008	-0.416784
O	-5.824655	0.727517	-0.881731
N	-3.787965	1.355958	-0.165407
C	-2.582505	1.024667	0.339707
N	-1.655235	1.976812	0.591713
C	-2.286261	-0.402727	0.616837
C	-3.263387	-1.346559	0.34223
C	-0.954078	-0.820173	1.17959
H	-6.377208	-1.757307	0.109143
H	-5.711551	-1.968731	-1.504379
H	-5.084031	-2.955442	-0.158156
H	-3.100946	-2.397799	0.525865
H	-0.701555	-0.276604	2.09335
H	-0.940553	-1.882873	1.418094
H	-0.160324	-0.642012	0.459052
H	-0.701457	1.723943	0.767071
H	-1.850508	2.908735	0.266893
H	1.671763	2.034495	0.337556
H	1.499253	-0.433282	0.097147
H	3.490993	-1.773896	-0.37853
H	5.862248	1.77505	-0.408242
H	3.86495	3.125832	0.080107
H	6.889301	-0.221084	-0.875569
H	5.890755	-1.396056	-1.689395
H	5.621113	-2.674317	0.460437
H	6.628401	-1.487598	1.280521
H	7.313757	-2.499205	0.005572

Table D5. Coordinates of phenylalanine and cytosine dimer

C	6.165099	2.005268	-0.387895
C	5.78607	0.97208	0.687317
C	4.580827	0.156088	0.371618
C	3.319039	0.742592	0.175248
C	4.70259	-1.211785	0.230407
C	2.214836	-0.053871	-0.120495
C	3.617507	-2.003982	-0.0611
C	2.351433	-1.412939	-0.23671
C	-5.474497	1.997073	0.191652
N	-4.367885	0.987422	-0.006738
C	-4.592375	-0.341067	0.358805
O	-5.699959	-0.664152	0.820875
N	-3.598098	-1.254293	0.184638
C	-2.435912	-0.892778	-0.334574
N	-1.492276	-1.849538	-0.467837
C	-2.165634	0.462813	-0.711167
C	-3.17127	1.368381	-0.539612
H	6.635724	0.309684	0.859216
H	5.625926	1.502434	1.631412
H	3.201865	1.81362	0.274293
H	5.674065	-1.671006	0.366545
H	1.24336	0.408934	-0.239996
H	3.728756	-3.075589	-0.154028
H	1.496214	-2.042183	-0.440043
H	5.357887	2.718905	-0.554014
H	-0.699973	-1.694615	-1.063677
H	-1.218342	0.765992	-1.129033
H	-3.064179	2.409408	-0.804566
H	-6.335455	1.699086	-0.399596
H	-5.757052	2.004996	1.240071
H	-5.106748	2.969983	-0.1196
H	6.374504	1.515526	-1.339676
H	7.0517	2.565093	-0.089403
H	-1.787042	-2.799422	-0.31631